A Bayesian Analysis of a Regime Switching Volatility Model

A dissertation presented for the degree of Doctor of Philosophy

By
Glen Livingston Jr
B.Com GradDipCA B.Math (Hons I)

School of Mathematical & Physical Sciences
The University of Newcastle

November 2016
Declaration of Originality

“This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited un the Universities Digital Repository, subject to the provisions of the Copyright Act 1968.”

Glen Livingston Jr
Acknowledgements

Without the assistance and advice of my supervisor Dr Darfiana Nur, the research presented in this thesis would not have been possible. She allowed me the freedom to investigate the topics I felt were important and trusted my judgement in choosing them. While at Newcastle, her office door was always open, and even after her move to Flinders University, was always available for discussions either by email or Skype.

I would like to thank the staff of the School of Mathematical and Physical Sciences that assisted me throughout my study. In particular, I owe a huge thank you to Dr David Allingham for his assistance in running simulation studies in parallel over the Newcastle Research Computer Grid. Without his assistance, I would still be waiting for a simulation study to finish on my laptop. For his advice and general discussions, as well as introducing me to the Bayesian perspective, I must also thank Dr Frank Tuyl.

For assistance in the final editing of this thesis, I thank Dr Teresa Bates. Her valuable suggestions, although often subtle, resulted in a substantial improvement overall.

I would also like to thank the Statistics research students who I had the pleasure of sharing an office with throughout my study. Particularly, I would like to thank Dr Salman Cheema, Sidra Zafar, and Yi-Fan Lin. Our discussions on everything from statistics to cultural practices made my time at the university significantly more enjoyable. Learning and experiencing even just a small part of their culture enriched my time in the office.

While they have never understood exactly what I do, my Mum and Dad have always supported my decision to do it. It is their influence that has directed me to value critical thinking, and for that, I thank them.

Finally, without the support of my wife Jae, this thesis would not have been started or completed. Even after I turned the lounge room into a home office and often complained to her about problems with my research, she always supported me and offered assistance when she thought she could provide it. I am eternally grateful to her.
For Jae.
## Contents

Declaration of Originality

Acknowledgements

Abstract

Notation

1 Introduction

2 Literature Review

2.1 Models

2.1.1 STAR Models

2.1.2 GARCH Models

2.1.3 STAR-GARCH Models

2.1.4 Multivariate STAR Models

2.1.5 Multivariate GARCH Models

2.1.6 Multivariate STAR-GARCH Models

2.2 Algorithms for Posterior Analysis

2.2.1 Gibbs Sampler

2.2.2 Metropolis-Hastings

2.2.3 Algorithms to Improve Acceptance Rates

2.2.4 Reversible Jump

2.3 Model Selection

2.4 Effective Data for RJMCMC

3 Univariate Models

3.1 STAR Model

3.1.1 Prior Distribution

3.1.2 Likelihood Function

4 Multivariate Models

4.1 Models

4.1.1 STAR Models

4.1.2 GARCH Models

4.1.3 STAR-GARCH Models

4.1.4 Multivariate STAR Models

4.1.5 Multivariate GARCH Models

4.1.6 Multivariate STAR-GARCH Models

5 Posterior Analysis

5.1 Algorithms

5.1.1 Gibbs Sampler

5.1.2 Metropolis-Hastings

5.1.3 Algorithms to Improve Acceptance Rates

5.1.4 Reversible Jump

5.2 Model Selection

5.3 Effective Data for RJMCMC

6 Conclusion

References

Appendices

Index
3.1.3 Posterior Distribution ........................................ 41
3.1.4 Posterior Simulator ........................................ 42

3.2 Generalised Autoregressive Conditional Heteroskedasticity Models ............... 48
3.2.1 Prior Distribution ........................................ 50
3.2.2 Likelihood Function ....................................... 50
3.2.3 Posterior Distribution .................................... 51
3.2.4 Posterior Simulator ....................................... 51

3.3 STAR-GARCH Models .......................................... 56
3.3.1 Prior Distribution ........................................ 58
3.3.2 Likelihood Function ....................................... 59
3.3.3 Posterior Distribution .................................... 59
3.3.4 Posterior Simulator ....................................... 60

3.4 Conclusion ................................................... 70

4 Multivariate Models ............................................. 71

4.1 Vector Autoregressive ......................................... 71
4.1.1 Likelihood Function ....................................... 72
4.1.2 Prior Distribution ........................................ 73
4.1.3 Posterior Distribution .................................... 74
4.1.4 Posterior Simulator ....................................... 76

4.2 Multivariate STAR ............................................ 80
4.2.1 Likelihood Function ....................................... 82
4.2.2 Prior Distribution ........................................ 83
4.2.3 Posterior Distribution .................................... 84
4.2.4 Posterior Simulator ....................................... 87

4.3 Multivariate GARCH .......................................... 93
4.3.1 VECH Conditional Covariance Equation .................... 95
4.3.2 BEKK Conditional Covariance Equation .................... 97
4.3.3 Prior Distribution ........................................ 98
4.3.4 Likelihood Function ....................................... 99
4.3.5 Posterior Distribution .................................... 100
4.3.6 Posterior Simulator ....................................... 100

4.4 Multivariate STAR-GARCH ................................. 111
4.4.1 Prior Distribution ........................................ 113
4.4.2 Likelihood Function ....................................... 114
List of Figures

2.1 Perspective plot of transition function. ............................................. 9

5.1 Posterior probability distributions for $N = 150$ (top left) and $N = 225$ (top right), and the posterior probability distribution for the difference in proportions where $N = 225$ (bottom). ................................................................. 139

5.2 Time plot of simulated STAR data $x_t$ from S-I (top), and time plot of the simulated transition variable over the transition function $F_t$ (grey) with the horizontal line at the true value of the location parameter $c$ (bottom). ................................................................. 141

5.3 An example set of parameter posterior distributions from S-I. The red lines indicate the positions of the true values for each parameter. .................................................................................. 141

5.4 Time plot of simulated STAR data $x_t$ from S-II (top) and time plot of the simulated transition variable over the transition function $F_t$ (grey) with the horizontal line at the true value of the location parameter $c$ (bottom). ................................................................. 143

5.5 An example set of parameter posterior distributions from S-II. The red lines indicate the positions of the true values for each parameter. .................................................................................. 144

5.6 Monthly US unemployment rate from 1960 to 2004 time series plots: the original rate $U_t$ (top); the transformed rate $x_t$ (middle); and the seasonally differenced rate $s_t$ (bottom). 145

5.7 Posterior histogram for $\gamma$ (left) and $c$ (right) showing the density of the respective prior distributions (dashed line) for the US unemployment data. ................................................................. 147

5.8 US unemployment rate $x_t$ with fitted values $\hat{x}_t$ (red), together with the corresponding rescaled fitted transition function $\hat{F}_t$ (grey). ................................................................. 147

5.9 Time plot of simulated GARCH data $x_t$ from G-I. .................................................. 148

5.10 An example set of parameter posterior distributions from G-I. The red lines indicate the positions of the true values for each parameter. ................................................................. 148

5.11 Time plot of simulated GARCH data $x_t$ from G-II. .................................................. 149

5.12 An example set of parameter posterior distributions from G-II. The red lines indicate the positions of the true values for each parameter. ................................................................. 150

5.13 Distribution of point estimates for $\alpha_0$ from G-II. .................................................. 150

5.14 Time plot of the daily DEM/GBP foreign exchange log returns between 3 January 1985 to 31 December 1991 (top) and the conditional variance calculated on a rolling window (bottom). ................................................................. 151

5.15 Posterior distributions for $N_j$ and the coefficient parameter vector $\alpha$ for the first 750 observations. ................................................................. 152

5.16 Posterior distributions for $N_j$ and the coefficient parameter vector $\alpha$ for the whole 1,974 observations. ................................................................. 152
5.17 Fitted conditional variance for the DEM/GBP exchange log returns using the whole $N = 1,974$ observations. ............................................................... 153
5.18 Time plot of simulated STAR-GARCH data $x_t$ from SG-I. ................. 154
5.19 An example set of parameter posterior distributions from SG-I. The red lines indicate the positions of the true values for each parameter. ......................... 156
5.20 Time plot of simulated STAR-GARCH data $x_t$ from SG-II. .................. 156
5.21 An example set of parameter posterior distributions from SG-II. The red lines indicate the positions of the true values for each parameter. ......................... 157
5.22 Time plot of simulated STAR-GARCH data $x_t$ from SG-III. .............. 159
5.23 An example set of parameter posterior distributions from SG-III. The red lines indicate the positions of the true values for each parameter. ......................... 160
5.24 Time plot of the monthly SOI from January 1876 to December 2015 (top) and the conditional variance calculated on a rolling window (bottom). ................. 161
5.25 Transition function from pilot run on the conditional mean for SOI data. ........ 161
5.26 Posterior distributions for $\gamma$ (left) and $c$ (right) from full algorithm applied to the monthly SOI data, together with the prior densities (dashed lines). ................................. 162
5.27 Posterior distributions for the monthly SOI data. .................................. 163
5.28 Time plot of the monthly SOI data $x_t$ with fitted values $\hat{x}_t$ (red), together with the corresponding rescaled fitted transition function $\hat{F}_t$ (grey), (top); and fitted conditional variance $\hat{h}_t$ (bottom). ...................................................... 164
5.29 Time plot of simulated M-STAR data $x_t$ from MS-I. ....................... 166
5.30 An example set of parameter posterior distributions from MS-I. The red lines indicate the positions of the true values for each parameter. ......................... 166
5.31 An example set of parameter posterior distributions from MS-I (continued). The red lines indicate the positions of the true values for each parameter. ......................... 167
5.32 An example set of parameter posterior distributions from MS-I (continued). The red lines indicate the positions of the true values for each parameter. ......................... 168
5.33 Time plot of simulated M-STAR data $x_t$ from MS-II. ..................... 170
5.34 An example set of parameter posterior distributions from MS-II. The red lines indicate the positions of the true values for each parameter. ......................... 172
5.35 An example set of parameter posterior distributions from MS-II (continued). The red lines indicate the positions of the true values for each parameter. ......................... 173
5.36 An example set of parameter posterior distributions from MS-II (continued). The red lines indicate the positions of the true values for each parameter. ......................... 174
5.37 Time plots for the Jökulsá river flow (top); Vatnsdalsá river flow (middle); and the precipitation (bottom). ...................................................... 174
5.38 Time plots of the MCMC chains for $\gamma$ (top); $c$ (middle); and $d$ (bottom). ......................... 176
5.39 Posterior distributions for the parameters $\gamma$ and $c$ for the river flow data. The prior densities are shown as a dashed line. ................................. 176
5.40 Posterior distributions for the M-STAR model fitted to the Icelandic river flow data. ... 177
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.41</td>
<td>Time plots for the actual (black) and the fitted (red) values for Jökulsá (top) and Vatnsdalsá (bottom) river flow.</td>
</tr>
<tr>
<td>5.42</td>
<td>Time plot of the fitted transition function $F_t$ (left) and the fitted transition function $F_{t,r}$ plotted against the transition variable $s_t$ (right).</td>
</tr>
<tr>
<td>5.43</td>
<td>Time plot of simulated data $x_t$ from the bivariate VECH model in (5.6.11).</td>
</tr>
<tr>
<td>5.44</td>
<td>An example set of parameter posterior distributions for the VECH model. The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.45</td>
<td>A typical set of MCMC chains for the VECH algorithm showing a failure of the chains to converge.</td>
</tr>
<tr>
<td>5.46</td>
<td>Time plot of simulated M-GARCH data $x_t$ from MG-I.</td>
</tr>
<tr>
<td>5.47</td>
<td>An example set of parameter posterior distributions from MG-I. The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.48</td>
<td>Time plot of simulated M-GARCH data $x_t$ from MG-II.</td>
</tr>
<tr>
<td>5.49</td>
<td>An example set of parameter posterior distributions from MG-II. The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.50</td>
<td>Daily share prices for BHP Limited and Rio Tinto Limited.</td>
</tr>
<tr>
<td>5.51</td>
<td>Time plots of the log returns $x_{1,t}$ and $x_{2,t}$, for BHP Limited and Rio Tinto Limited respectively.</td>
</tr>
<tr>
<td>5.52</td>
<td>Sample variances of a rolling window of length 10, for BHP Limited and Rio Tinto Limited.</td>
</tr>
<tr>
<td>5.53</td>
<td>Posterior distributions for the M-GARCH model fitted to the BHP Limited and Rio Tinto Limited data.</td>
</tr>
<tr>
<td>5.54</td>
<td>Estimated conditional covariances (top and bottom left) and correlation (bottom right) over the time period, for the BHP Limited and Rio Tinto Limited data.</td>
</tr>
<tr>
<td>5.55</td>
<td>Time plot of simulated M-STAR-GARCH data $x_t$ from MSG-I.</td>
</tr>
<tr>
<td>5.56</td>
<td>An example set of parameter posterior distributions from MSG-I. The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.57</td>
<td>An example set of parameter posterior distributions from MSG-I (continued). The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.58</td>
<td>An example set of parameter posterior distributions from MSG-I (continued). The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.59</td>
<td>Time plot of simulated M-STAR-GARCH data $x_t$ from MSG-II.</td>
</tr>
<tr>
<td>5.60</td>
<td>An example set of parameter posterior distributions from MSG-II. The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.61</td>
<td>An example set of parameter posterior distributions from MSG-II (continued). The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.62</td>
<td>An example set of parameter posterior distributions from MSG-II (continued). The red lines indicate the positions of the true values for each parameter.</td>
</tr>
<tr>
<td>5.63</td>
<td>Time plots for the Jökulsá river flow (top); Vatnsdalsá river flow (middle); and the precipitation (bottom).</td>
</tr>
<tr>
<td>5.64</td>
<td>Sample variances of a rolling window of length 10 for the Jökulsá river flow (top) and Vatnsdalsá river flow (bottom).</td>
</tr>
</tbody>
</table>
5.65 Posterior distributions for the M-STAR-GARCH model fitted to the Icelandic river flow data. .......................................................... 205

5.66 Posterior distributions for the M-STAR-GARCH model fitted to the Icelandic river flow data (continued). .......................................................... 206

5.67 Time plots for the actual (black) and the fitted (red) values for Jökulsá (top) and Vatnsdalsá (bottom) river flow. .......................................................... 207

5.68 Time plot of the fitted transition function $F_t$ (left) and the fitted transition function $F_{t_i}$, plotted against the transition variable $s_t$ (right). .......................................................... 208

5.69 Posterior distributions for the unconstrained M-STAR-GARCH model fitted to the Icelandic river flow data. .......................................................... 210

5.70 Posterior distributions for the unconstrained M-STAR-GARCH model fitted to the Icelandic river flow data (continued). .......................................................... 211

5.71 Fitted transition function over time (left); transition function plotted against the transition variable (right), both for unconstrained case. .......................................................... 212

5.72 Time plots for actual (black) and fitted (red) values for Jökulsá (top) and Vatnsdalsá (bottom) river flow for the unconstrained case. .......................................................... 213

5.73 Fitted conditional covariance and correlation over the time period for the unconstrained case. .......................................................... 213
## List of Tables

3.1 Index of GARCH($l, m$) models. ...................................................... 49  
3.2 Index of GARCH($l, m$) models. ...................................................... 58  
4.1 Index of M-GARCH models. ............................................................... 98  
4.2 Index of M-GARCH models. ............................................................... 123  
5.1 Summary statistics for the simulated STAR model of S-I. ...................... 142  
5.2 Summary statistics for the simulated STAR model of S-II. ..................... 143  
5.3 Summary statistics for the US unemployment data. ............................. 146  
5.4 Summary statistics for the simulated GARCH model of G-I. .................. 149  
5.5 Summary statistics for the simulated GARCH model of G-II. ................. 149  
5.6 Summary statistics for the DEM/GBP exchange log returns for the first $N = 750$ observations. ................................................................. 151  
5.7 Summary statistics for the DEM/GBP exchange log returns for the whole $N = 1,974$ observations. ................................................................. 153  
5.8 Summary statistics for the simulated STAR-GARCH model of SG-I. ......... 155  
5.9 Summary statistics for the simulated STAR-GARCH model of SG-II. ....... 158  
5.10 Summary statistics for the simulated STAR-GARCH model of SG-III. ....... 159  
5.11 Summary statistics for the SOI data. ............................................... 162  
5.12 Summary statistics for the simulated M-STAR model of MS-I. ............. 169  
5.13 Summary statistics for the simulated M-STAR model of MS-II. ............. 170  
5.14 Summary statistics for the simulated M-STAR model of MS-II (continued). 171  
5.15 Parameter estimates for the M-STAR model fitted to the Icelandic river flow data. ................................................................. 178  
5.16 Index of M-GARCH models. ............................................................... 183  
5.17 Summary statistics for the simulated M-GARCH model of MG-I. .......... 185  
5.18 Summary statistics for the simulated M-GARCH model of MG-II. .......... 187  
5.19 Point estimates and credible intervals for the BHP Limited and Rio Tinto Limited data. 191  
5.20 Summary statistics for the simulated M-STAR-GARCH model of MSG-I. ... 196  
5.21 Summary statistics for the simulated M-STAR-GARCH model of MSG-I (continued). 197
5.22 Summary statistics for the simulated M-STAR-GARCH model of MSG-II. ............. 200
5.23 Summary statistics for the simulated M-STAR-GARCH model of MSG-II (continued). 201
5.24 Parameter estimates for the M-STAR-GARCH model fitted to the Icelandic river flow data. ................................................................. 205
5.25 Parameter estimates for the M-STAR-GARCH model fitted to the Icelandic river flow data with unconstrained conditional covariance. ....................... 209
5.26 Comparison of point estimates for the contained and unconstrained M-STAR-GARCH models. ................................................................. 209
Abstract

Non-linear time series data is often generated by complex systems. While linear models provide a good first approximation of a system, often a more sophisticated non-linear model is required to properly account for the features of such data. Correctly accounting for these features should lead to the fitting of a more appropriate model.

