MODELLING THE COVARIANCE DYNAMICS
OF MULTIVARIATE FINANCIAL TIME SERIES

********************

A Bayesian Perspective on Parameter Estimation,
Model Selection and Financial Risk Management

by

BRENT G. HUDSON

A thesis submitted in partial fulfillment of the requirements for the degree
of Doctor of Philosophy

Discipline of Business Analytics
University of Sydney Business School

Submitted: 16 May 2011
Accepted: 17 January 2012
Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any other degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Brent Hudson.
Preface

This thesis presents several multivariate GARCH models, some new to the literature, which are used to model the covariance dynamics of multivariate financial stock return data via Bayesian techniques. Models of this type are used extensively in the field of financial econometrics, and are commonly estimated using maximum likelihood techniques. This thesis argues that the use of Bayesian techniques can improve the properties of the resulting estimators over the standard maximum likelihood approach. Applications in Bayesian model selection and financial risk management are also presented, to further support the findings.

I am greatly indebted to my supervisor, Associate Professor Richard Gerlach, who introduced me to the field of Bayesian econometrics and encouraged my research over the years. I would also like to thank Dr Darfiana Nur for developing my interest in time series analysis in an undergraduate course at the University of Newcastle. My thanks also go to my former managers at Essential Energy, Paul Maclean and Lawrence Clark, for the support they have given in allowing appropriate time away from the office to complete this thesis.

I would also like to thank my family, in particular my parents Gary and Angela, for their love and support throughout my life and guiding me in my career path. Finally, I sincerely thank my wife Dee for her love, friendship and patience during many weekends of thesis writing.

Brent Hudson.
This thesis is dedicated to my mother, Angela, who lost her battle with ovarian cancer during the examination period of this thesis. She always took a keen interest in the progression of my work, and was elated when this thesis was eventually submitted.

Rest in peace.

ANGELA HUDSON
11/05/1950 - 18/06/2011
## Contents

<table>
<thead>
<tr>
<th>Summary</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Univariate Volatility Modelling</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Multivariate Volatility Modelling</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Model Selection</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Parameter Estimation and Inference</td>
<td>7</td>
</tr>
<tr>
<td>1.5 Applications</td>
<td>8</td>
</tr>
<tr>
<td>1.6 Chapter Summary</td>
<td>10</td>
</tr>
<tr>
<td><strong>2 Bayesian Estimation Techniques</strong></td>
<td>11</td>
</tr>
<tr>
<td>2.1 The Metropolis Algorithm</td>
<td>12</td>
</tr>
<tr>
<td>2.2 The Metropolis-Hastings Algorithm</td>
<td>13</td>
</tr>
<tr>
<td>2.3 Delayed Rejection</td>
<td>15</td>
</tr>
<tr>
<td>2.4 Markov Chain Simulation</td>
<td>17</td>
</tr>
<tr>
<td>2.5 Gibbs Sampling</td>
<td>18</td>
</tr>
<tr>
<td>2.6 Chapter Summary</td>
<td>23</td>
</tr>
<tr>
<td><strong>3 A Bayesian Approach to Relaxing Parameter Restrictions in Multivariate GARCH Models</strong></td>
<td>24</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>25</td>
</tr>
<tr>
<td>3.2 Multivariate GARCH Models</td>
<td>27</td>
</tr>
</tbody>
</table>
4.2.2 The ADC Model ........................................... 78
4.2.3 The AG-DCC Model ..................................... 80
4.3 The APVECH Model Family ................................. 83
  4.3.1 Unconditional Properties and the Likelihood Function .... 87
  4.3.2 Necessary and Sufficient Conditions for PD and CS ....... 90
4.4 Bayesian Estimation Procedures ............................ 92
  4.4.1 The Gibbs Sampler ..................................... 92
  4.4.2 Prior and Posterior Distributions ....................... 93
  4.4.3 Conditional Posterior Sampling ......................... 94
4.5 Bayesian Model Selection .................................. 96
  4.5.1 Marginal Likelihood Approximations .................... 98
  4.5.2 Reversible-Jump MCMC ................................. 101
  4.5.3 Congdon’s Method .................................... 102
4.6 Simulation Study ........................................... 104
  4.6.1 The Bivariate APVECH Model Family .................. 104
  4.6.2 The Results ........................................... 107
4.7 Empirical Study ............................................ 111
  4.7.1 Estimation Results .................................... 112
4.8 Chapter Summary ........................................... 116
4.9 Appendix .................................................... 117
  4.9.1 Derivation of $E(I_{it}I_{jt})$ and $E(I_{it}I_{jt}\epsilon_{it}\epsilon_{jt})$ .... 117
  4.9.2 Simulation and Empirical Study Tables .................. 121

5 Estimating Portfolio Value at Risk using a Skew-\textit{t} Copula-GARCH Model ........................................... 138
  5.1 Introduction .............................................. 139
  5.2 Distributional Notation ................................... 144
  5.3 Multivariate Skew-\textit{t} Distributions ...................... 145
5.3.1 Jones (2001) ...................................................... 145
5.3.2 Azzalini and Capitanio (2003) ................................ 146
5.3.3 Sahu et al. (2003) .............................................. 146
5.3.4 Demarta and McNeil (2005) .................................. 147
5.3.5 Bauwens and Laurent (2005) ............................... 148
5.4 Copula-GARCH Models ......................................... 150
  5.4.1 Copula Properties ........................................... 150
  5.4.2 Jondeau and Rockinger (2006) ............................ 153
  5.4.3 Serban et al. (2007) ......................................... 156
  5.4.4 Ausin and Lopes (2010) .................................... 157
5.5 The ASTC-DTGARCH Model .................................... 160
  5.5.1 The Marginal Distributions ................................ 161
  5.5.2 The Skew-t Copula ......................................... 162
  5.5.3 Parameter Restrictions and the Likelihood Function .... 165
5.6 Bayesian Estimation Procedures ............................... 168
  5.6.1 The Gibbs Sampler .......................................... 168
  5.6.2 Prior and Posterior Distributions .......................... 169
  5.6.3 Conditional Posterior Sampling ........................... 170
5.7 Model Performance: A Stock Portfolio Example ............. 172
  5.7.1 The Data .................................................. 172
  5.7.2 Value at Risk .............................................. 176
  5.7.3 VaR Back-Testing ......................................... 179
  5.7.4 Competing Models ......................................... 185
  5.7.5 Parameter Estimation Results ............................. 190
  5.7.6 VaR Back-Testing Results ................................. 201
5.8 Chapter Summary .................................................. 204
5.9 Appendix .......................................................... 205
  5.9.1 Properties of the Univariate Skew-t Distribution ....... 205
Summary

This thesis presents several multivariate GARCH models, some new to the literature, which are used to model the covariance dynamics of multivariate financial stock return data via Bayesian techniques. There are three main goals for this thesis. The first goal is to generalise and improve Bayesian parameter estimation techniques for such models, which are typically estimated in the literature using a large number of unnecessary model restrictions. The second goal is to develop a test which identifies, locates and quantifies various forms of asymmetry in stock returns. The third goal is to improve risk forecasting in stock portfolios by considering additional stock return properties such as skewness and excess kurtosis in a new Bayesian model.

This thesis is separated into the following components:

- Chapter 1 introduces the thesis, provides necessary background and explains in further detail the contributions of the following chapters to the literature.

- Chapter 2 reviews the Bayesian model estimation techniques to be used in this thesis. Models proposed in this thesis will utilise the techniques from this chapter, with competing models (where applicable) fitted via the classical maximum likelihood approach.

- Chapter 3 presents a new approach to the fitting of multivariate GARCH models to stock returns from a Bayesian perspective, by relaxing the usual parameter restrictions typically imposed. A new general set of restrictions is developed, which has the ability to be applied to a wide variety of models, both univariate and mul-
tivariate. An example is illustrated in a bivariate setting containing two major international stock market indices.

- Chapter 4 presents a new family of multivariate GARCH models that accounts for the asymmetric behaviour commonly observed in stock returns. An effective model selection technique is also presented which is used to identify the type of asymmetry present in the data with minimal computational expense. An application is illustrated in a bivariate setting that utilises four major international stock market indices.

- Chapter 5 examines the issue of financial risk management for stock portfolios and presents a new Bayesian copula-GARCH model that attempts to improve risk forecasting over existing model specifications. This new model has the advantage of incorporating the stock return properties of asymmetry, skewness and excess kurtosis in the one multivariate model in a parsimonious setting. Risk forecasts are generated in an application using an equally-weighted portfolio containing five stocks, with comparisons made between existing univariate and multivariate specifications.

- Chapter 6 provides concluding remarks, and offers suggestions for future research.
Chapter 1

Introduction

Volatility plays an important role in financial markets. Accurate measurement and prediction of volatility is crucial in the effectiveness of various market activities such as trading/hedging strategies, portfolio optimisation, derivative pricing and risk management. The family of Autoregressive Conditional Heteroscedastic (ARCH) models introduced by Engle (1982) laid the foundation for a new approach to modelling volatility (or variance) whereby it was assumed that, conditionally, it was time-varying. This allowed periods of both high and low volatility to be captured in the one model, and also served as an extension to the Autoregressive Moving Average (ARMA) time series models proposed by Box and Jenkins (1970) which assumed a constant variance over time. ARCH models are typically used to model the volatility of returns of stocks, rather than the volatility of raw prices. There are two main reasons for this. First, for average investors, return of a stock is a complete and scale-free summary of the investment opportunity. Second, return series are easier to handle than price series because the former have more attractive statistical properties (Tsay, 2002, pp. 2), such as stationarity. Hence stock returns is the data source for all models proposed in this thesis.
1.1 Univariate Volatility Modelling

Let $y = \{y_1, \ldots , y_n\}$ denote a financial time series (e.g. stock returns) of sample size $n$. Mathematically, the ARCH($q$) model assumes that the conditional variance at time $t$ (denoted $h_t$) depends on a linear function of the squares of the past $q$ observations as follows:

$$y_t = z_t \sqrt{h_t},$$
$$h_t = c + \sum_{i=1}^{q} a_i y_{t-i}^2$$  \hspace{1cm} (1.1)

where $z_t$ is a standard normal random variable (i.e. $N(0,1)$) and $\{c, a_1, \ldots , a_q\}$ are parameters. A major drawback with ARCH models is that in many applications, a large number of lags (i.e. a large $q$) is required to model the financial time series accurately and therefore a large number of parameters to be estimated. Bollerslev (1986) alleviates this issue by generalising the structure in (1.1) so that $h_t$ also depends on a linear function of the past $p$ conditional variances. This is called the Generalised ARCH (or GARCH($p,q$)) model which has the form:

$$y_t = z_t \sqrt{h_t},$$
$$h_t = c + \sum_{i=1}^{q} a_i y_{t-i}^2 + \sum_{j=1}^{p} b_j h_{t-j}$$  \hspace{1cm} (1.2)

with $\{b_1, \ldots , b_p\}$ an additional set of parameters. In fact, when $p = q = 1$, the conditional variance equation in (1.2) is equivalent to an ARCH($\infty$) model, but with different ARCH effects, i.e.

$$h_t = c + \sum_{k=1}^{\infty} a_1 b_1^{k-1} y_{t-k}^2$$
$$= c + a_1 y_{t-1}^2 + a_1 b_1 y_{t-2}^2 + a_1 b_1^2 y_{t-3}^2 + \ldots$$
The model in (1.2) therefore keeps all previous observation lags in the model in a conve-
nient parsimonious structure. It is now clear that to ensure each $h_t$ is strictly positive, 
a sufficient set of restrictions $c > 0$, $a_i \geq 0$, $i = 1 \ldots , q$ and $b_j \geq 0$, $j = 1 \ldots , p$ can 
be applied. In addition, for the GARCH process to remain stationary, the unconditional 
variance needs to be strictly positive and finite. This is given by the following expectation:

$$E(h_t) = \frac{c}{1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j}.$$ 

For this expression to be strictly positive and finite, we must have $\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j < 1$ 
which is additionally imposed during estimation\(^1\). In most applications of GARCH mod-
els, $p = q = 1$ is enough to capture the volatility behaviour present in financial time 
series, compared to much higher values of $q$ for ARCH models (e.g. see Bollerslev et al., 

Engle (1982) also recognises the fact that other distributions for each error term $z_t$ 


can be used, depending on the properties of the financial time series to be captured. 


For example, Bollerslev (1987) proposes a *standardised Student-t distribution* for $z_t$ for 
a GARCH(1,1) model, while Hansen (1994) extends this to a *skew Student-t distribu-
tion*. These papers attempt to capture the skewness and the degree of excess kurtosis of 
stock price indices, foreign exchange and interest rate data not typically accounted for 
in GARCH models with normal errors. Other examples of error distributions used with 
GARCH models to capture these properties include the *generalised error distribution* (see 
for e.g. Nelson, 1991), the *Gaussian mixture distribution* (see for e.g. Ausin and Galeano, 
2007) and the *asymmetric Laplace distribution* (see for e.g. Chen et al., 2011b). Another 
property often overlooked in financial time series is that their mean is not always zero, as

\(^1\)A special case of the GARCH model is formed when $\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j = 1$: this is called the Integrated 
GARCH (IGARCH) model, and while having an infinite unconditional variance, is still often used in 
applications in the literature (see for e.g. Nelson, 1990; Choudhry, 1995 and Conrad and Haag, 2006).
implied by the models in (1.1) and (1.2). However we can easily include an ARMA-type structure in a GARCH model to capture any autocorrelation in the observations, and generalise the GARCH model as follows:

\[ y_t = m_0 + \sum_{k=1}^{r} m_k y_{t-k} + \varepsilon_t - \sum_{l=1}^{s} w_l \varepsilon_{t-l}, \quad \varepsilon_t|\psi_{t-1} \sim D(0, h_t), \]  
\[ h_t = c + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} b_j h_{t-j} \]  

(1.3)

where \( \psi_{t-1} \) denotes the information set up to time \( t-1 \), \( D \) denotes any distribution with mean 0 and variance \( h_t \), and \( \varepsilon_t = z_t \sqrt{h_t} \) is a residual (error term) with \( z_t \sim D(0,1) \). Therefore we call (1.3) the ARMA\((r,s)\)-GARCH\((p,q)\) model, and examples of applications of this general model can be found in Nakatsuma and Tsurumi (1999) and Trivez and Catalan (2009).

Black (1976) discusses the concept of volatility asymmetry, whereby stock returns tend to exhibit higher volatility following a negative “shock” than a positive shock of the same magnitude. To account for this asymmetry, the GARCH model has been extended and adjusted in various ways in the literature. Examples of this can be found in Engle (2004) who wrote a detailed survey. Perhaps the most widely used asymmetric GARCH models are: the Exponential GARCH model of Nelson (1991) where the standardised residual \( z_t \) is weighted depending on its sign; the GJR-GARCH model of Glosten et al. (1993) who apply an additional weighting to the coefficient of \( \varepsilon_{t-i}^2 \) when \( \varepsilon_{t-i} \) is negative; and the family of threshold (or regime-switching) GARCH models (TGARCH, DTARCH and DTGARCH) whereby more than one GARCH model is specified for a time series under a variety of different conditions - examples of which can be found in Li and Li (1996), Brooks (2001), Chen et al. (2005) and Chen and So (2006). All of the above models are examples of threshold GARCH models, an example of which will be utilised in Chapter 5 to capture asymmetry, skewness and excess kurtosis in a portfolio of stock returns.
1.2 Multivariate Volatility Modelling

When investors hold a portfolio of stocks, the question of market correlation arises. It is unlikely that the stock returns in a portfolio would move independently of each other, in which case there would be no correlation and the return volatilities for each stock would easily be modelled using the GARCH models discussed earlier. However, market correlation does exist and can potentially change over time. To capture this behaviour, Bollerslev et al. (1988) propose the diagonal VECH (DVECH) model which is a direct extension of the univariate GARCH model to more than one dimension (i.e. a multivariate GARCH model) in that it has GARCH equations for each conditional variance in the portfolio, plus GARCH-type equations for the conditional covariances between each stock return in the portfolio. For example, the bivariate DVECH(1,1) model has the following covariance equations:

\[
\begin{align*}
    h_{11,t} &= c_{11} + a_{11} \varepsilon_{1,t-1}^2 + b_{11} h_{11,t-1}, \\
    h_{12,t} &= c_{12} + a_{12} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + b_{12} h_{12,t-1}, \\
    h_{22,t} &= c_{22} + a_{22} \varepsilon_{2,t-1}^2 + b_{22} h_{22,t-1},
\end{align*}
\]

where the residual vector \((\varepsilon_{1t}, \varepsilon_{2t})\) conditionally follows a bivariate distribution \(D\) as follows:

\[
\varepsilon_{1t}, \varepsilon_{2t} | \psi_{t-1} \sim D \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{pmatrix} \right).
\]

To alleviate some of the estimation concerns of the DVECH model and reduce the number of parameters for large dimensions, other multivariate GARCH models have since been proposed - see Bauwens et al. (2006) for an extensive review. An additional review of some existing multivariate GARCH models is carried out in Chapter 3 of this thesis, however with the focus on model estimation issues such as parsimony (parameter numbers) and typical parameter restrictions used to enforce covariance stationarity. An extension to the diagonal VECH model is also presented in Chapter 3, with an application to two
international stock return indices.

Analogous to the univariate case, multivariate GARCH models have also been extended and adjusted in the literature to account for a variety of different properties such as skewness and excess kurtosis in the multivariate error distribution, multivariate volatility and covariance asymmetry. Kroner and Ng (1998) present the Asymmetric Dynamic Covariance (ADC) model that nests a variety of existing multivariate GARCH models while at the same time introducing basic asymmetric terms in the conditional covariance equations. De Goeij and Marquering (2004) directly extend the DVECH model of Bollerslev et al. (1988) to have asymmetric effects as in Glosten et al. (1993), but generalise further to allow asymmetric effects in covariances when residuals are of opposite sign. The concept of multivariate asymmetry is discussed in Chapter 4 of this thesis, as well as a review of some multivariate GARCH models exhibiting asymmetric effects. In addition, the model of De Goeij and Marquering (2004) is extended to incorporate residual and volatility spill-over effects (both symmetric and asymmetric) from other markets and applied to four international stock return indices.

1.3 Model Selection

To assist in deciding the model order of a GARCH model or the number of econometric effects to include, a number of model selection techniques that exist in the literature (both classical and Bayesian approaches) can be applied. Classical techniques include basic inspection of the sample (partial) autocorrelation functions for comparison with what is expected from the competing models (see for e.g. Cryer, 1986 and Brockwell and Davis, 1996) and information criterion functions such as the AIC by Akaike (1974) and the BIC by Schwarz (1978) which choose a model based on a penalised function of the likelihood of the estimated competing models. Examples of applications of such criterion functions to GARCH models can be found in Mitchell and McKenzie (2003), Daouk and Guo (2004) and An et al. (2002) in the multivariate case. Various Bayesian
model selection techniques also exist in the literature. The most popular are perhaps the techniques involving the approximation of marginal likelihoods (see for e.g. Newton and Raftery, 1994 and Chib, 1995) and also model search methods such as the reversible-jump sampler by Green (1995). The technique by Congdon (2006) is a simplified approach to Bayesian model selection in that it does not involve approximating the marginal likelihood or "jumping" between models of various dimensions but approximates model probabilities directly with minimal computation. This technique is applied in Chapter 4 of this thesis to decide on the number of asymmetric effects to include for a proposed multivariate GARCH model on international stock return indices.

1.4 Parameter Estimation and Inference

The method of maximum likelihood is perhaps the most common approach to estimate parameters of both univariate and multivariate GARCH models in the literature. This is mainly due to the technique being straightforward to implement and is also readily available in most statistical software packages. However, there are a number of drawbacks from using this approach on GARCH models. For example, constrained optimisation algorithms are typically utilised so that the resulting parameter estimates satisfy their restrictions (e.g. \( a_1 + b_1 < 1 \) for a GARCH(1, 1) model). These algorithms can become unstable and diverge when the parameter values are close to their restriction boundaries (see for e.g. Silvapulle and Sen, 2004). In fact, many GARCH models fit to stock returns have \( a_1 + b_1 \approx 1 \). This can lead to problems in calculating standard errors and hence confidence intervals for parameter estimates. Often techniques like the bootstrap of Efron (1982) are applied to estimate these standard errors which can require extensive work, even for simple time series models. These problems can also be amplified when the complexity of the model grows, for example, in the case of multivariate GARCH models and/or the addition of asymmetric effects. Bayesian methods for parameter estimation, while still computationally expensive, enjoy many advantages over maximum likelihood
and other classical methods. For example, information known for a problem before data is collected can be incorporated into the analysis in the form of a prior distribution, which can also include any parameter restrictions required. In addition, Bayesian methods account for parameter uncertainty and Markov chain Monte Carlo (MCMC) methods can be applied to simulate from parameter posterior distributions - standard errors can then be easily estimated from these simulations. Examples of successful applications of such techniques to GARCH-type models can be found in Bauwens and Lubrano (1998), Nakatsuma (2000) and Choy et al. (2008). This thesis will demonstrate these advantages and others by fitting multivariate GARCH models to stock returns via Bayesian estimation. Comparisons with classical maximum likelihood techniques will also be sometimes drawn. A more detailed description of the Bayesian methods applied in this thesis is given in Chapter 2.

1.5 Applications

As mentioned earlier, there are various applications of volatility models for investors in financial markets. Some specific examples include:

- **Option pricing**: If an investor holds (for example), European call or put options then these can be priced using the standard Black-Scholes option pricing formula (see Black and Scholes, 1973) where the conditional variance, often calculated from a GARCH model, is an input;

- **Portfolio optimisation**: An investor may wish to optimise their portfolio of stocks by owning a larger number of stocks that tend to be less volatile than others. This is known as optimising via the minimum-variance principles of Markowitz (1952), and the optimal weighting attributed to each stock can be determined from the calculation of return and covariance forecasts from a multivariate GARCH model;

- **Risk management**: Volatility forecasts from univariate/multivariate GARCH mod-
els are crucial inputs into the calculation of risk measures such as: Value at Risk (VaR) which is the minimum amount an investor can expect to lose over a given time horizon at a specified confidence level; and Expected Shortfall (ES) which is the expected loss, given that a loss exceeds the VaR threshold. If the VaR (and therefore ES) is high, then the investor knows to set aside this amount of capital to cover such losses and/or take action on their current position in order to reduce the size of the risk measures and minimise the risk of potentially significant portfolio losses.

In Chapter 5, a new multivariate GARCH model is proposed. Its performance will be examined by how well it can estimate VaR on a stock portfolio with known weights. It will be shown that this proposed model can be conveniently represented by its set of marginal distributions and a function known as a copula, used to link the marginals together to define their dependence structure. The theory on copulas perhaps began with the paper by Sklar (1959). However, the applications to GARCH models have only appeared recently in the literature (see for e.g. Patton, 2004; Jondeau and Rockinger, 2006 and Lee and Long, 2009). The proposed copula-GARCH model will also be compared with several existing univariate and multivariate GARCH models in VaR estimation performance. This will be performed by fitting the models to historical stock return data, creating portfolio VaR forecasts and comparing them with actual stock returns over the same time period. While risk measure back-testing has been applied in the literature on GARCH models in general (see for e.g. Wong, 2008; McAleer and Da Veiga, 2008 and Rombouts and Verbeek, 2009), to my knowledge there are no studies that have compared the performance between univariate and multivariate GARCH models in estimating risk measures. Chapter 5 aims to fill this important gap.
1.6 Chapter Summary

This chapter has served as a review of volatility modelling in the literature. Both univariate and multivariate GARCH models were discussed, and the variety of model selection techniques reviewed. The fitting of such models to financial time series was briefly described for both classical and Bayesian approaches, and a series of volatility modelling applications also reviewed.

The following chapter details the Bayesian estimation techniques to be applied to the multivariate GARCH models considered in this thesis.
Chapter 2

Bayesian Estimation Techniques

This chapter reviews some Bayesian estimation techniques that will be used in this thesis. Bayesian estimation is often viewed as a generalisation of classical estimation because prior information can be incorporated into a Bayesian analysis and used in conjunction with the collected data to make statistical inference. Bayesian estimation uses a posterior distribution which is a function of the collected data (in the form of a likelihood function) and a prior distribution reflecting the prior information: Let \( y = \{y_1, \ldots, y_n\} \) denote a collected sample of data with sample size \( n \) and let \( \theta = \{\theta_1, \ldots, \theta_m\} \) denote the \( m \) parameters of the assumed model for \( y \). The (joint) posterior distribution of \( \theta \) given \( y \) can be formed using Bayes’ rule as follows:

\[
p(\theta | y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta) \tag{2.1}
\]

where \( p(\theta, y) \) denotes the joint distribution of \( \theta \) and \( y \), \( p(y|\theta) \) represents the likelihood function, \( p(\theta) \) is the prior distribution of \( \theta \) and \( p(y) \) is a factor to make the posterior either integrate to 1 (under the assumption that \( \theta \) is continuous) or sum to 1 (under the assumption that \( \theta \) is discrete). In straightforward models containing mostly linear parameters and standard distributions (such as Gaussian errors), \( p(\theta|y) \) is known to follow a standard distributional form, and inference can be performed using Monte Carlo sampling from this posterior. Standard techniques are available to perform these simula-
tions (e.g. see Devroye, 1986). Otherwise, alternative techniques are typically employed. This chapter discusses some of these alternative techniques in calculating the posterior distribution for a set of parameters to be estimated for stock returns.

2.1 The Metropolis Algorithm

This is a rejection sampling method, introduced by Metropolis et al. (1953), which works quite efficiently. Given a target distribution \( p(\theta | y) \) that can be computed up to a normalising constant, the Metropolis algorithm creates a sequence of random points \((\theta^{[1]}, \theta^{[2]}, \ldots)\) whose distributions converge to the target distribution (Gelman et al., 2000, pp. 323). The algorithm proceeds as follows:

1. Begin with a starting value \( \theta^{[1]} \) (this may be a guess, or a random draw from a particular distribution).

2. For \( j = 2, \ldots, J \),
   
   (a) Draw a random sample \( \theta^p \) (called a proposed value) from a symmetric\(^1\) proposal distribution \( g(\theta) \).

   (b) Draw a random sample \( u^{[j]} \) from Unif[0,1] and set the following acceptance probability:
   
   \[
   \alpha^{[j]} = \min\left\{1, \frac{p(\theta^p | y)}{p(\theta^{[j-1]} | y)} \right\}.
   \]

   (c) If \( u^{[j]} < \alpha^{[j]} \), set \( \theta^{[j]} = \theta^p \). Otherwise, set \( \theta^{[j]} = \theta^{[j-1]} \).

The resulting sample \( \theta^{[1]}, \ldots, \theta^{[J]} \) will then be a dependent sample from the target distribution \( p(\theta | y) \).

The only restriction required here is that the proposal distribution \( g \) must be symmetric. Hastings (1970) demonstrates an improvement on this technique and shows that \( g \) needs (almost) no restriction at all.

\(^1\)A distribution \( g \) is symmetric if and only if \( g(\theta_a | \theta_b) = g(\theta_b | \theta_a) \), \( \forall \theta_a \) and \( \theta_b \).
2.2 The Metropolis-Hastings Algorithm

This is a generalisation of the Metropolis algorithm and is used more often because of the flexible requirements of the proposal distribution $g$. Symmetry is not required any more and the acceptance probability is slightly modified. The Metropolis-Hastings (MH) algorithm proceeds as follows:

1. Begin with a starting value $\theta^{[1]}$ (this may be a guess, or a random draw from a particular distribution).

2. For $j = 2, \ldots, J$,
   
   (a) Draw a proposed value $\theta^p$ from the proposal distribution $g(\theta)$.

   (b) Draw a random sample $u^{[j]}$ from Unif[0, 1] and set the following acceptance probability:
   
   $$
   \alpha^{[j]} = \min \left\{ 1, \frac{p(\theta^p | y)g(\theta^{[j-1]})}{p(\theta^{[j-1]} | y)g(\theta^p)} \right\}.
   $$

   (2.2)

   (c) If $u^{[j]} < \alpha^{[j]}$, set $\theta^{[j]} = \theta^p$. Otherwise, set $\theta^{[j]} = \theta^{[j-1]}$.

Once again, the resulting sample $\theta^{[1]}, \ldots, \theta^{[J]}$ will then be a dependent sample from the target distribution $p(\theta | y)$. When compared to the Metropolis algorithm, it is clear that acceptance probability differs through the inclusion of an additional ratio of proposal density values.

An important statistic of both Metropolis and MH algorithms often used in determining the quality of the sample generated from the target distribution is the acceptance rate (AR). This is an indicator of how many values from the proposal distribution were accepted/rejected during the algorithm, given by the formula:

$$
AR = \frac{1}{J} \sum_{j=1}^{J} I(\theta^{[j]} = \theta^p)
$$

where $I(A)$ takes the value 1 if $A$ is true, and 0 otherwise. The AR is also a measure of the performance and suitability of the proposal distribution being considered. For
example, ARs close to 1 indicate that either the target distribution $p(\theta|y)$ is “close” to
the proposal distribution, or that the proposal variance is too low (i.e. the tails of the
target distribution are not being captured). Conversely, ARs close to 0 indicate that
many of the values in the sample are being repeated and/or the proposal variance is too
high. To determine if a proposal distribution generates a “good” sample from the target
distribution will depend on the problem at hand. However, Roberts et al. (1997) suggest
an optimal acceptance rate of 0.234 for random walk algorithms (i.e. proposal means set
to $\theta^{[j-1]}$).

Another important statistic often used to determine sample quality is the potential
scale reduction factor (or PSRF - see for e.g. Gelman and Rubin, 1992 and Gelman
et al., 2000). This is an indicator of how well a generated sample can converge to, and is
a reasonable sample from, the target distribution. The PSRF is determined by running
numerous replications of the Metropolis or MH algorithm with over-dispersed starting
values and checking if each chain converges to the same target distribution: For the set
of $R$ algorithm replications $\{\theta^{[r]}_1, \ldots, \theta^{[r]}_J\}_{r=1}^R$ of length $J$, the between ($B$) and within
($W$) sample variances are first calculated as follows:

$$B = \frac{J}{R-1} \sum_{r=1}^R (\bar{\theta}_r - \bar{\theta})^2,$$

where $\bar{\theta}_r = \frac{1}{J} \sum_{j=1}^J \theta^{[j]}_r$, $\bar{\theta} = \frac{1}{R} \sum_{r=1}^R \bar{\theta}_r$, 

$$W = \frac{1}{R} \sum_{r=1}^R s^2_r,$$

where $s^2_r = \frac{1}{J-1} \sum_{j=1}^J (\theta^{[j]}_r - \bar{\theta}_r)^2$.

The PSRF is then given by

$$\text{PSRF} = \sqrt{\frac{W(J-1) + B}{WJ}} \quad (2.3)$$

which declines to 1 as $J \rightarrow \infty$, as described in Gelman et al. (2000). If a PSRF is close to
1, then the second halves of each of the $R$ replications can together be treated as a sample
from the target distribution. To be “close” to 1 will depend on the problem at hand,
however Gelman et al. (2000) suggest that values below 1.2 are acceptable. Otherwise,
the algorithm may need to be modified; for example, by increasing the value of \( J \) or adjusting the proposal distribution.

### 2.3 Delayed Rejection

Delayed rejection is a strategy that improves the MH algorithm, in that the resulting estimates have, uniformly, a smaller asymptotic variance on a sweep by sweep basis (Mira, 2001). Proposed by Tierney and Mira (1999), delayed rejection can be applied when many proposed values are rejected (i.e. when \( \theta^{[j]} = \theta^{[j-1]} \) is set) under a Metropolis or MH algorithm. Delayed rejection aims to reduce the number of these rejected proposals as follows: instead of retaining the same position and advancing time, as in the regular MH algorithm, propose a 2nd new candidate. The acceptance probability of the new candidate has to be adjusted in order to preserve the stationary distribution (Mira, 2001). Once we arrive at this second stage, if the corresponding candidate is also rejected, then we could retain the starting position or move on to a third stage, and so on. This section will state the acceptance probability of the general \( i \)th stage of the delayed rejection process.

The following is adapted from Mira (2001): Firstly, rewrite the standard MH acceptance probability from equation (2.2) as follows: assume that \( f(\theta) \) is the target distribution and draw a proposed value \( \theta^{p_1} \) from the proposal distribution \( g_1(\theta|\theta^{[j-1]}) \). The acceptance probability is then

\[
\alpha_1(\theta^{[j-1]}, \theta^{p_1}) = \min \left\{ 1, \frac{f(\theta^{p_1})g_1(\theta^{[j-1]}|\theta^{p_1})}{f(\theta^{[j-1]})g_1(\theta^{p_1}|\theta^{[j-1]})} \right\} = \min \left\{ 1, \frac{N_1}{D_1} \right\}. \quad (2.4)
\]

If we reject our proposed value \( \theta^{p_1} \), then draw a new candidate \( \theta^{p_2} \) from a new proposal distribution \( g_2(\theta|\theta^{[j-1]}, \theta^{p_1}) \). It can be shown that, to maintain the same stationary distribution, the acceptance probability of the new candidate \( \theta^{p_2} \) is

\[
\alpha_2(\theta^{[j-1]}, \theta^{p_1}, \theta^{p_2}) = \min \left\{ 1, \frac{N_2}{D_2} \right\}
\]
where

\[
\frac{N_2}{D_2} = \frac{f(\theta^{p_2})g_1(\theta^{p_1} | \theta^{p_2})g_2(\theta^{j-1} | \theta^{p_2}, \theta^{p_1})[1 - \alpha_1(\theta^{p_2}, \theta^{p_1})]}{f(\theta^{j-1})g_1(\theta^{p_1} | \theta^{j-1})g_2(\theta^{p_2} | \theta^{j-1}, \theta^{p_1})[1 - \alpha_1(\theta^{j-1}, \theta^{p_1})]}. \quad (2.5)
\]

If this second stage is reached, then \(N_1 < D_1\) must hold. So in equation (2.5) above, we can set \(\alpha_1(\theta^{j-1}, \theta^{p_1}) = N_1/D_1\) in \(D_2\), and after some algebra, obtain

\[
\alpha_2(\theta^{j-1}, \theta^{p_1}, \theta^{p_2}) = \min \left\{ 1, \frac{N_2}{g_2(\theta^{p_2} | \theta^{j-1}, \theta^{p_1})(D_1 - N_1)} \right\}.
\]

To find the acceptance probability for the \(i\)th stage, we use a form similar to \(\alpha_2\): If the candidate \(\theta^{p_{i-1}}\) proposed at stage \(i-1\) is rejected, draw a proposed value \(\theta^{p_i}\) from the proposal distribution \(g_i(\theta | \theta^{j-1}, \theta^{p_1}, \ldots, \theta^{p_{i-1}})\) and accept with probability

\[
\alpha_i(\theta^{j-1}, \theta^{p_1}, \ldots, \theta^{p_i}) = \min \left\{ 1, \frac{N_i}{D_i} \right\}
\]

where

\[
\frac{N_i}{D_i} = \frac{f(\theta^{p_i})g_1(\theta^{p_{i-1}} | \theta^{p_i})g_2(\theta^{p_{i-2}} | \theta^{p_i}, \theta^{p_{i-2}}) \ldots g_i(\theta^{j-1} | \theta^{p_i}, \theta^{p_{i-1}}, \ldots, \theta^{p_1})}{f(\theta^{j-1})g_1(\theta^{p_1} | \theta^{j-1})g_2(\theta^{p_2} | \theta^{j-1}, \theta^{p_1}) \ldots g_i(\theta^{p_i} | \theta^{j-1}, \theta^{p_1}, \ldots, \theta^{p_{i-1}})} \times \frac{[1 - \alpha_1(\theta^{p_i}, \theta^{p_{i-1}})][1 - \alpha_2(\theta^{p_i}, \theta^{p_{i-1}}, \theta^{p_{i-2}})] \ldots [1 - \alpha_{i-1}(\theta^{p_i}, \ldots, \theta^{p_1})]}{[1 - \alpha_1(\theta^{j-1}, \theta^{p_1})][1 - \alpha_2(\theta^{j-1}, \theta^{p_1}, \theta^{p_2})] \ldots [1 - \alpha_{i-1}(\theta^{j-1}, \theta^{p_1}, \ldots, \theta^{p_{i-1}})]}.
\]

Now if the \(i\)th stage is reached, then \(N_k < D_k\) must hold for \(k = 1, \ldots, i-1\), so \(\alpha_k(\theta^{j-1}, \theta^{p_1}, \ldots, \theta^{p_k})\) can be written as \(N_k/D_k\) (for each \(k\)) and obtain the recursive formula

\[
D_i = g_i(\theta^{p_i} | \theta^{j-1}, \theta^{p_1}, \ldots, \theta^{p_{i-1}})(D_{i-1} - N_{i-1})
\]

which leads to

\[
D_i = g_1(\theta^{p_1} | \theta^{j-1}, \theta^{p_1}, \ldots)[g_{i-1}(\theta^{p_{i-1}} | \theta^{j-1}, \theta^{p_1}, \ldots)[g_{i-2}(\theta^{p_{i-2}} | \theta^{j-1}, \theta^{p_1}, \ldots) \ldots [g_2(\theta^{p_2} | \theta^{j-1}, \theta^{p_1})]g_1(\theta^{p_1} | \theta^{j-1})f(\theta^{j-1}) - N_1] - N_2] - N_3] \ldots - N_{i-1}].
\]

An advantage of using delayed rejection is that it preserves the flexibility of the MH
algorithm in terms of allowing any proposal distribution \( g_1 \) to be considered, with the obvious restriction that the \( k \)th proposal \((g_k)\) will depend on all proposals before it (i.e. between \( k - 1 \) and 1). However, care still needs to be taken with the selection of \( g_1 \) since a “poor” \( g_1 \) may lead to a large number of rejection stages and a computationally expensive algorithm. In most practical applications of delayed rejection, only a small number of rejection stages are typically considered. In this thesis, it will be shown that \( i = 2 \) works well for the models proposed.

2.4 Markov Chain Simulation

The techniques discussed in Sections 2.1-2.3 all possess one common attribute in that they are all designed to produce dependent samples from the target distribution \( p(\theta|y) \), due to the dependence on \( \theta^{[j-1]} \) in the acceptance probabilities. This is an example of a Markov chain, where the sample at each iteration is dependent on the sample at the previous iteration. The idea of Markov chain simulation is to simulate a Markov process (defined below) on \( \theta \), which converges to a stationary transition distribution that is \( p(\theta|y) \) (Tsay, 2002, pp. 396).

The general definition of a Markov process is given as follows: Consider a stochastic process \( \{X_t\} \), where each \( X_t \) assumes a value in the space \( \Theta \). The process \( \{X_t\} \) is a Markov process if its conditional distribution function satisfies

\[
p(X_h|X_s, s \leq t) = p(X_h|X_t), \quad h > t
\]

as mentioned in Tsay (2002). This is simply saying that, given the value of \( X_t \), the values of \( X_h \) do not depend on the values in periods before \( t \). If \( \{X_t\} \) is a discrete-time stochastic process (like the processes considered in this thesis), then this property becomes

\[
p(X_h|X_t, X_{t-1}, \ldots) = p(X_h|X_t), \quad h > t.
\]
Now let $A$ be a subset of $\Theta$. The function

$$p_t(\theta, h, A) = p(X_h \in A, X_t = \theta), \quad h > t$$

is called the transition probability function of the Markov process. If the transition probability depends on $h - t$, but not on $t$, then the process has a stationary transition distribution (see Tsay, 2002).

In terms of posterior distributions, Andrieu et al. (2003) states that for any starting point, the chain will converge to the stationary transition distribution $p(\theta|y)$ so long as it obeys the following properties:

1. **Irreducibility**: For any state of the Markov chain, there is a positive probability of visiting all other states in a finite number of transition steps.

2. **Aperiodicity**: The chain should not get trapped in cycles, i.e. it does not hit the same spot in the distribution every $k$ (say) iterations.

Generally methods that use Markov chain simulation to obtain a sample from the distribution $p(\theta|y)$ are called *Markov chain Monte Carlo* (MCMC) methods. These will be referred to as MCMC methods throughout the rest of this thesis. MCMC methods are typically used when the posterior distribution $p(\theta|y)$ is not available in a standard distributional form, as demonstrated in the relevant sections of this chapter.

### 2.5 Gibbs Sampling

Geman and Geman (1984) and Gelfand and Smith (1990) consider this very popular MCMC method. It is often regarded as a special case of the MH algorithm - the proof of which can be found in Gelman et al. (2000). It is mainly used when there are numerous parameters in a model, so rather than attempting to simulate from the joint posterior distribution $p(\theta|y)$ directly (which can be problematic), it is often easier to split the
parameters into groups and simulate from conditional posterior distributions instead through the use of a Markov chain. The algorithm is described as follows.

Assume the data $y = \{y_1, \ldots, y_n\}$ has been collected and estimations of the parameters $\Theta = \{\theta_1, \ldots, \theta_m\}$ are sought based on the entertained model $M$. (Here the $\theta$’s are not necessarily scalars. The case of the $\theta$’s being vectors is considered later.) Also assume that knowledge exists of the full conditional posterior distributions, i.e.

$$p(\theta_1 | y, \theta_2, \theta_3, \ldots, \theta_m, M),$$

$$p(\theta_2 | y, \theta_1, \theta_3, \ldots, \theta_m, M),$$

$$\vdots$$

$$p(\theta_m | y, \theta_1, \theta_2, \ldots, \theta_{m-1}, M).$$

This knowledge can be obtained via Bayes’ rule using a specified likelihood function $p(y | \Theta, M)$ and a prior distribution $p(\theta_i | \theta_{\neq i}, M)$ for each $i = 1, \ldots, m$. (The notation $\theta_{\neq i}$ denotes all parameters that are not $\theta_i$.)

Let $\Theta^{[1]} = \{\theta_1^{[1]}, \theta_2^{[1]}, \ldots, \theta_m^{[1]}\}$ denote arbitrary starting values of $\Theta$. The Gibbs sampler simulates $\Theta^{[2]} = \{\theta_1^{[2]}, \theta_2^{[2]}, \ldots, \theta_m^{[2]}\}$ in the following sampling scheme:

1. Draw a random sample $\theta_1^{[2]}$ from $p(\theta_1 | y, \theta_2^{[1]}, \theta_3^{[1]}, \ldots, \theta_m^{[1]}, M),$

2. Draw a random sample $\theta_2^{[2]}$ from $p(\theta_2 | y, \theta_1^{[2]}, \theta_3^{[1]}, \ldots, \theta_m^{[1]}, M),$

$$\vdots$$

$m$. Draw a random sample $\theta_m^{[2]}$ from $p(\theta_m | y, \theta_1^{[2]}, \theta_2^{[2]}, \ldots, \theta_{m-1}^{[2]}, M).$

This completes what is called a Gibbs iteration. Then using $\Theta^{[2]}$ as new starting values, repeat the $m$ steps to obtain $\Theta^{[3]} = \{\theta_1^{[3]}, \theta_2^{[3]}, \ldots, \theta_m^{[3]}\}$ as the next Gibbs iteration. This can be repeated many times, so at step $j$, the sampling scheme is:

1. Draw a random sample $\theta_1^{[j]}$ from $p(\theta_1 | y, \theta_2^{[j-1]}, \theta_3^{[j-1]}, \ldots, \theta_m^{[j-1]}, M),$

2. Draw a random sample $\theta_2^{[j]}$ from $p(\theta_2 | y, \theta_1^{[j]}, \theta_3^{[j-1]}, \ldots, \theta_m^{[j-1]}, M),$

19
m. Draw a random sample $\theta_m^{[j]}$ from $p(\theta_m|y, \theta_1^{[j]}, \theta_2^{[j]}, \ldots, \theta_{m-1}^{[j]}, M)$.

Note that these samples will be dependent as they form a Markov chain. Doing this $J$ times (say), will result in the samples $\theta^{[1]}, \theta^{[2]}, \ldots, \theta^{[J]}$. For a chosen $w$ ($1 \leq w < J$) and large $J$, the samples $\left\{ \theta^{[j]} \right\}_{j=w+1}^{J}$ will be approximately equivalent to dependent simulations from the joint posterior distribution $p(\theta_1, \theta_2, \ldots, \theta_m|y, M)$. See Tierney (1994) for more information.

The first $w$ iterations are referred to as the burn-in period. The setting of $w$ mainly depends on the speed of mixing and efficiency of convergence of the Markov chain to the joint posterior distribution - the quicker the mixing speed, the smaller the $w$. This is discussed later in this thesis in greater detail.

The process just described is one version of the Gibbs sampler, where one single long Markov chain is used to obtain dependent samples. If the independence property is to be kept, then the following version can be used, which consists of running relatively short Markov chains using different starting values and a relatively small $J$. The following comes from Verdinelli and Wasserman (1991), however it is slightly modified to include the set of $m$ parameters:

Begin by considering $R$ groups of arbitrary starting values for each parameter $\theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$:

$$\begin{align*}
&\{(\theta_1)_1^{[1]}, (\theta_2)_1^{[1]}, \ldots, (\theta_m)_1^{[1]}\}, \\
&\{(\theta_1)_2^{[1]}, (\theta_2)_2^{[1]}, \ldots, (\theta_m)_2^{[1]}\}, \\
&\quad \vdots \\
&\{(\theta_1)_r^{[1]}, (\theta_2)_r^{[1]}, \ldots, (\theta_m)_r^{[1]}\}, \\
&\quad \vdots \\
&\{(\theta_1)_R^{[1]}, (\theta_2)_R^{[1]}, \ldots, (\theta_m)_R^{[1]}\}.
\end{align*}$$
For each of these $R$ groups, the $J$ Gibbs iterations are simulated in exactly the same way as before. So for the $r$th group of parameters, the following collection of random numbers is obtained:

$$
\begin{pmatrix}
\theta_1^{[1]}_r & \theta_2^{[1]}_r & \cdots & \theta_m^{[1]}_r \\
\theta_1^{[2]}_r & \theta_2^{[2]}_r & \cdots & \theta_m^{[2]}_r \\
\vdots & \vdots & \ddots & \vdots \\
\theta_1^{[J]}_r & \theta_2^{[J]}_r & \cdots & \theta_m^{[J]}_r
\end{pmatrix}.
$$

As a result, we obtain $R$ collections of $J$ sets of random numbers:

$$
\begin{pmatrix}
\theta_1^{[1]}_1 & \theta_2^{[1]}_1 & \cdots & \theta_m^{[1]}_1 \\
\theta_1^{[2]}_1 & \theta_2^{[2]}_1 & \cdots & \theta_m^{[2]}_1 \\
\vdots & \vdots & \ddots & \vdots \\
\theta_1^{[J]}_1 & \theta_2^{[J]}_1 & \cdots & \theta_m^{[J]}_1
\end{pmatrix}, \ldots , \begin{pmatrix}
\theta_1^{[1]}_2 & \theta_2^{[1]}_2 & \cdots & \theta_m^{[1]}_2 \\
\theta_1^{[2]}_2 & \theta_2^{[2]}_2 & \cdots & \theta_m^{[2]}_2 \\
\vdots & \vdots & \ddots & \vdots \\
\theta_1^{[J]}_2 & \theta_2^{[J]}_2 & \cdots & \theta_m^{[J]}_2
\end{pmatrix}, \ldots , (2.6)
$$

$$
\begin{pmatrix}
\theta_1^{[1]}_r & \theta_2^{[1]}_r & \cdots & \theta_m^{[1]}_r \\
\theta_1^{[2]}_r & \theta_2^{[2]}_r & \cdots & \theta_m^{[2]}_r \\
\vdots & \vdots & \ddots & \vdots \\
\theta_1^{[J]}_r & \theta_2^{[J]}_r & \cdots & \theta_m^{[J]}_r
\end{pmatrix}, \ldots , \begin{pmatrix}
\theta_1^{[1]}_R & \theta_2^{[1]}_R & \cdots & \theta_m^{[1]}_R \\
\theta_1^{[2]}_R & \theta_2^{[2]}_R & \cdots & \theta_m^{[2]}_R \\
\vdots & \vdots & \ddots & \vdots \\
\theta_1^{[J]}_R & \theta_2^{[J]}_R & \cdots & \theta_m^{[J]}_R
\end{pmatrix}. (2.7)
$$

Geman and Geman (1984) show that \{$(\theta_1)^{[J]}_1, (\theta_2)^{[J]}_2, \ldots, (\theta)^{[J]}_R$\} is a sample of size $R$ drawn from a cumulative distribution function (CDF) that converges to the joint posterior distribution $p(\theta_1, \theta_2, \ldots, \theta_m | y, M)$. That is, for large enough $J$, the last rows of the matrices in (2.6) and (2.7) will approximately be an independent sample of size $R$ randomly drawn from this posterior. The samples are in fact independent, since the samples utilise values from each $r$th group.

Estimates of the parameters in $\theta$ can be obtained in slightly different ways in both versions. In general, the two most common Bayesian parameter estimators are the maximum
a posteriori (MAP) and the minimum mean square error (MMSE) estimators defined by

\[
\theta_{\text{MAP}} = \arg \max p(\theta | y), \quad (2.8)
\]
\[
\theta_{\text{MMSE}} = \int \theta p(\theta | y) d\theta. \quad (2.9)
\]

Hence the estimators defined in (2.8) and (2.9) respectively equal the posterior mode and mean, as described in Peters and Sisson (2006). Approximations of these statistics can be obtained via the MCMC samples generated from the Gibbs sampler above: For the version using a long Markov chain and dependent samples,

\[
\hat{\theta}_{\text{MAP}} = \arg \max \{p(\theta^{[w+1]} | y), \ldots, p(\theta^{[J]} | y)\},
\]
\[
\hat{\theta}_{\text{MMSE}} = \frac{1}{J-w} \sum_{j=w+1}^{J} \theta^{[j]}.
\]

For the version using short Markov chains and independent samples,

\[
\hat{\theta}_{\text{MAP}} = \arg \max \{p((\theta_{1})^{[J]} | y), \ldots, p((\theta_{R})^{[J]} | y)\},
\]
\[
\hat{\theta}_{\text{MMSE}} = \frac{1}{R} \sum_{r=1}^{R} (\theta_{r}^{[J]}).
\]

Gibbs sampling, whichever version used, has the advantage to decompose a high dimensional estimation problem into several lower dimensional ones via full conditional distributions of the parameters. At the extreme, a high-dimensional problem with \(m\) parameters can be solved iteratively by using \(m\) univariate conditional distributions (Tsay, 2002, pp. 398). Tsay (2002) also makes the point that this is not always efficient: When parameters are highly correlated, it pays to draw them jointly. So if the parameters \(\theta_a\) and \(\theta_b\) are highly correlated, then the \(a\)th and \(b\)th step in a Gibbs iteration can be combined into the one step (\(c\), say) to read

\[
c. \text{ Draw a random sample } (\theta_a, \theta_b) \text{ from } p(\theta_a, \theta_b | y, \theta_{\neq a}, \theta_{\neq b}, M).
\]

Obviously, more than 2 parameters can be correlated with each other. This is known as
“blocking”. An early example of this can be found in Carter and Kohn (1994) who apply it to a state space model. The above step is adjusted appropriately to incorporate the number of “blocked” parameters. This process will be utilised in all models proposed in this thesis.

2.6 Chapter Summary

This chapter has reviewed the Bayesian estimation techniques from the literature that will be applied in this thesis. Bayes’ rule was introduced, which lays the foundation of Bayesian inference. The Metropolis, Metropolis-Hastings and Delayed Rejection algorithms were presented to demonstrate the process of simulating from joint posterior distributions of unknown form. The definition of a Markov chain was described, with the above algorithms satisfying its definition. Finally, the Gibbs sampler was illustrated to show that a high dimensional problem can be decomposed into smaller ones via simulating from conditional posterior distributions. Various combinations of these Bayesian estimation techniques will be applied in this thesis.

The time taken to run all Bayesian and classical estimation algorithms on the models presented in this thesis will be also stated, where applicable. The Intel Visual Fortran Compiler v8.1 is utilised for programming the algorithms, and are run on an Intel Quad Core i7 processor with a speed of 1.60GHz and 8GB of RAM.

The following chapter suggests a new approach to estimating multivariate GARCH models, by using the above Bayesian estimation techniques to assist in expanding the parameter space typically explored in the literature for such models.
Chapter 3

A Bayesian Approach to Relaxing Parameter Restrictions in Multivariate GARCH Models

This chapter is a more detailed version of the paper that appears as the following:


The original publication is available at www.springerlink.com.

In this chapter, a Bayesian prior formulation is proposed for a multivariate GARCH model that expands the allowable parameter space, implicitly enforcing both necessary and sufficient conditions for positive definiteness and covariance stationarity. This extends the standard approach of enforcing unnecessary parameter restrictions on these types of models. A VECH model specification is proposed that allows both parsimony and parameter interpretability, opposing existing specifications that achieve only one of these properties. A Markov chain Monte Carlo scheme, employing Metropolis-Hastings and Delayed Rejection algorithms, is designed. A simulation study shows favourable estimation and improved coverage of intervals, compared with classical methods applied to an equivalent BEKK model. Finally, some US and UK financial stock returns are

24
analysed, the results of which favour the proposed model specification.

3.1 Introduction

It is well documented that financial time series data exhibits time varying volatility. Engle (1982) and Bollerslev (1986) developed the Autoregressive Conditional Heteroscedastic (ARCH) and Generalised ARCH (GARCH) models to capture this empirical behaviour in univariate series. However, financial market theory has recently switched from the traditional view of market independence to the concept of market integration. A substantial amount of research now focuses on modelling co-movements of international stock markets and explores the dynamics of return covariances and spillover effects between markets using multivariate heteroscedastic models (e.g. see Ramchand and Susmel, 1998; Sim and Zurbruegg, 2001 and Fang and Miller, 2007).

The modelling of such time varying conditional covariance matrices for multiple asset returns has proven a challenge in the literature. Not only is it necessary to estimate the numerous parameters in the multivariate model, but simultaneously it must be ensured that the dynamic covariance matrices are positive definite at each time point. In addition, enforcement of stationarity conditions, as well as model parsimony and parameter interpretability, are major issues for multivariate models. Perhaps the simplest such model in the literature is the VECH model of Bollerslev et al. (1988), who directly extend the univariate GARCH model to more than one dimension. This model has high parameter interpretability, since each unknown parameter appears only once and in a linear fashion in each equation, but suffers from a lack of parsimony in parameter numbers and further it is difficult to ensure positive definiteness (PD) for the dynamic covariance matrices. A popular model proposed in response to these problems is the BEKK model introduced by Baba et al. (1990) and later published by Engle and Kroner (1995). This GARCH-type model ensures PD of each covariance matrix and has fewer parameters in general than the corresponding VECH model. However, parameter interpretation can be difficult with
the BEKK model, since many parameters appear in more than one equation and in a nonlinear fashion. A model specification is proposed in this chapter that is a special case of the VECH model, but permits parsimony and retains ease of interpretation and linearity in parameters.

Typically in the literature, only sufficient conditions for PD and covariance stationarity (CS) are imposed on GARCH models in general, as noted in Gourieroux (1997), Nelson and Cao (1992) and Chen et al. (2005); necessary conditions for these models have proven difficult to derive in closed form (see for e.g. Nelson, 1990; Ling and McAleer, 2002a,b). In practice then, GARCH models are often estimated enforcing sufficient conditions only, typically in the form of explicit or implied parameter restrictions, which may be restrictive and cut off part of the allowable parameter space; this is equally true under both Bayesian and classical approaches. At this point it is worth noting that CS concerns the properties of a model and is thus an unconditional property. PD, on the other hand, is not an unconditional property, but rather a constraint to be satisfied for estimates of covariance matrices, conditional upon the specific data set under analysis. This chapter will show that it is possible to find parameter values outside the usual sufficient conditions where PD and CS are still satisfied for real and simulated data, highlighting that the standard restrictions are not necessary.

A prior distribution is developed in this chapter for a multivariate GARCH model that will directly satisfy both necessary and sufficient conditions for PD and CS, while remaining diffuse and non-informative over the allowable parameter space, thus extending existing inference approaches in the literature. Markov chain Monte Carlo (MCMC) methods, including the Metropolis-Hastings (MH) algorithm, will be employed to help enforce the conditions in this prior. This approach can be utilised or extended to cover GARCH models in general; this chapter focuses on a specific multivariate model.

This chapter is set out as follows: Section 3.2 reviews some multivariate GARCH models that exist in the literature. Section 3.3 presents the proposed parsimonious VECH specification and discusses PD and CS in general. Section 3.4 presents the Bayesian
MCMC methods for estimation and presents the proposed prior formulation. Section 3.5 presents a simulation study, while Section 3.6 discusses an empirical study. Section 3.7 concludes the argument.

3.2 Multivariate GARCH Models

This section reviews some multivariate GARCH models existing in the literature. In particular, the VECH model of Bollerslev et al. (1988) and the BEKK model of Engle and Kroner (1995) are of particular importance to this chapter, while other models are reviewed to enhance understanding of models presented later in this thesis.

3.2.1 The VECH Model

Bollerslev et al. (1988) propose the multivariate VECH GARCH model: let \( \{ \epsilon_t \}_{t=1}^n = \{ (\epsilon_{1t}, \ldots, \epsilon_{Nt}) \}_{t=1}^n \) denote a set of \( N \)-dimensional residual vectors and \( \psi_{t-1} \) the information set at time \( t-1 \), so that \( E(\epsilon_t|\psi_{t-1}) = 0 \). Also let \( H_t \) be a dynamic conditional covariance matrix of order \( N \) with elements

\[
H_t = \begin{pmatrix}
    h_{11,t} & h_{12,t} & \cdots & h_{1N,t} \\
    h_{12,t} & h_{22,t} & \cdots & h_{2N,t} \\
    \vdots & \vdots & \ddots & \vdots \\
    h_{1N,t} & h_{2N,t} & \cdots & h_{NN,t}
\end{pmatrix}
\]

so that \( \text{Cov}(\epsilon_t|\psi_{t-1}) = H_t \). The \( N \)-dimensional VECH(\( p, q \)) model states that \( H_t \) evolves over time according to the following equation:

\[
\text{vech}(H_t) = \text{vech}(C) + \sum_{i=1}^{q} A_i \text{vech}(\epsilon_{t-i}\epsilon_{t-i}') + \sum_{j=1}^{p} B_j \text{vech}(H_{t-j}), \quad (3.1)
\]

where \( \text{vech}(\cdot) \) denotes the column stacking operator of the lower portion of a symmetric matrix, \( C \) is a square symmetric parameter matrix of order \( N \) and each \( A_i \) and \( B_j \) are
square parameter matrices of order \(N(N+1)/2\). The order numbers \(p\) and \(q\) are typically both set to 1, as this is usually sufficient in practice; see Bollerslev et al. (1992).

To illustrate this model, let \(N = 2\) and \(p = q = 1\). If \(C = [c_{ij}], A_1 = [a_{ij}]\) and \(B_1 = [b_{ij}]\), then equation (3.1) becomes:

\[
\begin{pmatrix}
h_{11,t} \\
h_{12,t} \\
h_{22,t}
\end{pmatrix} =
\begin{pmatrix}
c_{11} & c_{12} & c_{13} \\
c_{12} & c_{22} & c_{23} \\
c_{22} & c_{32} & c_{33}
\end{pmatrix}
\begin{pmatrix}
\varepsilon_{1,t-1}^2 \\
\varepsilon_{1,t-1}\varepsilon_{2,t-1} \\
\varepsilon_{2,t-1}^2
\end{pmatrix}
+ \begin{pmatrix}
b_{11} & b_{12} & b_{13} \\
b_{21} & b_{22} & b_{23} \\
b_{31} & b_{32} & b_{33}
\end{pmatrix}
\begin{pmatrix}
h_{11,t-1} \\
h_{12,t-1} \\
h_{22,t-1}
\end{pmatrix}
\]

which equates to

\[
\begin{align*}
h_{11,t} &= c_{11} + a_{11}\varepsilon_{1,t-1}^2 + a_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{13}\varepsilon_{2,t-1}^2 + b_{11}h_{11,t-1} + b_{12}h_{12,t-1} + b_{13}h_{22,t-1}, \\
h_{12,t} &= c_{12} + a_{21}\varepsilon_{1,t-1}^2 + a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{23}\varepsilon_{2,t-1}^2 + b_{21}h_{11,t-1} + b_{22}h_{12,t-1} + b_{23}h_{22,t-1}, \\
h_{22,t} &= c_{22} + a_{31}\varepsilon_{1,t-1}^2 + a_{32}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{33}\varepsilon_{2,t-1}^2 + b_{31}h_{11,t-1} + b_{32}h_{12,t-1} + b_{33}h_{22,t-1}.
\end{align*}
\]

The main advantages of this model are:

- Easy parameter interpretation - each contributes uniquely and linearly to a particular lagged residual or covariance term, similar to that of a univariate GARCH model. This can be seen in the bivariate example above.

- Residual and volatility spillover effects are easily captured, since the model is in its most general form (e.g. \(\varepsilon_{2,t-1}^2\) and \(h_{22,t-1}\) both appear in the equation for \(h_{11,t}\)). This is a common trait of financial markets.

However, this model suffers from some major afflictions:

- The model in its most general form in (3.1) contains a large number of parameters; there is a total of \(N(N+1)[N(N+1)(p+q) + 2]/4\) parameters required to be estimated (e.g. \(N = 3\) and \(p = q = 1\) gives 78 parameters).

- Restricting \(H_t\) to be PD for every \(t\) can be very difficult, as discussed in Engle and Kroner (1995).
To improve parsimony, Bollerslev et al. (1988) also discuss a special case, the diagonal \textit{VECH} (DVECH) model, where \( p = q = 1 \) and the matrices \( A_1 \) and \( B_1 \) are assumed diagonal. To demonstrate, redefine \( A_1 = [a_{ij}] \) and \( B_1 = [b_{ij}] \) as \( N \times N \) symmetric parameter matrices. In the DVECH(1,1) model, each covariance equation contained in \( H_t \) can now be written as

\[
h_{ij,t} = c_{ij} + a_{ij} \varepsilon_{i,t-1} \varepsilon_{j,t-1} + b_{ij} h_{ij,t-1}, \quad \forall \ i \leq j = 1, \ldots, N.
\] (3.2)

For \( N = 2 \), equation (3.2) gives rise to

\[
\begin{align*}
h_{11,t} &= c_{11} + a_{11} \varepsilon_{1,t-1}^2 + b_{11} h_{11,t-1}, \\
h_{12,t} &= c_{12} + a_{12} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + b_{12} h_{12,t-1}, \\
h_{22,t} &= c_{22} + a_{22} \varepsilon_{2,t-1}^2 + b_{22} h_{22,t-1}.
\end{align*}
\]

Thus only the own lagged moments and cross products appear in each of the conditional covariance equations (Bollerslev et al., 1988). Having this specification does result in a parameter size reduction, especially in small systems \((N \leq 5)\): there are \(3N(N + 1)/2\) parameters for the DVECH(1,1) model (e.g. \( N = 3 \) gives 18 parameters). However, the main drawback of this model is that it does not capture residual and/or volatility spillover effects between markets like the general VECH\((p,q)\) model in equation (3.1).

Covariance stationarity conditions for VECH models have been discussed extensively in the literature (e.g. see Engle and Kroner, 1995; Ledoit et al., 2003 and Bauwens et al., 2006). They each arrive at the same conclusions: assuming that \( H_t \) is positive definite, the VECH\((p,q)\) model from equation (3.1) is ensured to be covariance stationary if:

\[
c_{ii} > 0, \quad \forall \ i = 1, \ldots, N, \tag{3.3}
\]

All elements of \( A_i \) and \( B_j \) are \( \geq 0 \), \( \forall \ i = 1, \ldots, q; \ \forall \ j = 1, \ldots, p, \tag{3.4} \)

All eigenvalues of \( \sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j \) are less than one in modulus. \( \tag{3.5} \)
These sufficient conditions are typically enforced during parameter estimation for both classical and Bayesian approaches. However, there are no known closed-form necessary and sufficient conditions for PD in a general VECH model in the literature, including the DVECH model.

### 3.2.2 The BEKK Model

Engle and Kroner (1995) describe the BEKK model, a specification that automatically enforces each matrix $H_t$ to be PD. The BEKK$(p,q,K)$ model has the form

$$H_t = C^* C^* + \sum_{k=1}^{K} \sum_{i=1}^{q} A^*_{ik} \varepsilon_{t-i} \varepsilon_{t-i} A^*_i + \sum_{k=1}^{K} \sum_{j=1}^{p} B^*_{jk} H_{t-j} B^*_{jk}, \quad (3.6)$$

where $C^*$, and each $A^*_{ik}$ and $B^*_{jk}$ are $N \times N$ parameter matrices (with $C^*$ upper triangular), and the order $K$ denotes the generality of the model, often used to make the BEKK model equivalent to a corresponding VECH model. (The asterisk * has been added above to distinguish between VECH and BEKK parameters.) Similar to the VECH model, $p$ and $q$ are typically set to 1.

To illustrate this model, let $N = 2$ and $p = q = K = 1$. If $C^* = [c^*_{ij}]$, $A^*_{11} = [a^*_{ij}]$ and $B^*_{11} = [b^*_{ij}]$ then the BEKK$(1,1,1)$ model from equation (3.6) becomes

$$
\begin{pmatrix}
    h_{11,t} & h_{12,t} \\
    h_{12,t} & h_{22,t}
\end{pmatrix}
= \begin{pmatrix}
    c^*_{11} & c^*_{12} \\
    0 & c^*_{22}
\end{pmatrix}
\begin{pmatrix}
    c^*_{11} & c^*_{12} \\
    0 & c^*_{22}
\end{pmatrix}^t
+ \begin{pmatrix}
    a^*_{11} & a^*_{12} \\
    a^*_{21} & a^*_{22}
\end{pmatrix}
\begin{pmatrix}
    \varepsilon^2_{1,t-1} & \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\
    \varepsilon_{1,t-1} \varepsilon_{2,t-1} & \varepsilon^2_{2,t-1}
\end{pmatrix}
\begin{pmatrix}
    a^*_{11} & a^*_{12} \\
    a^*_{21} & a^*_{22}
\end{pmatrix}
+ \begin{pmatrix}
    b^*_{11} & b^*_{12} \\
    b^*_{21} & b^*_{22}
\end{pmatrix}
\begin{pmatrix}
    h_{11,t-1} & h_{12,t-1} \\
    h_{12,t-1} & h_{22,t-1}
\end{pmatrix}
\begin{pmatrix}
    a^*_{11} & a^*_{12} \\
    a^*_{21} & a^*_{22}
\end{pmatrix}
$$

30
which equates to

\[
\begin{align*}
    h_{11,t} &= c_{11}^2 + a_{11}^2 \varepsilon_{1,t-1} + 2a_{11}a_{21}^* \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{21}^2 \varepsilon_{2,t-1}^2 \\
    &+ b_{11}^2 h_{11,t-1} + 2b_{11}^* b_{21}^* h_{12,t-1} + b_{21}^2 h_{22,t-1}, \\
    h_{12,t} &= c_{12}^* + a_{11}^* a_{12}^* \varepsilon_{1,t-1} + (a_{11}^* a_{22} + a_{12}^* a_{21}^*) \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{21}^* a_{22}^* \varepsilon_{2,t-1}^2 \\
    &+ b_{11}^* b_{12}^* h_{11,t-1} + (b_{11}^* b_{22}^* + b_{12}^* b_{21}^*) h_{12,t-1} + b_{21}^* b_{22}^* h_{22,t-1}, \\
    h_{22,t} &= c_{22}^* + a_{12}^* \varepsilon_{1,t-1} + 2a_{12}^* a_{22}^* \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{22}^* \varepsilon_{2,t-1}^2 \\
    &+ b_{12}^* b_{12}^* h_{11,t-1} + 2b_{12}^* b_{22}^* h_{12,t-1} + b_{22}^* h_{22,t-1}.
\end{align*}
\]  

(3.7)  

(3.8)  

(3.9)

Similarly to the VECH model, a further reduction in parameter numbers can be achieved via a diagonal representation of (3.6):

\[
\begin{align*}
    h_{ij,t} &= c_{ij}^* + a_{ij}^* a_{jj}^* \varepsilon_{i,t-1} \varepsilon_{j,t-1} + b_{ij}^* b_{jj}^* h_{ij,t-1}, \quad \forall \ i \leq j = 1, \ldots, N,
\end{align*}
\]

(3.10)

which is obtained by setting \( p = q = K = 1 \), forcing the matrices \( A_{11}^* \) and \( B_{11}^* \) to be diagonal and setting \( C^* C^* = [c_{ij}^*] \). Thus the model in (3.10) is called a diagonal BEKK (DBEKK) model. Similar to the DVECH model, the DBEKK model does not allow residual and volatility spillover effects.

BEKK models are generally more parsimonious than VECH models of similar order. For a general BEKK(\( p, q, K \)) model there are \( N[N+1+2KN(p+q)]/2 \) parameters in total (e.g. \( N = 3 \) and \( p = q = K = 1 \) gives 24 parameters). Also \( H_t \) will be PD, albeit under very weak conditions (see Engle and Kroner, 1995). However, the parameters \( (a_{ij}^*, b_{ij}^*) \) in the BEKK model seem difficult to interpret, since each covariance equation for \( h_{ij,t} \) will include a coefficient that is not linear in the parameters and each parameter appears in more than one equation (see equations (3.7)-(3.9) for an example). This may set up hidden restrictions or relationships between parameters that can restrict some parts of the allowable parameter space. For example, when \( N = 2 \) in a DBEKK(1,1,1) model, if \( a_{11}^* = 0.05 \) and \( a_{22}^* = 0.15 \) (both likely values for real data), then the ARCH effect in
the equation for the covariance \( h_{12,t} \) must equal \( a_{11}^* a_{22}^* = 0.0075 \), which seems an unlikely and unnecessary restriction. As such general BEKK specifications may cause biases, or at least inefficiencies, in parameter estimates.