Determining the features exhibited by a particular data set is a difficult task, particularly for inexperienced modellers. Therefore, it is important to move towards a modelling paradigm where little to no user input is required, in order to open statistical modelling to users less experienced in MCMC. This sort of modelling process requires a general class of models that is able to account for the features found in most linear and non-linear data sets. One such class is the STAR-GARCH class of models. These are reasonably general models that permit regime changes in the conditional mean and allow for changes in the conditional covariance.

In this thesis, we develop original algorithms that combine the tasks of parameter estimation and model selection for univariate and multivariate STAR-GARCH models. The model order of the conditional mean and the model index of the conditional covariance equation are included as parameters for the model requiring estimation.

Combining the tasks of parameter estimation and model selection is facilitated through the Reversible Jump MCMC methodology. Other MCMC algorithms employed for the posterior distribution simulators are the Gibbs sampler, Metropolis-Hastings, Multiple-Try Metropolis and Delayed Rejection Metropolis-Hastings algorithms. The posterior simulation algorithms are successfully implemented in the statistical software program R, and their performance is tested in both extensive simulation studies and practical applications to real world data.

The current literature on multivariate extensions of STAR, GARCH, and STAR-GARCH models is quite limited from a Bayesian perspective. The implementation of a set of estimation algorithms that not only provide parameter estimates but is also able to automatically fit the model with highest posterior probability is a significant and original contribution. The impact of such a contribution will hopefully be a step forward on the path towards the automation of time series modelling.
Below are some details of the notation and functions used throughout this thesis unless stated otherwise.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_t$</td>
<td>Univariate data point.</td>
</tr>
<tr>
<td>$x$</td>
<td>Complete univariate data set.</td>
</tr>
<tr>
<td>$x_t$</td>
<td>Multivariate data point.</td>
</tr>
<tr>
<td>$X$</td>
<td>Complete multivariate data set.</td>
</tr>
<tr>
<td>$C, \Sigma, H$</td>
<td>Upper case symbol indicates a matrix.</td>
</tr>
<tr>
<td>$H, \Phi$</td>
<td>Bold and upper case indicates a matrix made up of either a vector of vectors, or a matrix of matrices.</td>
</tr>
<tr>
<td>$I_n$</td>
<td>$n \times n$ identity matrix.</td>
</tr>
<tr>
<td>$\Omega^T$</td>
<td>Transpose of the matrix $\Omega$.</td>
</tr>
<tr>
<td>$\text{tr}(\Omega)$</td>
<td>Trace of the square matrix $\Omega$.</td>
</tr>
<tr>
<td>$p(\Theta</td>
<td>\Omega)$</td>
</tr>
<tr>
<td>$\text{vec}(\Omega)$</td>
<td>Stacking the columns of the matrix $\Omega$ into a column vector.</td>
</tr>
<tr>
<td>$\text{vech}(\Omega)$</td>
<td>Stacking the lower triangle of the matrix $\Omega$ into a column vector.</td>
</tr>
<tr>
<td>$\exp[\omega]$</td>
<td>$e^{\omega}$.</td>
</tr>
<tr>
<td>$A \odot B$</td>
<td>The component-wise multiplication of the matrices $A$ and $B$.</td>
</tr>
<tr>
<td>$A \otimes B$</td>
<td>The Kronecker product of the matrices $A$ and $B$.</td>
</tr>
<tr>
<td>$\Gamma_p(n)$</td>
<td>The multivariate gamma function.</td>
</tr>
<tr>
<td>$\mathcal{N}(\mu, \sigma^2)$</td>
<td>Univariate normal distribution with mean $\mu$ and variance $\sigma^2$. The density of the distribution is given by $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$.</td>
</tr>
<tr>
<td>$\mathcal{N}_p(\mu, \Sigma)$</td>
<td>Multivariate normal distribution with mean $\mu$ and covariance matrix $\Sigma$ for $x$, a column vector of length $p$. The density of the distribution is given by $f(x) = (2\pi)^{-\frac{p}{2}}</td>
</tr>
<tr>
<td>$\mathcal{N}_{n,p}(M,U,V)$</td>
<td>Matrix normal distribution with $n \times p$ location matrix $M$, $n \times n$ scale matrix $U$, and $p \times p$ scale matrix $V$ for $X$, an $n \times p$ matrix. The density of the distribution is given by $f(X) = \frac{\exp\left[-\frac{1}{2} \text{tr} \left(V^{-1} (X-M)^T U^{-1} (X-M) \right)\right]}{(2\pi)^{-\frac{np}{2}}</td>
</tr>
<tr>
<td>$\mathcal{G}(\alpha, \beta)$</td>
<td>Gamma distribution with shape and rate parameters $\alpha$ and $\beta$, respectively. The density of the distribution is given by $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$.</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>$\mathcal{IG} (\alpha, \beta)$</td>
<td>Inverse Gamma distribution with shape and scale parameters $\alpha$ and $\beta$, respectively. The density of the distribution is given by $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$.</td>
</tr>
<tr>
<td>$\mathcal{W} (V, n)$</td>
<td>Wishart distribution with scale matrix $V$ and degrees of freedom $n$ for $X$, a $p \times p$ matrix. The density of the distribution is given by $f(X) = \frac{</td>
</tr>
<tr>
<td>$\mathcal{IW} (\Psi, \nu)$</td>
<td>Inverse Wishart distribution with scale matrix $\Psi$ and degrees of freedom $\nu$ for $X$, a $p \times p$ matrix. The density of the distribution is given by $f(X) = \frac{\Psi^\frac{\nu}{2}}{2^{\frac{\nu p}{2}} \Gamma_p \left(\frac{\nu}{2}\right)}</td>
</tr>
<tr>
<td>$\mathcal{T}_p (\mu, \Sigma)$</td>
<td>Multivariate $t$-distribution with location vector $\mu$, degrees of freedom $\nu$, and scale matrix $\Sigma$ for $x$, a column vector of length $p$. The density of the distribution is given by $f(x) = \frac{\Gamma \left(\frac{\nu + p}{2}\right)}{\Gamma \left(\frac{\nu}{2}\right) \pi^\frac{p}{2}</td>
</tr>
<tr>
<td>$\mathcal{U} (a, b)$</td>
<td>Uniform distribution over the range $a$ to $b$. The density of the distribution is given by $f(x) = \begin{cases} \frac{1}{b-a} &amp; \text{for } x \in [a, b] \ 0 &amp; \text{otherwise} \end{cases}$.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

In many instances, little to no theory surrounds the processes through which time related data is generated. By constructing time series models that capture the important features of such data, we may be able to provide information about its generation processes or predict its future values. Linear time series models, most commonly the Autoregressive Integrated Moving Average (ARIMA) model, have long been the first option applied to the analysis of time series data. The ARIMA model is an extension of the autoregressive (AR) model first implemented by Yule (1927) to study Wolfer’s sunspot numbers.

The popularity of ARIMA models is due, in part, to their relative simplicity. As a result, many statistical software packages are able to fit ARIMA models to a particular data set with little user input. As the AR and ARIMA models have been studied and applied for a considerable length of time, there is an extensive amount of research available on them. They have been applied in many research fields. As Tong (1990) observed, their prolonged survival indicates that they provide a good first approximation to the objective world.

Most strategies for building time series models have their roots in the methodology developed in the seminal work of Box and Jenkins (1976) (Box-Jenkins). The Box-Jenkins modelling methodology begins with the identification of potential model orders, often through the interpretation of graphical displays such as plots of the Autocorrelation function (ACF) and the Partial Autocorrelation Function (PACF). Once the potential models have been identified, the parameters for each of the candidate models are estimated. The candidate models are then checked using a suite of diagnostic tests and those that fail are disregarded. If all of the candidate models fail the diagnostic tests, the process begins again, using a more complicated class of models. This procedure is repeated until one or more models are found to be suitable.

“Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house.”

- Henri Poincaré
When more than one model is found to be adequate, the modeller must discriminate between these candidate models and choose the one that they think is best. There are several methods for doing this, such as choosing the model that has the best in-sample forecast performance. However, the most common technique is to employ a model scoring method such as the Akaike Information Criterion (AIC) (Akaike 1974) or Bayesian Information Criterion (BIC) (Schwarz 1978).

The Box-Jenkins time series modelling methodology can be thought of as a bottom up approach to time series modelling. That is, the modelling cycle starts with the simpler models towards the bottom of the modelling ladder, and only moves up the ladder to more complicated models when the models initially proposed are found to be inadequate.

One problem with this sort of methodology arises from the use of inspection of the ACF and PACF plots as the basis for the selection of the candidate models. These plots often display mixed information for data that arises from complicated processes (Kendall 1971). Consequently, the modeller may select too many models for investigation and waste resources on the estimation of parameters for models that will ultimately be disregarded (Chatfield and Prothero 1973). On the other hand, it is possible that the user may not include the best available model in the list of candidate models. This will ultimately lead to a failure to select the best available model.

It may be possible to fit an adequate ARIMA model to the data, even though a more complicated model is better suited to describe the characteristics of the data set, regardless of penalties imposed for additional complexity. In this situation, a user may not proceed to investigate the class of models with greater complexity and, instead, be satisfied with the simple ARIMA model. This is another potential problem that may be encountered when employing a bottom up modelling approach: models that would not be considered “optimal” may be selected. Examples in which linear models are fitted to a data set for which non-linear models would be more appropriate are provided in Tong (1990).

As previously mentioned, one appealing aspect of ARIMA models is their relative simplicity. This has the consequence that, at times, they will be fundamentally inadequate for describing the characteristics of a data set. When the underlying mechanisms generating data are significantly non-linear in nature, a more sophisticated non-linear time series model is required to capture the characteristics of the data.

According to Tong (1990), non-linear data is usually characterised by non-periodic highs and lows, and unusually high sudden bursts. A data set may possess either, or both of these features. To capture the feature of non-periodic highs and lows, regime changes in the conditional mean of a model may be introduced. Unusually high sudden bursts may be accounted for by allowing the conditional variance of the model to change over time. Data with these non-linear features arise in many fields of research and the ability to model such data effectively is an important objective. The literature contains details of several types of non-linear models that have been proposed to account for the above mentioned
characteristics of non-linear data.

Some models proposed to account for non-periodic highs and lows in data are the Markov Switching Autoregressive (MSAR) model (Hamilton 1989), the Threshold Autoregressive (TAR) model (Tong 1978), and the Smooth Transition Autoregressive (STAR) model (Chan and Tong 1986). In these models, regime changes give rise to a different equation for the conditional mean of the model.

The model that is most commonly suggested in the literature to capture unusually high sudden bursts in a data set is the Generalised Autoregressive Conditional Heteroskedastic (GARCH) model (Bollerslev 1986). This model is an extension of the innovative Autoregressive Conditional Heteroskedastic (ARCH) model introduced by Engle (1982). A simple ARIMA model assumes constant variance of the error over time, whereas GARCH-type models allow the conditional variance to change over time, thereby accommodating sudden bursts in volatility.

In situations where data exhibits both of the aforementioned characteristics, a combination of models may be required. For example, if a data set contains non-periodic highs and lows together with high or low sudden bursts, a user may choose to include a STAR-GARCH model that accounts for the respective features of the data as a possibility under their model selection strategy.

Two important aspects of any model building paradigm were noted by Tong (1990). The first is that a fairly wide class of models should be selected to ensure the ability of the final model to capture the important features of the observed data. The second is that the size of this class of models is ultimately constrained by the computational power available for the search. Tong points out that “recognition of these two aspects reinforces the belief that model building is as much an art as it is a science.”

Computational power has increased exponentially since 1990 and this is expected to continue. As a consequence, the restriction on the size of the class of models that can be considered has been greatly reduced. Consequently, future model building strategies are likely to move away from the somewhat romantic notion of an art form, towards the realm of science.

STAR-GARCH models form a relatively large class. By careful choice of the transition function incorporated in the conditional mean, the STAR-GARCH model may be seen to generalise several simpler models. For example, when the logistic function is used as the transition function, the conditional mean of the STAR-GARCH model generalises that of the logistic STAR, TAR and AR models (Luukkonen, Saikkonen, and Teräsvirta 1988). In general, the same conditional variance equation may be used to represent a GARCH or ARCH model, as well as a model with constant conditional variance (Bollerslev 1986).

Under the multivariate setting, the multivariate STAR-GARCH model can be seen to generalise several types of smaller models. Investigations into a full multivariate STAR-GARCH model are quite limited.
Most studies employing a multivariate GARCH model for the conditional covariance use a simpler vector autoregressive process for the conditional mean.

A possible reason for the limited amount of research into multivariate STAR-GARCH models could be the difficulty in estimating the parameters of these models. Even under a univariate setting, parameter estimation for these models is relatively complicated as there are several parameters of differing domains which must be estimated jointly. The smoothing parameter within the transition function can be notoriously difficult to estimate when it is relatively large (Dijk, Teräsvirta, and Franses 2002). When it is large, small perturbations in the smoothing parameter result in transition functions of very similar shapes. To obtain an accurate estimate, many observations close to the threshold variable are required.

Under a multivariate setting, the number of parameters requiring estimation is significantly increased. Assuming that both model orders are equal to one, the number of parameters in the bivariate GARCH model under the VECH formulation (Bollerslev, Engle, and Wooldridge 1988) is 21. When three series are under analysis, the number of parameters that must be estimated is increased to 78. This is a considerable increase over the number of parameters for two or three separate univariate GARCH models, and is an example of the curse of dimensionality. As a result of the increased number of parameters, the amount of data required to perform meaningful estimations increases, which, in turn, makes estimations by numerical methods more time consuming and difficult.

Several authors (Chan and McAleer 2002, 2003; Chan and Theoharakis 2011) have estimated STAR-GARCH models from a frequentist perspective. This is facilitated by the use of Quasi Maximum Likelihood Estimation (QMLE). However, Lundbergh and Teräsvirta (1998) and Dijk, Teräsvirta, and Franses (2002) noted that the convergence of the QMLE is sensitive to the initial values used in the search. When deterministic optimisation algorithms are employed, there is a risk the algorithm may locate a local maximum, and never find the global maximum.

The idea that the initial values are of critical importance under the frequentist setting is reinforced by Teräsvirta and Yang (2014b). They perform parameter estimation for the multivariate STAR model using non-linear least squares estimation techniques. They note that the likelihood function can be fairly flat in several directions and also contain several local optima. As a result, they search for suitable starting values over a grid of values.

From the Bayesian perspective, the risk of being trapped at a local maximum is reduced as an MCMC estimation procedure is stochastic in nature. That said, the initial values of the algorithm can play a significant role in the speed of convergence of an MCMC algorithm, as shown in Livingston Jr and Nur (2016a). Bayesian estimation has been applied to models similar to the STAR-GARCH model (see, for example, Gerlach and Chen (2008), Chen and So (2006), and Nieto, Zhang, and Li (2013)).
Another advantage of the Bayesian framework is the availability of the Reversible Jump MCMC (RJMCMC) algorithm (Green 1995). The use of RJMCMC algorithms allows the inclusion of the model orders for the conditional mean and conditional variance equations as parameters in the posterior distribution that require estimation. Therefore, a single estimation scheme allows the simultaneous execution of both parameter estimation and model selection.

By employing the STAR-GARCH class of models, we are able to propose a top down modelling procedure on a wide class of models that generalises several other models. A single estimation scheme employing the RJMCMC algorithm will combine parameter estimation and model selection into a single task. This means that all subsets of the STAR-GARCH model are considered in one scheme and resources are not wasted on models that will subsequently be discarded. In addition, as the STAR-GARCH model generalises the AR model, there is no need to undertake linearity testing on the data. The posterior distribution will determine how linear the data is.

The ultimate objective and original contribution of this thesis is the development of a set of algorithms that combine the task of parameter estimation and model selection for univariate and multivariate STAR-GARCH models. The model order of the conditional mean, and the model index of the conditional covariance equation will be included as parameters for the model requiring estimation. The aims of this research are as follows:

(1) To present a Bayesian analysis of STAR, GARCH, and STAR-GARCH models under a univariate setting.

(2) To present a Bayesian analysis of STAR, GARCH, and STAR-GARCH models under a multivariate setting.

(3) For each of the models mentioned in (1) and (2), to propose an automatic and efficient MCMC algorithm to estimate the parameters of the models using a combination of the Gibbs sampler, Metropolis-Hastings, Multiple-Try Metropolis, Delayed Rejection Metropolis-Hastings and RJMCMC algorithms.

(4) To illustrate the performance of the estimation scheme against simulated data as well as real world data.

In summary, the main contribution of this thesis is the development of algorithms combining the task of model selection and parameter estimation for the complicated class of non-linear STAR-GARCH models. The current literature on multivariate STAR models is limited and for multivariate STAR-GARCH models the literature is non-existent, from either a Bayesian or frequentist perspective. Therefore the implementation of an estimation algorithm that not only provides parameter estimates but also fits the model with highest posterior probability automatically is a significant and original contribution. The
impact of such a contribution will hopefully be a step forward on the path towards the automation of time series modelling.

The remainder of this thesis is organised as follows. Chapter 2 presents a literature review of the models considered, as well as a discussion of the well known MCMC algorithms that will be employed throughout this thesis. Chapters 3 and 4 present the posterior distributions, together with the design and structure of the algorithms for univariate and multivariate models, respectively. Through several simulation studies, the estimation schemes developed in chapters 3 and 4 are tested in Chapter 5 and applications to real world data are provided. The applications to real world data show that subsets of the model are able to identified. Chapter 6 presents a conclusion and some suggestions for future work.
Chapter 2

Literature Review

This chapter provides a review of the literature surrounding univariate time series models and a number of their multivariate extensions that are utilised in this thesis. In addition, we discuss some of the Markov chain Monte Carlo algorithms that are employed for the purposes of parameter estimation in the reviewed models.

2.1 Models

2.1.1 STAR Models

A popular model that is used to explain apparent regime switching behaviour in the conditional mean is the Smooth Transition Autoregressive (STAR) model. The STAR model evolved from the Threshold Autoregressive (TAR) model, introduced by Tong (1978), Tong and Lim (1980) and Tong (1983). The TAR model is effectively a piecewise linear model composed of two or more regimes of linear sub-models. The switch from one regime to another is implemented by an indicator function that takes the values zero or one, depending upon the values of a transition variable and a threshold parameter. If the transition variable is a lagged endogenous variable, the resulting model is a special case of the TAR model, called the Self Exiting Threshold Autoregressive (SETAR) model (Tong and Lim 1980).

The TAR and SETAR models have been widely applied in several fields, including, but not limited to, ecology, hydrology, finance and economics. In finance and economics, several indicators and indices such as exchange rates and the GDP have been modelled using the TAR model. The United States GNP was investigated by Potter (1995) who found that the TAR model outperformed a simple linear model. The forecast performance of the SETAR model was compared against a linear ARIMA model.
for Canadian GDP data from 1965 to 2000 by Feng and Liu (2003). It was concluded that the SETAR model performed better than the simpler ARIMA model. Boero and Marrocu (2002) applied the SETAR and other non-linear models to the US dollar, French franc, German mark and the Japanese yen in a comparison of their forecast performances against simple linear AR models. Their results confirmed a comparative advantage of non-linear models over linear models in forecasting under certain circumstances.

An application of the TAR model in epidemiology was presented by Watier and Richardson (1995). They applied a SETAR model to the monthly reported cases of salmonella in France from January 1978 to December 1988. A two regime SETAR model with an autoregressive order of twelve and a delay parameter of seven was best suited to modelling the data. Another example outside the fields of economics and finance is the application of a TAR model by Enders and Sandler (2002). They analysed a time series data set related to the number of deaths occurring in international terrorist attacks, and found that the regime switching model offered a better fit than an autoregressive time series model. In hydrology, Nieto, Zhang, and Li (2013) illustrated their estimation procedure for multiple regime TAR models using daily river flow as the observed data set and daily rainfall as the threshold variable.

The indicator function in the TAR model creates sudden jumps from one regime to the next. An immediate jump from one regime to another may not be the best representation of the underlying mechanism driving observed data. In a paper discussing the estimation of the transition between two intersecting lines, Bacon and Watts (1971) proposed a model that allows a smooth transition from one linear regime to the other. The idea of a smooth transition was adopted by Chan and Tong (1986), who suggested the replacement of the indicator function with a smooth transition function. In doing so, they created a new class of non-linear models, called the Smooth Transition Autoregressive (STAR) models.

A STAR model of order $k$, used to model a set of data, $\{x_1, \ldots, x_N\}$, can be expressed as follows:

$$x_t = \phi_{1.0} + \phi_{1.1}x_{t-1} + \cdots + \phi_{1.k}x_{t-k} + \phi_{2.0} + \phi_{2.1}x_{t-1} + \cdots + \phi_{2.k}x_{t-k} \cdot F_t + \varepsilon_t,$$

(2.1.1)

where $\varepsilon_t$ is the error term with $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ and $F_t$ is the transition function. The $\phi_{i,j} \in \mathbb{R}$ are the autoregressive parameters. The only condition on the transition function is that it is a smooth function that takes values in the range $[0, 1]$. A popular choice is the logistic function (Teräsvirta 1994), resulting in the Logistic STAR model. The logistic function commonly used is

$$F_t(s_t, \gamma, c) = \frac{1}{1 + \exp\left[-\gamma(s_t - d - c)\right]}.$$  

(2.1.2)
The parameters of the logistic function in (2.1.2) are the transition variable $s_{t-d} \in \mathbb{R}$, which is often the lagged endogenous variable $x_{t-d}$, with delay parameter $d \in \mathbb{N}$; the smoothing parameter, or shape parameter $\gamma \in \mathbb{R}^+$ and the location parameter $c \in \mathbb{R}$.

It is easy to see from (2.1.2) that changing the size of $\gamma$ causes the shape of $F_t$ to change. As $\gamma$ becomes larger, the transition function sharpens and approaches the step function. For small $\gamma$, $F_t$ flattens out, and when $\gamma = 0$, the transition function takes the value of one half for all $s_t$ and $c$. For some fixed $\gamma$, the values of the transition variable much larger or much smaller than the location parameter result in the transition function taking values close to 1 and 0, respectively. These features of the logistic function were observed by Teräsvirta (1994).

A visual representation for the transition function with $s_t$ and $\gamma$ both ranging from 0 to 10, and the location parameter fixed at 5 is shown in Figure 2.1.

![Perspective plot of transition function.](image)

Figure 2.1: Perspective plot of transition function.

The form of the logistic function used throughout this thesis is one that removes the scale of the transition variable $s_t$. This transition function was used in Gerlach and Chen (2008) and, as they observe, allows comparison of the smoothing parameters between models. Another advantage of this sort of transition function is that, since the scale of the transition variable is removed, the range of realistic values that can be taken by the smoothing parameter is much smaller. This will then be reflected in a prior distribution that limits its search to a smaller range of values. The form of this
transition function is as follows:

$$F_t(s_t, d, \gamma, c) = \frac{1}{1 + \exp\left[-\frac{\gamma}{S(s_t)}(s_t - d - c)\right]}.$$  \hspace{1cm} (2.1.3)

The variables of the logistic function in (2.1.3) are the same as those specified for (2.1.2), with the addition of $S(s_t)$, which denotes the standard deviation of the transition variable $s_t$.

A smooth transition between regimes may be of use when the transition from one regime to another arises as a result of several actions that take place over time. For example, the participants of markets make their product purchasing decisions on the basis of when they receive information, with each participant receiving information at different times. The culmination of a market participant’s actions may create a smooth transition from one regime to the next rather than a sudden, instantaneous one.

Like the TAR model, the STAR model is applicable in many fields. The Southern Oscillation Index, a commonly used measure of El Niño events, was modelled with a logistic STAR model by Hall, Skalin, and Teräsvirta (2001). As their model employed a lagged endogenous transition variable, it was unable to correctly predict the onset of an El Niño as the triggers for such an event are exogenous to the system. Their LSTAR model was found to be better suited to the description of turbulent periods in the data set than an autoregressive model. Once an El Niño event was under way, the model could be used for predictions of climate data a few months into the future.

An examination of real exchange rate data from the five major economies in the Association of the South-East Asian Nations was undertaken by Liew (2004). Previous authors had suggested using an exponential function for the transition function as the ESTAR model is better suited for investigating symmetrical positive and negative deviations from the equilibrium level. Allowing the data to “speak for itself”, Liew (2004) found that four of the five exchange rates were best modelled by the LSTAR model, a finding which confirmed the results of Liew, Baharumshah, and Lau (2002).

A similar assessment of the comparative suitabilities of the LSTAR model and the ESTAR model was performed by Giovanis (2009) on the Greece-USA spot exchange rate. ARCH effects were found in the time series and, consequently, the LSTAR model outperformed the ESTAR model for in sample and out of sample forecasting. Skalin and Teräsvirta (2002) applied the LSTAR model to unemployment data for OECD countries and found that the model was suitable for several of the examined series. They also found that the crucial parameters in the model had similar estimates across the series.

When investigating STAR modelling methodology, Teräsvirta (1994) proposed a Box-Jenkins style modelling cycle for logistic and exponential STAR models. Their cycle comprised three steps: model-type specification in which linearity was tested against STAR non-linearity, parameter estimation in a
frequentist setting using non-linear least squares, and the application of evaluation techniques to fitted
STAR models. They tested their techniques against simulated data, and then applied them to the
Canadian Lynx data set and the data obtained by taking the logarithms of the West German quarterly
production volumes from 1961 to 1986.

A comprehensive summary of the developments on STAR models was provided by Dijk, Teräsvirta,
and Franses (2002). They proposed new methods for testing non-linearity in addition to methods for
model evaluation and forecasting in the frequentist setting. STAR models were compared to other
non-linear time series models such as the SETAR model and the Markov Switching Autoregressive
(MSAR) model. They demonstrated their modelling cycle for the STAR model on data arising from
the United States unemployment rate.

The parameter estimation in Dijk, Teräsvirta, and Franses (2002) was performed using either Maximum
Likelihood Estimation (MLE), where \( \varepsilon_t \) is assumed to be normally distributed, or non-linear least
squares estimation, which can be interpreted as Quasi MLE (QMLE). It utilised the fact that, for fixed
values of \( \gamma \) and \( c \), the model is linear in the autoregressive parameters. A two step estimation procedure
was employed, in which the starting points for \( \gamma \) and \( c \) were found by searching over a grid of values.
Their estimation procedure did not incorporate model selection as, under a frequentist setting, it is a
difficult task to include the model order as a parameter for estimation. Under this paradigm, model
selection is usually performed as part of a Box-Jenkins style modelling cycle in which a model score
system such as AIC (Akaike 1974) or BIC (Schwarz 1978) is used to prioritise competing models.

Maringer and Meyer (2008) performed parameter estimation for the STAR model using a range of
optimisation heuristics. The methods employed included simulated annealing, threshold accepting and
differential evolution. They found that their methodology was relatively efficient in terms of CPU time
and outperformed deterministic parameter estimation algorithms, which are susceptible to becoming
trapped at local maxima.

From the Bayesian perspective, Péguin-Feissolle (1994) presented an estimation and forecasting pro-
cedure for a general non-linear model and then illustrated it using an LSTAR model with simulated
data. A comparison of non-linear models with regime changes in the conditional mean, namely Markov
Switching Autoregressive (MSAR) models, TAR and STAR models, was presented by Potter (1999).
The review included estimation techniques from a classical and Bayesian perspective as well as a review
of some of the parametric tests for non-linearity.

Another comparison of the STAR model to the MSAR model was performed by Deschamps (2008).
They estimated both STAR and MSAR model parameters from a Bayesian perspective, and calculated
the maximum likelihood estimates for comparison. The two models were also applied to a transforma-
tion of the United States unemployment rate data set used by Dijk, Teräsvirta, and Franses (2002).
The estimation methodology applied by Deschamps (2008) was similar to that implemented by Lopes and Salazar (2006). The key differences between the two methodologies lie in the techniques used for model selection and the estimation of implicit parameters for the transition function. Deschamps (2008) used marginal likelihoods for model selection, whereas Lopes and Salazar (2006) included the model order as a parameter for estimation using an RJMCMC (Green 1995) algorithm. To estimate the smoothing and threshold parameters for the transition function, Deschamps (2008) used an independence chain Metropolis-Hastings algorithm, whereas Lopes and Salazar (2006) employed a random walk Metropolis-Hastings algorithm. Deschamps (2008) argued that the independence chain Metropolis-Hastings algorithm was preferable to the random walk Metropolis-Hastings algorithm as it was easier to tune the acceptance rate of the Metropolis-Hastings step.

The Bayesian methodology outlined in Lopes and Salazar (2006) is an extension to STAR models of the RJMCMC procedure originally developed for autoregressive models by Troughton and Godsill (1997). Barbieri and O’Hagan (1996) and Vermaak et al. (2004) implemented RJMCMC algorithms for linear time series models in which the AR coefficients were parametrised using reflection coefficients, also known as partial correlation coefficients. This was done to facilitate the enforcement of stationarity into the algorithm. Vermaak et al. (2004) also provided an algorithm that used regular AR coefficients for situations in which there is no need to enforce stationarity. They assessed their algorithm for model selection against simulated AR(3) data and compared their results with other standard model selection criteria such as the Akaike Information Criterion (AIC) (Akaike 1974), the Bayesian Information Criterion (BIC) Schwarz (1978) and the Minimum Description Length (MDL) (Rissanen 1978). The Bayesian RJMCMC method outperformed the other methods. The paper also applied the model selection algorithm to the Southern Oscillation Index data set.

Following Troughton and Godsill (1997), Lopes and Salazar (2006) and Godsill (2001), Livingston Jr and Nur (2016a) performed a prior sensitivity analysis with a focus on the prior distributions for the smoothing and location parameters in the transition function. A Multiple-Try Metropolis algorithm was employed to estimate the smoothing parameter in an attempt to improve the acceptance rate and convergence of the algorithm.

Before reviewing the literature for STAR-GARCH models, we explore GARCH models.

### 2.1.2 GARCH Models

Traditional time series modelling assumed a constant conditional variance. The need to allow the conditional variance to change over time was the motivation for the introduction of the Autoregressive Conditional Heteroscedastic (ARCH) model by Engle (1982). This revolutionary new class of models could be applied to processes with non-constant variances, conditional on the past, but required con-
stant unconditional variances. This model was then generalised by Bollerslev (1986), who introduced past conditional variances into the model’s conditional variance equation. The resulting model is known as the Generalised Autoregressive Conditional Heteroscedastic (GARCH) model and has been widely applied in several fields, though most commonly in economics and finance.

GARCH models have been applied to Value at Risk estimation (Angelidis, Benos, and Degiannakis 2004; So and Yu 2006), as well as volatility estimation for financial markets and indices (Awartani and Corradi 2005; Gokcan 2000; Chong, Ahmad, and Abdullah 1999). Speech signals were modelled by Cohen (2004) using GARCH models. Radchenko (2005) used a GARCH model to analyse the effect of volatility in oil prices on the degree of asymmetry in the response of gasoline prices.

The GARCH model with zero mean is expressed mathematically as follows:

\[ x_t = \varepsilon_t \]
\[ \varepsilon_t = \sqrt{h_t} \eta_t \]
\[ h_t = \alpha_0 + \sum_{i=1}^{l} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{m} \beta_j h_{t-j} , \]

where \( l \) and \( m \) are the orders of the ARCH coefficients \( \alpha_i \) and GARCH coefficients \( \beta_j \), respectively; \( h_t \) is the conditional variance equation and \( \eta_t \sim N(0,1) \). In order to ensure that \( h_t > 0 \), it is required that \( \alpha_0 > 0, \alpha_i \geq 0 \) for \( i \in \{1, \ldots, l\} \) and \( \beta_j \geq 0 \) for \( j \in \{1, \ldots, m\} \). A final condition which ensures that the model is stationary is that \( \left( \sum_{i=1}^{l} \alpha_i + \sum_{j=1}^{m} \beta_j \right) < 1 \) (Bollerslev 1986).

The stationary conditions and autocorrelation structure of the GARCH model were derived by Bollerslev (1986) as a part of the model’s introduction. After estimating the parameters for the model using MLE, the author applied the model to the GNP deflater rate in the US. In addition, a simple GARCH(1,1) model was fitted to the conditional variance equation and its results were contrasted with those obtained using the more complicated ARCH(8) model proposed by Engle and Kraft (1983).

The basic structure of the GARCH model has been extended and modified to create several other models that are able to account for potential asymmetry in the conditional variance equation. Under the GARCH formulation, the conditional variance equation reacts symmetrically to past shocks, whether they are positive or negative. The EGARCH (Nelson 1991), GJR-GARCH (Glosten, Jagannathan, and Runkle 1993), TGARCH (Zakoian 1994) and STGARCH (Lubrano 2001) models are all examples of models within the GARCH family that seek to account for asymmetry in the conditional variance equation.

Estimation of the GARCH model parameters has also been performed in a Bayesian setting. Bauwens and Lubrano (1998) implemented a Griddy-Gibbs sampler (Ritter and Tanner 1992) for inference on the parameters of GARCH models with Student errors. They showed that their method is feasible and
competitive for importance sampling and the Metropolis-Hastings algorithm. The paper discusses the application of their algorithm to simulated data and the Brussels Stock Market index.

A review and comparison of several parameter estimation MCMC algorithms is provided by Asai (2006). Given that all methods reviewed converged to the same posterior distribution, the different methods were assessed with respect to criteria such as mixing, efficiency and the computational requirements of the algorithm. A custom method, based on an acceptance-rejection Metropolis-Hastings algorithm, performed the best out of all of the estimation methods they investigated.

Vrontos, Dellaportas, and Politis (2000) performed Bayesian inference for GARCH type models employing the RJMCMC methodology of Green (1995). Their algorithm allowed jumps between a GARCH(1,1) and an EGARCH(1,1) model. Two different approaches were outlined. They tested their methodology against simulated GARCH(1,1) and EGARCH(1,1) data as well as real life data in the form of the Athens Stock Exchange index. The RJMCMC methodology implemented allowed for subjective model selection as well as subsequent model averaging using the posterior probabilities as weights for the respective models.

An RJMCMC scheme testing for threshold non-linearity in GARCH models was proposed by So, Chen, and Chen (2005). This scheme allowed jumps between a regular GARCH model and a threshold GARCH model, using the posterior probabilities as evidence for asymmetry in the conditional variance. An extensive simulation study was provided in addition to an application of the model to 10 market indices.

The GARCH model and its extensions have been heavily applied in finance and economics, as well as other fields of research. While there is an extensive array of literature for GARCH models and their extensions, the literature surrounding the STAR-GARCH model is a little more limited.

### 2.1.3 STAR-GARCH Models

While the GARCH model is appropriate for modelling data sets that exhibit unusually high sudden bursts, data sets that also display regime changes in the conditional mean require models that can effectively accommodate this behaviour. The specification of an appropriate model for the conditional mean leads to more appropriate estimates of the conditional variance parameters (Lundbergh and Teräsvirta 1998). As STAR-GARCH models are a comparatively young class of models, little research has been performed on them to date. We now review the literature on STAR-GARCH and other related models.