To state the stationarity conditions for the general BEKK\((p, q, K)\) model, the following matrix operation first needs to be defined:

### The Kronecker Product

Let \( A = [a_{ij}] \) denote an \( m_1 \times n_1 \) matrix and \( B = [b_{ij}] \) an \( m_2 \times n_2 \) matrix. The *Kronecker product* \( A \otimes B \) is defined to be the following \( m_1 m_2 \times n_1 n_2 \) matrix:

\[
A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \ldots & a_{1n_1}B \\ a_{21}B & a_{22}B & \ldots & a_{2n_1}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m_11}B & a_{m_12}B & \ldots & a_{m_1n_1}B \end{pmatrix}.
\]


The BEKK\((p, q, K)\) model is covariance stationary if and only if all eigenvalues of the matrix \( \sum_{k=1}^{K} \sum_{i=1}^{q} (A_{ik}^* \otimes A_{ik}^*) + \sum_{k=1}^{K} \sum_{j=1}^{p} (B_{jk}^* \otimes B_{jk}^*) \) are less than one in modulus, as proven in Engle and Kroner (1995). For example, in the BEKK\((1, 1, 1)\) model in equations (3.7)-(3.9), the stationarity condition is that the eigenvalues of \( (A_{11}^* \otimes A_{11}^*) + (B_{11}^* \otimes B_{11}^*) \) are less than one in modulus, where

\[
A_{11}^* \otimes A_{11}^* = \begin{pmatrix} a_{11}^* A_{11}^* & a_{12}^* A_{11}^* \\ a_{21}^* A_{11}^* & a_{22}^* A_{11}^* \end{pmatrix} = \begin{pmatrix} a_{11}^* a_{11}^* & a_{11}^* a_{12}^* & a_{11}^* a_{12}^* & a_{12}^* a_{12}^* \\ a_{11}^* a_{21}^* & a_{11}^* a_{22}^* & a_{12}^* a_{21}^* & a_{12}^* a_{22}^* \\ a_{11}^* a_{21}^* & a_{12}^* a_{22}^* & a_{11}^* a_{22}^* & a_{12}^* a_{22}^* \\ a_{21}^* a_{21}^* & a_{21}^* a_{22}^* & a_{22}^* a_{22}^* & a_{22}^* a_{22}^* \end{pmatrix}
\]

and
\[ B^*_1 \otimes B^*_1 = \left( \begin{array}{ccc} b^*_{11}B^*_1 & b^*_1b^*_2B^*_1 \\ b^*_2B^*_1 & b^*_2b^*_1B^*_1 \end{array} \right) = \left( \begin{array}{cccc} b^2_{11} & b^*_{11}b^*_{12} & b^*_{11}b^*_{12} & b^2_{12} \\ b^*_{11}b^*_{21} & b^*_{11}b^*_{22} & b^*_{12}b^*_{21} & b^*_{12}b^*_{22} \\ b^*_{21}b^*_{11} & b^*_{21}b^*_{12} & b^*_{22}b^*_{11} & b^*_{22}b^*_{12} \\ b^*_{21}b^*_{22} & b^*_{21}b^*_{22} & b^*_{22}b^*_{22} \end{array} \right). \]

The stationarity conditions for both Vech and BEKK models are similar, but they are more difficult to enforce for the BEKK model, especially for large \( N \). If we were to rewrite the constant matrix \( C^*C^* \) as a single matrix (\( D^* \) say), then an added BEKK restriction to ensure PD would be that \( D^* \) is PD; the \( C^*C^* \) structure enforces PD of this matrix immediately. Another issue with BEKK models is avoiding observationally equivalent structures. For example, if \( i = j = 1 \) in model (3.10), then the coefficient of \( \varepsilon^2_{11,t-1} \) is \( a^*_{11} \). If the value of \( a^*_{11} \) is \(-0.3\), this would have the same effect on \( h_{11,t} \) as if \( a^*_{11} \) were \(0.3\). So to avoid this, a further condition usually imposed is positivity, i.e. that all elements of \( A^*_{ik} \) and \( B^*_{jk} \) are \( \geq 0 \), \( \forall i = 1, \ldots, q \), \( \forall j = 1, \ldots, p \) and \( \forall k = 1, \ldots, K \), which is required to achieve uniqueness in parameter representation.

### 3.2.3 The FARCH Model

Another multivariate GARCH model that enables a reduced number of parameters and ease of PD enforcement is the Factor ARCH (FARCH) model proposed by Engle et al. (1990), which specifies \( H_t \) as follows:

\[ H_t = C + \sum_{k=1}^{K} \lambda_k X_k^t \theta_{kt} \]

where \( C \) is an \( N \times N \) positive semi-definite matrix, the \( \lambda_k \)'s are \( N \times 1 \) vectors and the \( \theta_{kt} \)'s are positive random variables. This model assumes that the multivariate stock return evolution is driven by \( K \) underlying variables, or factors, which reduce the number of parameters to estimate when \( K < N \). Engle et al. (1990) make the point that each \( \theta_{kt} \) “can be any function of variables measurable with respect to the information set at time...”
but focus their attention on two practical restrictions:

- **The univariate portfolio representation**, which assumes that $\theta_{kt}$ is the conditional variance process of the weighted portfolio of stock returns $P_{kt} = w_k'y_t$ (with $w_k$ an $N \times 1$ vector) that follows a univariate GARCH model. For example,

$$\theta_{kt} = c_k + a_k u_{k,t-1}^2 + b_k \theta_{k,t-1}, \quad u_{kt} \mid \psi_{t-1} \sim N(0, \theta_{kt})$$

where $u_{kt} = w_k'\varepsilon_t$ and $c_k$, $a_k$ and $b_k$ are the usual GARCH parameters.

- **The recursive portfolio representation**, which assumes that $\theta_{kt}$ is the conditional variance process of $P_{kt}$ that depends on its own past conditional variances, plus the past conditional variances of the first $k - 1$ portfolios. For example

$$\theta_{kt} = c_k + a_{kk} u_{k,t-1}^2 + b_{kk} \theta_{k,t-1} + \sum_{j=1}^{k-1} [a_{kj} u_{j,t-1}^2 + b_{kj} \theta_{j,t-1}].$$

An advantage of this model is that $H_t$ is guaranteed to be positive semi-definite. In addition, the number of parameters are significantly reduced by allowing only a small number of factors to drive the multivariate stock return process rather than utilising the full $N$ dimensions. Because the portfolio representations are used for $\theta_{kt}$, only univariate methods are required for parameter estimation which are straightforward to implement. However, similar to the BEKK model, the parameters of the FARCH model seem difficult to interpret, since some parameters do not contribute uniquely to each model effect. While residual and volatility spillover effects are captured in the recursive portfolio representation above, it is at a portfolio level only and would be difficult to identify the contribution of each stock return to this effect.

### 3.2.4 The CCC Model

Another model introduced to simplify estimation and reduce parameter numbers is the Constant Conditional Correlation (CCC) model of Bollerslev (1990). As the name im-
plies, this model assumes that each conditional correlation \( \rho_{ij,t} = \frac{h_{ij,t}}{\sqrt{h_{ii,t}h_{jj,t}}} \), where \(-1 \leq \rho_{ij,t} \leq 1\), remains constant over time so that each conditional covariance equation \( h_{ij,t} \) can be represented as

\[
h_{ij,t} = \rho_{ij} \sqrt{h_{ii,t}h_{jj,t}}, \quad \forall \; i < j = 1, \ldots, N.
\]

Then \( H_t \) can be written as follows:

\[
H_t = D_t P D_t
\]

(3.11)

where \( D_t = \text{diag}(\sqrt{h_{11,t}}, \ldots, \sqrt{h_{NN,t}}) \) is an \( N \times N \) stochastic diagonal matrix and \( P \) is an \( N \times N \) time-invariant correlation matrix with typical element \( \rho_{ij} \).

Having this representation for a multivariate GARCH model leads to some attractive properties. Firstly, it follows that \( H_t \) will be PD if and only if each of the \( N \) conditional variances are well defined and \( P \) is PD. This is automatically satisfied if each \( h_{ii,t} \) is a GARCH\((p,q)\) model. This also implies easy parameter interpretation under this setting. Secondly, parameter estimation is much less computationally expensive since \( P \) is a constant matrix. However, in many applications, it has been proven that conditional correlation varies with time, thus rendering this restriction infeasible.

### 3.2.5 The DCC Model

Engle (2002) relaxes the constant conditional correlation assumption so that the matrix \( P \) in equation (3.11) is now time-varying, i.e.

\[
H_t = D_t P_t D_t
\]

(3.12)

where \( P_t = [\rho_{ij,t}] \). This is called the Dynamic Conditional Correlation (DCC) model. Engle (2002) describes various forms that \( \rho_{ij,t} \) could take, but perhaps the most commonly used form in the literature is the mean-reverting scalar equation: Let \( z_t = D_t^{-1} \epsilon_t \) denote
the standardised residual vector, $\bar{P} = E(z_t z_t')$ the unconditional correlation matrix of the standardised residuals and define $Q_t = [q_{i,j,t}]$. Then

$$Q_t = \bar{P}(1 - \alpha - \beta) + \alpha z_{t-1} z_{t-1}' + \beta Q_{t-1}$$  \hspace{1cm} (3.13)$$

where $\alpha$ and $\beta$ are scalar parameters. Each conditional correlation $\rho_{i,j,t}$ is then defined as

$$\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t} q_{j,j,t}}}, \hspace{0.5cm} \forall \ i \leq j = 1, \ldots, N.$$  

With the additional assumption that $\alpha + \beta < 1$, the scalar representation in (3.13) ensures that the long-run covariance matrix of the process is in fact $\bar{P}$, which also means that $Q_t$ will be PD. This is known as variance targeting (see for e.g. Engle and Mezrich, 1996).

Engle (2002) proposes a two-step approach to estimate the parameters of this model, by first writing the log-likelihood function $l$ as the sum of a volatility component $l_V$ and a correlation component $l_C$. The first step is to fit univariate GARCH models to each conditional variance $h_{i,i,t}$ by maximising $l_V$. Then, conditional on these estimates, estimate the parameters $\alpha$ and $\beta$ by maximising $l_C$. The unconditional correlation matrix $\bar{P}$ is set to its sample value. This is a straightforward procedure to implement. Moreover, for any value of $N$, the number of parameters remain the same since $\alpha$ and $\beta$ are used to explain all correlation dynamics in each dimension. However, we cannot distinguish between the correlation effects of the different return series since $\alpha$ and $\beta$ will be the same for each of them, thus resulting in difficult interpretation of these parameters. Of course, this problem can be alleviated by switching $\alpha$ and $\beta$ to symmetric matrices $A$ and $B$. However, this can significantly increase the number of parameters as $N$ becomes large and the two-step estimation procedure above becomes more difficult to implement.

**Summary**

Based on the multivariate GARCH models just discussed, it makes sense to propose a model that utilises all of their good properties. These include:
• Easy enforcement of $H_t$ to be PD, $\forall \ t = 1, \ldots, n$;
• Ease of parameter interpretation;
• Parsimony (i.e. a reduced number of parameters as $N$ increases);
• The capturing of residual and volatility spillover effects.

The model proposed in the next section will attempt to encapsulate all of these properties.

### 3.3 The PVECH Model

The model specification proposed here is a parsimonious representation of the VECH model in (3.1) for the conditional covariances, plus a vector autoregressive (AR) form (of order 1) in the mean equation. This is called the parsimonious VECH (PVECH) model:

Let $y = \{y_{it}\}_{t=1}^n = \{(y_{1t}, \ldots, y_{Nt})\}_{t=1}^n$ denote the $N$-dimensional multivariate time series of sample size $n$. In matrix form, the proposed PVECH model is given by:

\[
y_t = m_0 + \text{diag}(m_1)y_{t-1} + \epsilon_t, \quad \epsilon_t|\psi_{t-1} \sim N(0, H_t), \tag{3.14}
\]
\[
\text{vech}(H_t) = \text{vech}(C) + A_1\text{vech}(\epsilon_{t-1}\epsilon'_{t-1}) + B_1\text{vech}(H_{t-1}), \tag{3.15}
\]

where $m_0$ and $m_1$ are $N \times 1$ parameter vectors respectively containing the mean intercepts and autoregressive parameters for each series and $\text{diag}(\cdot)$ is a diagonal matrix with diagonal terms equal to the vector it operates on. The parsimony in this model is achieved by forcing a particular pattern on the parameter matrices $A_1 = [a_{ij}]$ and $B_1 = [b_{ij}]$, that maintains the spirit of a GARCH model representation, as follows: Restrict $A_1$ and $B_1$ so that only the following elements are non-zero:

\[
a_{ii}, b_{ii}, \quad \forall \ i = 1, \ldots, N(N + 1)/2;
\]
\[
a_{ij}, b_{ij}, \quad \forall \ i \neq j = 1, N + 1, 2N, 3N - 2, 4N - 5, \ldots, N(N + 1)/2.
\]
All other elements of $A_1$ and $B_1$ are set to 0. This model specification has $N(7N - 1)/2$ parameters (plus an additional $2N$ when including the mean parameters), which is always between the number of parameters required for a VECH(1,1) and a BEKK(1, 1, 1) model, but is much closer to that required for a BEKK(1, 1, 1). This can be shown in Table 3.1 and also graphically in Figure 3.1. The VECH(1, 1) parameter numbers essentially “explode” for values of $N$ greater than 3, while the PVECH and BEKK(1, 1, 1) models remain relatively stable. The DVECH(1, 1) and DBEKK(1, 1, 1) models have the lowest number of parameters (as expected), but as discussed in Section 3.2 they come at a cost of not being able to capture important data properties (e.g. volatility spillover effects).

Table 3.1: Parameter numbers for selected VECH and BEKK models under different dimensions

<table>
<thead>
<tr>
<th>$N$</th>
<th>PVECH</th>
<th>VECH(1,1)</th>
<th>DVECH(1,1)</th>
<th>BEKK(1, 1, 1)</th>
<th>DBEKK(1, 1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>13</td>
<td>21</td>
<td>9</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>78</td>
<td>18</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>54</td>
<td>210</td>
<td>30</td>
<td>42</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>85</td>
<td>465</td>
<td>45</td>
<td>65</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>123</td>
<td>903</td>
<td>63</td>
<td>93</td>
<td>33</td>
</tr>
<tr>
<td>7</td>
<td>168</td>
<td>1,596</td>
<td>84</td>
<td>126</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>220</td>
<td>2,628</td>
<td>108</td>
<td>164</td>
<td>52</td>
</tr>
<tr>
<td>9</td>
<td>279</td>
<td>4,095</td>
<td>135</td>
<td>207</td>
<td>63</td>
</tr>
<tr>
<td>10</td>
<td>345</td>
<td>6,105</td>
<td>165</td>
<td>255</td>
<td>75</td>
</tr>
</tbody>
</table>

To illustrate this model, consider the case $N = 3$. Defining the off-diagonal elements of $A_1$ and $B_1$ as $d_{ij}$ and $e_{ij}$ (respectively), the resulting matrices $A_1$ and $B_1$ are

$$A_1 = \begin{pmatrix} a_{11} & 0 & 0 & d_{12} & 0 & d_{13} \\ 0 & a_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{13} & 0 & 0 & 0 \\ d_{21} & 0 & 0 & a_{22} & 0 & d_{23} \\ 0 & 0 & 0 & 0 & a_{23} & 0 \\ d_{31} & 0 & 0 & d_{32} & 0 & a_{33} \end{pmatrix} \quad \text{and} \quad B_1 = \begin{pmatrix} b_{11} & 0 & 0 & e_{12} & 0 & e_{13} \\ 0 & b_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & b_{13} & 0 & 0 & 0 \\ e_{21} & 0 & 0 & b_{22} & 0 & e_{23} \\ 0 & 0 & 0 & 0 & b_{23} & 0 \\ e_{31} & 0 & 0 & e_{32} & 0 & b_{33} \end{pmatrix}.$$
In expanded form, the covariance equations are:

\[
\begin{align*}
    h_{11,t} &= c_{11} + a_{11} \varepsilon_{1,t-1}^2 + b_{11} h_{11,t-1} + d_{12} \varepsilon_{2,t-1}^2 + e_{12} h_{22,t-1} + d_{13} \varepsilon_{3,t-1}^2 + e_{13} h_{33,t-1}, \\
    h_{12,t} &= c_{12} + a_{12} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + b_{12} h_{12,t-1}, \\
    h_{13,t} &= c_{13} + a_{13} \varepsilon_{1,t-1} \varepsilon_{3,t-1} + b_{13} h_{13,t-1}, \\
    h_{22,t} &= c_{22} + a_{22} \varepsilon_{2,t-1}^2 + b_{22} h_{22,t-1} + d_{21} \varepsilon_{1,t-1}^2 + e_{21} h_{11,t-1} + d_{23} \varepsilon_{3,t-1}^2 + e_{23} h_{33,t-1}, \\
    h_{23,t} &= c_{23} + a_{23} \varepsilon_{2,t-1} \varepsilon_{3,t-1} + b_{23} h_{23,t-1}, \\
    h_{33,t} &= c_{33} + a_{33} \varepsilon_{3,t-1}^2 + b_{33} h_{33,t-1} + d_{31} \varepsilon_{1,t-1}^2 + e_{31} h_{11,t-1} + d_{32} \varepsilon_{2,t-1}^2 + e_{32} h_{22,t-1}.
\end{align*}
\]

We see that each covariance equation is equivalent to that from a DVECH(1, 1) model, while the variance equations include extra terms to capture volatility and squared shock spillovers between series.

The general model in equations (3.14) and (3.15) can be extended in a straightforward manner to include higher order mean and/or variance lags, non-normal residuals or other effects (e.g. volatility asymmetry). The methods and prior proposed in this chapter
extend easily to those cases, some of which are dealt with later in this thesis.

### 3.3.1 Unconditional Properties and the Likelihood Function

The lagged structure of the model means initial values must be set for both mean and covariance equations under the conditional likelihood approach. These are chosen to be set as the unconditional expectations from the model: Let \( \mu_1 \) and \( H_1 \) denote the unconditional mean vector and covariance matrix respectively, and define \( I_k \) as the \( k \times k \) identity matrix. The unconditional mean vector is derived from (3.14) as follows:

\[
\mu_1 = E(y_t) = E(m_0 + \text{diag}(m_1)y_{t-1} + \epsilon_t)
= m_0 + \text{diag}(m_1)E(y_t) \quad \text{(by stationarity)}
\Rightarrow \mu_1 = \left[I_N - \text{diag}(m_1)\right]^{-1}m_0. \tag{3.16}
\]

To derive the unconditional covariance matrix, we begin with

\[
\text{vech}(H_1) = \text{vech}[\text{Cov}(y_t)]
= \text{vech}[\text{Cov}(m_0 + \text{diag}(m_1)y_{t-1} + \epsilon_t)]
= \text{diag}[\text{vech}(m_1m_1')]\text{vech}[\text{Cov}(y_t)] + \text{vech}[\text{Cov}(\epsilon_t)] \quad \text{(by stationarity)}
\Rightarrow \text{vech}(H_1) = \left\{I_N^* - \text{diag}[\text{vech}(m_1m_1')]\right\}^{-1}\text{vech}[\text{Cov}(\epsilon_t)] \tag{3.17}
\]

where \( N^* = N(N+1)/2 \). Standard univariate GARCH theory can be extended to obtain:

\[
\text{vech}[\text{Cov}(\epsilon_t)] = \text{vech}[E(\epsilon_t\epsilon'_t)]
= \text{vech}\{E[E(\epsilon_t\epsilon'_t | \psi_{t-1})]\}
= \text{vech}\{E[\text{Cov}(\epsilon_t | \psi_{t-1})]\} = \text{vech}[E(H_t)]. \tag{3.18}
\]
For any square matrix $X = [x_{ij}]$ of order $N$, we have $\text{vech}[E(X)] = E[\text{vech}(X)]$ since

$$
\text{vech}[E(X)] = \text{vech} \left( \begin{pmatrix} E(x_{11}) & E(x_{12}) & \cdots & E(x_{1N}) \\
E(x_{21}) & E(x_{22}) & \cdots & E(x_{2N}) \\
\vdots & \vdots & \ddots & \vdots \\
E(x_{N1}) & E(x_{N2}) & \cdots & E(x_{NN}) \end{pmatrix} \right) \\
= [E(x_{11}) \ E(x_{21}) \ \cdots \ E(x_{N1}) \ E(x_{22}) \ \cdots]'
$$

$$
= E[\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{N1} & x_{22} & \cdots \end{pmatrix}']
= E[\text{vech}(X)].
$$

(3.19)

Therefore

$$
E[\text{vech}(H_t)] = E[\text{vech}(C) + A_1 \text{vech}(\varepsilon_{t-1}\varepsilon_{t-1}') + B_1 \text{vech}(H_{t-1})] \\
= \text{vech}(C) + A_1 E[\text{vech}(\varepsilon_t\varepsilon_t')] + B_1 E[\text{vech}(H_t)] \quad \text{(by stationarity)}
$$

$$
\Leftrightarrow \quad \text{vech}[\text{Cov}(\varepsilon_t)] = \text{vech}(C) + A_1 \text{vech}[\text{Cov}(\varepsilon_t)] + B_1 \text{vech}[\text{Cov}(\varepsilon_t)]
$$

(3.18)

$$
\quad \text{(by eqn's (3.18) and (3.19))}
$$

$$
\Rightarrow \quad \text{vech}[\text{Cov}(\varepsilon_t)] = (I_N - A_1 - B_1)^{-1}\text{vech}(C)
$$

and by using equation (3.17) we obtain the unconditional covariance matrix:

$$
\text{vech}(H_1) = \{I_N - \text{diag}[\text{vech}(m_1m_1')]\}^{-1}(I_N - A_1 - B_1)^{-1}\text{vech}(C).
$$

(3.20)

We then define the initial residual vector as $\varepsilon_1 = y_1 - \mu_1$.

Define the parameter vectors $m = [m_{ij}]$, $c = [c_{ij}]$, $a = [a_{ij}]$ and $b = [b_{ij}]$. As the residuals follow a multivariate normal distribution, the likelihood function is the product of $n$ multivariate normal probability density functions (PDF)s:

$$
p(y|m, c, a, b) = (2\pi)^{-\frac{nN}{2}} \prod_{t=1}^{\frac{n}{2}} |H_t|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \varepsilon_t' H_t^{-1} \varepsilon_t \right)
$$
with the log likelihood function taking the form

$$\ln p(y|m, c, a, b) = -\frac{Nn}{2}\ln(2\pi) - \frac{1}{2} \sum_{t=1}^{n} \left( \ln |H_t| + \varepsilon_t'H_t^{-1}\varepsilon_t \right).$$

The off-diagonal parameters $d_{ij}$ and $e_{ij}$ are also included in vectors $a$ and $b$, since they are still elements of the matrices $A_1$ and $B_1$.

### 3.3.2 Usual Parameter Restrictions

To enforce the PVECH model in (3.14) and (3.15) to be covariance stationary, the standard approach is to use a set of sufficient conditions through parameter restrictions. As it is a special case of the VECH($p, q$) model, a set of restrictions is given in (3.3)-(3.5), and for the autoregressive part of the model:

All eigenvalues of $\text{diag}(m_1)$ are less than one in modulus.

Examples of this can be seen in Silberberg and Pafka (2001) or Chen et al. (2005). We must also have positive definite conditional covariance matrices for each time point. However, there are no known conditions on parameters that guarantee PD for VECH-type models.

The conditions in (3.3)-(3.5) are only sufficient and are not necessary in general; other parameter values lying outside these restrictions can also achieve PD and CS. As an example, consider a simple bivariate DVECH(1, 1) model with the covariance equations

$$
\begin{align*}
h_{11,t} &= c_{11} + a_{11}\varepsilon_{1,t-1}^2 + b_{11}h_{11,t-1}, \\
h_{12,t} &= c_{12} + a_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + b_{12}h_{12,t-1}, \\
h_{22,t} &= c_{22} + a_{22}\varepsilon_{2,t-1}^2 + b_{22}h_{22,t-1}.
\end{align*}
$$
Using equation (3.20), the unconditional covariances for this model are

$$\text{vech}(E(H_t)) = \left( \frac{c_{11}}{1 - a_{11} - b_{11}}, \frac{c_{12}}{1 - a_{12} - b_{12}}, \frac{c_{22}}{1 - a_{22} - b_{22}} \right).$$  \hspace{1cm} (3.21)

According to the sufficient conditions in (3.4), all $a$'s and $b$'s must be positive. But if we were to set (for example)

$$(c_{11}, a_{11}, b_{11}) = (0.5, -0.05, 0.5),$$
$$(c_{12}, a_{12}, b_{12}) = (0.05, 0.05, 0.05),$$
$$(c_{22}, a_{22}, b_{22}) = (0.5, -0.05, 0.5),$$

calculating the unconditional covariances in (3.21) gives

$$\text{vech}(E(H_t)) = (0.9091, 0.0556, 0.9091).$$

So finite unconditional positive variances are obtained and $E(H_t)$ has positive eigenvalues: CS is achieved. Now for PD, each matrix $H_t$ is required to be PD, for $t = 1, \ldots, n$. This is a conditional property, which may or may not be violated for any realisation of data from this model. As a test, 1,000 replications of data were simulated from this model, with $n = 1,000$: 962 of these data sets achieved PD. Any one of these could correspond to the realisation of a real dataset, as noted by Nelson and Cao (1992) and Gourieroux (1997) for univariate GARCH models.

This example illustrates that the usual sufficient conditions in (3.3)-(3.5) are not in general necessary, implying that we can widen the allowable parameter space for multivariate GARCH models while still achieving PD and CS for the data at hand. The question is to what degree? Consider $a_{11}$ in the example above: How far negative can $a_{11}$ be before the CS or positive variance criteria are violated? This depends on what the other parameter values and the observed sample data are, but if $a_{11} < 0$, then the rest of the parameters need to be large enough so that a positive $h_{11,t}$ can be obtained.
and PD can hold for all $t$, while maintaining CS. Hence a set of more general restrictions are desirable. The definitions of PD and CS will now be considered and employed in a Bayesian prior distribution.

### 3.3.3 Proposed Model Restrictions

Below are the conditions of PD and CS:

- **Positive Definiteness:** Each conditional variance value and all eigenvalues of each conditional covariance matrix are strictly positive:

  $$h_{ii,t} > 0 \text{ and } \text{eigs}(H_t) > 0, \quad \forall \ t = 1, \ldots, n; \quad i = 1, \ldots, N.$$  
  (3.22)

- **Covariance Stationarity:** The unconditional mean and covariance matrices exist and are finite and positive definite:

  $$\mu_1, H_1 \text{ are finite; } \text{eigs}(H_1) > 0.$$  
  (3.23)

Because the definitions (3.22)-(3.23) do not place direct restrictions on the parameters themselves, the parameter space under these definitions has potentially increased over the usual sufficient conditions. For example, we can now have negative $a_{11}$ values as long as each $h_{11,t} > 0$, PD is achieved at each $t$ and $\mu_1, H_1$ are finite. These definitions are exactly the necessary and sufficient conditions for PD and CS. In the next section, these conditions will define the space of the prior distribution in a Bayesian analysis of the PVECH model. Note that the definition of PD in (3.22) is inherently conditional upon the data at hand, as long as each estimated $H_t$ is PD then the definition is completely satisfied for that dataset. It is irrelevant that some other realisation of data using the same parameter values may give rise to non PD matrices, because there are no conditions on parameter values that guarantee PD will be satisfied in all realisations from VECH models. Instead, the conditions (3.22)-(3.23) will be enforced directly in parameter estimation via the prior
distribution, and thus always satisfy PD and CS for every data set under analysis. The
generality of this approach also means that it has the advantage of being fully applicable
to any multivariate GARCH model requiring the monitoring of PD, including the full
VECH model from Section 3.2.

3.4 Bayesian Estimation Procedures

Estimation and inference will now be considered for the PVECH model defined by
(3.14) and (3.15) from a Bayesian perspective, using the joint posterior distribution
\( p(m, c, a, b|y) \). There is no straightforward way to search for parameter estimates that
satisfy the conditions (3.22)-(3.23) in a classical framework. Neither are the restrictions
on parameters known (in general) that are both necessary and sufficient for (3.22)-(3.23).

An advantage of the Bayesian approach here is that conditions (3.22)-(3.23) can form
(part of) the prior distribution, in a straightforward manner, allowing these conditions to
be both enforced and covered. In particular, under the MCMC approach, parameter val-
ues can be proposed and then accepted or rejected depending on whether they satisfy any
set of conditions. This simply constrains the posterior to the required parameter space
under the desired restrictions. This flexibility allows valid and efficient inference under
conditions (3.22)-(3.23), allowing a larger portion of the parameter space to be explored
than standard approaches. Such an approach will be taken in this chapter, employing the
Metropolis-Hastings (MH) algorithm from Section 2.2 combined with the Gibbs sampler
from Section 2.5. MCMC methods (but not this prior) have commonly been applied to
univariate GARCH-type models in the literature, see for example Geweke (1989), Chen

3.4.1 The Gibbs Sampler

As discussed in Section 2.5, the Gibbs sampler requires conditional posterior distributions,
set up to form a Markov chain, in order to obtain a dependent Monte Carlo sample
from the joint posterior \( p(m, c, a, b | y) \). This will allow numerical integration to estimate marginal posterior distributions for each parameter. The concept of “blocking” was also mentioned in Section 2.5 by Carter and Kohn (1994), where parameters are combined so that samples are taken from joint conditional posteriors. This reduces the correlation between MCMC iterates and speeds up convergence of the chain. This method will be employed for the PVECH model and an approach similar to that in Chen et al. (2005) for univariate GARCH models will be taken: Let \( \theta_{h_{ij,t}} \), \( i \leq j = 1, \ldots, N \) denote the vector of parameters contained in covariance equation \( h_{ij,t} \). The following joint conditional posterior distributions will be used to obtain a sample from \( p(m, c, a, b | y) \):

\[
\begin{align*}
1. & \quad p(m | y, c, a, b) \\
2. & \quad p(\theta_{h_{11,t}} | y, \theta_{\neq h_{11,t}}) \\
... & \quad : \\
N + 1. & \quad p(\theta_{h_{NN,t}} | y, \theta_{\neq h_{NN,t}}) \\
N + 2. & \quad p(\theta_{h_{12,t}} | y, \theta_{\neq h_{12,t}}) \\
... & \quad : \\
N(N + 1)/2 + 1. & \quad p(\theta_{h_{(N-1)N,t}} | y, \theta_{\neq h_{(N-1)N,t}}) 
\end{align*}
\]

Step 1 contains the autoregressive mean parameters, steps 2 to \( N + 1 \) contain the variance equation parameters and steps \( N + 2 \) to \( N(N + 1)/2 + 1 \) contain the covariance equation parameters. We simulate iteratively in turn from the above distributions a large number of times (\( J \) say) to obtain the MCMC sample \( \{m[j], c[j], a[j], b[j]\}_{j=1}^J \).

### 3.4.2 Prior and Posterior Distributions

In a Bayesian framework prior distributions must be placed on the model parameters. Here the prior is chosen to be flat or non-informative over the complete parameter range allowed by conditions (3.22)-(3.23), so that the observed data and likelihood dominate
inference. Even though we cannot explicitly write down this prior in terms of parameter values, it can easily be enforced in an MCMC analysis by simply rejecting any parameter values where (3.22)-(3.23) are not satisfied. The prior then implicitly constrains the posterior to lie only on the region defined by (3.22)-(3.23). Thus PD and CS are guaranteed, however no unnecessary parameter constraints are enforced.

Let $\theta_i, i = 1, \ldots, N(N + 1)/2 + 1$ denote the parameters in each joint conditional posterior above. The uniform priors to be used are:

$$p(\theta_i|\theta_{\neq i}) \propto I(\Theta), \quad i = 1, \ldots, N(N + 1)/2 + 1,$$

where $\Theta$ is the parameter space and $I$ indicates whether the parameters satisfy the conditions (3.22)-(3.23). To obtain posterior distributions, we apply Bayes’ rule:

$$p(\theta_i|y, \theta_{\neq i}) \propto p(y|\theta_i, \theta_{\neq i})p(\theta_i|\theta_{\neq i}), \quad i = 1, \ldots, N(N + 1)/2 + 1. \quad (3.24)$$

As each prior is uniform over (3.22)-(3.23), each posterior is proportional to the likelihood constrained to lie in the parameter space allowed by the conditions (3.22)-(3.23). These posteriors are not of a standard known form, as such we turn to MH methods.

### 3.4.3 Conditional Posterior Sampling

As discussed in Section 2.2, the Metropolis-Hastings (MH) algorithm is a rejection sampling method that allows simulation from distributions of unknown form. The MH algorithm, along with a delayed rejection (DR) step (see Section 2.3), will be used to simulate from each conditional posterior in the sampling scheme.

In order to describe the algorithm in detail, it is necessary to separate it into the following two parts:
The Burn-In Period

The period where conditional posterior samples are drawn for iterations \( j = 1, \ldots, w \), where \( 1 \leq w < J \), is called the \textit{burn-in} period. This period assists the Markov chain to converge to the target distribution.

As stated in Section 2.2, begin with some arbitrary starting values \( \theta_i^{[1]}, i = 1, \ldots, N(N+1)/2 + 1 \). For burn-in iterations \( j = 1, \ldots, w \) do the following:

1. Draw a random sample \( \theta_i^{[p1]} \) from the proposal distribution:

   \[ g_1(\theta_i | \theta_i^{[j-1]}) \sim N(v_i, c_i^{[j]} \Omega_i) \]

   i.e. a multivariate normal distribution with mean vector \( v_i \) and covariance matrix \( c_i^{[j]} \Omega_i \). Proposal means \( v_i \) can be set to the previous iterate (i.e. a random walk process) or an estimate such as that given by least squares or other method. The matrix \( \Omega_i \) is a diagonal matrix with “large” diagonal entries\(^1\), and \( c_i^{[j]} \) is a real number initially set to 1.

2. Draw a random sample \( u_i^{[j]} \) from \( \text{Unif}[0,1] \) and set the following MH acceptance probability:

   \[
   \alpha_1(\theta_i^{[j-1]}, \theta_i^{[p1]}) = \min \left\{ 1, \frac{p(\theta_i^{[p1]} | y, \theta_{\neq i}) g_1(\theta_i^{[j-1]} | \theta_i^{[p1]})}{p(\theta_i^{[j-1]} | y, \theta_{\neq i}) g_1(\theta_i^{[p1]} | \theta_i^{[j-1]})} \right\}. \quad (3.25)
   \]

3. If \( u_i^{[j]} < \alpha_1 \), set \( \theta_i^{[j]} = \theta_i^{[p1]} \) and return to Step 1 for the next \( i \). Otherwise, perform a delayed rejection step:

4. Draw a random sample \( \theta_i^{[p2]} \) from the proposal distribution:

   \[ g_2(\theta_i | \theta_i^{[j-1]}, \theta_i^{[p1]}) \sim N \left( \frac{1}{2}(v_i + \theta_i^{[j-1]}), \frac{1}{2} c_i^{[j]} \Omega_i \right) \]

---

\(^1\)Because the diagonal entries of a covariance matrix represent variances, the meaning of “large” in this context refers to these variances being large enough to cover as much of the range of each parameter as possible. A trial-and-error approach is typically applied to determine these values, since each model parameter can take on many different values.
5. Draw a random sample \( u_2^j \) from Unif[0, 1] and set the following MH acceptance probability:

\[
\alpha_2(\theta_i^{[j-1]}, \theta_i^{p1}, \theta_i^{p2}) = \min\left\{ 1, \frac{N_2}{D_2} \right\}
\]

where

\[
\frac{N_2}{D_2} = \frac{p(\theta_i^{p2} | y, \theta_{\neq i}) g_1(\theta_i^{p1} | \theta_i^{p2}) g_2(\theta_i^{[j-1]} | \theta_i^{p2}, \theta_i^{p1}) [1 - \alpha_1(\theta_i^{p2}, \theta_i^{p1})]}{p(\theta_i^{[j-1]} | y, \theta_{\neq i}) g_1(\theta_i^{p1} | \theta_i^{[j-1]}) g_2(\theta_i^{p2} | \theta_i^{[j-1]}, \theta_i^{p1}) [1 - \alpha_1(\theta_i^{[j-1]}, \theta_i^{p1})]}.
\]

6. If \( u_2^j < \alpha_2 \), set \( \theta_i^j = \theta_i^{p2} \). Otherwise, set \( \theta_i^j = \theta_i^{[j-1]} \). Return to Step 1 for the next \( i \).

Note the inclusion of the constant \( c_i^j \) with the proposal covariance matrix. This is important to achieving desirable acceptance rates (AR) for each of the \( N(N+1)/2 + 1 \) parameter groupings: At the end of every \( k \)th iteration (for a pre-specified \( k \)), calculate the AR from the previous \( k \) iterations as follows

\[
AR_i^j = \frac{1}{k} \sum_{j-k+1}^{j} I(\theta_i^j = \theta_i^{p1} \cup \theta_i^j = \theta_i^{p2}) \quad \forall \ j = k, 2k, \ldots, w-k.
\]

For the next set of \( k \) iterations set the following values for \( c_i^{[j+1]} \), \( \forall \ j = k, 2k, \ldots, w-k \):

\[
c_i^{[j+1]} = \begin{cases} 
\left( \frac{1}{2} + \frac{j}{2w} \right) c_i^j & \text{if } AR_i^j < 0.15, \\
\frac{1}{2} c_i^j & \text{if } 0.15 \leq AR_i^j \leq 0.5, \\
\left( 2 - \frac{j}{w} \right) c_i^j & \text{if } AR_i^j > 0.5.
\end{cases}
\]

This process is one form of variance tuning, as the proposal variances in \( \Omega_i \) are increased / decreased by a factor of \( c_i^j \) after \( k \) iterations have been completed. Other examples of tuning proposal variances can be found in the literature, some examples are Haario et al. (2001) and Roberts and Rosenthal (2009). The algorithm above has been designed to allow for more aggressive tuning to take place during early iterations (where knowledge
is minimal about the ideal proposal variance) and less aggressive as the burn-in period approaches completion. It also enables the AR during the burn-in period to be controlled so that it lies between two threshold values, here 0.15 and 0.5. These threshold values were chosen to give the final ARs a greater chance of reaching the optimal AR of 0.234 by Roberts et al. (1997) as mentioned in Section 2.2.

The Sampling Period

Once iteration \( w \) has been completed, the sampling period begins, which is where inference is made on the parameters of the PVECH model. The proposal distributions are now modified slightly to reflect the information contained in the burn-in period: Let \( \hat{\theta}_i \) and \( \hat{\Sigma}_i \) denote the sample means and covariances of the iterations in the burn-in period for each parameter group \( i \). For iterations \( j = w + 1, \ldots, J \) do the following:

1. Draw a random sample \( \theta_i^{p_1} \) from the proposal distribution:
   \[
   g_1(\theta_i | \theta_i^{j-1}) \sim N(\hat{\theta}_i, \hat{\Sigma}_i).
   \]

2. Draw a random sample \( u_{i1}^{[j]} \) from \( \text{Unif}[0, 1] \) and set the MH acceptance probability as in equation (3.25).

3. If \( u_{i1}^{[j]} < \alpha_1 \), set \( \theta_i^{p_1} = \theta_i^{p_i} \) and return to Step 1 for the next \( i \). Otherwise, perform a delayed rejection step:

4. Draw a random sample \( \theta_i^{p_2} \) from the proposal distribution:
   \[
   g_2(\theta_i, \theta_i^{j-1}, \theta_i^{p_i}) \sim N\left(\frac{1}{2}(\theta_i^{j-1} + \theta_i^{j-1}), \frac{1}{2}\hat{\Sigma}_i\right).
   \]

5. Draw a random sample \( u_{i2}^{[j]} \) from \( \text{Unif}[0, 1] \) and set the MH acceptance probability as in equation (3.26).
6. If \( u_2^{[j]} < \alpha_2 \), set \( \theta_i^{[j]} = \theta_i^{p_2} \). Otherwise, set \( \theta_i^{[j]} = \theta_i^{[j-1]} \). Return to Step 1 for the next \( i \).

The set of simulations \( \left\{ \theta_i^{[w+1]}, \ldots, \theta_i^{[j]} \right\}_{i=1}^{N(N+1)/2+1} \) is then a dependent sample from the joint posterior distribution \( p(m, c, a, b | y) \) of the PVECH model. This technique is known as \textit{adaptive MCMC sampling}, as the proposal distributions are modified during the algorithm. Chen \textit{et al.} (2005) illustrated the favourable performance of this technique for a non-linear univariate GARCH model.

### 3.5 Simulation Study

The methods in Sections 3.3 and 3.4 will now be empirically examined via a simulation study. Results from the proposed MCMC method on a bivariate PVECH model \((N = 2)\) will be compared with those obtained by applying standard large sample classical maximum likelihood theory to an equivalent bivariate BEKK model. Data from two parameter sets is to be simulated with sample size \( n = 1,000 \) and each estimated using the Bayesian methods from Section 3.4, and the BEKK model via maximum likelihood. This procedure will be replicated 1,000 times to examine the properties of estimators such as bias and coverage of confidence intervals. Finally, two simulated data sets of length \( n = 50,000 \) will be analysed in the same way, to illustrate the consistency of estimation for each method.

#### 3.5.1 The Bivariate PVECH and BEKK Models

Assuming that \( \varepsilon_t | \psi_{t-1} \sim N(0, H_t) \), the bivariate PVECH model is:

\[
\begin{align*}
y_{1t} &= m_{10} + m_{11} y_{1,t-1} + \varepsilon_{1t}, \\
y_{2t} &= m_{20} + m_{22} y_{2,t-1} + \varepsilon_{2t},
\end{align*}
\]
\[ h_{11,t} = c_{11} + a_{11} \varepsilon^2_{1,t-1} + b_{11} h_{11,t-1} + d_{12} \varepsilon^2_{2,t-1} + e_{12} h_{22,t-1}, \]
\[ h_{12,t} = c_{12} + a_{12} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + b_{12} h_{12,t-1}, \]
\[ h_{22,t} = c_{22} + a_{22} \varepsilon^2_{2,t-1} + b_{22} h_{22,t-1} + d_{21} \varepsilon^2_{1,t-1} + e_{21} h_{11,t-1}. \]

Data from this model is to be simulated using two sets of parameter values:

- **Model 1**: This model has parameters that lie inside the usual parameter restrictions:
  
  \[
  (m_{10}, m_{20}, m_{11}, m_{22}) = (0.01, 0.01, 0.2, 0.2), \\
  (c_{11}, a_{11}, b_{11}, d_{12}, e_{12}) = (0.02, 0.3, 0.4, 0.05, 0.05), \\
  (c_{22}, a_{22}, b_{22}, d_{21}, e_{21}) = (0.01, 0.3, 0.225, 0.05, 0.05), \\
  (c_{12}, a_{12}, b_{12}) = (0.005, 0.35, 0.35). 
  \]

- **Model 2**: This model has parameters that lie outside the usual parameter restrictions but still permit stationary and positive definite realisations:
  
  \[
  (m_{10}, m_{20}, m_{11}, m_{22}) = (0.1, 0.15, 0.3, 0.25), \\
  (c_{11}, a_{11}, b_{11}, d_{12}, e_{12}) = (0.2, 0.36, 0.04, 0.09, 0), \\
  (c_{22}, a_{22}, b_{22}, d_{21}, e_{21}) = (-0.01, 0.16, 0.64, -0.01, 0.1), \\
  (c_{12}, a_{12}, b_{12}) = (0.01, 0.21, 0.16). 
  \]

These parameter values were chosen for a number of reasons:

- The parameters in Model 1 lie well inside the usual parameter restrictions for VECH models, with some typically obtained from real financial return data.

- To demonstrate the exploration of a wider parameter space during estimation, some of the parameters in Model 2 were chosen to lie outside the usual parameter restrictions (e.g. \( c_{22}, d_{21} < 0 \)). In fact, variance intercept parameters \( c_{ii} \) are typically...
restricted to be strictly positive in the literature. This is a reasonable assumption
for simple models (e.g. the DVECH model) since having zero or negative values for
this parameter would typically result in zero or negative unconditional variances.
However, this does not affect the PVECH model, as demonstrated by the PD and
CS calculations below.

• In order to improve the estimation efficiency of the equivalent BEKK model to be
estimated on the same data, some of the covariance equation parameters in Models
1 and 2 were chosen to satisfy the relationship between corresponding variance
equation parameters (see equations (3.27) and (3.28) to follow).

To be certain that both sets of parameter values satisfy PD and CS, the unconditional
covariances of the model need to be examined - using equation (3.20) it can be shown
that

\[
\begin{align*}
\text{Var}(y_{1t}) & = \frac{c_{11}(1 - a_{22} - b_{22}) + c_{22}(d_{12} + e_{12})}{[(1 - a_{11} - b_{11})(1 - a_{22} - b_{22}) - (d_{12} + e_{12})(d_{21} + e_{21})](1 - m_{11}^2)}, \\
\text{Cov}(y_{1t}, y_{2t}) & = \frac{c_{12}}{(1 - a_{12} - b_{12})(1 - m_{11}m_{22})}, \\
\text{Var}(y_{2t}) & = \frac{c_{22}(1 - a_{11} - b_{11}) + c_{11}(d_{21} + e_{21})}{[(1 - a_{11} - b_{11})(1 - a_{22} - b_{22}) - (d_{12} + e_{12})(d_{21} + e_{21})](1 - m_{22}^2)}. \\
\end{align*}
\]

For Model 1, we have

\[
\text{vech}(H_1) = (0.0825, 0.0174, 0.0393), \quad \text{eigs}(H_1) = (0.0887, 0.0332)
\]

and for Model 2, we have

\[
\text{vech}(H_1) = (0.3840, 0.0172, 0.1144), \quad \text{eigs}(H_1) = (0.3851, 0.1133).
\]

Therefore models with the above parameter values satisfy PD and CS, even when some
parameters lie outside the usual restrictions.
The corresponding bivariate BEKK model will be fitted to the simulated datasets from Models 1 and 2 through maximum likelihood. The BEKK orders were chosen so that vech\((H_t)\) has exactly the same form as the bivariate PVECH model, giving an AR(1)-BEKK(1, 1, 2) model, which has the form:

\[
y_t = m_0^* + \text{diag}(m_t^*)y_{t-1} + \epsilon_t, \quad \epsilon_t | \psi_{t-1} \sim N(0, H_t)
\]

\[
H_t = C^*C^* + \sum_{k=1}^{2} A^*_{1k}\epsilon_{t-1}\epsilon_{t-1}'A_{1k}^* + \sum_{k=1}^{2} B_{1k}^*H_{t-1}B_{1k}^*
\]

where

\[
C^* = \begin{pmatrix} c_{11}^* & c_{12}^* \\ 0 & c_{22}^* \end{pmatrix}, \quad A_{11}^* = \begin{pmatrix} a_{11}^* & 0 \\ 0 & a_{22}^* \end{pmatrix}, \quad A_{12}^* = \begin{pmatrix} 0 & a_{12}^* \\ a_{21}^* & 0 \end{pmatrix},
\]

\[
B_{11}^* = \begin{pmatrix} b_{11}^* & 0 \\ 0 & b_{22}^* \end{pmatrix}, \quad B_{12}^* = \begin{pmatrix} 0 & b_{12}^* \\ b_{21}^* & 0 \end{pmatrix}.
\]

In expanded form, this model becomes

\[
y_{1t} = m_{10}^* + m_{11}^*y_{1,t-1} + \epsilon_{1t},
\]

\[
y_{2t} = m_{20}^* + m_{21}^*y_{2,t-1} + \epsilon_{2t},
\]

\[
h_{11,t} = c_{11}^* + a_{11}^*\epsilon_{1,t-1}^2 + b_{11}^2h_{11,t-1} + a_{21}^2\epsilon_{2,t-1}^2 + b_{21}^2h_{22,t-1}, \quad (3.27)
\]

\[
h_{12,t} = c_{12}^* + (a_{11}^*a_{22}^* + a_{12}^*a_{21}^*)\epsilon_{1,t-1}\epsilon_{2,t-1} + (b_{11}^*b_{22}^* + b_{12}^*b_{21}^*)h_{12,t-1},
\]

\[
h_{22,t} = c_{22}^* + a_{22}^2\epsilon_{2,t-1}^2 + b_{22}^2h_{22,t-1} + a_{12}^2\epsilon_{1,t-1}^2 + b_{12}^2h_{11,t-1}.
\]

These PVECH and BEKK models are now equivalent. They only differ in parameter coefficients, which can be equated as follows:

\[
m_{10} \equiv m_{10}^*, \quad m_{11} \equiv m_{11}^*, \quad m_{20} \equiv m_{20}^*, \quad m_{22} \equiv m_{22}^*;
\]

\[
c_{11} \equiv c_{11}^*, \quad a_{11} \equiv a_{11}^*, \quad b_{11} \equiv b_{11}^*, \quad d_{12} \equiv a_{21}^*, \quad e_{12} \equiv b_{21}^*;
\]

\[
c_{12} \equiv c_{12}^*, \quad a_{12} \equiv a_{12}^* + a_{12}^*a_{21}^*, \quad b_{12} \equiv b_{11}^*b_{22} + b_{12}^*b_{21}^*;
\]

\[
c_{22} \equiv c_{22}^* + c_{22}^2, \quad a_{22} \equiv a_{22}^2, \quad b_{22} \equiv b_{22}^2, \quad d_{21} \equiv a_{12}^*, \quad e_{21} \equiv b_{12}^*; \quad (3.28)
\]
Once the BEKK model is estimated, we make the transformations above for comparison; the Delta method (see Oehlert, 1992) is then used to make inferences on the parameters. The following subsection contains further details on this procedure.

### 3.5.2 Maximum Likelihood on the BEKK Model

Let \( \theta^* = \{m^*, c^*, a^*, b^*\} \) denote the parameters in the AR(1)-BEKK(1, 1, 2) model above. The maximum likelihood estimates of these parameters are given by

\[
\hat{\theta}^* = \arg \max_{\theta^*} \sum_{t=1}^{n} \ln p(y_t|m^*, c^*, a^*, b^*)
\]

where \( p(y_t|m^*, c^*, a^*, b^*) \) denotes the bivariate normal distribution:

\[
p(y_t|m^*, c^*, a^*, b^*) = \frac{1}{2\pi|H_t|^\frac{1}{2}} \exp \left( -\frac{1}{2} \varepsilon_t' H_t^{-1} \varepsilon_t \right).
\]

The initial values for the estimation are set as \( \varepsilon_1 = y_1 - \bar{y} \) and \( H_1 = \bar{H} \), where \( \bar{y} \) and \( \bar{H} \) denote the sample mean vector and covariance matrix (respectively) of the data \( y \). The remaining values of \( \varepsilon_t \) and \( H_t \) then follow the BEKK equations as given by the equations contained in (3.27).

To perform the estimation, the Fortran optimisation routine UMINF is utilised from the IMSL Fortran Library v5.0. This routine minimises a multivariate function using a Quasi-Newton method to converge to the minimum: Define \( f \) as follows

\[
f(\theta^*) = -\sum_{t=1}^{n} \ln p(y_t|m^*, c^*, a^*, b^*)
\]

i.e. the negative log-likelihood function. Given a starting point \( \theta^*_{[1]} \), estimate the gradient function \( g_{[1]} \) (i.e. vector of first-order derivatives) using a finite difference approximation and set the matrix \( B_{[1]} \) equal to some positive definite matrix (e.g. the identity matrix). (This is called the Hessian matrix and is designed to converge to a matrix of second-order derivatives - see below). For each iteration \( j \), the algorithm proceeds as follows:
1. Calculate the search direction $\mathbf{d}_j = -B_{\mathbf{[j]}}^{-1} \mathbf{g}_j$.

2. Find the new point $\mathbf{\theta}_j^*$ via a line search, i.e.

$$
\mathbf{\theta}_{j+1}^* = \mathbf{\theta}_j^* + \lambda \mathbf{d}_j, \quad \lambda > 0
$$

such that

$$
f(\mathbf{\theta}_{j+1}^*) \leq f(\mathbf{\theta}_j^*) + \alpha \mathbf{g}_j' \mathbf{d}_j, \quad 0 < \alpha < 0.5.
$$

3. If convergence is achieved, then $\hat{\mathbf{\theta}}_* = \mathbf{\theta}_j^*$ is the solution and set $s = j$, the total number of iterations in the algorithm. Otherwise, increment $j$ by 1 and estimate $\mathbf{g}_{j+1}$ using finite differences and $B_{j+1}$ using the BFGS formula (e.g. see Broyden, 1970)

$$
B_{j+1} = B_j - \frac{B_j (\mathbf{\theta}_{j+1}^* - \mathbf{\theta}_j^*) (\mathbf{\theta}_{j+1}^* - \mathbf{\theta}_j^*)' B_j}{(\mathbf{\theta}_{j+1}^* - \mathbf{\theta}_j^*)' B_j (\mathbf{\theta}_{j+1}^* - \mathbf{\theta}_j^*)} + \frac{(\mathbf{g}_{j+1} - \mathbf{g}_j)' (\mathbf{g}_{j+1} - \mathbf{g}_j)}{(\mathbf{g}_{j+1} - \mathbf{g}_j)' (\mathbf{\theta}_{j+1}^* - \mathbf{\theta}_j^*)}
$$

and return to Step 1.

See Dennis and Schnabel (1983, Appendix A) for more details of the above algorithm.

UMINF is an unconstrained minimisation procedure in that it assumes that the parameters $\mathbf{\theta}^*$ can take any real value on $(-\infty, \infty)$. As mentioned in Section 3.2 and in Engle and Kroner (1995), parameters in BEKK models need to be restricted so that CS is achieved - for the AR(1)-BEKK(1,1,2) model above we require that

$$
\text{eigs} \left[ \sum_{k=1}^{2} (A_{1k}^* \otimes A_{1k}^*) + \sum_{k=1}^{2} (B_{1k}^* \otimes B_{1k}^*) \right] \text{ are less than one in modulus.} \quad (3.30)
$$

To include this restriction in the above algorithm, when a new $\mathbf{\theta}_j^*$ is generated in Step 2, the eigenvalues of the matrix in (3.30) are calculated. If any are greater than 1 in modulus, then $\mathbf{\theta}_j^*$ is re-calculated with a different value of $\lambda$ until the restriction is satisfied.

To transform these BEKK estimates into PVECH estimates for comparison, the Delta
method is applied which is an intuitive technique for approximating the moments of functions of random variables (Oehlert, 1992): Let \( g(\theta_i^*) \) denote the function defining the \( i \)th transformation in (3.28), for \( i = 1, \ldots, 17 \) (note that \( \theta_i^* \) can be either univariate or multivariate). The assumption is made that \( g(\theta_i^*) \) can be approximated by a first-order Taylor polynomial as follows\(^2\):

\[
g(\theta_i^*) \approx g(\hat{\theta}_i^*) + g'(\hat{\theta}_i^*) \odot (\theta_i^* - \hat{\theta}_i^*)
\]  

(3.31)

where \( g' \) is a vector of partial derivatives and the operator \( \odot \) is called the Hadamard product, defined as element-by-element matrix multiplication\(^3\). The corresponding PVECH estimates and their standard errors can be approximated by the first two moments of equation (3.31):

\[
E[g(\theta_i^*)] \approx g(\hat{\theta}_i^*),
\]  

(3.32)

\[
\text{Var}[g(\theta_i^*)] \approx [g'(\hat{\theta}_i^*)]^T \text{Cov}(\theta_i^*) g'(\hat{\theta}_i^*).
\]  

(3.33)

The expectation in equation (3.32) is obtained by simply transforming each BEKK estimate via (3.28), resulting in corresponding PVECH parameter estimates. To obtain corresponding PVECH standard errors, we refer to the final Hessian matrix \( B_{[s]} \) calculated in the estimation procedure above. The following is now performed:

1. Extract the block matrix \( B_{[s]}^i \) from \( B_{[s]} \).

2. Now \( \text{Cov}(\theta_i^*) \) is approximated by \( \text{Cov}(\hat{\theta}_i^*) \) which is given by the inverse of the block matrix calculated in Step 1:

\[
\text{Cov}(\hat{\theta}_i^*) = (B_{[s]}^i)^{-1}.
\]

\(^2I\) change notation temporarily here to deal with derivatives and matrix transposes appearing together: a derivative of \( f(x) \) is \( f'(x) \), while the transpose of matrix \( A \) is now \( A^T \). This will continue until the end of this subsection, when a matrix transpose will return to the \( A' \) notation.

\(^3F\)or matrices \( A = [a_{ij}] \) and \( B = [b_{ij}] \), the Hadamard product \( A \odot B \) has elements given by \( [a_{ij}b_{ij}] \).

57
3. Calculate \( \text{Var}[g(\hat{\theta}_i^*)] \) as given in equation (3.33).

4. The approximate standard error of \( g(\hat{\theta}_i^*) \) is given by

\[
\text{se}[g(\hat{\theta}_i^*)] \approx \sqrt{\text{Var}[g(\hat{\theta}_i^*)]}. 
\]

Hence 95\% confidence intervals for each \( g(\hat{\theta}_i^*) \) are then calculated as \( g(\hat{\theta}_i^*) \pm 1.96\text{se}[g(\hat{\theta}_i^*)], \)
\( i = 1, \ldots, 17. \)

### 3.5.3 The Results

A burn-in period of \( w = 5,000 \) iterations is chosen for the MCMC procedure, along with a sampling period of \( J - w = 20,000 \) iterations. Note that \( N(N + 1)/2 + 1 = 4 \), hence there are 4 steps in the MCMC sampling scheme. For all simulated datasets, elements of each proposal mean \( \mathbf{v}_i \) for the burn-in period only are chosen to correspond to either the previous MCMC iterate (i.e. a random walk MH framework), a function of the unconditional moments of the model or a pre-defined fixed value. Hence for iterations \( j = 1, \ldots, w \), proposal means are chosen as follows:

\[
\mathbf{v}_1 = (\hat{\bar{m}}_{10}, \hat{\bar{m}}_{20}, \hat{\bar{m}}_{11}, \hat{\bar{m}}_{22}) = \left( \bar{y}_1(1 - m_{11}^{[j-1]}), \bar{y}_2(1 - m_{22}^{[j-1]}), m_{11}^{[j-1]}, m_{22}^{[j-1]} \right), 
\]

\[
\mathbf{v}_2 = (\hat{\bar{c}}_{11}, \hat{\bar{a}}_{11}, \hat{\bar{b}}_{11}, \hat{\bar{d}}_{12}, \hat{\bar{b}}_{12}) = \left( c_{11}^{[j-1]}, a_{11}^{[j-1]}, b_{11}^{[j-1]}, 0, 0 \right), 
\]

\[
\mathbf{v}_3 = (\hat{\bar{c}}_{22}, \hat{\bar{a}}_{22}, \hat{\bar{b}}_{22}, \hat{\bar{d}}_{21}, \hat{\bar{e}}_{21}) = \left( c_{22}^{[j-1]}, a_{22}^{[j-1]}, b_{22}^{[j-1]}, 0, 0 \right), 
\]

\[
\mathbf{v}_4 = (\hat{\bar{c}}_{12}, \hat{\bar{a}}_{12}, \hat{\bar{b}}_{12}) = \left( \bar{\gamma}(1 - m_{11}^{[j]}m_{22}^{[j]})(1 - a_{12}^{[j-1]} - b_{12}^{[j-1]}), a_{12}^{[j-1]}, b_{12}^{[j-1]} \right) 
\]

where \( \bar{y}_k \) denotes the corresponding sample means within each dataset, and \( \bar{\gamma} \) denotes the sample covariance within each bivariate dataset. These equations are derived from the unconditional model moments shown earlier, with sample moments substituted for theoretical moments. Proposal covariance matrices are initially chosen as:

\[
c_1^{[j]} \Omega_1 = \text{diag} \left( \frac{2\sigma_1^2}{n}, \frac{2\sigma_2^2}{n}, \frac{1}{16}, \frac{1}{16} \right), \quad c_2^{[j]} \Omega_2 = c_3^{[j]} \Omega_3 = \text{diag} \left( \frac{1}{4900}, \frac{1}{49}, \frac{1}{49}, \frac{1}{100}, \frac{1}{100} \right),
\]

58
\[ c_4^{[j]} \Omega_4 = \text{diag} \left( \frac{1}{25}, \frac{1}{25}, \frac{1}{25} \right) \]

with \( s_k^2 \) defined as the sample variance of each series and each \( c_4^{[j]} \) initially set to 1, with the tuning algorithm applied as described in Section 3.4.

Tables 3.4-3.7 presented in the Appendix to this chapter contain the results from simulating 1,000 datasets from Models 1 and 2 with estimation using the MCMC methods in Section 3.4 and maximum likelihood (ML). Tables 3.4 and 3.6 contain the parameter estimates for both Bayesian and classical methods: the Bayesian results are averages of the 1,000 posterior means and medians, while the MLE columns contain the mean and median of the 1,000 BEKK estimates, transformed to match the estimates in the corresponding PVECH model. Tables 3.5 and 3.7 contain empirical coverages of 95% interval estimates (CI) across the simulations. Two Bayesian intervals are presented - the first by taking the 2.5th and 97.5th percentiles of the 1,000 posterior means for each parameter, labelled “P’tiles of Mean”; the second is labelled “Mean P’tiles” and is the average of the 95% credible intervals for each parameter over the 1,000 replications. The “Bayesian Coverage” column is the percentage of replications where the true parameter value lay within the MCMC 95% credible interval. Two ML intervals are also presented - the first by taking the 2.5th and 97.5th percentiles of the 1,000 transformed MLEs, while the second was calculated by the Delta method, as described in Section 3.5.2. Finally, the Delta coverage column is the percentage of times that the true parameter value lay within each Delta interval, over the 1,000 replications. The Bayesian and Delta coverage closest to the nominal 95% for each model have been written in **bold** text, while any coverages falling below 80% have been written in **red** text. The average run time to fit the PVECH model to each dataset via MCMC methods is approximately 1 min 15 sec, while fitting the BEKK model to each dataset takes on average 2 sec, using the machine and software described in Section 2.6.

For unbiased estimation, we expect the mean and median columns of Tables 3.4 and 3.6 to be “close” to the true parameter values. Table 3.4 shows this is the case for both Bayesian and ML results from Model 1. However, major differences between the
Bayesian and ML results can be seen in the CI and coverage columns of Table 3.5. While the Bayesian and ML estimates again seem unbiased, as each true value is contained inside the 95% intervals, the average Delta CIs are generally a lot narrower than the Bayesian intervals and the Delta coverages are quite poor compared to the nominal 95% level. All ML interval estimators in fact achieve below 85% coverage, some below 80% (as given in red), with an average coverage of 81% across all parameters in Model 1. The Bayesian interval estimates have achieved near nominal 95% coverage across all Model 1 parameters, the lowest being 84%, with average coverage of 93.3%. Note that three variance equation parameters in Table 3.5 have lower confidence limits that are negative for the Bayesian method, namely $d_{12}, e_{12}$ and $e_{21}$. Two of these limits are negative for the ML Delta method, but not for the intervals generated by the percentiles of the 1,000 MLEs. This is because in the transformation from the BEKK model the GARCH parameters in the variance equations have been squared (as shown above). While it is possible to have negative limits through the Delta method, the percentile CIs are strictly positive intervals.

Tables 3.6 and 3.7 display results for Model 2. The Bayesian estimates appear unbiased and close to their true values, these being inside the average Bayesian interval estimates. However the MLEs seem unbiased only when the true parameter values in the variance equation are strictly positive. When the true parameter value is zero or negative, as for $c_{22}$, $d_{21}$ and $e_{12}$, the ML results seem to display some bias, in the positive direction. The ML intervals estimates for the two negative parameters display extremely low coverage, close to 27% for a nominal 95% interval. Lower than nominal coverage is not restricted to these “unusual” parameters, however, with most ML nominal 95% interval coverages below 90%; the ML intervals for $e_{21}$ have only 43% coverage. The average coverage for the ML intervals is 77% across parameters. Using the Bayesian method, however, negative posterior mean estimates were obtained for $c_{22}$ and $d_{21}$, along with interval estimate coverage close to the nominal 95% coverage for all parameters, including $c_{22}$, $d_{21}$ and $e_{12}$; average coverage for the Bayesian intervals being 94.5%. The less favourable ML
results here are partially because the transformations, e.g. $c_{22} = c_{12}^* + c_{22}^*$ and $d_{21} = a_{12}^*$, force the variance parameter estimates to be positive. Another reason the Delta method has poor interval coverage may be that it is based on a first-order Taylor approximation, so the standard errors may not be accurate. Furthermore, Silvapulle and Sen (2004) point out the problems classical approaches have with inference under parameter constraints. In fact, optimising a likelihood can be a difficult task numerically and results can be unreliable, particularly in multivariate models. This may be the major factor to explain these simulation results, the normal approximation in the large sample methods may not be appropriate in this model, for the sample size considered (n = 1,000).

Tables 3.8 and 3.9 in the Appendix display results for one simulated dataset, from each of Model 1 and 2, of length $n = 50,000$. Results formed from MCMC estimation takes approximately 1 hour to complete for this sample size, while the ML estimation takes approximately 5 sec to complete. As the sample size increases, we should expect the standard error to become smaller, hence the parameter estimates to become more accurate, thus displaying a consistency result. Table 3.8 illustrates that this has happened for Model 1 under both methods - all 95% CIs have narrowed in size, and are almost comparable across methods, with all estimates close to the true parameter values. However, closer inspection reveals that 6 of the 17 ML intervals do not contain the true parameter value (we would expect between 0 and 4 intervals not to contain the true value from 17 independent intervals, at a 99% level). Only two Bayesian intervals do not contain their true value. Clearer differences among methods come from observing Table 3.9 for Model 2 - the Bayesian results seem similar to Model 1, with only one true parameter value not contained inside the 95% interval estimates. However, the MLEs seem biased and not tending towards the true values, with eight interval estimates not containing their true value; this includes the two negative, but not the zero, parameters in the volatility equations.

In summary, the Bayesian method proposed in this chapter on the given bivariate PVECH model seems to provide close to unbiased and consistent parameter estimates
for the models considered. Interval estimates for parameters also seemed to have close to nominal 95% coverage. In contrast, the ML method gave some apparently biased and inconsistent parameter estimates. It also gave interval estimates that were rarely close to the nominal 95% coverage, most intervals being lower than 85% and displaying similar poor performance across both models considered. In particular, the ML method was not able to accurately estimate CIs regardless of whether true values were outside or inside the usual sufficient restrictions; also with significantly lower coverage performance for negative parameters, as expected.

3.6 Empirical Study

This section analyses a real financial stock return dataset, using the Bayesian and ML methods discussed in Section 3.5. The data consists of bivariate daily closing values for two stock indices from 2 November 1987 through to 30 June 2005, resulting in a total of \( n = 4,374 \) bivariate stock returns. The two indices are the Standard & Poor (S&P) 500 index (United States) and the FTSE 100 (United Kingdom) index. Following the conventional approach, the actual data used are the bivariate percentage daily log returns, formed by the equation

\[
y_t = 100(\ln p_t - \ln p_{t-1}),
\]

where \( p_t \) is the bivariate daily closing value at time \( t \). In this example, \( y_{1t} \) will denote the S&P 500 index and \( y_{2t} \) will denote the FTSE 100 index.

A time series plot and scatterplot of the returns are given in Figure 3.2. Note the seemingly stationary mean and time varying volatility in both series from plot (a), implying a GARCH-like model structure. In a bivariate sense we also notice that both high and low periods of volatility generally appear at close to the same time in each series. This implies that a bivariate model seems appropriate. Plot (b) further supports this, demonstrating clear positive correlation (approximately 0.4387) between the two return series and also indicated by the slope of the fitted blue line.
Figure 3.2: (a) Time series plot, and (b) Scatterplot of daily log returns for S&P 500 and FTSE 100 indices between 2 November 1987 and 30 June 2005.
The bivariate PVECH model from Section 3.5 is fitted to this data, and its results compared with the equivalent BEKK model. These results are given in Table 3.2. The MCMC estimation takes approximately 4 min 30 sec to run, while the ML estimation takes approximately 3 sec. These results represent mostly well-known findings, e.g. each return series displays a strong and clear persistence in volatility, since the estimates of GARCH parameters give \( a_{ii} + b_{ii} \approx 1 \), \( i = 1, 2 \). However, there are some significant differences between the Bayesian and classical estimates and intervals, these being especially clear for the parameters \( e_{12} \) and \( e_{21} \) - the Bayesian estimates are negative and significantly less than zero, while the classical estimates are close to zero, without going below. An important point to note is that because the classical estimates of these parameters are so close to zero, the Delta confidence intervals will usually contain a positive and negative limit, implying that MLEs for \( e_{12} \) and \( e_{21} \) are insignificant. The Bayesian results suggest that these parameters are actually significantly negative. This is particularly relevant since these two parameters represent the volatility spillover between the two markets. In particular, this implies that there is a small but significant negative partial spillover effect from the US market volatility to the UK market volatility and vice versa.

### Table 3.2: Parameter estimation results for the PVECH and BEKK models fitted to the S&P 500/FTSE 100 bivariate returns.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bayesian Results</th>
<th>Classical Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Posterior Mean</td>
<td>Posterior Median</td>
</tr>
<tr>
<td>( m_{10} )</td>
<td>0.0537</td>
<td>0.0539</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.0449</td>
<td>0.0449</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>-0.1037</td>
<td>-0.1037</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>0.0048</td>
<td>0.0050</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.0157</td>
<td>0.0157</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.0499</td>
<td>0.0499</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.9433</td>
<td>0.9445</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.0296</td>
<td>0.0295</td>
</tr>
<tr>
<td>( e_{12} )</td>
<td>-0.0374</td>
<td>-0.0374</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>0.0178</td>
<td>0.0178</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.0607</td>
<td>0.0605</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.9147</td>
<td>0.9152</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>0.0296</td>
<td>0.0295</td>
</tr>
<tr>
<td>( e_{21} )</td>
<td>-0.0210</td>
<td>-0.0212</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.0091</td>
<td>0.0091</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.0578</td>
<td>0.0577</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.9162</td>
<td>0.9164</td>
</tr>
</tbody>
</table>
The Bayesian results also include two additional columns: the acceptance rates (AR) for each of the 4 steps in the sampling scheme (as a percentage), as well as the potential scale reduction factor (PSRF) for each parameter, as described in Section 2.2. The highest AR is for the autoregressive parameters (89%), while the lowest AR is for the GARCH parameters in the US market volatility equation (35%). The variability in these ARs demonstrates the ability for the MCMC scheme to capture posterior distributions of different shapes, with a generous proportion of values still being accepted. Each PSRF is calculated using \( R = 5 \) replications of the sampling scheme based on 5,000 MCMC iterations after the burn-in period. It can be seen that all PSRF values are close to 1, with 6 out of the 17 parameters having PSRFs slightly greater than 1.2 (the largest being 1.32 for the \( b_{11} \) parameter). This is proof of a fairly efficient and convergent algorithm. To further support these findings, Figure 3.3 displays the MCMC iterates of the parameters \( m_{22}, c_{11}, a_{22} \) and \( b_{12} \) for the \( R = 5 \) replications. While the speed of the convergence is different for each parameter, convergence is achieved (at least graphically) by the time the iterations reach the end of the burn-in period for all starting values considered. This is further proof of the efficiency and convergence of this sampling scheme.

As further evidence that the PVECH model is favoured over the BEKK model for this dataset, Table 3.3 displays model selection results for the two models; the value of the log-likelihood function evaluated at the parameter estimates, and the information criterion functions of Akaike (1974) (the AIC) and Schwarz (1978) (the BIC), which are given by the formulas

\[
AIC = -2 \ln(L) + 2p,
\]

\[
BIC = -2 \ln(L) + p \ln(n),
\]

where \( L \) denotes the likelihood function and \( p \) is the number of model parameters. These information criterion functions were discussed briefly in Chapter 1, and will be discussed in more detail in Chapter 4.
Figure 3.3: MCMC iterates for the parameters $m_{22}$, $c_{11}$, $a_{22}$ and $b_{12}$ of the PVECH model fitted to the S&P 500/FTSE 100 bivariate returns, each with 5 different starting values.
The model that has the smallest AIC or BIC is typically the model that is favoured - this has occurred for the PVECH model in each case. The PVECH model also has a larger log-likelihood function, thus providing further evidence that the PVECH model fits the data better over the equivalent BEKK model.