A TAR-ARCH model was proposed by Li and Lam (1995) and applied to the Hang Seng index. Parameter estimation was performed in a frequentist setting with model selection through AIC. Following on
Section 2.1: Models

from this, Li and Li (1996) introduced a double threshold GARCH model. The parameter estimation for this model was performed using MLE and the model was also applied to the Hang Seng index.

From a Bayesian perspective, Chen and So (2006) introduced a TAR-GARCH model that allowed exogenous variables in the conditional mean equation. Another contribution for TAR-GARCH models was the application of the RJMCMC algorithm by Nieto, Zhang, and Li (2013), who extended the work of Campbell (2004). It included the number of regimes in the TAR model and the model order for each regime as parameters for estimation. Both these discrete variables impact upon the total number of variables in the parameter space. In Miazhynskaia and Dorffner (2006), an AR-GARCH model was used to review popular model selection tools in the Bayesian setting. It was found that, for the simulated data used, the Bayesian model selection methods outperformed likelihood based model selection methods.

Chan and Theoharakis (2011) tackled the difficult task of parameter estimation for Multiple Regime STAR-GARCH (MRSTAR-GARCH) models. They noted that, in the frequentist setting, parameter estimation is problematic due to numerical difficulties in optimising the likelihood function. An investigation of the likelihood function for MRSTAR-GARCH models was performed in order to determine the nature of these numerical difficulties. It was found that the shape parameter in the transition function caused the problems. They proposed a simple parameter transformation in order to alleviate these difficulties.

In Gerlach and Chen (2008), a STAR-Smooth Transition GARCH (STAR-STGARCH) model was presented, which aimed to capture asymmetries in financial markets. To deal with the model identification problem as the smoothing parameter approaches zero, a prior distribution structure for the conditional mean coefficient parameter $\phi$ that takes into account the magnitude of the smoothing parameter was employed. An extensive simulation study and an application to six daily share market indices were provided. Model selection was performed using Bayes factors.

In their proposal of a STAR-STGARCH model for modelling economic time series, Lundbergh and Teräsvirta (1998) observed that appropriately modelling the conditional mean is critical in avoiding misspecification of the conditional variance. This observation, which also applies to the STAR-GARCH model as a subset of the STAR-STGARCH model, was supported by an experiment in Chan and McAlaer (2002).

The mathematical form of the STAR-GARCH model considered in this thesis is the following combi-
nation of the STAR and GARCH models shown in (2.1.1) and (2.1.4), respectively:

\[ x_t = \phi_{1,0} + \phi_{1,1} x_{t-1} + \cdots + \phi_{1,k} x_{t-k} + \]

\[ (\phi_{2,0} + \phi_{2,1} x_{t-1} + \cdots + \phi_{2,k} x_{t-k}) F_t + \varepsilon_t \]

\[ \varepsilon_t = \sqrt{h_t} \eta_t \]

\[ h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_l \varepsilon_{t-l}^2 + \beta_1 h_{t-1} + \cdots + \beta_m h_{t-m}. \]

In (2.1.5), \( \eta_t \sim D(0,1) \) is an error term, where \( D(0,1) \) is an error distribution with mean 0 and variance 1. The transition function is any smooth function that ranges between zero and one. Here it is prescribed to be the logistic function shown in (2.1.3).

Chan and McAleer (2002) provided guidance on the MLE of STAR-GARCH models using evidence from simulated data. They described statistical and structural properties of the model as well as its application to the S&P500. Following this, Chan and McAleer (2003) highlighted some problems with estimation for STAR-GARCH and STAR-STGARCH models. They found that the QMLE was sensitive to the choice of optimisation algorithm employed and that parameter estimates were sensitive to extreme observations.

Boero and Marrocu (2002) applied STAR-GARCH models in economics and finance. They employed separate models to investigate stock market returns for countries in the Pacific basin using either the S&P500, or the Nikkei225 as the alternative threshold variables. This provided evidence of Japan’s leading role in the period leading up to the Japanese crisis, which was followed by a period in which the region was more closely linked to US markets.

While some more complicated extensions of the STAR-GARCH model exist, some restraint was applied when deciding how general the model should be for the purposes of combining parameter estimation and model selection into a single task. This was due to the understanding that, when moving to the multivariate setting, this task would be considerably more complicated.

### 2.1.4 Multivariate STAR Models

The Multivariate STAR model, often called the Vector STAR model, is an extension of the Vector AR (VAR) model in the same sense that the univariate STAR model is an extension of the AR model. The VAR was initially introduced by Sims (1980) and has been used successfully to model data in a broad range of fields such as energy, civil engineering and macroeconomics (Kulshreshtha and Parikh 2000; Chandra and Al-Deek 2009; Litterman 1986).

The VAR model was extended to include a threshold in the conditional mean by Tsay (1998), creating a
multivariate piecewise linear model. The authors proposed a threshold VAR model building procedure and a test for detecting threshold non-linearity in multivariate data. This test is an extension of the threshold non-linearity test from Tsay (1989) into a more general setting. The model building strategy was then applied to monthly US interest rate data and daily river flow data from two Icelandic rivers. The second of these real-world data sets is analysed in Chapter 5.

The threshold VAR model has commonly been applied in the analysis of economic data. To analyse the impact of changes in oil price and volatility on industrial production and real stock returns, Huang, Hwang, and Peng (2005) applied a threshold VAR model to data from the US, Canada and Japan for the years from 1970 to 2002.

The work of Blanchard and Perotti (2002) was extended by Baum and Koester (2011) to investigate the effects of fiscal shocks, and whether their effects vary for different phases of the business cycle. The study used quarterly economic data from 1976 to 2009 from Germany. Another study that investigated fiscal shocks was Afonso, Baxa, and Slavík (2011). A threshold VAR model was employed to analyse whether the effects of fiscal policy on economic activity differ, depending on financial market conditions. A financial stress variable was employed as the threshold variable. The data analysed was macroeconomic and financial data from 1980 to 2009.

More recently, Ferraresi, Roventini, and Fagiolo (2015) investigated how the effects of fiscal shocks on output dynamics depend on the state of credit markets through the implementation of a threshold VAR model. This study and those mentioned above found, by testing for threshold non-linearity, that a threshold VAR model was appropriate for the economic data under analysis. In order to test whether shocks to a South African financial conditions index had an asymmetric effect on output, interest rates and inflation, Balcilar et al. (2016) applied two logistic M-STAR models. One model included inflation as the transition variable, and the other incorporated separate transition variables for each series as in Teräsvirta and Yang (2014b). It was found that the South African economy is strongly non-linear in its responses to financial shocks.

From a Bayesian perspective, the VAR model has shown good forecasting performance (Doan, Litterman, and Sims 1984; Litterman 1986; Bańbura, Giannone, and Reichlin 2007, 2010). Carriero, Kapetanios, and Marcellino (2009) employed a Bayesian VAR model to forecast a large panel of 33 exchange rates and found the methodology efficient for handling large data sets. Korobilis (2013) studied Bayesian VAR variable selection from the perspective of forecast performance, based on an application to UK macroeconomic data.

The next logical extension to the threshold VAR model is to allow smooth transitions from one regime to the next. The resulting model has several names in the literature. Here we shall refer to it as the Multivariate STAR model (M-STAR) in order to maintain consistency with the nomenclature.
surrounding multivariate GARCH models.

A highly general form of the M-STAR($k$) model is presented in Teräsvirta and Yang (2014a). The model presented there allows for differing transition functions between series as well as multiple regimes. The two regime parametrisation of that model with no intercept terms can be expressed as follows:

$$x_t = \left[ \tilde{B}_T^1 + \tilde{F}_t \tilde{B}_T^2 \right] \tilde{z}_t + \varepsilon_t, \quad (2.1.6)$$

where

$$\tilde{B}_T^1 = [\Phi_{1,1}, \ldots, \Phi_{1,k}]$$

$$\tilde{B}_T^2 = [\Phi_{2,1}, \ldots, \Phi_{2,k}]$$

$$\tilde{F}_t = \begin{bmatrix}
F_{1,t} \left( s_{1,t-d_1}, \gamma_1, c_1 \right) & 0 \\
0 & \ddots \\
0 & F_{p,t} \left( s_{p,t-d_p}, \gamma_p, c_p \right)
\end{bmatrix}$$

$$\tilde{z}_t = [x_{t-1}^T, \ldots, x_{t-k}^T]^T.$$

The matrices $\tilde{B}_T^1$ and $\tilde{B}_T^2$ are composed of individual coefficient parameter matrices, $\Phi_{i,j}$, $i \in \{1, 2\}$, $j \in \{1, \ldots, k\}$. The matrix $\tilde{F}_t$ is diagonal: its diagonal entries are transition functions for each time series within the multivariate data set. The error $\varepsilon_t$ for the process is a $p \times 1$ vector that is often assumed to be distributed via a multivariate normal distribution.

The M-STAR model with logistic transition function first appeared in Anderson and Vahid (1998) where it was used to test linearity in multivariate time series against an M-STAR alternative with a logistic transition function. Weise (1999) used a form of the model to investigate whether positive and negative monetary shocks have asymmetric effects on output and prices, and whether the effects of monetary shocks vary over the business cycle. In addition, Weise (1999) investigated whether the effects of monetary shocks vary disproportionately with the size of the shock. The data sets modelled were the quarterly US time series of first differences of the logarithmic industrial production, the consumer price index and M1 (a measure of US money supply) over the period from 1960 to 1995.

Camacho (2004) extended the linearity, model selection and model adequacy tests recently proposed for univariate smooth transition regression models to the multivariate context. Using the testing and selection procedure developed, Camacho (2004) investigated several candidate models for forecasting both output growth and the business-cycle phases of the US economy, based on an index made up of leading economic indicators. A logistic M-STAR model was applied to study asymmetries in the effects of oil price and monetary policy shocks to output growth in the United States in Rahman and Serletis (2010).
The work of Rothman, Dijk, and Franses (2001) used a Smooth Transition Vector Error Correction Model (STVECM), a model that differs slightly from, but is related to the M-STAR model. They applied their model to data derived from industrial production, money, prices and interest rates in order to show that a certain form of non-linear Granger causality from money to output exists. Serra et al. (2011) also employed the STVECM to investigate price transmission patterns in the US corn, ethanol, oil and gasoline price system over the last two decades. Their study identified several interesting relationships regarding how prices are transferred between the variables used.

As previously mentioned with regard to univariate models, the review of STAR models by Dijk, Teräsvirta, and Franses (2002) focused on the modelling procedure for univariate STAR models with an application to the US unemployment rate. In addition, a short review of the literature to date for the M-STAR and STVECM models was provided.

A deviation from the M-STAR model was proposed by Dueker et al. (2011). This model, the Contemporaneous-threshold Multivariate STAR (C-MSTAR) model differs from the M-STAR model in that, rather than allowing for a smooth transition from one extreme linear model to another extreme linear model, the C-MSTAR is effectively a $p$ dimensional mixture model that is weighted together by time varying probabilities. The model was used to investigate the relationship between US stock prices and interest rates.

A survey of threshold VAR and M-STAR models of varying types was provided by Hubrich and Teräsvirta (2013). Estimation, evaluation, linearity testing, forecasting, and other facets of modelling for a range of threshold and smooth transition vector models were reviewed.

The model shown in (2.1.6) is a simpler version of the model presented in Teräsvirta and Yang (2014b). Their multivariate STAR model is a multiple regime model that allows for differing transition functions between series. The model was applied to the Icelandic river flow data set as well as data relating to gasoline prices and consumption. Teräsvirta and Yang (2014a) discussed linearity and misspecification tests in the case in which a single transition variable controls the transitions of the whole system. The authors also tested constancy of the error covariance matrix.

Each of the investigations of M-STAR models mentioned above tackled parameter estimation from a frequentist perspective. Model selection was achieved through the use of either the AIC or the BIC model scoring technique, and/or sequentially removing parameters that were found to be below some pre-determined level of significance. From a Bayesian perspective, Balcombe and Rapsomanikis (2008) investigated the relationship between sugar, ethanol and oil in Brazil using STVECM and threshold VECM models, among others. Their model selection was performed using Bayes factors.

Links in the business cycles between the US, UK, France and Germany were studied by Gefang and
Strachan (2010). This work was an extension of the univariate work of Lubrano (2000, 2001) to a multivariate setting. The model selection was performed through the use of Bayes factors, and the logistic M-STAR model was found to be superior to the VAR model for describing the relationships within the data.

Another investigation into the causal effects from money to output using US data was undertaken by Gefang (2012). This application took a Bayesian perspective and used Bayes factors for model selections. A total of 3,138 models were considered, including several forms of the VAR, M-STAR, VECM and STVECM models. Out of the 3,138 models, the individual model with the highest posterior probability had a probability of less than 10%, while a total of 114 models had non-negligible posterior probability. They concluded that it could be misleading to draw conclusions from a single model. The posterior support for a non-linear model was 96.12%.

Auerbach and Gorodnichenko (2012) investigated the differences between fiscal policy use during recessions and expansions to determine when adjustments are more effective. Due to the non-linearity of the problem, it is possible for several local maxima to exist. In addition, the calculation of confidence intervals for the parameter estimates can be problematic. To overcome these problems, their parameter estimation scheme used an MCMC approach to classical estimation as outlined in Chernozhukov and Hong (2003).

2.1.5 Multivariate GARCH Models

Within a multivariate setting it is just as important, if not more so, to allow the conditional covariance to change over time. By doing so, it is possible to gauge how changes in volatility for one series affect the volatility of another. This transmission of volatility from one series to another is sometimes referred to as a spill over effect. Financial contagion is an example of this type of effect.

The structure of Multivariate GARCH (M-GARCH) models is similar to that of the univariate GARCH models presented in (2.1.4). The structure outlined in Bauwens, Laurent, and Rombouts (2006) for a model with zero conditional mean is as follows:

\[
x_t = \varepsilon_t \\
\varepsilon_t = H_t^{\frac{1}{2}} \eta_t.
\]

(2.1.7)

Here, \( \eta_t \) is an independent identically distributed random vector of length \( p \) such that \( E[\eta] = 0 \) and \( E[\eta \eta^T] = I_p \). The conditional covariance matrix at time \( t \) is given by \( H_t \in \mathbb{R}^{p \times p} \), and is required to be positive definite for all \( t \). The \( p \times 1 \) vector \( \varepsilon_t \) is a vector of errors for the process. Under the assumption that \( \eta \) is distributed as a multivariate normal distribution, the errors of the process are
distributed via a multivariate normal distribution, that is, $\varepsilon_t \sim N_p(0, H_t)$.

For univariate GARCH models, there are several representations of the conditional variance equation, $h_t$. Some examples are the regular GARCH (Bollerslev 1986), the EGARCH (Nelson 1991) and the GJR-GARCH (Glosten, Jagannathan, and Runkle 1993) models. For the M-GARCH model, as one might expect, there are a number of different representations of the conditional covariance equation, $H_t$.

One branch of M-GARCH models is related to modelling the correlation between the series over time. The first of these is the Constant Conditional Correlation (CCC) model of Bollerslev (1990). For a CCC model, the conditional correlations are constant, meaning that the conditional covariances are proportional to the product of the corresponding conditional standard deviations. The relatively low number of parameters in the model makes it a parsimonious option among M-GARCH models.

A Lagrange Multiplier (LM) test was introduced by Tse (2000) to test for constant correlation in an M-GARCH model. Several simulation studies were performed in order to examine the methodology proposed, and applications to real data were also provided. The test was shown to perform comparatively well against another constant correlation test.

The CCC model, due to its parsimonious structure, has been heavily applied in finance. One application tested the hypothesis that correlation among international equity returns was constant for the period, 1960 to 1990. The study, by Longin and Solnik (1995), concluded that international correlation between markets over the 30 year period increased, and that correlation rises in periods where the conditional volatility is large.

An extension to the CCC model, the smooth transition conditional correlation GARCH model, was introduced by Silvennoinen and Teräsvirta (2005). This model allows the correlations to vary smoothly between two extreme states in a similar manner to the way in which the STAR model varies the conditional mean between two extreme states with a smooth transition function. Following on from this, the double smooth transition conditional correlation GARCH model of Silvennoinen and Teräsvirta (2009a) used two transition variables, rather than one, to allow time variation in the conditional correlations.

A generalisation of the CCC model was independently proposed by Engle (2002), Christodoulakis and Satchell (2002) and Tse and Tsui (2002). The new model, called the dynamic conditional correlation (DCC) model, allows the conditional correlation matrix to be time dependent. This new model, together with the CCC model, was used to analyse ten share prices traded on the NYSE with a focus on forecast performance.

From a Bayesian perspective, Fioruci, Ehlers, and Filho (2014) estimated the parameters for the DCC
model under differing error distributions, specifically those distributions that allow for skew and heavy tails. They stated a preference for the multivariate skew density of Bauwens and Laurent (2005) due to its simplicity and generality.

Further extensions to the DCC model have been proposed in the literature. Several extend the model through the introduction of thresholds, or smooth transition functions, as well as by allowing for asymmetry. Some examples of these extensions are found in Billio, Caporin, and Gobbo (2006), Cappiello, Engle, and Sheppard (2006) and Audrino and Trojani (2011).

The other branch of M-GARCH models focuses on modelling the conditional covariance, as opposed to employing a structure related to the correlation of the series. The covariance structure known as the VECH model is the most natural extension of the univariate GARCH model into the multivariate setting. It was originally proposed by Bollerslev, Engle, and Wooldridge (1988). Under the VECH conditional covariance equation, each element of the conditional covariance matrix is a linear function of the lagged squared errors, cross-products of errors and lagged values of the conditional covariance matrices.

The conditional covariance equation for the VECH model is:

$$\text{vech}(H_t) = c + \sum_{i=1}^{l} A_i \text{vech}(\varepsilon_{t-i} \varepsilon_{T-i}^T) + \sum_{j=1}^{m} B_j \text{vech}(H_{t-j}),$$  \hspace{1cm} (2.1.8)

where $A_i, B_j$ are the $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ coefficient parameter matrices and $c$ is an intercept vector of length $\frac{p(p+1)}{2}$. The parameters $l$ and $m$ are the model orders for the ARCH and GARCH parts of the model.

The advantages of using the VECH formulation are that the model is sufficiently general, and the interpretations of the parameters $A, B$ and $c$ are fairly straightforward. Unfortunately, the number of parameters requiring estimation is $(l + m) \left( \frac{p(p+1)}{2} \right)^2 + \frac{p(p+1)}{2}$, which is large, even for small $p, l$ and $m$. A high level of computational power is required for estimating all of the parameters of the model.

Hudson and Gerlach (2008) proposed a novel approach that restricted the number of parameters in the VECH model to $\frac{3}{2} p (p + 1) + 2 p (p - 1)$. This amount is always smaller than the number of parameters for the full VECH model, but larger than the number of parameters for the BEKK model. Estimation of the model’s parameters in Hudson and Gerlach (2008) was from a Bayesian perspective and the prior distribution was used to enforce positive definiteness of the conditional covariance matrices.

Another formulation of the conditional covariance equation is the BEKK model proposed by Engle and Kroner (1995). The name is an acronym made out of the initials of those who contributed to the development of multivariate GARCH models, namely, Baba, Engle, Kraft and Kroner. The conditional
covariance equation was originally defined as follows:

\[ H_t = C_0 C_0^T + \sum_{q=1}^{Q} \sum_{i=1}^{l} A_{qi} \varepsilon_{t-i} \varepsilon_{t-i}^T A_{qi}^T + \sum_{q=1}^{Q} \sum_{j=1}^{m} B_{qj} H_{t-j} B_{qj}^T, \]  

(2.1.9)

where \( A_{qi}, B_{qj} \) and \( C_0 \) are \( p \times p \) parameter matrices, with \( C_0 \) being lower triangular. The parameter \( Q \) is included in the model to accommodate more general representations of the conditional covariance equation. When \( Q > 1 \), an identification problem arises due to the existence of several parametrisations that lead to the same model. Conditions to eliminate redundant, observationally equivalent, models were provided by Engle and Kroner (1995).

The formulation of the BEKK conditional covariance equation ensures that the conditional covariance matrix \( H_t \) is positive definite. Another advantage of this formulation is that the number of parameters that must be estimated for the model is always smaller than that required for the VECH model.

The number of parameters that must be estimated for the full BEKK model, in general, is \((l + m) p^2 Q + \frac{p}{2} (p + 1)\). While this is fewer than for the VECH model, even for small \( p \), it still requires a considerable amount of computational effort. The diagonal BEKK (DBEKK) model is a simpler form of the full BEKK model in which the coefficient matrices are diagonal. It is evident that this form of the BEKK model has fewer parameters again: in general, the number of parameters for the DBEKK model is \((l + m) p Q + \frac{p}{2} (p + 1)\). For some multivariate data sets, the more parsimonious DBEKK model is sufficient to account for the conditional covariance of the data.

In the literature, it is common for the DCC model to be applied to forecast conditional correlations, whereas the BEKK model is used to forecast conditional covariances. Caporin and McAleer (2012) investigated the reasons why the DCC and BEKK models coexist, when one model can do almost everything the other model can do. They found that, when a large number of series are under analysis, the curse of dimensionality can be reduced through the use of targeting. This makes the BEKK model a reasonable alternative to the DCC model in such circumstances.

Practical applications of the BEKK formulation lie in the fields of finance and economics. Often, the subject of investigation is the transfer of volatility between markets. Kearney and Patton (2000) looked at the transfer of volatility in exchange rate data. A spill over of volatility from the Hong Kong stock exchange to the exchanges in mainland China was found to be unidirectional in a study by Li (2007) that applied the BEKK model. Saleem (2009) measured the effect of the 1998 Russian financial crisis on other international financial markets using several bivariate BEKK models.

Several multivariate ARCH and GARCH type models were applied to bivariate foreign exchange rate data in Vrontos, DellaPortas, and Politis (2003). Differences in inference between the Bayesian method and classical estimation were found for scenarios in which the posterior distribution was non-normal.
The posterior predictive distribution was used to compare the models applied: an M-GARCH model with diagonal covariance was found to be preferable, based on one step ahead prediction.

A Bayesian non-parametric modelling approach for M-GARCH models was proposed by Jensen and Maheu (2013). Their approach was applied to a ten asset portfolio, with model selection performed using Bayes factors. They found that their best semi-parametric model performed significantly better in density forecasts than models with Student-$t$ innovations. The paper also provides guidance on expanding their approach to include a VAR model as the conditional mean.

There have been several valuable reviews and surveys of the rapidly changing literature on M-GARCH models. Some examples are Bauwens, Laurent, and Rombouts (2006), Bauwens, Hafner, and Laurent (2012), Silvennoinen and Teräsvirta (2009b), Francq and Zakoïan (2010) and Tsay (2013).

Another separate class of volatility models that has not been reviewed in detail is the Stochastic Volatility (SV) models. This class of models has been extended into the multivariate framework and the models are referred to as Multivariate Stochastic Volatility (MSV) models. Worthwhile reviews of the literature on MSV models were provided by Asai, McAleer, and Yu (2006) and Chib, Omori, and Asai (2009), who covered methods of estimation and model selection. The SV model is defined in a similar way to the GARCH models and, within the univariate framework, has commonly been applied to financial time series.

### 2.1.6 Multivariate STAR-GARCH Models

Accounts in the literature of specific applications of an M-GARCH model in which the conditional mean is modelled by an M-STAR process are non-existent. Indeed, Tong (2015) states that “TAR models for multivariate time series are still in their infancy”. There are, however, occasional discussions of modelling the conditional covariance using an M-GARCH process with a conditional mean structure that is something other than an M-STAR process.

Within the Bayesian framework, Uhlig (1997) proposed a methodology for estimating the parameters for a VAR model in which the conditional covariance is modelled as an MSV process. Another investigation that used a VAR structure for the conditional mean was Li and Majerowska (2008). They used an M-GARCH model with a BEKK conditional covariance equation for the conditional covariance. The purpose of the paper was to investigate linkages between the US, German, Polish and Hungarian share markets.

In an investigation into the effect of the conditional mean specification in M-GARCH modelling, Osiewalski and Pipień (2004b) used a VAR model with M-GARCH errors to model a set of bivariate foreign exchange data. Their approach was similar to that of Osiewalski and Pipień (2004a).
total of ten models of the conditional covariance were proposed, including CCC, VECH and BEKK models, both conditionally Student-\(t\) distributed and normally distributed. Estimation was performed from a Bayesian perspective, and model selection was made through the use of Bayes factors. Under four scenarios, a combination of long and short data lengths, and with or without the use of exogenous variables, the \(t\)-BEKK specification was always found to be the best model. They found that an appropriate specification of the conditional mean was crucial for obtaining a good model fit.

Goeij and Marquering (2004) investigated the interaction of bond and share markets through the use of a modified form of the diagonal VECH model in which the conditional mean was modelled as a VAR process. They chose the diagonal VECH model over the full VECH model as they expected spill over effects to be minimal. Their modified form of the diagonal VECH model was a generalisation of the approach from Glosten, Jagannathan, and Runkle (1993) to a multivariate setting. It was found that their asymmetric diagonal VECH model was better suited than the original diagonal VECH model for modelling data sets that exhibit significant asymmetries.

To analyse the transmission of volatility between the Canadian and US share markets, Karolyi (1995) applied a VAR model in which the conditional covariance was modelled by a BEKK model. The BEKK covariance equation was slightly modified to account for the effects of holidays and weekends on the volatility. Within the Bayesian framework, and also focusing on the transmission of volatility, Polasek and Ren (2000) analysed daily exchange rates between the US, Germany and Japan around the time of the Asia crisis. The structure of the conditional mean chosen was a VAR process that took the conditional covariance into account. This sort of model is referred to as GARCH in mean. The conditional covariance was modelled by a VECH conditional covariance equation. The model’s parameters were estimated within the Bayesian framework.

### 2.2 Algorithms for Posterior Analysis

Within the Bayesian paradigm, the information of interest regarding the parameters of a model is contained within the posterior distribution. The posterior distribution is the primary source (though not necessarily the sole source) of information for estimating parameters and determining the levels of uncertainty surrounding those estimates.

The posterior distribution is analysed by the application of Bayes’ theorem, which states:

\[
p(\Theta | x) = \frac{p(x|\Theta) p(\Theta)}{p(x)}. \tag{2.2.10}
\]

On the right hand side of (2.2.10), \(p(x|\Theta)\) is the likelihood function for the model as commonly defined.
in statistics, and \( p(\Theta) \) is the prior distribution. The prior distribution is a function of \( \Theta \) that reflects
the researcher’s beliefs about the parameter before reviewing the data. This function can be set up
to provide very little information about the parameter, or, if the researcher has experience or has
performed other research involving the parameter, it may be set up to be rather informative.

The denominator \( p(x) \) in (2.2.10) is not a function of \( \Theta \) and is therefore constant for differing values
of \( \Theta \). It is typically ignored and the proportional relationship between the posterior distribution, the
likelihood function and the prior distribution is analysed. Therefore Bayes’ theorem simplifies to

\[
p(\Theta|x) \propto p(x|\Theta)p(\Theta).
\]

(2.2.11)

For simple models, the posterior distribution will be relatively simple to analyse. It may be possible
to obtain a closed form analytical solution to the posterior distribution. For more complicated models,
the posterior distribution may be difficult to analyse directly if the posterior distribution is of a non-
standard form. It may require a numerical approximation of the distribution through simulations
which can be performed via the application of a Markov chain Monte Carlo (MCMC) algorithm. Some
models may require the application of more than one type of MCMC algorithm in order to simulate
the posterior distribution.

For situations in which the conditional posterior distributions of the parameters of a model are of
standard form, a simple algorithm that can be used to simulate the posterior distribution is the Gibbs
Sampler.

### 2.2.1 Gibbs Sampler

The Gibbs sampler was originally proposed by Geman and Geman (1984) and is named after the
American physicist Josiah Willard Gibbs. Some other detailed presentations of the application of the
algorithm in various situations can be found in Tanner and Wong (1987), Gelfand et al. (1990) and
Gelfand and Smith (1990).

The Gibbs sampler is most useful when we have a multivariate parameter of interest \( \theta = (\theta_1, \ldots, \theta_n) \),
and the full conditional distributions for \( \theta_j \) are able to be derived and sampled from. To implement the
algorithm, we either randomly or deterministically select some initial values \( \theta^{(0)} \). At the \( i \)th iteration
of the algorithm, we cycle through the parameter vector, sampling each of the variables \( \theta_j \) from their
conditional distributions. During execution of the algorithm, at the beginning of the \( i \)th iteration, the
state of the parameter vector will be \( \theta^{(i-1)} = \{\theta_1^{(i-1)}, \ldots, \theta_n^{(i-1)}\} \). The algorithm is then:
Section 2.2: Algorithms for Posterior Analysis

1. For all $j \in \{1, \ldots, n\}$, 
   \[
   \theta_j \sim p \left( \theta_j | \theta_1^{(i)}, \ldots, \theta_j^{(i)}, \theta_{j+1}^{(i-1)}, \ldots, \theta_n^{(i-1)} \right).
   \]

2. Set $i = i + 1$, repeat until the algorithm has converged.

If we allow $i$ to become sufficiently large, the algorithm will converge to the joint distribution $p(\theta)$ (Geman and Geman 1984) and the stored simulated parameter values will be a numerical representation of the joint distribution. The influence of the initial values can be removed by ignoring values simulated at the beginning of the algorithm. This is known as the “burn in” period.

The Gibbs sampler algorithm has been extensively applied for the purposes of posterior simulation for various types of models. Some early applications to autoregressive like models are found in Chib (1993), Albert and Chib (1993), Chib and Greenberg (1994) and McCulloch and Tsay (1994). Interestingly, the Gibbs sampler algorithm is actually a special case of another very popular MCMC algorithm, the Metropolis-Hastings algorithm.

2.2.2 Metropolis-Hastings

The Metropolis algorithm was first introduced by Metropolis et al. (1953) and later generalised in Hastings (1970) to form the Metropolis-Hastings algorithm. Interestingly, there is some controversy over which authors of Metropolis et al. (1953) were responsible for the discovery of the algorithm (Rosenbluth 2003).

To use the Metropolis-Hastings algorithm, suppose we have some non-standard target distribution $\pi(\theta)$ that we wish to simulate from, where $\theta$ is either a multivariate parameter vector or simply a univariate parameter of interest. The Metropolis-Hastings algorithm allows us to obtain samples from the target distribution by making proposals from a known standard distribution, $q(\theta)$ that is easily simulated from. We then accept or reject the proposed value on the basis of a calculated acceptance probability, or acceptance ratio. We would expect that the closer the proposal distribution ($q(\theta)$) is to the target distribution, the better the algorithm will perform.

After specifying the initial values $\theta^{(0)}$ and assuming that the most recently simulated value is $\theta^{(i-1)}$, the $i$th iteration of the algorithm is as follows:

1. Generate a proposed value $\theta^*$ from the proposal density $q \left( \theta^* | \theta^{(i-1)} \right)$.

2. Calculate the acceptance probability, $r$:

   \[
   r = \min \left( 1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{q(\theta^{(i-1)} | \theta^*)}{q(\theta^* | \theta^{(i-1)})} \right). \tag{2.2.12}
   \]
3. If \( r > u \), where \( u \sim U(0, 1) \), set \( \theta^{(i)} = \theta^* \), else set \( \theta^{(i)} = \theta^{(i-1)} \).

4. Repeat steps one to three until the algorithm has converged.

When the proposal distribution is symmetric, that is, \( q(\theta^*|\theta^{(i-1)}) = q(\theta^{(i-1)}|\theta^*) \), the ratio of the proposal distributions is equal to one and is removed from the acceptance ratio formula in (2.2.12). In Metropolis et al. (1953), there was a requirement that the proposal distribution be symmetric. It was the contribution of Hastings (1970) that allowed the removal of the need for symmetry in the proposal distribution and thus extended the algorithm to the more general case.

The algorithm described above is an example of a Random Walk Metropolis-Hastings Algorithm. This is because the proposal distribution is dependent upon the most recently simulated value, \( \theta^{(i-1)} \). This, however, is not necessarily required. When the proposal distribution is not dependent upon the most recently simulated value, the algorithm is known as an independence Metropolis-Hastings algorithm.

### 2.2.3 Algorithms to Improve Acceptance Rates

Applications of the Metropolis-Hastings algorithm in either a high dimensional setting or situations where there is little information within the data regarding a parameter of interest can lead to low acceptance rates for the algorithm. In order to ensure satisfactory acceptance rates, one may need to specify a proposal distribution with relatively low variance. This can lead to slow convergence and/or the need for many iterations to allow proper exploration of the parameter space.

By creating an algorithm that leads to improved acceptance rates over the standard Metropolis-Hastings algorithm, a user can afford to use a larger variance in the proposal distribution, leading to faster convergence and/or the need for fewer iterations of the algorithm to explore the parameter space. It is important that the new algorithm is not too computationally demanding, for otherwise any additional advantage may be lost due to computational processing time.

There are many algorithms that have been suggested in the literature to improve acceptance rates for the Metropolis-Hastings algorithm. Two such algorithms employed in this thesis are the Multiple-Try Metropolis (Liu, Liang, and Wong 2000) and the Delayed Rejection Metropolis-Hastings (Mira 2001a) algorithms.

**Multiple-Try Metropolis**

The Multiple-Try Metropolis algorithm (Liu, Liang, and Wong 2000) allows several candidate parameters to be proposed at each iteration. An individual candidate parameter is randomly chosen using probabilities that are calculated based upon a user defined function. The individual candidate param-
eter is then tested for acceptance. The algorithm requires the addition of a modified acceptance ratio formula to the original Metropolis-Hastings formula to ensure convergence to the target distribution.

The Multiple-Try Metropolis algorithm was further generalised in (Pandolfi, Bartolucci, and Friel 2010) and Pandolfi, Bartolucci, and Friel (2014). The generalisation also allows the application of the algorithm to model selection problems using a Reversible Jump-style (Green 1995) algorithm.

As in the Metropolis-Hastings algorithm, we have a target distribution $\pi(\theta)$ and a proposal distribution $q(\theta | \psi)$. Here $\psi$ is usually set to be the most recent value simulated. In addition, we have an arbitrary function $w(\theta, \psi)$ which satisfies $w(\theta, \psi) > 0$. The generalised Multiple-Try Metropolis (MTM) algorithm proposed in Pandolfi, Bartolucci, and Friel (2010) is as follows:

1. Draw $k$ independent trial proposals $\theta^*_{1}, \ldots, \theta^*_k$ from the proposal distribution $q(\theta^* | \theta^{(i-1)})$.
2. Select a point $\theta^*$ from the set $\{\theta^*_1, \ldots, \theta^*_k\}$ with probability
   \[
   p_\theta = \frac{w(\theta^*, \theta^{(i-1)})}{\sum_{j=1}^{k} w(\theta^*_j, \theta^{(i-1)})}.
   \]
3. Draw realisations $\psi^*_1, \ldots, \psi^*_{k-1}$ from the proposal distribution $q(\psi^* | \theta^*)$ and set $\psi^*_k = \theta^{(i-1)}$.
4. Define
   \[
   p_\psi = \frac{w(\theta^{(i-1)}, \theta^*)}{\sum_{j=1}^{k} w(\psi^*_j, \theta^*)}.
   \]
5. Calculate the acceptance ratio, $r$:
   \[
   r = \min \left( 1, \frac{\pi(\theta^*) q(\theta^{(i-1)} | \theta^*) p_\psi}{\pi(\theta^{(i-1)}) q(\theta^* | \theta^{(i-1)}) p_\theta} \right). \tag{2.2.13}
   \]
6. If $r > u$, where $u \sim U(0, 1)$, set $\theta^{(i)} = \theta^*$, else, set $\theta^{(i)} = \theta^{(i-1)}$.
7. Repeat steps one to six until the algorithm has converged.