Table 3.3: Model selection results for the PVECH and BEKK models fitted to the S&P 500/FTSE 100 bivariate returns.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>PVECH Model</th>
<th>BEKK Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Likelihood</td>
<td>−11,182</td>
<td>−11,203</td>
</tr>
<tr>
<td>AIC</td>
<td>22,398</td>
<td>22,440</td>
</tr>
<tr>
<td>BIC</td>
<td>22,506</td>
<td>22,549</td>
</tr>
</tbody>
</table>

Figure 3.4 plots the estimated dynamic volatilities, covariances and correlations between the markets. The first three plots contain the estimated variances ($\hat{h}_{11,t}$ and $\hat{h}_{22,t}$) and estimated covariances ($\hat{h}_{12,t}$) as a result of the Bayesian and classical fits to the return data, while the last plot contains the estimated correlations ($\hat{\rho}_{12,t}$), calculated through the equation

$$\hat{\rho}_{12,t} = \frac{\hat{h}_{12,t}}{\sqrt{\hat{h}_{11,t}\hat{h}_{22,t}}}, \quad t = 1, \ldots, n.$$  

The Bayesian and classical fits have been calculated in different ways: the Bayesian fit uses a posterior mean estimate of the terms $h_{11,t}$, $h_{22,t}$ and $h_{12,t}$ for each $t$ over the MCMC sample, i.e.

$$\hat{h}_{ij,t} = \frac{1}{f - w} \sum_{j=w+1}^{f} h_{ij,t}^{[j]}, \quad t = 1, \ldots, n.$$  

The classical estimates plug the parameter MLEs from Table 3.2 directly into the variance-covariance equations. Differences between methods seem apparent in periods of high volatility, with the Bayesian method generating higher volatility estimates. These may be caused by biases in parameter estimates from the MLE method or other reasons.

To further discuss the parameter estimates in a practical sense, each return series displays a strong and clear persistence in volatility, since the estimates of GARCH parameters give $a_{ii} + b_{ii} \approx 1, \ i = 1, 2$. There is also very strong persistence in covariance between US and UK returns, as measured by $a_{12} + b_{12} \approx 1$ around a positive mean $c_{12} > 0$.  

67
Figure 3.4: Estimated volatility, covariance and correlation for the S&P 500/FTSE 100 bivariate returns, using the Bayesian MCMC and classical maximum likelihood methods from the PVECH and BEKK models (respectively).

— Bayesian Fit — Classical Fit
The plotted correlations over time fluctuate between small negative values up to about 0.7, mostly they are close to about 0.5. This indicates clear positive correlation between US and UK stock returns, but that the correlation changes with time and can be negative (observing Figure 3.4), although it is hard to see the pattern of behaviour at or around these negative correlation estimates.

3.7 Chapter Summary

This chapter proposed a prior formulation on a multivariate VECH GARCH model that relaxes the usual parameter restrictions imposed on VECH and BEKK models. The prior directly allows both necessary and sufficient conditions for covariance stationarity and positive definiteness of covariance matrices to be satisfied via their definitions. Results from a simulation study, under two sets of model parameters, indicated that the proposed Bayesian MCMC method was favourable in terms of parameter estimate bias, consistency and coverage of interval estimates, as well as extending the allowable parameter space, over the large sample ML method. The comparison was even more favourable for the Bayesian method when the true parameter values lay outside the usual parameter restrictions. When fit to bivariate stock return data from the US and UK markets, significant differences were found between these two competing models. Some parameter estimates, namely the volatility spillovers, under the Bayesian approach were significantly negative, but due to the usual parameter restrictions the competing MLE estimates were insignificant. Differences were also observed in volatility estimation, especially in times of high volatility, with MLE estimates being lower than the corresponding Bayesian estimates at those times.

A major advantage in using the Bayesian techniques described in this chapter is that they can be extended to apply to any type of multivariate GARCH model, due to the generality of the conditions required. The following chapter proposes a generalisation of the PVECH model, with an application to model selection.
### 3.8 Appendix

#### 3.8.1 Tables of Results for Simulation Study with \( n = 1,000 \)

**Table 3.4:** Parameter estimates for PVECH and (transformed) BEKK models fitted to simulated data from Model 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bayesian Results</th>
<th>Classical Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Value</td>
<td>Posterior Mean</td>
</tr>
<tr>
<td>( m_{10} )</td>
<td>0.01</td>
<td>0.0099</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.01</td>
<td>0.0099</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.2</td>
<td>0.1973</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>0.2</td>
<td>0.1976</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.02</td>
<td>0.0225</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.3</td>
<td>0.3132</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.4</td>
<td>0.3535</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.05</td>
<td>0.0706</td>
</tr>
<tr>
<td>( e_{12} )</td>
<td>0.05</td>
<td>0.0477</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>0.01</td>
<td>0.0105</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.3</td>
<td>0.3995</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.225</td>
<td>0.1943</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>0.05</td>
<td>0.0534</td>
</tr>
<tr>
<td>( e_{21} )</td>
<td>0.05</td>
<td>0.0587</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.005</td>
<td>0.0057</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.35</td>
<td>0.3558</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.35</td>
<td>0.3058</td>
</tr>
</tbody>
</table>

**Table 3.5:** 95% credible intervals and coverages for PVECH and (transformed) BEKK models fitted to simulated data from Model 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bayesian Results</th>
<th>Classical Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Value</td>
<td>95% CI P'tiles of Mean</td>
</tr>
<tr>
<td>( m_{10} )</td>
<td>0.01</td>
<td>[-0.0048, 0.0244]</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.01</td>
<td>[0.0005, 0.0195]</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.2</td>
<td>[0.1364, 0.2614]</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>0.2</td>
<td>[0.1347, 0.2567]</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.02</td>
<td>[0.0149, 0.0313]</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.3</td>
<td>[0.2321, 0.4023]</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.4</td>
<td>[0.2129, 0.4830]</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.05</td>
<td>[-0.0084, 0.1573]</td>
</tr>
<tr>
<td>( e_{12} )</td>
<td>0.05</td>
<td>[-0.0785, 0.1675]</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>0.01</td>
<td>[0.0060, 0.0152]</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.3</td>
<td>[0.2213, 0.3996]</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.225</td>
<td>[0.0544, 0.3584]</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>0.05</td>
<td>[0.0209, 0.0885]</td>
</tr>
<tr>
<td>( e_{21} )</td>
<td>0.05</td>
<td>[-0.0084, 0.1337]</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.005</td>
<td>[0.0037, 0.0082]</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.35</td>
<td>[0.2801, 0.4375]</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.35</td>
<td>[0.1733, 0.4182]</td>
</tr>
</tbody>
</table>
Table 3.6: Parameter estimates for PVECH and (transformed) BEKK models fitted to simulated data from Model 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Bayesian Results</th>
<th>Classical Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Posterior Mean</td>
<td>Posterior Median</td>
</tr>
<tr>
<td>( m_{10} )</td>
<td>0.1</td>
<td>0.1003</td>
<td>0.1003</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.15</td>
<td>0.1499</td>
<td>0.1499</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.3</td>
<td>0.2978</td>
<td>0.2980</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>0.25</td>
<td>0.2487</td>
<td>0.2487</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.2</td>
<td>0.1968</td>
<td>0.1965</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.36</td>
<td>0.3555</td>
<td>0.3540</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.04</td>
<td>0.0684</td>
<td>0.0636</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.09</td>
<td>0.0785</td>
<td>0.0765</td>
</tr>
<tr>
<td>( e_{12} )</td>
<td>0.2</td>
<td>0.0808</td>
<td>0.052808</td>
</tr>
</tbody>
</table>

Table 3.7: 95% credible intervals and coverages for PVECH and (transformed) BEKK models fitted to simulated data from Model 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Bayesian Results</th>
<th>Classical Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>95% CI P’tiles</td>
<td>95% CI P’tiles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean P’tiles</td>
<td>Mean P’tiles</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>-0.01</td>
<td>-0.0464</td>
<td>-0.0348</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.16</td>
<td>0.1803</td>
<td>0.1786</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.64</td>
<td>0.5647</td>
<td>0.6592</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>-0.0053</td>
<td>-0.0059</td>
</tr>
<tr>
<td>( e_{21} )</td>
<td>0.1</td>
<td>0.0974</td>
<td>0.0966</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.01</td>
<td>0.0116</td>
<td>0.0113</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.21</td>
<td>0.2065</td>
<td>0.2070</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.16</td>
<td>0.0611</td>
<td>0.0680</td>
</tr>
</tbody>
</table>
### 3.8.2 Tables of Results for Simulation Study with \( n = 50,000 \)

**Table 3.8:** Parameter estimation results from fitting PVECH and BEKK models to a simulated dataset from Model 1 of sample size \( n = 50,000 \)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Posterior Mean</th>
<th>Posterior Median</th>
<th>95% CI Percentiles</th>
<th>ML Estimate</th>
<th>Delta Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{10} )</td>
<td>0.01</td>
<td>0.0113</td>
<td>0.0112</td>
<td>[0.0101, 0.0126]</td>
<td>0.0104</td>
<td>0.0089, 0.0118</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.01</td>
<td>0.0108</td>
<td>0.0108</td>
<td>[0.0096, 0.0120]</td>
<td>0.0107</td>
<td>0.0087, 0.0117</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.2</td>
<td>0.1918</td>
<td>0.1918</td>
<td>[0.1835, 0.2001]</td>
<td>0.1922</td>
<td>0.1862, 0.1983</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>0.2</td>
<td>0.1965</td>
<td>0.1965</td>
<td>[0.1891, 0.2039]</td>
<td>0.1963</td>
<td>0.1903, 0.2023</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.02</td>
<td>0.0209</td>
<td>0.0209</td>
<td>[0.0199, 0.0219]</td>
<td>0.0206</td>
<td>0.0198, 0.0214</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.3</td>
<td>0.2956</td>
<td>0.2956</td>
<td>[0.2855, 0.3060]</td>
<td>0.2907</td>
<td>0.2821, 0.2993</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.4</td>
<td>0.3979</td>
<td>0.3981</td>
<td>[0.3775, 0.4181]</td>
<td>0.4008</td>
<td>0.3852, 0.4164</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.05</td>
<td>0.0544</td>
<td>0.0543</td>
<td>[0.0396, 0.0699]</td>
<td>0.0549</td>
<td>0.0462, 0.0636</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.05</td>
<td>0.0251</td>
<td>0.0251</td>
<td>[0.0103, 0.0610]</td>
<td>0.0307</td>
<td>0.0118, 0.0496</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>0.01</td>
<td>0.0998</td>
<td>0.0998</td>
<td>[0.0995, 0.1001]</td>
<td>0.0999</td>
<td>0.0994, 0.1003</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.3</td>
<td>0.2943</td>
<td>0.2942</td>
<td>[0.2862, 0.3031]</td>
<td>0.2854</td>
<td>0.2769, 0.2939</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.225</td>
<td>0.2410</td>
<td>0.2411</td>
<td>[0.2251, 0.2561]</td>
<td>0.2449</td>
<td>0.2286, 0.2613</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>0.05</td>
<td>0.0484</td>
<td>0.0484</td>
<td>[0.0460, 0.0507]</td>
<td>0.0468</td>
<td>0.0435, 0.0501</td>
</tr>
<tr>
<td>( e_{21} )</td>
<td>0.05</td>
<td>0.0494</td>
<td>0.0494</td>
<td>[0.0496, 0.0505]</td>
<td>0.0499</td>
<td>0.0441, 0.0584</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.0065</td>
<td>0.0000</td>
<td>0.0000</td>
<td>[0.0000, 0.0000]</td>
<td>0.0000</td>
<td>0.0000, 0.0000</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.35</td>
<td>0.3472</td>
<td>0.3472</td>
<td>[0.3392, 0.3550]</td>
<td>0.3387</td>
<td>0.3314, 0.3469</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.35</td>
<td>0.3450</td>
<td>0.3449</td>
<td>[0.3318, 0.3581]</td>
<td>0.3525</td>
<td>0.3410, 0.3639</td>
</tr>
</tbody>
</table>

**Table 3.9:** Parameter estimation results from fitting PVECH and BEKK models to a simulated dataset from Model 2 of sample size \( n = 50,000 \)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Posterior Mean</th>
<th>Posterior Median</th>
<th>95% CI Percentiles</th>
<th>ML Estimate</th>
<th>Delta Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{10} )</td>
<td>0.1</td>
<td>0.0981</td>
<td>0.0981</td>
<td>[0.0943, 0.1025]</td>
<td>0.0980</td>
<td>0.0945, 0.1015</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.15</td>
<td>0.1506</td>
<td>0.1506</td>
<td>[0.1486, 0.1527]</td>
<td>0.1504</td>
<td>0.1482, 0.1526</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.3</td>
<td>0.3027</td>
<td>0.3027</td>
<td>[0.2943, 0.3112]</td>
<td>0.3030</td>
<td>0.2967, 0.3093</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>0.25</td>
<td>0.2594</td>
<td>0.2594</td>
<td>[0.2519, 0.2668]</td>
<td>0.2598</td>
<td>0.2532, 0.2665</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.2</td>
<td>0.2024</td>
<td>0.2024</td>
<td>[0.1967, 0.2083]</td>
<td>0.1903</td>
<td>0.1848, 0.1959</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.36</td>
<td>0.3673</td>
<td>0.3673</td>
<td>[0.3584, 0.3761]</td>
<td>0.3876</td>
<td>0.3769, 0.3983</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.04</td>
<td>0.0533</td>
<td>0.0554</td>
<td>[0.0500, 0.0592]</td>
<td>0.0474</td>
<td>0.0436, 0.0587</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.09</td>
<td>0.0984</td>
<td>0.0985</td>
<td>[0.0965, 0.0994]</td>
<td>0.1056</td>
<td>0.0912, 0.1200</td>
</tr>
<tr>
<td>( e_{12} )</td>
<td>0.0</td>
<td>0.0002</td>
<td>0.0003</td>
<td>[-0.0262, 0.0341]</td>
<td>0.0247</td>
<td>-0.0030, 0.0224</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>-0.01</td>
<td>-0.0099</td>
<td>-0.0099</td>
<td>[-0.0103, 0.0096]</td>
<td>0.0004</td>
<td>0.0006, 0.0002</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.16</td>
<td>0.1593</td>
<td>0.1593</td>
<td>[0.1584, 0.1602]</td>
<td>0.1578</td>
<td>0.1527, 0.1647</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.64</td>
<td>0.6403</td>
<td>0.6403</td>
<td>[0.6393, 0.6413]</td>
<td>0.6482</td>
<td>0.6382, 0.6582</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>-0.0100</td>
<td>-0.0100</td>
<td>[-0.0110, -0.0087]</td>
<td>0.0003</td>
<td>-0.0001, 0.0007</td>
</tr>
<tr>
<td>( e_{21} )</td>
<td>0.1</td>
<td>0.0990</td>
<td>0.0991</td>
<td>[0.0968, 0.1014]</td>
<td>0.0559</td>
<td>0.0533, 0.0584</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.01</td>
<td>0.0005</td>
<td>0.0005</td>
<td>[0.0002, 0.0008]</td>
<td>0.0002</td>
<td>0.0002, 0.0004</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.21</td>
<td>0.2146</td>
<td>0.2146</td>
<td>[0.2045, 0.2245]</td>
<td>0.2124</td>
<td>0.2063, 0.2185</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.16</td>
<td>0.1572</td>
<td>0.1570</td>
<td>[0.1128, 0.2019]</td>
<td>0.1381</td>
<td>0.1152, 0.1609</td>
</tr>
</tbody>
</table>
Chapter 4

A Test for Multivariate Asymmetry in Financial Time Series using Approximate Bayesian Model Selection

An approximate Bayesian model selection procedure is discussed in this chapter that is adapted to simultaneously: (i) test for asymmetry in multivariate financial time series, and (ii) identify an appropriate heteroscedastic model from a proposed extension to a family of multivariate nonlinear GARCH models. The proposed model family captures the major stylised features of financial stock returns: fat tails, excess kurtosis, time-varying volatility, volatility clustering and covariance asymmetry, via a parsimonious specification with easily identifiable and interpretable parameters. Both the necessary and sufficient conditions for stationarity and positive definiteness are enforced empirically via their definitions, through a specific but diffuse prior distribution (as in Chapter 3). An adaptive Markov chain Monte Carlo (MCMC) approach is employed for estimation and inference, and combined with a computationally efficient, but approximate, model selection approach. A simulation study illustrates favourable estimation properties and very strong
model selection performance, indicating that an accurate approximation has been obtained, for practical purposes. An empirical study on stock returns from international markets is presented and posterior model probabilities are estimated. In each case they reveal clear volatility and covariance asymmetry in general across markets.

4.1 Introduction

Chapter 3 of this thesis reviewed a number of multivariate GARCH models in the literature, particularly the VECH model of Bollerslev et al. (1988) and the BEKK model of Engle and Kroner (1995). All of these fall under the category of “symmetric” models, since they exhibit the same covariance (and/or mean) process following good and bad market news. A major stylised trait discovered by Black (1976) is known as volatility asymmetry, which is the phenomenon of higher volatility following bad market news compared to good market news of the same magnitude. Nonlinear univariate GARCH models, such as the EGARCH (Nelson, 1991) and GJR-GARCH (Glosten et al., 1993), have been proposed for volatility asymmetry, while double threshold GARCH models, e.g. DT-ARCH of Li and Li (1996), can capture both mean and volatility asymmetry.

Asymmetry has also been included in many extensions to symmetric multivariate GARCH models: the Asymmetric Dynamic Covariance (ADC) model of Kroner and Ng (1998) is an early example which directly extends the BEKK model, but suffers from a lack of parameter interpretability like the BEKK model (as discussed in Chapter 3); the DVECH model was extended by De Goeij and Marquering (2004) to capture cross-asset and interacting asymmetry, but being a diagonal model ensures that no spill-over effects are captured; Cappiello et al. (2006) extend the DCC model to allow asymmetric volatility and correlation, though asymmetry is in a very restricted form and again parameters are repeated across equations; while more recently Kwan et al. (2010) combine the DCC and univariate threshold GARCH models into the one asymmetric multivariate GARCH model. In addition, these models do not alleviate the issues of simultaneously allowing
important estimation properties such as ease of PD enforcement and maximising the allowable parameter space under CS, as discussed in Chapter 3. An asymmetric family of VECH models is proposed in this chapter, extending and adapting the symmetric PVECH model proposed in Chapter 3. The challenges involved are: (i) to capture important asymmetries, including “self-exciting” forms (asymmetry in an asset’s own volatility equation) as well as interacting or cross-asset asymmetry in both volatility and covariance equations; (ii) maintain parsimony and parameter interpretation in all equations, and (iii) ensure necessary and sufficient conditions for PD and CS can be satisfied in a straightforward manner in the general model. There is an absence of models in the literature that satisfy all three of these properties. The model and approach to be proposed in this chapter intends to fill this important gap.

Some recent work has been done on asymmetric model selection: see Chen et al. (2006a) for the univariate case; while Osiewalski et al. (2007) compare bivariate GARCH and stochastic volatility models, employing the marginal likelihood estimated by the harmonic mean (see Newton and Raftery, 1994). This latter method is known to experience numerical problems, since it involves the inverse of the likelihood function, and can be practically un-implementable. This chapter considers the adaption of the approximate Bayesian model selection method by Congdon (2006) to select among a family of possibly non-nested multivariate asymmetric models. Compared to other powerful and well-known Bayesian model selection methods such as that by Chib (1995), Kass and Raftery (1995) and the reversible jump (RJ) method of Green (1995), Congdon’s method is computationally less expensive and simpler to implement. Rather than examining the marginal likelihood (like in the examples above), Congdon’s method approximates posterior model probabilities directly using Monte Carlo estimates. Through a simulation study, I demonstrate the strong model selection performance of this method for reasonable sample size, thus adding to its success as illustrated by other authors (see for e.g. Min and Czado, 2010 and Chen et al., 2011a).

The outline of this chapter is as follows: Section 4.2 of this chapter reviews some
Section 4.3 describes the proposed model family and discusses its properties. Section 4.4 presents the Bayesian MCMC procedure and how it will be adapted in this chapter, while Section 4.5 reviews the Bayesian model selection procedure of Congdon (2006) and its comparison to other popular approaches. Section 4.6 presents a simulation study, and Section 4.7 presents an empirical study on several pairs of bivariate data from international stock markets. Section 4.8 concludes.

4.2 Asymmetric Multivariate GARCH Models

As mentioned in Chapter 1, asymmetry in stock returns was pointed out by Black (1976), and is the phenomenon of higher volatility occurring during negative shocks compared to positive shocks of the same magnitude. Asymmetric effects have been considered extensively in the literature in both univariate and multivariate cases. While different approaches have been taken to model asymmetry, I focus on multivariate extensions of the univariate GJR-GARCH model by Glosten et al. (1993), which extends the univariate GARCH(1,1) model by introducing an indicator variable $I_t$ that is a function of the sign of the residual $\varepsilon_t$ such that

$$I_t = \begin{cases} 
1 & \text{if } \varepsilon_t < 0, \\
0 & \text{otherwise}.
\end{cases}$$

The conditional variance $h_t$ is then defined as

$$h_t = c_1 + (a_1 + \alpha_1 I_{t-1})\varepsilon_{t-1}^2 + b_1 h_{t-1}. \quad (4.1)$$

Thus the GJR-GARCH model in (4.1) assumes that the conditional variance is inflated by a scalar $\alpha_1$ (with $0 \leq \alpha_1 < 1$) whenever a negative residual occurs. Of course, the standard GARCH(1,1) model is obtained when $\alpha_1 = 0$.

Examples of multivariate extensions of this model will now be reviewed.
4.2.1 The ADVECH Model

The diagonal VECH model has been extended to incorporate GJR asymmetries by De Goeij and Marquering (2004) who propose the asymmetric diagonal VECH (ADVECH) model: As in Chapter 3, let \( \{ \varepsilon_t \}_{t=1}^n = \{ (\varepsilon_{1t}, \ldots, \varepsilon_{Nt}) \}_{t=1}^n \) denote a set of \( N \)-dimensional residual vectors, \( \psi_{t-1} \) the information set at time \( t-1 \) and \( H_t \) the conditional covariance matrix of order \( N \). Before stating this model, first define

\[
I_{it} = \begin{cases} 
1 & \text{if } \varepsilon_{it} < 0 \\
0 & \text{otherwise} 
\end{cases}, \quad i = 1, \ldots, N, \tag{4.2}
\]

the vectors

\[
\varepsilon_t^- = \begin{pmatrix} I_{1t} \varepsilon_{1t} \\ \vdots \\ I_{Nt} \varepsilon_{Nt} \end{pmatrix} \quad \text{and} \quad \varepsilon_t^+ = \begin{pmatrix} (1 - I_{1t})\varepsilon_{1t} \\ \vdots \\ (1 - I_{Nt})\varepsilon_{Nt} \end{pmatrix}
\]

and define the matrix operator \( T \) which substitutes the lower triangular elements of a matrix with the upper triangular elements. The asymmetric DVECH model can be written in matrix form as follows:

\[
H_t = C + A_1 \odot \varepsilon_{t-1} \varepsilon_{t-1}' + A_2 \odot \varepsilon_{t-1}^- \varepsilon_{t-1}^- + A_3 \odot T(\varepsilon_{t-1}^- \varepsilon_{t-1}^+) + A_4 \odot T(\varepsilon_{t-1}^+ \varepsilon_{t-1}^-) + B_1 \odot H_{t-1}, \tag{4.3}
\]

where \( C = [c_{ij}] \), \( A_1 = [a_{ij,1}] \), \( A_2 = [a_{ij,2}] \), \( A_3 = [a_{ij,3}] \), \( A_4 = [a_{ij,4}] \) and \( B_1 = [b_{ij}] \) are \( N \times N \) symmetric parameter matrices. This model is much clearer when written in terms of the elements of \( H_t \) as follows:

\[
h_{ij,t} = c_{ij} + a_{ij,1} + a_{ij,2} I_{i,t-1} I_{j,t-1} + a_{ij,3} I_{i,t-1}(1 - I_{j,t-1}) \\
+ a_{ij,4}(1 - I_{i,t-1})I_{j,t-1} \varepsilon_{t-1} \varepsilon_{t-1} + b_{ij} h_{ij,t-1}, \quad \forall \ i < j = 1, \ldots, N.
\]

77
This model contains the usual GJR asymmetries in the conditional variance equations.
However, for the conditional covariances the effects from the four different residual sign
combinations are included, i.e. $a_{ij,1}$ when both residuals are positive, $a_{ij,2}$ when both
residuals are negative, $a_{ij,3}$ when residual $i$ is negative and residual $j$ is positive, and $a_{ij,4}$
when residual $i$ is positive and residual $j$ is negative.

The main advantages of this model are that parameter interpretation remains straight-
forward because of its VECH notation. Also covariance asymmetry is captured for resid-
uals of opposite sign, which many models in the literature neglect to include. The pa-
rameters are also reduced by restricting the model to a diagonal form. However, this also
means that no residual and/or volatility spillover effects are captured in the model, nor
are potential asymmetries from these effects included. Difficulties can also occur when
estimating this model, i.e. ensuring that $H_t$ is PD; De Goeij and Marquering (2004)
state that they enforced PD by only allowing parameter values that ensure the one-step
ahead forecast of $H_t$ was PD, and admit that this does not imply PD of multi-step ahead
forecasts of $H_t$. While PD is satisfied for the bivariate datasets they considered, this may
not be the case for others. This then calls for other approaches to be considered.

4.2.2 The ADC Model

Proposed by Kroner and Ng (1998), the Asymmetric Dynamic Covariance (ADC) model
nests many well-known multivariate GARCH models plus allows for asymmetric effects.
Let $\eta_t = I_t\varepsilon_t$, where $I_t$ is defined in equation (4.2), and let $\eta_t = (\eta_{1t}, \ldots, \eta_{Nt})$. The
conditional covariance matrix $H_t$ is defined as

$$H_t = D_tPD_t + \Phi \odot \Theta_t$$

where $D_t = \text{diag}(\sqrt{\theta_{11,t}}, \ldots, \sqrt{\theta_{NN,t}})$, $P = [\rho_{ij}]$ is a correlation matrix, $\Phi = [\phi_{ii}]$ is a
scalar matrix with $\phi_{ii} = 0$, and $\Theta_t = [\theta_{ij,t}]$ has elements defined as
where $C = [c_{ij}]$ is a scalar matrix, and each $a_i = (a_{i1}, \ldots, a_{iN})$, $g_i = (g_{i1}, \ldots, g_{iN})$ and $b_i = (b_{i1}, \ldots, b_{iN})$ are $N \times 1$ parameter vectors. On its own, each $\theta_{ij,t}$ can be viewed as a conditional covariance equation of an asymmetric BEKK model, with the ADC model reverting to this under certain conditions (see below).

As mentioned earlier, the ADC model is general enough to nest other multivariate GARCH models, namely the asymmetric versions of the VECH, BEKK, FARCH and CCC models. To easily observe this, consider the bivariate case $N = 2$. These four models will result under the following conditions:

- $\rho_{12} = a_{12} = g_{12} = b_{12} = a_{21} = g_{21} = b_{21} = 0$ and $\phi_{12} = 1$ gives an asymmetric VECH model\(^1\):

\[
\begin{align*}
    h_{11,t} &= c_{11} + (a_{11}^2 + g_{11}^2 I_{1,t-1}^2)\epsilon_{1,t-1}^2 + b_{11}^2 h_{11,t-1}, \\
    h_{12,t} &= c_{12} + (a_{11}a_{22} + g_{11}g_{22} I_{1,t-1} I_{2,t-1})\epsilon_{1,t-1}\epsilon_{2,t-1} + b_{11}b_{22} h_{12,t-1}, \\
    h_{22,t} &= c_{22} + (a_{22}^2 + g_{22}^2 I_{2,t-1}^2)\epsilon_{2,t-1}^2 + b_{22}^2 h_{22,t-1}.
\end{align*}
\]

- $\rho_{12} = \phi_{12} = 1$ gives an asymmetric BEKK model:

\[
\begin{align*}
    h_{11,t} &= c_{11} + (a_{11}^2 + g_{11}^2 I_{1,t-1}^2)\epsilon_{1,t-1}^2 + 2(a_{11}a_{12} + g_{11}g_{12} I_{1,t-1} I_{2,t-1})\epsilon_{1,t-1}\epsilon_{2,t-1} \\
    &\quad + (a_{22}^2 + g_{22}^2 I_{2,t-1}^2)\epsilon_{2,t-1}^2 + b_{11}^2 h_{11,t-1} + 2b_{11}b_{22} h_{12,t-1} + b_{22}^2 h_{22,t-1}, \\
    h_{12,t} &= c_{12} + (a_{11}a_{21} + g_{11}g_{21} I_{1,t-1}^2)\epsilon_{1,t-1}^2 + (a_{12}a_{22} + g_{12}g_{22} I_{2,t-1}^2)\epsilon_{2,t-1}^2 \\
    &\quad + [(a_{11}a_{22} + a_{12}a_{21}) + (g_{11}g_{22} + g_{12}g_{21}) I_{1,t-1} I_{2,t-1}]\epsilon_{1,t-1}\epsilon_{2,t-1} \\
    &\quad + b_{11}b_{21} h_{11,t-1} + (b_{11}b_{22} + b_{12}b_{21}) h_{12,t-1} + b_{12}b_{22} h_{22,t-1},
\end{align*}
\]

\(^1\)In fact, this model is also an asymmetric DBEKK(1, 1, 1) model - they are observationally equivalent under this setting.
\[ h_{22,t} = c_{22} + (a_{22}^2 + g_{22}^2 I_{2,t-1}^2) \varepsilon_{2,t-1}^2 + 2(a_{21} a_{22} + g_{22} g_{11} I_{1,t-1} I_{2,t-1}) \varepsilon_{1,t-1} \varepsilon_{2,t-1} \]
\[ + (a_{21}^2 + g_{21}^2 I_{1,t-1}) \varepsilon_{1,t-1}^2 + b_{22}^2 h_{22,t-1} + 2b_{21} b_{22} h_{12,t-1} + b_{21}^2 h_{11,t-1}. \]

- \(\rho_{12} = 0\), \(\Phi = \lambda \lambda'\), \(c_{ij} = c\), \(a_i' \varepsilon_{t-1} \varepsilon_{t-1}' a_j = a(\mathbf{w}' \mathbf{\varepsilon}_{t-1})^2\), \(g_i' \mathbf{\eta}_{t-1} \mathbf{\eta}_{t-1}' g_j = \gamma (\mathbf{w}' \mathbf{\eta}_{t-1})^2\), \(b'_i H_{t-1} b_j = \beta (\mathbf{w}' H_{t-1} \mathbf{w})\) and \(\sigma_{ij} = \omega_{ij} - \lambda_i \lambda_j \mathbf{w}' \mathbf{C} \mathbf{w}\) (\(\forall i, j = 1, 2\)) gives an asymmetric FARCH model:

\[ h_{ij,t} = \sigma_{ij} + \lambda_i \lambda_j [c + a(\mathbf{w}' \mathbf{\varepsilon}_{t-1})^2 + \gamma (\mathbf{w}' \mathbf{\eta}_{t-1})^2 + \beta (\mathbf{w}' H_{t-1} \mathbf{w})], \quad \forall i, j = 1, 2. \]

- \(a_{12} = g_{12} = b_{12} = a_{21} = g_{21} = b_{21} = 0\) and \(\phi_{12} = 0\) gives an asymmetric CCC model:

\[ h_{11,t} = c_{11} + (a_{11}^2 + g_{11}^2 I_{1,t-1}^2) \varepsilon_{1,t-1}^2 + b_{11}^2 h_{11,t-1}, \]
\[ h_{12,t} = \rho_{12} \sqrt{h_{11,t} h_{22,t}}, \]
\[ h_{22,t} = c_{22} + (a_{22}^2 + g_{22}^2 I_{2,t-1}^2) \varepsilon_{2,t-1}^2 + b_{22}^2 h_{22,t-1}. \]

Having four seemingly different multivariate GARCH models in the one general model is obviously an advantage of the ADC model, and can make techniques such as model selection much easier. However, in all forms of the model, parameter interpretation is difficult due to the general BEKK notation. In addition, the ADC model only allows for asymmetric effects in conditional covariances based on residuals of the same sign and thus ignore potential asymmetries for residuals of opposite signs.

### 4.2.3 The AG-DCC Model

Cappiello et al. (2006) extends the DCC model by Engle (2002) (reviewed in Chapter 3) to incorporate asymmetric effects. This is called the Asymmetric Generalised DCC (AG-DCC) model, and the form of the conditional covariance matrix \(H_t\) remains the
same as that for the DCC model:

\[ H_t = D_t P_t D_t \]

where \( D_t = \text{diag}(\sqrt{h_{11,t}}, \ldots, \sqrt{h_{NN,t}}) \) and \( P_t = [\rho_{ij,t}] \) is a correlation matrix. The difference comes in the structure of \( P_t \): let \( z_t = D_t^{-1} \varepsilon_t \) denote the standardised residual vector and \( \bar{P} = E(z_t z_t') \) the unconditional correlation matrix of the standardised residuals, as with the DCC model. In addition, define \( \eta_{it} = I_{it} z_{it} \) where

\[
I_{it} = \begin{cases} 
1 & \text{if } z_{it} < 0, \\
0 & \text{otherwise} 
\end{cases} \quad i = 1\ldots,N
\]

with the vector \( \eta_t = (\eta_{1t}, \ldots, \eta_{Nt}) \) and let \( \bar{\Upsilon} = E(\eta_t \eta_t') \). As with the DCC model, the conditional correlations are defined by

\[
\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t}q_{jj,t}}} \quad \forall \ i = j = 1, \ldots, N
\]

but now the matrix \( Q_t = [q_{ij,t}] \) is defined as

\[
Q_t = \bar{P} - A' \bar{P} A - B' \bar{P} B - G' \bar{\Upsilon} G + A' z_{t-1} z_{t-1}' A + G' \eta_{t-1} \eta_{t-1}' G + B' Q_{t-1} B, \quad (4.4)
\]

where \( A, G \) and \( B \) are \( N \times N \) parameter matrices. Each \( \eta_{it} \) will either be zero or a negative residual, and so the vector \( \eta_t \) will only impact on \( Q_t \) when the standardised residuals \( z_{it} \) and \( z_{jt} \) are both negative. When this occurs, \( Q_t \) will be increased by \( G' \eta_{t-1} \eta_{t-1}' G \).

While this notation results in a significant parameter number increase over the standard DCC model, restrictions can be placed on the parameter matrices to render the model more parsimonious. An example of this is the assumption that \( A, G \) and \( B \) are diagonal, which reduces \( Q_t \) to the following form:

\[
Q_t = \bar{P} \odot (\mathbf{u}' - \mathbf{a}' - \mathbf{b}' - \bar{\Upsilon} \odot \mathbf{g} \mathbf{g}' + \mathbf{a} \mathbf{a}' \odot z_{t-1} z_{t-1}' + \mathbf{g} \mathbf{g}' \odot \eta_{t-1} \eta_{t-1}' + \mathbf{b} \mathbf{b}' \odot Q_{t-1} \quad (4.5)
\]

81
where $\mathbf{1}$ is a vector of ones and $\mathbf{a}$, $\mathbf{g}$ and $\mathbf{b}$ are vectors containing the diagonal elements of the matrices $A$, $G$ and $B$ respectively. A further reduction in parameter numbers is achieved by considering the scalar representation of equation (4.4), which simply extends the DCC model to include asymmetric effects (called the A-DCC model):

$$Q_t = (1 - a^2 - b^2)\bar{P} - g^2\bar{\Upsilon} + a^2z_{t-1}z'_{t-1} + g^2\eta_{t-1}\eta'_{t-1} + b^2Q_{t-1} \tag{4.6}$$

where $a$, $g$ and $b$ are scalars. Cappiello et al. (2006) recommend the use of the scalar form in equation (4.6) for problems of high dimension.

As with the DCC model, estimation of the parameters of the AG-DCC model has the same two-step procedure as described in Section 3.2, i.e. univariate models are fitted to obtain estimates of each $h_{ii,t}$, then, by setting $\bar{P}$ and $\bar{\Upsilon}$ to their sample values, estimating the parameters contained in $Q_t$. A set of sufficient conditions to ensure that $Q_t$ is PD for all $t$ during estimation in all forms of the AG-DCC model are as follows:

- **AG-DCC**: $\bar{P} - A'\bar{P}A - B'\bar{P}B - G'\bar{\Upsilon}G$ is positive semi-definite and $Q_1$ is PD;
- **AG-DCC Diagonal**: $\bar{P} \odot (\mathbf{1}\mathbf{1}' - \mathbf{a}\mathbf{a}' - \mathbf{b}\mathbf{b}') - \bar{\Upsilon} \odot \mathbf{g}\mathbf{g}'$ is positive semi-definite and $Q_1$ is PD;
- **A-DCC**: $(1 - a^2 - b^2)\bar{P} - g^2\bar{\Upsilon}$ is positive semi-definite and $Q_1$ is PD.

The A-DCC condition will hold when $a^2 + \delta g^2 + b^2 < 1$, where $\delta = \max\{eigs(\bar{P}^{-\frac{1}{2}}\bar{\Upsilon}\bar{P}^{-\frac{1}{2}})\}$.

Estimation is quite straightforward with the AG-DCC model, since a two-step procedure is used and conditions are given (as above) to ensure the PD of $Q_t$. Having the flexibility of either a more general model (capturing more effects) or a more restrictive model (reducing parameter numbers) is also an advantage. However, similar to the ADC model, the asymmetric effects in conditional correlations only exist when the standardised residuals are of the same sign. Thus they do not consider asymmetry when these residuals are of opposite sign. Also, the parameters are difficult to interpret, because of the BEKK notation that places non-linear parameters on each correlation effect in $Q_t$. 

82
Summary

As with the symmetric multivariate GARCH models reviewed in Chapter 3, the asymmetric multivariate GARCH models reviewed here all have “good” and “bad” properties. To encapsulate the “good” properties of these models, in the following section an extension to the PVECH model from Chapter 3 will be proposed that incorporates asymmetric effects. This will generalise the asymmetric diagonal VECH model to incorporate both symmetric and asymmetric effects from other markets (e.g. spillover parameters). Parsimony will also be achieved by imposing a specific order of markets in the dataset - details of which will follow. Using Bayesian model selection, I decide on the different forms of asymmetry that is present in some selected international stock market indices.

4.3 The APVECH Model Family

In order to identify the different types of multivariate asymmetry in stock returns, while maintaining desirable features such as simplicity, parsimony and the ability to satisfy PD and CS in a straightforward manner, an extended family of multivariate GARCH models is presented in this section. The parsimonious symmetric VECH model proposed in Chapter 3 will be extended by adapting similar forms to the model of De Goeij and Marquering (2004), but improving on these forms via the inclusion of additional asymmetric terms in covariance intercepts and spill-over effects in the volatility equations.

Let $y_t = (y_{1t}, \ldots, y_{Nt})$ denote the $N$-dimensional time series of sample size $n$ at time $t$. The time series labelled 1, i.e. $y_{1t} = (y_{11}, \ldots, y_{1n})$ should be set as the largest or most significant or most influential series among those considered. The second most influential or second largest series should occupy label 2. The remaining labels are irrelevant in our model. Among practitioners and analysts alike, such choices are usually straightforward and non-controversial. In studies combining data across markets, the US market would typically be 1 and Japan might be 2, while among individual assets the role of 1 and 2 could be decided by market capitalisation rankings without much controversy. The
residual vector $\varepsilon_t$ is defined analogously to $y_t$ and $\psi_{t-1}$ denotes the information set up to time $t-1$. Motivated from many empirical studies of stock returns, starting with Bollerslev (1987), the error distribution $\varepsilon_t|\psi_{t-1}$ is chosen as a multivariate standardised Student-t distribution with $\nu$ degrees of freedom and time-varying covariance matrix $H_t = [h_{ij,t}]$:

$$
\varepsilon_t|\psi_{t-1} \sim t^*_\nu(0, H_t), \quad \varepsilon_t = \sqrt{\frac{\nu-2}{\nu}} H_t^{1/2} z_t,
$$

where the vector $z_t$ follows the multivariate Student-t distribution with mean vector 0 and covariance matrix $\frac{\nu}{\nu-2} I_N$ ($I_N$ is the $N \times N$ identity matrix). Standardising this distribution as in equation (4.7) ensures that $\varepsilon_t|\psi_{t-1}$ has a covariance matrix equal to $H_t$. The asterisk $^*$ has been included to distinguish between the distributions. Note that this is not the usual choice for multivariate models. All the models reviewed so far have employed Gaussian errors.

Asymmetry in the model will be driven, as usual, by the indicator vector $I_t = (I_{1t}, \ldots, I_{Nt})$, defined as:

$$
I_{it} = \begin{cases} 
1 & \text{if } \varepsilon_{it} < 0 \\
0 & \text{otherwise}
\end{cases}, \quad i = 1, \ldots, N,
$$

as in equation (4.2). $H_t$ contains the $N$ conditional variances plus the $N(N-1)/2$ conditional covariances between each pair $y_{it}, y_{jt}$ (i.e. $h_{ij,t} \ \forall \ i, j, t$), each of which can contain numerous parameters to capture autoregressive, spill-over and asymmetric effects.

In order to maintain parsimony, with the addition of nonlinearity to the PVECH model family from Chapter 3, the following questions were considered:

1. How does volatility in the most influential stock (market) affect volatility in smaller stocks?

2. How do movements or shocks in the most influential stock (market) nonlinearily affect the volatility in smaller stocks?
3. How do movements or shocks in the second most influential stock (market) nonlinearly affect volatility in the most influential stock (market)?

4. How do movements in pairs of markets nonlinearly and interactively affect their dynamic covariance?

These are common questions asked by analysts when modeling multivariate stock return data. By restricting our attention to these questions only, we can significantly reduce the number of parameters in the model by simply setting certain likely insignificant asymmetric effects to be zero.

The proposed model family, which is called the Asymmetric PVECH (APVECH) model, can be described by the following set of equations (∀ i ≠ j = 2, . . . , N):

\[
\begin{align*}
y_{1t} &= m_{10} + m_{11}y_{1,t-1} + \varepsilon_{1t}, \\
y_{it} &= m_{i0} + m_{ii}y_{i,t-1} + \varepsilon_{it}, \\
h_{11,t} &= c_{11} + \gamma_{11}I_{1,t-1} + (a_{11} + \alpha_{11}I_{1,t-1})\varepsilon_{1,t-1}^2 + b_{11}h_{11,t-1} \\
&\quad + (d_{12} + \delta_{12}I_{2,t-1})\varepsilon_{2,t-1}^2 + e_{12}h_{22,t-1}, \\
h_{ii,t} &= c_{ii} + \gamma_{ii}I_{i,t-1} + (a_{ii} + \alpha_{ii}I_{i,t-1})\varepsilon_{i,t-1}^2 + b_{ii}h_{ii,t-1} \\
&\quad + (d_{ii} + \delta_{ii}I_{i,t-1})\varepsilon_{1,t-1}^2 + e_{ii}h_{11,t-1}, \\
h_{ij,t} &= c_{ij} + \gamma_{ij1}I_{i,t-1}I_{j,t-1} + \gamma_{ij2}I_{i,t-1}(1 - I_{j,t-1}) + \gamma_{ij3}(1 - I_{i,t-1})I_{j,t-1} + b_{ij}h_{ij,t-1} \\
&\quad + [a_{ij} + \alpha_{ij1}I_{i,t-1}I_{j,t-1} + \alpha_{ij2}I_{i,t-1}(1 - I_{j,t-1}) + \alpha_{ij3}(1 - I_{i,t-1})I_{j,t-1}]\varepsilon_{i,t-1}\varepsilon_{j,t-1}.
\end{align*}
\]

Below are some comments about this model:

- Volatility includes the usual GARCH(1, 1) equation with intercept \(c_{ii}\), ARCH effect \(a_{ii}\) and GARCH effect \(b_{ii}\). For \(i = 1\) spill-over effects are allowed from the second stock’s volatility (via \(e_{12}\)) and squared shock (\(d_{12}\)). For \(i > 1\) equivalent spill-over effects come from stock \(i = 1\). Asymmetry is allowed via threshold nonlinear effects from the stock itself, allowing the intercept to change to \(c_{ii} + \gamma_{ii}\) and the ARCH effect to become \(a_{ii} + \alpha_{ii}\) when \(I_{i,t-1} = 1 \equiv \varepsilon_{i,t-1} < 0\). Asymmetric
spill-over effects from stock 1 are allowed, driven by stock 1, via changing \((d_{i1})\) to \((d_{i1} + \delta_{i1})\) when \(I_{1,t-1} = 1\). There are four possible regimes here, depending on the combinations \((I_{1,t-1}, I_{i,t-1}) = (0,0); (0,1); (1,0); (1,1)\). Thus different responses in volatility are allowed in each stock depending on whether stock \(i\) and stock 1 are both rising unexpectedly, both falling unexpectedly, or one falling and the other rising unexpectedly.

- Covariance for \(y_{it}\) and \(y_{jt}\), denoted \(h_{ij,t}\), again includes simple GARCH(1,1)-type dynamics with intercept, “ARCH” and “GARCH” parameters \((c_{ij}, a_{ij}\) and \(b_{ij}\)). Asymmetry is now driven by the interaction between the two markets in question via the indicators \((I_{i,t-1}, I_{j,t-1})\). Again, the intercept and “ARCH” effects are allowed to change, in response to combinations of each stock unexpectedly rising and/or falling; with four regimes present again.

- Each mean equation exhibits a simple AR(1) process, as in the PVECH model of Chapter 3. I have chosen to focus on volatility and covariance asymmetry in the APVECH model, rather than mean asymmetry, which will be examined in detail in Chapter 5.

Such asymmetric and interactive volatility effects between large markets (e.g. US) and domestic markets have been found in the univariate analyses of Chen et al. (2005) and Chen et al. (2006b). De Goeij and Marquering (2004) finds asymmetric and interactive volatility and covariances between stock and bond markets, and also assume the effects \(b_{ii}, e_{ii}, b_{ij}\) are symmetric only, for parsimony. These studies motivated the choices of which parameters to allow to have nonlinear asymmetric effects. The APVECH model is similar, in spirit only, to that in De Goeij and Marquering (2004). For the model proposed here: the parameters all enter linearly, are easily interpretable and appear in one equation only, as opposed to the BEKK extended form in Kroner and Ng (1998) and Cappiello et al. (2006) whose parameters are repeated across equations, appear as products with other parameters (i.e. nonlinear) and are generally difficult to interpret quantitatively. Apart
from the simple form of De Goeij and Marquering (2004), asymmetric VECH models, such as the one proposed here, have not appeared in the literature, to the best of my knowledge, mainly due to the difficulty in enforcing PD and CS in this model form and maintaining parsimony.

4.3.1 Unconditional Properties and the Likelihood Function

The unconditional properties of this model family will now be investigated. Later, these are used to help enforce the necessary and sufficient conditions for PD and CS.

Because the APVECH model has independent AR(1) equations for each mean, the unconditional mean vector $\mu_1 = E(y_t)$ for this model is the same as that for the PVECH model in equation (3.16), i.e.

$$\mu_1 = [I_N - \text{diag}(m_1)]^{-1}m_0$$

where $m_0 = (m_{10}, \ldots, m_{N0})$ and $m_1 = (m_{11}, \ldots, m_{NN})$ are $N \times 1$ parameter vectors. Equivalently in element form,

$$\mu_i = \frac{m_{i0}}{1 - m_{ii}}, \quad \forall \ i = 1, \ldots, N.$$

The unconditional covariance matrix $E(H_t)$ can be derived by examining the conditional variances and covariances separately: To derive the unconditional variances, we begin with

$$\text{Var}(y_{it}) = \text{Var}(m_{i0} + m_{ii}y_{i,t-1} + \varepsilon_{it})$$

$$= m_{ii}^2 \text{Var}(y_{it}) + \text{Var}(\varepsilon_{it})$$

$$\Rightarrow \text{Var}(y_{it}) = \frac{\text{Var}(\varepsilon_{it})}{1 - m_{ii}^2}, \quad \forall \ i = 1, \ldots, N.$$

Now $\text{Var}(\varepsilon_{it}) = E(\varepsilon_{it}^2) = E(h_{it,t})$, which gives
\[ E(h_{11,t}) = c_{11} + \gamma_{11}E(I_{1t}) + (a_{11} + \alpha_{11}E(I_{1t}))E(\varepsilon_{11}^2) + b_{11}E(h_{11,t}) + (d_{12} + \delta_{12}E(I_{2t}))E(\varepsilon_{22}^2) + e_{12}E(h_{22,t}), \]

\[ E(h_{ii,t}) = c_{ii} + \gamma_{ii}E(I_{it}) + (a_{ii} + \alpha_{ii}E(I_{it}))E(\varepsilon_{ii}^2) + b_{ii}E(h_{ii,t}) + (d_{ii} + \delta_{ii}E(I_{it}))E(\varepsilon_{ii}^2) + e_{ii}E(h_{11,t}). \]

Under the assumption that \( \varepsilon_{it} \) (and therefore \( z_{it} \)) is symmetric about zero and knowing that \( I_{it} \) is a Bernoulli random variable, we have \( E(I_{it}) = \frac{1}{2} \). Therefore

\[ E(h_{11,t}) = c_{11} + \frac{1}{2}\gamma_{11} + (a_{11} + \frac{1}{2}\alpha_{11} + b_{11})E(h_{11,t}) + (d_{12} + \frac{1}{2}\delta_{12} + e_{12})E(h_{22,t}), \]

\[ E(h_{ii,t}) = c_{ii} + \frac{1}{2}\gamma_{ii} + (a_{ii} + \frac{1}{2}\alpha_{ii} + b_{ii})E(h_{ii,t}) + (d_{ii} + \frac{1}{2}\delta_{ii} + e_{ii})E(h_{11,t}). \]

Let \( \lambda_{rr} = c_{rr} + \frac{1}{2}\gamma_{rr}, \omega_{rr} = a_{rr} + \frac{1}{2}\alpha_{rr} + b_{rr} \) and \( \kappa_{rs} = d_{rs} + \frac{1}{2}\delta_{rs} + e_{rs} \) for any positive integers \( r \) and \( s \). By rearrangement of the above equations, we obtain

\[ E(h_{11,t}) = \frac{\lambda_{11}(1 - \omega_{22}) + \lambda_{22}\kappa_{12}}{(1 - \omega_{11})(1 - \omega_{22}) - \kappa_{12}\kappa_{21}} \quad \text{and} \quad E(h_{ii,t}) = \frac{\lambda_{ii} + \kappa_{ii}E(h_{11,t})}{1 - \omega_{ii}} \]

and hence the unconditional variances for the model are

\[ \text{Var}(y_{1t}) = \frac{\lambda_{11}(1 - \omega_{22}) + \lambda_{22}\kappa_{12}}{[(1 - \omega_{11})(1 - \omega_{22}) - \kappa_{12}\kappa_{21}](1 - \omega_{11})}, \]

\[ \text{Var}(y_{it}) = \frac{\lambda_{ii} + \kappa_{ii}E(h_{11,t})}{(1 - \omega_{ii})(1 - \omega_{11})}, \quad \forall \ i = 2, \ldots, N. \]

Now for the unconditional covariances, we begin with

\[ \text{Cov}(y_{1t}, y_{jt}) = \text{Cov}(m_{00} + m_{1t}y_{it-1} + \varepsilon_{it}, m_{0j} + m_{jj}y_{jt-1} + \varepsilon_{jt}) = m_{1j}m_{jj}\text{Cov}(y_{1t}, y_{jt}) + \text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) \]

\[ \Rightarrow \text{Cov}(y_{1t}, y_{jt}) = \frac{\text{Cov}(\varepsilon_{it}, \varepsilon_{jt})}{1 - m_{ii}m_{jj}}, \quad \forall \ i < j = 1, \ldots, N. \]

Now \( \text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = E(\varepsilon_{it}\varepsilon_{jt}) = E(h_{ij,t}) \), which gives
\[
E(h_{ij,t}) = c_{ij} + \gamma_{ij,1}E(I_{it}I_{jt}) + \gamma_{ij,2}E[I_{it}(1 - I_{jt})] + \gamma_{ij,3}E[(1 - I_{it})I_{jt}]
+ a_{ij}E(\varepsilon_{it}\varepsilon_{jt}) + \alpha_{ij,1}E[I_{it}I_{jt}\varepsilon_{it}\varepsilon_{jt}] + \alpha_{ij,2}E[I_{it}(1 - I_{jt})\varepsilon_{it}\varepsilon_{jt}]
+ \alpha_{ij,3}E[(1 - I_{it})I_{jt}\varepsilon_{it}\varepsilon_{jt}] + b_{ij}E(h_{ij,t}),
\]

(4.8)

Analogous to each \(I_{it}\) determining the position of each residual either side of zero, each \(I_{it}\) and \(I_{jt}\) together determine which “quadrant” the residual vector \((\varepsilon_{it}, \varepsilon_{jt})\) appears in the corresponding 2-dimensional plane. In fact, determining the exact unconditional covariance for the APVECH model is a very difficult problem - this is because each of the expectations in equation (4.8) involving \(I_{it}\) turn out to be non-linear functions of the conditional covariances which do not simplify as well as that in the unconditional variance calculations (for example). To see why this is the case, refer to the Appendix of this chapter for derivations of the expectations in equation (4.8). For the purposes of fitting the APVECH model to stock returns and enforcement of covariance stationarity, it is much easier to make the following assumption, which does not appear to affect the model’s performance overall (e.g. see Sections 4.6 and 4.7): In the ADVECH model by De Goeij and Marquering (2004) (reviewed in Section 4.2), a stationarity condition imposed during estimation was presented in this paper “assuming that errors are equally distributed around zero for the quadrants”. This assumptions means that there are an equal number of residual observations in each of the four quadrants, implying that

\[
E[(1 - I_{it})(1 - I_{jt})] = E(I_{it}I_{jt}) = E[I_{it}(1 - I_{jt})] = E[(1 - I_{it})I_{jt}] = \frac{1}{4}.
\]

Applying this restriction to the APVECH model, equation (4.8) then becomes

\[
E(h_{ij,t}) = c_{ij} + \frac{1}{4}\gamma_{ij,1} + \frac{1}{4}\gamma_{ij,2} + \frac{1}{4}\gamma_{ij,3} + a_{ij}E(h_{ij,t}) + \frac{1}{4}\alpha_{ij,1}E(h_{ij,t})
+ \frac{1}{4}\alpha_{ij,2}E(h_{ij,t}) + \frac{1}{4}\alpha_{ij,3}E(h_{ij,t}) + b_{ij}E(h_{ij,t})
\Rightarrow E(h_{ij,t}) = \frac{c_{ij} + \frac{1}{4}\gamma_{ij,1} + \frac{1}{4}\gamma_{ij,2} + \frac{1}{4}\gamma_{ij,3}}{1 - a_{ij} - \frac{1}{4}\alpha_{ij,1} - \frac{1}{4}\alpha_{ij,2} - \frac{1}{4}\alpha_{ij,3} - b_{ij}}.
\]

89
Hence, under the assumption of De Goeij and Marquering (2004), the unconditional covariance of the model can be approximated by

$$\text{Cov}(y_{it}, y_{jt}) = \frac{c_{ij} + \frac{1}{4} \gamma_{ij,1} + \frac{1}{4} \gamma_{ij,2} + \frac{1}{4} \gamma_{ij,3}}{(1 - a_{ij} - \frac{1}{4} \alpha_{ij,1} - \frac{1}{4} \alpha_{ij,2} - \frac{1}{4} \alpha_{ij,3} - b_{ij})(1 - m_{ii} m_{jj})}, \quad \forall \ i < j = 1, \ldots, N.$$ 

Therefore

$$E(H_t) = \begin{cases} \text{Var}(y_{it}) & \text{if } i = j, \\ \text{Cov}(y_{it}, y_{jt}) & \text{if } i \neq j. \end{cases}$$

We set the initial covariance matrix as $H_1 = E(H_t)$, which again is independent of $t$. The conditional covariance matrix $H_t, t = 2, \ldots, n$ is then defined in (4.7).

Let $\mathbf{y}$ denote the data and $\mathbf{\theta}$ denote the set of all parameters in the model. Using equation (4.7), the likelihood function is the product of the $n$ multivariate standardised Student-$t$ PDFs:

$$p(\mathbf{y} | \mathbf{\theta}) = \left\{ \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) [(\nu-2)\pi]^\frac{N}{2}} \right\}^n \prod_{t=1}^{n} |H_t|^{-\frac{1}{2}} \left(1 + \frac{\varepsilon_t^\prime H_t^{-1} \varepsilon_t}{\nu - 2}\right)^{-\frac{\nu+N}{2}} $$

with the log likelihood function taking the form

$$\ln p(\mathbf{y} | \mathbf{\theta}) = n \ln \left\{ \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) [(\nu-2)\pi]^\frac{N}{2}} \right\} - \frac{1}{2} \sum_{t=1}^{n} \ln |H_t|$$

$$- \frac{\nu + N}{2} \sum_{t=1}^{n} \ln \left(1 + \frac{\varepsilon_t^\prime H_t^{-1} \varepsilon_t}{\nu - 2}\right).$$

The initial values $\varepsilon_1$ and $H_1$ are set to their expectations above.

### 4.3.2 Necessary and Sufficient Conditions for PD and CS

The (sufficient) conditions usually imposed to ensure CS on multivariate GARCH-type models are directed on the parameters themselves (see for e.g. Bollerslev et al., 1988 and De Goeij and Marquering, 2004), while no (parameter) conditions exist on the VECH model in general to ensure PD. Instead, conditions usually imposed are that most pa-
rameters are restricted to be positive. However, these conditions may not be necessary for PD and/or CS: parameter values lying outside these restrictions could still achieve PD and CS data realisations - this was demonstrated in Chapter 3. As in Chapter 3, the conditions of PD and CS will be applied directly when fitting the APVECH model. These conditions are given in (3.22) and (3.23) in Chapter 3, but are replicated below for convenience:

- **Positive Definiteness**: Each conditional variance value and all eigenvalues of each conditional covariance matrix are strictly positive:

  \[ h_{ii,t} > 0 \quad \text{and} \quad \text{eigs}(H_t) > 0, \forall \ t = 1, \ldots, n; \ i = 1, \ldots, N. \]

- **Covariance Stationarity**: The unconditional mean and covariance matrices exist and are finite and positive definite:

  \[ \mu_1, H_1 \text{ are finite}; \quad \text{eigs}(H_1) > 0. \]

These conditions can be directly enforced during estimation by incorporating them into the prior distribution under a Bayesian approach; i.e. the prior distribution will be defined only on the region where (3.22) and (3.23) are satisfied, as detailed in the next section.

Under an MCMC approach, parameter values can be proposed and then accepted or rejected depending on whether they satisfy any set of conditions: this is a simple way of constraining the posterior to the required parameter space. Thus, each set of MCMC iterates produced can be simply rejected if PD and CS are not satisfied via (3.22) and (3.23). In other words, PD and CS will be measured and satisfied computationally, by ensuring that each accepted set of MCMC parameter iterates empirically satisfy (3.22) and (3.23). This flexibility allows valid and efficient inference under the definitions for PD and CS. Modern computer programming languages, such as Fortran 95, Matlab, etc, are highly capable of numerically measuring whether the elements of a vector or matrix
are practically finite or instead practically infinite. Any set of parameters generated via
the MCMC method (described in the next section) that lead to $\mu_1$, $H_1$ containing any
practically infinite values, or any of $\text{eigs}(H_t) < 0$ or $h_{ii,t} < 0$ will simply be rejected.
Thus inference will be performed by implicitly considering the region where (3.22) and
(3.23) are satisfied. Thus inference on the necessary and sufficient parameter region, even
though it cannot be written down explicitly, will be performed.

4.4 Bayesian Estimation Procedures

Bayesian inference is made via the joint posterior distribution $p(\theta|y)$. An adaptive
MCMC sampling scheme is designed, incorporating the random walk Metropolis and
independent proposal Metropolis-Hastings methods, as well as the delayed rejection sam-
pler of Tierney and Mira (1999), to ensure efficient mixing and quick convergence to the
stationary distribution. Parameters are generated in blocks, where practical and possible,
which also aids mixing (see for e.g. Carter and Kohn, 1994). This sampling scheme is an
extension of that proposed in Chapter 3.

4.4.1 The Gibbs Sampler

Let $\theta_k$ denote the parameter vector corresponding to the $k$th equation of the model family
(i.e. mean, variance and covariance). The contribution here is in finding the optimal way
to distribute the parameters among the blocks. Too many parameters in each block can
lead to very low MH acceptance rates, while too few in each block can lead to slow mixing.
Further, different ways of combining parameters among blocks can significantly affect
mixing and convergence rates in the MCMC sampler. After much experimentation, I
found that combining parameters from each individual volatility and covariance equations
as separate blocks to be the most efficient design. As such, simulation is undertaken
recursively from the following joint conditional posterior distributions:
We thus simulate recursively all the mean equation parameters across the $N$ equations as a block; the variance parameters in each of the $N$ volatility equations in $N$ blocks; then the covariance parameters in each of the $N(N-1)/2$ covariance equations in $N(N-1)/2$ blocks; and finally the inverse of the degrees of freedom parameter. We invert $\nu$, as in Chen et al. (2005), since the resulting parameter space has a finite range - we infer on $\nu$ by simply inverting the MCMC iterates. Performing the scheme for a total of $D$ iterations, after burn-in, results in a sample from $p(\theta|y)$.

### 4.4.2 Prior and Posterior Distributions

From Bayes’ rule, conditional posterior distributions are given by:

$$p(\theta_k|y, \theta_{\neq k}) \propto p(y|\theta)p(\theta_k|\theta_{\neq k}), \quad k = 1, \ldots, N(N + 1)/2 + 2;$$

where $p(\theta_k|\theta_{\neq k})$ is the prior distribution for each parameter grouping. Using an approach similar to that of Chapter 3, the prior distributions will be assumed uniform:

$$p(\theta_k|\theta_{\neq k}) \propto I(A), \quad k = 1, \ldots, N(N + 1)/2 + 2 \quad (4.10)$$

where $A$ is the region implicitly defined by (3.22)-(3.23), plus the region $0 < \tau < 0.5$ which
ensures the likelihood is finite in \( \nu \). Thus each posterior distribution is simply proportional to the likelihood function, allowing the data to dominate our inferences, inside the region permitted by PD and CS being satisfied. Even though we cannot explicitly write down this prior in terms of parameter values, it can easily be enforced in an MCMC analysis by simply rejecting any parameter values where PD and CS are not satisfied (empirically), as described above. The prior then implicitly constrains the posterior to lie only on the region defined by (3.22)-(3.23). Thus PD and CS are guaranteed, while no unnecessary parameter constraints are enforced.

4.4.3 Conditional Posterior Sampling

As in Chapter 3, the Metropolis-Hastings (MH) and Delayed Rejection (DR) algorithms will be applied to simulate from each conditional posterior given in Section 4.4.1, since these distributions are of unknown form. Adaptive MCMC sampling will be applied, as a burn-in period will be used to assist in the convergence of parameter simulations to the target distribution, with a sampling period used to make inference - this approach was also used in Chapter 3.

Each conditional posterior distribution will be simulated as follows:

The Burn-In Period

Begin with some initial guesses \( \theta^{[1]}_i, i = 1, \ldots, N(N+1)/2 + 2 \). For the burn-in iterations \( j = 1, \ldots, w \) do the following:

1. Draw a random sample \( \theta_{\pi}^{[j]} \) from the proposal distribution:

\[
g_1(\theta_i|\theta^{[j-1]}_i) \sim N(\theta^{[j-1]}_i, c^{[j]}_i \Omega_i)
\]

i.e. a multivariate normal distribution with mean vector \( \theta^{[j-1]}_i \) and covariance matrix \( c^{[j]}_i \Omega_i \). Having the proposal mean set to the previous MCMC iterate means that this is a random walk Metropolis step. The matrix \( \Omega_i \) is a diagonal matrix with
“large” diagonal entries (see Chapter 3), and $c_i^{[j]}$ is a real number initially set to 1.

2. Draw a random sample $u_i^{[j]}$ from Unif[0, 1] and set the following Metropolis acceptance probability:

$$\alpha_1(\theta_i^{[j-1]}, \theta_i^{p1}) = \min \left\{ 1, \frac{p(\theta_i^{p1}|y, \theta_\neq i)}{p(\theta_i^{[j-1]}|y, \theta_\neq i)} \right\}. \quad (4.11)$$

3. If $u_i^{[j]} < \alpha_1$, set $\theta_i^{[j]} = \theta_i^{p1}$. Otherwise, set $\theta_i^{[j]} = \theta_i^{[j-1]}$, then return to Step 1 for the next $i$.

As in Chapter 3, at the end of every $k$th iteration (for a pre-specified $k$), calculate the AR from the previous $k$ iterations as follows

$$AR_i^{[j]} = \frac{1}{k} \sum_{j-k+1}^j I(\theta_i^{[j]} = \theta_i^{p1} \cup \theta_i^{[j]} = \theta_i^{p2}) \quad \forall \ j = k, 2k, \ldots, w-k.$$ 

For the next set of $k$ iterations set the following values for $c_i^{[j+1]}$, $\forall \ j = k, 2k, \ldots, w-k$:

$$c_i^{[j+1]} = \begin{cases} 
\left( \frac{1}{2} + \frac{j}{2w} \right) c_i^{[j]} & \text{if } AR_i^{[j]} < 0.15, \\
\frac{1}{2} c_i^{[j]} & \text{if } 0.15 \leq AR_i^{[j]} \leq 0.5, \\
\left( 2 - \frac{j}{w} \right) c_i^{[j]} & \text{if } AR_i^{[j]} > 0.5.
\end{cases}$$

As in Chapter 3, the threshold values of 0.15 and 0.5 were chosen to give the final ARs a greater chance of reaching the optimal AR of 0.234 by Roberts et al. (1997).

The Sampling Period

Let $\bar{\theta}_i$ and $\hat{\Sigma}_i$ denote the sample means and covariances of the iterations in the burn-in period for each parameter group $i$. For iterations $j = w+1, \ldots, J$ do the following:

1. Draw a random sample $\theta_i^{p1}$ from the proposal distribution:

$$g_1(\theta_i|\theta_i^{[j-1]}) \sim N(\bar{\theta}_i, \hat{\Sigma}_i).$$
2. Draw a random sample $u_1^{[j]}$ from Unif$[0,1]$ and set the following MH acceptance probability:

$$
\alpha_1(\theta_1^{[j-1]}, \theta_1^{p_1}) = \min \left\{ 1, \frac{p(\theta_1^{p_1} | y, \theta_1 \neq i)g_1(\theta_1^{[j-1]} | \theta_1^{p_1})}{p(\theta_1^{[j-1]} | y, \theta_1 \neq i)g_1(\theta_1^{p_1} | \theta_1^{[j-1]})} \right\}.
$$

3. If $u_1^{[j]} < \alpha_1$, set $\theta_1^{[j]} = \theta_1^{p_1}$ and return to Step 1 for the next $i$. Otherwise, perform a delayed rejection step:

4. Draw a random sample $\theta_1^{p_2}$ from the proposal distribution:

$$
g_2(\theta_1 | \theta_1^{[j-1]}, \theta_1^{p_1}) \sim N \left( \frac{1}{2} (\bar{\theta}_1 + \theta_1^{[j-1]}), \frac{1}{2} \hat{\Sigma}_1 \right).
$$

5. Draw a random sample $u_2^{[j]}$ from Unif$[0,1]$ and set the following MH acceptance probability:

$$
\alpha_2(\theta_1^{[j-1]}, \theta_1^{p_1}, \theta_1^{p_2}) = \min \left\{ 1, \frac{N_2}{D_2} \right\}
$$

where

$$
\frac{N_2}{D_2} = \frac{p(\theta_1^{p_2} | y, \theta_1 \neq i)g_1(\theta_1^{p_1} | \theta_1^{p_2})g_2(\theta_1^{[j-1]} | \theta_1^{p_2}, \theta_1^{p_1})[1 - \alpha_1(\theta_1^{p_2}, \theta_1^{p_1})]}{p(\theta_1^{[j-1]} | y, \theta_1 \neq i)g_1(\theta_1^{p_1} | \theta_1^{[j-1]})g_2(\theta_1^{p_2} | \theta_1^{[j-1]}, \theta_1^{p_1})[1 - \alpha_1(\theta_1^{[j-1]}, \theta_1^{p_1})]}.
$$

6. If $u_2^{[j]} < \alpha_2$, set $\theta_1^{[j]} = \theta_1^{p_2}$. Otherwise, set $\theta_1^{[j]} = \theta_1^{[j-1]}$. Return to Step 1 for the next $i$.

The set of simulations $\left\{ \theta_i^{[w+1]}, \ldots, \theta_i^{[j]} \right\}_{i=1}^{N(N+1)/2+2}$ is then a sample from the joint posterior distribution $p(\theta | y)$ of the APVECH model.

### 4.5 Bayesian Model Selection

As mentioned in Chapter 1, there are many techniques existing in the literature that can determine the “best” model in a problem from a group of competing models, both
from classical and Bayesian perspectives. Perhaps the most commonly used classical model selection techniques applied to GARCH models are the ones involving information criterion functions: Let $\hat{\theta}_k$ denote the parameter estimates calculated from fitting model $k$ to the data $y$, where $k = 1, \ldots, K$ (a total of $K$ competing models). Akaike (1974) proposes the Akaike Information Criterion (AIC) function which is defined by

$$\text{AIC}_k = -2 \ln p(y|\hat{\theta}_k) + 2P_k$$

where $p(y|\hat{\theta}_k)$ is the likelihood function value for model $k$, while $P_k$ is the number of parameters in model $k$. The model that has the smallest AIC value is typically the model selected for the problem. The purpose of the $2P_k$ is to “penalise” models containing too many parameters. The majority of information criterion functions differ based on this penalty. For example, the Bayesian Information Criterion (BIC) function of Schwarz (1978) is defined by

$$\text{BIC}_k = -2 \ln p(y|\hat{\theta}_k) + P_k \ln(n)$$

where $n$ is the sample size. Thus the BIC calculated for models with a large number of parameters will have a larger penalty than that from the AIC.