The acceptance ratio in (2.2.13) ensures that the Markov chain converges to the stationary distribution, $\pi(\theta)$.

Some authors have proposed extensions and modifications to the MTM algorithm to further improve acceptance rates and/or the speed of the algorithm. For example, Craiu and Lemieux (2007) proposed a modification to the algorithm that allowed the use of correlated proposals. Casarin, Craiu, and Leisen (2013) extended the algorithm to allow the use of different proposal distributions in the multiple try stage of the algorithm. Martino and Read (2013) provided a short review of the MTM algorithm, together with simulation studies exploring the effectiveness of differing designs for these algorithms.
Delayed Rejection Metropolis-Hastings

The Delayed Rejection Metropolis-Hastings (DRMH) algorithm was first proposed by Tierney and Mira (1999), and later presented by Mira (2001a, 2001b). An extension of the DRMH algorithm to a model selection setting was explored in Green and Mira (2001).

The idea behind the DRMH algorithm is that once a candidate parameter is rejected, we propose and test a new candidate parameter for acceptance, rather than adopting the previous value in the chain as the current value. As in the MTM algorithm, there is a modified acceptance probability formula for the second proposed value. This maintains the condition of detailed balance. The sampler may be set up to allow several rejections at each iteration of the algorithm.

One of the advantages of the DRMH algorithm is that it allows partial local adaptation of the proposal at each iteration. This is done by specifying proposals that depend upon the candidates that have already been rejected in the current iteration of the algorithm. The user is able to specify a larger variance for the first proposal and then reduce the variance of the proposal distribution on subsequent proposals. This allows for a potentially faster and more efficient exploration of the state space. Haario et al. (2006) describe the strategy as “a sort of ‘first bold’ versus ‘second timid’ tennis serving strategy”.

The implementation of the algorithm in this thesis is limited to delaying the rejection only once at each iteration. Therefore, at each iteration, the maximum number of possible proposals is two. We need to specify proposal distributions at these steps and for the times that the algorithm is employed within this thesis. For the first proposal, we use the most recent simulated value as a location parameter. For the second proposal, we allow the rejected candidate to be used in determining the location parameter. The notation for the first and second proposal distributions will be \( q_1(\theta^*|\theta^{(i-1)}) \) and \( q_2(\theta^*|\theta^{*},\theta^{(i-1)}) \), respectively.

Assuming that we are at the \( i \)th iteration, the current state of the parameter vector will be \( \theta^{(i-1)} \), and the current iteration of the algorithm executes the following steps:

1. Generate a proposed value, \( \theta^* \) from the proposal density, \( q_1(\theta^*|\theta^{(i-1)}) \).
2. Calculate the acceptance ratio, \( r_1 \):
   \[
   r_1(\theta^*,\theta^{(i-1)}) = \min \left( 1, \frac{\pi(\theta^*) q_1(\theta^{(i-1)}|\theta^{(i-1)})}{\pi(\theta^{(i-1)}) q_1(\theta^*|\theta^{(i-1)})} \right).
   \]
3. Simulate \( u \), where \( u \sim U(0,1) \).
   (a) If \( r_1 > u \), accept the proposed value and set \( \theta^{(i)} = \theta^* \).
and θ the parameter space that require estimation. Therefore, an RJMCMC algorithm is useful here, as it
the model order is a variable that is unknown and its value impacts upon the number of terms within
The algorithm has been applied in a wide range of fields. Within the context of time series analysis,
Hastie and Green (2012).
RJMCMC algorithm is useful in situations where “the number of things you don’t know is one of the
algorithm proposed by Green (1995). There have been several papers explaining this algorithm and
An extension of the Metropolis-Hastings algorithm to an even more general setting is the RJMCMC
θ′ from q2(θ′ | θ∗, θ(i−1)).
(i) Calculate the acceptance ratio for the proposed parameter θ′, r2:
\[
r_2(θ′, θ∗, θ(i−1)) = \min \left( 1, \frac{π(θ′)}{π(θ(i−1))} \frac{q_1(θ′)}{q_1(θ∗ | θ(i−1))} \frac{q_2(θ′ | θ(i−1))}{q_2(θ∗ | θ(i−1), θ∗)} \frac{1 − r_1(θ∗, θ′)}{1 − r_1(θ∗, θ(i−1))} \right).
\]
(2.2.14)
(ii) Simulate a new u, again where \( u \sim U(0, 1) \). If \( r_2 > u \), accept the new proposed value
θ′ and set θ(i) = θ′. Else, set θ(i) = θ(i−1).
4. Repeat steps one to three until the algorithm has converged.
Mira (2001a) provides the acceptance probability for the \( j \)th stage of a DRMH algorithm with several
stages, as opposed to the simple two stage algorithm presented above. Assuming that the proposal at
the \( j \)th stage is θj, the acceptance probability formula will be
\[
r_j(θ_j, θ^{(i−1)}, θ_1, \ldots, θ_{j−1}) = \min \left( 1, \frac{π(θ_j | θ^{(i−1)}, θ_1, \ldots, θ_{j−1})}{π(θ^{(i−1)}, θ_1, \ldots, θ_{j−1})} \frac{q_1(θ_j | θ^{(i−1)}, θ_1, \ldots, θ_{j−1})}{q_1(θ^{(i−1)}, θ_1, \ldots, θ_{j−1})} \right.
\]
\[
\times \frac{q_2(θ_j | θ^{(i−1)}, θ_1, \ldots, θ_{j−1})}{q_2(θ^{(i−1)}, θ_1, \ldots, θ_{j−1})} \frac{1 − r_1(θ^{(i−1)}, θ_1, \ldots, θ_{j−1})}{1 − r_1(θ^{(i−1)}, θ_1, \ldots, θ_{j−1})} \left( 1 − r_2(θ_j, θ^{(i−1)}, θ_1, \ldots, θ_{j−1}) \right)
\]
\[
\left( 1 − r_1(θ^{(i−1)}, θ_1, \ldots, θ_{j−1}) \right) \left( 1 − r_2(θ_j, θ^{(i−1)}, θ_1, \ldots, θ_{j−1}) \right) \left( 1 − r_1(θ_j, θ^{(i−1)}, θ_1, \ldots, θ_{j−2}) \right)
\]
(2.2.15)
If we denote the proposed parameters at the first and second stages of the DRMH algorithm by θ∗
and θ′, respectively, and substitute them into the formula in (2.2.15), we obtain the formula shown in
(2.2.14).

2.2.4 Reversible Jump

An extension of the Metropolis-Hastings algorithm to an even more general setting is the RJMCMC
algorithm proposed by Green (1995). There have been several papers explaining this algorithm and
providing examples of its use (Green 2003; Green and Hastie 2009; Hastie and Green 2012). The
RJMCMC algorithm is useful in situations where “the number of things you don’t know is one of the
things that you don’t know” (Hastie and Green 2012).
The algorithm has been applied in a wide range of fields. Within the context of time series analysis,
the model order is a variable that is unknown and its value impacts upon the number of terms within
the parameter space that require estimation. Therefore, an RJMCMC algorithm is useful here, as it
allows us to include the model order as one of the parameters for estimation.

The original formulation of the Reversible Jump methodology does not limit itself to problems involving model selection between nested models. The methodology is flexible enough to enable jumps from one model type to a completely different model type, where the length of the parameter vectors may be the same or different. An example from time series analysis is discussed in Vrontos, Dellaportas, and Politis (2000), where two model types, GARCH and EGARCH models are proposed, and the estimation procedure is executed for the posterior probabilities and other parameters of each model.

The algorithm is structured similarly to the Metropolis-Hastings algorithm in that, after proposals are made, the acceptance of those proposals is tested using an acceptance probability or acceptance ratio formula. Suppose that we have a set of models, \( \mathcal{M} \), and the current state of the RJMCMC algorithm at the beginning of the \( i \)th iteration is model \( m \) with corresponding parameter vector \( \theta_m \). A proposal to move to new model \( m^* \) is made from the jumping distribution \( j(m^* | m) \) with corresponding coefficient vector \( \theta_{m^*} \). A vector \( u^* \) is generated from a proposal distribution \( q(u^* | \theta_m, m, m^*) \) and we set \( g_{m,m^*}(\theta, u^*) = g_{m,m^*}(\theta_{m^*}, m, m^*) \) for some specified invertible function \( g_{m,m^*} \) such that \( g_{m,m^*}^{-1} = g_{m,m^*} \).

This is a function that transforms the variables of one model into variables for the other. The dimension matching criterion is met here in that \( d(\theta_m) + d(u^*) = d(\theta_{m^*}) + d(u) \), where \( d(u) \) denotes the dimension of \( u \).

According to Green (1995), the acceptance probability of moving to the new model \( m^* \) is

\[
r = \min \left( 1, \frac{\pi(x | \theta_{m^*}, m^*) \pi(\theta_{m^*} | m^*) j(m^* | m^*) q(u^* | \theta_{m^*}, m^*)}{\pi(x | \theta_m, m) \pi(\theta_m | m) j(m^* | m) q(u^* | \theta_m, m)} \right),
\]

(2.2.16)

where \( \pi(m^*) \) is the prior model probability for model \( m^* \) and \( J \) is the Jacobian of the transformation such that \( J = \frac{\partial(\theta_{m^*}, u^*)}{\partial(\theta_m, u)} \). The acceptance probability in (2.2.16) ensures the chain satisfies the condition of detailed balance and therefore converges to the target distribution of \( \pi(\theta, m | x) \).

The choices of proposal distributions and function \( g_{m,m^*} \) do not affect the results of the algorithm. However, depending on the model, they can be crucial to rate of the convergence of the Markov chain. As previously mentioned, the function \( g_{m,m^*} \) transforms the parameters of one model to parameters for the other. For the models used throughout this thesis, we have found it is best to propose all the elements of the new model coefficient vector directly and not use the transformation function \( g_{m,m^*} \).

This is because most of our model moves take place between nested models. In this situation, the coefficient parameter values that are common to both models can change significantly from one model to the other, as noted by Vrontos, Dellaportas, and Politis (2000). The parameters that are common to the current and the proposed model may represent regions of low posterior probability, leading to slow convergence.
The advantage of simulating the complete parameter vector is highlighted in Troughton and Godsill (1997) through differences in the speeds of convergence between the two types of proposals they use, partial versus full. The method of proposing model moves using the birth and death procedure converges significantly more slowly than the procedure of proposing a completely new coefficient vector.

Therefore, ignoring the use of the transformation function and proposing the complete parameter vector for the proposed model, the required condition of dimension matching is still met and the Jacobian in (2.2.16) becomes one. The acceptance probability in (2.2.16) also contains the jumping distribution $j(m^*|m)$. For nested models, we define this distribution so that the proposed jumps brought about by $j(m^*|m)$ are small, but on occasions are large enough to allow good mixing. With this in mind, Troughton and Godsill (1997) chose the following discretised Laplacian density:

$$j(m^*|m) \propto \exp \left[ -\Delta_m |m^* - m| \right] ,$$

for $m^* \in \{1, \ldots, m_{\text{max}}\}$. In this thesis, we shall employ the above distribution as the jumping distribution when we are dealing with nested models. One advantage of such a choice is that $j(m^*|m) = j(m|m^*)$ and therefore disappears from the acceptance probability calculation. For situations where the models are not nested, the probability of model moves will be proportional to one, that is, the probabilities of jumping to each model will be equal. Again, in this situation, the jumping densities disappear from the acceptance probability and the acceptance probability in (2.2.16) simplifies to

$$r = \min \left( 1, \frac{\pi(x|\theta_{m^*}, m^*) \pi(\theta^*|m^*) \pi(m^*) q(\theta^*|\theta, m, m^*)}{\pi(x|\theta_m, m) \pi(\theta|m) \pi(m) q(\theta|\theta, m, m^*)} \right)$$

$$= \min \left( 1, \frac{\pi(\theta_{m^*}, m^*|x) q(\theta^*|\theta, m, m^*)}{\pi(\theta_m, m|x) q(\theta^*|\theta, m, m^*)} \right) .$$

The Reversible Jump algorithm allows the model selection step of the modelling cycle to be included as part of the parameter estimation stage. This simplifies the modelling cycle and will be the model selection tool used throughout this thesis. There are several other model selection tools available. Some of those relevant to the Bayesian framework are discussed in the following section.

2.3 Model Selection

Model selection is an important part of the time series modelling process. A wide range of selection methods are available to allow the modeller to discriminate between candidate models. A common approach to model selection is the application of some sort of scoring technique, where models are tested for fit and penalised for model complexity. From the frequentist perspective, examples of this process are the Akaike Information Criterion (AIC) (Akaike 1974) and the Bayesian Information Criterion.
(BIC) (Schwarz 1978). Recently, Resende and Dorea (2016) proposed the Efficient Determination Criterion as a generalisation of these criteria, providing an application to M-GARCH models with BEKK covariance equations.

A model scoring technique that lends itself well to a Bayesian methodology, where the posterior distribution is simulated through the use of an MCMC algorithm, is the Deviance Information Criterion (DIC) (Spiegelhalter, Best, and Carlin 1998). The DIC assesses model fit while penalising for model complexity and, like the AIC and BIC, is a tool for comparing competing models that does not necessarily identify the “correct” model.

The DIC is calculated as follows:

\[
D_{IC} = 2 \bar{D} - D(\bar{\theta}),
\]

where

\[
\bar{D} = E_\theta[D(\theta)]
\]

\[
D(\theta) = -2 \ln(p(X|\theta)).
\]

Suppose that \(\{\theta_1, \ldots, \theta_N\}\) is a sample from the posterior distribution obtained after discarding some of the initial values of the Markov chain. To calculate \(\bar{D}\), we simply take the simulated parameter(s) \(\hat{\theta}_i\) and calculate \(D(\theta)\) for each \(i = 1 \ldots N\). The mean of these values is \(\bar{D}\). For the parameter estimates \(\bar{\theta}\), either the posterior mean or median may be used. Then \(D(\bar{\theta})\) is simply \(-2\) times the log likelihood using the posterior parameter estimates.

Mathematically, \(\bar{D}\) and \(D(\bar{\theta})\) are expressed as follows:

\[
\bar{D} = \frac{1}{N} \sum_{i=1}^{N} D(\theta_i)
\]

\[
D(\bar{\theta}) = D\left(\frac{1}{N} \sum_{i=1}^{N} \theta_i\right).
\]

As it is simple to calculate the DIC when an MCMC sample has been used, we shall employ this index later in this thesis. As discussed in Section 2.2.4, the RJMCMC methodology of Green (1995) may be employed for both parameter estimation and model selection. There are some obvious advantages of merging these tasks into one process. Depending upon how one calculates the interval estimates from the MCMC output, the degree of uncertainty surrounding the model may be incorporated into the uncertainty surrounding the other parameters of the model. This can potentially provide a fairer indication of the actual uncertainty for the final fitted model. In addition, the process of model estimation and model selection has the potential to be fully automated, creating a “black box” approach to modelling.
Section 2.4: Effective Data for RJMCMC

There are, however, other types of model selection tools employed within the Bayesian setting. For example, Carlin and Chib (1995) introduced a scheme that uses a product space for all possible model parameters and a model indexing variable. This methodology was linked to the RJMCMC methodology in (Green 1995) through the unifying framework presented in Godsill (2001). A connection between the methodology of Green (1995) and Carlin and Chib (1995) was also identified in Dellaportas, Forster, and Ntzoufras (2002). In Miazhynskaia and Dorffner (2006), a univariate AR-GARCH model was used to review popular model selection tools in the Bayesian setting including RJMCMC.

2.4 Effective Data for the Reversible Jump Algorithm

Within a time series setting that employs an RJMCMC algorithm to move between models of varying dimensions, the amount of data available to calculate the acceptance probability of the move between models will often depend on the orders of the models being jumped to and from.

For example, let us assume that we are dealing with a data set of length $N$ and are seeking to move from a simple AR($k$) model to a candidate AR($k^*$) model, where $k \neq k^*$ and $k \in \{1, \ldots, k_{\text{max}}\}$. The form of the acceptance probability for making such a move is detailed in Section 2.2.4 and includes the calculation of the posterior probabilities for each of the models. This, in turn, involves calculating the likelihood function for each model as a function of the data.

For time series data, the length of data that is available for use in calculating the likelihood function often depends upon the model order. For example, to calculate the likelihood function for an AR($k$) model, it is common practice to treat the first $k$ data points as initial conditions. Therefore, the length of the available data, or effective data length, is $n = N - k$.

The question that then arises is this: when calculating the likelihood functions for two models of differing dimensions, say $k$ and $k^*$, what data should be made available for each of the calculations?

One option is to use the maximum amount of data available for each calculation. That is, for the calculation of the likelihood functions for the AR($k$) and AR($k^*$) models, we condition on the first $k$ and $k^*$ data points, respectively, and use the following $n = N - k$ and $n^* = N - k^*$ points, respectively, for the likelihood function calculations. This will mean that the numbers of data points used in these calculations are not consistent with each other.

The other option is to calculate each likelihood function, conditional on the same initial values. Here, the same $s^* = \max(k, k^*)$ data points would be assigned as initial conditions and both likelihood functions would be calculated using the same $n^* = N - s^*$ data points.

One solution to the above question is not to condition on any data points at the start of the data set,
but instead to create an initial condition parameter. This initial condition parameter vector would be included as another parameter requiring estimation. Under this scenario, the complete data set would be available for calculation of the likelihood function for both the current model and the candidate model.

There are a number of points to note regarding this solution. First, for more complicated models, the estimation of the initial condition parameter may not be entirely straight forward. Second, some extra assumptions may be required to facilitate the estimation of the initial condition parameter vector. Third, the estimation of the initial condition parameter vector is of little to no interest to the practitioner and, as a result, rarely occurs in practice. One reason why the first few data points can be sacrificed as initial conditions is that data sets are now usually sufficiently large enough to withstand the loss of a few data points.

Given the impracticality of the above solution, the choice of whether to use the same data or the maximum amount of data available when calculating both likelihood functions is not immediately obvious. Journal articles will often present the mathematical form of the acceptance probability for an RJMCMC algorithm, but rarely go into detail regarding its calculation. The only reference to the amount of data that should be included in each likelihood calculation was found in an unpublished technical report (Troughton and Godsill 1997) in which it was noted that

“To eliminate the dependence on the scale of the signal, the same vector $y_1$, and hence length $n_e$, is used for both model orders being considered, i.e., all probabilities are conditional on the first $\max(k, k')$ values of $y$.”

This would appear somewhat intuitive as one would think that using the same data in each likelihood calculation should reduce any bias for one particular model over another, based on the length of the data set used. This notion was, however, brought into doubt after receiving comments from a reviewer stating that to use the same data in each calculation was not standard practice. Investigation of the matter led to communications with several experts in time series who have published work using the RJMCMC methodology. This communication resulted in conflicting advice pertaining to the method that should be employed. In fact, two co-authors of a paper employing an RJMCMC scheme offered differing opinions on how the calculation should be performed. There appears to be no consensus regarding the amount of data that should be used in each likelihood function calculation.

In practice, when $N$ is large and the difference between $k$ and $k^*$ is relatively small, there will be an insignificant difference between the ratios of the likelihood functions calculated by conditioning on the same initial data points and those calculated using the maximum amount of data available. This, alone, is potentially the reason for the lack of detail in research papers regarding how the calculation
should be performed. For situations where $N$ is relatively small and the difference between $k$ and $k^*$ could possibly be relatively large, it is important to know which method is best. Even when $N$ is large, if one method is better than another, it might as well be employed.

In order to find the most effective method to use, a simulation study to investigate the estimation performance of both methods was conducted. The details of this are given in Section 5.1. In summary, the study found that the intuitive solution discussed in Troughton and Godsill (1997) performed better in identifying the true model. Consequently, the practice of conditioning the data on the initial $s' = \max(k, k^*)$ data points and using the same $n' = N - s'$ data points in each of the likelihood calculations will be employed throughout this thesis.
Chapter 3

Univariate Models

This chapter contains a Bayesian analysis of the posterior distributions for the univariate STAR, GARCH and STAR-GARCH models, and details the design of separate hybrid MCMC estimation algorithms. The analysis of each model includes a definition of the structure of the prior distribution and the derivation of the likelihood function. The joint posterior distribution and the conditional posterior distributions required for use in the MCMC estimation algorithms are then derived. Our MCMC estimation procedures will employ a mixture of the MCMC algorithms discussed in Chapter 2.

3.1 STAR Model

Suppose we have an observed data set of length $N$, denoted by $x_{1:N} = \{x_1, \ldots, x_N\}^T$. An objective of analysing this data is to fit a model $M_k$ of order $k$, for some $k \in \mathcal{K} = \{1, \ldots, k_{\text{max}}\}$. We shall use $\mathcal{K}$ to denote the set of model orders that are considered to be potentially appropriate.

The initial value vector $x_0 = \{x_{1-k}, \ldots, x_0\}^T$ of the process is an unknown parameter vector that can be estimated. In practice, however, as the initial condition parameter vector provides little information of interest to a practitioner, it is often ignored during the estimation procedure. This is possible as the majority of data sets are relatively long, so their first few data points can be sacrificed to provide initial conditions for the model.

For the STAR model, the length of the initial condition vector is a function of the model order $k$ and the delay parameter $d$ of the transition function. By setting $s = \max(k, d)$, the initial condition of the model can be represented by a vector, $\{x_1, \ldots, x_s\}^T$. Therefore, the effective data will be the truncated data set $x_{s+1:N} = \{x_{s+1}, \ldots, x_N\}^T$. The length of the effective data set is then $n = N - s$. 

“He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may cast.”

- Leonardo da Vinci
In the following, whenever no subscript is specified for the data variable, it is assumed to be the effective data. That is, \( x = x_{s+1,N} \).

Recall from (2.1.1) that the STAR model of order \( k \) is represented by

\[
x_t = \phi_{1,0} + \phi_{1,1}x_{t-1} + \cdots + \phi_{1,k}x_{t-k} + \\
(\phi_{2,0} + \phi_{2,1}x_{t-1} + \cdots + \phi_{2,k}x_{t-k})F_t + \varepsilon_t,
\]

where \( \varepsilon_t \) is the error term with \( \varepsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,\sigma^2) \), \( \phi_{i,j} \in \mathbb{R} \) are the autoregressive parameters, and \( F_t \) is the transition function.

The transition function used is the scale free logistic function defined in (2.1.3). This function has the form,

\[
F_t(s_t,d,\gamma,c) = \frac{1}{1 + \exp\left(\frac{\gamma S(s_t)(s_t - d - c)}{S(s_t)}\right)}.
\]

For the logistic function described in (3.1.2), \( s_{t-d} \in \mathbb{R} \) is the transition variable. Here \( d \) is referred to as the delay parameter. The set of values considered for the delay parameters are \( d \in \mathcal{D} = \{1,\ldots,d_{\text{max}}\} \).

The transition variable \( s_{t-d} \) is sometimes set equal to the lagged endogenous variable \( x_{t-d} \). The parameters \( \gamma \in \mathbb{R}^+ \) and \( c \in \mathbb{R} \) are referred to as the smoothing parameter and the location parameter, respectively. The standard deviation of the transition variable \( S(s_t) \) is simply the sample standard deviation of the complete transition variable vector.

The expression for the LSTAR model given in (3.1.1) may be condensed using vector matrix notation as follows:

\[
x = Z_1\phi + \varepsilon,
\]

where \( Z_1 \in \mathbb{R}^{n \times 2(k+1)} \) is the design matrix, \( \phi \in \mathbb{R}^{2(k+1)} \) is the coefficient vector and \( \varepsilon \in \mathbb{R}^n \) is a vector of error terms. These vectors and matrices have the forms:

\[
Z_1 = \begin{bmatrix}
1 & x_s & \cdots & x_{s+1-k} & F_{s+1} & x_{s}F_{s+1} & \cdots & x_{s+1-k}F_{s+1} \\
1 & x_{s+1} & \cdots & x_{s+2-k} & F_{s+2} & x_{s+1}F_{s+2} & \cdots & x_{s+2-k}F_{s+2} \\
& \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_{N-1} & \cdots & x_{N-k} & F_N & x_{N-1}F_N & \cdots & x_{N-k}F_N
\end{bmatrix}
\]

\[
\phi = [\phi_{1,0}, \phi_{1,1}, \ldots, \phi_{1,k}, \phi_{2,0}, \phi_{2,1}, \ldots, \phi_{2,k}]^T
\]

\[
\varepsilon = [\varepsilon_{s+1}, \ldots, \varepsilon_N].
\]

In summary, the collection of parameters that we wish to estimate is \( \Theta \). The elements of \( \Theta \) that are of most interest are the model index \( \mathcal{M}_k \), which is based upon the model order \( k \); the coefficient
parameter $\phi$; the variance $\sigma^2$ of the error; and the transition function parameters $d, \gamma, \text{ and } c$. Therefore, we may write $\Theta = (\mathcal{M}_k, \phi, \sigma^2, d, \gamma, c)$. A negative subscript is used to denote the situation in which the corresponding parameter has been removed from the parameter vector $\Theta$. For example, $\Theta_{-\mathcal{M}_k} = (\phi, \sigma^2, d, \gamma, c)$.

The proportional relationship for the posterior distribution was explained in Chapter 2 and detailed in (2.2.11). Applying the parameters from the STAR model to this relationship results in the expression

$$p(\Theta | x) \propto p(x | \Theta) p(\Theta). \quad (3.1.3)$$

The joint prior distribution $p(\Theta) = p(M_k, \phi, \sigma^2, d, \gamma, c)$ shown in equation (3.1.3) is complicated by the differing domains of its parameters. Consequently, the prior distribution needs to be broken down and some assumptions made about the relationships between its parameters.

### 3.1.1 Prior Distribution

To simplify the analysis of the posterior distribution of the STAR model, we shall break the joint prior distribution $p(\Theta) = p(M_k, \phi, \sigma^2, d, \gamma, c)$ down into the following hierarchical structure:

$$p(\Theta) = p(M_k, \phi, \sigma^2, d, \gamma, c)$$

$$= p(M_k, \phi, \sigma^2) p(d, \gamma, c) \quad (3.1.4)$$

$$= p(\phi | M_k) p(M_k) p(\sigma^2) p(d) p(\gamma) p(c).$$

In the derivation of the hierarchical structure of the prior distribution in (3.1.4), it has been assumed that $\sigma^2, d, \gamma, \text{ and } c$ are all independent. The individual prior distributions for each parameter in (3.1.4) must be flexible enough to accommodate a range of beliefs about the corresponding parameter as well as to ensure that the posterior distribution is in a tractable form. In an attempt to meet both of these requirements, we assign the following individual prior distributions to each of the parameters:

$$p(\phi | M_k) \sim N_{(2k+2)}(0, \Sigma_{\phi})$$

$$p(\sigma^2) \sim IG(\alpha_0, \beta_0)$$

$$p(M_k) \propto \left[ \frac{\Lambda_k^{\gamma_k} \Gamma_k}{k!} \right] \mathbb{1}_K(k)$$

$$p(d) \propto \mathbb{1}_D(d)$$

$$p(\gamma) \sim G(\alpha_\gamma, \beta_\gamma)$$

$$p(c) \sim N(\mu_c, \sigma_c^2).$$

(3.1.5)
The function \( I_J(j) \) in equation (3.1.5) denotes an indicator function that takes the value 1 if \( j \in J \), and 0 otherwise. The quantities \( (\alpha_0, \beta_0, \Lambda_k, \tau_k, \alpha_\gamma, \beta_\gamma, \mu_c, \sigma_c^2) \) in the prior distributions and the prior covariance matrix \( \Sigma_\phi \) are assumed fixed and known. They are chosen to reflect the degree of prior belief about their respective parameters.

The prior \( p(M_k) \) for \( M_k \) is proportional to a Poisson distribution raised to a power. The reason for raising the Poisson distribution to a power is to add some additional flexibility into the prior distribution. By setting both \( \tau_k \) and \( \Lambda_k \) less than one, the prior distribution \( p(M_k) \) is a function of decreasing density as \( k \) increases. As \( \tau_k \) approaches zero, this function approaches the discrete uniform distribution.

### 3.1.2 Likelihood Function

The likelihood function \( p(x|\Theta) \) for the model arises from a random variable transformation from \( \varepsilon \) to \( x \). If the individual elements of \( \varepsilon \) are distributed such that \( \varepsilon_t \sim N(0, \sigma^2) \) and \( \varepsilon = Z_1 \phi - x \), then \( x \sim N_n(Z_1 \phi, \sigma^2 I_n) \). Here \( I_n \) represents an \( n \times n \) identity matrix. The likelihood function is therefore

\[
p(x|\Theta) = [2\pi]^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2}(x - Z_1 \phi)^T (\sigma^2 I_n)^{-1} (x - Z_1 \phi) \right]. \tag{3.1.6}
\]

### 3.1.3 Posterior Distribution

By combining the prior distributions in equation (3.1.5) with the likelihood function in equation (3.1.6), we arrive at the following joint posterior distribution:

\[
p(\Theta|x) \propto p(x|M_k, \phi, \sigma^2, d, \gamma, c) p(M_k) p(\phi) p(\sigma^2) p(d) p(\gamma) p(c) \\
\propto p(x|M_k, \phi, \sigma^2, d, \gamma, c) p(\phi|M_k) p(\sigma^2|M_k) p(d) p(\gamma) p(c) \\
\propto [2\pi]^{-\frac{2(d+1)}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2}(x - Z_1 \phi)^T (\sigma^2 I_n)^{-1} (x - Z_1 \phi) \right] \times \\
[2\pi]^{-\frac{2(d+1)}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \phi^T \Sigma_\phi^{-1} \phi \right] \times \\
\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} (\sigma^2)^{-\alpha_0-1} \exp \left[ -\frac{\beta_0}{\sigma^2} \right] \left( \frac{\Lambda_k}{k!} \right)^{\tau_k} \propto \\
\frac{\beta_\gamma^{\alpha_\gamma}}{\Gamma(\alpha_\gamma)} (\sigma_\gamma^2)^{-\alpha_\gamma-1} \exp \left[ -\beta_\gamma \right] \left( \frac{\mu_c}{2\sigma_c^2} \right)^{\frac{\tau_k}{2}} \propto \\
\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\beta_\gamma^{\alpha_\gamma}}{\Gamma(\alpha_\gamma)} \frac{\beta_0}{\sigma^2} \frac{\beta_\gamma}{\sigma_\gamma^2} \exp \left[ -\frac{(c - \mu_c)^2}{2\sigma_c^2} \right]. \tag{3.1.7}
\]

The expression for the joint posterior in equation (3.1.7) is of a non-standard form and, consequently,
we are unable to sample from this distribution directly. In order to draw samples from the joint posterior distribution, a custom MCMC algorithm is required. We employ an algorithm that follows the structure of a Gibbs sampler. The algorithm has an embedded RJMCMC step to account for the change in the dimension of the coefficient vector, as well as an embedded MTM step for simulation of the implicit parameters, $\gamma$ and $c$. The MTM algorithm is employed in an attempt to improve the acceptance rate for the simulation of the smoothing parameter, $\gamma$.

The execution of a Gibbs sampler algorithm requires conditional posterior distributions for each of the parameters. The conditional posterior distributions required for the algorithm are derived in detail in Appendix A.1. We summarise these results in equation (3.1.8). For each of the parameters, we have

$$
p(M_k, \phi|x, \Theta_{-M_k, \phi}) \propto [2\pi]^{-\frac{k}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (x - Z_1 \phi)^T (\sigma^2 I_n)^{-1} (x - Z_1 \phi) \right] 
$$

$$
p(\phi|x, \Theta_{-\phi}) \sim \mathcal{N}_d(\hat{\phi}, C_\phi)
$$

$$
p(\sigma^2|x, \Theta_{-\sigma^2}) \sim IG(\alpha_k, \beta_k)
$$

$$
p(d|x, \Theta_{-d}) \propto \exp \left[-\frac{1}{2\sigma^2} (x - Z_1 \phi)^T (x - Z_1 \phi) \right] \| \mathcal{D}(d) \]

$$
p(\gamma,c|x, \Theta_{-(\gamma,c)}) \propto \gamma^{\alpha_\gamma - 1} \exp \left[-\frac{1}{2\sigma^2} (x - Z_1 \phi)^T (x - Z_1 \phi) - \beta_\gamma - \left(\frac{c - \mu_c}{2\sigma^2_c}\right)^2 \right],
$$

where

$$
\hat{\phi} = \frac{C_\phi Z_1^T x}{\sigma^2},
$$

$$
C_\phi = \left[ \frac{Z_1^T Z_1}{\sigma^2} + \Sigma_\phi^{-1} \right]^{-1},
$$

$$
\alpha_k = \alpha_0 + \frac{n}{2},
$$

$$
\beta_k = \beta_0 + \frac{1}{2} (x - Z_1 \phi)^T (x - Z_1 \phi).
$$

The prior hyper-parameters $(\Sigma_\phi, \alpha_0, \beta_0, \alpha_\gamma, \beta_\gamma, \mu_c, \sigma^2_c)$ are assumed to be fixed and known.

### 3.1.4 Posterior Simulator

As previously mentioned, the posterior simulator will be structured as a Gibbs sampler with embedded RJMCMC and MTM steps for the joint conditional distributions of $(M_k, \phi)$ and $(\gamma, c)$, respectively.

#### STAR - Main Algorithm

1. Initialise $(M_0^{(0)}, \phi^{(0)}, \sigma^2(0), d^{(0)}, \gamma^{(0)}, c^{(0)})$ either randomly or deterministically and set $i = 1$. 

42
2. At the \( i \)th iteration, execute the following steps employing the conditional posterior distributions obtained in equation (3.1.8):

(a) (i) **RJ\textsubscript{M} Step** - Simulate the STAR model order and coefficient parameter vector using

\[
\begin{align*}
(M_k^{(i)}, \phi^{(i)}) & \sim p(M_k, \phi|x, \sigma^2\gamma, \epsilon) \\
\end{align*}
\]

(ii) If the proposal in step 2(a)(i) is rejected, simulate the coefficient parameter vector using

\[
\phi^{(i)} \sim p(\phi|M_k^{(i)}, \sigma^2\gamma, \epsilon)
\]

(b) Simulate the variance of the error and delay parameters using

\[
\begin{align*}
\sigma^{2(i)} & \sim p(\sigma^2|x, M_k^{(i)}, \phi^{(i)}) \\
d^{(i)} & \sim p(d|M_k^{(i)}, \phi^{(i)}, \sigma^{2(i)}, \gamma^{(i)}, \epsilon)
\end{align*}
\]

(c) **MTM\textsubscript{\gamma,c} Step** - Simulate the smoothing and location parameters of the transition function using

\[
(\gamma, c)^{(i)} \sim p(\gamma, c|x, M_k^{(i)}, \phi^{(i)}, \sigma^{2(i)})
\]

3. Set \( i = i + 1 \). If all the MCMC chains have converged, STOP, else, return to step 2.

There are two steps of the main algorithm for which further elaboration is required. These are the RJ\textsubscript{M} Step and the MTM\textsubscript{\gamma,c} Step. The following sections describe their algorithms in detail.