From a Bayesian perspective, model selection is performed by calculating posterior model probabilities $\Pr(M = k|y)$, for $k = 1, \ldots, K$, and the model with maximum posterior probability is chosen. Under the discrete assumption of Bayes’ rule, this probability can be written as

$$\Pr(M = k|y) = \frac{p(y|M = k) \Pr(M = k)}{\sum_{i=1}^K p(y|M = i) \Pr(M = i)} \quad (4.12)$$

where $\Pr(M = i)$ denotes the prior model probability for model $i$ and $p(y|M = i)$ denotes the corresponding marginal likelihood function. As the setting of $\Pr(M = i)$ is typically subjective, the calculation of the marginal likelihood $p(y|M = i)$ has achieved
the majority of focus in the literature. Analytically it can be calculated as

\[ p(y|M = i) = \int_{\theta_i} p(y|\theta_i, M = i)p(\theta_i|M = i)d\theta_i \]  \hspace{1cm} (4.13)

where \( \theta_i \) denotes the vector of parameters for model \( i \). However, exact analytic evaluation of the integral in (4.13) is possible only in elementary cases, for example exponential family distributions with conjugate priors (see Kass and Raftery, 1995). Various numerical approximations of the marginal likelihood has thus been explored, some examples of which are given below:

### 4.5.1 Marginal Likelihood Approximations

To approximate the marginal likelihood \( p(y|M = i) \) for model \( i \), the following approaches have been considered:

**Monte Carlo Approximations**

By simply averaging the likelihoods from sampled parameter values, a simple Monte Carlo estimate of \( p(y|M = i) \) can be obtained as follows:

\[ \hat{p}(y|M = i) = \frac{1}{J} \sum_{j=1}^{J} p(y|\theta_i^{[j]}, M = i) \]  \hspace{1cm} (4.14)

where \( \theta_i^{[j]} \) represents the \( j \)th parameter value simulated from the prior distribution \( p(\theta_i|M = i) \), for \( j = 1, \ldots, J \). A similar estimator is the harmonic mean of likelihood values proposed by Newton and Raftery (1994), given by

\[ \hat{p}(y|M = i) = \left[ \frac{1}{J} \sum_{j=1}^{J} \left( \frac{1}{p(y|\theta_i^{[j]}, M = i)} \right) \right]^{-1}. \]  \hspace{1cm} (4.15)

Kass and Raftery (1995) mention that the variance of the estimate in (4.14) can be quite large, especially when the posterior is concentrated relative to the prior, resulting in
many small likelihood values in combination with a few large likelihood values. Similarly, Chib (1995) mentions the instability of the estimate in (4.15) as the inverse likelihood does not have finite variance. An alternative estimator that can avoid these issues is the one calculated via importance sampling, which uses a weighted average based on a user-defined function \( p^* (\theta^{[j]}_i | M = i) \) as follows:

\[
\hat{p}(y|M = i) = \frac{\sum_{j=1}^{J} w_j p(y | \theta^{[j]}_i, M = i)}{\sum_{j=1}^{J} w_j} \tag{4.16}
\]

where \( w_j = p(\theta^{[j]}_i | M = i)/p^* (\theta^{[j]}_i | M = i) \). The variance of (4.16) can thus be controlled by choosing an appropriate function for \( p^* (\theta^{[j]}_i | M = i) \). Gerlach et al. (1999) use importance sampling to estimate marginal likelihood terms for a family of autoregressive time series models.

**Approximation from Gibbs Sampler Output**

Chib (1995) proposes to approximate marginal likelihoods by utilising output from a Gibbs sampler that has been used to generate samples from the joint posterior distribution \( p(\theta_i | y, M = i) \). The first step of this is to make the marginal likelihood the subject of the Bayes’ rule formula as follows:

\[
p(y|M = i) = \frac{p(y | \theta_i, M = i)p(\theta_i | M = i)}{p(\theta_i | y, M = i)}
\]

and then convert to the “computationally convenient” logarithm scale:

\[
\ln p(y|M = i) = \ln p(y | \theta_i, M = i) + \ln p(\theta_i | M = i) - \ln p(\theta_i | y, M = i) \tag{4.17}
\]

This approach requires that the samples from the joint posterior distribution \( p(\theta_i | y, M = i) \) were generated from \( B \) blocks of conditional posteriors in a Gibbs sampling scheme: Define the parameter vector \( \theta_i \) for the \( i \)th model in such a way that it can be decomposed
into $B$ parameter blocks, i.e. $\theta_i = \{\theta_{i1}, \ldots, \theta_{iB}\}$. Hence the conditional posteriors are

1. $p(\theta_{i1}|y, \theta_{\neq i1}, M = i)$
   
   $\vdots$

   $b$. $p(\theta_{ib}|y, \theta_{\neq ib}, M = i)$
   
   $\vdots$

   $B$. $p(\theta_{iB}|y, \theta_{\neq iB}, M = i)$. 

Now the joint posterior can be written as the product of the conditional posteriors

$$p(\theta_i|y, M = i) = p(\theta_{i1}|y, M = i) \times p(\theta_{i2}|y, \theta_{i1}, M = i) \times \ldots \times p(\theta_{iB}|y, \theta_{i1}, \ldots, \theta_{i,B-1}, M = i).$$

A Monte Carlo estimate of each conditional posterior $p(\theta_{ib}|y, \theta_{i1}, \ldots, \theta_{i,b-1}, M = i)$ is

$$p(\theta_{ib}|y, \theta_{i1}, \ldots, \theta_{i,b-1}, M = i) = \frac{1}{J} \sum_{j=1}^{J} p(\theta_{ib}|y, \theta_{i1}, \ldots, \theta_{i,b-1}, \theta_{i,b+1}^{[j]}, \ldots, \theta_{iB}^{[j]}, M = i)$$

and when substituted into equation (4.17), we obtain an approximation for the (log) marginal likelihood:

$$\ln \hat{p}(y|M = i) = \ln p(y|\theta_i, M = i) + \ln p(\theta_i|M = i) - \sum_{b=1}^{B} \ln p(\theta_{ib}|y, \theta_{i1}, \ldots, \theta_{i,b-1}, M = i).$$

This technique eliminates the need to control the variance of marginal likelihood estimate, since information from the joint posterior distribution is included in the calculations. In applications, however, this technique requires knowledge of the normalising constants of the full conditional posteriors. This means that the posteriors must be of a known
form (e.g. Gaussian, Student-t, etc). Chib and Jeliazkov (2001) extend this technique to include the case where knowledge of the conditional posteriors was obtained via MH methods. This simply adjusts equation (4.18) to include relevant acceptance probabilities and proposal distributions obtained from the full conditional posterior.

4.5.2 Reversible-Jump MCMC

Green (1995) proposes the reversible-jump MCMC approach to calculating posterior model probabilities. The idea is to generate samples from each posterior distribution \( p(\theta_k | y, M = k) \) and \( \Pr(M = k | y) \), \( k = 1, \ldots, K \) via a Metropolis-Hastings algorithm that “jumps” between each of the \( K \) model spaces. There is essentially two additional requirements compared to that of the standard MH acceptance probability from equation (2.2) in order to move from model \( l \) to model \( m \):

- A one-to-one bijective transformation that defines exactly how the sampling scheme should move between model spaces. This transformation is given by the functions \( u_l = f_l(\theta_l, \theta_m) \) and \( u_m = f_m(\theta_l, \theta_m) \).

- The probability of making a jump of this type, denoted by \( \Pr_J(\theta_l, M = l) \) and \( \Pr_J(\theta_m, M = m) \).

So to jump between model \( l \) and model \( m \) in a sampling scheme at position \( j \), the following is performed:

1. Draw a proposed value \( \theta^p_m \) from the proposal distribution \( g_l(u_l) \).

2. Draw a random sample \( v \) from \( \text{Unif}[0, 1] \) and set the following acceptance probability:

\[
\alpha = \min \left\{ 1, \frac{p(\theta^p_m | y, M = m) \Pr_J(\theta_m, M = m) g_m(u^m_{j-1})}{p(\theta^p_j^{j-1} | y, M = l) \Pr_J(\theta_l, M = l) g_l(u^l_j)} \left| \frac{\partial(\theta_m, u_m)}{\partial(\theta_l, u_l)} \right| \right\}
\]

where \( \left| \frac{\partial(\theta_m, u_m)}{\partial(\theta_l, u_l)} \right| \) denotes the Jacobian of the bijection transformation.
3. If \( v < \alpha \), jump to model space \( m \) and update \( \theta_m \). Otherwise, remain at model space \( l \) and update \( \theta_l \).

Papers by Vrontos et al. (2000) and Chen et al. (2005) have successfully implemented this technique in GARCH model selection.

### 4.5.3 Congdon’s Method

The Bayesian model selection techniques just discussed have been highly successful in the literature. However (i) all involve certain subjective choices that the final results can be highly sensitive to, and (ii) most involve running the MCMC chain for more iterations, or multiple times, for each model. The approximate approach by Congdon (2006) will be employed on the APVECH model family proposed in this chapter to detect for asymmetric effects in stock returns. This method simply relies on having available independently sampled MCMC chains from each model under consideration and no extra complexity is introduced by having more than two models under comparison.

Let \( \theta = \left\{ \theta^{[j]}_1, \ldots, \theta^{[j]}_K \right\}_{j=1}^J \) denote MCMC samples taken independently from the posterior distributions \( \{p(\theta_k|y, M = k)\}_{k=1}^K \) of the \( K \) considered models. The posterior model probability \( \Pr(M = k|y) \) can be written in integral notation as

\[
\Pr(M = k|y) = \int \Pr(M = k|y, \theta)p(\theta|y)d\theta.
\]

Under some apparently non-standard and controversial assumptions (discussed below), Congdon (2006) shows that, evaluated at MCMC iteration \( j \), this integral can be approximately estimated via:

\[
\Pr(M = k|y, \theta^{[j]}_k) = \frac{p(y|\theta^{[j]}_k, M = k)p(\theta^{[j]}_k|M = k)\Pr(M = k)}{\sum_{i=1}^K p(y|\theta^{[j]}_i, M = i)p(\theta^{[j]}_i|M = i)\Pr(M = i)}.
\] (4.19)

An approximate Monte Carlo estimate of \( \Pr(M = k|y) \) is then:
\[
\hat{Pr}(M = k|y) = \frac{1}{J} \sum_{j=1}^{J} Pr(M = k|y, \theta^j_k)
\] (4.20)

i.e. the average of the conditional model posteriors over the iterations.

Robert and Marin (2008) point out that this method is approximate since sampling from the distribution \(p(\theta_k|y, M = k)\) independently is not achieving a sample from the joint model posterior \(p(\theta|y)\) (not conditioning on any model) as required in the integral above. This point is irrefutable. However, they also found the approximation involved to be quite accurate in the toy examples they considered. It is also beyond doubt that many (often Gaussian) approximations applied in statistical/econometric practice are not true or accurate in the finite samples they are applied to, but instead the sampling properties of such estimators/tests are (sometimes examined and found to be or often simply assumed to be) acceptable, in the results they deliver. I will provide yet more evidence that Congdon’s method fits into this category: practically it is a very accurate and powerful model selection tool, regardless of the mathematical appropriateness of its questionable assumptions.

Congdon’s method requires the following assumptions to be made:

(i) Given model \(M = k\), the parameters between the \(K\) models are a priori independent; see equation (2) in Congdon (2006) and Carlin and Chib (1995, pp 475).

(ii) The posterior distribution \(p(\theta_k|y, M = k)\) is independent of \(\theta_j\) where \(k \neq j\). i.e. given the data and \(M = k\), the parameters of model \(M = k\) are independent of the parameters of model \(M = j\). See also Scott (2002).

(iii) The prior distribution \(p(\theta_j|M = k)\) is uniform and proportional to 1. This is equation (3) in Congdon (2006) and is an interesting assumption. If the current model is \(M = k\), a priori this tells us nothing about the parameters of model \(M = j\).

Once again, these assumptions seem somewhat strange. However, it is not the purpose of this chapter to justify or suggest these assumptions are reasonable. Logically, perhaps
they are not, as again are many Gaussian, and other, assumptions applied in statistical and econometric practice. Instead, the purpose of this chapter is to highlight the finite sample performance of this approximate model selection tool in an extensive simulation study. The results delivered are remarkable and illustrate the robustness of this method to these apparently illogical assumptions and the approximation discussed above.

4.6 Simulation Study

The methods in Section 4.4 and 4.5 will now be examined in a simulation study. Four different models from the APVECH model family given in Section 4.3 will be considered, in a bivariate setting ($N = 2$). I proceed to estimate the parameters using simulated data, and use Congdon’s model selection procedure in Section 4.5 to evaluate model selection.

4.6.1 The Bivariate APVECH Model Family

When $N = 2$, the APVECH model family from Section 4.3 becomes:

\[
\begin{align*}
    y_{1t} &= m_{10} + m_{11}y_{1,t-1} + \varepsilon_{1t}, \\
    y_{2t} &= m_{20} + m_{22}y_{2,t-1} + \varepsilon_{2t}, \\
    h_{11,t} &= c_{11} + \gamma_{11}I_{1,t-1} + (a_{11} + \alpha_{11}I_{1,t-1})\varepsilon_{1,t-1}^2 + b_{11}h_{11,t-1} \\
                  &\quad + (d_{12} + \delta_{12}I_{2,t-1})\varepsilon_{2,t-1}^2 + c_{12}h_{22,t-1}, \\
    h_{22,t} &= c_{22} + \gamma_{22}I_{2,t-1} + (a_{22} + \alpha_{22}I_{2,t-1})\varepsilon_{2,t-1}^2 + b_{22}h_{22,t-1} \\
                  &\quad + (d_{21} + \delta_{21}I_{1,t-1})\varepsilon_{1,t-1}^2 + c_{21}h_{11,t-1}, \\
    h_{12,t} &= c_{12} + \gamma_{12,1}I_{1,t-1}I_{2,t-1} + \gamma_{12,2}I_{2,t-1}(1 - I_{2,t-1}) + \gamma_{12,3}(1 - I_{1,t-1})I_{2,t-1} \\
                  &\quad + [a_{12} + \alpha_{12,1}I_{1,t-1}I_{2,t-1} + \alpha_{12,2}I_{1,t-1}(1 - I_{2,t-1}) + \alpha_{12,3}(1 - I_{1,t-1})I_{2,t-1}]\varepsilon_{1,t-1}\varepsilon_{2,t-1} + b_{12}h_{12,t-1}.
\end{align*}
\]

The four models chosen from this family have a variety of asymmetric behaviour:
Define the following parameter vectors

\[ \theta^a_L = \{ \gamma_{11}, \alpha_{11}, \gamma_{22}, \alpha_{22} \} , \]
\[ \theta^a_F = \{ \delta_{12}, \delta_{21} \} , \]
\[ \theta^a_C = \{ \gamma_{12,1}, \gamma_{12,2}, \gamma_{12,3}, \alpha_{12,1}, \alpha_{12,2}, \alpha_{12,3} \} \]

which define the model parameters representing asymmetry, i.e local (L), foreign (F) and covariance (C) asymmetric parameters. Also let \( \mathbf{0}_i \) denote the \( i \)th order zero vector.

The following table outlines the four models under investigation in this simulation study:

<table>
<thead>
<tr>
<th>Model</th>
<th>Asymmetry Inclusions</th>
<th>Asymmetric Parameter Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Symmetric (PVECH)</td>
<td>( \theta^a_L = \mathbf{0}_4, \theta^a_F = \mathbf{0}_2, \theta^a_C = \mathbf{0}_6 )</td>
</tr>
<tr>
<td>2</td>
<td>Local Variance Asymmetries</td>
<td>( \theta^a_L \neq \mathbf{0}_4, \theta^a_F = \mathbf{0}_2, \theta^a_C = \mathbf{0}_6 )</td>
</tr>
<tr>
<td>3</td>
<td>All Variance Asymmetries</td>
<td>( \theta^a_L \neq \mathbf{0}_4, \theta^a_F \neq \mathbf{0}_2, \theta^a_C = \mathbf{0}_6 )</td>
</tr>
<tr>
<td>4</td>
<td>All Var and Cov Asymmetries</td>
<td>( \theta^a_L \neq \mathbf{0}_4, \theta^a_F \neq \mathbf{0}_2, \theta^a_C \neq \mathbf{0}_6 )</td>
</tr>
</tbody>
</table>

Hence a total of \( K = 4 \) models are available for selection, with each describing asymmetry in a slightly different way. Model 1 is a symmetric bivariate PVECH model as in Chapter 3 (but with bivariate standardised Student-\( t \) errors), while Model 4 describes the most general APVECH model in the family.

Equation (4.19) will now be examined in more detail for these models. First, denote:

\[ L_k^{[j]} = p(y|M = k, \theta^{[j]}_k) p(\theta^{[j]}_k|M = k) \Pr(M = k). \]  

(4.21)

\( L_k^{[j]} \) is the product of a model likelihood, from equation (4.9), a prior distribution and a prior model probability. The assumption is made that the prior model probabilities \( \Pr(M = k) \) are equal, i.e. \( \Pr(M = k) = \frac{1}{4}, k = 1, 2, 3, 4. \)

Under the approach in Section 4.4.2, the assumption is also made that the prior distribution for each model’s set of parameters is uniform over the parameter space restricted by (3.22) and (3.23), as in (4.10). This prior will usually be improper, since the parameter
space is potentially infinite. This prior improperness is not a problem during parameter estimation, for reasonable sample size, since the likelihood dominates inference. However (4.21) requires the evaluation of \( p(\theta^{[j]}_k|M = k) \); i.e. we must choose a proper density for each model.

To do this, we restrict the range of each parameter to create a finite prior space, while still maintaining a flat prior over the intersection of this restricted range and the space allowed by (3.22) and (3.23). Since the above models are nested, the choices remain consistent across the models. Parameter limits have been chosen well outside the ranges usually observed in empirical studies and that would be expected to give rise to PD and CS datasets. For sensitivity purposes, two specific choices have been made. The first is given as:

<table>
<thead>
<tr>
<th>Parameter Group</th>
<th>Restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_{yi} )</td>
<td>( \theta_{yi} \in (-2, 2) )</td>
</tr>
<tr>
<td>( \theta_{hi,t} )</td>
<td>( c_{ii}, \gamma_{ii} \in (-1, 1); \theta_{\neq c_{ii}, \gamma_{ii}} \in (-1.5, 1.5) )</td>
</tr>
<tr>
<td>( \theta_{hi,j,t} )</td>
<td>( c_{ij}, \gamma_{ij,k} \in (-0.5, 0.5); \theta_{\neq c_{ij}, \gamma_{ij,k}} \in (-1.5, 1.5) )</td>
</tr>
<tr>
<td>( \tau )</td>
<td>( \tau \in (0, 2) )</td>
</tr>
</tbody>
</table>

The region \( A \) is thus specified by the intersection of these ranges and (3.22) and (3.23) for each model. I note that throughout simulations and real data examples, I have never observed any simulated parameter iterates to be outside these ranges, when they were not enforced. These restrictions result in the following uniform prior distribution for the chosen models:

**Prior Set 1**

\[
p(\theta^{[j]}_1|M = 1) = \frac{1}{4^4 \times 3^{10} \times 2^3}, \quad p(\theta^{[j]}_2|M = 2) = \frac{p(\theta^{[j]}_1|M = 1)}{2^2 \times 3^2}, \quad p(\theta^{[j]}_3|M = 3) = \frac{p(\theta^{[j]}_2|M = 2)}{3^2}, \quad p(\theta^{[j]}_4|M = 4) = \frac{p(\theta^{[j]}_3|M = 3)}{3^3}.
\]
To demonstrate that model selection results are not significantly affected by small movements in prior distribution boundaries, another set of prior ranges is examined:

<table>
<thead>
<tr>
<th>Parameter Group</th>
<th>Restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{y_{it}}$</td>
<td>$\theta_{y_{it}} \in (-1.5, 1.5)$</td>
</tr>
<tr>
<td>$\theta_{h_{iti,t}}$, $c_{ii} \in (-1.5, 1.5)$; $\gamma_{ii} \in (-1, 1)$; $\theta \neq c_{ii}, \gamma_{ii} \in (-1.25, 1.25)$</td>
<td></td>
</tr>
<tr>
<td>$\theta_{h_{ij,t}}$, $c_{ij}, \gamma_{ij,k} \in (-0.75, 0.75)$; $\theta \neq c_{ij}, \gamma_{ij,k} \in (-1.5, 1.5)$</td>
<td></td>
</tr>
<tr>
<td>$\tau \in (0, 1)$</td>
<td></td>
</tr>
</tbody>
</table>

Comparing these priors to the original set, the widths of all boundaries have either slightly increased, decreased or remained the same. The prior distribution now becomes:

**Prior Set 2**

\[
p(\theta^j_1 | M = 1) = \frac{1}{3^8 \times 1.5 \times 2.5^8}, \quad p(\theta^j_2 | M = 2) = \frac{p(\theta^j_1 | M = 1)}{2^2 \times 2.5^2}, \quad p(\theta^j_3 | M = 3) = \frac{p(\theta^j_2 | M = 2)}{2.5^2}, \quad p(\theta^j_4 | M = 4) = \frac{p(\theta^j_3 | M = 3)}{3^3 \times 1.5^3}.
\]

We see that simpler models are a priori favoured over more complex models for both sets of prior ranges. Since the ranges are quite wide and MCMC iterates are never close to either set of boundaries (in my experience), resulting parameter estimates and inferences will be unaffected by these choices. Calculations in equations (4.19) and (4.20) can now be performed. Both prior sets will be applied in the simulation and empirical studies in the following sections.

### 4.6.2 The Results

Data is simulated with 100 replications, over two separate sample sizes: $n = 1,000$ and $n = 3,000$ bivariate observations. This simulation study is repeated for each of the four possible models. For each of the resulting 400 datasets, all four models considered are estimated to allow comparison across models and model selection to be conducted.
For each dataset, the MCMC sampler is run for $J = 30,000$ iterations, including a burn-in period of $w = 10,000$ iterations. Prior to the results appearing in this chapter, convergence of the proposed sampler was checked extensively for some of the simulated data sets by running the chain multiple times with different starting points and then observing MCMC trace plots, noting convergence to the same stationary distribution (within acceptable Monte Carlo sampling error). These experiments revealed that convergence was always clearly achieved well before 10,000 iterations were reached.

Tables 4.1-4.8 (located in the Appendix of this chapter) contain the parameter estimation results from the simulation study for $n = 1,000$ and $n = 3,000$. These contain the mean of the 100 posterior means for each model parameter, plus the means of the estimated 95% credible intervals (CIs) from each replication (presented as: Estimate[Lower,Upper]). Each CI is simply the 2.5th and 97.5th percentiles of each parameter’s MCMC chain, post burn-in. The first two columns identify the parameter (Par) and its true value (TV) from the model used to simulate the data and whose results are displayed in blue text. The remaining columns contain results from the other models fitted to the data, purely for comparative purposes. Any 95% CI not containing the corresponding TV in these tables is displayed in red text. Each MCMC replication run for the most complex model (Model 4) at the highest sample size ($n = 3,000$) takes on average 6 min 45 sec to complete on the machine and software described in Section 2.6.

Since the chosen priors are flat, the results in blue in these tables should be (approximately) unbiased: this is indeed the case since all estimates lie close to their TVs, while their corresponding CIs all contain the TV. This includes the parameters whose TVs are negative (e.g. $c_{22}$ and $d_{21}$) which, under usual parameter restrictions, might be assumed positive. It is also expected that where additional parameters with TV zero in more complex models are estimated, these should be close to zero. This can be seen in these tables also (e.g. in Table 4.1 and 4.5, the parameter estimates for $\gamma_{11}$, $\alpha_{11}$ and $\delta_{12}$ in Model 3 for both sample sizes are close to zero, which is also contained in their CIs). Only two parameters do not satisfy this trait and only when data is simulated from Model 1 (Tables
4.1 and 4.5), i.e. parameters $\alpha_{22}$ and $\alpha_{12,1}$ from Model 4. This correctly indicates that a simpler model is probably sufficient in these datasets. Alternately, Tables 4.2-4.4 and 4.6-4.8 indicate some bias and precision problems when simpler models are fit to data simulated from more complex versions (as given by the results in red text); e.g. in Table 4.8 seven of the eighteen parameters in Model 1 are biased and the TV is not included in their CI. This is especially clear for the parameters $a_{11}$ and $a_{22}$ whose estimates are $0.448[0.388, 0.509]$ and $0.304[0.261, 0.349]$ (respectively), which are nowhere near their corresponding TVs of 0.3 and 0.2. The conclusion here is that usually over-fitting in this model family with the proposed estimators will not cause bias, but will simply reduce precision, as expected. However, when not accounting for asymmetries present in the data, an under-fit model will have biased parameter estimates.

Tables 4.1-4.8 also demonstrate how parameter estimation properties improve with an increased sample size: the parameter estimates generally lie closer to their TVs and are contained within a tighter CI. However, the biases and errors when fitting models that are too simple did not improve at all with sample size. Clearly, effects like asymmetry need to be identified and modelled appropriately so that no biases or large errors in parameter estimation occur.

Model selection results from this study are contained in Tables 4.9 and 4.10 in the Appendix. Shown are the mean of each posterior model probability $\hat{Pr}(M = k|y)$, $k = 1, 2, 3, 4$ over the 100 replications, and the number of times each model was selected out of 100 as having the highest posterior probability. The first column indicates the true model and the second denotes the prior used, while the top row indicates which model was fit to the datasets. The model having the highest (mean) posterior probability and selected the most times out of the 100 datasets has been displayed in blue text in these tables. Running Congdon’s model selection procedure to obtain the mean posterior model probabilities and selection numbers for this simulation study takes approximately 14 sec. The results of this model selection procedure are quite interesting. Firstly, if Model 1 or 2 was the true model, then the same corresponding model was selected the
most times out of 100 datasets and had the highest average posterior probability, at
both sample sizes and over both priors. The results are not particularly sensitive to the
two priors considered, which is consistent across all entries in both tables. However, for
\( n = 1,000 \), when Model 3 or Model 4 was the true model, Model 2 had the highest
probabilities and selection numbers. This can be mainly attributed to the fact that the
majority of additional parameters estimated for Models 3 and 4 were insignificant, i.e. the
corresponding CIs contained zero (see Tables 4.3-4.4). This means that the procedure
has seen minimal difference between Models 2-4 in this case and thus has chosen the
most simple model. These results may have been different had the true parameter values
been chosen differently in this simulation study. Generally, \( n = 1,000 \) is a very small
sample size for even a univariate GARCH model to be applied to, but the results are still
couraging at this level.

The model selection results have improved for the far more reasonable, though still
practically small, sample size of \( n = 3,000 \) (see Table 4.10). When the true model
was Models 1, 2 and 4, the same corresponding model was selected between 94 and 99
times for each of the models, while the off-diagonal entries only exhibited small numbers,
ranging between 0 and 6 times selected. Prior sensitivity has also reduced at this sample
size. As in the \( n = 1,000 \) case, Model 2 was chosen when Model 3 was the true model,
although the probabilities and selection numbers have improved somewhat. Once again,
the additional parameters estimated in Model 3 (i.e. \( \delta_{12} \) and \( \delta_{21} \)) were insignificant when
\( n = 3,000 \) (see Table 4.3), but the probabilities and selection numbers improved since
the estimates are more precise (i.e. smaller CI width). Finally I note that, impressively,
the simplest model (Model 1), is correctly chosen 99 times under this measure. Clearly
the posterior model probability (approximation) can account and correct for model size
and complexity very effectively. This may be due to the fact that under the uniform prior
assumption, each \( p(\theta|\theta^0|M = k) \) decreases with increasing parameter numbers, which in
turn will decrease the value of the conditional probability in equation (4.19) for more
complex models (assuming the other components of this equation are fixed). Hence these
priors could be viewed as a “penalty” function, similar to that of the information criterion functions in classical methods. Despite Congdon’s method being an approximation to the true MC estimate of posterior model probability, in this model family its performance seems highly and completely satisfactory, suggesting that it is a very good approximation here, or regardless, at the least a highly effective model selection tool in its own right.

4.7 Empirical Study

The estimation and model selection procedures will now be applied to real financial stock return data. These consist of four international stock market indices: S&P500 (US), FTSE100 (UK), All Ordinaries (AU) and Nikkei 225 (JP). Bivariate vectors of percentage daily log returns are formed via:

\[ y_t = 100(\ln p_t - \ln p_{t-1}), \]

where \( p_t \) is the bivariate daily closing value at time \( t \).

The following table displays each bivariate return series considered, its time span and sample size (\( n \)):

<table>
<thead>
<tr>
<th>Bivariate Returns</th>
<th>Duration</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P500/FTSE100</td>
<td>Apr 3, 84 - Aug 4, 06</td>
<td>5,535</td>
</tr>
<tr>
<td>S&amp;P500/All Ordinaries</td>
<td>Aug 6, 84 - Aug 4, 06</td>
<td>5,434</td>
</tr>
<tr>
<td>S&amp;P500/Nikkei 225</td>
<td>Jan 5, 84 - Aug 4, 06</td>
<td>5,397</td>
</tr>
<tr>
<td>FTSE100/All Ordinaries</td>
<td>Aug 6, 84 - Aug 4, 06</td>
<td>5,492</td>
</tr>
<tr>
<td>FTSE100/Nikkei 225</td>
<td>Apr 3, 84 - Aug 4, 06</td>
<td>5,357</td>
</tr>
<tr>
<td>All Ordinaries/Nikkei 225</td>
<td>Aug 6, 84 - Aug 4, 06</td>
<td>5,267</td>
</tr>
</tbody>
</table>

Note that in a bivariate setting, ordering the two markets from largest to smallest is irrelevant because all possible covariances are captured in the model in this dimension. Ordering of the markets needs to be done only for \( N > 2 \).
4.7.1 Estimation Results

Table 4.11 in the Appendix displays the model selection results for all six market pairings across the two priors. The typical times to run each model are up to approximately 11 min 45 sec - this is for the most complex model with the largest sample size (i.e. Model 4 fitted to the US/UK returns with $n = 5,535$). Again the results are not too sensitive to these choices. The probabilities in the blue text are the highest across the models and indicate the model selected in each case. Model 4 was clearly favoured for all market pairings, except for the US/AU pair which favoured Model 3. These results tell us that variance and covariance asymmetry occurs and is strongly favoured (over the symmetric Model 1) in all market pairings, except for the US/AU pair which favours only variance asymmetry.

Tables 4.12-4.23 in the Appendix contain the parameter estimation results for each bivariate series. A number of observations can be made from these tables. Firstly, the results show the usual high persistence in volatility for stock returns, since $a_{ii} + b_{ii}$ has a high value ($> 0.75$) in each market pair, while persistence in covariance is still strong ($a_{12} + b_{12} > 0.85$) for five of the pairs (US/UK, US/JP, UK/AU, UK/JP and AU/JP), but slightly weaker for the US/AU pair.

Volatility asymmetry is also quite evident across markets. Volatility intercepts displayed this trait slightly, with some $\gamma_{ii}$ significant from zero ($\gamma_{ii} \approx 0.003 - 0.046$) following negative local returns. Volatility increased significantly in all markets in response to that market falling the day before ($\alpha_{ii} > 0$). Volatility spill-over effects ($e_{12}, e_{21}$) were mostly negative, many of these significantly; while return squared spill-over effects ($d_{12}, d_{21}$) were mostly positive and some were significantly asymmetric ($\delta_{12}, \delta_{21}$): the other market dropping unexpectedly in price had a significantly larger and positive effect ($\delta_{12}, \delta_{21} > 0$) in volatility in the local market.

Asymmetry in the covariances was also evident. Covariance intercepts only showed this effect in the UK/JP series, with $\gamma_{12,3}$ significantly different from zero. The $\alpha_{12,1}$ parameter measures the change in the ARCH covariance effect when both markets fell
on the previous day, and was significantly positive in four of the six market pairs. This indicates an increase in conditional covariance (correlation) following unexpected drops in both markets. In addition, the $\alpha_{12,2}$ and $\alpha_{12,3}$ parameters which measure the change in the ARCH covariance effect when one market rose and the other market fell are both significantly negative for the US/UK and US/JP market pairs, ranging between $-0.04$ and $-0.07$. This indicates a reduced ARCH covariance effect in these market pairs when each market moved in opposite directions. These significant asymmetric covariance parameters also support the model selection results in Table 4.11 where Model 4 had the highest posterior model probability in most market pairs.

The estimates for the degrees of freedom $\nu$ demonstrate the presence of excess kurtosis in these stock returns - all ranged from 7.21 to 11.08, showing that normality of the residuals should not be assumed. It is also worth noting the negative estimates of some parameters that are usually restricted to be positive in the literature. For example, the estimates of $e_{12}$ are significantly negative in four of six market pairs. This trait was also observed in the empirical study of Chapter 3.

Acceptance rates (ARs) for each of the 5 steps in the sampling scheme are also presented in the Appendix, below the tables of corresponding parameter estimates. These ARs for the sampling period only. The ARs do not differ too much between models within each dataset - this is expected since the models are from the same family. All ARs across all models and datasets range between 14% and 99%, indicating that a variety of posterior distribution shapes are present for each parameter.

To demonstrate efficiency and convergence of the sampling scheme, Table 4.24 displays potential scale reduction factors (PSRFs) for Model 4 fitted to the AU/JP bivariate returns. As in Chapter 3, each PSRF is calculated using $R = 5$ replications of MCMC iterates and 5,000 iterations after the burn-in period. MCMC iterates from the parameters $\gamma_{11}$, $b_{22}$, $\alpha_{12,3}$ and $\nu$ are displayed in Figure 4.2. From Table 4.24, it can be seen that all PSRFs are close to 1, with only one parameter having their PSRF greater than 1.2 ($1.23$ for $\alpha_{22}$). In addition, Figure 4.2 shows different speeds of convergence for the
parameters displayed; parameters $b_{22}$ and $\nu$ have almost instantaneous convergence, while $\gamma_{11}$ and $\alpha_{12,3}$ have slower convergence, but iterations after burn-in still have a relatively small variance between replications. Hence the sampling scheme is quite efficient, with convergence achieved as demonstrated by these results.

Figure 4.1 displays estimated conditional correlation plots for each pair of bivariate return datasets. These correlations come from the model whose posterior model probability was the highest (see Table 4.11), and are posterior means of the correlation series across all sampling period MCMC iterates - these were calculated by first forming a correlation MCMC iterate via:

$$\hat{\rho}_t = \frac{h_{12,t}^{[j]}}{\sqrt{h_{11,t}^{[j]}h_{22,t}^{[j]}}}, \quad t = 1, \ldots, n$$

where each $h_{i,j,t}^{[j]}$ is the calculated (co)variance based on the chosen model and the $j$th MCMC iterate parameter values. The estimated conditional correlations were then calculated as

$$\hat{\rho}_t = \frac{1}{J-w} \sum_{j=w+1}^{J} \hat{\rho}_t^{[j]}, \quad t = 1, \ldots, n.$$ 

The following information can be gathered from these plots: Firstly, it can be seen that correlations are mostly significantly positive (i.e. $\hat{\rho}_t > 0$ for most $t$) across markets, as expected for international financial markets. While negative correlations were obtained in some instances (the lowest being around $-0.05$), the majority are between 0 and 0.82. Secondly, these correlations are not constant - they all seem to fluctuate in a “wave-like” persistent pattern over time. This property is also confirmed by significant parameter estimates for the $h_{12,t}$ equation in Tables 4.12-4.22.
Figure 4.1: Plots of estimated conditional correlations $\hat{\rho}_t$ for each bivariate combination of S&P 500 (US), FTSE 100 (UK), All Ordinaries (AU) and Nikkei 225 (JP) returns between 4 January 1984 and 4 August 2006.
4.8 Chapter Summary

A family of parsimonious multivariate GARCH models was proposed that allows for excess kurtosis and asymmetry properties generally found in stock return data. Model parameters were estimated via a Bayesian MCMC approach that implicitly enforces necessary and sufficient conditions for PD and CS via a computational algorithm and implicit prior distribution. The Bayesian approximate model selection approach by Congdon (2006) was applied to decide between four models that had varying asymmetric effects. A simulation study showed that the model selection procedure selected the two simpler models out of the four considered when data was simulated a high proportion of times, the reasons for this being a small sample size \( n = 1,000 \) resulting in insignificant parameter estimates for the more complex models. These results improved for the larger sample size \( n = 3,000 \) but still had the ability to select a simpler model when required. An empirical study of four international markets confirmed the presence of excess kurtosis in stock returns as well as volatility and covariance asymmetry, including asymmetry from interactions between pairs of markets. Finally, estimated conditional correlations illustrated their time-varying and asymmetric nature, for the stock returns chosen.

Asymmetry is extended further in the next chapter to be incorporated into mean, variance and correlation components of a proposed copula-GARCH model. An application in risk forecasting for stock portfolios is also presented.
4.9 Appendix

4.9.1 Derivation of $E(I_{it}I_{jt})$ and $E(I_{it}I_{jt}\varepsilon_{it}\varepsilon_{jt})$

In this sub-section, I show how to derive the expectations $E(I_{it}I_{jt})$ and $E(I_{it}I_{jt}\varepsilon_{it}\varepsilon_{jt})$ to a point where non-linear terms develop in the equation for the unconditional covariance of the APVECH model, i.e.

$$E(h_{ij,t}) = c_{ij} + \gamma_{ij,1}E(I_{it}I_{jt}) + \gamma_{ij,2}E[I_{it}(1 - I_{jt})] + \gamma_{ij,3}E[(1 - I_{it})I_{jt}]$$

$$+ \alpha_{ij}E(I_{it}I_{jt}\varepsilon_{it}\varepsilon_{jt}) + \alpha_{ij,2}E[I_{it}(1 - I_{jt})\varepsilon_{it}\varepsilon_{jt}]$$

$$+ \alpha_{ij,3}E[(1 - I_{it})I_{jt}\varepsilon_{it}\varepsilon_{jt}] + b_{ij}E(h_{ij,t}).$$

As mentioned in Section 4.3, the value of each $I_{it}$ and $I_{jt}$ determine which quadrant the vector $(\varepsilon_{it}, \varepsilon_{jt})$ lies, hence the approach of calculating expectations involving $(1 - I_{it})$ and $(1 - I_{jt})$ only differs via integration limits.

Deriving $E(I_{it}I_{jt})$

Using the law of total probability, we have

$$E(I_{it}I_{jt}) = \sum_{r=0}^{1} \sum_{s=0}^{1} E(I_{it}I_{jt}|I_{it} = r, I_{jt} = s) Pr(I_{it} = r, I_{jt} = s)$$

$$= E(I_{it}I_{jt}|I_{it} = 1, I_{jt} = 1) Pr(I_{it} = 1, I_{jt} = 1) \text{ (all other terms zero)}$$

$$= Pr(\varepsilon_{it} < 0, \varepsilon_{jt} < 0) = Pr(z_{it} < 0, z_{jt} < 0)$$

where $z_{it} = \varepsilon_{it}/\sqrt{h_{it}}$, i.e. the standardised residual. Using the law of total expectation, we can write $E(I_{it}I_{jt}) = E[E(I_{it}I_{jt}|\psi_{t-1})]$, which gives

$$E[E(I_{it}I_{jt}|\psi_{t-1})] = E[Pr(z_{it} < 0, z_{jt} < 0|\psi_{t-1})]$$

$$= E \left[ \int_{-\infty}^{0} \int_{-\infty}^{0} p(z_{it}, z_{jt}|\psi_{t-1}) dz_{it} dz_{jt} \right]$$
where \( p(z_{it}, z_{jt} | \psi_{t-1}) \) denotes the joint PDF of \( z_{it} \) and \( z_{jt} \). Under equation (4.7), this is in fact a bivariate standardised Student-\( t \) distribution with zero mean vector, unit variances and correlation \( \rho_{ij,t} = h_{ij,t} / \sqrt{h_{ii,t} h_{jj,t}} \). Let \( I_1 \) denote the integral above. We then have

\[
I_1 = \frac{\nu}{2\pi(\nu - 2)\sqrt{1 - \rho_{ij,t}^2}} \int_{-\infty}^{0} \int_{-\infty}^{0} \left[ 1 + \frac{z_{it}^2 - 2\rho_{ij,t} z_{it} z_{jt} + z_{jt}^2}{(\nu - 2)(1 - \rho_{ij,t}^2)} \right]^{-\frac{\nu+2}{2}} dz_{it} dz_{jt}
\]

(4.22)

where \( T_2(z | \mu, P_t, \nu) \) denotes the bivariate standardised Student-\( t \) CDF at point \( z \) with mean vector \( \mu \), correlation matrix \( P_t = [\rho_{ij,t}] \) and degrees of freedom \( \nu \). In summary, we have

\[
E(I_{it} I_{jt}) = E[T_2(0|0, P_t, \nu)].
\]

(4.23)

**Deriving \( E(I_{it} I_{jt} \varepsilon_{it} \varepsilon_{jt}) \)**

Using the law of total expectation, we have

\[
E(I_{it} I_{jt} \varepsilon_{it} \varepsilon_{jt}) = E[E(I_{it} I_{jt} \varepsilon_{it} \varepsilon_{jt} | I_{it} = 1, I_{jt} = 1, \psi_{t-1})]
\]

\[
= E[E(\varepsilon_{it} \varepsilon_{jt} | \varepsilon_{it} < 0, \varepsilon_{jt} < 0, \psi_{t-1})]
\]

\[
= E[E(\sqrt{h_{ii,t} h_{jj,t}} z_{it} z_{jt} | z_{it} < 0, z_{jt} < 0, \psi_{t-1})]
\]

\[
= E[E(\sqrt{h_{ii,t} h_{jj,t}} | z_{it} < 0, z_{jt} < 0, \psi_{t-1}) E(z_{it} z_{jt} | z_{it} < 0, z_{jt} < 0, \psi_{t-1})]
\]

\[
= E(\sqrt{h_{ii,t} h_{jj,t}}) E[E(z_{it} z_{jt} | z_{it} < 0, z_{jt} < 0, \psi_{t-1})]
\]

where the last two steps were taken due to each \( z_{it} \) and \( z_{jt} \) being independent of each conditional standard deviation \( \sqrt{h_{ii,t}} \) and \( \sqrt{h_{jj,t}} \). We then have

\[
E[E(z_{it} z_{jt} | z_{it} < 0, z_{jt} < 0, \psi_{t-1})] = E \left[ \int_{-\infty}^{0} \int_{-\infty}^{0} z_{it} z_{jt} p(z_{it}, z_{jt} | \psi_{t-1}) dz_{it} dz_{jt} \right]
\]

(4.24)

where \( p(z_{it}, z_{jt} | \psi_{t-1}) \) is the bivariate standardised Student-\( t \) PDF, as described above.
While the integrals in equations (4.23) and (4.24) can be evaluated numerically (e.g. see Genz and Bretz, 2002 for a comparison of methods for the unstandardised multivariate Student-$t$ distribution), to my knowledge there is no exact representation of these integrals in the literature in terms of conditional correlations and/or variances for this distribution. To further understand the complexity of this problem, assume now that $z_{it}, z_{jt} | \psi_{t-1}$ follows a bivariate normal distribution with zero mean vector, unit variances and correlation matrix $P_t = [\rho_{ij,t}]$. Then the integrals in equations (4.22) and (4.24) become

$$I_1 = \frac{1}{2\pi \sqrt{1 - \rho_{ij,t}^2}} \int_{-\infty}^{0} \int_{-\infty}^{0} \exp \left[ - \frac{z_{it}^2 - 2\rho_{ij,t} z_{it} z_{jt} + z_{jt}^2}{2(1 - \rho_{ij,t}^2)} \right] \, dz_{it}dz_{jt}, \quad (4.25)$$

$$I_2 = \frac{1}{2\pi \sqrt{1 - \rho_{ij,t}^2}} \int_{-\infty}^{0} \int_{-\infty}^{0} z_{it}z_{jt} \exp \left[ - \frac{z_{it}^2 - 2\rho_{ij,t} z_{it} z_{jt} + z_{jt}^2}{2(1 - \rho_{ij,t}^2)} \right] \, dz_{it}dz_{jt}. \quad (4.26)$$

Under this multivariate normal assumption, an exact solution of the integrals in (4.25) and (4.26) actually exists. Kamat (1953) presents what are called “incomplete moments” of the multivariate normal distribution and gives results for special cases. These incomplete moments are defined as

$$[m_1, m_2, m_3, \ldots] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} x_1^{m_1} x_2^{m_2} x_3^{m_3} \ldots p(x) \, dx$$

where $p(x)$ is the standardised multivariate normal PDF at the vector $x = (x_1, x_2, x_3, \ldots)$. In a bivariate space, integrals taken in diagonally opposite quadrants will be the same due to the symmetry of the distribution, hence

$$[m_1, m_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^{m_1} x_2^{m_2} p(x_1, x_2) \, dx_1 dx_2 = \int_{-\infty}^{0} \int_{-\infty}^{0} x_1^{m_1} x_2^{m_2} p(x_1, x_2) \, dx_1 dx_2.$$

Now the integrals $I_1$ and $I_2$ are equivalent to the incomplete moments $[0, 0]$ and $[1, 1]$ (respectively), and using the exact expressions of Kamat (1953), we obtain

$$I_1 = \frac{1}{2\pi} \left( \frac{\pi}{2} + \sin^{-1} \rho_{ij,t} \right), \quad I_2 = \frac{1}{2\pi} \left[ \rho_{ij,t} \left( \frac{\pi}{2} + \sin^{-1} \rho_{ij,t} \right) + \sqrt{1 - \rho_{ij,t}^2} \right].$$
and so the required expectations are given by

\[
E(I_{it}I_{jt}) = E \left[ \frac{1}{2\pi} \left( \frac{\pi}{2} + \sin^{-1} \rho_{ij,t} \right) \right] = \frac{1}{4} + \frac{1}{2\pi} E(\sin^{-1} \rho_{ij,t}), \tag{4.27}
\]

\[
E(I_{it}I_{jt} \varepsilon_{it} \varepsilon_{jt}) = E(\sqrt{h_{ii,t}h_{jj,t}}) E \left\{ \frac{1}{2\pi} \left[ \rho_{ij,t} \left( \frac{\pi}{2} + \sin^{-1} \rho_{ij,t} \right) + \sqrt{1 - \rho_{ij,t}^2} \right] \right\}
\]

\[
= E(\sqrt{h_{ii,t}h_{jj,t}}) \times \left[ \frac{1}{4} E(\rho_{ij,t}) + \frac{1}{2\pi} E \left( \rho_{ij,t} \sin^{-1} \rho_{ij,t} \right) + \frac{1}{2\pi} E \left( \sqrt{1 - \rho_{ij,t}^2} \right) \right]. \tag{4.28}
\]

Several observations can be made from these results:

- In order to obtain an exact expression for the unconditional covariance of the APVECH model, the expectations in (4.27) and (4.28) would require to be linear in terms of $h_{ij,t}$. But since $\rho_{ij,t} = h_{ij,t}/\sqrt{h_{ii,t}h_{jj,t}}$, we obtain non-linear terms in the expectation and hence an exact unconditional covariance is not feasible.

- To approximate the unconditional covariance, highly sophisticated techniques would need to be applied, due to the variety of non-linear forms appearing in equations (4.27) and (4.28). For example, approximating the expectation $E(\sqrt{h_{ii,t}h_{jj,t}})$ using a first-order Taylor polynomial (i.e. Delta method) implies that $E(\sqrt{h_{ii,t}h_{jj,t}}) \approx \sqrt{E(h_{ii,t}h_{jj,t})}$. This is in fact a poor approximation because for any random variable $X$, $E(\sqrt{X}) < \sqrt{E(X)}$, as proven by Murthy and Pillai (1966). Hence higher-order Taylor polynomials and/or other approximation techniques would need to be applied, which could potentially increase the computational expense when checking this property in model estimation.

- The expectations in (4.27) and (4.28) become even more complex when the distribution is switched back to a bivariate standardised Student-$t$ distribution, since the degrees of freedom parameter $\nu$ now needs to be considered, even though no exact expression of incomplete moments of the multivariate Student-$t$ distribution has been derived as yet in the literature (to my knowledge).
Based on the above information, I apply a simple assumption to the unconditional covariance, similar to that in De Goeij and Marquering (2004), which does not appear to affect model performance in the simulation and empirical studies of Chapter 4.

### 4.9.2 Simulation and Empirical Study Tables

**Table 4.1:** Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 1 with \( n = 1,000 \)

<table>
<thead>
<tr>
<th>Par</th>
<th>TV</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{10} )</td>
<td>0.06</td>
<td>0.060[0.034,0.087]</td>
<td>0.060[0.034,0.087]</td>
<td>0.061[0.034,0.088]</td>
<td>0.059[0.032,0.086]</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.04</td>
<td>0.041[0.028,0.055]</td>
<td>0.041[0.028,0.054]</td>
<td>0.040[0.027,0.054]</td>
<td>0.040[0.026,0.053]</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.03</td>
<td>0.027[-0.033,0.087]</td>
<td>0.027[-0.033,0.087]</td>
<td>0.028[-0.032,0.087]</td>
<td>0.028[-0.031,0.088]</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>-0.1</td>
<td>-0.103[-0.158,-0.047]</td>
<td>-0.102[-0.157,-0.046]</td>
<td>-0.101[-0.158,-0.044]</td>
<td>-0.101[-0.157,-0.045]</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.1</td>
<td>0.103[0.070,0.141]</td>
<td>0.106[0.068,0.148]</td>
<td>0.107[0.068,0.149]</td>
<td>0.117[0.075,0.165]</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.3</td>
<td>0.286[0.214,0.366]</td>
<td>0.312[0.219,0.417]</td>
<td>0.318[0.224,0.421]</td>
<td>0.308[0.210,0.414]</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.4</td>
<td>0.363[0.215,0.505]</td>
<td>0.340[0.197,0.479]</td>
<td>0.342[0.203,0.479]</td>
<td>0.323[0.182,0.462]</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.09</td>
<td>0.118[0.010,0.241]</td>
<td>0.119[0.003,0.248]</td>
<td>0.169[0.049,0.298]</td>
<td>0.139[0.067,0.284]</td>
</tr>
<tr>
<td>( e_{12} )</td>
<td>0.074[-0.084,0.243]</td>
<td>0.103[-0.048,0.260]</td>
<td>0.086[-0.042,0.217]</td>
<td>0.078[-0.059,0.219]</td>
<td></td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>0.001[-0.035,0.037]</td>
<td>0.001[-0.035,0.037]</td>
<td>-0.003[-0.045,0.039]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{11} )</td>
<td>-0.002[-0.109,0.104]</td>
<td>-0.014[-0.117,0.090]</td>
<td>0.032[-0.085,0.150]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta_{12} )</td>
<td>-0.026[-0.177,0.131]</td>
<td>0.042[0.122,0.212]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>-0.01</td>
<td>-0.011[-0.021,-0.001]</td>
<td>-0.009[-0.019,0.002]</td>
<td>-0.007[-0.018,0.004]</td>
<td>-0.005[-0.016,0.006]</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.2</td>
<td>0.204[0.144,0.270]</td>
<td>0.230[0.146,0.326]</td>
<td>0.216[0.141,0.300]</td>
<td>0.199[0.122,0.287]</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.6</td>
<td>0.529[0.406,0.636]</td>
<td>0.511[0.379,0.626]</td>
<td>0.485[0.352,0.601]</td>
<td>0.501[0.373,0.615]</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>-0.006[-0.013,0.004]</td>
<td>-0.005[-0.013,0.005]</td>
<td>-0.003[-0.013,0.012]</td>
<td>-0.003[-0.014,0.011]</td>
</tr>
<tr>
<td>( e_{21} )</td>
<td>0.1</td>
<td>0.119[0.078,0.165]</td>
<td>0.110[0.070,0.156]</td>
<td>0.114[0.079,0.151]</td>
<td>0.102[0.067,0.140]</td>
</tr>
<tr>
<td>( \gamma_{22} )</td>
<td>0.001[-0.008,0.010]</td>
<td>-0.001[-0.011,0.009]</td>
<td>-0.002[-0.012,0.009]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{22} )</td>
<td>-0.006[-0.101,0.089]</td>
<td>0.039[-0.024,0.105]</td>
<td>0.071[0.007,0.137]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta_{31} )</td>
<td>0.001[-0.016,0.019]</td>
<td>0.003[-0.015,0.023]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.01</td>
<td>0.011[0.005,0.017]</td>
<td>0.011[0.006,0.017]</td>
<td>0.011[0.006,0.018]</td>
<td>0.016[0.002,0.029]</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.3</td>
<td>0.287[0.231,0.342]</td>
<td>0.303[0.240,0.365]</td>
<td>0.313[0.247,0.378]</td>
<td>0.259[0.179,0.337]</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.5</td>
<td>0.462[0.331,0.574]</td>
<td>0.453[0.326,0.565]</td>
<td>0.446[0.318,0.559]</td>
<td>0.420[0.287,0.540]</td>
</tr>
<tr>
<td>( \gamma_{12,1} )</td>
<td>-0.004[-0.025,0.017]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma_{12,2} )</td>
<td>-0.004[-0.026,0.018]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma_{12,3} )</td>
<td>-0.005[-0.028,0.017]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12,1} )</td>
<td>0.083[0.003,0.164]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12,2} )</td>
<td>0.034[-0.059,0.125]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12,3} )</td>
<td>0.037[-0.053,0.126]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 2 with \( n = 1,000 \)

<table>
<thead>
<tr>
<th>( \text{Par} )</th>
<th>( \text{TV} )</th>
<th>( \text{Model 1} )</th>
<th>( \text{Model 2} )</th>
<th>( \text{Model 3} )</th>
<th>( \text{Model 4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{10} )</td>
<td>0.06</td>
<td>0.086[0.045,0.128]</td>
<td>0.062[0.023,0.101]</td>
<td>0.062[0.023,0.101]</td>
<td>0.060[0.021,0.099]</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.04</td>
<td>0.054[0.020,0.089]</td>
<td>0.039[0.005,0.072]</td>
<td>0.039[0.005,0.072]</td>
<td>0.038[0.004,0.071]</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.03</td>
<td>0.021[-0.043,0.085]</td>
<td>0.023[-0.037,0.083]</td>
<td>0.024[-0.036,0.084]</td>
<td>0.025[-0.034,0.084]</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>-0.1</td>
<td>-0.101[-0.163,-0.038]</td>
<td>-0.099[-0.159,-0.039]</td>
<td>-0.099[-0.158,-0.040]</td>
<td>-0.099[-0.157,-0.039]</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.1</td>
<td>0.165[0.110,0.227]</td>
<td>0.115[0.055,0.180]</td>
<td>0.116[0.056,0.181]</td>
<td>0.124[0.062,0.192]</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.3</td>
<td>0.448[0.345,0.555]</td>
<td>0.313[0.215,0.420]</td>
<td>0.308[0.212,0.413]</td>
<td>0.305[0.205,0.414]</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.4</td>
<td>0.350[0.220,0.476]</td>
<td>0.356[0.238,0.470]</td>
<td>0.355[0.235,0.467]</td>
<td>0.346[0.229,0.460]</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.09</td>
<td>0.091[0.027,0.162]</td>
<td>0.092[0.031,0.159]</td>
<td>0.103[0.029,0.187]</td>
<td>0.101[0.024,0.187]</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0</td>
<td>0.037[-0.051,0.139]</td>
<td>0.031[-0.049,0.123]</td>
<td>0.030[-0.050,0.123]</td>
<td>0.030[-0.051,0.124]</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>0.1</td>
<td>0.110[0.011,0.211]</td>
<td>0.108[0.009,0.209]</td>
<td>0.109[0.006,0.214]</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{11} )</td>
<td>0.3</td>
<td>0.271[0.096,0.452]</td>
<td>0.275[0.102,0.453]</td>
<td>0.294[0.107,0.484]</td>
<td></td>
</tr>
<tr>
<td>( \delta_{12} )</td>
<td>-0.002[-0.110,0.108]</td>
<td>-0.002[-0.110,0.108]</td>
<td>-0.002[-0.110,0.108]</td>
<td>-0.002[-0.110,0.108]</td>
<td></td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>-0.01</td>
<td>0.023[-0.009,0.057]</td>
<td>0.000[-0.032,0.036]</td>
<td>0.000[-0.032,0.037]</td>
<td>0.004[-0.030,0.042]</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.2</td>
<td>0.300[0.225,0.387]</td>
<td>0.220[0.142,0.306]</td>
<td>0.216[0.140,0.298]</td>
<td>0.217[0.137,0.304]</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.6</td>
<td>0.559[0.454,0.653]</td>
<td>0.554[0.456,0.643]</td>
<td>0.554[0.456,0.643]</td>
<td>0.549[0.450,0.640]</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>0.066[-0.017,0.037]</td>
<td>0.004[-0.018,0.032]</td>
<td>0.000[-0.020,0.046]</td>
<td>0.000[-0.021,0.044]</td>
</tr>
<tr>
<td>( c_{21} )</td>
<td>0.1</td>
<td>0.104[0.036,0.178]</td>
<td>0.102[0.042,0.168]</td>
<td>0.101[0.037,0.176]</td>
<td>0.100[0.036,0.169]</td>
</tr>
<tr>
<td>( \gamma_{22} )</td>
<td>0.05</td>
<td>0.056[-0.001,0.114]</td>
<td>0.055[-0.002,0.114]</td>
<td>0.056[-0.003,0.116]</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{22} )</td>
<td>0.2</td>
<td>0.175[0.042,0.312]</td>
<td>0.178[0.051,0.308]</td>
<td>0.183[0.044,0.325]</td>
<td></td>
</tr>
<tr>
<td>( \delta_{21} )</td>
<td>-0.001[-0.044,0.045]</td>
<td>-0.001[-0.044,0.045]</td>
<td>-0.001[-0.044,0.045]</td>
<td>-0.001[-0.044,0.045]</td>
<td></td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.01</td>
<td>0.013[-0.004,0.031]</td>
<td>0.012[-0.004,0.029]</td>
<td>0.012[-0.004,0.029]</td>
<td>0.019[-0.021,0.059]</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.3</td>
<td>0.309[0.230,0.385]</td>
<td>0.297[0.229,0.362]</td>
<td>0.297[0.229,0.362]</td>
<td>0.278[0.185,0.363]</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.5</td>
<td>0.441[0.283,0.573]</td>
<td>0.446[0.306,0.567]</td>
<td>0.446[0.305,0.566]</td>
<td>0.427[0.289,0.549]</td>
</tr>
<tr>
<td>( \gamma_{12,1} )</td>
<td>0</td>
<td>0.000[-0.079,0.079]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma_{12,2} )</td>
<td>0</td>
<td>-0.010[-0.084,0.063]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma_{12,3} )</td>
<td>0</td>
<td>-0.013[-0.084,0.058]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12,1} )</td>
<td>0</td>
<td>0.012[-0.183,0.203]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12,2} )</td>
<td>0</td>
<td>0.017[-0.144,0.171]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12,3} )</td>
<td>0</td>
<td>-0.004[-0.164,0.150]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3: Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 3 with $n = 1,000$

<table>
<thead>
<tr>
<th>Par</th>
<th>TV</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{10}$</td>
<td>0.06</td>
<td>0.085 [0.041, 0.129]</td>
<td>0.058 [0.017, 0.100]</td>
<td>0.059 [0.018, 0.101]</td>
<td>0.057 [0.016, 0.098]</td>
</tr>
<tr>
<td>$m_{20}$</td>
<td>0.04</td>
<td>0.056 [0.019, 0.092]</td>
<td>0.039 [0.004, 0.074]</td>
<td>0.038 [0.003, 0.073]</td>
<td>0.037 [0.002, 0.072]</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>0.03</td>
<td>0.020 [-0.044, 0.084]</td>
<td>0.023 [-0.039, 0.084]</td>
<td>0.023 [-0.038, 0.084]</td>
<td>0.023 [-0.037, 0.083]</td>
</tr>
<tr>
<td>$m_{22}$</td>
<td>-0.1</td>
<td>-0.102 [-0.163, -0.042]</td>
<td>-0.102 [-0.159, -0.045]</td>
<td>-0.101 [-0.159, -0.044]</td>
<td>-0.102 [-0.159, -0.045]</td>
</tr>
<tr>
<td>$c_{11}$</td>
<td>0.1</td>
<td>0.167 [0.112, 0.228]</td>
<td>0.118 [0.055, 0.187]</td>
<td>0.116 [0.055, 0.184]</td>
<td>0.125 [0.061, 0.195]</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.3</td>
<td>0.444 [0.343, 0.548]</td>
<td>0.312 [0.213, 0.419]</td>
<td>0.312 [0.214, 0.418]</td>
<td>0.312 [0.211, 0.421]</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.4</td>
<td>0.358 [0.232, 0.481]</td>
<td>0.351 [0.233, 0.465]</td>
<td>0.354 [0.238, 0.468]</td>
<td>0.349 [0.233, 0.461]</td>
</tr>
<tr>
<td>$d_{12}$</td>
<td>0.09</td>
<td>0.141 [0.068, 0.220]</td>
<td>0.132 [0.062, 0.207]</td>
<td>0.106 [0.033, 0.188]</td>
<td>0.105 [0.031, 0.187]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0</td>
<td>0.023 [-0.066, 0.126]</td>
<td>0.036 [-0.050, 0.134]</td>
<td>0.027 [-0.056, 0.122]</td>
<td>0.025 [-0.058, 0.119]</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.1</td>
<td>0.108 [-0.001, 0.218]</td>
<td>0.109 [0.004, 0.217]</td>
<td>0.112 [0.004, 0.222]</td>
<td>0.112 [0.004, 0.222]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.3</td>
<td>0.281 [0.105, 0.459]</td>
<td>0.263 [0.101, 0.427]</td>
<td>0.277 [0.099, 0.456]</td>
<td>0.277 [0.099, 0.456]</td>
</tr>
<tr>
<td>$\delta_{12}$</td>
<td>0.09</td>
<td>0.085 [-0.035, 0.206]</td>
<td>0.092 [-0.032, 0.219]</td>
<td>0.092 [-0.032, 0.219]</td>
<td>0.092 [-0.032, 0.219]</td>
</tr>
<tr>
<td>$c_{22}$</td>
<td>-0.01</td>
<td>0.025 [-0.007, 0.060]</td>
<td>0.003 [-0.032, 0.042]</td>
<td>0.004 [-0.030, 0.043]</td>
<td>0.007 [-0.029, 0.046]</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.2</td>
<td>0.302 [0.226, 0.383]</td>
<td>0.210 [0.134, 0.294]</td>
<td>0.214 [0.139, 0.296]</td>
<td>0.214 [0.136, 0.300]</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>0.6</td>
<td>0.560 [0.459, 0.651]</td>
<td>0.564 [0.469, 0.650]</td>
<td>0.562 [0.468, 0.648]</td>
<td>0.561 [0.466, 0.647]</td>
</tr>
<tr>
<td>$d_{21}$</td>
<td>-0.01</td>
<td>0.004 [-0.019, 0.036]</td>
<td>0.002 [-0.020, 0.030]</td>
<td>0.006 [-0.022, 0.042]</td>
<td>0.006 [-0.023, 0.042]</td>
</tr>
<tr>
<td>$c_{21}$</td>
<td>0.1</td>
<td>0.107 [0.040, 0.180]</td>
<td>0.101 [0.041, 0.165]</td>
<td>0.096 [0.037, 0.160]</td>
<td>0.095 [0.036, 0.158]</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.05</td>
<td>0.055 [-0.008, 0.119]</td>
<td>0.057 [-0.003, 0.119]</td>
<td>0.058 [-0.004, 0.122]</td>
<td>0.058 [-0.004, 0.122]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.2</td>
<td>0.191 [0.061, 0.323]</td>
<td>0.174 [0.057, 0.292]</td>
<td>0.176 [0.052, 0.300]</td>
<td>0.176 [0.052, 0.300]</td>
</tr>
<tr>
<td>$\delta_{21}$</td>
<td>0</td>
<td>0.005 [-0.038, 0.053]</td>
<td>0.007 [-0.039, 0.058]</td>
<td>0.007 [-0.039, 0.058]</td>
<td>0.007 [-0.039, 0.058]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0.01</td>
<td>0.013 [-0.005, 0.033]</td>
<td>0.013 [-0.005, 0.032]</td>
<td>0.013 [-0.005, 0.031]</td>
<td>0.017 [-0.025, 0.059]</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.3</td>
<td>0.303 [0.224, 0.381]</td>
<td>0.300 [0.230, 0.365]</td>
<td>0.299 [0.228, 0.365]</td>
<td>0.284 [0.194, 0.367]</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>0.5</td>
<td>0.447 [0.287, 0.580]</td>
<td>0.446 [0.301, 0.570]</td>
<td>0.444 [0.300, 0.568]</td>
<td>0.418 [0.272, 0.546]</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>0</td>
<td>0.008 [-0.074, 0.090]</td>
<td>0.007 [-0.085, 0.070]</td>
<td>0.007 [-0.148, 0.143]</td>
<td>0.004 [-0.142, 0.146]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0</td>
<td>-0.005 [-0.182, 0.170]</td>
<td>-0.001 [-0.149, 0.143]</td>
<td>0.004 [-0.142, 0.146]</td>
<td>0.004 [-0.142, 0.146]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0</td>
<td>-0.005 [-0.182, 0.170]</td>
<td>-0.001 [-0.149, 0.143]</td>
<td>0.004 [-0.142, 0.146]</td>
<td>0.004 [-0.142, 0.146]</td>
</tr>
</tbody>
</table>
Table 4.4: Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 4 with \( n = 1,000 \)

<table>
<thead>
<tr>
<th>Par</th>
<th>TV</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{10} )</td>
<td>0.06</td>
<td>0.087[0.045,0.129]</td>
<td>0.065[0.025,0.106]</td>
<td>0.065[0.024,0.106]</td>
<td>0.059[0.019,0.100]</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.04</td>
<td>0.062[0.027,0.097]</td>
<td>0.049[0.015,0.083]</td>
<td>0.048[0.014,0.082]</td>
<td>0.042[0.008,0.075]</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.03</td>
<td>0.029[-0.034,0.091]</td>
<td>0.031[-0.030,0.091]</td>
<td>0.031[-0.029,0.091]</td>
<td>0.033[-0.027,0.092]</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>-0.1</td>
<td>-0.104[-0.163,-0.044]</td>
<td>-0.103[-0.160,-0.045]</td>
<td>-0.102[-0.160,-0.044]</td>
<td>-0.100[-0.157,-0.043]</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.1</td>
<td>0.172[0.118,0.233]</td>
<td>0.123[0.062,0.190]</td>
<td>0.124[0.063,0.190]</td>
<td>0.131[0.069,0.199]</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.3</td>
<td>0.45[0.354,0.551]</td>
<td>0.347[0.248,0.454]</td>
<td>0.347[0.247,0.455]</td>
<td>0.318[0.218,0.427]</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.4</td>
<td>0.339[0.217,0.458]</td>
<td>0.338[0.224,0.449]</td>
<td>0.336[0.222,0.447]</td>
<td>0.329[0.217,0.438]</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.09</td>
<td>0.143[0.070,0.222]</td>
<td>0.141[0.071,0.217]</td>
<td>0.128[0.047,0.217]</td>
<td>0.112[0.034,0.199]</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0</td>
<td>0.033[-0.058,0.137]</td>
<td>0.035[-0.050,0.132]</td>
<td>0.034[-0.051,0.131]</td>
<td>0.034[-0.051,0.129]</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>0.1</td>
<td>0.105[0.003,0.210]</td>
<td>0.106[0.004,0.211]</td>
<td>0.110[0.004,0.220]</td>
<td>0.110[0.004,0.220]</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.3</td>
<td>0.212[0.053,0.374]</td>
<td>0.203[0.050,0.358]</td>
<td>0.277[0.109,0.445]</td>
<td>0.277[0.109,0.445]</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.4</td>
<td>0.339[0.217,0.458]</td>
<td>0.338[0.224,0.449]</td>
<td>0.336[0.222,0.447]</td>
<td>0.329[0.217,0.438]</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>0.002[-0.019,0.031]</td>
<td>0.001[-0.019,0.029]</td>
<td>0.008[-0.019,0.043]</td>
<td>0.004[-0.022,0.038]</td>
</tr>
<tr>
<td>( c_{21} )</td>
<td>0.1</td>
<td>0.113[0.049,0.183]</td>
<td>0.109[0.049,0.175]</td>
<td>0.105[0.045,0.170]</td>
<td>0.103[0.044,0.166]</td>
</tr>
<tr>
<td>( \gamma_{22} )</td>
<td>0.05</td>
<td>0.052[-0.006,0.111]</td>
<td>0.053[-0.004,0.110]</td>
<td>0.058[-0.002,0.118]</td>
<td>0.058[-0.002,0.118]</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.2</td>
<td>0.137[0.014,0.261]</td>
<td>0.128[0.020,0.236]</td>
<td>0.183[0.065,0.301]</td>
<td>0.183[0.065,0.301]</td>
</tr>
<tr>
<td>( \delta_{21} )</td>
<td>0</td>
<td>0.046[-0.074,0.168]</td>
<td>0.079[-0.045,0.206]</td>
<td>0.003[-0.032,0.041]</td>
<td>0.003[-0.032,0.041]</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>-0.01</td>
<td>0.021[-0.010,0.054]</td>
<td>-0.001[-0.035,0.037]</td>
<td>0.001[-0.034,0.038]</td>
<td>0.003[-0.032,0.041]</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.2</td>
<td>0.308[0.235,0.384]</td>
<td>0.242[0.164,0.329]</td>
<td>0.244[0.168,0.328]</td>
<td>0.217[0.141,0.301]</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.6</td>
<td>0.554[0.455,0.642]</td>
<td>0.554[0.460,0.638]</td>
<td>0.553[0.459,0.638]</td>
<td>0.551[0.458,0.636]</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>0.002[-0.019,0.031]</td>
<td>0.001[-0.019,0.029]</td>
<td>0.008[-0.019,0.043]</td>
<td>0.004[-0.022,0.038]</td>
</tr>
<tr>
<td>( c_{21} )</td>
<td>0.1</td>
<td>0.113[0.049,0.183]</td>
<td>0.109[0.049,0.175]</td>
<td>0.105[0.045,0.170]</td>
<td>0.103[0.044,0.166]</td>
</tr>
<tr>
<td>( \gamma_{22} )</td>
<td>0.05</td>
<td>0.052[-0.006,0.111]</td>
<td>0.053[-0.004,0.110]</td>
<td>0.058[-0.002,0.118]</td>
<td>0.058[-0.002,0.118]</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.2</td>
<td>0.137[0.014,0.261]</td>
<td>0.128[0.020,0.236]</td>
<td>0.183[0.065,0.301]</td>
<td>0.183[0.065,0.301]</td>
</tr>
<tr>
<td>( \delta_{21} )</td>
<td>0</td>
<td>0.046[-0.074,0.168]</td>
<td>0.079[-0.045,0.206]</td>
<td>0.003[-0.032,0.041]</td>
<td>0.003[-0.032,0.041]</td>
</tr>
</tbody>
</table>


124
Table 4.5: Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 1 with \( n = 3,000 \)

<table>
<thead>
<tr>
<th>Par</th>
<th>TV</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.06[0.045,0.075]</td>
<td>0.060[0.044,0.075]</td>
<td>0.060[0.044,0.075]</td>
<td>0.058[0.043,0.074]</td>
</tr>
<tr>
<td>( m_{10} )</td>
<td>0.04</td>
<td>0.040[0.033,0.048]</td>
<td>0.040[0.032,0.048]</td>
<td>0.040[0.032,0.048]</td>
<td>0.039[0.031,0.047]</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.03</td>
<td>0.031[-0.003,0.065]</td>
<td>0.031[-0.003,0.065]</td>
<td>0.031[-0.003,0.065]</td>
<td>0.031[-0.003,0.065]</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>-0.1</td>
<td>-0.099[-0.132,-0.067]</td>
<td>-0.099[-0.132,-0.067]</td>
<td>-0.099[-0.131,-0.066]</td>
<td>-0.099[-0.132,-0.066]</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.1</td>
<td>0.097[0.079,0.117]</td>
<td>0.099[0.078,0.121]</td>
<td>0.097[0.075,0.120]</td>
<td>0.099[0.077,0.122]</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.3</td>
<td>0.292[0.248,0.340]</td>
<td>0.297[0.246,0.352]</td>
<td>0.285[0.236,0.337]</td>
<td>0.266[0.215,0.319]</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.4</td>
<td>0.379[0.294,0.461]</td>
<td>0.368[0.285,0.448]</td>
<td>0.376[0.287,0.464]</td>
<td>0.377[0.287,0.464]</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.09</td>
<td>0.096[0.036,0.162]</td>
<td>0.100[0.040,0.166]</td>
<td>0.132[0.054,0.218]</td>
<td>0.098[0.021,0.180]</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0</td>
<td>0.047[-0.052,0.152]</td>
<td>0.063[-0.019,0.147]</td>
<td>0.056[-0.069,0.189]</td>
<td>0.061[-0.060,0.189]</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>0.001</td>
<td>0.001[-0.018,0.020]</td>
<td>0.002[-0.018,0.021]</td>
<td>-0.001[-0.022,0.019]</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{11} )</td>
<td>0.003</td>
<td>0.003[-0.049,0.056]</td>
<td>0.001[-0.053,0.056]</td>
<td>0.029[-0.034,0.093]</td>
<td></td>
</tr>
<tr>
<td>( \delta_{12} )</td>
<td>0</td>
<td>-0.038[-0.139,0.066]</td>
<td>0.011[-0.092,0.116]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>-0.01</td>
<td>-0.009[-0.014,-0.005]</td>
<td>-0.010[-0.015,-0.004]</td>
<td>-0.009[-0.015,-0.004]</td>
<td>-0.009[-0.016,-0.003]</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.2</td>
<td>0.200[0.163,0.240]</td>
<td>0.211[0.165,0.263]</td>
<td>0.185[0.157,0.213]</td>
<td>0.170[0.141,0.200]</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.6</td>
<td>0.580[0.518,0.637]</td>
<td>0.564[0.498,0.624]</td>
<td>0.571[0.503,0.631]</td>
<td>0.568[0.496,0.632]</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>-0.008[-0.011,-0.004]</td>
<td>-0.008[-0.011,-0.003]</td>
<td>-0.007[-0.011,0.000]</td>
<td>-0.008[-0.012,-0.002]</td>
</tr>
<tr>
<td>( e_{21} )</td>
<td>0.1</td>
<td>0.104[0.082,0.127]</td>
<td>0.108[0.087,0.129]</td>
<td>0.108[0.084,0.134]</td>
<td>0.110[0.085,0.137]</td>
</tr>
<tr>
<td>( \gamma_{22} )</td>
<td>0.000</td>
<td>0.000[-0.005,0.005]</td>
<td>-0.001[-0.006,0.004]</td>
<td>-0.001[-0.006,0.004]</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{22} )</td>
<td>0</td>
<td>0.000[-0.004,0.000]</td>
<td>0.025[-0.010,0.064]</td>
<td>0.056[0.015,0.098]</td>
<td></td>
</tr>
<tr>
<td>( \delta_{21} )</td>
<td>0</td>
<td>0.000[-0.008,0.007]</td>
<td>0.001[-0.006,0.008]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0.01</td>
<td>0.009[0.007,0.012]</td>
<td>0.009[0.006,0.012]</td>
<td>0.009[0.006,0.012]</td>
<td>0.011[0.004,0.018]</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.3</td>
<td>0.290[0.257,0.324]</td>
<td>0.296[0.260,0.334]</td>
<td>0.284[0.255,0.312]</td>
<td>0.246[0.207,0.282]</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.5</td>
<td>0.486[0.421,0.546]</td>
<td>0.484[0.419,0.546]</td>
<td>0.493[0.428,0.553]</td>
<td>0.492[0.423,0.556]</td>
</tr>
<tr>
<td>( \gamma_{12},1 )</td>
<td>0</td>
<td>-0.002[-0.013,0.009]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma_{12},2 )</td>
<td>0</td>
<td>-0.002[-0.013,0.009]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma_{12},3 )</td>
<td>0</td>
<td>-0.002[-0.013,0.009]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12},1 )</td>
<td>0</td>
<td>0.053[0.001,0.106]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12},2 )</td>
<td>0</td>
<td>0.011[-0.047,0.067]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_{12},3 )</td>
<td>0</td>
<td>0.029[-0.030,0.085]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.6: Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 2 with $n = 3,000$

<table>
<thead>
<tr>
<th>Par</th>
<th>TV</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.06 [0.063, 0.112]</td>
<td>0.061 [0.040, 0.083]</td>
<td>0.061 [0.039, 0.083]</td>
<td>0.061 [0.039, 0.082]</td>
</tr>
<tr>
<td>$n_{10}$</td>
<td>0.04</td>
<td>0.058 [0.038, 0.078]</td>
<td>0.041 [0.023, 0.060]</td>
<td>0.041 [0.023, 0.060]</td>
<td>0.041 [0.022, 0.059]</td>
</tr>
<tr>
<td>$n_{20}$</td>
<td>0.03</td>
<td>0.028 [-0.009, 0.065]</td>
<td>0.032 [-0.003, 0.066]</td>
<td>0.032 [-0.003, 0.066]</td>
<td>0.031 [-0.003, 0.066]</td>
</tr>
<tr>
<td>$n_{32}$</td>
<td>-0.1</td>
<td>-0.102 [-0.139, -0.066]</td>
<td>-0.101 [-0.135, -0.066]</td>
<td>-0.101 [-0.135, -0.066]</td>
<td>-0.101 [-0.135, -0.066]</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>0.1</td>
<td>0.155 [0.125, 0.188]</td>
<td>0.106 [0.072, 0.140]</td>
<td>0.106 [0.073, 0.140]</td>
<td>0.109 [0.075, 0.144]</td>
</tr>
<tr>
<td>$m_{20}$</td>
<td>0.3</td>
<td>0.453 [0.388, 0.519]</td>
<td>0.304 [0.247, 0.365]</td>
<td>0.301 [0.245, 0.361]</td>
<td>0.296 [0.236, 0.358]</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>0.4</td>
<td>0.376 [0.302, 0.448]</td>
<td>0.382 [0.313, 0.448]</td>
<td>0.383 [0.316, 0.448]</td>
<td>0.381 [0.313, 0.447]</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>0.09</td>
<td>0.096 [0.058, 0.137]</td>
<td>0.095 [0.059, 0.133]</td>
<td>0.095 [0.052, 0.141]</td>
<td>0.092 [0.048, 0.139]</td>
</tr>
<tr>
<td>$c_{11}$</td>
<td>0</td>
<td>0.010 [-0.034, 0.060]</td>
<td>0.008 [-0.032, 0.054]</td>
<td>0.007 [-0.034, 0.051]</td>
<td>0.007 [-0.034, 0.052]</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.1</td>
<td>0.104 [0.049, 0.159]</td>
<td>0.103 [0.048, 0.159]</td>
<td>0.102 [0.046, 0.159]</td>
<td>0.102 [0.046, 0.159]</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.3</td>
<td>0.293 [0.190, 0.398]</td>
<td>0.294 [0.193, 0.398]</td>
<td>0.310 [0.201, 0.420]</td>
<td>0.310 [0.201, 0.420]</td>
</tr>
<tr>
<td>$\delta_{11}$</td>
<td>0.09</td>
<td>0.055 [0.070]</td>
<td>0.012 [-0.051, 0.077]</td>
<td>0.012 [-0.051, 0.077]</td>
<td>0.012 [-0.051, 0.077]</td>
</tr>
<tr>
<td>$c_{22}$</td>
<td>-0.01</td>
<td>0.015 [-0.002, 0.033]</td>
<td>-0.008 [-0.025, 0.011]</td>
<td>-0.008 [-0.025, 0.011]</td>
<td>-0.006 [-0.024, 0.013]</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.2</td>
<td>0.302 [0.255, 0.350]</td>
<td>0.204 [0.160, 0.250]</td>
<td>0.203 [0.160, 0.248]</td>
<td>0.198 [0.154, 0.245]</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>0.6</td>
<td>0.585 [0.530, 0.636]</td>
<td>0.590 [0.540, 0.638]</td>
<td>0.590 [0.541, 0.638]</td>
<td>0.590 [0.541, 0.637]</td>
</tr>
<tr>
<td>$d_{21}$</td>
<td>-0.01</td>
<td>-0.004 [-0.014, 0.011]</td>
<td>-0.005 [-0.014, 0.008]</td>
<td>-0.002 [-0.014, 0.015]</td>
<td>-0.003 [-0.015, 0.014]</td>
</tr>
<tr>
<td>$e_{21}$</td>
<td>0.1</td>
<td>0.104 [0.068, 0.142]</td>
<td>0.101 [0.069, 0.133]</td>
<td>0.100 [0.067, 0.134]</td>
<td>0.099 [0.067, 0.133]</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.05</td>
<td>0.052 [0.021, 0.083]</td>
<td>0.052 [0.022, 0.083]</td>
<td>0.052 [0.021, 0.083]</td>
<td>0.052 [0.021, 0.083]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.2</td>
<td>0.188 [0.112, 0.265]</td>
<td>0.186 [0.114, 0.258]</td>
<td>0.196 [0.121, 0.273]</td>
<td>0.196 [0.121, 0.273]</td>
</tr>
<tr>
<td>$\delta_{21}$</td>
<td>-0.002 [-0.022, 0.020]</td>
<td>0.000 [-0.021, 0.022]</td>
<td>0.000 [-0.021, 0.022]</td>
<td>0.000 [-0.021, 0.022]</td>
<td>0.000 [-0.021, 0.022]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0.01</td>
<td>0.011 [0.002, 0.020]</td>
<td>0.010 [0.001, 0.018]</td>
<td>0.010 [0.001, 0.018]</td>
<td>0.013 [0.008, 0.034]</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.3</td>
<td>0.304 [0.257, 0.351]</td>
<td>0.294 [0.255, 0.332]</td>
<td>0.295 [0.256, 0.332]</td>
<td>0.278 [0.227, 0.325]</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>0.5</td>
<td>0.490 [0.408, 0.563]</td>
<td>0.490 [0.419, 0.555]</td>
<td>0.490 [0.419, 0.554]</td>
<td>0.483 [0.409, 0.550]</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>-0.001 [-0.044, 0.042]</td>
<td>-0.006 [-0.045, 0.034]</td>
<td>-0.005 [-0.044, 0.033]</td>
<td>-0.027 [-0.080, 0.033]</td>
<td>-0.012 [-0.076, 0.097]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.01</td>
<td>0.490 [0.419, 0.555]</td>
<td>0.490 [0.419, 0.554]</td>
<td>0.490 [0.419, 0.554]</td>
<td>0.483 [0.409, 0.550]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.12</td>
<td>0.011 [-0.074, 0.093]</td>
<td>0.011 [-0.074, 0.093]</td>
<td>0.011 [-0.074, 0.093]</td>
<td>0.011 [-0.074, 0.093]</td>
</tr>
</tbody>
</table>
Table 4.7: Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 3 with $n = 3,000$

<table>
<thead>
<tr>
<th>Par $\nu$</th>
<th>TV</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{10}$</td>
<td>0.06</td>
<td>0.088 [0.063, 0.113]</td>
<td>0.061 [0.038, 0.084]</td>
<td>0.061 [0.038, 0.085]</td>
<td>0.061 [0.038, 0.084]</td>
</tr>
<tr>
<td>$m_{20}$</td>
<td>0.04</td>
<td>0.059 [0.039, 0.080]</td>
<td>0.042 [0.023, 0.062]</td>
<td>0.041 [0.022, 0.061]</td>
<td>0.041 [0.022, 0.060]</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>0.03</td>
<td>0.028 [-0.010, 0.065]</td>
<td>0.031 [-0.004, 0.066]</td>
<td>0.031 [-0.004, 0.066]</td>
<td>0.031 [-0.003, 0.066]</td>
</tr>
<tr>
<td>$m_{22}$</td>
<td>-0.1</td>
<td>-0.103 [-0.139, -0.068]</td>
<td>-0.101 [-0.135, -0.068]</td>
<td>-0.101 [-0.134, -0.067]</td>
<td>-0.100 [-0.134, -0.067]</td>
</tr>
<tr>
<td>$c_{11}$</td>
<td>0.1</td>
<td>0.154 [0.123, 0.187]</td>
<td>0.106 [0.071, 0.142]</td>
<td>0.105 [0.071, 0.141]</td>
<td>0.109 [0.074, 0.146]</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.3</td>
<td>0.448 [0.385, 0.511]</td>
<td>0.305 [0.249, 0.365]</td>
<td>0.306 [0.250, 0.367]</td>
<td>0.300 [0.242, 0.362]</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.4</td>
<td>0.386 [0.311, 0.458]</td>
<td>0.383 [0.315, 0.450]</td>
<td>0.387 [0.319, 0.454]</td>
<td>0.386 [0.317, 0.452]</td>
</tr>
<tr>
<td>$d_{12}$</td>
<td>0.09</td>
<td>0.136 [0.091, 0.182]</td>
<td>0.127 [0.086, 0.170]</td>
<td>0.095 [0.051, 0.142]</td>
<td>0.094 [0.048, 0.142]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0</td>
<td>0.005 [-0.046, 0.063]</td>
<td>0.013 [-0.035, 0.066]</td>
<td>0.006 [-0.041, 0.059]</td>
<td>0.006 [-0.041, 0.059]</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.1</td>
<td>0.102 [0.044, 0.162]</td>
<td>0.102 [0.044, 0.162]</td>
<td>0.100 [0.040, 0.160]</td>
<td>0.100 [0.040, 0.160]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.3</td>
<td>0.292 [0.191, 0.395]</td>
<td>0.281 [0.181, 0.381]</td>
<td>0.295 [0.190, 0.401]</td>
<td>0.295 [0.190, 0.401]</td>
</tr>
<tr>
<td>$\delta_{12}$</td>
<td>0.09</td>
<td>0</td>
<td>0.082 [0.012, 0.153]</td>
<td>0.085 [0.011, 0.160]</td>
<td>0.085 [0.011, 0.160]</td>
</tr>
<tr>
<td>$c_{22}$</td>
<td>-0.01</td>
<td>0.016 [-0.001, 0.034]</td>
<td>-0.006 [-0.024, 0.014]</td>
<td>-0.006 [-0.024, 0.013]</td>
<td>-0.005 [-0.023, 0.015]</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.2</td>
<td>0.305 [0.260, 0.352]</td>
<td>0.202 [0.159, 0.248]</td>
<td>0.207 [0.164, 0.253]</td>
<td>0.201 [0.158, 0.247]</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>0.6</td>
<td>0.577 [0.522, 0.628]</td>
<td>0.583 [0.533, 0.631]</td>
<td>0.583 [0.533, 0.630]</td>
<td>0.584 [0.535, 0.631]</td>
</tr>
<tr>
<td>$d_{21}$</td>
<td>-0.01</td>
<td>-0.003 [-0.014, 0.013]</td>
<td>-0.004 [-0.014, 0.009]</td>
<td>-0.003 [-0.016, 0.012]</td>
<td>-0.004 [-0.016, 0.012]</td>
</tr>
<tr>
<td>$e_{21}$</td>
<td>0.1</td>
<td>0.104 [0.068, 0.141]</td>
<td>0.101 [0.070, 0.134]</td>
<td>0.099 [0.068, 0.132]</td>
<td>0.098 [0.066, 0.132]</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.05</td>
<td>0.049 [0.017, 0.081]</td>
<td>0.051 [0.019, 0.083]</td>
<td>0.049 [0.017, 0.082]</td>
<td>0.049 [0.017, 0.082]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.2</td>
<td>0.201 [0.125, 0.277]</td>
<td>0.186 [0.113, 0.259]</td>
<td>0.197 [0.120, 0.274]</td>
<td>0.197 [0.120, 0.274]</td>
</tr>
<tr>
<td>$\delta_{21}$</td>
<td>0</td>
<td></td>
<td>0.003 [-0.018, 0.026]</td>
<td>0.004 [-0.017, 0.027]</td>
<td>0.004 [-0.017, 0.027]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0.01</td>
<td>0.011 [0.002, 0.021]</td>
<td>0.011 [0.002, 0.020]</td>
<td>0.011 [0.002, 0.020]</td>
<td>0.014 [0.008, 0.036]</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.3</td>
<td>0.306 [0.259, 0.352]</td>
<td>0.301 [0.263, 0.338]</td>
<td>0.299 [0.261, 0.335]</td>
<td>0.285 [0.236, 0.330]</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>0.5</td>
<td>0.487 [0.406, 0.560]</td>
<td>0.484 [0.412, 0.549]</td>
<td>0.481 [0.411, 0.546]</td>
<td>0.476 [0.403, 0.543]</td>
</tr>
<tr>
<td>$\gamma_{12,1}$</td>
<td>-0.004 [-0.049, 0.041]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{12,2}$</td>
<td>-0.005 [-0.047, 0.036]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{12,3}$</td>
<td>-0.004 [-0.044, 0.037]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{12,1}$</td>
<td>0.016 [-0.094, 0.126]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{12,2}$</td>
<td>0.010 [-0.076, 0.092]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{12,3}$</td>
<td>0.012 [-0.075, 0.096]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.8: Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 4 with \( n = 3,000 \)

<table>
<thead>
<tr>
<th>Par</th>
<th>TV</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{10} )</td>
<td>0.06</td>
<td>0.088[0.064,0.113]</td>
<td>0.066[0.043,0.089]</td>
<td>0.066[0.043,0.090]</td>
<td>0.060[0.038,0.083]</td>
</tr>
<tr>
<td>( m_{20} )</td>
<td>0.04</td>
<td>0.061[0.041,0.081]</td>
<td>0.047[0.028,0.066]</td>
<td>0.046[0.027,0.066]</td>
<td>0.041[0.022,0.059]</td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>0.03</td>
<td>0.025[-0.011,0.062]</td>
<td>0.027[-0.008,0.062]</td>
<td>0.027[-0.008,0.062]</td>
<td>0.027[-0.006,0.061]</td>
</tr>
<tr>
<td>( m_{22} )</td>
<td>-0.1</td>
<td>-0.101[-0.135,-0.066]</td>
<td>-0.100[-0.133,-0.066]</td>
<td>-0.099[-0.133,-0.066]</td>
<td>-0.099[-0.132,-0.066]</td>
</tr>
<tr>
<td>( c_{11} )</td>
<td>0.1</td>
<td>0.158[0.128,0.191]</td>
<td>0.111[0.076,0.147]</td>
<td>0.111[0.076,0.147]</td>
<td>0.113[0.078,0.149]</td>
</tr>
<tr>
<td>( a_{11} )</td>
<td>0.3</td>
<td>0.448[0.388,0.509]</td>
<td>0.335[0.279,0.395]</td>
<td>0.334[0.277,0.394]</td>
<td>0.294[0.238,0.354]</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>0.4</td>
<td>0.372[0.300,0.443]</td>
<td>0.372[0.305,0.439]</td>
<td>0.374[0.306,0.441]</td>
<td>0.376[0.310,0.442]</td>
</tr>
<tr>
<td>( d_{12} )</td>
<td>0.09</td>
<td>0.133[0.089,0.178]</td>
<td>0.129[0.088,0.172]</td>
<td>0.108[0.062,0.157]</td>
<td>0.091[0.047,0.138]</td>
</tr>
<tr>
<td>( c_{12} )</td>
<td>0</td>
<td>0.017[-0.034,0.074]</td>
<td>0.019[-0.030,0.073]</td>
<td>0.016[-0.033,0.069]</td>
<td>0.016[-0.033,0.070]</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>0.3</td>
<td>0.226[0.134,0.319]</td>
<td>0.111[0.076,0.147]</td>
<td>0.111[0.076,0.147]</td>
<td>0.113[0.078,0.149]</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.4</td>
<td>0.372[0.300,0.443]</td>
<td>0.372[0.305,0.439]</td>
<td>0.374[0.306,0.441]</td>
<td>0.376[0.310,0.442]</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>-0.002[-0.013,0.012]</td>
<td>-0.003[-0.013,0.011]</td>
<td>0.000[-0.013,0.017]</td>
<td>-0.004[-0.016,0.011]</td>
</tr>
<tr>
<td>( c_{21} )</td>
<td>0.1</td>
<td>0.103[0.068,0.140]</td>
<td>0.100[0.068,0.133]</td>
<td>0.100[0.068,0.133]</td>
<td>0.101[0.068,0.136]</td>
</tr>
<tr>
<td>( a_{22} )</td>
<td>0.2</td>
<td>0.226[0.134,0.319]</td>
<td>0.111[0.076,0.147]</td>
<td>0.111[0.076,0.147]</td>
<td>0.113[0.078,0.149]</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.6</td>
<td>0.579[0.526,0.629]</td>
<td>0.583[0.533,0.630]</td>
<td>0.581[0.532,0.629]</td>
<td>0.583[0.535,0.629]</td>
</tr>
<tr>
<td>( d_{21} )</td>
<td>-0.01</td>
<td>-0.002[-0.013,0.012]</td>
<td>-0.003[-0.013,0.011]</td>
<td>0.000[-0.013,0.017]</td>
<td>-0.004[-0.016,0.011]</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>0.05</td>
<td>0.046[0.014,0.078]</td>
<td>0.048[0.017,0.080]</td>
<td>0.048[0.016,0.080]</td>
<td>0.048[0.016,0.080]</td>
</tr>
</tbody>
</table>

Table 4.9: Mean posterior model probabilities and model selection counts for the APVECH family of models in the simulation study with \( n = 1,000 \)

<table>
<thead>
<tr>
<th>Data</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set</td>
<td>Mean Prob</td>
<td>Mean Prob</td>
<td>Mean Prob</td>
<td>Mean Prob</td>
</tr>
<tr>
<td>1</td>
<td>0.748</td>
<td>0.204</td>
<td>0.042</td>
<td>0.006</td>
</tr>
<tr>
<td>2</td>
<td>0.725</td>
<td>0.218</td>
<td>0.053</td>
<td>0.004</td>
</tr>
<tr>
<td>1</td>
<td>0.005</td>
<td>0.579</td>
<td>0.300</td>
<td>0.016</td>
</tr>
<tr>
<td>2</td>
<td>0.004</td>
<td>0.575</td>
<td>0.337</td>
<td>0.296</td>
</tr>
<tr>
<td>1</td>
<td>0.009</td>
<td>0.527</td>
<td>0.335</td>
<td>0.108</td>
</tr>
<tr>
<td>2</td>
<td>0.008</td>
<td>0.525</td>
<td>0.373</td>
<td>0.094</td>
</tr>
<tr>
<td>1</td>
<td>0.027</td>
<td>0.435</td>
<td>0.190</td>
<td>0.348</td>
</tr>
<tr>
<td>2</td>
<td>0.025</td>
<td>0.453</td>
<td>0.226</td>
<td>0.296</td>
</tr>
</tbody>
</table>

128
### Table 4.10: Mean posterior model probabilities and model selection counts for the APVECH family of models in the simulation study with \( n = 3,000 \)

<table>
<thead>
<tr>
<th>Data</th>
<th>Prior Set</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Prob</td>
<td># of Sel</td>
<td>Mean Prob</td>
<td># of Sel</td>
</tr>
<tr>
<td>Model 1</td>
<td>1</td>
<td>0.763</td>
<td>99</td>
<td>0.154</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.739</td>
<td>99</td>
<td>0.166</td>
<td>1</td>
</tr>
<tr>
<td>Model 2</td>
<td>1</td>
<td>1E-11</td>
<td>0</td>
<td>0.601</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9E-12</td>
<td>0</td>
<td>0.598</td>
<td>94</td>
</tr>
<tr>
<td>Model 3</td>
<td>1</td>
<td>4E-14</td>
<td>0</td>
<td>0.466</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3E-14</td>
<td>0</td>
<td>0.461</td>
<td>59</td>
</tr>
<tr>
<td>Model 4</td>
<td>1</td>
<td>5E-11</td>
<td>0</td>
<td>0.073</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5E-11</td>
<td>0</td>
<td>0.084</td>
<td>3</td>
</tr>
</tbody>
</table>

### Table 4.11: Posterior model probabilities for the APVECH family of models fitted to each bivariate combination of S&P 500 (US), FTSE 100 (UK), All Ordinaries (AU) and Nikkei 225 (JP) returns.

<table>
<thead>
<tr>
<th>Data</th>
<th>Prior Set</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P500/</td>
<td>1</td>
<td>0.1371</td>
<td>0.1737</td>
<td>0.0388</td>
<td>0.6505</td>
</tr>
<tr>
<td>FTSE100</td>
<td>2</td>
<td>0.1340</td>
<td>0.1815</td>
<td>0.0431</td>
<td>0.6414</td>
</tr>
<tr>
<td>S&amp;P500/</td>
<td>1</td>
<td>0.1207</td>
<td>0.1613</td>
<td>0.4149</td>
<td>0.3032</td>
</tr>
<tr>
<td>All Ordinaries</td>
<td>2</td>
<td>0.1169</td>
<td>0.1638</td>
<td>0.4380</td>
<td>0.2814</td>
</tr>
<tr>
<td>S&amp;P500/</td>
<td>1</td>
<td>0.0148</td>
<td>0.1457</td>
<td>0.2091</td>
<td>0.6304</td>
</tr>
<tr>
<td>Nikkei 225</td>
<td>2</td>
<td>0.0147</td>
<td>0.1471</td>
<td>0.2206</td>
<td>0.6177</td>
</tr>
<tr>
<td>FTSE100/</td>
<td>1</td>
<td>0.2118</td>
<td>0.1066</td>
<td>0.1602</td>
<td>0.5214</td>
</tr>
<tr>
<td>All Ordinaries</td>
<td>2</td>
<td>0.2033</td>
<td>0.1102</td>
<td>0.1734</td>
<td>0.5131</td>
</tr>
<tr>
<td>FTSE100/</td>
<td>1</td>
<td>0.0216</td>
<td>0.1844</td>
<td>0.1575</td>
<td>0.6365</td>
</tr>
<tr>
<td>Nikkei 225</td>
<td>2</td>
<td>0.0214</td>
<td>0.1847</td>
<td>0.1634</td>
<td>0.6306</td>
</tr>
<tr>
<td>All Ordinaries/</td>
<td>1</td>
<td>0.0058</td>
<td>0.2346</td>
<td>0.1026</td>
<td>0.6571</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0055</td>
<td>0.2380</td>
<td>0.1089</td>
<td>0.6476</td>
</tr>
</tbody>
</table>
Table 4.12: Parameter posterior means and 95% credible intervals for the APVECH family of models fitted to the S&P 500/FTSE 100 bivariate returns

<table>
<thead>
<tr>
<th>Par</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{10}$</td>
<td>0.069[0.049,0.089]</td>
<td>0.055[0.034,0.076]</td>
<td>0.049[0.029,0.071]</td>
<td>0.051[0.031,0.071]</td>
</tr>
<tr>
<td>$m_{20}$</td>
<td>0.062[0.041,0.081]</td>
<td>0.052[0.032,0.072]</td>
<td>0.053[0.036,0.070]</td>
<td>0.046[0.027,0.065]</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>-0.079[-0.103,-0.053]</td>
<td>-0.070[-0.095,-0.045]</td>
<td>-0.073[-0.097,-0.048]</td>
<td>-0.069[-0.097,-0.041]</td>
</tr>
<tr>
<td>$m_{22}$</td>
<td>0.007[-0.018,0.031]</td>
<td>0.005[-0.021,0.030]</td>
<td>0.005[-0.018,0.028]</td>
<td>0.006[-0.019,0.032]</td>
</tr>
<tr>
<td>$c_{11}$</td>
<td>0.010[0.007,0.014]</td>
<td>0.018[0.000,0.036]</td>
<td>0.019[0.007,0.034]</td>
<td>0.022[0.009,0.053]</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.040[0.031,0.050]</td>
<td>0.015[-0.002,0.032]</td>
<td>0.030[0.015,0.043]</td>
<td>0.014[-0.011,0.041]</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.949[0.936,0.962]</td>
<td>0.850[0.817,0.883]</td>
<td>0.935[0.883,0.970]</td>
<td>0.791[0.749,0.830]</td>
</tr>
<tr>
<td>$d_{12}$</td>
<td>0.015[0.009,0.021]</td>
<td>0.004[-0.011,0.020]</td>
<td>0.027[0.014,0.042]</td>
<td>-0.006[-0.024,0.021]</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>-0.018[-0.027,-0.008]</td>
<td>0.058[0.031,0.086]</td>
<td>-0.040[-0.052,-0.027]</td>
<td>0.096[0.060,0.132]</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.022[-0.023,0.106]</td>
<td>0.020[-0.010,0.050]</td>
<td>0.032[-0.045,0.147]</td>
<td>0.032[-0.045,0.147]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.086[0.055,0.118]</td>
<td>0.032[0.009,0.057]</td>
<td>0.120[0.076,0.162]</td>
<td>0.120[0.076,0.162]</td>
</tr>
<tr>
<td>$\delta_{12}$</td>
<td>0.012[-0.012,0.037]</td>
<td>0.012[-0.012,0.037]</td>
<td>0.029[-0.008,0.058]</td>
<td>0.029[-0.008,0.058]</td>
</tr>
<tr>
<td>$c_{22}$</td>
<td>0.016[0.011,0.022]</td>
<td>0.017[0.009,0.024]</td>
<td>0.019[0.002,0.040]</td>
<td>0.022[0.009,0.037]</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.060[0.048,0.072]</td>
<td>0.041[0.022,0.063]</td>
<td>0.059[0.035,0.085]</td>
<td>0.042[0.019,0.066]</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>0.908[0.886,0.928]</td>
<td>0.906[0.844,0.952]</td>
<td>0.867[0.810,0.912]</td>
<td>0.900[0.834,0.956]</td>
</tr>
<tr>
<td>$d_{21}$</td>
<td>0.025[0.017,0.033]</td>
<td>0.039[0.027,0.052]</td>
<td>0.009[-0.003,0.023]</td>
<td>0.035[0.020,0.050]</td>
</tr>
<tr>
<td>$e_{21}$</td>
<td>-0.009[-0.019,0.002]</td>
<td>-0.024[-0.051,0.011]</td>
<td>0.009[-0.012,0.034]</td>
<td>-0.034[-0.077,0.013]</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.012[0.000,0.025]</td>
<td>0.026[0.004,0.047]</td>
<td>0.003[-0.017,0.022]</td>
<td>0.003[-0.017,0.022]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.030[0.012,0.049]</td>
<td>0.021[-0.007,0.050]</td>
<td>0.046[0.020,0.072]</td>
<td>0.046[0.020,0.072]</td>
</tr>
<tr>
<td>$\delta_{21}$</td>
<td>0.033[0.018,0.048]</td>
<td>0.033[0.018,0.048]</td>
<td>0.024[0.008,0.040]</td>
<td>0.024[0.008,0.040]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0.005[0.003,0.006]</td>
<td>0.008[0.006,0.012]</td>
<td>0.011[0.005,0.020]</td>
<td>0.007[-0.007,0.021]</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.036[0.029,0.044]</td>
<td>0.048[0.034,0.062]</td>
<td>0.045[0.032,0.066]</td>
<td>0.040[0.016,0.065]</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>0.945[0.934,0.955]</td>
<td>0.919[0.897,0.939]</td>
<td>0.913[0.863,0.947]</td>
<td>0.883[0.848,0.912]</td>
</tr>
<tr>
<td>$\gamma_{12,1}$</td>
<td>-0.005[-0.031,0.021]</td>
<td>0.009[-0.020,0.037]</td>
<td>-0.006[-0.024,0.012]</td>
<td>-0.006[-0.024,0.012]</td>
</tr>
<tr>
<td>$\gamma_{12,2}$</td>
<td>0.062[0.035,0.087]</td>
<td>-0.066[-0.114,-0.019]</td>
<td>0.062[0.035,0.087]</td>
<td>-0.066[-0.114,-0.019]</td>
</tr>
<tr>
<td>$\gamma_{12,3}$</td>
<td>-0.041[-0.082,-0.003]</td>
<td>0.014[0.020,0.052]</td>
<td>-0.041[-0.082,-0.003]</td>
<td>0.014[0.020,0.052]</td>
</tr>
</tbody>
</table>

Table 4.13: Acceptance rates for the APVECH family of models fitted to the S&P 500/FTSE 100 bivariate returns

<table>
<thead>
<tr>
<th>Step</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98.1%</td>
<td>88.3%</td>
<td>83.2%</td>
<td>86.7%</td>
</tr>
<tr>
<td>2</td>
<td>68.5%</td>
<td>77.3%</td>
<td>50.8%</td>
<td>76.1%</td>
</tr>
<tr>
<td>3</td>
<td>81.2%</td>
<td>46.9%</td>
<td>42.1%</td>
<td>53.9%</td>
</tr>
<tr>
<td>4</td>
<td>81.6%</td>
<td>53.9%</td>
<td>41.0%</td>
<td>58.6%</td>
</tr>
<tr>
<td>5</td>
<td>28.0%</td>
<td>85.3%</td>
<td>75.8%</td>
<td>85.9%</td>
</tr>
</tbody>
</table>
Table 4.14: Parameter posterior means and 95% credible intervals for the APVECH family of models fitted to the S&P 500/All Ordinaries bivariate returns

<table>
<thead>
<tr>
<th>Par</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{10}$</td>
<td>0.065[0.045,0.085]</td>
<td>0.056[0.028,0.081]</td>
<td>0.035[0.013,0.066]</td>
<td>0.047[0.027,0.069]</td>
</tr>
<tr>
<td>$m_{20}$</td>
<td>0.060[0.041,0.078]</td>
<td>0.053[0.036,0.071]</td>
<td>0.059[0.041,0.077]</td>
<td>0.065[0.047,0.083]</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>-0.020[-0.047,0.006]</td>
<td>-0.020[-0.053,0.011]</td>
<td>-0.022[-0.050,0.010]</td>
<td>-0.024[-0.052,0.003]</td>
</tr>
<tr>
<td>$m_{22}$</td>
<td>0.094[0.067,0.121]</td>
<td>0.098[0.070,0.127]</td>
<td>0.098[0.069,0.126]</td>
<td>0.097[0.067,0.125]</td>
</tr>
<tr>
<td>$c_{11}$</td>
<td>0.023[0.017,0.030]</td>
<td>0.035[0.014,0.068]</td>
<td>0.008[0.002,0.018]</td>
<td>0.039[0.031,0.047]</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.054[0.044,0.066]</td>
<td>0.035[0.019,0.052]</td>
<td>0.038[0.021,0.057]</td>
<td>0.031[0.018,0.045]</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.947[0.934,0.960]</td>
<td>0.929[0.889,0.952]</td>
<td>0.927[0.905,0.946]</td>
<td>0.922[0.890,0.949]</td>
</tr>
<tr>
<td>$d_{12}$</td>
<td>0.017[0.003,0.030]</td>
<td>0.006[-0.006,0.016]</td>
<td>-0.003[-0.015,0.009]</td>
<td>0.025[0.015,0.035]</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>-0.055[-0.081,-0.029]</td>
<td>-0.044[-0.068,-0.016]</td>
<td>-0.021[-0.036,-0.004]</td>
<td>-0.074[-0.103,-0.045]</td>
</tr>
<tr>
<td>$c_{11}$</td>
<td>0.001[-0.053,0.077]</td>
<td>0.040[0.015,0.062]</td>
<td>0.025[-0.017,0.108]</td>
<td></td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.059[0.033,0.083]</td>
<td>0.050[0.033,0.066]</td>
<td>0.084[0.047,0.123]</td>
<td></td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.947[0.934,0.960]</td>
<td>0.929[0.889,0.952]</td>
<td>0.927[0.905,0.946]</td>
<td>0.922[0.890,0.949]</td>
</tr>
<tr>
<td>$d_{22}$</td>
<td>0.017[0.003,0.030]</td>
<td>0.006[-0.006,0.016]</td>
<td>-0.003[-0.015,0.009]</td>
<td>0.025[0.015,0.035]</td>
</tr>
<tr>
<td>$e_{22}$</td>
<td>-0.028[-0.037,-0.019]</td>
<td>-0.018[-0.039,0.007]</td>
<td>-0.024[-0.038,-0.010]</td>
<td>-0.030[-0.044,-0.016]</td>
</tr>
<tr>
<td>$c_{22}$</td>
<td>0.046[0.009,0.084]</td>
<td>0.038[0.008,0.069]</td>
<td>0.030[-0.002,0.061]</td>
<td></td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.035[-0.020,0.090]</td>
<td>0.032[-0.018,0.085]</td>
<td>0.018[-0.032,0.069]</td>
<td></td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>0.776[0.737,0.816]</td>
<td>0.626[0.442,0.776]</td>
<td>0.694[0.592,0.787]</td>
<td>0.715[0.691,0.738]</td>
</tr>
<tr>
<td>$d_{21}$</td>
<td>0.055[0.043,0.067]</td>
<td>0.082[0.051,0.117]</td>
<td>0.043[0.019,0.069]</td>
<td>0.032[0.015,0.049]</td>
</tr>
<tr>
<td>$e_{21}$</td>
<td>-0.028[-0.037,-0.019]</td>
<td>-0.018[-0.039,0.007]</td>
<td>-0.024[-0.038,-0.010]</td>
<td>-0.030[-0.044,-0.016]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0.013[0.005,0.021]</td>
<td>0.005[0.002,0.008]</td>
<td>0.023[0.014,0.032]</td>
<td>0.004[-0.020,0.028]</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.005[-0.016,0.027]</td>
<td>0.005[-0.007,0.017]</td>
<td>0.007[-0.022,0.036]</td>
<td>0.032[-0.020,0.084]</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>0.718[0.570,0.879]</td>
<td>0.898[0.842,0.954]</td>
<td>0.502[0.408,0.602]</td>
<td>0.650[0.580,0.724]</td>
</tr>
<tr>
<td>$\gamma_{12,1}$</td>
<td>0.002[-0.038,0.042]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{12,2}$</td>
<td>0.009[-0.032,0.050]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{12,3}$</td>
<td>0.003[-0.033,0.040]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{12,1}$</td>
<td>0.012[-0.059,0.085]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{12,2}$</td>
<td>-0.067[-0.152,0.021]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{12,3}$</td>
<td>-0.068[-0.148,0.014]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.15: Acceptance rates for the APVECH family of models fitted to the S&P 500/All Ordinaries bivariate returns

<table>
<thead>
<tr>
<th>Step</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>97.4%</td>
<td>81.1%</td>
<td>79.8%</td>
<td>82.8%</td>
</tr>
<tr>
<td>2</td>
<td>14.6%</td>
<td>65.9%</td>
<td>59.5%</td>
<td>68.1%</td>
</tr>
<tr>
<td>3</td>
<td>37.6%</td>
<td>74.1%</td>
<td>80.5%</td>
<td>70.2%</td>
</tr>
<tr>
<td>4</td>
<td>80.7%</td>
<td>81.7%</td>
<td>80.7%</td>
<td>88.7%</td>
</tr>
<tr>
<td>5</td>
<td>93.0%</td>
<td>81.4%</td>
<td>86.1%</td>
<td>83.2%</td>
</tr>
</tbody>
</table>
Table 4.16: Parameter posterior means and 95% credible intervals for the APVECH family of models fitted to the S&P 500/Nikkei 225 bivariate returns

<table>
<thead>
<tr>
<th>Par</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>m_{10}</td>
<td>0.068[0.048,0.088]</td>
<td>0.047[0.024,0.070]</td>
<td>0.050[0.031,0.071]</td>
<td>0.048[0.028,0.079]</td>
</tr>
<tr>
<td>m_{20}</td>
<td>0.068[0.043,0.093]</td>
<td>0.052[0.024,0.079]</td>
<td>0.054[0.028,0.079]</td>
<td>0.058[0.032,0.083]</td>
</tr>
<tr>
<td>m_{11}</td>
<td>-0.021[-0.048,0.005]</td>
<td>-0.013[-0.042,0.016]</td>
<td>-0.017[-0.045,0.011]</td>
<td>-0.027[-0.055,0.001]</td>
</tr>
<tr>
<td>m_{22}</td>
<td>0.019[-0.008,0.045]</td>
<td>0.023[-0.005,0.051]</td>
<td>0.022[-0.004,0.049]</td>
<td>0.023[-0.004,0.049]</td>
</tr>
<tr>
<td>c_{11}</td>
<td>0.014[0.009,0.020]</td>
<td>0.010[-0.001,0.022]</td>
<td>0.014[0.005,0.022]</td>
<td>0.011[0.002,0.021]</td>
</tr>
<tr>
<td>a_{11}</td>
<td>0.061[0.048,0.074]</td>
<td>0.033[0.018,0.049]</td>
<td>0.029[0.014,0.043]</td>
<td>0.038[0.015,0.064]</td>
</tr>
<tr>
<td>b_{11}</td>
<td>0.927[0.910,0.942]</td>
<td>0.902[0.879,0.923]</td>
<td>0.918[0.899,0.937]</td>
<td>0.901[0.873,0.924]</td>
</tr>
<tr>
<td>d_{12}</td>
<td>0.002[-0.001,0.006]</td>
<td>0.002[-0.002,0.006]</td>
<td>0.001[-0.002,0.005]</td>
<td>0.001[-0.002,0.005]</td>
</tr>
<tr>
<td>e_{12}</td>
<td>-0.004[-0.008,0.000]</td>
<td>-0.004[-0.008,0.000]</td>
<td>-0.005[-0.009,0.000]</td>
<td>-0.005[-0.010,-0.001]</td>
</tr>
<tr>
<td>a_{11}</td>
<td>0.082[0.061,0.102]</td>
<td>0.066[0.038,0.095]</td>
<td>0.073[0.040,0.103]</td>
<td>0.073[0.040,0.103]</td>
</tr>
<tr>
<td>b_{12}</td>
<td>0.927[0.910,0.942]</td>
<td>0.902[0.879,0.923]</td>
<td>0.918[0.899,0.937]</td>
<td>0.901[0.873,0.924]</td>
</tr>
<tr>
<td>c_{22}</td>
<td>0.021[0.014,0.028]</td>
<td>0.015[0.001,0.029]</td>
<td>0.019[0.003,0.035]</td>
<td>0.021[0.007,0.036]</td>
</tr>
<tr>
<td>a_{22}</td>
<td>0.061[0.048,0.074]</td>
<td>0.033[0.018,0.049]</td>
<td>0.029[0.014,0.043]</td>
<td>0.038[0.015,0.064]</td>
</tr>
<tr>
<td>b_{22}</td>
<td>0.927[0.910,0.942]</td>
<td>0.902[0.879,0.923]</td>
<td>0.918[0.899,0.937]</td>
<td>0.901[0.873,0.924]</td>
</tr>
<tr>
<td>d_{21}</td>
<td>0.002[-0.001,0.006]</td>
<td>0.002[-0.002,0.006]</td>
<td>0.001[-0.002,0.005]</td>
<td>0.001[-0.002,0.005]</td>
</tr>
<tr>
<td>e_{21}</td>
<td>-0.004[-0.008,0.000]</td>
<td>-0.004[-0.008,0.000]</td>
<td>-0.005[-0.009,0.000]</td>
<td>-0.005[-0.010,-0.001]</td>
</tr>
<tr>
<td>a_{12}</td>
<td>0.082[0.061,0.102]</td>
<td>0.066[0.038,0.095]</td>
<td>0.073[0.040,0.103]</td>
<td>0.073[0.040,0.103]</td>
</tr>
<tr>
<td>b_{12}</td>
<td>0.927[0.910,0.942]</td>
<td>0.902[0.879,0.923]</td>
<td>0.918[0.899,0.937]</td>
<td>0.901[0.873,0.924]</td>
</tr>
<tr>
<td>c_{12}</td>
<td>0.012[0.001,0.032]</td>
<td>0.011[0.004,0.019]</td>
<td>0.004[0.001,0.014]</td>
<td>0.004[0.001,0.014]</td>
</tr>
<tr>
<td>a_{12}</td>
<td>0.020[0.007,0.040]</td>
<td>0.023[0.007,0.041]</td>
<td>0.016[0.005,0.029]</td>
<td>0.020[0.011,0.055]</td>
</tr>
<tr>
<td>b_{12}</td>
<td>0.838[0.605,0.975]</td>
<td>0.842[0.758,0.927]</td>
<td>0.829[0.857,0.978]</td>
<td>0.852[0.773,0.916]</td>
</tr>
<tr>
<td>γ_{11}</td>
<td>-0.002[-0.033,0.029]</td>
<td>-0.011[-0.045,0.023]</td>
<td>-0.004[-0.036,0.030]</td>
<td>-0.006[-0.031,0.029]</td>
</tr>
<tr>
<td>γ_{12}</td>
<td>0.014[-0.021,0.050]</td>
<td>0.014[-0.021,0.050]</td>
<td>0.013[-0.016,0.041]</td>
<td>0.013[-0.016,0.041]</td>
</tr>
<tr>
<td>γ_{21}</td>
<td>0.103[0.076,0.130]</td>
<td>0.105[0.075,0.136]</td>
<td>0.099[0.068,0.130]</td>
<td>0.099[0.068,0.130]</td>
</tr>
<tr>
<td>γ_{22}</td>
<td>0.030[0.011,0.049]</td>
<td>0.030[0.011,0.049]</td>
<td>0.040[0.018,0.062]</td>
<td>0.040[0.018,0.062]</td>
</tr>
</tbody>
</table>

Table 4.17: Acceptance rates for the APVECH family of models fitted to the S&P 500/Nikkei 225 bivariate returns

<table>
<thead>
<tr>
<th>Step</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>97.1%</td>
<td>82.5%</td>
<td>84.4%</td>
<td>83.6%</td>
</tr>
<tr>
<td>2</td>
<td>31.7%</td>
<td>71.3%</td>
<td>70.6%</td>
<td>70.5%</td>
</tr>
<tr>
<td>3</td>
<td>46.0%</td>
<td>69.8%</td>
<td>67.3%</td>
<td>69.0%</td>
</tr>
<tr>
<td>4</td>
<td>52.7%</td>
<td>78.4%</td>
<td>75.1%</td>
<td>88.5%</td>
</tr>
<tr>
<td>5</td>
<td>92.3%</td>
<td>83.1%</td>
<td>83.2%</td>
<td>83.4%</td>
</tr>
</tbody>
</table>
Table 4.18: Parameter posterior means and 95% credible intervals for the APVECH family of models fitted to the FTSE 100/All Ordinaries bivariate returns

<table>
<thead>
<tr>
<th>Par</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{10}$</td>
<td>0.067[0.046,0.087]</td>
<td>0.052[0.032,0.073]</td>
<td>0.049[0.029,0.071]</td>
<td>0.048[0.027,0.069]</td>
</tr>
<tr>
<td>$m_{20}$</td>
<td>0.062[0.044,0.079]</td>
<td>0.056[0.038,0.074]</td>
<td>0.059[0.041,0.077]</td>
<td>0.059[0.042,0.075]</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>-0.042[-0.069,-0.016]</td>
<td>-0.041[-0.068,-0.014]</td>
<td>-0.036[-0.063,-0.010]</td>
<td>-0.044[-0.071,-0.018]</td>
</tr>
<tr>
<td>$m_{22}$</td>
<td>0.099[0.072,0.126]</td>
<td>0.101[0.074,0.128]</td>
<td>0.103[0.076,0.130]</td>
<td>0.107[0.081,0.134]</td>
</tr>
<tr>
<td>$c_{11}$</td>
<td>0.021[0.015,0.026]</td>
<td>0.016[0.001,0.030]</td>
<td>0.019[0.007,0.033]</td>
<td>0.021[0.008,0.033]</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.067[0.054,0.082]</td>
<td>0.043[0.026,0.061]</td>
<td>0.039[0.024,0.058]</td>
<td>0.039[0.027,0.050]</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.923[0.905,0.939]</td>
<td>0.914[0.887,0.939]</td>
<td>0.896[0.835,0.934]</td>
<td>0.896[0.802,0.937]</td>
</tr>
<tr>
<td>$d_{12}$</td>
<td>0.022[0.014,0.030]</td>
<td>0.007[-0.010,0.024]</td>
<td>0.004[-0.014,0.021]</td>
<td>0.004[-0.013,0.021]</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>-0.039[-0.050,-0.029]</td>
<td>-0.011[-0.046,0.029]</td>
<td>-0.009[-0.037,0.021]</td>
<td>-0.036[-0.070,-0.002]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0.048[0.029,0.067]</td>
<td>0.059[0.037,0.080]</td>
<td>0.056[0.038,0.074]</td>
<td>0.056[0.038,0.074]</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.084[0.065,0.104]</td>
<td>0.073[0.045,0.102]</td>
<td>0.078[0.049,0.101]</td>
<td>0.073[0.042,0.106]</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>0.847[0.810,0.880]</td>
<td>0.826[0.762,0.882]</td>
<td>0.825[0.771,0.876]</td>
<td>0.811[0.743,0.869]</td>
</tr>
<tr>
<td>$d_{21}$</td>
<td>0.035[0.025,0.046]</td>
<td>0.035[0.023,0.048]</td>
<td>0.024[0.006,0.042]</td>
<td>0.014[0.000,0.029]</td>
</tr>
<tr>
<td>$e_{21}$</td>
<td>-0.026[-0.036,-0.016]</td>
<td>-0.021[-0.037,-0.004]</td>
<td>-0.026[-0.037,-0.014]</td>
<td>-0.017[-0.033,0.000]</td>
</tr>
<tr>
<td>$c_{22}$</td>
<td>0.028[0.004,0.061]</td>
<td>0.016[-0.021,0.054]</td>
<td>0.016[0.008,0.068]</td>
<td>0.014[0.008,0.066]</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.008[0.005,0.012]</td>
<td>0.008[0.005,0.013]</td>
<td>0.009[0.005,0.013]</td>
<td>-0.001[-0.014,0.013]</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>0.031[0.019,0.044]</td>
<td>0.031[0.019,0.044]</td>
<td>0.030[0.017,0.043]</td>
<td>0.006[0.025,0.038]</td>
</tr>
<tr>
<td>$b_{21}$</td>
<td>0.906[0.870,0.934]</td>
<td>0.905[0.862,0.939]</td>
<td>0.904[0.862,0.938]</td>
<td>0.852[0.791,0.898]</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>0.005[0.019,0.031]</td>
<td>0.009[0.013,0.032]</td>
<td>0.022[0.002,0.046]</td>
<td>0.083[0.038,0.126]</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.004[0.061,0.067]</td>
<td>0.001[-0.019,0.021]</td>
<td>0.003[-0.019,0.021]</td>
<td>-0.035[-0.093,0.020]</td>
</tr>
<tr>
<td>$b_{21}$</td>
<td>0.030[0.009,0.052]</td>
<td>0.030[0.011,0.061]</td>
<td>0.030[0.013,0.061]</td>
<td>0.041[0.021,0.061]</td>
</tr>
</tbody>
</table>

Table 4.19: Acceptance rates for the APVECH family of models fitted to the FTSE 100/All Ordinaries bivariate returns

<table>
<thead>
<tr>
<th>Step</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>97.5%</td>
<td>88.2%</td>
<td>88.1%</td>
<td>84.2%</td>
</tr>
<tr>
<td>2</td>
<td>77.0%</td>
<td>68.2%</td>
<td>71.0%</td>
<td>62.9%</td>
</tr>
<tr>
<td>3</td>
<td>77.2%</td>
<td>71.2%</td>
<td>73.1%</td>
<td>68.4%</td>
</tr>
<tr>
<td>4</td>
<td>90.7%</td>
<td>75.9%</td>
<td>77.8%</td>
<td>76.8%</td>
</tr>
<tr>
<td>5</td>
<td>30.6%</td>
<td>83.3%</td>
<td>85.2%</td>
<td>81.3%</td>
</tr>
</tbody>
</table>
Table 4.20: Parameter posterior means and 95% credible intervals for the APVECH family of models fitted to the FTSE 100/Nikkei 225 bivariate returns

<table>
<thead>
<tr>
<th>Par</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m_{10})</td>
<td>0.070 [0.049, 0.091]</td>
<td>0.053 [0.033, 0.073]</td>
<td>0.052 [0.029, 0.075]</td>
<td>0.046 [0.024, 0.067]</td>
</tr>
<tr>
<td>(m_{20})</td>
<td>0.064 [0.038, 0.090]</td>
<td>0.043 [0.017, 0.070]</td>
<td>0.047 [0.020, 0.073]</td>
<td>0.044 [0.017, 0.070]</td>
</tr>
<tr>
<td>(m_{11})</td>
<td>-0.012 [-0.039, 0.016]</td>
<td>-0.008 [-0.036, 0.018]</td>
<td>-0.007 [-0.034, 0.021]</td>
<td>-0.009 [-0.037, 0.019]</td>
</tr>
<tr>
<td>(m_{22})</td>
<td>0.026 [-0.001, 0.052]</td>
<td>0.030 [0.003, 0.056]</td>
<td>0.030 [0.003, 0.056]</td>
<td>0.033 [0.007, 0.059]</td>
</tr>
<tr>
<td>(c_{11})</td>
<td>0.023 [0.015, 0.031]</td>
<td>0.023 [0.009, 0.037]</td>
<td>0.030 [0.018, 0.042]</td>
<td>0.041 [0.024, 0.059]</td>
</tr>
<tr>
<td>(a_{11})</td>
<td>0.086 [0.070, 0.103]</td>
<td>0.058 [0.034, 0.085]</td>
<td>0.052 [0.027, 0.079]</td>
<td>0.054 [0.035, 0.074]</td>
</tr>
<tr>
<td>(b_{11})</td>
<td>0.893 [0.872, 0.913]</td>
<td>0.888 [0.852, 0.919]</td>
<td>0.890 [0.857, 0.920]</td>
<td>0.867 [0.834, 0.897]</td>
</tr>
<tr>
<td>(d_{12})</td>
<td>0.002 [-0.002, 0.007]</td>
<td>0.002 [-0.002, 0.006]</td>
<td>0.003 [-0.003, 0.008]</td>
<td>0.001 [-0.004, 0.006]</td>
</tr>
<tr>
<td>(e_{12})</td>
<td>-0.003 [-0.008, 0.002]</td>
<td>-0.004 [-0.010, 0.003]</td>
<td>-0.005 [-0.012, 0.002]</td>
<td>-0.006 [-0.016, 0.005]</td>
</tr>
<tr>
<td>(\gamma_{11})</td>
<td>0.024 [-0.007, 0.073]</td>
<td>0.012 [-0.034, 0.085]</td>
<td>0.016 [-0.039, 0.119]</td>
<td></td>
</tr>
<tr>
<td>(\alpha_{11})</td>
<td>0.056 [0.029, 0.082]</td>
<td>0.067 [0.039, 0.095]</td>
<td>0.094 [0.054, 0.135]</td>
<td></td>
</tr>
<tr>
<td>(\delta_{12})</td>
<td>0.002 [-0.007, 0.011]</td>
<td>0.010 [-0.001, 0.022]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c_{22})</td>
<td>0.026 [0.017, 0.036]</td>
<td>0.019 [0.004, 0.033]</td>
<td>0.023 [0.013, 0.033]</td>
<td>0.027 [0.015, 0.039]</td>
</tr>
<tr>
<td>(a_{22})</td>
<td>0.081 [0.066, 0.096]</td>
<td>0.047 [0.032, 0.064]</td>
<td>0.046 [0.032, 0.061]</td>
<td>0.045 [0.026, 0.064]</td>
</tr>
<tr>
<td>(b_{22})</td>
<td>0.905 [0.887, 0.922]</td>
<td>0.890 [0.867, 0.910]</td>
<td>0.890 [0.862, 0.914]</td>
<td>0.876 [0.843, 0.905]</td>
</tr>
<tr>
<td>(d_{21})</td>
<td>0.043 [0.026, 0.059]</td>
<td>0.040 [0.018, 0.062]</td>
<td>0.021 [-0.006, 0.046]</td>
<td>0.030 [0.014, 0.046]</td>
</tr>
<tr>
<td>(e_{21})</td>
<td>-0.045 [-0.066, -0.025]</td>
<td>-0.039 [-0.063, -0.016]</td>
<td>-0.031 [-0.057, -0.003]</td>
<td>-0.049 [-0.072, -0.026]</td>
</tr>
<tr>
<td>(\gamma_{22})</td>
<td>0.035 [0.016, 0.054]</td>
<td>0.018 [-0.001, 0.039]</td>
<td>0.029 [0.003, 0.056]</td>
<td></td>
</tr>
<tr>
<td>(\alpha_{22})</td>
<td>0.084 [0.058, 0.110]</td>
<td>0.086 [0.059, 0.113]</td>
<td>0.115 [0.080, 0.152]</td>
<td></td>
</tr>
<tr>
<td>(\delta_{21})</td>
<td>0.027 [0.000, 0.054]</td>
<td>0.041 [0.011, 0.071]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c_{12})</td>
<td>0.008 [0.004, 0.012]</td>
<td>0.010 [0.005, 0.016]</td>
<td>0.009 [0.004, 0.015]</td>
<td>0.006 [-0.008, 0.021]</td>
</tr>
<tr>
<td>(a_{12})</td>
<td>0.034 [0.023, 0.047]</td>
<td>0.037 [0.024, 0.050]</td>
<td>0.036 [0.023, 0.050]</td>
<td>0.019 [-0.008, 0.046]</td>
</tr>
<tr>
<td>(b_{12})</td>
<td>0.924 [0.893, 0.947]</td>
<td>0.911 [0.868, 0.942]</td>
<td>0.913 [0.867, 0.944]</td>
<td>0.835 [0.766, 0.888]</td>
</tr>
<tr>
<td>(\gamma_{12,1})</td>
<td>-0.002 [-0.031, 0.025]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\gamma_{12,2})</td>
<td>-0.029 [-0.062, 0.005]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\gamma_{12,3})</td>
<td>0.035 [0.004, 0.066]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{12,1})</td>
<td>0.092 [0.050, 0.135]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{12,2})</td>
<td>-0.058 [-0.122, 0.005]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha_{12,3})</td>
<td>-0.045 [-0.102, 0.010]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.21: Acceptance rates for the APVECH family of models fitted to the FTSE 100/Nikkei 225 bivariate returns

<table>
<thead>
<tr>
<th>Step</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98.2%</td>
<td>87.6%</td>
<td>87.8%</td>
<td>84.8%</td>
</tr>
<tr>
<td>2</td>
<td>45.8%</td>
<td>70.6%</td>
<td>74.1%</td>
<td>72.0%</td>
</tr>
<tr>
<td>3</td>
<td>32.4%</td>
<td>66.4%</td>
<td>66.2%</td>
<td>76.0%</td>
</tr>
<tr>
<td>4</td>
<td>61.3%</td>
<td>71.8%</td>
<td>72.1%</td>
<td>78.7%</td>
</tr>
<tr>
<td>5</td>
<td>27.5%</td>
<td>85.8%</td>
<td>85.0%</td>
<td>82.6%</td>
</tr>
</tbody>
</table>
Table 4.22: Parameter posterior means and 95% credible intervals for the APVECH family of models fitted to the All Ordinaries/Nikkei 225 bivariate returns

<table>
<thead>
<tr>
<th>Par Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>m_{10}</td>
<td>0.068 [0.050, 0.086]</td>
<td>0.062 [0.044, 0.079]</td>
<td>0.064 [0.046, 0.082]</td>
<td>0.058 [0.040, 0.076]</td>
</tr>
<tr>
<td>m_{20}</td>
<td>0.075 [0.049, 0.101]</td>
<td>0.063 [0.036, 0.087]</td>
<td>0.064 [0.039, 0.089]</td>
<td>0.054 [0.029, 0.081]</td>
</tr>
<tr>
<td>m_{11}</td>
<td>0.060 [0.034, 0.086]</td>
<td>0.062 [0.037, 0.088]</td>
<td>0.064 [0.039, 0.089]</td>
<td>0.067 [0.041, 0.092]</td>
</tr>
<tr>
<td>m_{22}</td>
<td>0.001 [-0.024, 0.026]</td>
<td>0.004 [-0.021, 0.029]</td>
<td>0.005 [-0.019, 0.029]</td>
<td>0.006 [-0.020, 0.031]</td>
</tr>
<tr>
<td>c_{11}</td>
<td>0.015 [0.010, 0.022]</td>
<td>0.016 [0.008, 0.024]</td>
<td>0.018 [0.008, 0.029]</td>
<td>0.016 [0.007, 0.027]</td>
</tr>
<tr>
<td>a_{11}</td>
<td>0.060 [0.047, 0.074]</td>
<td>0.054 [0.033, 0.078]</td>
<td>0.057 [0.032, 0.087]</td>
<td>0.053 [0.030, 0.077]</td>
</tr>
<tr>
<td>b_{11}</td>
<td>0.921 [0.899, 0.939]</td>
<td>0.910 [0.880, 0.937]</td>
<td>0.904 [0.864, 0.937]</td>
<td>0.893 [0.863, 0.922]</td>
</tr>
<tr>
<td>d_{12}</td>
<td>0.004 [0.001, 0.007]</td>
<td>0.006 [0.001, 0.010]</td>
<td>0.006 [0.001, 0.011]</td>
<td>0.005 [0.001, 0.009]</td>
</tr>
<tr>
<td>e_{12}</td>
<td>-0.005 [-0.008, -0.002]</td>
<td>-0.007 [-0.012, -0.002]</td>
<td>-0.006 [-0.011, 0.001]</td>
<td>-0.007 [-0.011, -0.003]</td>
</tr>
<tr>
<td>γ_{11}</td>
<td>0.015 [-0.003, 0.033]</td>
<td>0.013 [-0.006, 0.036]</td>
<td>0.026 [0.002, 0.068]</td>
<td>-0.006 [-0.011, -0.003]</td>
</tr>
<tr>
<td>α_{11}</td>
<td>0.002 [0.007, 0.038]</td>
<td>0.018 [-0.002, 0.042]</td>
<td>0.053 [0.030, 0.077]</td>
<td>-0.002 [-0.010, 0.005]</td>
</tr>
<tr>
<td>δ_{12}</td>
<td>-0.002 [-0.010, 0.005]</td>
<td>0.023 [0.011, 0.035]</td>
<td>0.002 [0.004, 0.006]</td>
<td>0.024 [0.008, 0.039]</td>
</tr>
<tr>
<td>c_{22}</td>
<td>0.024 [0.017, 0.030]</td>
<td>0.019 [0.011, 0.028]</td>
<td>0.023 [0.011, 0.035]</td>
<td>0.024 [0.008, 0.039]</td>
</tr>
<tr>
<td>a_{22}</td>
<td>0.083 [0.071, 0.095]</td>
<td>0.046 [0.032, 0.061]</td>
<td>0.046 [0.031, 0.062]</td>
<td>0.040 [0.024, 0.057]</td>
</tr>
<tr>
<td>b_{22}</td>
<td>0.907 [0.894, 0.920]</td>
<td>0.903 [0.885, 0.919]</td>
<td>0.903 [0.887, 0.918]</td>
<td>0.891 [0.871, 0.910]</td>
</tr>
<tr>
<td>d_{21}</td>
<td>-0.001 [-0.007, 0.006]</td>
<td>-0.003 [-0.008, 0.003]</td>
<td>-0.002 [-0.016, 0.011]</td>
<td>-0.008 [-0.021, 0.004]</td>
</tr>
<tr>
<td>e_{21}</td>
<td>-0.005 [-0.012, 0.001]</td>
<td>-0.006 [-0.012, 0.000]</td>
<td>-0.004 [-0.011, 0.003]</td>
<td>-0.006 [-0.013, 0.002]</td>
</tr>
<tr>
<td>γ_{22}</td>
<td>0.026 [0.004, 0.050]</td>
<td>0.013 [-0.010, 0.035]</td>
<td>0.020 [0.013, 0.054]</td>
<td>0.020 [0.013, 0.054]</td>
</tr>
<tr>
<td>α_{22}</td>
<td>0.073 [0.053, 0.093]</td>
<td>0.075 [0.056, 0.094]</td>
<td>0.111 [0.083, 0.139]</td>
<td>-0.001 [-0.018, 0.016]</td>
</tr>
<tr>
<td>δ_{21}</td>
<td>-0.001 [-0.018, 0.016]</td>
<td>0.010 [-0.006, 0.028]</td>
<td>0.010 [-0.006, 0.028]</td>
<td>0.010 [-0.006, 0.028]</td>
</tr>
</tbody>
</table>

Table 4.23: Acceptance rates for the APVECH family of models fitted to the All Ordinaries/Nikkei 225 bivariate returns

<table>
<thead>
<tr>
<th>Step</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98.0%</td>
<td>89.9%</td>
<td>91.1%</td>
<td>86.9%</td>
</tr>
<tr>
<td>2</td>
<td>25.0%</td>
<td>53.1%</td>
<td>47.6%</td>
<td>69.5%</td>
</tr>
<tr>
<td>3</td>
<td>40.9%</td>
<td>43.9%</td>
<td>42.5%</td>
<td>56.7%</td>
</tr>
<tr>
<td>4</td>
<td>47.4%</td>
<td>57.0%</td>
<td>53.7%</td>
<td>61.0%</td>
</tr>
<tr>
<td>5</td>
<td>21.6%</td>
<td>84.7%</td>
<td>82.5%</td>
<td>81.3%</td>
</tr>
</tbody>
</table>
Table 4.24: Potential scale reduction factors for Model 4 of the APVECH family of models fitted to the All Ordinaries/Nikkei 225 bivariate returns

<table>
<thead>
<tr>
<th>Par</th>
<th>PSRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{10}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$m_{20}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$m_{22}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$c_{11}$</td>
<td>1.06</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>1.11</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>1.10</td>
</tr>
<tr>
<td>$d_{12}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>1.05</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>1.19</td>
</tr>
<tr>
<td>$\delta_{12}$</td>
<td>1.03</td>
</tr>
<tr>
<td>$c_{22}$</td>
<td>1.05</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>1.06</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$d_{21}$</td>
<td>1.10</td>
</tr>
<tr>
<td>$e_{21}$</td>
<td>1.06</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>1.10</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>1.23</td>
</tr>
<tr>
<td>$\delta_{21}$</td>
<td>1.12</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>1.03</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>1.07</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>1.08</td>
</tr>
<tr>
<td>$\gamma_{12,1}$</td>
<td>1.08</td>
</tr>
<tr>
<td>$\gamma_{12,2}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$\gamma_{12,3}$</td>
<td>1.05</td>
</tr>
<tr>
<td>$\alpha_{12,1}$</td>
<td>1.13</td>
</tr>
<tr>
<td>$\alpha_{12,2}$</td>
<td>1.05</td>
</tr>
<tr>
<td>$\alpha_{12,3}$</td>
<td>1.03</td>
</tr>
<tr>
<td>$\nu$</td>
<td>1.03</td>
</tr>
</tbody>
</table>
Figure 4.2: MCMC iterates for the parameters $\gamma_{11}$, $b_{22}$, $\alpha_{12,3}$ and $\nu$ for Model 4 of the APVECH family of models fitted to the All Ordinaries/Nikkei 225 bivariate returns, each with 5 different starting values.
Chapter 5

Estimating Portfolio Value at Risk using a Skew-$t$ Copula-GARCH Model

The goal of this chapter is to improve existing techniques in risk forecasting for stock portfolios. To assist in reaching this goal, the vast properties of multivariate stock returns need to be accurately captured. A copula-GARCH model is proposed in this chapter that simultaneously captures stock return asymmetry in both conditional mean and variance equations of its marginal distributions, as well as asymmetry in its dependence structure via a skew-$t$ copula. In addition, the marginal distributions are generalised to incorporate both skewness and excess kurtosis, thus enabling information on the first four moments of stock returns to be available in the one model. The performance of the proposed model is compared to a variety of existing univariate and multivariate GARCH models by estimating Value at Risk (VaR) for multiple time horizons on a stock portfolio consisting of five Dow Jones Industrial (DJI) stocks. While VaR back-testing has already been applied extensively in the literature, it has always been applied to compare models of the same dimension. Under the assumption that the portfolio weights are known, comparisons can be drawn between univariate and multivariate models together in the one study. In
this example, the proposed skew-$t$ copula-GARCH model outperforms the competing
models in many cases, and also demonstrates significant asymmetries that are present in
the data. These results clearly indicate the need to accurately capture the tail behaviour
in stock returns for risk management purposes.

5.1 Introduction

Holding a portfolio of stocks attracts many risks. A substantial focus of such risks in the
literature has been on market risk, i.e. the risk of a financial loss due to unfavourable
movements in stock prices. In order to quantify the amount of market risk in a stock
portfolio at a point in time, the common approach is to use Value at Risk (VaR). Established at J.P. Morgan in 1993 as part of the “Weatherstone 4:15pm” daily risk assessment
report, VaR is perhaps the most widely used measure of market risk. It is one number
that represents the minimum amount that could be lost over a given time horizon at
a specified confidence level. Knowledge of VaR enables institutions to take action on
“risky” portfolio positions that could potentially result in huge financial losses. There
have been many investigations of VaR in the literature, some examples can be found in

VaR can be estimated from holding either one stock (using one return series), or a
portfolio of stocks (using multiple return series). For one stock, VaR can be estimated
from its return series \( \{y_t^i\}_{t=1}^n \) using volatility forecasts from a univariate GARCH model
(see for e.g. Giot and Laurent, 2003; Wong and So, 2003 and So and Yu, 2006). When a
portfolio of stocks is held, portfolio VaR can be estimated from the multivariate return se-
ries \( \{(y_{1t}, \ldots, y_{Nt})\}_{t=1}^n \) using both volatility and correlation forecasts from a multivariate
GARCH model, plus the portfolio weights \( \{w_i\}_{i=1}^N \) (see for e.g. Palaro and Hotta, 2006;
McAleer and Da Veiga, 2008 and Rombouts and Verbeek, 2009). However, when the
portfolio weights are known, they can be combined with the multivariate return series to
give portfolio returns via \( \{w_1y_{1t} + \ldots + w_Ny_{Nt}\}_{t=1}^n \) which can be estimated via more par-
simonious and easier to estimate univariate GARCH models. Hence, under this setting, the performance of univariate and multivariate GARCH models can be compared when estimating VaR for a stock portfolio. There is evidence in the literature that simpler models forecast volatility better than more complex models (see for e.g. Dimson and Marsh, 1990 and Cheong, 2009). In addition, fitting a univariate GARCH model to portfolio returns may implicitly capture the multivariate dependence structure more accurately than a poorly specified multivariate GARCH model. This theory is tested in this chapter by comparing a series of univariate and multivariate GARCH models in estimating VaR for multiple time horizons on a Dow Jones Industrial (DJI) stock portfolio.

In order to test the adequacy of a model in VaR estimation, several back-testing procedures have been proposed in the literature. Perhaps the earliest one is that recommended in the revised Basel Capital Accord (Basel II), which recommends examining the number of instances that VaR is violated (i.e. actual portfolio losses exceeding the estimated VaR) in history. Kupiec (1995) formalises this as the unconditional coverage (UC) test, where the number of observed VaR violations is compared to the significance level used in estimating each VaR. The percentage of VaR violations should be close to the significance level for a “reasonable” VaR model. However, this back-test does not take into account the timing of each VaR violation; a VaR model should have violations close to the significance level and each violation should occur at independent times over the back-testing period. Christoffersen (1998) combines the UC test together with a simple test of independence between VaR violations known as the conditional coverage (CC) test. However, dependence between violations can only be identified in this test when they occur consecutively (i.e. one violation after the other). Engle and Manganelli (2004) generalise this concept through the dynamic quantile (DQ) test, which enables independence testing for potentially non-consecutive violations, and thus is a more powerful test (see for e.g Chen et al., 2011b and Berkowitz et al., 2011). A non-binary test that can be used to back-test VaR is the criterion function of Koenker and Bassett (1978), which calculates differences between actual returns and the estimated VaR. The model which
minimises this function is the model preferred\(^1\). These VaR back-testing procedures will be utilised in this chapter in order to compare model performance in estimating VaR.

Accurate volatility and correlation forecasting requires a well-specified multivariate GARCH model. Ever since the introduction of the VECH multivariate GARCH model by Bollerslev et al. (1988), there has been an extensive amount of research appearing in the literature on improving the way conditional correlation is modelled in multivariate financial time series. Much of this research deals with simplifying estimation of model parameters, since typically many restrictions are required in order to achieve positive definiteness (PD) and covariance stationarity (CS) of the dynamic covariance matrices, as Chapters 3 and 4 of this thesis illustrate (e.g. the models of Engle et al., 1990; Bollerslev, 1990; Engle and Kroner, 1995 and Kroner and Ng, 1998). Chapters 3 and 4 also illustrate the Dynamic Conditional Correlation (DCC) model of Engle (2002) and the Asymmetric Generalised DCC (AG-DCC) model of Cappiello et al. (2006), the advantages of which are the ability to estimate each individual dataset through univariate GARCH models, then use this information to estimate the time-varying correlation component of the model without the need for a large number of restrictions. These models represent examples of estimation by separating a model’s marginal distributions from the dependence structure.

This is precisely the methodology behind *copulas*. A copula is a multivariate cumulative distribution function (CDF) that has marginals following uniform distributions on the \([0, 1]\) interval. The copula itself can take on many forms, which is usually determined by the dependence structure needing to be captured. The idea itself is not new, dating back (at least) to the results of Sklar (1959). More recent texts such as Joe (1997), Nelsen (1999) and Cherubini et al. (2004) provide excellent theory and applications behind different copula structures. Papers such as Patton (2004), Jondeau and Rockinger (2006), Serban et al. (2007), Lai et al. (2009) and Lee and Long (2009) are examples of copula-GARCH applications to financial time series. Perhaps the most common types of copulas implemented in the literature are elliptical copulas (e.g. the Gaussian and

---

\(^1\)This process is mathematically equivalent to maximising the likelihood function under an asymmetric Laplace distribution (see for e.g. Koenker and Machado, 1999 and Yu and Moyeed, 2001).
Student-$t$ copulas), which are based on their respective multivariate distributions. These are popular in financial applications because the dependence structure of these copulas is defined in terms of correlations, which are easily interpreted. However, they do not have the ability to capture asymmetric dependence since they are based on symmetric multivariate distributions. Archimedean copulas (e.g. Clayton, Frank and Gumbel copulas) which have closed-form CDF’s can offer an asymmetric dependence alternative to the Gaussian and Student-$t$ forms. Another type of copula that has recently developed is the skew-$t$ copula, with various forms existing due to different forms of the multivariate skew-$t$ distributions appearing in the literature (see for e.g Jones, 2001, Azzalini and Capitanio, 2003 and Sahu et al., 2003). Chan and Kroese (2010) provide an application of a skew-$t$ copula to portfolio credit risk, while Smith et al. (2011) propose a new skew-$t$ copula and use Bayesian techniques to estimate it. Sun et al. (2008) combine a skew-$t$ copula with an ARMA-GARCH model.

The majority of papers on copula-GARCH models in the literature attempt to capture the asymmetry, skewness and excess kurtosis from return series in their models in various ways, but not in all components of their models. For example, Jondeau and Rockinger (2006) have univariate skew-$t$ distributions for each marginal, but do not allow for asymmetric dependence in their copulas; Serban et al. (2007) offer a high-dimensional copula model but do not consider asymmetry or skewness at all, and Lai et al. (2009) consider copulas with asymmetric dependence but do not consider skewness in their marginals. Using the multivariate skew-$t$ distribution of Bauwens and Laurent (2005), a new skew-$t$ copula-GARCH model is proposed in this chapter which simultaneously allows for asymmetry, skewness and excess kurtosis in the marginal distributions, and (potentially) asymmetric dependence via a skew-$t$ copula, thus extending existing approaches. This model is compared to other existing univariate and multivariate GARCH models in estimating VaR, as discussed earlier.

The common ground between the copula-GARCH (and many other) papers in this area is that the majority use a two-step estimation procedure similar to that of Engle
marginal distributions are estimated first via maximum likelihood, then the dependence structure between them is estimated via maximising a copula likelihood based on information from the marginals. Patton (2006) shows that resulting parameter estimates are consistent and asymptotically normally distributed (under certain regularity conditions), but they are not efficient asymptotically as information in the marginals is lost from the first step. Ausin and Lopes (2010) propose a one-step Bayesian approach in estimating a bivariate copula-GARCH model, which alleviates this issue by considering the joint posterior distribution of all model parameters. While this approach demonstrates to work well in this paper, it becomes quite computationally expensive once the dimension of the model increases. Ausin and Lopes (2010) also suggest a two-step Bayesian approach that is similar to the classical two-step approach but the joint posterior distribution is split into marginal posteriors and a copula posterior for simulation. This enables faster Bayesian estimation (although not fully Bayesian) and can be much more convenient in large dimensions. This estimation approach is taken in this chapter to fit the proposed skew-t copula-GARCH model to the returns from a DJI portfolio containing five stocks.

This chapter is organised as follows: Section 5.2 briefly discusses the notation of distributions to be used in this chapter. Section 5.3 reviews some multivariate skew-t distributions that exist in the literature, while Section 5.4 reviews some influential copula-GARCH models existing in the literature. Section 5.5 introduces the proposed skew-t copula-GARCH model and discusses its properties. Section 5.6 discusses the Bayesian estimation procedures to be applied to the model. Section 5.7 applies the model in estimating VaR for a 5-dimensional stock portfolio, with its performance compared with a series of existing univariate and multivariate GARCH models. Section 5.8 concludes this chapter.
5.2 Distributional Notation

It is at this point that I mention distributional notation for the remainder of this chapter:

- Unless otherwise stated, all multivariate distributions will assume the general dimension $N$. This has been the notation throughout this thesis and hence will continue. For example, if $f$ is a multivariate PDF, then $f : \mathbb{R}^N \to \mathbb{R}^+$, where $\mathbb{R}^+$ denotes the set of positive real numbers.

- Multivariate PDF's and CDF's will be displayed as **boldface** to distinguish between univariate PDF's and CDF's. In addition, CDF’s will be given the upper-case notation, with a “$-1$” superscript included for inverse CDF’s. Examples of this are given in the table below:

<table>
<thead>
<tr>
<th>Notation</th>
<th>Distribution Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>Univariate PDF</td>
</tr>
<tr>
<td>$f$</td>
<td>Multivariate PDF</td>
</tr>
<tr>
<td>$F$</td>
<td>Univariate CDF</td>
</tr>
<tr>
<td>$F$</td>
<td>Multivariate CDF</td>
</tr>
<tr>
<td>$F^{-1}$</td>
<td>Univariate Inverse CDF</td>
</tr>
</tbody>
</table>

- Any parameters representing skewness or excess kurtosis will be placed as subscripts with the distribution, with mean (vectors) and (co)variances placed in brackets. For any standardised distributions, the parameters in the brackets will simply be omitted. For example, $f_{\nu}(x|\mu, \Sigma)$ is a multivariate PDF at $x$ with mean vector $\mu$, covariance matrix $\Sigma$ and degrees of freedom $\nu$, while $F_{\lambda,\nu}(x)$ is a univariate CDF at $x$ with zero mean, unit variance, skewness parameter $\lambda$ and degrees of freedom $\nu$.

Typically vectors have been given the boldface notation throughout this thesis - this will continue, however the distinction between multivariate distributions and vectors should be clear.
5.3 Multivariate Skew-$t$ Distributions

This section briefly reviews some of the multivariate skew-$t$ distributions that currently exist in the literature. See Kotz and Nadarajah (2004) for other generalisations and related distributions.

Assume that $X = (X_1, \ldots, X_N)$ is a vector of random variables, with $x = (x_1, \ldots, x_N)$ the observation vector. Unlike the unique multivariate normal and multivariate Student-$t$ distributions given by the PDF’s

$$n(x|\mu, \Sigma) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right]$$ (5.1)

and

$$t_\nu(x|\mu, \Sigma) = \frac{\Gamma \left( \frac{\nu + N}{2} \right)}{\Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{N}{2} \right)} \left[ 1 + \frac{1}{\nu} (x - \mu)' \Sigma^{-1} (x - \mu) \right]^{-\frac{\nu + N}{2}} ,$$ (5.2)

there are various ways to represent a multivariate skew-$t$ distribution:

5.3.1 Jones (2001)

The multivariate skew-$t$ distribution of Jones (2001) forms as an association with the multivariate $F$ distribution and beta distributions, the properties of which are detailed in this paper. The PDF is given by

$$skt_\nu(x) = \frac{\Gamma(s)}{\Gamma(\nu_0) \cdots \Gamma(\nu_N)} \prod_{i=1}^{N} \left[ \frac{2 \left( x_i + \sqrt{w_i + x_i^2} \right)^{2\nu_i}}{w_i \sqrt{w_i + x_i^2}} \right] ^{\frac{s}{2}} \left[ 1 + \sum_{j=1}^{N} \left( \frac{x_j + \sqrt{w_j + x_j^2}}{w_j} \right)^{2\nu_j} \right]^{-\frac{s + \nu_0 + \cdots + \nu_N}{2}},$$ (5.3)

where $\nu = (\nu_0, \ldots, \nu_N)$ are the $N+1$ degrees of freedom parameters, $s = \nu_0 + \cdots + \nu_N$ and $w_i = \nu_0 + \nu_i$, $i = 1, \ldots, N$. The skewness in this distribution is controlled by the values of each $w_i$, therefore no unique skewness parameters exist in equation (5.3). Hence this is a parsimonious distribution when fitted, but skewness is difficult to interpret.
5.3.2 Azzalini and Capitanio (2003)

The multivariate skew-t distribution of Azzalini and Capitanio (2003) offers a direct extension to the multivariate skew-normal distribution studied in Azzalini and Dalla Valle (1996). Its PDF is given by

\[ skt_{\lambda,\nu}(x|\mu, \Sigma) = 2 t_\nu(x|\mu, \Sigma) T_{\nu+N} \left( \lambda' D^{-1}(x - \mu) \sqrt{\frac{\nu + N}{Q_x + \nu}} \right) \]  

(5.4)

where \( t_\nu \) denotes the multivariate Student-t PDF (as in equation (5.2)) with mean vector \( \mu \), covariance matrix \( \Sigma = [\sigma_{ij}] \) and degrees of freedom \( \nu \), while \( T_{\nu+N} \) denotes the univariate Student-t CDF with degrees of freedom \( \nu + N \). The additional parameters of this distribution are contained in the parameter vector \( \lambda \), with \( Q_x = (x - \mu)'\Sigma^{-1}(x - \mu) \) and \( D \) is a diagonal matrix such that \( D = \text{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{NN}}) \). This is an improvement over the Jones (2001) distribution in (5.3), mainly because of the existence of the distinct “shape” parameters contained in \( \lambda \). However, should many calls to this function exist in a computer program (e.g. in a Bayesian analysis), then the Student-t CDF needs to be calculated each time which can be computationally expensive.

5.3.3 Sahu et al. (2003)

The multivariate skew-t distribution of Sahu et al. (2003) has a similar form to that of Azzalini and Capitanio (2003) in equation (5.4), and has the PDF

\[ skt_{\delta,\nu}(x|\mu, \Sigma) = 2^N t_\nu(x|\mu, \Sigma + D^2) \Pr(V > 0) \]  

(5.5)

where \( \mu \), \( \Sigma \) and \( \nu \) are defined as in equation (5.4), and \( D = \text{diag}(\delta) = \text{diag}(\delta_1, \ldots, \delta_N) \) is a diagonal matrix defining the skewness parameters. The random vector \( V \) follows a multivariate Student-t distribution with mean vector \( \mu_V \), covariance matrix \( \Sigma_V \) and degrees of freedom \( \nu + N \), where
\[ \mu_V = D(\Sigma + D^2)^{-1}(x - \mu), \quad \Sigma_V = \frac{\nu + Q_x}{\nu + N} [I_N - D(\Sigma + D^2)^{-1}D] \]

with \(Q_x = (x - \mu)'(\Sigma + D^2)(x - \mu)\) and \(I_N\) the \(N \times N\) identity matrix. Like the distribution of Azzalini and Capitanio (2003) in equation (5.4), unique parameters (i.e. \(\delta\)) represent the “shape” of the distribution. However, Smith et al. (2011) state that evaluation of the probability \(\Pr(V > 0)\) needs to be undertaken numerically, and further state that “this is still difficult in even moderate dimensions”. Hence this distribution would be impractical in many applications.

### 5.3.4 Demarta and McNeil (2005)

The multivariate skew-t distribution of Demarta and McNeil (2005) derives as a special case of the multivariate normal-mixture distribution. Its PDF is given by

\[
skt_{\gamma, \nu}(x|\mu, \Sigma) = c \frac{K_{\nu+\nu} \left( \sqrt{(\nu + Q_x)\gamma^\prime \Sigma^{-1} \gamma} \right) \exp \left[ (x - \mu)'\Sigma^{-1}(x - \mu) \right]}{\left( \sqrt{(\nu + Q_x)\gamma^\prime \Sigma^{-1} \gamma} \right)^{\nu + \nu} (1 + \frac{Q_x}{\nu})^{\nu+\nu}}
\]

where \(K(\cdot)\) denotes the modified Bessel function of the third kind (see Abramowitz and Stegun, 1972, ch 9,10), \(Q_x = (x - \mu)'\Sigma^{-1}(x - \mu)\) and \(c\) is a normalising constant equal to

\[
c = \frac{2^{2-\nu-N}}{\Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{\nu/2} |\Sigma|^{1/2}}.
\]

The same definition is given to the parameters \(\mu, \Sigma, \gamma\) and \(\nu\) as in the distribution of Azzalini and Capitanio (2003). This distribution is easy to implement, however Banachewicz and van der Vaart (2008) and Fung and Seneta (2010) illustrate that its tail dependence coefficient (i.e. the additional amount of weighting of the distribution in its tails compared to that of a multivariate normal distribution) has trivial values (0 or 1). In fact, Fung and Seneta (2010) state that this distribution “may not be an appropriate skew-encompassing extension of the multivariate symmetric t distribution, which itself possess nontrivial values of the tail dependence coefficient”.

---

147
5.3.5 Bauwens and Laurent (2005)

The multivariate skew-t distribution of Bauwens and Laurent (2005) acts as a natural extension to the univariate skew-

\[ t \]

distribution formed by the approach of Fernandez and Steel (1998): they generate a skewed distribution from a PDF \( f \) (which is unimodal and symmetric around zero), using the parameter \( \gamma \in (0, \infty) \) as follows:

\[
\text{sk}_{\gamma}(\varepsilon) = \frac{2\gamma}{1 + \gamma^2} \left[ f \left( \frac{\varepsilon}{\gamma} \right) I(\varepsilon \geq 0) + f(\gamma \varepsilon) I(\varepsilon < 0) \right]
\]

(5.7)

where \( I \) is the indicator equal to 1 if the condition is true and 0 otherwise. Under the assumption that \( f \) is the standardised Student-t distribution, i.e.

\[
f_{\nu}(x|\mu, \sigma^2) = \frac{\Gamma \left( \frac{\nu + 1}{2} \right)}{\Gamma \left( \frac{\nu}{2} \right) \sqrt{\nu - 2} \pi \sigma^2} \left[ 1 + \frac{(x - \mu)^2}{\nu - 2} \right]^{-\frac{\nu + 1}{2}},
\]

applying (5.7) to \( f \) results in the following skew-t distribution:

\[
\text{skt}_{\gamma,\nu}(x|\mu, \sigma^2) = \frac{2\gamma s \Gamma \left( \frac{\nu + 1}{2} \right)}{(1 + \gamma^2) \Gamma \left( \frac{\nu}{2} \right) \sqrt{\nu - 2} \pi \sigma^2} \left[ 1 + \frac{\gamma^{-2I}[m\sigma + s(x - \mu)]^2}{(\nu - 2)\sigma^2} \right]^{-\frac{\nu + 1}{2}}
\]

(5.8)

where

\[
I = \begin{cases} 
1 & \text{if } x \geq \mu - \frac{m\sigma}{s} \\
-1 & \text{if } x < \mu - \frac{m\sigma}{s}
\end{cases}
\]

and

\[
m = \frac{\Gamma \left( \frac{\nu + 1}{2} \right) \sqrt{\nu - 2}}{\Gamma \left( \frac{\nu}{2} \right) \sqrt{\pi}} \left( \gamma - \frac{1}{\gamma} \right), \quad s^2 = \gamma^2 + \frac{1}{\gamma^2} - 1 - m^2.
\]

The constants \( m \) and \( s \) ensure that the distribution is standardised to have mean \( \mu \) and variance \( \sigma^2 \), which are included after application of equation (5.7). Bauwens and Laurent (2005) extend this idea to the multivariate space, and consider a skewed alternative to the multivariate standardised Student-t distribution:

\[
f_{\eta}(x|\mu, \Sigma) = \frac{\Gamma \left( \frac{\eta + N}{2} \right) |\Sigma|^{-\frac{1}{2}}}{\Gamma \left( \frac{N}{2} \right) \eta^{-N/2}} \left[ 1 + \frac{(x - \mu)'\Sigma^{-1}(x - \mu)}{\eta - 2} \right]^{-\frac{\eta + N}{2}}.
\]
Their multivariate skew-\( t \) PDF is given by

\[ \text{skt}_{\gamma, \eta}(x | \mu, \Sigma) = \left( \frac{2}{\sqrt{\pi}} \right)^N \left( \prod_{i=1}^{N} \frac{\gamma_i s_i}{1 + \gamma_i^2} \right) \frac{\Gamma \left( \frac{\eta+N}{2} \right) \sqrt{\eta-2}}{\Gamma \left( \frac{\eta}{2} \right) (\eta-2)^{\frac{N}{2}}} \left( 1 + \frac{a^\prime a}{\eta-2} \right)^{-\frac{\eta+N}{2}} \]  

(5.9)

where \( \mu = [\mu_i] \) and \( \Sigma = [\sigma_{ij}] \) are the mean vector and covariance matrix (respectively), \( \gamma = (\gamma_1, \ldots, \gamma_N) \) is a vector of skewness parameters and \( \eta \) is the degrees of freedom parameter. The additional vector \( a = (a_1, \ldots, a_N) \) has elements defined as \( a_i = \gamma_i^{-1}(m_i + s_i x_i^*) \), with \( x^* = (x_1^*, \ldots, x_N^*) = \Sigma^{-\frac{1}{2}}(x - \mu) \) and

\[ I_{it} = \begin{cases} 
1 & \text{if } x_i \geq \mu_i - \frac{m_i \sqrt{\sigma_{ii}}}{s_i}, \\
-1 & \text{if } x_i < \mu_i - \frac{m_i \sqrt{\sigma_{ii}}}{s_i},
\end{cases} \]

\[ m_i = \frac{\Gamma \left( \frac{\eta-1}{2} \right) \sqrt{\eta-2}}{\Gamma \left( \frac{\eta}{2} \right) \sqrt{\pi}} \left( \gamma_i - \frac{1}{\gamma_i} \right), \quad s_i^2 = \gamma_i^2 + 1 - m_i^2. \]

The Bauwens and Laurent (2005) distribution from equation (5.9) will be employed as the basis for a skew-\( t \) copula in this chapter, mainly due to the following reasons:

- Its marginal distributions revert to the univariate skew-\( t \) distribution in equation (5.8), whose PDF, CDF and inverse CDF are available in many statistical computer packages (e.g. R Development Core Team, 2010) or can be easily programmed\(^2\).

- There is no dependence on a CDF in the calculation of the multivariate PDF in (5.9), yielding faster computations.

- The distribution is in a standardised format, in order to eliminate the impact of \( \gamma \) and \( \eta \) on the covariance matrix so that \( E(X) = \mu \) and \( \text{Cov}(X) = \Sigma \).

- The univariate skew-\( t \) distribution in equation (5.8) has been included with GARCH models in applications to financial time series (e.g. see Alberg et al., 2008 and Tu

\(^2\)The actual derivations of the CDF and inverse CDF of the univariate skew-\( t \) distribution are detailed in the Appendix of this chapter.
et al., 2008). Hence it seems natural for an extension to higher dimensions with an application to multivariate financial time series.

5.4 Copula-GARCH Models

This section reviews some of the main copula-GARCH model contributions to the literature. However, it is necessary to first briefly review the definition of a copula and some of its properties before introducing each contribution.

5.4.1 Copula Properties

A copula is an $N$-dimensional multivariate CDF defined on the unit $N$-cube $[0, 1]^N$ that has uniformly distributed marginal distributions. Sklar (1959) states and proves many properties of copulas, two of which are mentioned below for convenience:

Let $X_i, i = 1, \ldots, N$ denote random variables and let $F$ denote any multivariate CDF with marginals given by $u_i = F_i(x_i) = \Pr(X_i \leq x_i), i = 1, \ldots, N$.

1. $F$ can be represented by its univariate marginals and a copula $C$ such that

$$F(x_1, \ldots, x_N) = C(F_1(x_1), \ldots, F_N(x_N)). \quad (5.10)$$

2. If $C$ is a copula and each $F_i$ are marginal CDF’s, then $F$ is a joint distribution with each $F_i$ as its marginals:

$$C(u_1, \ldots, u_N) = F(F_1^{-1}(u_1), \ldots, F_N^{-1}(u_N)). \quad (5.11)$$

Now let $\{y_t\}_{t=1}^n = \{(y_{1t}, \ldots, y_{Nt})\}_{t=1}^n$ denote a set of $N$-dimensional observations of sample size $n$ and $\psi_{t-1}$ the information set at time $t - 1$. A copula-GARCH model can be defined by first specifying a time-varying copula $C_t$ for the time-varying CDF $F_t$ of $y_t$, conditional on $\psi_{t-1}$. This can be achieved by utilising the copula property from equation
as follows:

\[ y_t | \psi_{t-1} \sim F_t(y_t | \psi_{t-1}) = C_t(F_{1t}(y_{1t} | \psi_{t-1}), \ldots, F_{Nt}(y_{Nt} | \psi_{t-1}) | \psi_{t-1}) \]  

(5.12)

where \( F_{it} \) denotes the marginal CDF of each \( y_{it} | \psi_{t-1}, \forall i = 1, \ldots, N \). The PDF of \( y_t | \psi_{t-1} \) can easily be derived from equation (5.12) through differentiation:

\[ f_t(y_t | \psi_{t-1}) = c_t(F_{1t}^{-1}(y_{1t} | \psi_{t-1}), \ldots, F_{Nt}^{-1}(y_{Nt} | \psi_{t-1}) | \psi_{t-1}) \prod_{i=1}^N f_{it}(y_{it} | \psi_{t-1}) \]  

(5.13)

where \( c_t \) denotes the PDF of the copula \( C_t \), and \( f_{it} \) denotes the marginal PDF of each \( y_{it} | \psi_{t-1}, \forall i = 1, \ldots, N \). The copula PDF can also be derived from differentiation, this time using equation (5.11) and setting \( u_{it} = F_{it}(y_{it} | \psi_{t-1}) \):

\[ c_t(u_{1t}, \ldots, u_{Nt} | \psi_{t-1}) = \frac{f_t(F_{1t}^{-1}(u_{1t}), \ldots, F_{Nt}^{-1}(u_{Nt}) | \psi_{t-1})}{\prod_{i=1}^N f_{it}(F_{it}^{-1}(u_{it}) | \psi_{t-1})}. \]  

(5.14)

Perhaps the most commonly used copulas in the literature for financial time series are the Gaussian and Student-\( t \) copulas, which are based on their respective multivariate distributions. Using their PDFs in equations (5.1) and (5.2) in addition to the copula equations in (5.11) and (5.14) these copulas can be written as follows (omitting the time dependence and \( \psi_{t-1} \) condition for convenience):

\[ C^{Gaussian}(u_{1t}, \ldots, u_{Nt}) = \Phi(\Phi^{-1}(u_{1t}), \ldots, \Phi^{-1}(u_{Nt}) | 0, P_t), \]  

(5.15)

\[ c^{Gaussian}(u_{1t}, \ldots, u_{Nt}) = |P_t|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} x_t'(P_t^{-1} - I_N)x_t \right], \]  

(5.16)

\[ C^{Student}(u_{1t}, \ldots, u_{Nt}) = T_\eta(T_\eta^{-1}(u_{1t}), \ldots, T_\eta^{-1}(u_{Nt}) | 0, P_t), \]  

(5.17)

\[ c^{Student}(u_{1t}, \ldots, u_{Nt}) = \frac{|P_t|^{-\frac{1}{2}} \Gamma \left( \frac{\eta + N}{2} \right) \Gamma \left( \frac{\eta}{2} \right)^{N-1} \left( 1 + \frac{1}{\eta} x_t'P_t^{-1}x_t \right)^{-\frac{\eta + N}{2}}}{\Gamma \left( \frac{\eta + 1}{2} \right) \prod_{i=1}^N \left( 1 + \frac{x_{it}^2}{\eta} \right)^{-\frac{\eta + 1}{2}}}, \]  

(5.18)
where $\Phi$ and $T_{\eta}$ denote the multivariate normal and Student-\(t\) CDFs with zero mean vector, correlation matrix $P_t = [\rho_{ij,t}]$ and degrees of freedom $\eta$, $\Phi^{-1}$ and $T_{\eta}^{-1}$ denote the univariate normal and Student-\(t\) inverse CDFs, $x_t = (x_{1t}, \ldots, x_{Nt})$, $x_{it} = \Phi^{-1}(u_{it})$ (Gaussian), $x_{it} = T_{\eta}^{-1}(u_{it})$ (Student-\(t\)) and $I_N$ is the $N \times N$ identity matrix.

Let $\theta_M$ denote the parameters of the $i$th marginal, and let $\theta_C$ denote the parameters of the copula. As with a standard multivariate GARCH model, the likelihood function of a copula-GARCH model can be formed using the product of the model PDFs for each observation: If $\theta$ denotes the full model parameters, from equation (5.13) we obtain

$$L(\theta) = \prod_{t=1}^{n} f_t(y_t|\psi_{t-1}, \theta)$$

$$= \prod_{t=1}^{n} \left[ c_t(u_{1t}, \ldots, u_{Nt}|\psi_{t-1}, \theta_C) \prod_{i=1}^{N} f_{it}(y_{it}|\psi_{t-1}, \theta_{Mi}) \right]$$

with log-likelihood

$$l(\theta) = \sum_{t=1}^{n} \left[ \ln c_t(u_{1t}, \ldots, u_{Nt}|\psi_{t-1}, \theta_C) + \sum_{i=1}^{N} \ln f_{it}(y_{it}|\psi_{t-1}, \theta_{Mi}) \right]. \quad (5.19)$$

In order to estimate the parameters of a copula-GARCH model, it has been typical in the literature to apply the two-step technique proposed by Joe and Xu (1996), where the marginals and copula are estimated separately. Estimation proceeds as follows:

1. Estimate the parameters of each marginal via maximum likelihood:

$$\hat{\theta}_M = \arg \max \sum_{t=1}^{n} \ln f_{it}(y_{it}|\psi_{t-1}, \theta_{Mi}), \quad \forall i = 1, \ldots, N;$$

2. Based on the standardised residuals $\hat{z}_{it}$ formed from each $\hat{\theta}_M$, calculate each marginal CDF $\hat{u}_{it} = F_{it}(\hat{z}_{it}|\psi_{t-1}, \hat{\theta}_{Mi})$ and estimate the parameters of the copula via maximum likelihood:

$$\hat{\theta}_C = \arg \max \sum_{t=1}^{n} \ln c_t(F_{1t}^{-1}(\hat{u}_{1t}), \ldots, F_{Nt}^{-1}(\hat{u}_{Nt})|\psi_{t-1}, \theta_C).$$
The literature contains many publications on copula-GARCH models. However, the skew-\(t\) copula-GARCH model to be proposed in this chapter has been motivated by the following selection of references:

1. Jondeau and Rockinger (2006);
2. Serban et al. (2007);
3. Ausin and Lopes (2010);

each of which will now be reviewed.

### 5.4.2 Jondeau and Rockinger (2006)

Jondeau and Rockinger (2006) propose a bivariate copula-GARCH model for capturing the dependence between a set of international stock market indices. Assuming that \(\{y_t\}_{t=1}^n\) is a set of univariate time series observations, their model for each marginal distribution can be described as follows:

\[
y_t = \mu_t + \varepsilon_t, \\
h_t = a_0 + b_0^+ (\varepsilon_{t-1}^+)^2 + b_0^- (\varepsilon_{t-1}^-)^2 + c_0 h_{t-1},
\]

where \(\mu_t\) is the conditional mean function (assuming the form of an AR(10) model), \(\{\varepsilon_t\}_{t=1}^n\) are the residuals and \(h_t\) is the conditional variance function with parameters \(a_0, b_0^+, b_0^-\) and \(c_0\). Asymmetry in the marginals is captured via \(b_0^+\) and \(b_0^-\), with \(\varepsilon_t^+ = \text{max}(\varepsilon_t, 0)\) and \(\varepsilon_t^- = \text{max}(-\varepsilon_t, 0)\), similar to that of a GJR-GARCH model. Each standardised residual \(z_t = \varepsilon_t / \sqrt{h_t}\) is chosen to follow the skew-\(t\) distribution of Hansen (1994), which has the PDF

\[
skt_{\lambda, \nu}(z_t) = \begin{cases} 
bc \left[ 1 + \frac{1}{\nu - 2} \left( \frac{bz_t + a}{1 - \lambda} \right)^2 \right]^{-\frac{\nu + 1}{2}} & \text{if } z_t < -a/b \\
bc \left[ 1 + \frac{1}{\nu - 2} \left( \frac{bz_t + a}{1 + \lambda} \right)^2 \right]^{-\frac{\nu + 1}{2}} & \text{if } z_t \geq -a/b
\end{cases}
\]

(5.20)
where

\[ a = 4\lambda c \left( \frac{\nu - 2}{\nu - 1} \right), \quad b^2 = 1 + 3\lambda^2 - a^2 \quad \text{and} \quad c = \frac{\Gamma \left( \frac{\nu + 1}{2} \right)}{\Gamma \left( \frac{\nu}{2} \right) \sqrt{\nu - 2}}. \]

This distribution has skewness and degrees of freedom parameters given by \( \lambda \) and \( \nu \) (respectively). As an addition to the model, these two parameters are also assumed to be time-varying so that \( \varepsilon_t | \psi_{t-1} \sim \text{skt}_{\lambda_t, \nu_t}(0, h_t) \), with

\[
\begin{align*}
\tilde{\lambda}_t &= a_1 + b_1^+ \varepsilon_{t-1}^+ + b_1^- \varepsilon_{t-1}^- + c_1 \tilde{\lambda}_{t-1}, \quad \lambda_t = g_{[L, U]}(\tilde{\lambda}_t) \\
\tilde{\nu}_t &= a_2 + b_2^+ \varepsilon_{t-1}^+ + b_2^- \varepsilon_{t-1}^- + c_2 \tilde{\nu}_{t-1}, \quad \nu_t = g_{[L, U]}(\tilde{\nu}_t)
\end{align*}
\]

where \( a_i, b_i^+, b_i^- \) and \( c_i \) are parameters, and \( g_{[L, U]}(x) = L + (U - L)[1 + \exp(-x)]^{-1} \) is called the logistic map, which restricts \( x \) to the interval \([L, U] \).

Let the CDF of the marginals be defined as \( u_{it} = F_{it}(z_{it} | \psi_{t-1}), i = 1, 2 \). Two copulas are chosen to model the dependence structure, namely the Gaussian and Student-\( t \) copulas as given by equations (5.15)-(5.18). The correlations \( \rho_{12,t} \) in these copulas are given three different structures:

1. A semi-parametric approach where \( \rho_{12,t} \) depends on the position of past joint realisations of \( z_{1t} \) and \( z_{2t} \) in the unit square:

\[
\rho_{12,t} = \sum_{j=1}^{16} d_j I[(z_{1,t-1}, z_{2,t-1}) \in A_j]
\]

where \( A_j \) is the \( j \)th element of the unit square grid, \( I \) is the indicator variable that is equal to 1 if the condition is true and 0 otherwise, and \( d_j \in [-1, 1] \).

2. A persistence specification similar to that of a GARCH model:

\[
\rho_{12,t} = \tilde{\rho}_{12}(1 - \alpha - \beta) + \alpha \xi_{12,t-1} + \beta \rho_{12,t-1}
\]

where
\[ \xi_{12,t} = \frac{\sum_{k=1}^{M} z_{1,t-k+1} z_{2,t-k+1}}{\sqrt{\left( \sum_{k=1}^{M} z_{1,t-k+1}^2 \right) \left( \sum_{k=1}^{M} z_{2,t-k+1}^2 \right)}} \]

represents the correlation between the residuals over the \( M \) most recent observations (see for e.g. Tse and Tsui, 2002), with \( \hat{\rho}_{12}, \alpha \) and \( \beta \) parameters.

3. A regime-switching specification for both the correlation \( \rho_{12,t} \) and degrees of freedom \( \eta \) (for the Student-\( t \) copula):

\[
\begin{align*}
\rho_{12,t} &= \rho_0 S_t + \rho_1 (1 - S_t), \\
\eta_t &= \eta_0 S_t + \eta_1 (1 - S_t),
\end{align*}
\]

where \( \rho_i \) and \( \eta_i \) are parameters and \( S_t \) denotes the unobserved regime of the system at time \( t \), assumed to follow a two-state Markov process with transition probability matrix \( Q \) having elements \( q_0 = \Pr(S_t = 0|S_{t-1} = 0) \) and \( q_1 = \Pr(S_t = 1|S_{t-1} = 1) \), where

\[
Q = \begin{pmatrix} q_0 & 1 - q_0 \\ 1 - q_1 & q_1 \end{pmatrix}.
\]

This model has the advantage of capturing skewness and kurtosis of the marginal distributions via the skew-\( t \) distribution in equation (5.20), as well as allowing the first four conditional moments to be time-varying. Flexibility in the specification of the correlations is also considered. However, some of the main drawbacks of the model are:

- Asymmetric dependence is not considered, as only Gaussian and Student-\( t \) copulas are implemented;
- Each correlation structure has a simple specification on their own, and do not allow for dependence properties together (e.g. there could be multiple regimes in the persistence specification);
- While the application could be extended from the bivariate case, this may result in parsimony issues with the model specification.
5.4.3 Serban et al. (2007)

Serban et al. (2007) propose a 17-dimensional copula GARCH model and apply it to index funds. The marginal distributions of this model have the following conditional mean and variance equations (for $i = 1, \ldots, 17$):

$$
y_{it} = \phi_i y_{i,t-1} + \varepsilon_{it}, \quad \varepsilon_{it} | \psi_{t-1} \sim t^*_{\nu_i}(0, h_{it})
$$

$$
h_{it} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1}
$$

where $\phi_i$, $\omega_i$, $\alpha_i$ and $\beta_i$ are parameters and $t^*_{\nu_i}$ denotes the standardised Student-$t$ distribution with degrees of freedom parameter $\nu_i$. The copula used to link these marginals is the Student-$t$ copula given by equations (5.17) and (5.18), with the following specification for the correlation matrix $P_t = [\rho_{ij,t}]$

$$
\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t}q_{jj,t}}}, \quad \forall \ i, j = 1, \ldots, 17
$$

where the matrix $Q_t = [q_{ij,t}]$ is given two specifications:

1. Scalar DCC equations (with parameter $\lambda$):

   $$
   q_{ij,t} = (1 - \lambda) z_{i,t-1} z_{j,t-1} + \lambda q_{ij,t-1}
   $$

2. Generalised DCC equations:

   $$
   Q_t = \bar{Q} - A' \bar{Q} A - B' \bar{Q} B + A' z_{t-1} z_{t-1}' A + B' Q_{t-1} B
   $$

   with $\bar{Q}$, $A$ and $B$ parameter matrices, and $z_t = (z_{1t}, \ldots, z_{17t})$ a vector of standardised residuals.

They show that the scalar DCC specification (87 parameters) is preferred over the generalised DCC specification (120 parameters) in their application from the results of the BIC.
in model selection. The obvious advantage with this model is the application in a high-dimensional space \((N = 17)\) with a small number of parameters. However, the mean, variance and correlation components of the model are of a simple form: the means are AR(1) models with zero intercept terms, the variances are GARCH\((1, 1)\) models, and the correlations have one parameter describing the dependence between all residual pairs in the scalar DCC specification, and a simple BEKK model in the generalised specification. Hence there is no allowance for any forms of asymmetry in any elements of the model. In addition, there is also no allowance for any skewness in the marginal components, as only fat tails are captured via the univariate and multivariate Student-\(t\) distributions.

### 5.4.4 Ausin and Lopes (2010)

Aisin and Lopes (2010) propose a similar model to that of Serban et al. (2007), but use Bayesian methods to estimate parameters in a bivariate setting: their conditional mean and variance equations for the marginals are

\[
\begin{align*}
   y_{it} &= \mu_i + \sqrt{h_{it}}z_{it}, \\
   h_{it} &= \omega_i + \alpha_i(y_{i,t-1} - \mu_i)^2 + \beta_i h_{i,t-1}
\end{align*}
\]

where \(\mu_i, \omega_i, \alpha_i\) and \(\beta_i\) are parameters, and \(z_{it}\) is chosen to follow the (unstandardised) Student-\(t\) distribution with degrees of freedom \(\nu_i\), zero mean and variance \(\nu_i/(\nu_i-2)\). The copula chosen to link these marginals is the Student-\(t\) copula with \(\eta\) degrees of freedom from equations (5.17) and (5.18), with the following specification for the correlation matrix \(P_t = [\rho_{ij,t}]\):

\[
P_t = \bar{P}(1 - a - b) + a\xi_{t-1} + bP_{t-1}
\]

where \(a\) and \(b\) are nonnegative parameters, \(\bar{P}\) is a correlation parameter matrix and \(\xi_t = [\xi_{ij,t}]\) is defined as
\[
\xi_{ij,t} = \frac{\sum_{k=1}^{M} x_{i,t-k+1} x_{j,t-k+1}}{\sqrt{\left(\sum_{k=1}^{M} x_{i,t-k+1}^2\right) \left(\sum_{k=1}^{M} x_{j,t-k+1}^2\right)}}
\]

where \(x_{it} = T^{-1}(z_{it})\). This is similar to that used in the persistence structure of Jondeau and Rockinger (2006), but the sample correlations are calculated on the \(x_{it}\)'s rather than the \(z_{it}\)'s.

Rather than using a classical two-step estimation process (as described earlier), Ausin and Lopes (2010) develop a one-step Bayesian approach where the parameters from the entire model are estimated via simulating from the joint posterior distribution of all parameters. This is briefly described below:

**One-Step Bayesian Approach**

Let \(p(\theta|y)\) denote the joint posterior distribution of the parameters \(\theta\) in the model. To obtain simulations from this posterior, Ausin and Lopes (2010) develop a Gibbs sampler that simulates from posteriors that are conditional on other parameters in the model (as in Chapters 3-4 of this thesis). These conditional posteriors are given in the sampling scheme below:

1. \(p(\mu_1, \omega_1, \alpha_1, \beta_1|y, \theta \neq \mu_1, \omega_1, \alpha_1, \beta_1)\)
2. \(\vdots\)
3. \(N. p(\mu_N, \omega_N, \alpha_N, \beta_N|y, \theta \neq \mu_N, \omega_N, \alpha_N, \beta_N)\)
4. \(N + 1. p(\nu_1|y, \theta \neq \nu_1)\)
5. \(\vdots\)
6. \(2N. p(\nu_N|y, \theta \neq \nu_N)\)
7. \(2N + 1. p(\eta|y, \theta \neq \eta)\)
8. \(2N + 2. p(\bar{P}, a, b|y, \theta \neq \bar{P}, a, b)\)
The priors for each conditional posterior above are assumed uniform, restricted within the range allowed by the usual GARCH parameter restrictions. This is not the case, however, for the degrees of freedom parameters $\nu_i$ and $\eta_i$ whose priors are chosen to be proportional to $1/(1 + \nu_i^2)$ and $1/(1 + \eta_i^2)$, with $\nu_i, \eta_i > 0$ (the half-right side Cauchy distribution). Samples are then drawn from each conditional posterior above via the MH algorithm, using random walk multivariate normal proposal distributions.

**Two-Step Bayesian Approach**

Analogous to the classical two-step approach, from a Bayesian perspective the estimation is split into the marginal and copula components, and an MCMC algorithm is undertaken on each component. Let $\theta_{Mi} = (\mu_i, \omega_i, \alpha_i, \beta_i, \nu_i)$ denote the parameter vector for marginal $M_i$, and let $\theta_C = (\bar{P}, a, b, \eta)$ denote the copula parameters. The two-step Bayesian estimation can be carried out as follows:

1. Simulate from the joint posterior distribution $p(\theta_{Mi}|y)$ for each marginal, resulting in the sample $\hat{\theta}_{Mi} = \{\theta_{Mi}^{[j]}\}_{j=1}^J$.

2. After forming the estimates

$$\hat{z}_{it} = \frac{1}{J} \sum_{j=1}^J y_{it} - \mu_{i}^{[j]} \sqrt{h_{it}^{[j]}}$$

and

$$\hat{\nu}_i = \frac{1}{J} \sum_{j=1}^J \nu_{i}^{[j]}$$

simulate from the joint posterior $p(\theta_C|y, \hat{\theta}_{M1}, \ldots, \hat{\theta}_{MN})$.

Step 1 is simply Bayesian estimation on a univariate GARCH model with Student-$t$ errors, and simulation is performed using the relevant conditional posteriors in the Gibbs sampler discussed above. Step 2 utilises the standardised residual estimates and the estimate of each $\nu_i$ which are substituted into the copula posterior, then simulation is also performed via the relevant conditional posteriors in the Gibbs sampler above.

The main advantage of this copula-GARCH model is that using a one-step Bayesian approach considers parameter estimation in the one step, which improves efficiency and
also considers parameter uncertainty through simulation of their posterior distributions. While this approach demonstrates to work well in the Ausin and Lopes (2010) paper, it is only a bivariate example, as is many applications of copula-GARCH models. The one-step Bayesian approach becomes quite computationally expensive once the dimension of the problem increases. In addition, both marginals and copula components of the model do not consider asymmetry or skewness at all, which has shown to exist in stock returns.

Summary

As mentioned in the introduction to this chapter, the series of copula-GARCH models proposed in the literature thus far (including the ones just reviewed) have incorporated asymmetry, skewness and excess kurtosis in various ways, but not in all components of their models. The skew-$t$ copula-GARCH model to be proposed in the following section will attempt to fill this important gap by generalising both marginal and copula components to have all of these properties, and it will be demonstrated that these additional effects are significant at least in the stock portfolio application to be presented. In addition, parameters will be estimated via the two-step Bayesian approach of Ausin and Lopes (2010), which, while not as efficient as the one-step Bayesian approach, has a reduced computational expense in higher dimensions and thus is more useful in portfolios of many stocks.

5.5 The ASTC-DTGARCH Model

Let $\{y_t\}_{t=1}^n = \{(y_{1t}, \ldots, y_{Nt})\}_{t=1}^n$ denote a set of $N$-dimensional observations of sample size $n$ and $\psi_{t-1}$ the information set at time $t - 1$. The Asymmetric Skew-$t$ Copula - Double Threshold GARCH (ASTC-DTGARCH) model allows for conditional mean and variance asymmetry, as well as skewness and excess kurtosis in its marginal distributions. The dependence structure is captured via a skew-$t$ copula formed from the multivariate skew-$t$ distribution of Bauwens and Laurent (2005).
5.5.1 The Marginal Distributions

The conditional mean and variance equations of the marginal distributions are of a similar form to that of the DTGARCH model of Chen et al. (2005): Let \( \{ (\varepsilon_{1t}, \ldots, \varepsilon_{Nt}) \}_{t=1}^{n} \) denote a set of residuals, with \( \{ (h_{1t}, \ldots, h_{Nt}) \}_{t=1}^{n} \) defining the conditional variances. For \( i = 1, \ldots, N \), we have

\[
y_{it} = \begin{cases} 
  m_{i0}^{(1)} + m_{i1}^{(1)} y_{i,t-1} + \varepsilon_{it} & \text{if } y_{i,t-1} \leq r_{i}, \\
  m_{i0}^{(2)} + m_{i1}^{(2)} y_{i,t-1} + \varepsilon_{it} & \text{if } y_{i,t-1} > r_{i},
\end{cases}
\]

\[
h_{it} = \begin{cases} 
  c_{i1}^{(1)} + a_{i1}^{(1)} \varepsilon_{i,t-1}^{2} + b_{i1}^{(1)} h_{i,t-1} & \text{if } y_{i,t-1} \leq r_{i}, \\
  c_{i1}^{(2)} + a_{i1}^{(2)} \varepsilon_{i,t-1}^{2} + b_{i1}^{(2)} h_{i,t-1} & \text{if } y_{i,t-1} > r_{i},
\end{cases}
\]

where each \( \varepsilon_{it} | \psi_{t-1} \) has been chosen to follow the univariate skew-\( t \) distribution using the approach of Fernandez and Steel (1998) from equation (5.8), and in terms of the errors is given by the PDF

\[
skt_{\lambda_i, \nu_i}(\varepsilon_{it}|0, h_{it}) = \frac{2 \lambda_i s_i \Gamma(\frac{\nu_i+1}{2})}{(1 + \lambda_i^2 \nu_i)} \sqrt{\nu_i - 2} \pi h_{it} \left[ 1 + \frac{\lambda_i^{-2I_{it}} (m_i \sqrt{h_{it}} + s_i \varepsilon_{it})^2}{(\nu_i - 2)h_{it}} \right]^{-\frac{\nu_i+1}{2}}
\]

where

\[
I_{it} = \begin{cases} 
  1 & \text{if } \varepsilon_{it} \geq -\frac{m_i \sqrt{h_{it}}}{s_i}, \\
  -1 & \text{if } \varepsilon_{it} < -\frac{m_i \sqrt{h_{it}}}{s_i},
\end{cases}
\]

and

\[
m_i = \frac{\Gamma(\frac{\nu_i-1}{2})}{\Gamma(\frac{\nu_i}{2})} \sqrt{\nu_i - 2} \left( \lambda_i - \frac{1}{\lambda_i} \right), \quad s_i^2 = \lambda_i^2 + \frac{1}{\lambda_i^2} - 1 - m_i^2.
\]

The parameters of the marginals can be represented by the following vectors (for \( i = 1, \ldots, N \)):

\[
\mathbf{m}_i = (m_{i0}^{(1)}, m_{i0}^{(2)}, m_{i1}^{(1)}, m_{i1}^{(2)}), \quad \mathbf{c}_i = (c_{i1}^{(1)}, c_{i1}^{(2)}), \quad \mathbf{a}_i = (a_{i1}^{(1)}, a_{i1}^{(2)}), \quad \mathbf{b}_i = (b_{i1}^{(1)}, b_{i1}^{(2)}), \quad \mathbf{r} = (r_1, \ldots, r_N), \quad \mathbf{\lambda} = (\lambda_1, \ldots, \lambda_N), \quad \mathbf{\nu} = (\nu_1, \ldots, \nu_N).
\]
Hence the autoregressive parameters are contained in the vector $m_i$; the GARCH parameters are contained in $c_i$, $a_i$ and $b_i$; the threshold parameters are contained in $r$; while the skewness and degrees of freedom parameters are contained in $\lambda$ and $\nu$ (respectively). Note that these marginals revert to the symmetric standardised Student-t distribution when $\lambda_i = 1$. In fact, $\lambda_i^2$ is a skewness measure determining the skewness of each marginal (as mentioned in Bauwens and Laurent, 2005), and the sign of $\ln \lambda_i$ indicates the direction of the skewness: if $\ln \lambda_i < 0$ ($> 0$) then the distribution is negatively (positively) skewed. See the Appendix of this chapter for some properties of the univariate skew-t distribution.

Let the standardised residuals be defined as $z_{it} = \varepsilon_{it}/\sqrt{h_{it}}$, $\forall i = 1, \ldots, N$. Then the marginal CDF of $y_{it} | \psi_{t-1}$ can be written in short-hand as $u_{it} = F_{it}(y_{it} | \psi_{t-1}) = \text{Skt}_{\lambda_i, \nu}(z_{it})$ (see equation (5.41) in the Appendix).

### 5.5.2 The Skew-t Copula

A skew-t copula is chosen to link the above marginal distributions, which will form the ASTC-DTGARCH model. This skew-t copula is based on the multivariate skew-t distribution of Bauwens and Laurent (2005), and using equation (5.11), this becomes

$$C_t(u_{1t}, \ldots, u_{Nt} | \psi_{t-1}) = \text{Skt}_{\gamma, \eta}^{-1}(\text{Skt}_{\gamma_1, \eta}(u_{1t}), \ldots, \text{Skt}_{\gamma_N, \eta}(u_{Nt}) | 0, P_t)$$

(5.22)

where $\text{Skt}_{\gamma, \eta}$ denotes the CDF of the multivariate skew-t distribution with zero mean vector, correlation matrix $P_t$ and parameters given by $\gamma$ and $\eta$, while each $\text{Skt}_{\gamma_i, \eta}^{-1}$ denotes the inverse CDF of the univariate skew-t distribution with skewness parameter $\gamma_i$ and degrees of freedom $\eta$ (see equation (5.44) in the Appendix). Refer to the Appendix of this chapter for more properties of this skew-t copula.

It is important to note at this point that the parameters $\gamma$ and $\eta$ normally define the skewness vector and degrees of freedom (respectively) of the multivariate skew-t distribution. However, because this distribution is used to form a copula, they do not have this
interpretation in the ASTC-DTGARCH model. This is because copulas are used only to “link” marginal distributions together to form the joint CDF for the problem at hand. Instead, each $\gamma_i$ has an impact on the amount of non-linearity in the copula’s dependence structure - the further each $\gamma_i$ is away from 1, the more non-linear the dependence structure. Refer to the examples in Figure 5.7 and its corresponding commentary. Each of the marginal parameters $\lambda_i$ and $\nu_i$ instead can be used to form $N$ univariate skewness and excess kurtosis measures.

As with the marginals, the copula PDF can be obtained by differentiation in equation (5.22):

$$c_t(u_1, \ldots, u_N|\psi_{t-1}) = \frac{\text{skt}_{\eta}(\text{Skt}_{\eta}^{-1}(u_{1t}), \ldots, \text{Skt}_{\eta}^{-1}(u_{Nt}|0, P_t))}{\prod_{i=1}^{N} \text{skt}_{\eta}(\text{Skt}_{\eta}^{-1}(u_{it}))}$$

(5.23)

where $\text{skt}$ and $\text{skt}$ denote the corresponding univariate and multivariate skew-t PDFs (respectively).

Now let $x_t = (x_{1t}, \ldots, x_{Nt})$, where $x_{it} = \text{Skt}_{\eta}^{-1}(u_{it})$, $\forall i = 1, \ldots, N$ and define $x^*_t = P_t^{-\frac{1}{2}} x_t$. Using the multivariate PDF expression in equation (5.9), the copula PDF from equation (5.23) can be written as

$$c_t(u_1, \ldots, u_N|\psi_{t-1}) = \frac{|P_t|^{-\frac{1}{2}} \Gamma\left(\frac{\eta + N}{2}\right) \Gamma\left(\frac{\eta}{2}\right)^{N-1} \left(1 + \frac{a_t^t a_t}{\eta - 2}\right)^{-\frac{\eta N}{2}}}{\Gamma\left(\frac{\eta + 1}{2}\right)^N \prod_{i=1}^{N} \left[1 + \frac{\gamma_i^{-2} I_{i}^{C^*} (m_i^C + s_i^C x^*_i)^2}{\eta - 2}\right]^{-\frac{\eta + 1}{2}}}$$

(5.24)

where the vector $a_t = (a_{1t}, \ldots, a_{Nt})$ has each element as $a_{it} = \gamma_i^{-1} I_{i}^{C^*} (m_i^C + s_i^C x^*_i)$, with

$$I_{i}^{C^*} = \begin{cases} 1 & \text{if } x^*_i \geq -\frac{m_i^C}{s_i^C}, \\ -1 & \text{if } x^*_i < -\frac{m_i^C}{s_i^C} \\ \end{cases}, \quad I_{i}^{C} = \begin{cases} 1 & \text{if } x_{it} \geq -\frac{m_i^C}{s_i^C}, \\ -1 & \text{if } x_{it} < -\frac{m_i^C}{s_i^C} \\ \end{cases}$$

and
\[ m_i^C = \frac{\Gamma \left( \frac{\eta - 2}{2} \right)}{\Gamma \left( \frac{\eta}{2} \right) \sqrt{\pi}} \left( \gamma_i - \frac{1}{\gamma_i} \right), \quad (s_i^C)^2 = \gamma_i^2 + \frac{1}{\gamma_i^2} - 1 - (m_i^C)^2. \]

Time-varying correlations in the dependence structure, represented by the symmetric matrix \( P_t = [\rho_{i,j,t}] \) in the copula above, will follow a similar form to that of Tse and Tsui (2002), with additional terms to allow for correlation asymmetry: Let \( \xi_t = [\xi_{i,j,t}] \), where

\[ \xi_{i,j,t} = \frac{\sum_{k=1}^M x_{i,t-k+1} x_{j,t-k+1}}{\sqrt{\left( \sum_{k=1}^M x_{i,t-k+1}^2 \right) \left( \sum_{k=1}^M x_{j,t-k+1}^2 \right)}} , \quad \forall \, i, j = 1, \ldots, N. \tag{5.25} \]

Equation (5.25) represents the sample correlation between each \( x_{it} \) and \( x_{jt} \), using the most recent \( M \) observations, as utilised by Ausin and Lopes (2010). The correlation matrix \( P_t = [\rho_{i,j,t}] \) then has elements defined as follows:

\[
\begin{align*}
\rho_{i,j,t} = & \begin{cases} 
\hat{\rho}_{ij}^{(1)} (1 - \alpha^{(1)} - \beta^{(1)}) + \alpha^{(1)} \xi_{i,j,t-1} + \beta^{(1)} \rho_{i,j,t-1} & \text{if } z_{i,t-1}, z_{j,t-1} \geq 0, \\
\hat{\rho}_{ij}^{(2)} (1 - \alpha^{(2)} - \beta^{(2)}) + \alpha^{(2)} \xi_{i,j,t-1} + \beta^{(2)} \rho_{i,j,t-1} & \text{if } z_{i,t-1} \geq 0, z_{j,t-1} < 0, \\
\hat{\rho}_{ij}^{(3)} (1 - \alpha^{(3)} - \beta^{(3)}) + \alpha^{(3)} \xi_{i,j,t-1} + \beta^{(3)} \rho_{i,j,t-1} & \text{if } z_{i,t-1} < 0, z_{j,t-1} \geq 0, \\
\hat{\rho}_{ij}^{(4)} (1 - \alpha^{(4)} - \beta^{(4)}) + \alpha^{(4)} \xi_{i,j,t-1} + \beta^{(4)} \rho_{i,j,t-1} & \text{if } z_{i,t-1}, z_{j,t-1} < 0.
\end{cases}
\end{align*}
\]

The parameters contained in the copula can be represented by the following matrices:

\[
\bar{\rho}^{(1)} = \begin{bmatrix} \hat{\rho}_{ij}^{(1)} \end{bmatrix}, \quad \bar{\rho}^{(2)} = \begin{bmatrix} \hat{\rho}_{ij}^{(2)} \end{bmatrix}, \quad \bar{\rho}^{(3)} = \begin{bmatrix} \hat{\rho}_{ij}^{(3)} \end{bmatrix}, \quad \bar{\rho}^{(4)} = \begin{bmatrix} \hat{\rho}_{ij}^{(4)} \end{bmatrix},
\]

\[
\alpha = (\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \alpha^{(4)}), \quad \beta = (\beta^{(1)}, \beta^{(2)}, \beta^{(3)}, \beta^{(4)}), \quad \gamma = (\gamma_1, \ldots, \gamma_N)
\]

and \( \eta \) is a scalar parameter.

There are some interesting points to note about the ASTC-DTGARCH model:

- Asymmetry is captured in both mean and variance equations of the marginals by having an additional autoregressive and GARCH process on either side of a threshold value \( r_i \). The skewness and excess kurtosis of each univariate time series is captured via each \( \lambda_i \) and \( \nu_i \) in the univariate skew-\( t \) distribution, thus extending
the DTGARCH model of Chen et al. (2005) who do not consider skewness.

- The copula correlation equations have a similar mean-reverting structure to that of Tse and Tsui (2002), but extended with asymmetry captured via four sets of parameters according to the signs of the previous standardised residuals $z_{i,t-1}$ and $z_{j,t-1}$. This extends models such as the ADC model of Kroner and Ng (1998) and AG-DCC model of Cappiello et al. (2006) to allow for correlation asymmetry when the residuals are of opposite sign.

- For each of the four correlation regimes, each equation is subject to the variance targeting constraint (see for e.g. Engle and Mezrich, 1996) where the long-run correlation matrix is equal to $\bar{P}^{(k)}$, $k = 1, 2, 3, 4$. In a Bayesian analysis, there is added flexibility whereby these matrices can either be estimated, or set to the following sample values:

$$
\bar{P}^{(1)} = \frac{1}{n_1} \sum_{t=1}^{n_1} z_{i,t}^+ z_{i,t}^{+\prime}, \quad \bar{P}^{(2)} = \frac{1}{n_2} \sum_{t=1}^{n_2} z_{i,t}^+ z_{i,t}^{-\prime},
$$

$$
\bar{P}^{(3)} = \frac{1}{n_3} \sum_{t=1}^{n_3} z_{i,t}^- z_{i,t}^{+\prime}, \quad \bar{P}^{(4)} = \frac{1}{n_4} \sum_{t=1}^{n_4} z_{i,t}^- z_{i,t}^{-\prime},
$$

where $z_{i,t}^+$ and $z_{i,t}^-$ denote the vectors of positive and negative standardised residuals (respectively), while $n_k$ denotes the sample size of each corresponding sign combination, with $n = n_1 + n_2 + n_3 + n_4$. A similar technique has been applied to symmetric forms of conditional correlation (e.g. see Engle, 2002; Tse and Tsui, 2002 and Cappiello et al., 2006 for multivariate GARCH models). In the application of the model in this chapter, these correlation matrices will be set to these sample values above.

### 5.5.3 Parameter Restrictions and the Likelihood Function

To ensure PD and CS in any copula-GARCH model, restrictions are typically placed on the parameters. In previous chapters of this thesis, the approach taken has been to
“relax” these typical parameter restrictions by utilising the unconditional properties of each multivariate GARCH model and ensuring that each conditional covariance matrix at each time point had positive eigenvalues, hence generating necessary and sufficient conditions. However, for the ASTC-DTGARCH model, standard sufficient parameter restrictions that are typically used in the literature will be applied instead. This is because the model is defined in terms of univariate GARCH models (the marginals) and a copula - sufficient conditions typically placed on GARCH model parameters do not cut off a significant part of the parameter space, and relaxing them to be necessary also would further increase computational expense for minimal estimation improvement. As an example, consider a simple GARCH(1,1) model with conditional variance equation

\[ h_t = c + a \varepsilon_{t-1}^2 + bh_{t-1}. \]

Standard sufficient parameter restrictions \( c > 0, a, b \geq 0 \) and \( a + b < 1 \) are typically applied in the literature to ensure that the unconditional variance \( E(h_t) = c/(1 - a - b) \) is finite and strictly positive. Necessary and sufficient restrictions can be obtained by only enforcing each \( E(h_t) \) to be finite and strictly positive, which now means that \( c \) can be negative, as long as \( a + b > 1 \). The problem with this is that discontinuities arise in the parameter space, i.e. at \( c = 0 \) and \( a + b = 1 \), and so estimation algorithms may have difficulty in converging to a solution for a minimal parameter space expansion. Since the marginals of the ASTC-DTGARCH model have a similar form, standard sufficient parameter restrictions similar to that above will be applied during estimation.

The parameter restrictions used for the ASTC-DTGARCH model are given below (for \( i = 1, \ldots, N \)):

- \( |m_{i1}^{(1)}|, |m_{i1}^{(2)}| < 1 \),
- \( c_{i1}^{(1)}, c_{i1}^{(2)} > 0, a_{i1}^{(1)}, a_{i1}^{(2)}, b_{i1}^{(1)}, b_{i1}^{(2)} \geq 0, a_{i1}^{(1)} + b_{i1}^{(1)} < 1, a_{i1}^{(2)} + b_{i1}^{(2)} < 1 \),
- \( P_{i,q} < r_i < P_{i,100-q}, \lambda_i > 0, \nu_i > 2 \),
- \( \alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \alpha^{(4)}, \beta^{(1)}, \beta^{(2)}, \beta^{(3)}, \beta^{(4)} \geq 0 \),
\[
\alpha^{(1)} + \beta^{(1)} \leq 1, \quad \alpha^{(2)} + \beta^{(2)} \leq 1, \quad \alpha^{(3)} + \beta^{(3)} \leq 1, \quad \alpha^{(4)} + \beta^{(4)} \leq 1,
\]
\[
\gamma_i > 0, \quad \eta > 2,
\]
where \( P_{i,q} \) denotes the \( q \)th sample percentile of dataset \( y_{it} \). This restriction ensures that enough data points are considered in each regime to generate meaningful inferences, as in Chen et al. (2005) and Chen and So (2006). If the correlation matrices \( \bar{P}^{(k)}, k = 1, 2, 3, 4 \) are to be estimated, then an additional restriction would be that each \( \bar{P}^{(k)} \) is PD. The assumption of non-negativity of the parameters \( \alpha^{(k)} \) and \( \beta^{(k)} \) has been used to coincide with similar restrictions used for the one-regime models of Tse and Tsui (2002) and Ausin and Lopes (2010).

Let \( \theta_{Mi} = (\mu_i, c_i, a_i, b_i, r_i, \lambda_i, \nu_i) \) denote the parameter vector for each marginal and let \( \theta_C = (\bar{P}^{(1)}, \bar{P}^{(2)}, \bar{P}^{(3)}, \bar{P}^{(4)}, \alpha, \beta, \gamma, \eta) \) denote the parameter vector for the copula, with \( \theta = (\theta_{M1}, \ldots, \theta_{MN}, \theta_C) \) denoting the full parameter vector. As in equation (5.19), the (log) likelihood function for the ASTC-DTGARCH model can be written in terms of its marginal and copula components as follows:

\[
\ln p(y|\theta) = \sum_{t=1}^{n} \left[ \ln c_t(u_{1t}, \ldots, u_{Nt}|\psi_{t-1}, \theta_C) + \sum_{i=1}^{N} \ln f_{it}(y_{it}|\psi_{t-1}, \theta_{Mi}) \right] = \ln p(u|\theta_C) + \sum_{i=1}^{N} \ln p(y_i|\theta_{Mi}),
\]

where \( u = \{(u_{1t}, \ldots, u_{Nt})\}_{t=1}^{n}, \ y_i = (y_{i1}, \ldots, y_{in}), \) and each \( \ln p(y_i|\theta_{Mi}) \) is the marginal log likelihood given by

\[
\ln p(y_i|\theta_{Mi}) = n \ln(2\lambda_is_i) + n \ln \Gamma\left(\frac{\nu_i + 1}{2}\right) - n \ln(1 + \lambda_i^2) - n \ln \Gamma\left(\frac{\nu_i}{2}\right)
- \frac{n}{2} \ln[(\nu_i - 2)\pi] - \frac{1}{2} \sum_{t=1}^{n} \ln(h_{it}) - \frac{\nu_i + 1}{2} \sum_{t=1}^{n} \ln \left[ 1 + \frac{\lambda_i^{-2h_t}(m_i\sqrt{h_t} + s_i\varepsilon_{it})^2}{(\nu_i - 2)h_t} \right],
\]

while \( \ln p(u|\theta_C) \) is the copula log likelihood given by
\[
\ln p(u|\theta_C) = \frac{-1}{2} \sum_{t=1}^{n} \ln |P_t| + n \ln \Gamma \left( \frac{\eta + N}{2} \right) + n(N - 1) \ln \Gamma \left( \frac{\eta}{2} \right) \\
- \frac{\eta + N}{2} \sum_{t=1}^{n} \ln \left( 1 + \frac{a_t^i a_t}{\eta - 2} \right) - nN \ln \Gamma \left( \frac{\eta + 1}{2} \right) \\
+ \frac{\eta + 1}{2} \sum_{t=1}^{n} \sum_{i=1}^{N} \ln \left[ 1 + \frac{-2I_t^c (m^C_i + s^C_i x_{it})^2}{\eta - 2} \right].
\]

5.6 Bayesian Estimation Procedures

As in Chapters 3 and 4, estimating parameters of the ASTC-DTGARCH model from a Bayesian perspective involves simulation from the joint posterior distribution \(p(\theta|y)\). While Ausin and Lopes (2010) utilises a one-step Bayesian approach for this simulation on a bivariate series, extending this to larger dimensions would increase computational expense, rendering this estimation infeasible for uses such as in a trading environment. Instead, their two-step Bayesian approach will be applied which, analogous to the two-step classical approach, splits the problem into the model’s marginal and copula components. Simulations via a Gibbs sampler are then performed from each marginal posterior distribution, whose information is used to then simulate from the copula posterior distribution.

5.6.1 The Gibbs Sampler

This two-step Bayesian procedure can be summarised as follows (\(\forall i = 1, \ldots, N\)):

1. Simulate from each marginal posterior distribution \(p(\theta_{M_i}|y)\), resulting in the set of samples \(\hat{\theta}_M = \left\{ \theta^{[j]}_{M_1}, \ldots, \theta^{[j]}_{M_N} \right\}_{j=1}^{J}\). To obtain this sample, the following joint conditional posteriors have been chosen for the simulation, after consideration of the mixing and correlations in the MCMC iterates (i.e. “blocking”):

   1. \(p(m^{(1)}_{i0}, m^{(2)}_{i0}, m^{(1)}_{i1}, m^{(2)}_{i1}|y, \theta_{\neq m_0})\)
   2. \(p(c^{(1)}_{i1}, c^{(2)}_{i1}, a^{(1)}_{i1}, a^{(2)}_{i1}, b^{(1)}_{i1}, b^{(2)}_{i1}|y, \theta_{\neq c,a,b_0})\)
   3. \(p(r_i, \lambda_i, \tau_i|y, \theta_{\neq r,\lambda,\tau})\), \(\tau_i = \nu_i^{-1}\)
As in Chapter 4, the degrees of freedom parameter \( \nu_i \) has been inverted because \( \tau_i \) can take values on the finite interval \((0, 0.5)\). Inference on \( \nu_i \) is then made by simply inverting the MCMC iterations for \( \tau_i \).

2. Form estimates of the standardised residuals and the marginal skewness and degrees of freedom parameters by averaging over the MCMC iterations, i.e.

\[
\hat{z}_{it} = \frac{1}{J} \sum_{j=1}^{J} \frac{\varepsilon_{ij}^{[j]}}{\sqrt{h_{it}^{[j]}}}, \quad \hat{\lambda}_i = \frac{1}{J} \sum_{j=1}^{J} \Lambda_i^{[j]}, \quad \hat{\nu}_i = \frac{1}{J} \sum_{j=1}^{J} \nu_i^{[j]} = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{\tau_i^{[j]}}.
\]

Upon forming the marginal CDF values \( \hat{u}_{it} = \text{Skt}_{\hat{\lambda}_i, \hat{\nu}_i}(\hat{z}_{it}) \), set \( \hat{x}_{it} = \text{Skt}_{-1}(\hat{u}_{it}) \) and simulate from the copula posterior distribution \( p(\theta_C | y, \hat{\theta}_M) \), resulting in the set of samples \( \hat{\theta}_C = \{ \theta_C^{[j]} \}_{j=1}^{J} \). This simulation is performed by using the following joint conditional posteriors:

1. \( p(\alpha^{(1)}, \beta^{(1)}, \alpha^{(2)}, \beta^{(2)}, \alpha^{(3)}, \beta^{(3)}, \alpha^{(4)}, \beta^{(4)} | y, \theta \neq \alpha, \beta, \hat{\theta}_M) \)

2. \( p(\gamma_1, \ldots, \gamma_N | y, \theta \neq \gamma, \hat{\theta}_M) \)

3. \( p(\kappa | y, \theta \neq \kappa, \hat{\theta}_M), \quad \kappa = \eta^{-1} \)

As in the marginal posteriors, the copula degrees of freedom parameter \( \eta \) has been inverted to obtain values on a finite interval.

The simulations \( (\hat{\theta}_M, \hat{\theta}_C) \) will then be an approximate dependent sample from the joint posterior distribution \( p(\theta | y) \).

### 5.6.2 Prior and Posterior Distributions

Let \( p(\theta_k | y, \theta \neq k), \ k = 1, \ldots, 3N + 3 \) denote each conditional posterior described above. Using Bayes’ rule, the posteriors for the marginals and copula can be written as
\[ p(\theta_k | y, \theta_{\neq k}) \propto p(y_i | \theta_{M_i}) p(\theta_k | \theta_{\neq k}), \quad k = 1, \ldots, 3N; \]
\[ p(\theta_k | y, \theta_{\neq k}) \propto p(\hat{u} | \theta_{C}, \hat{\theta}_M) p(\theta_k | \theta_{\neq k}), \quad k = 3N + 1, \ldots, 3N + 3, \]

where \( p(y_i | \theta_{M_i}) \) and \( p(\hat{u} | \theta_{C}, \hat{\theta}_M) \) are the marginal and copula likelihoods (respectively), and \( p(\theta_k | \theta_{\neq k}) \) is each corresponding prior distribution. As in Chapters 3 and 4, these priors are chosen to be uniform, i.e.

\[ p(\theta_k | \theta_{\neq k}) \propto I(A), \quad k = 1, \ldots, 3N + 3 \]

where \( A \) is given by the corresponding parameter restrictions from Section 5.5.3. As each conditional posterior is now proportional to either a marginal or copula likelihood of unknown form, MH and DR algorithms are now required for simulation:

### 5.6.3 Conditional Posterior Sampling

As in Chapters 3 and 4, adaptive MCMC sampling is to be performed that applies the Metropolis-Hastings (MH) and Delayed Rejection (DR) algorithms in order to simulate from the above conditional posterior distributions. Exactly the same procedure will be used as in Chapter 4, but it is repeated here briefly for convenience. For further details, refer to Section 4.4 of this thesis. For notational purposes, the \( j \)th simulation for the \( k \)th parameter block will be called \( \theta_k^{[j]} \).

#### The Burn-In Period

Begin with some initial values \( \theta_k^{[1]}, \quad k = 1, \ldots, 3N + 3 \). For the burn-in iterations \( j = 1, \ldots, w \) do the following:

1. Draw a random sample \( \theta_k^{p1} \) from the proposal distribution

\[ g_1(\theta_k | \theta_k^{[j-1]}) \sim N(\theta_k^{[j-1]}, \Omega_k) \]
where $\Omega_k$ is a diagonal matrix with “large” diagonal entries (see Chapters 3 and 4), and $c_{kj}^i$ is a real number initially set to 1.

2. Draw a random sample $u_i^{[j]}$ from Unif[0, 1] and set the following Metropolis acceptance probability:

$$
\alpha_1(\theta_k^{[j-1]}, \theta_k^{p_1}) = \min \left\{ 1, \frac{p(\theta_k^{p_1} | y, \theta_{\neq k})}{p(\theta_k^{[j-1]} | y, \theta_{\neq k})} \right\}.
$$

3. If $u_i^{[j]} < \alpha_1$, set $\theta_k^{[j]} = \theta_k^{p_1}$. Otherwise, set $\theta_k^{[j]} = \theta_k^{[j-1]}$. Return to Step 1 for the next $k$.

Tuning of the proposal covariance matrix $\Omega_k$ is also carried out by changing the value of $c_{kj}^i$ at regular intervals so that acceptance rates fall between 0.15 and 0.5 - refer to Chapters 3 and 4 for the necessary details.

The Sampling Period

Let $\tilde{\theta}_k$ and $\hat{\Sigma}_k$ denote the sample means and covariances of the iterations in the burn-in period for each parameter group $k$. For iterations $j = w + 1, \ldots, J$ do the following:

1. Draw a random sample $\theta_k^{p_1}$ from the proposal distribution:

$$
g_1(\theta_k | \theta_{[j-1]}^k) \sim N(\tilde{\theta}_k, \hat{\Sigma}_k).
$$

2. Draw a random sample $u_i^{[j]}$ from Unif[0, 1] and set the following MH acceptance probability:

$$
\alpha_1(\theta_k^{[j-1]}, \theta_k^{p_1}) = \min \left\{ 1, \frac{p(\theta_k^{p_1} | y, \theta_{\neq k})g_1(\theta_k^{[j-1]} | \theta_k^{p_1})}{p(\theta_k^{[j-1]} | y, \theta_{\neq k})g_1(\theta_k^{p_1} | \theta_k^{[j-1]})} \right\}.
$$

3. If $u_i^{[j]} < \alpha_1$, set $\theta_k^{[j]} = \theta_k^{p_1}$ and return to Step 1 for the next $k$. Otherwise, perform a delayed rejection step:
4. Draw a random sample $\theta_k^{p2}$ from the proposal distribution:

$$g_2(\theta_k | \theta_k^{[j-1]}, \theta_k^{p1}) \sim N \left( \theta_k^{[j-1]}, \frac{1}{2} \hat{\Sigma}_k \right).$$

5. Draw a random sample $u_{[j]}^2$ from Unif[0, 1] and set the following acceptance probability:

$$\alpha_2(\theta_k^{[j-1]}, \theta_k^{p1}, \theta_k^{p2}) = \min\left\{ 1, \frac{N_2}{D_2} \right\},$$

where

$$N_2 = \frac{p(\theta_k^{p2} | y, \theta \neq k)g_1(\theta_k^{p1} | \theta_k^{p2})g_2(\theta_k^{[j-1]} | \theta_k^{p2}, \theta_k^{p1})[1 - \alpha_1(\theta_k^{p2}, \theta_k^{p1})]}{p(\theta_k^{[j-1]} | y, \theta \neq k)g_1(\theta_k^{p1} | \theta_k^{[j-1]})g_2(\theta_k^{p2} | \theta_k^{[j-1]}, \theta_k^{p1})[1 - \alpha_1(\theta_k^{[j-1]}, \theta_k^{p1})]}.$$

6. If $u_{[j]}^2 < \alpha_2$, set $\theta_k^{[j]} = \theta_k^{p2}$. Otherwise, set $\theta_k^{[j]} = \theta_k^{[j-1]}$. Return to Step 1 for the next $k$.

The set of samples $\{\theta_k^{[w+1]}, \ldots, \theta_k^{[j]}\}_{k=1}^{3N+3}$ is then an approximate dependent sample from the joint posterior distribution $p(\theta | y)$ of the ASTC-DTGARCH model.

### 5.7 Model Performance: A Stock Portfolio Example

In this section, the ASTC-DTGARCH model will be fitted to a portfolio of stock returns, and its performance measured by its ability to calculate Value at Risk (VaR). Several VaR back-testing procedures will be utilised for this analysis, with comparisons drawn between other existing models in the literature.

#### 5.7.1 The Data

The data consists of the following five highly-traded Dow Jones Industrial (DJI) stocks: Cisco Systems, Inc. (CSCO), Exxon Mobil Corporation (XOM), General Electric Company (GE), Intel Corporation (INTC) and Microsoft Corporation (MSFT). Stock returns
are formed from the daily closing prices of these stocks via the equation

$$y_t = \ln P_t - \ln P_{t-1},$$

where $y_t = (CSCO_t, XOM_t, GE_t, INTC_t, MSFT_t)$ and $P_t$ is the multivariate daily closing price at time $t$. The data is obtained from Yahoo Finance for the period 11 February 1994 to 31 December 2009, resulting in a total of $n = 4,000$ stock returns. For this example, an equally-weighted portfolio is considered, so that each stock has the same impact on the portfolio. Hence portfolio returns (PR) can be formed via the equation

$$p_t = 0.2(y_{1t} + y_{2t} + y_{3t} + y_{4t} + y_{5t}).$$

Figure 5.1 displays the plots of each of the six stock return series as formed above. Various properties of financial time series can be observed from these plots, such as stationarity, volatility clustering, outliers from extreme market events and skewness. It is also worth noting that by holding a portfolio of the given stocks, the impact of the outliers are less extreme, as indicated by the more stable portfolio return plot.

Table 5.1 displays sample statistics of the six stock return series. Here we can see large negative skewness measures (as low as $-17$) and large positive kurtosis measures (as large as $539.9$) for all stocks, indicating that these moments must be accounted for in any model for these returns. The portfolio returns, as indicated by their plot in Figure 5.1, have much less skewness and kurtosis, but are still significant and in the same direction as the other stocks.

Table 5.2 displays sample correlation coefficients between each stock return series. Three types are presented here:

1. *Pearson’s Product Moment*: This is the standard correlation measure for testing the strength of a linear relationship between two variables;
Figure 5.1: Plots of the DJI stock returns CSCO, XOM, GE, INTC and MSFT, as well as equally-weighted portfolio returns PR between 11 February 1994 and 31 December 2009.
Table 5.1: Sample statistics for the DJI stock returns of CSCO, XOM, GE, INTC and MSFT, as well as their corresponding equally-weighted portfolio returns PR between 11 February 1994 and 31 December 2009.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>CSCO</th>
<th>XOM</th>
<th>GE</th>
<th>INTC</th>
<th>MSFT</th>
<th>PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>−0.0003</td>
<td>0.0000</td>
<td>−0.0005</td>
<td>−0.0003</td>
<td>−0.0002</td>
<td>−0.0003</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0014</td>
<td>0.0005</td>
<td>0.0009</td>
<td>0.0012</td>
<td>0.0010</td>
<td>0.0004</td>
</tr>
<tr>
<td>Skewness</td>
<td>−6.5075</td>
<td>−13.5658</td>
<td>−17.037</td>
<td>−7.7021</td>
<td>−10.467</td>
<td>−1.5838</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>117.83</td>
<td>412.80</td>
<td>539.85</td>
<td>153.04</td>
<td>215.35</td>
<td>18.3868</td>
</tr>
</tbody>
</table>

Table 5.2: Sample correlation coefficients for each bivariate combination of the DJI stock returns CSCO, XOM, GE, INTC and MSFT between 11 February 1994 and 31 December 2009.

<table>
<thead>
<tr>
<th>Return</th>
<th>Correlation</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson</td>
<td></td>
</tr>
<tr>
<td>CSCO</td>
<td>0.1299</td>
<td>0.2276</td>
</tr>
<tr>
<td></td>
<td>Spearman</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Kendall</td>
<td>0.1600</td>
</tr>
<tr>
<td>XOM</td>
<td>Pearson</td>
<td>0.1738</td>
</tr>
<tr>
<td></td>
<td>Spearman</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Kendall</td>
<td>0.2491</td>
</tr>
<tr>
<td>GE</td>
<td>Pearson</td>
<td>0.2170</td>
</tr>
<tr>
<td></td>
<td>Spearman</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Kendall</td>
<td>0.2798</td>
</tr>
<tr>
<td>INTC</td>
<td>Pearson</td>
<td>0.2916</td>
</tr>
<tr>
<td></td>
<td>Spearman</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Kendall</td>
<td>0.4151</td>
</tr>
<tr>
<td>MSFT</td>
<td>Pearson</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spearman</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Kendall</td>
<td></td>
</tr>
</tbody>
</table>

2. *Spearman’s ρ*: This is a non-parametric association measure, whereby Pearson’s product moment correlation is calculated on the ranks of the observations within each variable. The relationship can be non-linear;

3. *Kendall’s τ*: This is another non-parametric association measure, which is a standardised difference between the number of “concordant” pairs (bivariate observations moving in the same direction) and “discordant” pairs (bivariate observations moving in opposite directions). The relationship can be non-linear.
The Pearson correlations are the weakest in this sample, whereas Spearman’s \( \rho \) and Kendall’s \( \tau \) are much higher. This is evidence of the existence of non-linear relationships between each pair of returns in this sample. The highest correlation values can be seen for the return pair of CSCO and INTL, giving 0.37 for Pearson, 0.61 for Spearman and 0.45 for Kendall. Hypothesis testing across all return pairs confirms significance of each correlation. This indicates non-linear dependence between each stock return series, which should be captured in an appropriate multivariate model.

5.7.2 Value at Risk

Value at Risk (VaR) is perhaps the most widely used measure of market risk by risk managers and practitioners. It is defined to be the minimum potential loss of a stock portfolio that could be realised over a specified time horizon with a given level of confidence. For any set of portfolio log returns \( \{p_t\}_{t=1}^n = \{w'y_t\}_{t=1}^n \) of sample size \( n \), VaR can be estimated mathematically for a long position\(^3\) as follows:

\[
\text{VaR}_{\alpha,S} = A \times q_{\alpha,S}
\]

where \( A \) is the investment amount, \( S \) is the time horizon and \( q_{\alpha,S} \) is the \( \alpha \)% quantile of the distribution of \( p_{[S]} = p_{n+1} + p_{n+2} + \ldots + p_{n+S} \). In the analysis to follow, the assumption is made that \( A \) is $1, so that VaR is simply the relevant quantile.

Rather than having a point estimate of VaR as in classical techniques, the estimation of VaR from a Bayesian perspective requires knowledge of the posterior distribution \( p(\text{VaR}_{\alpha,S}|y) \). This distribution can be written as follows:

\[
p(\text{VaR}_{\alpha,S}|y) = \int p(\text{VaR}_{\alpha,S}|y, \theta)p(\theta|y)d\theta. \tag{5.29}
\]

Now \( p(\theta|y) \) is simply the parameter posterior distribution, from which simulations have

\(^3\)When an investor holds a long position, an unfavourable impact on their portfolio occurs when stock prices fall.
already been obtained. Hence simulations \( \left\{ \text{VaR}_{\alpha, S}^{(j)} \right\}_{j=1}^J \) from the distribution in (5.29) can be made using the simulations \( \left\{ \theta^{(j)} \right\}_{j=1}^J \) from \( p(\theta | y) \), then a point estimate of VaR can be formed by using the sample mean of the VaR sample:\(^4\)

\[
\text{VaR}_{\alpha, S} = E(\text{VaR}_{\alpha, S} | y) \approx \frac{1}{J} \sum_{j=1}^J \text{VaR}_{\alpha, S}^{(j)}. \quad (5.30)
\]

What is left is the calculation of each \( \text{VaR}_{\alpha, S}^{(j)} \). For the ASTC-DTGARCH model, this can be achieved by utilising a Monte Carlo simulation procedure as in Wong and So (2003), Bauwens and Storti (2009) and Ausin and Lopes (2010) (for example). At time \( n + 1 \), the information set \( I_n \) is known, which includes values for \( y_{in}, \varepsilon_{in}, h_{in}, z_{in}, \xi_{ik,n} \) and \( \rho_{ik,n} \). Now for each parameter simulation \( \theta^{(j)} = \left( \theta_{M_1}^{(j)}, \ldots, \theta_{M_N}^{(j)}, \theta_C^{(j)} \right) \), \( \forall \ j = w + 1, \ldots, J \) the one-step ahead conditional variance and copula correlation values can be formed using the relevant model equations as follows:

\[
h_{i,n+1}^{(j)} = \begin{cases} (c_{i1}^{(1)})^{(j)} + (a_{i1}^{(1)})^{(j)} \varepsilon_{in}^2 + (b_{i1}^{(1)})^{(j)} h_{in} & \text{if } y_{in} \leq r_i^{(j)}, \\ (c_{i1}^{(2)})^{(j)} + (a_{i1}^{(2)})^{(j)} \varepsilon_{in}^2 + (b_{i1}^{(2)})^{(j)} h_{in} & \text{if } y_{in} > r_i^{(j)}, \end{cases}
\]

\[
\rho_{ik,n+1}^{(j)} = \begin{cases} \rho_{ik}^{(1)} [1 - (\alpha^{(1)})^{(j)} - (\beta^{(1)})^{(j)}] + (\alpha^{(1)})^{(j)} \xi_{ik,n} + (\beta^{(1)})^{(j)} \rho_{ik,n} & \text{if } z_{in}, z_{kn} \geq 0, \\ \rho_{ik}^{(2)} [1 - (\alpha^{(2)})^{(j)} - (\beta^{(2)})^{(j)}] + (\alpha^{(2)})^{(j)} \xi_{ik,n} + (\beta^{(2)})^{(j)} \rho_{ik,n} & \text{if } z_{in} \geq 0, z_{kn} < 0, \\ \rho_{ik}^{(3)} [1 - (\alpha^{(3)})^{(j)} - (\beta^{(3)})^{(j)}] + (\alpha^{(3)})^{(j)} \xi_{ik,n} + (\beta^{(3)})^{(j)} \rho_{ik,n} & \text{if } z_{in} < 0, z_{kn} \geq 0, \\ \rho_{ik}^{(4)} [1 - (\alpha^{(4)})^{(j)} - (\beta^{(4)})^{(j)}] + (\alpha^{(4)})^{(j)} \xi_{ik,n} + (\beta^{(4)})^{(j)} \rho_{ik,n} & \text{if } z_{in}, z_{kn} < 0. \end{cases}
\]

Then, for \( l = 1, \ldots, L \), do the following:

1. Simulate \( \left( x_{1,n+1}^{(j)}, \ldots, x_{N,n+1}^{(j)} \right) \) from the multivariate skew-t distribution\(^5\) with zero mean vector, correlation matrix \( F_{n+1}^{(j)} = [\rho_{ik,n+1}^{(j)}] \), skewness vector \( \gamma^{(j)} \) and degrees

---

\(^4\)The quantities Median(\( \text{VaR}_{\alpha, S} | y \)) (the posterior median) or even Mode(\( \text{VaR}_{\alpha, S} | y \)) (the posterior mode) are also location measures of VaR, whose corresponding sample values \( P_{\alpha,S} \{ \text{VaR}_{\alpha, S}^{(j)} \}_{j=1}^J \) and \( \arg \max \{ \text{VaR}_{\alpha, S}^{(j)} \}_{j=1}^J \) could also be used here to obtain point estimates.

\(^5\)The procedure of generating random numbers from this distribution is detailed in the Appendix of this chapter.
of freedom \( \eta^{[j]} \).

2. Set \( \varepsilon_{i,n+1}^{[j]} = \text{Skt}_{\chi_i^{[j]}, \nu_i^{[j]}}^{-1}\left(\text{Skt}_{\gamma_i^{[j]}, \eta_i^{[j]}(x_{i,n+1}^{[j]})}\right) \sqrt{h_{i,n+1}^{[j]}} \) and form return simulations using the DTGARCH equations from the marginals:

\[
y_{i,n+1}^{[j]} = \begin{cases} 
(m_{i0}^{(1)})^{[j]} + (m_{i1}^{(1)})^{[j]} y_{i,n} + \varepsilon_{i,n+1}^{[j]} & \text{if } y_{i,n} \leq r_i^{[j]}, \\
(m_{i0}^{(2)})^{[j]} + (m_{i1}^{(2)})^{[j]} y_{i,n} + \varepsilon_{i,n+1}^{[j]} & \text{if } y_{i,n} > r_i^{[j]}.
\end{cases}
\]

3. Form portfolio return simulations via \( p_{n+1}^{[j]} = w_1 y_{1,n+1}^{[j]} + \ldots + w_N y_{N,n+1}^{[j]} \).

This then yields the set of portfolio return simulations \( \left\{ p_{n+1}^{[j]} \right\}_{i=1}^L \), and \( \text{VaR}_{\alpha,i}^{[j]} \) is simply the \( \alpha \% \) sample quantile of these \( L \) simulations.

To determine each \( \text{VaR}_{\alpha,i}^{[j]} \) where \( s = 2, \ldots, S \), a similar procedure is used, but with the assumption that all variables at the previous time horizon are known values. Hence the following additional steps are utilised:

4. Substitute \( x_{i,n+s-1}^{[j]} \), \( i = 1, \ldots, N \) appropriately into equation (5.25) to obtain \( \xi_{1k,n+s-1}^{[j]} \).

5. Substitute each \( y_{i,n+s-1}^{[j]}, \varepsilon_{i,n+s-1}^{[j]}, h_{i,n+s-1}^{[j]}, \varepsilon_{1k,n+s-1}^{[j]}, \xi_{1k,n+s-1}^{[j]} \) and \( \rho_{1k,n+s-1}^{[j]} \) appropriately into the conditional variance and copula correlation equations above to obtain \( h_{i,n+s}^{[j]} \) and \( \rho_{1k,n+s}^{[j]} \).

6. Simulate \( \left( x_{1,n+s}^{[j]}, \ldots, x_{N,n+s}^{[j]} \right) \) from the multivariate skew-\( t \) distribution with zero mean vector, correlation matrix \( P_{n+s}^{[j]} = [\rho_{1k,n+s}^{[j]}] \), skewness vector \( \gamma^{[j]} \) and degrees of freedom \( \eta^{[j]} \).

7. Set \( \varepsilon_{i,n+s}^{[j]} = \text{Skt}_{\chi_i^{[j]}, \nu_i^{[j]}}^{-1}\left(\text{Skt}_{\gamma_i^{[j]}, \eta_i^{[j]}(x_{i,n+s}^{[j]})}\right) \sqrt{h_{i,n+s}^{[j]}} \) and form return simulations using the DTGARCH equations from the marginals:

\[
y_{i,n+s}^{[j]} = \begin{cases} 
(m_{i0}^{(1)})^{[j]} + (m_{i1}^{(1)})^{[j]} y_{i,n+s-1} + \varepsilon_{i,n+s}^{[j]} & \text{if } y_{i,n+s-1} \leq r_i^{[j]}, \\
(m_{i0}^{(2)})^{[j]} + (m_{i1}^{(2)})^{[j]} y_{i,n+s-1} + \varepsilon_{i,n+s}^{[j]} & \text{if } y_{i,n+s-1} > r_i^{[j]}.
\end{cases}
\]
8. Form portfolio return simulations at the $s$-period horizon using $p_{n+s}^{[j,l]} = w_1 y_{1,n+s}^{[j,l]} + \ldots + w_N y_{N,n+s}^{[j,l]}$ and the equation

$$p_n[s]^{[j,l]} = \sum_{k=1}^{s} p_{n+k}^{[j,l]}.$$ 

Then VaR$_{\alpha,s}^{[j]}$ is simply the $\alpha\%$ quantile of the simulations $\{p_n[s]^{[j,l]}\}_{l=1}^L$.

### 5.7.3 VaR Back-Testing

To monitor the ASTC-DTGARCH model’s performance in estimating VaR for the DJI stock portfolio, a series of VaR back-testing procedures is utilised. The requirements of these back-tests are a set of VaR estimates, as well as actual portfolio returns at the times represented by each VaR. The time period from 24 January 2002 to 31 December 2009 is used as the back-testing period, which corresponds to the last $K = 2,000$ observations in the sample. The following table displays the time horizons considered in the analysis, with the number of non-overlapping VaR estimates generated over the back-testing period:

<table>
<thead>
<tr>
<th>Horizon ($s$)</th>
<th>No. of VaR Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,000</td>
</tr>
<tr>
<td>5</td>
<td>400</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
</tr>
</tbody>
</table>

Two significance levels $\alpha = 0.01, 0.05$ are also considered$^6$. These combinations will assist in examining the model’s short and long-term forecasting performance, as well as its tail behaviour in forecast distributions.

In summary, the following is recursively performed:

---

$^6$A much larger sampling window would be required in order to efficiently perform back-testing for “extreme” risk (i.e. $\alpha < 0.01$) using the Monte Carlo approach considered here.
1. Fit the ASTC-DTGARCH model to the set of returns \{y_t\}_{t=1}^{2000} using the Bayesian estimation procedures from Section 5.6, resulting in the sample \( \hat{\theta} = \{\theta[j]\}_j^{J+1} \) from the joint posterior distribution \( p(\theta | y) \).

2. For each \( k = 1, \ldots, K - 20 \) do the following:

   (a) Simulate from the VaR posterior distribution \( p(\text{VaR}_{\alpha,s} | y) \) using the sample \( \hat{\theta} \) and the procedures from Section 5.7.2, then obtain the VaR estimate \( \text{VaR}_{\alpha,s}^{(K+k)} \) using equation (5.30). This is performed for each time horizon \( s = 1, 5, 10, 20 \) and significance level \( \alpha = 0.01, 0.05 \).

   (b) Obtain the equally-weighted portfolio returns \( p_{K+k}[s] = \sum_{i=1}^{s} p_{K+k+i} \) for each \( s \).

3. If \( k \) is a multiple of 20, return to Step 1 and re-fit the ASTC-DTGARCH model to the set of returns \( \{y_t\}_{t=1+K}^{K+k} \) to generate a new set of samples \( \hat{\theta} \). Otherwise, return to Step 2 for the next \( k \).

This procedure will then generate sets of non-overlapping VaR estimates and actual portfolio returns for each \( \alpha, s \) and time period considered above.

For a VaR model to perform well, there are two main requirements:

- The percentage of times that the actual portfolio return exceeds the corresponding VaR estimate (i.e. when \( p_{K+k}[s] < \text{VaR}_{\alpha,s}^{(K+k)} \), called a “VaR violation”) should not be significantly different from \( \alpha \), i.e. \( \Pr\left(p_{K+k}[s] < \text{VaR}_{\alpha,s}^{(K+k)}\right) = \alpha \);

- VaR violations should occur at independent times over the back-testing period.

These requirements have been formalised in four VaR back-testing procedures in the literature, each of which will now be described. (Further properties of these tests can be found in Berkowitz et al., 2011 and Gaglianone et al., 2011, for example):

**The Unconditional Coverage Test**

Kupiec (1995) formalises the Unconditional Coverage (UC) test, which is a likelihood ratio test that the mean level of VaR violations is equal to the coverage probability (the
significance level). This test first requires the following indicator series from the collection of VaR estimates and portfolio returns:

$$V_{k+1} = \begin{cases} 
1 & \text{if } p_{K+k}^s < \text{VaR}_{\alpha,s}^{(K+k)}, \\
0 & \text{otherwise}, 
\end{cases} \quad k = 0, \ldots, K - 1.$$ 

Given the significance level $\alpha$, the test has the following null ($H_0$) and alternative ($H_1$) hypotheses:

$$H_0 : \ E(V_k) = \alpha,$$

$$H_1 : \ E(V_k) \neq \alpha.$$ 

The test statistic for this test uses a ratio of binomial likelihoods for the true and estimated coverage probabilities: Let $\hat{\alpha} = n_1/K$, where $n_1 = \sum_{k=1}^{K} V_k$. Using the binomial likelihoods

$$L(\alpha) = \alpha^{n_1}(1 - \alpha)^{K-n_1} \quad \text{and} \quad L(\hat{\alpha}) = \hat{\alpha}^{n_1}(1 - \hat{\alpha})^{K-n_1},$$

the test statistic $TS_{uc}$ is given by

$$TS_{uc} = -2 \ln \left[ \frac{L(\alpha)}{L(\hat{\alpha})} \right] \sim \chi_1^2$$

where $\chi_1^2$ denotes the chi-squared distribution with $\nu$ degrees of freedom. Therefore the $p$-value for this test can be calculated as $p_{uc} = 1 - F_{\chi_1^2}(TS_{uc})$, where $F$ denotes the CDF of the corresponding chi-squared distribution. If $p_{uc} < \alpha$, then the null hypothesis $H_0$ is rejected in favour of the alternative $H_1$.

**The Conditional Coverage Test**

While the UC test examines the number of VaR violations against the corresponding significance level, it does not take into account the timing of each violation, i.e. when
each VaR violation occurs. The Conditional Coverage (CC) test of Christoffersen (1998) extends the previous test by additionally testing for independence between VaR violations contained in \( \{V_k\} \).

The null and alternative hypotheses are

\[
H_0 : \quad E(V_k) = \alpha \quad \text{and} \quad \{V_k\} \text{ is independent of } \{V_{k-1}\},
\]

\[
H_1 : \quad E(V_k) \neq \alpha \quad \text{or} \quad \{V_k\} \text{ is not independent of } \{V_{k-1}\}.
\]

This testing involves checking whether or not \( \{V_k\} \) is a first-order Markov chain with the following transition probability matrix:

\[
A = \begin{pmatrix}
1 - \alpha_{01} & \alpha_{01} \\
1 - \alpha_{11} & \alpha_{11}
\end{pmatrix}
\]

where \( \alpha_{ij} = \Pr(V_k = j|V_{k-1} = i) \). For independence, we require that \( \alpha_{01} = \alpha_{11} \), which makes the probability of obtaining a violation at time \( k \) constant and therefore independent of what occurred at time \( k - 1 \). Estimates of these probabilities can be obtained by first counting the different occurrences of \((V_k, V_{k-1})\) as follows:

\[
n_{00} = \sum_{k=2}^{K} I(V_k = 0|V_{k-1} = 0), \quad n_{01} = \sum_{k=2}^{K} I(V_k = 1|V_{k-1} = 0),
\]

\[
n_{10} = \sum_{k=2}^{K} I(V_k = 0|V_{k-1} = 1), \quad n_{11} = \sum_{k=2}^{K} I(V_k = 1|V_{k-1} = 1),
\]

where \( n_{00} + n_{01} + n_{10} + n_{11} = K - 1 \). Then form the relevant ratios to obtain the following estimates:

\[
\hat{\alpha}_{01} = \frac{n_{01}}{n_{00} + n_{01}} \quad \text{and} \quad \hat{\alpha}_{11} = \frac{n_{11}}{n_{10} + n_{11}}.
\]

The test statistic again contains a ratio of likelihoods, but this time extended via a multinomial likelihood:
\[ L(\hat{\alpha}_{01}, \hat{\alpha}_{11}) = \hat{\alpha}_{01}^{n_{01}} (1 - \hat{\alpha}_{01})^{n_{00}} \hat{\alpha}_{11}^{n_{11}} (1 - \hat{\alpha}_{11})^{n_{10}}. \]

Combining this with the binomial likelihood in equation (5.31), the test statistic \( TS_{cc} \) is given by

\[
TS_{cc} = -2 \ln \left[ \frac{L(\alpha)}{L(\hat{\alpha}_{01}, \hat{\alpha}_{11})} \right] \sim \chi^2_2.
\]

The p-value for this test is then \( p_{cc} = 1 - F_{\chi^2_2}(TS_{cc}) \). If \( p_{cc} < \alpha \), then the null hypothesis \( H_0 \) is rejected in favour of the alternative \( H_1 \).

**The Dynamic Quantile Test**

The disadvantage of using the CC test is that dependence is only identified when \( V_k = V_{k-1} \); situations where \( V_k = V_{k-2} \) are thus deemed independent when this is most likely not the case over a large back-testing sample. Engle and Manganelli (2004) propose the Dynamic Quantile (DQ) test which has the ability to test for independence between as many lags as required. It uses the idea of a “hit” function which subtracts the significance level \( \alpha \) away from each \( V_k \) as follows:

\[
\text{hit}_k = \begin{cases} 
1 - \alpha & \text{if } V_k = 1, \\
-\alpha & \text{otherwise}.
\end{cases}
\]

If \( \text{hit}_k \) is uncorrelated with its own lagged values and VaR estimates, then there will be no autocorrelation in the “hits” - see Engle and Manganelli (2004) for more details.

The simplest way to test this is through a multiple regression analysis using the required explanatory variables to be tested. Four “hit” lags and one VaR lag have been chosen for this analysis, as recommended by Engle and Manganelli (2004) and also used by Chen et al. (2011b). This means that the multiple regression equation is

\[
\text{hit}_k = \beta_0 + \beta_1 \text{hit}_{k-1} + \beta_2 \text{hit}_{k-2} + \beta_3 \text{hit}_{k-3} + \beta_4 \text{hit}_{k-4} + \beta_5 \text{VaR}_{\alpha,s}^{(k-1)} + \epsilon_k, \quad (5.32)
\]
where $\beta_0, \ldots, \beta_5$ are the regression parameters and $\varepsilon_k$ is a normally distributed residual with zero mean and constant variance. Hence the DQ test reduces to testing if each of the $\beta$ parameters above are significantly different from zero, giving the following null and alternative hypotheses:

$$
H_0 : \quad \beta_0 = \beta_1 = \ldots = \beta_5 = 0,
$$

$$
H_1 : \quad \text{Not all } \beta_i\text{'s are zero.}
$$

As in standard multiple regression analysis, we can write equation (5.32) in matrix form as $\text{Hit} = X\beta + \varepsilon$, where

$$
\text{Hit} = (\text{hit}_5, \ldots, \text{hit}_K), \quad \beta = (\beta_0, \ldots, \beta_5), \quad \varepsilon = (\varepsilon_5, \ldots, \varepsilon_K),
$$

$$
X = \begin{pmatrix}
1 & \text{hit}_4 & \text{hit}_3 & \text{hit}_2 & \text{hit}_1 & \text{VaR}_{\alpha,s}^{(K+4)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & \text{hit}_{K-1} & \text{hit}_{K-2} & \text{hit}_{K-3} & \text{hit}_{K-4} & \text{VaR}_{\alpha,s}^{(2K-1)}
\end{pmatrix}.
$$

The test statistic $TS_{dq}$ for this test is given by

$$
TS_{dq} = \frac{\text{Hit}'X'(X'X)^{-1}X'\text{Hit}}{\alpha(1 - \alpha)} \sim \chi^2_6.
$$

The $p$-value for this test is then $p_{dq} = 1 - F_{\chi^2_6}(TS_{dq})$. If $p_{dq} < \alpha$, then the null hypothesis $H_0$ is rejected in favour of the alternative $H_1$.

**The VaR Criterion Function**

In addition to the DQ test, Engle and Manganelli (2004) also specify the evolution of the VaR quantile over time using a model based on an autoregressive process - the Conditional Autoregressive Value at Risk (CAViaR) model. This model is estimated in this paper using the regression quantile framework of Koenker and Bassett (1978), whereby parameters are estimated based on the minimum of a VaR criterion function. In fact,
this criterion function can be applied to any model used for VaR estimation.

For the sample of VaR estimates and portfolio returns given earlier, the VaR criterion function of Koenker and Bassett (1978) is given by

$$
CF = \alpha \sum_{k=1}^{K+1} \left(p_{K+k}[s] - \text{VaR}_{\alpha,s}^{(K+k)}\right) \\
+ (1 - \alpha) \sum_{k=1}^{K+1} \left(\text{VaR}_{\alpha,s}^{(K+k)} - p_{K+k}[s]\right)
$$

and will be minimised if the series \(\left\{\text{VaR}_{\alpha,s}^{(K+k)}\right\}_{k=0}^{K-1}\) is the true \(\alpha\)% quantile of the series \(\left\{p_{K+k}[s]\right\}_{k=0}^{K-1}\).

5.7.4 Competing Models

In order to decide the best VaR model for the given sample of DJI stocks, a series of competing models is also fitted to the data and VaR estimated using classical techniques. Both univariate and multivariate GARCH models are considered, ranging from simpler models (e.g. symmetric with normal errors) to more complex ones (e.g. allowing for volatility asymmetry, skewness and excess kurtosis). Please note that some of these models will already have been discussed in this thesis, but are repeated here for convenience.

The GARCH Model

For the portfolio return series \(\{p_t\} = \{0.2(y_{1t} + y_{2t} + y_{3t} + y_{4t} + y_{5t})\}\), the GARCH model of Bollerslev (1986) is given by (with AR(1) mean effects):

$$
p_t = m_0 + m_1 p_{t-1} + \varepsilon_t, \quad \varepsilon_t|\psi_{t-1} \sim D(0, h_t) \\
h_t = c_1 + a_1 \varepsilon_{t-1}^2 + b_1 h_{t-1},
$$

where \(D\) denotes either the normal, standardised Student-\(t\) or skew-\(t\) distributions with zero mean and variance \(h_t\). In the results that follow, these models are given the labels “GARCH\(_N\)”, “GARCH\(_T\)” and “GARCH\(_S\)” (respectively).
The GJR-GARCH Model

This model is an extension of the GARCH model above to allow for asymmetric effects in volatility, as in Glosten et al. (1993). It is given by (with AR(1) mean effects):

\[ p_t = m_0 + m_1 p_{t-1} + \varepsilon_t, \quad \varepsilon_t | \psi_{t-1} \sim D(0, h_t) \]

\[ h_t = c_1 + (a_1 + \alpha_1 I_{t-1}) \varepsilon_{t-1}^2 + b_1 h_{t-1}, \]

where \( D \) denotes either the normal, standardised Student-\( t \) or skew-\( t \) distributions with zero mean and variance \( h_t \), while \( I_t \) is the indicator variable such that

\[ I_t = \begin{cases} 1 & \text{if } \varepsilon_t < 0, \\ 0 & \text{otherwise}. \end{cases} \]

In the results that follow, these models are given the labels “GJR-GARCH\( N \)”, “GJR-GARCH\( T \)” and “GJR-GARCH\( S \)” (respectively).

The DCC Model

The DCC model of Engle (2002) specifies the conditional distribution of the \( N \)-dimensional vector \( y_t = (y_{1t}, \ldots, y_{Nt}) \) as follows

\[ y_t = m_0 + \text{diag}(m_1) y_{t-1} + \varepsilon_t, \quad \varepsilon_t | \psi_{t-1} \sim N(0, D_t P_t D_t) \]

where \( m_0 \) and \( m_1 \) are \( N \times 1 \) parameter vectors, \( D_t = \text{diag}(\sqrt{h_{11,t}}, \ldots, \sqrt{h_{NN,t}}) \) is a diagonal matrix of conditional standard deviations (with each \( h_{ii,t} \) defined as a GARCH(1,1) model as in (5.34)), while \( P_t = [\rho_{ij,t}] \) is a correlation matrix defined as

\[ \rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}, \]

where

\[ q_{ij,t} = \tilde{\rho}_{ij} (1 - \alpha - \beta) + \alpha z_{i,t-1} z_{j,t-1} + \beta q_{ij,t-1} \]
and each \( z_{it} = \varepsilon_{it}/\sqrt{h_{ii,t}} \) is a standardised residual. The two-step estimation approach as detailed in Engle (2002) is applied to this model in the analysis to come - also refer to Section 3.2 of this thesis for more details of this model. In the results that follow, this model is given the label “DCC”.

**The GJR-ADCC Model**

The GJR-ADCC model is an extension of the DCC model to have GJR asymmetries in the conditional variances, as well as asymmetries in the conditional correlations, as in Cappiello et al. (2006). The model has the same AR(1) mean structure as in the DCC model from (5.37), but with the following structure for the correlations:

\[
q_{ij,t} = \bar{\rho}_{ij}(1 - \alpha^2 - \beta^2) - \gamma^2 \bar{\eta}_{ij} + \bar{\rho}_{ij}(1 - \alpha^2) \varepsilon_{it} - 1 \varepsilon_{jt} - 1 + \gamma^2 \bar{\eta}_{it-1} \bar{\eta}_{jt-1} + \beta^2 q_{ij,t-1}
\]

where each \( \eta_{it} = I_{it}z_{it} \), with \( I_{it} \) defined in equation (5.36). A similar two-step estimation approach as for the DCC model is applied to this model also - see Cappiello et al. (2006) and also Section 4.2 for further details. In the results that follow, this model is given the label “GJR-ADCC”.

**The Copula-GJR-GARCH Model**

There are three Copula-GJR-GARCH models represented in the analysis:

- **Gaussian Copula**: The marginals of this model are each represented by the GJR-GARCH model in equation (5.35) with normal errors, while the copula used for the dependence structure is the Gaussian copula given in equations (5.15) and (5.16).

- **Student-t Copula**: The marginals of this model are each represented by the GJR-GARCH model in equation (5.35) with standardised Student-\( t \) errors, while the copula used for the dependence structure is the Student-\( t \) copula given in equations (5.17) and (5.18).
• **Skew-t Copula:** The marginals of this model are each represented by the GJR-GARCH model in equation (5.35) with skew-t errors, while the copula used for the dependence structure is the skew-t copula given in equations (5.22) and (5.24).

The dependence structure of these copulas is defined in terms of the correlation matrix \( P_t = [\rho_{ij,t}] \), whose structure for each of these models is the following:

\[
\rho_{ij,t} = \tilde{\rho}_{ij}(1 - \alpha - \beta) + \alpha \xi_{ij,t-1} + \beta \rho_{ij,t-1}
\]

(5.38)
as in Ausin and Lopes (2010). In the results that follow, these models are given the labels “COP-GJR-GARCH\(_N\)”, “COP-GJR-GARCH\(_T\)” and “COP-GJR-GARCH\(_S\)” (respectively).

**The Asymmetric Copula-GJR-GARCH Model**

This group of models has the same structure as the above Copula-GJR-GARCH models, but extended so that the correlation structure in equation (5.38) allows for asymmetry. This is done via a basic two-regime correlation structure similar to that of the ASTC-DTGARCH model, but only allowing for asymmetry when the standardised residuals are of the same sign. The correlation structure is then

\[
\rho_{ij,t} = \begin{cases} 
\tilde{\rho}_{ij}(1 - \alpha^{(1)} - \beta^{(1)}) + \alpha^{(1)} \xi_{ij,t-1} + \beta^{(1)} \rho_{ij,t-1} & \text{if } z_{i,t-1}, z_{j,t-1} \geq 0, \\
\tilde{\rho}_{ij}(1 - \alpha^{(2)} - \beta^{(2)}) + \alpha^{(2)} \xi_{ij,t-1} + \beta^{(2)} \rho_{ij,t-1} & \text{if } z_{i,t-1}, z_{j,t-1} < 0.
\end{cases}
\]

In the results that follow, these models are given the labels “ACOP-GJR-GARCH\(_N\)”, “ACOP-GJR-GARCH\(_T\)” and “ACOP-GJR-GARCH\(_S\)” (respectively).

**VaR Estimation from a Classical Perspective**

To estimate VaR for the above competing models, a similar Monte Carlo approach to that of the ASTC-DTGARCH model can be employed, but based on the parameter estimate vector \( \theta \) containing only one estimate for each parameter (instead of \( J \) Bayesian
The calculations will vary depending on the model, but can be summarised as follows:

- **Univariate GARCH Models**: For $l = 1, \ldots, L$ and $s = 1, \ldots, S$ do the following:
  1. Calculate $h_{n+s}$ under the assumption that $\varepsilon_{n+s-1}, I_{n+s-1}$ and $h_{n+s-1}$ are known.
  2. Simulate $p_{n+s}^{[l]}$ from the corresponding error distribution with mean $\hat{m}_0 + \hat{m}_1 p_{n+s-1}^{[l]}$ and variance $h_{n+s}$.
  3. Form portfolio return simulations via $p_{n+s}^{[l]} = \sum_{k=1}^{s} p_{n+k}^{[l]}$.

Then VaR_{\alpha, s} is simply the $\alpha\%$ quantile of the simulations $\{p_{n+s}^{[l]}\}_{l=1}^{L}$.

- **DCC and GJR-ADCC Models**: For $l = 1, \ldots, L$ and $s = 1, \ldots, S$ do the following:
  1. Calculate $h_{i,n+s}$ under the assumption that $\varepsilon_{i,n+s-1}, I_{i,n+s-1}$ and $h_{i,n+s-1}$ are known (for each $i = 1, \ldots, N$) and form the diagonal matrix $D_{n+s}$.
  2. Calculate $P_{n+s} = [\rho_{i,j,n+s}]$ under the assumption that $z_{i,n+s-1}, z_{j,n+s-1}$ and $q_{i,j,n+s-1}$ are known.
  3. Simulate $y_{n+s}^{[l]}$ from the multivariate normal distribution with mean vector $\hat{m}_0 + \text{diag}(\hat{m}_1) y_{n+s-1}^{[l]}$ and covariance matrix $D_{n+s} P_{n+s} D_{n+s}$.
  4. Form portfolio return simulations via $p_{n+s}^{[l]} = \sum_{k=1}^{s} w^\prime y_{n+k}^{[l]}$, where $w=(0.2, 0.2, 0.2, 0.2, 0.2)$.

Then VaR_{\alpha, s} is simply the $\alpha\%$ quantile of the simulations $\{p_{n+s}^{[l]}\}_{l=1}^{L}$.

- **Copula-GARCH Models**: For $l = 1, \ldots, L$ and $s = 1, \ldots, S$ do the following:
  1. Calculate $h_{i,n+s}$ under the assumption that $\varepsilon_{i,n+s-1}, I_{i,n+s-1}$ and $h_{i,n+s-1}$ are known (for each $i = 1, \ldots, N$).
  2. Calculate $P_{n+s} = [\rho_{i,j,n+s}]$ under the assumption that $z_{i,n+s-1}, z_{j,n+s-1}, \xi_{i,j,n+s-1}$ and $\rho_{i,j,n+s-1}$ are known.
3. Simulate $x_{i,n+s}^{[l]}$ from the relevant multivariate distribution that the copula was based on, with zero mean vector, correlation matrix $P_{n+s}$ and any skewness and/or degrees of freedom parameters.

4. Set $\epsilon_{i,n+s}^{[l]} = F_{\theta_{M_i}}^{-1}\left(F_{\theta_C}\left(x_{i,n+s}^{[l]}\right)\right) \sqrt{h_{i,n+s}}$ for $i = 1, \ldots, N$ and let $y_{n+s}^{[l]} = \hat{m}_0 + \text{diag}(\hat{m}_1)y_{n+s-1}^{[l]} + \epsilon_{n+s}^{[l]}$, where $\theta_{M_i}$ and $\theta_C$ are the relevant marginal and copula skewness and degrees of freedom parameters, with $F$ and $F^{-1}$ denoting the corresponding univariate CDF and inverse CDF (respectively).

5. Form portfolio return simulations via $p_{n}[s]^{[l]} = \sum_{k=1}^{s} w' y_{n+k}^{[l]}$, where $w=(0.2, 0.2, 0.2, 0.2, 0.2)$.

Then VaR$_{\alpha,s}$ is simply the $\alpha\%$ quantile of the simulations $\{p_{n}[s]^{[l]}\}_{t=1}^{L}$.

### 5.7.5 Parameter Estimation Results

Fitting of the ASTC-DTGARCH model for the VaR back-testing uses a total of $J = 15,000$ iterations for each run, with a burn-in period of $w = 5,000$ for both marginal and copula components. To yield good acceptance rates in the MCMC algorithm, the model is fitted to the percentage log returns (i.e. $100y_t$), and so VaR’s are then obtained by simply dividing the corresponding quantile by 100. The prior for each threshold value $r_i$ in the marginals is chosen so that at least 200 observations are in each regime, giving $q = 10$ and $r_i \in (P_{1,10}, P_{1,90})$ for the 2,000 length rolling window. In the copula correlations, the number of historical $x_{it}'s$ chosen in the calculation of each $\xi_{ij,t}$ is $M = 5$, which corresponds to the minimum requirement for $\xi_t$ to be PD (see Tse and Tsui, 2002). A total of $L = 1,000$ iterations are used in calculation of each VaR$_{\alpha,s}$, as in Ausin and Lopes (2010). The average run time to fit the ASTC-DTGARCH model to the given DJI stock returns is approximately 1 hr 10 min, using the machine and software described in Section 2.7.

Fitting of the competing models is done via maximum likelihood, utilising the Fortran optimisation routine UMINF - details of which are contained in Section 3.5.2, where it
was used to estimate a bivariate BEKK model. For the DCC and GJR-ADCC models, a two-step estimation approach is utilised that first maximises a variance likelihood, then a correlation likelihood - see Engle (2002) and Cappiello et al. (2006) for the necessary details. For each copula-GARCH model (except ASTC-DTGARCH), marginals are estimated first using the relevant univariate GARCH models, then parameters of the copula are estimated via maximising a copula likelihood - as proposed by Joe and Xu (1996).

Tables 5.3-5.6 at the end of this sub-section display parameter estimation results from fitting the ASTC-DTGARCH model to the DJI stock returns. Values in each “Mean” column are simply the average of the 100 posterior means for each parameter (Par), while intervals in each “Percentiles” column are the 2.5th and 97.5th percentiles of the corresponding 100 posterior means. The “AR” column represents the mean acceptance rate over the 100 replications for that particular group of parameters. Important observations can be made from these results. From Table 5.3, we notice the usual high persistence in volatility in stock returns, with each $a_{i1}^{(k)} + b_{i1}^{(k)} > 0.9$. Both mean and volatility asymmetry is present in the data, which is mainly noticeable in the mean and variance intercept terms; the mean intercept $m_{i0}^{(1)}$ is greater than $m_{i0}^{(2)}$ in all stocks except CSCO, indicating that the conditional mean is higher following a return below the estimated threshold $r_i$ than a return above $r_i$ for these stocks. The same applies for the conditional variances as given by the intercept terms $c_{i1}^{(1)}$ and $c_{i1}^{(2)}$. (e.g. the means for $m_{i0}^{(1)}$ and $m_{i0}^{(2)}$ for the INTC stock are 0.32 and 0.02, while the means for $c_{i1}^{(1)}$ and $c_{i1}^{(2)}$ for the INTC stock are 0.64 and 0.11). Table 5.4 shows that mean values of the threshold parameter $r_i$ are all negative, with 95% of posterior means for these parameters less than zero in 2 of the 5 intervals. This contradicts the assumption made by the GJR-GARCH model of zero thresholds, and supports the similar findings of Chen et al. (2005). This table also illustrates the presence of negative skewness and excess kurtosis in the data, with each $\lambda_i$ for the stocks CSCO, XOM and INTC less than 1 and not containing 1 in their intervals, as well as each mean $\nu_i$ for all stocks less than 7. These properties are also present in the copula parameters $\gamma_i$ and $\eta$ in Table 5.5, but not to the same extent. There also
does not appear to be any significant asymmetry given by the copula correlation parameters $\alpha^{(k)}$ and $\beta^{(k)}$, since their corresponding means have similar values. However, Table 5.6, which gives the estimates of each $\tilde{P}^{(k)}$, clearly demonstrates that correlation changes based on the sign of the residuals - positive correlations resulted (up to 0.43 for $\tilde{\rho}_{14}^{(1)}$, i.e. the sample correlation between the positive residuals of the stocks CSCO$_t$ and INTC$_t$) whenever the residuals were either both positive or both negative (i.e. $\tilde{\rho}_{ij}^{(3)}$ or $\tilde{\rho}_{ij}^{(4)}$), with all intervals under these scenarios not containing zero. This is a typical result for models of similar form. Smaller correlations resulted when residuals were of opposite sign (i.e. $\tilde{\rho}_{ij}^{(2)}$ or $\tilde{\rho}_{ij}^{(3)}$), but 5 out of these 20 corresponding intervals did not contain zero. Hence some correlation asymmetry is present under this scenario, even though correlations are relatively small.

To further support the findings of significant skewness and excess kurtosis, Figure 5.2 illustrates the evolution of the marginal distribution parameters $\lambda_i$ and $\nu_i$ ($i = 1, \ldots, 5$), for each of the 100 replications of estimating the ASTC-DTGARCH model. The three stocks CSCO$_t$, XOM$_t$ and INTC$_t$ all have values of $\lambda_i$ below 1 (indicating a degree of negative skewness), while GE$_t$ and MSFT$_t$ have values of $\lambda_i$ mainly above 1 (indicating a degree of positive skewness). Similarly, values of $\nu_i$ for all stocks are below 12, mostly below 8. Clearly skewness and excess kurtosis needs to be captured for this dataset, and based on this evidence the ASTC-DTGARCH model is quite capable of doing this.

The acceptance rates (ARs) reported in these tables also vary significantly. ARs for the autoregressive parameters are quite high (89-93%), while for the GARCH parameters in both marginals and correlation equations ARs are much lower (17-43%). The marginal error distribution parameters ($r_i, \lambda_i, \nu_i$) have reasonably large ARs (62-74%), while the copula parameter groupings ($\gamma_1, \ldots, \gamma_5$) and $\eta$ each have ARs above 90%. This demonstrates the flexibility of the MCMC sampling scheme to generate from posterior distributions of varying shapes, as also discovered in Chapters 3 and 4.

Table 5.7 displays potential scale reduction factors (PSRFs) for parameters of the ASTC-DTGARCH model fitted to an arbitrary set of 2,000 observations of the DJI
stock portfolio returns. As in Chapters 3 and 4, each PSRF is calculated using $R = 5$ replications of MCMC iterates and 5,000 iterations after the burn-in period. For the majority of parameters, each PSRF is reasonably close to 1, with most below 1.05 (including parameters contained in the copula), indicating a degree of convergence to the corresponding posterior distribution. However, the variance intercept parameter in the 1st regime of each marginal distribution has PSRFs ranging between 1.66 and 2.79. This does not necessarily mean non-convergence: Figure 5.3 displays plots of MCMC iterations for a selection of 4 parameters, one including the $c_{21}^{(1)}$ parameter. It appears that convergence is not quite achieved for all $R = 5$ replications, but may be improved by either increasing the value of $R$, or increasing the size of the burn-in and/or sampling period. Other interesting points to note regarding this figure: convergence is almost instantaneous for the $b_{41}^{(2)}$ and $\gamma_1$ parameters, while the $r_1$ parameter offers interesting shapes. Some iterates converge on a value for a period of time, then rise/fall suddenly and converge on a different value for a period of time. The end result being a posterior distribution with a high variance and potentially multi-modal, with convergence still achieved. This is further proof that the presented MCMC sampling scheme can capture posterior distributions of varying shapes, with efficiency and convergence present for most parameters.

Figure 5.4 displays plots of the ASTC-DTGARCH estimates of $P_t = [\rho_{ij,t}]$ for a particular sample of 2,000 observations. As with estimated correlations from Chapters 3 and 4, these are calculated by averaging over the MCMC iterates $\rho_{ij,t}^{[j]}$ using corresponding parameter values. In each case, significant positive correlations have resulted, with all correlations falling between 0 and 0.5. High persistence in correlation is also apparent from these plots, which is also demonstrated by the fact that each $\alpha^{(k)} + \beta^{(k)} > 0.8$ in Table 5.5. This indicates the need to accurately capture the dependence structure in a copula model.
Table 5.3: Mean parameter estimates of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio: Autoregressive and GARCH parameters of the marginal distributions.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Par</th>
<th>Mean</th>
<th>Percentiles</th>
<th>Par</th>
<th>Mean</th>
<th>Percentiles</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSCO</td>
<td>$m_{10}^{(1)}$</td>
<td>0.0412</td>
<td>[-0.2001,0.3085]</td>
<td>$m_{10}^{(2)}$</td>
<td>0.0704</td>
<td>[0.0187,0.2019]</td>
<td>89.5%</td>
</tr>
<tr>
<td></td>
<td>$m_{11}^{(1)}$</td>
<td>-0.0238</td>
<td>[-0.0764,0.0366]</td>
<td>$m_{11}^{(2)}$</td>
<td>-0.0351</td>
<td>[-0.0568,0.0013]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{11}^{(1)}$</td>
<td>0.3943</td>
<td>[0.3281,0.4814]</td>
<td>$c_{11}^{(2)}$</td>
<td>0.0851</td>
<td>[0.0068,0.1935]</td>
<td>17.6%</td>
</tr>
<tr>
<td></td>
<td>$a_{11}^{(1)}$</td>
<td>0.0082</td>
<td>[0.0008,0.0397]</td>
<td>$a_{11}^{(2)}$</td>
<td>0.0384</td>
<td>[0.0156,0.0695]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{11}^{(1)}$</td>
<td>0.9886</td>
<td>[0.9559,0.9970]</td>
<td>$b_{11}^{(2)}$</td>
<td>0.9495</td>
<td>[0.9036,0.9753]</td>
<td></td>
</tr>
<tr>
<td>XOM</td>
<td>$m_{20}^{(1)}$</td>
<td>0.1031</td>
<td>[0.0189,0.2414]</td>
<td>$m_{20}^{(2)}$</td>
<td>0.0591</td>
<td>[0.0098,0.0813]</td>
<td>92.6%</td>
</tr>
<tr>
<td></td>
<td>$m_{21}^{(1)}$</td>
<td>-0.0150</td>
<td>[-0.1206,0.0003]</td>
<td>$m_{21}^{(2)}$</td>
<td>-0.1025</td>
<td>[-0.1319,-0.0814]</td>
<td>35.6%</td>
</tr>
<tr>
<td></td>
<td>$c_{21}^{(1)}$</td>
<td>0.4021</td>
<td>[0.1623,0.4413]</td>
<td>$c_{21}^{(2)}$</td>
<td>0.0682</td>
<td>[0.0176,0.1702]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{21}^{(1)}$</td>
<td>0.0108</td>
<td>[0.0006,0.1301]</td>
<td>$a_{21}^{(2)}$</td>
<td>0.0664</td>
<td>[0.0330,0.1170]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{21}^{(1)}$</td>
<td>0.9675</td>
<td>[0.8280,0.9885]</td>
<td>$b_{21}^{(2)}$</td>
<td>0.8881</td>
<td>[0.8039,0.9462]</td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>$m_{30}^{(1)}$</td>
<td>0.0736</td>
<td>[-0.1885,0.3061]</td>
<td>$m_{30}^{(2)}$</td>
<td>0.0036</td>
<td>[-0.0325,0.0495]</td>
<td>89.8%</td>
</tr>
<tr>
<td></td>
<td>$m_{31}^{(1)}$</td>
<td>-0.0101</td>
<td>[-0.0963,0.0402]</td>
<td>$m_{31}^{(2)}$</td>
<td>0.0197</td>
<td>[-0.0377,0.0743]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{31}^{(1)}$</td>
<td>0.3055</td>
<td>[0.2750,0.3246]</td>
<td>$c_{31}^{(2)}$</td>
<td>0.1086</td>
<td>[0.0025,0.3881]</td>
<td>27.2%</td>
</tr>
<tr>
<td></td>
<td>$a_{31}^{(1)}$</td>
<td>0.0097</td>
<td>[0.0002,0.0623]</td>
<td>$a_{31}^{(2)}$</td>
<td>0.0581</td>
<td>[0.0137,0.1189]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{31}^{(1)}$</td>
<td>0.9837</td>
<td>[0.9280,0.9959]</td>
<td>$b_{31}^{(2)}$</td>
<td>0.9084</td>
<td>[0.7976,0.9736]</td>
<td></td>
</tr>
<tr>
<td>INTC</td>
<td>$m_{40}^{(1)}$</td>
<td>0.3199</td>
<td>[0.0493,0.7640]</td>
<td>$m_{40}^{(2)}$</td>
<td>0.0178</td>
<td>[-0.0227,0.0938]</td>
<td>89.6%</td>
</tr>
<tr>
<td></td>
<td>$m_{41}^{(1)}$</td>
<td>0.0276</td>
<td>[-0.0238,0.1106]</td>
<td>$m_{41}^{(2)}$</td>
<td>-0.0018</td>
<td>[-0.0359,0.0278]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{41}^{(1)}$</td>
<td>0.6381</td>
<td>[0.4348,0.7616]</td>
<td>$c_{41}^{(2)}$</td>
<td>0.1128</td>
<td>[0.0094,0.3094]</td>
<td>22.5%</td>
</tr>
<tr>
<td></td>
<td>$a_{41}^{(1)}$</td>
<td>0.0065</td>
<td>[0.0016,0.0295]</td>
<td>$a_{41}^{(2)}$</td>
<td>0.0372</td>
<td>[0.0175,0.0682]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{41}^{(1)}$</td>
<td>0.9878</td>
<td>[0.9615,0.9947]</td>
<td>$b_{41}^{(2)}$</td>
<td>0.9459</td>
<td>[0.9041,0.9695]</td>
<td></td>
</tr>
<tr>
<td>MSFT</td>
<td>$m_{50}^{(1)}$</td>
<td>0.2011</td>
<td>[0.0816,0.3073]</td>
<td>$m_{50}^{(2)}$</td>
<td>-0.0175</td>
<td>[-0.0802,0.1333]</td>
<td>90.5%</td>
</tr>
<tr>
<td></td>
<td>$m_{51}^{(1)}$</td>
<td>0.0103</td>
<td>[-0.0157,0.0388]</td>
<td>$m_{51}^{(2)}$</td>
<td>-0.0272</td>
<td>[-0.0735,0.0118]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{51}^{(1)}$</td>
<td>0.2978</td>
<td>[0.2394,0.4046]</td>
<td>$c_{51}^{(2)}$</td>
<td>0.0101</td>
<td>[0.0026,0.0210]</td>
<td>21.3%</td>
</tr>
<tr>
<td></td>
<td>$a_{51}^{(1)}$</td>
<td>0.0010</td>
<td>[0.0007,0.0021]</td>
<td>$a_{51}^{(2)}$</td>
<td>0.0510</td>
<td>[0.0356,0.1046]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_{51}^{(1)}$</td>
<td>0.9930</td>
<td>[0.9732,0.9962]</td>
<td>$b_{51}^{(2)}$</td>
<td>0.9411</td>
<td>[0.8649,0.9607]</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.4: Mean parameter estimates of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio: Threshold and error parameters of the marginal distributions.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Parameter</th>
<th>Mean</th>
<th>Percentiles</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSCO_t</td>
<td>r_1</td>
<td>-1.8230</td>
<td>[-2.6043,0.1608]</td>
<td>73.7%</td>
</tr>
<tr>
<td></td>
<td>\lambda_1</td>
<td>0.9571</td>
<td>[0.9223,0.9886]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>\nu_1</td>
<td>5.3866</td>
<td>[4.7442,6.2064]</td>
<td></td>
</tr>
<tr>
<td>XOM_t</td>
<td>r_2</td>
<td>-1.0958</td>
<td>[-1.7640,0.2722]</td>
<td>68.4%</td>
</tr>
<tr>
<td></td>
<td>\lambda_2</td>
<td>0.9248</td>
<td>[0.8532,0.9749]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>\nu_2</td>
<td>6.4011</td>
<td>[5.3519,11.2169]</td>
<td></td>
</tr>
<tr>
<td>GE_t</td>
<td>r_3</td>
<td>-1.3084</td>
<td>[-1.9446,-0.5261]</td>
<td>62.1%</td>
</tr>
<tr>
<td></td>
<td>\lambda_3</td>
<td>1.0280</td>
<td>[1.0009,1.0581]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>\nu_3</td>
<td>5.7419</td>
<td>[4.4870,7.5560]</td>
<td></td>
</tr>
<tr>
<td>INTC_t</td>
<td>r_4</td>
<td>-2.2393</td>
<td>[-2.8542,-1.1139]</td>
<td>70.7%</td>
</tr>
<tr>
<td></td>
<td>\lambda_4</td>
<td>0.9851</td>
<td>[0.9732,0.9969]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>\nu_4</td>
<td>5.5964</td>
<td>[4.3215,7.2805]</td>
<td></td>
</tr>
<tr>
<td>MSFT_t</td>
<td>r_5</td>
<td>-1.0452</td>
<td>[-1.4663,0.6212]</td>
<td>64.5%</td>
</tr>
<tr>
<td></td>
<td>\lambda_5</td>
<td>1.0096</td>
<td>[0.9927,1.0311]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>\nu_5</td>
<td>4.4250</td>
<td>[4.1574,4.7660]</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.5: Mean parameter estimates of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio: GARCH parameters \(\alpha^{(k)}\), \(\beta^{(k)}\), \(k = 1, 2, 3, 4\) and copula distribution parameters.

<table>
<thead>
<tr>
<th>Par</th>
<th>Mean</th>
<th>Percentiles</th>
<th>Par</th>
<th>Mean</th>
<th>Percentiles</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>\alpha^{(1)}</td>
<td>0.0162</td>
<td>[0.0063,0.0444]</td>
<td>\beta^{(1)}</td>
<td>0.8079</td>
<td>[0.1255,0.9891]</td>
<td>42.9%</td>
</tr>
<tr>
<td>\alpha^{(2)}</td>
<td>0.0067</td>
<td>[0.0025,0.0153]</td>
<td>\beta^{(2)}</td>
<td>0.9810</td>
<td>[0.9609,0.9919]</td>
<td></td>
</tr>
<tr>
<td>\alpha^{(3)}</td>
<td>0.0142</td>
<td>[0.0047,0.0231]</td>
<td>\beta^{(3)}</td>
<td>0.9612</td>
<td>[0.9403,0.9749]</td>
<td></td>
</tr>
<tr>
<td>\alpha^{(4)}</td>
<td>0.0146</td>
<td>[0.0093,0.0226]</td>
<td>\beta^{(4)}</td>
<td>0.9455</td>
<td>[0.8047,0.9842]</td>
<td></td>
</tr>
<tr>
<td>\gamma_1</td>
<td>1.0006</td>
<td>[0.9453,1.0378]</td>
<td>\gamma_2</td>
<td>0.9920</td>
<td>[0.8040,1.1279]</td>
<td>90.4%</td>
</tr>
<tr>
<td>\gamma_3</td>
<td>1.1582</td>
<td>[1.0612,1.2599]</td>
<td>\gamma_4</td>
<td>1.0083</td>
<td>[0.9784,1.0467]</td>
<td></td>
</tr>
<tr>
<td>\gamma_5</td>
<td>0.9874</td>
<td>[0.9717,0.9992]</td>
<td>\eta</td>
<td>7.6078</td>
<td>[6.7194,9.2315]</td>
<td>97.3%</td>
</tr>
</tbody>
</table>
Table 5.6: Mean parameter estimates of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio: Correlation parameters $\hat{\rho}^{(k)}$, $k = 1, 2, 3, 4$ of the copula.

<table>
<thead>
<tr>
<th>Parameter $\hat{\rho}^{(1)}$</th>
<th>Mean</th>
<th>Percentiles</th>
<th>Parameter $\hat{\rho}^{(1)}$</th>
<th>Mean</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_{12}$</td>
<td>0.1390</td>
<td>[0.0980, 0.2207]</td>
<td>$\hat{\rho}_{13}$</td>
<td>0.3626</td>
<td>[0.2738, 0.4058]</td>
</tr>
<tr>
<td>$\hat{\rho}_{14}$</td>
<td>0.4291</td>
<td>[0.3538, 0.5068]</td>
<td>$\hat{\rho}_{15}$</td>
<td>0.3614</td>
<td>[0.2923, 0.4341]</td>
</tr>
<tr>
<td>$\hat{\rho}_{23}$</td>
<td>0.2022</td>
<td>[0.1574, 0.2520]</td>
<td>$\hat{\rho}_{24}$</td>
<td>0.1632</td>
<td>[0.0862, 0.2814]</td>
</tr>
<tr>
<td>$\hat{\rho}_{25}$</td>
<td>0.1941</td>
<td>[0.0904, 0.3583]</td>
<td>$\hat{\rho}_{34}$</td>
<td>0.3477</td>
<td>[0.2906, 0.4096]</td>
</tr>
<tr>
<td>$\hat{\rho}_{35}$</td>
<td>0.3035</td>
<td>[0.2330, 0.3404]</td>
<td>$\hat{\rho}_{45}$</td>
<td>0.4105</td>
<td>[0.3570, 0.4579]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter $\hat{\rho}^{(2)}$</th>
<th>Mean</th>
<th>Percentiles</th>
<th>Parameter $\hat{\rho}^{(2)}$</th>
<th>Mean</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_{12}$</td>
<td>-0.0166</td>
<td>[-0.0385, 0.0127]</td>
<td>$\hat{\rho}_{13}$</td>
<td>0.0208</td>
<td>[0.0036, 0.0475]</td>
</tr>
<tr>
<td>$\hat{\rho}_{14}$</td>
<td>-0.0160</td>
<td>[-0.0498, 0.0436]</td>
<td>$\hat{\rho}_{15}$</td>
<td>-0.1343</td>
<td>[-0.1935, -0.0809]</td>
</tr>
<tr>
<td>$\hat{\rho}_{23}$</td>
<td>-0.0059</td>
<td>[-0.0322, 0.0198]</td>
<td>$\hat{\rho}_{24}$</td>
<td>0.0807</td>
<td>[0.0643, 0.0912]</td>
</tr>
<tr>
<td>$\hat{\rho}_{25}$</td>
<td>0.0206</td>
<td>[-0.0135, 0.0498]</td>
<td>$\hat{\rho}_{34}$</td>
<td>-0.0268</td>
<td>[-0.1053, 0.0566]</td>
</tr>
<tr>
<td>$\hat{\rho}_{35}$</td>
<td>-0.0611</td>
<td>[-0.1437, 0.0024]</td>
<td>$\hat{\rho}_{45}$</td>
<td>-0.0975</td>
<td>[-0.1742, -0.0268]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter $\hat{\rho}^{(3)}$</th>
<th>Mean</th>
<th>Percentiles</th>
<th>Parameter $\hat{\rho}^{(3)}$</th>
<th>Mean</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_{12}$</td>
<td>-0.0284</td>
<td>[-0.1044, 0.0132]</td>
<td>$\hat{\rho}_{13}$</td>
<td>-0.0467</td>
<td>[-0.1359, 0.0379]</td>
</tr>
<tr>
<td>$\hat{\rho}_{14}$</td>
<td>0.0622</td>
<td>[0.0163, 0.0940]</td>
<td>$\hat{\rho}_{15}$</td>
<td>-0.1675</td>
<td>[-0.2631, 0.0158]</td>
</tr>
<tr>
<td>$\hat{\rho}_{23}$</td>
<td>-0.0035</td>
<td>[-0.0407, 0.0282]</td>
<td>$\hat{\rho}_{24}$</td>
<td>-0.0347</td>
<td>[-0.0545, 0.0405]</td>
</tr>
<tr>
<td>$\hat{\rho}_{25}$</td>
<td>-0.0263</td>
<td>[-0.0406, 0.0136]</td>
<td>$\hat{\rho}_{34}$</td>
<td>-0.0472</td>
<td>[-0.1382, 0.0863]</td>
</tr>
<tr>
<td>$\hat{\rho}_{35}$</td>
<td>-0.0173</td>
<td>[-0.1060, 0.0589]</td>
<td>$\hat{\rho}_{45}$</td>
<td>-0.0362</td>
<td>[-0.1452, 0.0604]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter $\hat{\rho}^{(4)}$</th>
<th>Mean</th>
<th>Percentiles</th>
<th>Parameter $\hat{\rho}^{(4)}$</th>
<th>Mean</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_{12}$</td>
<td>0.1118</td>
<td>[0.0340, 0.2650]</td>
<td>$\hat{\rho}_{13}$</td>
<td>0.1265</td>
<td>[0.0655, 0.2465]</td>
</tr>
<tr>
<td>$\hat{\rho}_{14}$</td>
<td>0.1476</td>
<td>[0.0467, 0.3627]</td>
<td>$\hat{\rho}_{15}$</td>
<td>0.1265</td>
<td>[0.0491, 0.2482]</td>
</tr>
<tr>
<td>$\hat{\rho}_{23}$</td>
<td>0.2979</td>
<td>[0.1881, 0.3591]</td>
<td>$\hat{\rho}_{24}$</td>
<td>0.1269</td>
<td>[0.0552, 0.2378]</td>
</tr>
<tr>
<td>$\hat{\rho}_{25}$</td>
<td>0.1799</td>
<td>[0.0999, 0.2975]</td>
<td>$\hat{\rho}_{34}$</td>
<td>0.1042</td>
<td>[0.0577, 0.2320]</td>
</tr>
<tr>
<td>$\hat{\rho}_{35}$</td>
<td>0.1033</td>
<td>[0.0205, 0.2938]</td>
<td>$\hat{\rho}_{45}$</td>
<td>0.1908</td>
<td>[0.0803, 0.2444]</td>
</tr>
</tbody>
</table>
Table 5.7: Potential scale reduction factors (PSRFs) for parameters of the ASTC-DTGARCH model fitted to an arbitrary set of 2,000 observations of the DJI stock portfolio returns

<table>
<thead>
<tr>
<th>Par</th>
<th>PS-RF</th>
<th>Par</th>
<th>PS-RF</th>
<th>Par</th>
<th>PS-RF</th>
<th>Par</th>
<th>PS-RF</th>
<th>Par</th>
<th>PS-RF</th>
<th>Par</th>
<th>PS-RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{10}^{(1)}$</td>
<td>1.00</td>
<td>$m_{10}^{(2)}$</td>
<td>1.05</td>
<td>$m_{20}^{(1)}$</td>
<td>1.00</td>
<td>$m_{20}^{(2)}$</td>
<td>1.03</td>
<td>$m_{30}^{(1)}$</td>
<td>1.01</td>
<td>$m_{30}^{(2)}$</td>
<td>1.02</td>
</tr>
<tr>
<td>$m_{11}^{(1)}$</td>
<td>1.00</td>
<td>$m_{11}^{(2)}$</td>
<td>1.03</td>
<td>$m_{21}^{(1)}$</td>
<td>1.00</td>
<td>$m_{21}^{(2)}$</td>
<td>1.00</td>
<td>$m_{31}^{(1)}$</td>
<td>1.04</td>
<td>$m_{31}^{(2)}$</td>
<td>1.01</td>
</tr>
<tr>
<td>$c_{11}^{(1)}$</td>
<td>1.66</td>
<td>$c_{11}^{(2)}$</td>
<td>1.27</td>
<td>$c_{21}^{(1)}$</td>
<td>1.81</td>
<td>$c_{21}^{(2)}$</td>
<td>1.08</td>
<td>$c_{31}^{(1)}$</td>
<td>1.98</td>
<td>$c_{31}^{(2)}$</td>
<td>1.23</td>
</tr>
<tr>
<td>$a_{11}^{(1)}$</td>
<td>1.02</td>
<td>$a_{11}^{(2)}$</td>
<td>1.02</td>
<td>$a_{21}^{(1)}$</td>
<td>1.00</td>
<td>$a_{21}^{(2)}$</td>
<td>1.00</td>
<td>$a_{31}^{(1)}$</td>
<td>1.02</td>
<td>$a_{31}^{(2)}$</td>
<td>1.05</td>
</tr>
<tr>
<td>$a_{12}^{(1)}$</td>
<td>1.02</td>
<td>$a_{12}^{(2)}$</td>
<td>1.02</td>
<td>$a_{22}^{(1)}$</td>
<td>1.00</td>
<td>$a_{22}^{(2)}$</td>
<td>1.00</td>
<td>$a_{32}^{(1)}$</td>
<td>1.02</td>
<td>$a_{32}^{(2)}$</td>
<td>1.05</td>
</tr>
<tr>
<td>$b_{11}^{(1)}$</td>
<td>1.36</td>
<td>$b_{11}^{(2)}$</td>
<td>1.20</td>
<td>$b_{12}^{(1)}$</td>
<td>1.46</td>
<td>$b_{12}^{(2)}$</td>
<td>1.05</td>
<td>$b_{21}^{(1)}$</td>
<td>1.09</td>
<td>$b_{21}^{(2)}$</td>
<td>1.05</td>
</tr>
<tr>
<td>$b_{12}^{(1)}$</td>
<td>1.20</td>
<td>$b_{12}^{(2)}$</td>
<td>1.20</td>
<td>$b_{22}^{(1)}$</td>
<td>1.05</td>
<td>$b_{22}^{(2)}$</td>
<td>1.05</td>
<td>$b_{32}^{(1)}$</td>
<td>1.05</td>
<td>$b_{32}^{(2)}$</td>
<td>1.05</td>
</tr>
<tr>
<td>$r_{1}$</td>
<td>1.16</td>
<td>$r_{2}$</td>
<td>1.19</td>
<td>$r_{3}$</td>
<td>1.04</td>
<td>$r_{4}$</td>
<td>1.04</td>
<td>$r_{5}$</td>
<td>1.06</td>
<td>$r_{5}$</td>
<td>1.06</td>
</tr>
<tr>
<td>$\lambda_{1}$</td>
<td>1.00</td>
<td>$\lambda_{2}$</td>
<td>1.01</td>
<td>$\lambda_{3}$</td>
<td>1.00</td>
<td>$\lambda_{4}$</td>
<td>1.00</td>
<td>$\lambda_{5}$</td>
<td>1.00</td>
<td>$\lambda_{5}$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\nu_{1}$</td>
<td>1.00</td>
<td>$\nu_{2}$</td>
<td>1.02</td>
<td>$\nu_{3}$</td>
<td>1.01</td>
<td>$\nu_{4}$</td>
<td>1.00</td>
<td>$\nu_{5}$</td>
<td>1.14</td>
<td>$\nu_{5}$</td>
<td>1.14</td>
</tr>
</tbody>
</table>

197
Figure 5.2: Plots of estimates of $\lambda_i$ and $\nu_i$, $i = 1, \ldots, 5$, from the marginal distributions of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio.
Figure 5.3: MCMC iterates for the parameters $r_1^{(1)}$, $c_21^{(1)}$, $b_{41}^{(2)}$, and $\gamma_1$ for the ASTC-DTGARCH model fitted to an arbitrary set of 2,000 observations of the D11 stock portfolio returns, each with 5 different starting values.
Figure 5.4: Plots of the estimated correlation matrix $P_t = [\rho_{ij,t}]$ over a selection of 2,000 observations for the dependence structure of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio.
5.7.6 VaR Back-Testing Results

Tables 5.8-5.11 in the Appendix contain the VaR back-testing results performed on the DJI stock portfolio for the four time horizons $s = 1, 5, 10, 20$ days. Each table contains the number of VaR violations in the sample, an estimate of the coverage (or significance level) $\hat{\alpha}$, the ratio $\hat{\alpha}/\alpha$, and the $p$-values of the UC, CC and DQ tests for each model. Values of $\hat{\alpha}/\alpha$ closest to 1 are indicated in **bold**, while any $p$-values less than the corresponding significance level are indicated in red. Whenever an “N/A” appears in the columns for the DQ test $p$-value ($p_{dq}$), this indicates no VaR violations and the test is unable to be performed. The average run time to compute each VaR estimate using the MCMC iterates generated from the ASTC-DTGARCH model is approximately 7 min, using the machine and software described in Section 2.7.

We can see from these tables that the majority of models have insignificant $p$-values for both the UC and CC tests at the $\alpha = 0.01$ level for all time horizons, indicating a reasonable percentage of VaR violations with respect to $\alpha$ in the relevant samples, and violations are not occurring consecutively. A total of 6 out of the 15 models have significant $p$-values for the DQ test for $s = 1$ day at the $\alpha = 0.01$ level, but do not seem to be restricted to any model grouping (i.e. univariate or multivariate). However, clear differences can be seen at the $\alpha = 0.05$ level, with the majority of multivariate models based on the multivariate normal distribution having significant $p$-values across all horizons, and are generally over-estimating VaR (as indicated by the value of $\hat{\alpha}/\alpha < 1$). This is particularly noticeable for the DCC, GJR-ADCC and COP-GJR-GARCH$_N$ models, where all $p$-values at the $s = 1, 5$ day horizons being significant at this $\alpha$ level. This also occurs for the ACOP-GJR-GARCH$_N$ model, but to a lesser extent. These results may be due to the normal error distribution in these models not being able to accurately capture the tail behaviour of the portfolio return distributions. It is important to note that the $p$-values for the ASTC-DTGARCH model are all insignificant, except at $s = 1$ day at the $\alpha = 0.05$ level. Evidence now exists of univariate GARCH models performing better at VaR estimation than mis-specified multivariate GARCH models.
As mentioned in Section 5.1, whenever the weights are known in a stock portfolio, univariate GARCH models used for VaR estimation may outperform multivariate GARCH models as correlations between each stock return series are “implicitly” captured in the historical portfolio return series, whereas the correlations need to be modelled in multivariate GARCH models. This property is highlighted in Tables 5.12-5.13 which display values of the Koenker and Bassett (1978) VaR criterion function for each model, as well as their ranking based on the lowest function value. We can see that the above property is prevalent in these results, with all six univariate GARCH models generally ranking the highest compared to their multivariate counterparts. However, the ASTC-DTGARCH model also performs quite well, and is comparable to the univariate GARCH models - it ranks 1st a total of two times plus a 2nd and a 4th, and is not restricted to the one significance level. Even for $s = 1$ at the $\alpha = 0.05$ level where the ASTC-DTGARCH model has significant $p$-values, it still ranks as the 3rd best multivariate GARCH model. It is clearly the best performing multivariate GARCH model in this group. The worst performing models are the multivariate models with normal errors, i.e. DCC, GJR-ADCC, COP-GJR-GARCH$_N$ and ACOP-GJR-GARCH$_N$, which always rank in the bottom four models in each case and fully support the findings of the UC, CC and DQ tests described earlier. Moreover, as mentioned in Chapter 3, the numerical optimisation of the likelihood in the competing models may be unreliable compared to the numerical robustness of the MCMC approach for the ASTC-DTGARCH model. This is a further indication that the error distribution and estimation approach must be carefully chosen in a multivariate model in order to accurately estimate VaR.

Figures 5.8-5.11 display each estimated VaR and portfolio returns for $\alpha = 0.01, 0.05$ and each time horizon $s = 1, 5, 10, 20$ over the last 2,000 observations in the sample. Some models are combined together in these plots for ease of viewing and to highlight some important properties. Based on the VaR back-testing results above and visual checks, the following models are grouped together by simply calculating the mean of their VaR estimates at the same time horizon:
• **Univariate:** This line represents the mean of the VaR estimates for all 6 univariate GARCH models considered in the study, i.e. GARCH\(_N\), GARCH\(_T\), GARCH\(_S\), GJR-GARCH\(_N\), GJR-GARCH\(_T\) and GJR-GARCH\(_S\).

• **DCC and GJR-ADCC:** This line represents the mean of the VaR estimates for the DCC and GJR-ADCC models.

• **Gaussian Copulas:** This line represents the mean of the VaR estimates of the two multivariate models using Gaussian copulas, i.e. COP-GJR-GARCH\(_N\) and ACOP-GJR-GARCH\(_N\).

• **Student-t and Skew-t Copulas:** This line represents the mean of the VaR estimates of the remaining copula models, i.e. COP-GJR-GARCH\(_T\), COP-GJR-GARCH\(_S\), ACOP-GJR-GARCH\(_T\) and ACOP-GJR-GARCH\(_S\).

The ASTC-DTGARCH model has been left on its own to examine its performance individually. In each of the plots, we see differences between each model group. Firstly, the ASTC-DTGARCH model compares quite well with the univariate model group, with its VaR estimates generally lying closer to the portfolio returns under all scenarios. It also responds to periods of both high and low volatility at the correct times, reflecting corresponding values of the VaR criterion function in Tables 5.12-5.13. Secondly, the DCC and GJR-ADCC model group performs quite poorly; it does not respond to periods of low volatility in the middle of the back-testing period, thus over-estimating VaR for all time horizons. The same can be said for the Gaussian copula group, but is more prevalent for the longer time horizons \(s = 5, 10, 20\). Thirdly, improvements are apparent when the error distribution is changed from Gaussian to Student-\(t\) or skew-\(t\) for the copula models; while not as accurate as the univariate or ASTC-DTGARCH models, VaR estimates from this group are closer to the portfolio returns than that of their Gaussian counterparts. Combined with earlier VaR back-testing results, this clearly indicates that multivariate GARCH models must allow for asymmetry, skewness and excess kurtosis across the entire model in order to accurately estimate VaR, as in the ASTC-DTGARCH model.
5.8 Chapter Summary

The ASTC-DTGARCH model has been proposed in this chapter, which is a copula-GARCH model that encapsulates asymmetry in conditional means and variances in the marginal distributions, as well as asymmetry in the correlations of the dependence structure of the copula. Skewness and excess kurtosis were captured in the marginal distributions via the skew-$t$ distribution formed from the methods of Fernandez and Steel (1998), while asymmetric dependence was also captured in the copula via the multivariate skew-$t$ distribution of Bauwens and Laurent (2005). This extends existing approaches in the literature whereby copula-GARCH models have not had asymmetry, skewness and excess kurtosis represented in all of their model components. This model was applied to a multivariate stock return series comprising of a portfolio of DJI stocks. Parameter estimation results showed that asymmetry exists in both conditional mean and variance equations for most stocks, as demonstrated by differences in corresponding mean and variance intercept estimates. Asymmetry was also apparent in the copula dependence structure via each parameter matrix $\bar{P}^{(k)}$, where correlations changed values based on the sign of the residuals. Skewness and excess kurtosis were also discovered in the marginal distributions, thus demonstrating the need to capture these properties in the one multivariate model. An application in back-testing Value at Risk (VaR) was also presented on the same DJI stock portfolio, where the ASTC-DTGARCH model outperformed other existing multivariate GARCH models in estimating VaR and in some tests outperformed existing univariate GARCH models that are expected to perform better in the setting where the portfolio weights are known.
5.9 Appendix

5.9.1 Properties of the Univariate Skew-t Distribution

The PDF

Let $Z$ denote a random variable. If $Z$ follows the univariate skew-$t$ distribution, its PDF is given by

$$
skt_{\lambda, \nu}(z) = \frac{2\lambda s}{(1 + \lambda^2) \Gamma\left(\frac{\nu}{2}\right) \sqrt{(\nu - 2)\pi}} \left[ 1 + \frac{\lambda^{-2I}(m + sz)^2}{\nu - 2} \right]^{-\frac{\nu + 1}{2}}
$$

where $\lambda$ is the skewness parameter, $\nu$ is the degrees of freedom parameter, with

$$
m = \frac{\Gamma\left(\frac{\nu-1}{2}\right) \sqrt{\nu - 2}}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi}} \left( \lambda - \frac{1}{\lambda} \right), \quad s^2 = \lambda^2 + \frac{1}{\lambda^2} - 1 - m^2
$$

and

$$
I = \begin{cases} 
-1 & \text{if } z < -\frac{m}{s}, \\
1 & \text{if } z \geq -\frac{m}{s}.
\end{cases}
$$

This is a skewed version of the standardised Student-$t$ distribution using the approach of Fernandez and Steel (1998). It is standardised to have a zero mean (i.e. $E(Z) = 0$) and unit variance (i.e. $\text{Var}(Z) = 1$).

Figure 5.5 displays some examples of the univariate skew-$t$ PDF for various values of $\lambda$ and $\nu$. Plot (a) shows the PDF for various values of $\lambda$ (for fixed $\nu = 8$), while plot (b) shows the PDF for various values of $\nu$ (for fixed $\lambda = 0.7$). The standard normal PDF is superimposed on both plots for comparison. From plot (a), we can see that negative skewness exists for $\lambda < 1$ and positive skewness exists for $\lambda > 1$, as per the definition of the distribution. There is also a significant difference in tail behaviour between each PDF here, with heavier tails occurring when $\lambda$ moves further away from 1. As expected, each skew-$t$ PDF has heavier tails than that of the standard normal PDF on the side where the skewness occurs. For the fixed value of $\lambda$, plot (b) shows that when $\nu$ increases, the...
mode of the PDF decreases. There is not a significant amount of difference in the tails of each PDF, but are still heavier than the standard normal PDF, as expected.

Figure 5.5: Plots of the univariate skew-t PDF with different values of the parameters $\lambda$ and $\nu$. The fixed values in plots (a) and (b) are $\nu = 8$ and $\lambda = 0.7$ (respectively).

The derivations in the sections to follow assume that the CDF and inverse CDF of the (unstandardised) Student-t distribution (i.e. $T_{\nu}(z)$ and $T_{\nu}^{-1}(u)$) are known, where

$$T_{\nu}(z) = \frac{\Gamma \left(\frac{\nu+1}{2}\right)}{\Gamma \left(\frac{\nu}{2}\right) \sqrt{\nu \pi}} \int_{-\infty}^{z} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt.$$ 

This is usually not a problem, since both of these functions are typically available in most statistical computing packages.

The CDF

Let $\text{Skt}_{\lambda, \nu}(z)$ denote the CDF of the skew-$t$ distribution. This CDF needs to be derived over the two separate regions on either side of the point $-m/s$, as defined in the PDF.
If $z < -m/s$, then

$$
\text{Skt}_{\lambda, \nu}(z) = \frac{2\lambda s \Gamma\left(\frac{\nu+1}{2}\right)}{(1 + \lambda^2)\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu - 2)\pi}} \int_{-\infty}^{z} \left[ 1 + \frac{\lambda^2(m + st)^2}{\nu - 2} \right]^{-\frac{\nu+1}{2}} dt.
$$

To calculate this integral, we make the following substitutions:

- Let $w = \sqrt{\frac{\nu}{\nu - 2}}[\lambda(m + st)]$. Differentiating yields $dw = \sqrt{\frac{\nu}{\nu - 2}}sdt$.
- When $t = -\infty$, $w = -\infty$ and when $t = z$, $w = \sqrt{\frac{\nu}{\nu - 2}}[\lambda(m + sz)]$

Therefore

$$
\text{Skt}_{\lambda, \nu}(z) = \frac{2\lambda s \Gamma\left(\frac{\nu+1}{2}\right)}{(1 + \lambda^2)\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu - 2)\pi}} \int_{-\infty}^{z} \left[ 1 + \frac{\lambda^2(m + st)^2}{\nu - 2} \right]^{-\frac{\nu+1}{2}} \sqrt{\frac{\nu}{\nu - 2}} s dt
\quad = \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{(1 + \lambda^2)\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}} \int_{-\infty}^{\sqrt{\frac{\nu}{\nu - 2}}[\lambda(m + sz)]} \left( 1 + \frac{w^2}{\nu} \right)^{-\frac{\nu+1}{2}} dw,
$$

(5.39)

If $z \geq -m/s$, then the CDF is made up of two integrals $I_1$ and $I_2$ as follows:

$$
\text{Skt}_{\lambda, \nu}(z) = \frac{2\lambda s \Gamma\left(\frac{\nu+1}{2}\right)}{(1 + \lambda^2)\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu - 2)\pi}} \int_{-\infty}^{-m} \left[ 1 + \frac{\lambda^2(m + st)^2}{\nu - 2} \right]^{-\frac{\nu+1}{2}} dt
\quad + \frac{2\lambda s \Gamma\left(\frac{\nu+1}{2}\right)}{(1 + \lambda^2)\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu - 2)\pi}} \int_{-m}^{z} \left[ 1 + \frac{(m + st)^2}{\lambda^2(\nu - 2)} \right]^{-\frac{\nu+1}{2}} dt
\quad = I_1 + I_2.
$$

By substituting $z = -m/s$ into the CDF in equation (5.39), we obtain $I_1:

$$
I_1 = \frac{2}{1 + \lambda^2} T_\nu(0) = \frac{1}{1 + \lambda^2}.
$$

To calculate $I_2$, we make the following substitutions:

- Let $w = \sqrt{\frac{\nu}{\nu - 2}}\left[\frac{1}{\lambda}(m + st)\right]$. Differentiating yields $dw = \sqrt{\frac{\nu}{\nu - 2}} \lambda dt$. 

207
When \( t = -\frac{m}{s} \), \( w = 0 \) and when \( t = z \), \( w = \sqrt{\frac{\nu}{\nu - 2}} \left[ \frac{1}{\lambda}(m + sz) \right] \)

Therefore

\[
I_2 = \frac{2\lambda s \Gamma \left( \frac{\nu + 1}{2} \right)}{(1 + \lambda^2) \Gamma \left( \frac{\nu}{2} \right) \sqrt{\nu - 2\pi}} \int_{-\frac{m}{s}}^{z} \left[ 1 + \frac{(m + st)^2}{\nu^2} \right] \frac{\nu}{\nu - 2} \sqrt{\frac{\nu}{2\pi}} \lambda^2 (m + sz) dt \\
= \frac{2\lambda^2 \Gamma \left( \frac{\nu + 1}{2} \right)}{(1 + \lambda^2) \Gamma \left( \frac{\nu}{2} \right) \sqrt{\nu - 2\pi}} \int_0^{\frac{w}{\sqrt{\nu - 2}}[\frac{1}{\lambda}(m + sz)]} \left[ 1 + \frac{w^2}{\nu} \right]^{-\frac{\nu + 1}{2}} dw \\
= \frac{2\lambda^2}{1 + \lambda^2} \left[ T_{\nu} \left( \sqrt{\frac{\nu}{\nu - 2}} \left[ \frac{1}{\lambda}(m + sz) \right] \right) - T_{\nu}(0) \right] \\
= \frac{2\lambda^2}{1 + \lambda^2} T_{\nu} \left( \sqrt{\frac{\nu}{\nu - 2}} \left[ \frac{1}{\lambda}(m + sz) \right] \right) - \frac{\lambda^2}{1 + \lambda^2}. 
\]

Hence if \( z \geq -m/s \), we have

\[
\text{Skt}_{\lambda,\nu}(z) = I_1 + I_2 = \frac{1 - \lambda^2}{1 + \lambda^2} + \frac{2\lambda^2}{1 + \lambda^2} T_{\nu} \left( \sqrt{\frac{\nu}{\nu - 2}} \left[ \frac{1}{\lambda}(m + sz) \right] \right). 
\] (5.40)

Combining equations (5.39) and (5.40), the CDF of the skew-t distribution is given by

\[
\text{Skt}_{\lambda,\nu}(z) = \begin{cases} 
\frac{2}{1 + \lambda^2} T_{\nu} \left( \sqrt{\frac{\nu}{\nu - 2}}[\lambda(m + sz)] \right) & \text{if } z < -\frac{m}{s}, \\
1 - \lambda^2 + \frac{2\lambda^2}{1 + \lambda^2} T_{\nu} \left( \sqrt{\frac{\nu}{\nu - 2}} \left[ \frac{1}{\lambda}(m + sz) \right] \right) & \text{if } z \geq -\frac{m}{s}.
\end{cases} 
\] (5.41)

The Inverse CDF

Let \( q = \text{Skt}_{\lambda,\nu}^{-1}(u) \) denote the inverse CDF of the skew-t distribution, where \( u \in [0, 1] \).

The first step is to determine the equivalent change point to \(-m/s\) on the inverse scale.

Using the CDF in equation (5.41), we find that

\[
\text{Skt}_{\lambda,\nu}(-\frac{m}{s}) = \frac{1}{1 + \lambda^2}. 
\]

This means that the region \( z < -m/s \) is equivalent to \( u < 1/(1 + \lambda^2) \) and \( z \geq -m/s \) is equivalent to \( u \geq 1/(1 + \lambda^2) \) on the inverse scale.

The next step is to interchange the variables in the CDF and solve the resulting
equation to obtain the inverse CDF: Let $u = \text{Skt}_{\lambda, \nu}(z)$ and $q = z$ in equation (5.41). Over the region $u < 1/(1 + \lambda^2)$, we have

$$u = \frac{2}{1 + \lambda^2} T_{\nu} \left( \sqrt{\frac{\nu}{\nu - 2} \left[ \lambda (m + sq) \right]} \right)$$

and solving this equation for $q$ yields

$$q = \frac{1}{s} \left[ \frac{1}{\lambda} \sqrt{\frac{\nu - 2}{\nu}} T_{\nu}^{-1} \left( \frac{1 + \lambda^2}{2} u \right) - m \right]. \quad (5.42)$$

Over the region $u \geq 1/(1 + \lambda^2)$, we have

$$u = \frac{1 - \lambda^2}{1 + \lambda^2} + \frac{2\lambda^2}{1 + \lambda^2} T_{\nu} \left( \sqrt{\frac{\nu}{\nu - 2} \left[ \frac{1}{\lambda} (m + sq) \right]} \right)$$

and solving this equation for $q$ yields

$$q = \frac{1}{s} \left[ \lambda \sqrt{\frac{\nu - 2}{\nu}} T_{\nu}^{-1} \left( \frac{1 + \lambda^2}{2\lambda^2} u - \frac{1 - \lambda^2}{2\lambda^2} \right) - m \right]. \quad (5.43)$$

Combining equations (5.42) and (5.43), the inverse CDF of the skew-$t$ distribution is given by

$$\text{Skt}_{\lambda, \nu}^{-1}(u) = \begin{cases} \frac{1}{s} \left[ \frac{1}{\lambda} \sqrt{\frac{\nu - 2}{\nu}} T_{\nu}^{-1} \left( \frac{1 + \lambda^2}{2} u \right) - m \right] & \text{if } u < \frac{1}{1 + \lambda^2}, \\ \frac{1}{s} \left[ \lambda \sqrt{\frac{\nu - 2}{\nu}} T_{\nu}^{-1} \left( \frac{1 + \lambda^2}{2\lambda^2} u - \frac{1 - \lambda^2}{2\lambda^2} \right) - m \right] & \text{if } u \geq \frac{1}{1 + \lambda^2}. \end{cases} \quad (5.44)$$

**Random Number Generation**

As the inverse CDF of the skew-$t$ distribution was derived in the previous section, random number generation is straightforward, as the inversion method can be used (see Devroye, 1986): Given a random number $u$ from the Unif[0, 1] distribution, a random number $x$ from the skew-$t$ distribution with mean $\mu$, variance $\sigma^2$, skewness parameter $\lambda$ and degrees
of freedom \( \nu \) can be obtained by

\[
x = \mu + \sigma \text{Skt}_{\lambda,\nu}^{-1}(u).
\]

Generating random numbers from the multivariate skew-t distribution from equation (5.9) is also straightforward, since the univariate skew-t distribution above are its marginals. Given an \( N \)-dimensional random vector \( \mathbf{u} = (u_1, \ldots, u_N) \) with elements coming from the Unif[0, 1] distribution, a random vector \( \mathbf{x} \) from the multivariate skew-t distribution with mean vector \( \mu = (\mu_1, \ldots, \mu_N) \), covariance matrix \( \Sigma = [\sigma_{ij}] \), skewness parameter vector \( \lambda = (\lambda_1, \ldots, \lambda_N) \) and degrees of freedom \( \nu \) can be obtained by the following steps:

1. Generate the random vector \( \mathbf{z} = (z_1, \ldots, z_N) \) using independent univariate skew-t distributions via
   \[
z_i = \text{Skt}_{\lambda_i,\nu}^{-1}(u_i), \quad i = 1, \ldots, N.
\]
2. Calculate the Cholesky decomposition of the covariance matrix \( \Sigma \) so that \( AA' = \Sigma \).
3. The random vector \( \mathbf{x} \) is then formed by the equation \( \mathbf{x} = \mathbf{\mu} + A\mathbf{z} \).

**Moments**

Moments of the univariate skew-t distribution are easily derived via its unstandardised version: Let \( Z^* \) denote the unstandardised version of \( Z \), i.e. \( Z^* = m + sZ \). Its PDF is given by

\[
\text{skt}_{\lambda,\nu}^{*}(z^*) = \frac{2\lambda \Gamma \left( \frac{\nu+1}{2} \right)}{(1 + \lambda^2) \Gamma \left( \frac{\nu}{2} \right) \sqrt{(\nu - 2)\pi}} \left[ 1 + \frac{\lambda^{-2}z^2}{\nu - 2} \right]^{-\frac{\nu+2}{2}}.
\]

Fernandez and Steel (1998) state a formula for computing the \( k \)th moment of (5.45), which is given by

\[
E(Z^k) = M_k \frac{\lambda^{k+1} + (-1)^k}{\lambda + \frac{1}{\lambda}}
\]

where

\[
M_k = \int_{0}^{\infty} 2u^k f(u) du
\]
and \( f(\cdot) \) denotes the standardised Student-\( t \) distribution. We can then obtain the moments of the univariate skew-\( t \) distribution via (5.46) and using expectation properties for functions of random variables: since \( Z = (Z^* - m)/s \), we have the following:

\[
E(Z^k) = \frac{E(Z^* - m)^k}{s^k}. \tag{5.47}
\]

As examples, the first four moments become

\[
\begin{align*}
E(Z) &= \frac{E(Z^*) - m}{s}, \\
E(Z^2) &= \frac{E(Z^*^2) - 2mE(Z^*) + m^2}{s^2}, \\
E(Z^3) &= \frac{E(Z^*^3) - 3mE(Z^*^2) + 3m^2E(Z^*) - m^3}{s^3}, \\
E(Z^4) &= \frac{E(Z^*^4) - 4mE(Z^*^3) + 6m^2E(Z^*^2) - 4m^3E(Z^*) + m^4}{s^4}.
\end{align*}
\]

### 5.9.2 Properties of the Skew-\( t \) Copula

Figure 5.6 displays contour plots of various bivariate PDFs using a combination of marginals and a copula, as described in equation (5.13). Each PDF is formed using standard normal marginal distributions and the corresponding copula, with dependence parameter \( \rho = 0.8 \) and degrees of freedom \( \eta = 3 \) (for non-Gaussian distributions). The top two plots display the bivariate normal and Student-\( t \) PDFs, and represent well-known findings in the literature, i.e. the Student-\( t \) has higher kurtosis than the normal. The remaining four plots display the bivariate skew-\( t \) distribution with four skewness parameter combinations for \((\gamma_1, \gamma_2)\): \((0.6, 0.8)\), \((0.6, 1.2)\), \((1.4, 0.8)\) and \((1.4, 1.2)\). Immediately, we see the flexibility of this distribution with these different parameter combinations. For the plots with \((\gamma_1, \gamma_2) = (0.6, 0.8), (1.4, 1.2)\) (indicating both negative/both positive skewness respectively), we notice the PDF becomes quite “pointy” towards the direction of the skewness, increasing the heaviness of the tails. For the plots with \((\gamma_1, \gamma_2) = (0.6, 1.2), (1.4, 0.8)\), we notice that the PDF becomes quite non-linear,
particularly in its tails due to there being a combination of positive and negative skewness represented by $\gamma_1$ and $\gamma_2$. The dependence parameter $\rho = 0.8$ does not appear to be affected by the different values of $(\gamma_1, \gamma_2)$. These plots clearly indicate the additional variety of dependence structures that can be captured using a skew-$t$ copula, compared to that of a Gaussian or Student-$t$ copula.

Figure 5.7 displays 1,000 data simulations from a variety of bivariate copulas $C(u_1, u_2)$ with dependence parameter $\rho = 0.9$ and degrees of freedom $\eta = 3$ (for non-Gaussian copulas). The top two plots display data simulated from the Gaussian and Student-$t$ copulas, and once again represent well-known findings, i.e. strong symmetric dependence, with data from the Student-$t$ copula being more widely dispersed in the centre. The remaining plots display simulations from the skew-$t$ copula, and we find some interesting properties. Firstly, there is obvious dependence between $u_1$ and $u_2$, with a wide dispersion of points similar to that of the Student-$t$ copula. Secondly, the dependence seems to be asymmetric, as the points seem to follow a “curved” shape with increasing $u_1$ and $u_2$. The shape of these “curves” is dependent on the values of the skewness parameters $(\gamma_1, \gamma_2)$, and in each case there is a clearly different shape. There is also a large cluster of points in the tails, indicating the existence of tail dependence similar to that of the Student-$t$ copula. These simulations also clearly display the flexibility of the skew-$t$ copula, compared to that of a Gaussian or Student-$t$ copula.
Figure 5.6: Contour plots of various bivariate PDFs using the given copula and standard normal marginal distributions, with dependence parameter $\rho = 0.8$ and degrees of freedom parameter $\eta = 3$ (for non-Gaussian distributions).
Figure 5.7: Simulations from various bivariate copulas with dependence parameter $\rho = 0.9$ and degrees of freedom parameter $\eta = 3$ (for non-Gaussian copulas).
5.9.3 VaR Back-Testing: Tables of Results

Table 5.8: VaR back-testing results for the unconditional, conditional and dynamic quantile tests for various models fitted to the returns of the DJI stock portfolio: 1-day horizon

<table>
<thead>
<tr>
<th></th>
<th>Violations</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\alpha}/\alpha )</th>
<th>( p_{uc} )</th>
<th>( p_{cc} )</th>
<th>( p_{dq} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.01 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH(_N)</td>
<td>24</td>
<td>0.0120</td>
<td>1.20</td>
<td>0.3835</td>
<td>0.5100</td>
<td>0.0146</td>
</tr>
<tr>
<td>GARCH(_T)</td>
<td>13</td>
<td>0.0065</td>
<td>0.65</td>
<td>0.0928</td>
<td>0.2245</td>
<td>0.0000</td>
</tr>
<tr>
<td>GARCH(_S)</td>
<td>10</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.0129</td>
<td>0.0433</td>
<td>0.0000</td>
</tr>
<tr>
<td>GJR-GARCH(_N)</td>
<td>23</td>
<td>0.0115</td>
<td>1.15</td>
<td>0.5102</td>
<td>0.6151</td>
<td>0.2246</td>
</tr>
<tr>
<td>GJR-GARCH(_T)</td>
<td>12</td>
<td>0.0060</td>
<td>0.60</td>
<td>0.0521</td>
<td>0.1416</td>
<td>0.0826</td>
</tr>
<tr>
<td>GJR-GARCH(_S)</td>
<td>9</td>
<td>0.0045</td>
<td>0.45</td>
<td>0.0056</td>
<td>0.0207</td>
<td>0.0078</td>
</tr>
<tr>
<td>DCC</td>
<td>17</td>
<td>0.0085</td>
<td>0.85</td>
<td>0.4889</td>
<td>0.6813</td>
<td>0.1671</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>18</td>
<td>0.0090</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COP-GJR-GARCH(_N)</td>
<td>17</td>
<td>0.0085</td>
<td>0.85</td>
<td>0.4889</td>
<td>0.2564</td>
<td>0.1365</td>
</tr>
<tr>
<td>COP-GJR-GARCH(_T)</td>
<td>11</td>
<td>0.0055</td>
<td>0.55</td>
<td>0.0270</td>
<td>0.0820</td>
<td>0.0000</td>
</tr>
<tr>
<td>COP-GJR-GARCH(_S)</td>
<td>15</td>
<td>0.0075</td>
<td>0.75</td>
<td>0.2397</td>
<td>0.4484</td>
<td>0.0001</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH(_N)</td>
<td>15</td>
<td>0.0075</td>
<td>0.75</td>
<td>0.2397</td>
<td>0.1301</td>
<td>0.0666</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH(_T)</td>
<td>11</td>
<td>0.0055</td>
<td>0.55</td>
<td>0.0270</td>
<td>0.0820</td>
<td>0.0410</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH(_S)</td>
<td>11</td>
<td>0.0055</td>
<td>0.55</td>
<td>0.0270</td>
<td>0.0820</td>
<td>0.0399</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>26</td>
<td>0.0130</td>
<td>1.30</td>
<td>0.1974</td>
<td>0.3084</td>
<td>0.0206</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH(_N)</td>
<td>93</td>
<td>0.0465</td>
<td>0.93</td>
<td>0.4676</td>
<td>0.5593</td>
<td>0.0183</td>
</tr>
<tr>
<td>GARCH(_T)</td>
<td>101</td>
<td>0.0505</td>
<td>1.01</td>
<td>0.9184</td>
<td>0.9148</td>
<td>0.0848</td>
</tr>
<tr>
<td>GARCH(_S)</td>
<td>89</td>
<td>0.0445</td>
<td>0.89</td>
<td>0.2505</td>
<td>0.4526</td>
<td>0.0445</td>
</tr>
<tr>
<td>GJR-GARCH(_N)</td>
<td>94</td>
<td>0.0470</td>
<td>0.94</td>
<td>0.5342</td>
<td>0.6235</td>
<td>0.0709</td>
</tr>
<tr>
<td>GJR-GARCH(_T)</td>
<td>100</td>
<td>0.0500</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9010</td>
<td>0.1646</td>
</tr>
<tr>
<td>GJR-GARCH(_S)</td>
<td>88</td>
<td>0.0440</td>
<td>0.88</td>
<td>0.2092</td>
<td>0.2628</td>
<td>0.0321</td>
</tr>
<tr>
<td>DCC</td>
<td>58</td>
<td>0.0290</td>
<td>0.58</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>63</td>
<td>0.0315</td>
<td>0.63</td>
<td>0.0000</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td>COP-GJR-GARCH(_N)</td>
<td>59</td>
<td>0.0295</td>
<td>0.59</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>COP-GJR-GARCH(_T)</td>
<td>94</td>
<td>0.0470</td>
<td>0.94</td>
<td>0.5342</td>
<td>0.7943</td>
<td>0.3598</td>
</tr>
<tr>
<td>COP-GJR-GARCH(_S)</td>
<td>91</td>
<td>0.0455</td>
<td>0.91</td>
<td>0.3487</td>
<td>0.6460</td>
<td>0.2981</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH(_N)</td>
<td>62</td>
<td>0.0310</td>
<td>0.62</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH(_T)</td>
<td>92</td>
<td>0.0460</td>
<td>0.92</td>
<td>0.4057</td>
<td>0.6611</td>
<td>0.4287</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH(_S)</td>
<td>91</td>
<td>0.0455</td>
<td>0.91</td>
<td>0.3487</td>
<td>0.5909</td>
<td>0.2332</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>136</td>
<td>0.0680</td>
<td>1.36</td>
<td>0.0004</td>
<td>0.0013</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Table 5.9: VaR back-testing results for the unconditional, conditional and dynamic quantile tests for various models fitted to the returns of the DJI stock portfolio: 5-day horizon

<table>
<thead>
<tr>
<th>Model</th>
<th>Violations</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\alpha}/\alpha$</th>
<th>$p_{uc}$</th>
<th>$p_{cc}$</th>
<th>$p_{dq}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH$_N$</td>
<td>4</td>
<td>0.0100</td>
<td><strong>1.00</strong></td>
<td>1.0000</td>
<td>0.9603</td>
<td>0.9754</td>
</tr>
<tr>
<td>GARCH$_T$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9696</td>
</tr>
<tr>
<td>GARCH$_S$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9641</td>
</tr>
<tr>
<td>GJR-GARCH$_N$</td>
<td>3</td>
<td>0.0075</td>
<td>0.75</td>
<td>0.5991</td>
<td>0.8535</td>
<td>0.9882</td>
</tr>
<tr>
<td>GJR-GARCH$_T$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9670</td>
</tr>
<tr>
<td>GJR-GARCH$_S$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9662</td>
</tr>
<tr>
<td>DCC</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td><strong>0.0000</strong></td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>1</td>
<td>0.0025</td>
<td>0.25</td>
<td>0.0714</td>
<td>0.1979</td>
<td>0.8444</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_N$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9840</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_T$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9415</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_S$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9463</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_N$</td>
<td>1</td>
<td>0.0025</td>
<td>0.25</td>
<td>0.0714</td>
<td>0.1979</td>
<td>0.8613</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_T$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9534</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_S$</td>
<td>2</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.2660</td>
<td>0.5359</td>
<td>0.9421</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>6</td>
<td>0.0150</td>
<td>1.50</td>
<td>0.3494</td>
<td>0.5859</td>
<td>0.0150</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH$_N$</td>
<td>18</td>
<td>0.0450</td>
<td>0.90</td>
<td>0.6409</td>
<td>0.8817</td>
<td>0.7410</td>
</tr>
<tr>
<td>GARCH$_T$</td>
<td>20</td>
<td>0.0500</td>
<td><strong>1.00</strong></td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.6802</td>
</tr>
<tr>
<td>GARCH$_S$</td>
<td>18</td>
<td>0.0450</td>
<td>0.90</td>
<td>0.6409</td>
<td>0.8817</td>
<td>0.7139</td>
</tr>
<tr>
<td>GJR-GARCH$_N$</td>
<td>17</td>
<td>0.0425</td>
<td>0.85</td>
<td>0.4804</td>
<td>0.3686</td>
<td>0.6284</td>
</tr>
<tr>
<td>GJR-GARCH$_T$</td>
<td>20</td>
<td>0.0500</td>
<td><strong>1.00</strong></td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.7009</td>
</tr>
<tr>
<td>GJR-GARCH$_S$</td>
<td>19</td>
<td>0.0475</td>
<td>0.95</td>
<td>0.8171</td>
<td>0.9709</td>
<td>0.6489</td>
</tr>
<tr>
<td>DCC</td>
<td>5</td>
<td>0.0125</td>
<td>0.25</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>8</td>
<td>0.0200</td>
<td>0.40</td>
<td>0.0018</td>
<td>0.0027</td>
<td>0.0022</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_N$</td>
<td>10</td>
<td>0.0250</td>
<td>0.50</td>
<td>0.0114</td>
<td>0.0324</td>
<td>0.0116</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_T$</td>
<td>13</td>
<td>0.0325</td>
<td>0.65</td>
<td>0.0871</td>
<td>0.1520</td>
<td>0.5988</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_S$</td>
<td>12</td>
<td>0.0300</td>
<td>0.60</td>
<td>0.0481</td>
<td>0.0998</td>
<td>0.1330</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_N$</td>
<td>9</td>
<td>0.0225</td>
<td>0.45</td>
<td>0.0048</td>
<td>0.0158</td>
<td>0.0559</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_T$</td>
<td>13</td>
<td>0.0325</td>
<td>0.65</td>
<td>0.0871</td>
<td>0.1520</td>
<td>0.5971</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_S$</td>
<td>11</td>
<td>0.0275</td>
<td>0.55</td>
<td>0.0245</td>
<td>0.0597</td>
<td>0.3526</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>27</td>
<td>0.0675</td>
<td>1.35</td>
<td>0.1265</td>
<td>0.3025</td>
<td>0.2445</td>
</tr>
</tbody>
</table>
Table 5.10: VaR back-testing results for the unconditional, conditional and dynamic quantile tests for various models fitted to the returns of the DJI stock portfolio: 10-day horizon

<table>
<thead>
<tr>
<th>Model</th>
<th>Violations</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\alpha}/\alpha )</th>
<th>( p_{uc} )</th>
<th>( p_{cc} )</th>
<th>( p_{dq} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH( N )</td>
<td>2</td>
<td>0.0100</td>
<td><strong>1.00</strong></td>
<td>1.0000</td>
<td>0.9799</td>
<td>0.8262</td>
</tr>
<tr>
<td>GARCH( T )</td>
<td>1</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.4315</td>
<td>0.7339</td>
<td>0.9799</td>
</tr>
<tr>
<td>GARCH( S )</td>
<td>1</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.4315</td>
<td>0.7339</td>
<td>0.9774</td>
</tr>
<tr>
<td>GJR-GARCH( N )</td>
<td>2</td>
<td>0.0100</td>
<td><strong>1.00</strong></td>
<td>1.0000</td>
<td>0.9799</td>
<td>0.8386</td>
</tr>
<tr>
<td>GJR-GARCH( T )</td>
<td>1</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.4315</td>
<td>0.7339</td>
<td>0.9822</td>
</tr>
<tr>
<td>GJR-GARCH( S )</td>
<td>1</td>
<td>0.0050</td>
<td>0.50</td>
<td>0.4315</td>
<td>0.7339</td>
<td>0.9815</td>
</tr>
<tr>
<td>DCC</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.0450</td>
<td>0.1353</td>
<td>N/A</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.0450</td>
<td>0.1353</td>
<td>N/A</td>
</tr>
<tr>
<td>COP-GJR-GARCH( N )</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.0450</td>
<td>0.1353</td>
<td>N/A</td>
</tr>
<tr>
<td>COP-GJR-GARCH( T )</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.0450</td>
<td>0.1353</td>
<td>N/A</td>
</tr>
<tr>
<td>COP-GJR-GARCH( S )</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.0450</td>
<td>0.1353</td>
<td>N/A</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH( N )</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.0450</td>
<td>0.1353</td>
<td>N/A</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH( T )</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.0450</td>
<td>0.1353</td>
<td>N/A</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH( S )</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.0450</td>
<td>0.1353</td>
<td>N/A</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>3</td>
<td>0.0150</td>
<td>1.50</td>
<td>0.5082</td>
<td>0.7634</td>
<td>0.9626</td>
</tr>
<tr>
<td>GARCH( N )</td>
<td>8</td>
<td>0.0400</td>
<td>0.80</td>
<td>0.5020</td>
<td>0.5769</td>
<td>0.7675</td>
</tr>
<tr>
<td>GARCH( T )</td>
<td>11</td>
<td>0.0550</td>
<td>1.10</td>
<td>0.7493</td>
<td>0.4964</td>
<td>0.8911</td>
</tr>
<tr>
<td>GARCH( S )</td>
<td>9</td>
<td>0.0450</td>
<td>0.90</td>
<td>0.7416</td>
<td>0.6214</td>
<td>0.8929</td>
</tr>
<tr>
<td>GJR-GARCH( N )</td>
<td>8</td>
<td>0.0400</td>
<td>0.80</td>
<td>0.5020</td>
<td>0.5769</td>
<td>0.3639</td>
</tr>
<tr>
<td>GJR-GARCH( T )</td>
<td>10</td>
<td>0.0500</td>
<td><strong>1.00</strong></td>
<td>1.0000</td>
<td>0.5889</td>
<td>0.8833</td>
</tr>
<tr>
<td>GJR-GARCH( S )</td>
<td>8</td>
<td>0.0400</td>
<td>0.80</td>
<td>0.5020</td>
<td>0.5769</td>
<td>0.9416</td>
</tr>
<tr>
<td>DCC</td>
<td>4</td>
<td>0.0200</td>
<td>0.40</td>
<td><strong>0.0275</strong></td>
<td>0.0838</td>
<td>0.0000</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>4</td>
<td>0.0200</td>
<td>0.40</td>
<td><strong>0.0275</strong></td>
<td>0.0838</td>
<td>0.0000</td>
</tr>
<tr>
<td>COP-GJR-GARCH( N )</td>
<td>5</td>
<td>0.0250</td>
<td>0.50</td>
<td>0.0737</td>
<td>0.1822</td>
<td><strong>0.0011</strong></td>
</tr>
<tr>
<td>COP-GJR-GARCH( T )</td>
<td>10</td>
<td>0.0500</td>
<td><strong>1.00</strong></td>
<td>1.0000</td>
<td>0.5889</td>
<td>0.4512</td>
</tr>
<tr>
<td>COP-GJR-GARCH( S )</td>
<td>7</td>
<td>0.0350</td>
<td>0.70</td>
<td>0.3047</td>
<td>0.4646</td>
<td>0.7692</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH( N )</td>
<td>5</td>
<td>0.0250</td>
<td>0.50</td>
<td>0.0737</td>
<td>0.1822</td>
<td>0.4337</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH( T )</td>
<td>8</td>
<td>0.0400</td>
<td>0.80</td>
<td>0.5020</td>
<td>0.5769</td>
<td>0.1918</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH( S )</td>
<td>8</td>
<td>0.0400</td>
<td>0.80</td>
<td>0.5020</td>
<td>0.5769</td>
<td>0.8571</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>11</td>
<td>0.0550</td>
<td>1.10</td>
<td>0.7493</td>
<td>0.4964</td>
<td>0.1171</td>
</tr>
</tbody>
</table>
Table 5.11: VaR back-testing results for the unconditional, conditional and dynamic quantile tests for various models fitted to the returns of the DJI stock portfolio: 20-day horizon

<table>
<thead>
<tr>
<th>Model</th>
<th>Violations</th>
<th>α̂</th>
<th>α̂/α</th>
<th>p uc</th>
<th>p cc</th>
<th>p dq</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH$_N$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>1.0000</td>
</tr>
<tr>
<td>GARCH$_T$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>1.0000</td>
</tr>
<tr>
<td>GARCH$_S$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.9999</td>
</tr>
<tr>
<td>GJR-GARCH$_N$</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.1563</td>
<td>0.3697</td>
<td>N/A</td>
</tr>
<tr>
<td>GJR-GARCH$_T$</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.1563</td>
<td>0.3697</td>
<td>N/A</td>
</tr>
<tr>
<td>GJR-GARCH$_S$</td>
<td>0</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.1563</td>
<td>0.3697</td>
<td>N/A</td>
</tr>
<tr>
<td>DCC</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.6542</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.4420</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_N$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.8710</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_T$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.9992</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_S$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.9992</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_N$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.9159</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_T$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.9994</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_S$</td>
<td>1</td>
<td>0.0100</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.9898</td>
<td>0.9995</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>2</td>
<td>0.0200</td>
<td>2.00</td>
<td>0.3763</td>
<td>0.6422</td>
<td>0.2395</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH$_N$</td>
<td>6</td>
<td>0.0600</td>
<td>1.20</td>
<td>0.6560</td>
<td>0.5699</td>
<td>0.7745</td>
</tr>
<tr>
<td>GARCH$_T$</td>
<td>6</td>
<td>0.0600</td>
<td>1.20</td>
<td>0.6560</td>
<td>0.5699</td>
<td>0.5517</td>
</tr>
<tr>
<td>GARCH$_S$</td>
<td>6</td>
<td>0.0600</td>
<td>1.20</td>
<td>0.6560</td>
<td>0.5699</td>
<td>0.5182</td>
</tr>
<tr>
<td>GJR-GARCH$_N$</td>
<td>6</td>
<td>0.0600</td>
<td>1.20</td>
<td>0.6560</td>
<td>0.6081</td>
<td>0.0484</td>
</tr>
<tr>
<td>GJR-GARCH$_T$</td>
<td>5</td>
<td>0.0500</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.7662</td>
<td>0.7307</td>
</tr>
<tr>
<td>GJR-GARCH$_S$</td>
<td>5</td>
<td>0.0500</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.7662</td>
<td>0.6923</td>
</tr>
<tr>
<td>DCC</td>
<td>1</td>
<td>0.0100</td>
<td>0.20</td>
<td>0.0261</td>
<td>0.0869</td>
<td>0.6759</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>2</td>
<td>0.0200</td>
<td>0.40</td>
<td>0.1191</td>
<td>0.0181</td>
<td>0.0039</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_N$</td>
<td>3</td>
<td>0.0300</td>
<td>0.60</td>
<td>0.3230</td>
<td>0.1022</td>
<td>0.2030</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_T$</td>
<td>3</td>
<td>0.0300</td>
<td>0.60</td>
<td>0.3230</td>
<td>0.1022</td>
<td>0.2145</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_S$</td>
<td>2</td>
<td>0.0200</td>
<td>0.40</td>
<td>0.1191</td>
<td>0.2939</td>
<td>0.9254</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_N$</td>
<td>2</td>
<td>0.0200</td>
<td>0.40</td>
<td>0.1191</td>
<td>0.2939</td>
<td>0.9074</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_T$</td>
<td>2</td>
<td>0.0200</td>
<td>0.40</td>
<td>0.1191</td>
<td>0.2939</td>
<td>0.9219</td>
</tr>
<tr>
<td>ACOP-GJR-GARCH$_S$</td>
<td>2</td>
<td>0.0200</td>
<td>0.40</td>
<td>0.1191</td>
<td>0.2939</td>
<td>0.9061</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>7</td>
<td>0.0700</td>
<td>1.40</td>
<td>0.3855</td>
<td>0.5283</td>
<td>0.5778</td>
</tr>
</tbody>
</table>

218
Table 5.12: Values of the VaR criterion function of Koenker and Bassett (1978) for various models fitted to the returns of the DJI stock portfolio: 1 and 5-day horizons

<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CF Rank</td>
<td>CF Rank</td>
<td>CF Rank</td>
<td>CF Rank</td>
</tr>
<tr>
<td>GARCH$_N$</td>
<td>0.9638 3</td>
<td>3.3761 6</td>
<td>0.3493 3</td>
<td>1.3373 6</td>
</tr>
<tr>
<td>GARCH$_T$</td>
<td>0.9070 5</td>
<td>3.3472 4</td>
<td>0.3485 2</td>
<td>1.3343 5</td>
</tr>
<tr>
<td>GARCH$_S$</td>
<td>1.0161 10</td>
<td>3.3587 5</td>
<td>0.3639 6</td>
<td>1.3415 7</td>
</tr>
<tr>
<td>GJR-GARCH$_N$</td>
<td>0.9520 1</td>
<td>3.3433 3</td>
<td>0.3534 4</td>
<td>1.3214 1</td>
</tr>
<tr>
<td>GJR-GARCH$_T$</td>
<td>0.9755 4</td>
<td>3.3293 1</td>
<td>0.3572 5</td>
<td>1.3290 2</td>
</tr>
<tr>
<td>GJR-GARCH$_S$</td>
<td>1.0108 8</td>
<td>3.3352 2</td>
<td>0.3717 7</td>
<td>1.3334 3</td>
</tr>
<tr>
<td>DCC</td>
<td>1.2273 15</td>
<td>4.1866 15</td>
<td>0.6648 15</td>
<td>2.0872 15</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>1.1793 14</td>
<td>4.0526 14</td>
<td>0.5508 14</td>
<td>1.8464 14</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_N$</td>
<td>1.1421 13</td>
<td>3.9296 13</td>
<td>0.4734 13</td>
<td>1.7116 13</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_T$</td>
<td>1.0107 7</td>
<td>3.4354 8</td>
<td>0.3858 8</td>
<td>1.3728 8</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_S$</td>
<td>1.0141 9</td>
<td>3.4660 11</td>
<td>0.4021 11</td>
<td>1.3984 11</td>
</tr>
<tr>
<td>GJR-GARCH$_N$</td>
<td>1.1216 12</td>
<td>3.8866 12</td>
<td>0.4678 12</td>
<td>1.6903 12</td>
</tr>
<tr>
<td>GJR-GARCH$_T$</td>
<td>1.0186 11</td>
<td>3.4334 7</td>
<td>0.3999 10</td>
<td>1.3926 9</td>
</tr>
<tr>
<td>GJR-GARCH$_S$</td>
<td>1.0071 6</td>
<td>3.4516 10</td>
<td>0.3967 9</td>
<td>1.3956 10</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>0.9560 2</td>
<td>3.4375 9</td>
<td>0.3452 1</td>
<td>1.3341 4</td>
</tr>
</tbody>
</table>

Table 5.13: Values of the VaR criterion function of Koenker and Bassett (1978) for various models fitted to the returns of the DJI stock portfolio: 10 and 20-day horizons

<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CF Rank</td>
<td>CF Rank</td>
<td>CF Rank</td>
<td>CF Rank</td>
</tr>
<tr>
<td>GARCH$_N$</td>
<td>0.2296 2</td>
<td>0.8931 1</td>
<td>0.1911 6</td>
<td>0.6730 3</td>
</tr>
<tr>
<td>GARCH$_T$</td>
<td>0.2376 3</td>
<td>0.9128 5</td>
<td>0.1787 4</td>
<td>0.6787 5</td>
</tr>
<tr>
<td>GARCH$_S$</td>
<td>0.2516 5</td>
<td>0.9157 6</td>
<td>0.1824 5</td>
<td>0.6847 6</td>
</tr>
<tr>
<td>GJR-GARCH$_N$</td>
<td>0.2542 6</td>
<td>0.8979 2</td>
<td>0.1689 2</td>
<td>0.6679 2</td>
</tr>
<tr>
<td>GJR-GARCH$_T$</td>
<td>0.2428 4</td>
<td>0.9023 4</td>
<td>0.1612 1</td>
<td>0.6622 1</td>
</tr>
<tr>
<td>GJR-GARCH$_S$</td>
<td>0.2580 7</td>
<td>0.9021 3</td>
<td>0.1763 3</td>
<td>0.6761 4</td>
</tr>
<tr>
<td>DCC</td>
<td>0.4982 15</td>
<td>1.6039 15</td>
<td>0.3929 15</td>
<td>1.2819 15</td>
</tr>
<tr>
<td>GJR-ADCC</td>
<td>0.4065 14</td>
<td>1.4222 14</td>
<td>0.3397 14</td>
<td>1.0911 14</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_N$</td>
<td>0.3524 12</td>
<td>1.2393 13</td>
<td>0.2969 12</td>
<td>0.9859 12</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_T$</td>
<td>0.2638 8</td>
<td>0.9389 7</td>
<td>0.2089 9</td>
<td>0.7208 8</td>
</tr>
<tr>
<td>COP-GJR-GARCH$_S$</td>
<td>0.2785 11</td>
<td>0.9649 10</td>
<td>0.2140 11</td>
<td>0.7621 11</td>
</tr>
<tr>
<td>GJR-GARCH$_N$</td>
<td>0.3533 13</td>
<td>1.2279 12</td>
<td>0.3005 13</td>
<td>0.9987 13</td>
</tr>
<tr>
<td>GJR-GARCH$_T$</td>
<td>0.2729 9</td>
<td>0.9484 9</td>
<td>0.2065 8</td>
<td>0.7303 9</td>
</tr>
<tr>
<td>GJR-GARCH$_S$</td>
<td>0.2761 10</td>
<td>0.9667 11</td>
<td>0.2125 10</td>
<td>0.7426 10</td>
</tr>
<tr>
<td>ASTC-DTGARCH</td>
<td>0.2168 1</td>
<td>0.9471 8</td>
<td>0.2025 7</td>
<td>0.6905 7</td>
</tr>
</tbody>
</table>
Figure 5.8: Plots of mean VaR estimates of the given model groups against actual DJI portfolio returns between 24 January 2002 and 31 December 2009: 1-day horizon with (a) $\alpha = 0.01$; (b) $\alpha = 0.05$. 

(a) 

(b)
Figure 5.9: Plots of mean VaR estimates of the given model groups against actual DJI portfolio returns between 24 January 2002 and 31 December 2009: 5-day horizon with (a) $\alpha = 0.01$; (b) $\alpha = 0.05$. 

(a) 

(b)
Figure 5.10: Plots of mean VaR estimates of the given model groups against actual DJI portfolio returns between 24 January 2002 and 31 December 2009: 10-day horizon with (a) $\alpha = 0.01$; (b) $\alpha = 0.05$. 

(a) 

(b) 

222
Figure 5.11: Plots of mean VaR estimates of the given model groups against actual DJI portfolio returns between 24 January 2002 and 31 December 2009: 20-day horizon with (a) $\alpha = 0.01$; (b) $\alpha = 0.05$. 

(a)

(b)
Chapter 6

Conclusion

Stock market analysts and researchers are always investigating ways to improve their modelling and prediction of stock market volatility and correlation, particularly during or after times of extreme market events (e.g. the Global Financial Crisis of 2008). This thesis has presented a series of multivariate GARCH models that can assist in this prediction when an investor holds a portfolio of stocks. In particular, Bayesian techniques were applied to estimate the parameters of the proposed models which has additional flexibility over classical maximum likelihood methods by incorporating prior information into the analysis and the ability to account for parameter uncertainty.

The work was introduced in Chapter 1 where the background of volatility modelling, model selection and parameter estimation techniques were discussed. Applications of volatility modelling were then described, demonstrating its importance in various market activities.

In Chapter 2, the main Bayesian techniques existing in the literature were reviewed that are used to estimate model parameters. Bayes’ rule was stated, which is the foundation of Bayesian inference and shows how to derive the joint posterior distribution of model parameters. Different ways of simulating from this joint posterior distribution, namely the Metropolis-Hastings (MH) and Delayed Rejection (DR) algorithms were then described. The Gibbs sampler was also described, which splits the joint posterior into
conditional posteriors for simulation and links them using the idea of a Markov chain, which was also discussed. This approach is called Markov chain Monte Carlo (MCMC) simulation and was used in the remainder of this thesis for the proposed models.

In Chapter 3, a multivariate GARCH model (the PVECH model) was presented whose parameters were estimated via Bayesian techniques. Instead of estimating parameters using standard parameter restrictions, a new technique was proposed that “relaxes” these standard restrictions so that a wider parameter space could be explored. This was done by utilising the unconditional expectations of the model and simply rejecting any parameter value in an MCMC algorithm that did not satisfy positive definiteness (PD) and covariance stationarity (CS) directly. A simulation study was presented which showed favourable estimation performance for the PVECH model against an equivalent BEKK model fitted via maximum likelihood, particularly when true parameter values lied outside the standard parameter restrictions. Finally, an application to stock returns of two international stock market indices was presented, where some PVECH parameter estimates were obtained that lied outside the standard parameter restrictions, while the equivalent BEKK model showed that corresponding estimates were insignificant. The PVECH model also gave higher volatility estimates than that of the equivalent BEKK model.

In Chapter 4, the PVECH model from Chapter 3 was extended to allow for well-known asymmetric effects, resulting in the APVECH model family. It extended the asymmetric DVECH model of De Goeij and Marquering (2004) by including residual and volatility spill-over effects and asymmetries, plus additional asymmetric terms for (co)variance intercepts. A Gibbs sampler was designed to estimate the parameters of the model, applying the same technique to relax the standard parameter restrictions usually imposed, as for the PVECH model. To determine the number of asymmetric effects to include for a dataset, the approximate model selection technique of Congdon (2006) was implemented. A simulation study on four selected APVECH models showed unbiased parameter estimates and the ability for the model selection technique to select a simpler
model if required; the precision improved with a larger sample size. An application of the
APVECH model family to four international stock market indices revealed the presence
of volatility and covariance asymmetry in the stock returns, with the most complex model
in the APVECH model family chosen in model selection for the majority of market pairs.

In Chapter 5, the ASTC-DTGARCH model was proposed which is a copula-GARCH
model that incorporates asymmetry, skewness and excess kurtosis in all components of
the model. This extends current copula-GARCH models in the literature which allow
for these three properties in only some model components. The multivariate skew-\(t\)
distribution of Bauwens and Laurent (2005) was used to construct a skew-\(t\) copula for
the model. A two-step Bayesian approach was implemented to estimate the parameters
of the model (as in Ausin and Lopes, 2010), firstly by simulating from the joint posterior
of the parameters for each marginal distribution, then simulating from the joint posterior
of the copula parameters, conditional on estimates from the marginals. This reduces
computational cost for models of high dimension. An application of the model to stock
returns from a DJI portfolio showed the presence of asymmetry in mean, variance and
correlation components, as well as skewness and excess kurtosis in all marginals. The
ASTC-DTGARCH model also performed well in estimating VaR for the portfolio against
a variety of other univariate and multivariate GARCH models, based on the results of
various back-testing procedures.

6.1 Future Research

Suggestions for future research in this area include, but are not limited to, the following:

- An approximation was used for the unconditional covariances of the APVECH
  model family in Chapter 4, and showed that even under residual normality that an
  exact expression would be difficult, if not impossible to derive. While this approx-
  imation did not appear to impact on the results in the simulation and empirical
  studies, further work in this area could involve deriving incomplete moments of the

226
multivariate Student-\( t \) distribution, or obtaining better approximations of unconditional covariances through the use of regression and/or Newton’s method.

- The error distribution for the APVECH model family could be extended to a multivariate skew-\( t \) distribution, as utilised in Chapter 5.

- Additional asymmetric terms could be included in the mean equations and/or with lagged covariances to fully generalise the APVECH model family. However, this would impact on the unconditional expectations of the model and may lead to approximations to be sought (as discussed above).

- The ASTC-DTGARCH model could be extended to have more than two regimes in the marginals, and using the two-step Bayesian approach for estimation, would thus have minimal impact on computational expense.

- The back-testing from Chapter 5 could be extended to incorporate Expected Shortfall (ES) which is the expected loss, given that a portfolio return has exceeded VaR. This is a coherent risk measure (see for e.g. Artzner et al., 1997, 1999), while VaR is not, and also has more attractive statistical properties than VaR. This would be particularly useful for financial institutions that use small significance levels (i.e. \( \alpha \leq 0.01 \)).
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Parameter numbers for selected VECH and BEKK models under different dimensions</td>
<td>38</td>
</tr>
<tr>
<td>3.2</td>
<td>Parameter estimation results for the PVECH and BEKK models fitted to the S&amp;P 500/FTSE 100 bivariate returns.</td>
<td>64</td>
</tr>
<tr>
<td>3.3</td>
<td>Model selection results for the PVECH and BEKK models fitted to the S&amp;P 500/FTSE 100 bivariate returns.</td>
<td>67</td>
</tr>
<tr>
<td>3.4</td>
<td>Parameter estimates for PVECH and (transformed) BEKK models fitted to simulated data from Model 1</td>
<td>70</td>
</tr>
<tr>
<td>3.5</td>
<td>95% credible intervals and coverages for PVECH and (transformed) BEKK models fitted to simulated data from Model 1</td>
<td>70</td>
</tr>
<tr>
<td>3.6</td>
<td>Parameter estimates for PVECH and (transformed) BEKK models fitted to simulated data from Model 2</td>
<td>71</td>
</tr>
<tr>
<td>3.7</td>
<td>95% credible intervals and coverages for PVECH and (transformed) BEKK models fitted to simulated data from Model 2</td>
<td>71</td>
</tr>
<tr>
<td>3.8</td>
<td>Parameter estimation results from fitting PVECH and BEKK models to a simulated dataset from Model 1 of sample size $n = 50,000$</td>
<td>72</td>
</tr>
<tr>
<td>3.9</td>
<td>Parameter estimation results from fitting PVECH and BEKK models to a simulated dataset from Model 2 of sample size $n = 50,000$</td>
<td>72</td>
</tr>
<tr>
<td>4.1</td>
<td>Means of the 100 parameter posterior means and 95% credible intervals for models from the APVECH model family: Data simulated from Model 1 with $n = 1,000$</td>
<td>121</td>
</tr>
</tbody>
</table>
4.2 Means of the 100 parameter posterior means and 95% credible intervals
for models from the APVECH model family: Data simulated from Model
2 with $n = 1,000$ ......................................................... 122
4.3 Means of the 100 parameter posterior means and 95% credible intervals
for models from the APVECH model family: Data simulated from Model
3 with $n = 1,000$ ......................................................... 123
4.4 Means of the 100 parameter posterior means and 95% credible intervals
for models from the APVECH model family: Data simulated from Model
4 with $n = 1,000$ ......................................................... 124
4.5 Means of the 100 parameter posterior means and 95% credible intervals
for models from the APVECH model family: Data simulated from Model
1 with $n = 3,000$ ......................................................... 125
4.6 Means of the 100 parameter posterior means and 95% credible intervals
for models from the APVECH model family: Data simulated from Model
2 with $n = 3,000$ ......................................................... 126
4.7 Means of the 100 parameter posterior means and 95% credible intervals
for models from the APVECH model family: Data simulated from Model
3 with $n = 3,000$ ......................................................... 127
4.8 Means of the 100 parameter posterior means and 95% credible intervals
for models from the APVECH model family: Data simulated from Model
4 with $n = 3,000$ ......................................................... 128
4.9 Mean posterior model probabilities and model selection counts for the
APVECH family of models in the simulation study with $n = 1,000$ .... 128
4.10 Mean posterior model probabilities and model selection counts for the
APVECH family of models in the simulation study with $n = 3,000$ .... 129
4.11 Posterior model probabilities for the APVECH family of models fitted to
each bivariate combination of S&P 500 (US), FTSE 100 (UK), All Ordinaries
(AU) and Nikkei 225 (JP) returns. .................................... 129
4.12 Parameter posterior means and 95% credible intervals for the APVECH
family of models fitted to the S&P 500/FTSE 100 bivariate returns . . .

130

4.13 Acceptance rates for the APVECH family of models fitted to the S&P
500/FTSE 100 bivariate returns . . . . . . . . . . . . . . . . . . . . . . .

130

4.14 Parameter posterior means and 95% credible intervals for the APVECH
family of models fitted to the S&P 500/All Ordinaries bivariate returns .

131

4.15 Acceptance rates for the APVECH family of models fitted to the S&P
500/All Ordinaries bivariate returns . . . . . . . . . . . . . . . . . . . . .

131

4.16 Parameter posterior means and 95% credible intervals for the APVECH
family of models fitted to the S&P 500/Nikkei 225 bivariate returns . . .

132

4.17 Acceptance rates for the APVECH family of models fitted to the S&P
500/Nikkei 225 bivariate returns . . . . . . . . . . . . . . . . . . . . . . .

132

4.18 Parameter posterior means and 95% credible intervals for the APVECH
family of models fitted to the FTSE 100/All Ordinaries bivariate returns

133

4.19 Acceptance rates for the APVECH family of models fitted to the FTSE
100/All Ordinaries bivariate returns . . . . . . . . . . . . . . . . . . . . .

133

4.20 Parameter posterior means and 95% credible intervals for the APVECH
family of models fitted to the FTSE 100/Nikkei 225 bivariate returns . .

134

4.21 Acceptance rates for the APVECH family of models fitted to the FTSE
100/Nikkei 225 bivariate returns . . . . . . . . . . . . . . . . . . . . . . .

134

4.22 Parameter posterior means and 95% credible intervals for the APVECH
family of models fitted to the All Ordinaries/Nikkei 225 bivariate returns

135

4.23 Acceptance rates for the APVECH family of models fitted to the All Ordinaries/Nikkei 225 bivariate returns . . . . . . . . . . . . . . . . . . . .

135

4.24 Potential scale reduction factors for Model 4 of the APVECH family of
models fitted to the All Ordinaries/Nikkei 225 bivariate returns . . . . .

230

136


5.1 Sample statistics for the DJI stock returns of CSCO, XOM, GE, INTC and MSFT, as well as their corresponding equally-weighted portfolio returns PR between 11 February 1994 and 31 December 2009.

5.2 Sample correlation coefficients for each bivariate combination of the DJI stock returns CSCO, XOM, GE, INTC and MSFT between 11 February 1994 and 31 December 2009.

5.3 Mean parameter estimates of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio: Autoregressive and GARCH parameters of the marginal distributions.

5.4 Mean parameter estimates of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio: Threshold and error parameters of the marginal distributions.

5.5 Mean parameter estimates of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio: GARCH parameters $\alpha^{(k)}, \beta^{(k)}$, $k = 1, 2, 3, 4$ and copula distribution parameters.

5.6 Mean parameter estimates of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio: Correlation parameters $\bar{P}^{(k)}$, $k = 1, 2, 3, 4$ of the copula.

5.7 Potential scale reduction factors (PSRFs) for parameters of the ASTC-DTGARCH model fitted to an arbitrary set of 2,000 observations of the DJI stock portfolio returns.

5.8 VaR back-testing results for the unconditional, conditional and dynamic quantile tests for various models fitted to the returns of the DJI stock portfolio: 1-day horizon.

5.9 VaR back-testing results for the unconditional, conditional and dynamic quantile tests for various models fitted to the returns of the DJI stock portfolio: 5-day horizon.
5.10 VaR back-testing results for the unconditional, conditional and dynamic quantile tests for various models fitted to the returns of the DJI stock portfolio: 10-day horizon ........................................ 217
5.11 VaR back-testing results for the unconditional, conditional and dynamic quantile tests for various models fitted to the returns of the DJI stock portfolio: 20-day horizon ........................................ 218
5.12 Values of the VaR criterion function of Koenker and Bassett (1978) for various models fitted to the returns of the DJI stock portfolio: 1 and 5-day horizons .................................................. 219
5.13 Values of the VaR criterion function of Koenker and Bassett (1978) for various models fitted to the returns of the DJI stock portfolio: 10 and 20-day horizons .................................................. 219
List of Figures

3.1 Plot of parameter numbers for selected VEC and BEKK models under different dimensions ......................................................... 39

3.2 (a) Time series plot, and (b) Scatterplot of daily log returns for S&P 500 and FTSE 100 indices between 2 November 1987 and 30 June 2005. . . . 63

3.3 MCMC iterates for the parameters $m_{22}$, $c_{11}$, $a_{22}$ and $b_{12}$ of the PVECH model fitted to the S&P 500/FTSE 100 bivariate returns, each with 5 different starting values. ............................................... 66

3.4 Estimated volatility, covariance and correlation for the S&P 500/FTSE 100 bivariate returns, using the Bayesian MCMC and classical maximum likelihood methods from the PVECH and BEKK models (respectively). . 68

4.1 Plots of estimated conditional correlations $\hat{\rho}_t$ for each bivariate combination of S&P 500 (US), FTSE 100 (UK), All Ordinaries (AU) and Nikkei 225 (JP) returns between 4 January 1984 and 4 August 2006. ............. 115

4.2 MCMC iterates for the parameters $\gamma_{11}$, $b_{22}$, $\alpha_{12,3}$ and $\nu$ for Model 4 of the APVECH family of models fitted to the All Ordinaries/Nikkei 225 bivariate returns, each with 5 different starting values. .................. 137

5.1 Plots of the DJI stock returns CSCO, XOM, GE, INTC and MSFT, as well as equally-weighted portfolio returns PR between 11 February 1994 and 31 December 2009. ......................................................... 174
5.2 Plots of estimates of $\lambda_i$ and $\nu_i$, $i = 1, \ldots, 5$, from the marginal distributions of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio. ................................................. 198

5.3 MCMC iterates for the parameters $r_1$, $c_2^{(1)}$, $b_4^{(2)}$, and $\gamma_1$ for the ASTC-DTGARCH model fitted to an arbitrary set of 2,000 observations of the DJI stock portfolio returns, each with 5 different starting values. ........ 199

5.4 Plots of the estimated correlation matrix $P_t = [\rho_{ij,t}]$ over a selection of 2,000 observations for the dependence structure of the ASTC-DTGARCH model fitted to the returns of the DJI stock portfolio. ...................... 200

5.5 Plots of the univariate skew-t PDF with different values of the parameters $\lambda$ and $\nu$. The fixed values in plots (a) and (b) are $\nu = 8$ and $\lambda = 0.7$ (respectively). ................................................................. 206

5.6 Contour plots of various bivariate PDFs using the given copula and standard normal marginal distributions, with dependence parameter $\rho = 0.8$ and degrees of freedom parameter $\eta = 3$ (for non-Gaussian distributions). 213

5.7 Simulations from various bivariate copulas with dependence parameter $\rho = 0.9$ and degrees of freedom parameter $\eta = 3$ (for non-Gaussian copulas). 214

5.8 Plots of mean VaR estimates of the given model groups against actual DJI portfolio returns between 24 January 2002 and 31 December 2009: 1-day horizon with (a) $\alpha = 0.01$; (b) $\alpha = 0.05$. ......................... 220

5.9 Plots of mean VaR estimates of the given model groups against actual DJI portfolio returns between 24 January 2002 and 31 December 2009: 5-day horizon with (a) $\alpha = 0.01$; (b) $\alpha = 0.05$. ......................... 221

5.10 Plots of mean VaR estimates of the given model groups against actual DJI portfolio returns between 24 January 2002 and 31 December 2009: 10-day horizon with (a) $\alpha = 0.01$; (b) $\alpha = 0.05$. ......................... 222
5.11 Plots of mean VaR estimates of the given model groups against actual DJI portfolio returns between 24 January 2002 and 31 December 2009: 20-day horizon with (a) $\alpha = 0.01$; (b) $\alpha = 0.05$. 

\[223\]
References


243


247