**Reversible Jump - RJ\textsubscript{M} Step**

The length of the coefficient parameter \( \phi \) is \( 2(k + 1) \). Therefore, as \( k \) changes, the dimensionality of \( \phi \) changes and, for this reason, an RJMCMC (Green 1995) algorithm is required at this step. If the main algorithm is currently at the beginning of the \( i \)th iteration, the most recently simulated values for the model order and coefficient parameter will be \( k^{(i-1)} \) and \( \phi^{(i-1)} \), respectively.

The jump to a new model order \( k^* \) arises from a proposal or jumping distribution \( j(M_{k^*}|M_k^{(i-1)}) \). In addition, a proposal for the coefficient parameter \( \phi \) is taken from a pre-defined proposal distribution.

Recall that in Section 2.2.4, we simplified the full acceptance probability formula as reported in Green (1995). The formula shown in equation (2.2.18) is recalled here:

\[
\begin{align*}
r & = \min \left( 1, \frac{\pi(\theta^{*}, m^*|x)q(\theta^{*}, m, m^*)}{\pi(\theta, m|x)q(\theta, m^*, m^*)} \right)
\end{align*}
\]

(3.1.9)
The target distribution within the acceptance probability in equation (3.1.9) is \( \pi(\theta, m|x) \). For our application, this target distribution is the joint conditional posterior distribution for \((M_k, \phi)\). Substituting this notation into the acceptance probability in equation (3.1.9) yields

\[
 r_{(M_k, \phi^*)} = \min \left( 1, \frac{\pi(M_k^*, \phi^*|x, \Theta_{-(M_k, \phi)}) q(\phi^*|M_k^*, \phi^*)}{\pi(M_k, \phi|x, \Theta_{-(M_k, \phi)}) q(\phi^*|\phi, M_k^*, \phi^*)} \right). \tag{3.1.10}
\]

Substitution of the candidate model order \( k^* \) and proposed coefficient parameter \( \phi^* \), together with the current values \( k(i-1) \) and \( \phi(i-1) \), directly into the acceptance probability formula in equation (3.1.10) will produce a Markov chain that converges to the desired target density.

In practice, when dealing with large data sets, one often encounters computational issues associated with the calculation of the acceptance probability. These computational issues usually arise as a result of arithmetic overflow and underflow. For a relatively large data set length \( N \), the calculation of the likelihood function involves the multiplication of close to \( N \) terms. If each of these terms is either relatively large or small, the product can be a very large or very small number, respectively. If the result is either larger or smaller than a predetermined quantity, the software will declare the result to be equal to infinity or zero, respectively. This then leads to problems in the acceptance probability formula caused by an attempt to divide infinity by infinity or zero by zero.

Some computational issues can be resolved by changing the order in which calculations are made or by taking the logarithm of the acceptance probability. Moreover, simplification of the acceptance probability through the Candidate’s Identity (Besag 1989), together with a careful choice of the proposal distribution for the coefficient parameter will aid in avoiding computational issues. The careful manipulation of the acceptance probability from equation (3.1.10) is shown in Appendix A.2.1. The results are summarised in equation (3.1.11) below:

\[
 r_{(M_k, \phi^*)} = \min \left( 1, \frac{|\Sigma\phi^*|^{-\frac{1}{2}} C_{\phi^*} \exp \left[ -\frac{1}{2} \left( \frac{x^T x}{\sigma^2} - \left( \phi^* \right)^T C_{\phi}^{-1} \phi^* \right) \right]}{|\Sigma\phi|^{-\frac{1}{2}} C_{\phi} \exp \left[ -\frac{1}{2} \left( \frac{x^T x}{\sigma^2} - \phi^T C_{\phi}^{-1} \phi \right) \right]} \right)^{\frac{1}{2}}, \tag{3.1.11}
\]

Equation (3.1.11) shows the acceptance probability for the joint conditional posterior distribution \( p(M_k, \phi|x, \Theta_{-(M_k, \phi)}) \). Interestingly, however, the acceptance probability in equation (3.1.11) is not a function of \( \phi \) and, therefore, does not actually require the simulation of a random \( \phi^* \). This has been previously noted by Stark, Fitzgerald, and Hladky (1997). This means that, in the event that a move from the current model \( k(i-1) \) to the proposed model \( k^* \) is rejected, resources are not wasted in the simulation of a vector that is not required.
Section 3.1: STAR Model

As discussed in Section 5.1, the model order impacts upon the amount of data that is available for use in the acceptance probability formula. Consistently with the results of Section 5.1, as well as the practice described in Troughton and Godsill (1997), the likelihood calculations are conditional on the first \( s^* = \max(k, k^*, d) \) data points. The data that should be used in the calculation is therefore \( x_{(s^*+1):N} \), which is a vector of length \( N - s^* = n^* \). The calculations of \( \hat{\phi} \) and \( C_{\phi} \) should also take into account the new effective length of the data.

The overall structure of the \( RJ_{\mathcal{M}} \) Step can now be described.

**RJ\( \mathcal{M} \) Algorithm**

1. The current state is \( \left( \mathcal{M}^{(i-1)}_k, \phi^{(i-1)}, \sigma^2(i-1), d(i-1), \gamma(i-1), c(i-1) \right) \).

2. Propose a candidate model index \( \mathcal{M}_{k^*} \) from \( j(M_{k^*}|M_k) \) using the discretised Laplacian shown in equation (2.2.17).

3. Calculate the acceptance probability \( r(\mathcal{M}_k, US) \) per equation (3.1.11).

4. Simulate \( u \sim U(0, 1) \):
   
   (a) If \( r(\mathcal{M}_k, US) > u \), then:
   
   (i) Accept the candidate model index \( \mathcal{M}_{k^*} \), and set \( \mathcal{M}^{(i)}_k = \mathcal{M}_{k^*} \).
   
   (ii) Simulate the coefficient parameter vector using the full condition posterior distribution:
   
   \[
   \phi^{(i)} \sim p \left( \phi|\mathbf{x}, \mathcal{M}^{(i)}_k, \sigma^2(i-1), d(i-1), \gamma(i-1), c(i-1) \right).
   \]

   (b) If \( r(\mathcal{M}_k, US) < u \), then:

   (i) Reject the candidate model index \( \mathcal{M}_{k^*} \), and set \( \mathcal{M}^{(i)}_k = \mathcal{M}^{(i-1)}_k \).

   (ii) Simulate the coefficient parameter vector using the full conditional posterior distribution per step 2(a)(ii) of the STAR - Main Algorithm.

**Multiple-Try Metropolis - MTM\((\gamma,c)\) Step**

The embedded MTM step is used to simulate the transition function’s smoothing and location parameters \( \gamma \) and \( c \), respectively. These parameters can sometimes be difficult to estimate using the standard Metropolis-Hastings algorithm as their initial values impact upon both the acceptance rate of the algorithm and the speed of convergence. With this in mind, an embedded MTM step can be employed in order to improve the acceptance rate and the speed of convergence of the algorithm.

The MTM algorithm requires proposal distributions for \( \gamma \) and \( c \). As \( \gamma \) is a strictly positive parameter, a Gamma distribution will be used, centred at the most recently simulated value. The proposal distri-
bution for \( c \) will be a normal distribution, located at the most recently simulated value. Specifically, the proposal distributions are as follows:

\[
q \left( \gamma^* | \gamma^{(i-1)} \right) \sim \mathcal{G} \left( \frac{\left( \gamma^{(i-1)} \right)^2}{\Delta_{\gamma}}, \frac{\gamma^{(i-1)}}{\Delta_{\gamma}} \right)
\]

\[
q \left( c^* | c^{(i-1)} \right) \sim \mathcal{N} \left( c^{(i-1)}, \Delta_c \right).
\]

The variances \( \Delta_{\gamma} \) and \( \Delta_c \) of the proposal distributions specified in equation (3.1.12) are used to tune the acceptance rate of the algorithm. Adjustments are made to these tuning parameters with the aim of achieving an acceptance rate of somewhere between 10% and 50%.

The implementation of the MTM algorithm requires the definition of an arbitrary function \( w(\theta, \psi) \) which satisfies \( w(\theta, \psi) > 0 \), and the selection of the number \( n_{\text{MTM}} \) of “Try’s” that will be performed. The function \( w(\theta, \psi) \) employed in this thesis has the form

\[
w(\theta, \psi) = \pi(\theta) q(\psi | \theta),
\]

where \( \pi(\theta) \) is the target distribution and \( q(\psi | \theta) \) is the density of the current value, given the proposed value from the proposal distribution. This function leads to what Liu, Liang, and Wong (2000) refer to as the MTM-I scheme.

In Section 2.2.3, the acceptance probability for an MTM algorithm was defined in equation (2.2.13) by

\[
r = \min \left( 1, \frac{\pi(\theta^*) q(\theta^{(i-1)} | \theta^*) p_\psi}{\pi(\theta^{(i-1)}) q(\theta^* | \theta^{(i-1)}) p_\theta} \right),
\]

where

\[
p_\theta = \frac{w(\theta^*, \theta^{(i-1)})}{\sum_{j=1}^{n_{\text{MTM}}} w(\theta_j^*, \theta^{(i-1)})}, \quad p_\psi = \frac{w(\theta^{(i-1)}, \theta^*)}{\sum_{j=1}^{n_{\text{MTM}}} w(\psi_j^*, \theta^*)}.
\]

Combining the information in equations (3.1.14) and (3.1.15) with the form of the function given in equation (3.1.13) results in

\[
\frac{\pi(\theta^*) q(\theta^{(i-1)} | \theta^*) p_\psi}{\pi(\theta^{(i-1)}) q(\theta^* | \theta^{(i-1)}) p_\theta} = \frac{\pi(\theta^*) q(\theta^{(i-1)} | \theta^*)}{\sum_{j=1}^{n_{\text{MTM}}} w(\theta_j^*, \theta^*)} \frac{w(\theta^{(i-1)}, \theta^*)}{\sum_{j=1}^{n_{\text{MTM}}} w(\theta_j^*, \theta^{(i-1)})}
\]

\[
= \frac{\pi(\theta^*) q(\theta^{(i-1)} | \theta^*)}{\pi(\theta^{(i-1)}) q(\theta^* | \theta^{(i-1)})} \frac{\pi(\theta^{(i-1)}) q(\theta^* | \theta^{(i-1)})}{\pi(\theta^*) q(\theta^{(i-1)} | \theta^*)}
\]

\[
= \frac{\pi(\theta^*) q(\theta^{(i-1)} | \theta^*)}{\pi(\theta^{(i-1)}) q(\theta^* | \theta^{(i-1)})}
\]
\[
\text{MTM} = \frac{\sum_{j=1}^{n_{\text{MTM}}} \pi (\theta_j^*) q (\theta_j^{(i-1)}) \theta_j^*}{\sum_{j=1}^{n_{\text{MTM}}} \pi (\psi_j^*) q (\theta_j^*|\psi_j^*)}.
\] (3.1.16)

If we begin with the simplified form of the MTM acceptance probability expressed in equation (3.1.16) and then substitute in the target and proposal distributions, we obtain the following expression for the acceptance probability:

\[
r_{(\gamma,c,US)}(\gamma,c) = \min \left( 1, \left[ \sum_{j=1}^{n_{\text{MTM}}} (\gamma_j^*)^{\alpha_j-1} \exp \left[ -\frac{1}{2\sigma^2} (x - Z_{1(j)}^\phi)^T (x - Z_{1(j)}^\phi) - \beta_j \gamma_j^* - \frac{(c_j^* - \mu_c)^2}{2\sigma_c^2} \right] \right] \right)
\]

We simplify the expression for the acceptance probability given in equation (3.1.17) in Appendix A.2.2. The structure of the MTM(\gamma,c) Step can then be described as follows.

**MTM(\gamma,c) Algorithm**

1. The current state is \( (M^{(i)}_k, \phi^{(i)}, \sigma^{2(i)}, d^{(i)}, \gamma^{(i-1)}, c^{(i-1)}) \).

2. Draw \( n_{\text{MTM}} \) independent trial proposals \( \{\gamma_1^*, \ldots, \gamma_{n_{\text{MTM}}}^*\} \) and \( \{c_1^*, \ldots, c_{n_{\text{MTM}}}^*\} \) from the proposal distributions \( q (\gamma^*|\gamma^{(i-1)}) \) and \( q (c^*|c^{(i-1)}) \), respectively. The proposal distributions are defined in equation (3.1.12). Form the set \( \{\theta_1^*, \ldots, \theta_{n_{\text{MTM}}}^*\} \) of trials.

3. Select a point \( \theta^* = (\gamma^*, c^*) \) from the set \( \{\theta_1^*, \ldots, \theta_{n_{\text{MTM}}}^*\} \), with probability

\[
p_{\theta} = \frac{w (\theta^*, \theta^{(i-1)})}{\sum_{j=1}^{n_{\text{MTM}}} w (\theta_j^*, \theta^{(i-1)})},
\]
where $\theta^{(i-1)} = (\gamma^{(i-1)}, c^{(i-1)})$.

4. Draw $n_{MTM} - 1$ realisations $\{\gamma_1', \ldots, \gamma_{n_{MTM}-1}'\}$ and $\{c_1', \ldots, c_{n_{MTM}-1}'\}$ from the proposal distributions $q(\gamma' | \gamma^*)$ and $q(c' | c^*)$, respectively. The proposal distributions are those defined in equation (3.1.12). Form the set $\psi_1 = (\gamma_1', c_1')$, $\ldots$, $\psi_{n_{MTM}-1} = (\gamma_{n_{MTM}-1}', c_{n_{MTM}-1}')$ and set $\psi_{n_{MTM}} = \theta^{(i-1)}$.

5. Define

$$p_{\psi} = \frac{w(\theta^{(i-1)}, \theta^*)}{\sum_{j=1}^{n_{MTM}} w(\psi_j, \theta^*)}.$$  

6. Calculate the acceptance probability $r_{(\gamma,c,US)}$ per equation (3.1.17).

7. Simulate $u \sim U(0, 1)$ using the following steps:

   (a) If $r_{(\gamma,c,US)} > u$, then:
      (i) Accept the proposed smoothing parameter and set $\gamma^{(i)} = \gamma^*$.
      (ii) Accept the proposed location parameter and set $c^{(i)} = c^*$.

   (b) If $r_{(\gamma,c,US)} < u$, then:
      (i) Reject the proposed smoothing parameter and set $\gamma^{(i)} = \gamma^{(i-1)}$.
      (ii) Reject the proposed location parameter and set $c^{(i)} = c^{(i-1)}$.

### 3.2 Generalised Autoregressive Conditional Heteroskedasticity Models

We wish to apply similar techniques to those outlined for STAR models in Section 3.1 to fit a GARCH model $\mathcal{N}_j$ to our data set $x$.

We indexed our search space of models for the STAR model on the model order, but use a different approach for the GARCH model: rather than treat the model orders $l$ and $m$ as separate parameters for estimation, we combine them into a single model index. The models are indexed by $\mathcal{N}_j$ for $j \in \{1, \ldots, 10\}$, with ARCH and GARCH model orders as given in Table 3.1.

The maximum orders for the ARCH and GARCH parameters are arbitrarily set to a relatively low value of three. As observed by Bollerslev (1986), one advantage of the GARCH model is its ability to be applied to a data set using lower model orders than those required for a pure ARCH model. It is trivial to increase or decrease the range of model orders included in the model search.

Note that Table 3.1 includes the model index $\mathcal{N}_1$, where $l = 0$ and $m = 0$. This represents the situation in which the data is simply independent and identically distributed from a normal distribution. This
index will later allow the pure STAR model to be fitted as part of the STAR-GARCH estimation algorithm.

Recall from Chapter 2, equation (2.1.4) that the mathematical form of the GARCH model is

\[ x_t = \varepsilon_t \]
\[ \varepsilon_t = \sqrt{h_t} \eta_t \]  
\[ h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_l \varepsilon_{t-l}^2 + \beta_1 h_{t-1} + \cdots + \beta_m h_{t-m}. \]  

In equation (3.2.18), \( \varepsilon_t \sim D(0,1) \) is an error term, where \( D(0,1) \) is a distribution with mean 0 and variance 1.

The model from equation (3.2.18) may be rewritten using vector matrix notation as

\[ x = \varepsilon, \]
\[ \varepsilon = h^{1/2} \otimes \eta \]
\[ h = Z_2 \alpha, \]

where \( \varepsilon \in \mathbb{R}^n \) is a vector of error terms; \( h \in \mathbb{R}^n \) is a vector of conditional variance terms; \( \alpha \in \mathbb{R}^{1+m+l} \); \( Z_2 \in \mathbb{R}^{n \times 1+m+l} \) is the variance design matrix; \( \eta \in \mathbb{R}^n \) is a vector of independent identically distributed random variables from the error distribution \( D(0,1) \); \( n \) is the effective data length such that \( n = N - s \); and \( s = \max(l,m) \). These matrices and vectors have to following representations:

\[ \varepsilon = [\varepsilon_s, \varepsilon_{s+1}, \ldots, \varepsilon_N]^T \]
\[ h = [h_s, h_{s+1}, \ldots, h_N]^T \]
\[ h^{1/2} = [h_s^{1/2}, h_{s+1}^{1/2}, \ldots, h_N^{1/2}]^T \]
\[ \alpha = [\alpha_0, \alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_m]^T \]
\[ Z_2 = \begin{bmatrix}
1 & \varepsilon_s^2 & \ldots & \varepsilon_{s+1}^2 & h_s & \ldots & h_{s-m} \\
1 & \varepsilon_{s+1}^2 & \ldots & \varepsilon_{s+2}^2 & h_{s+1} & \ldots & h_{s+1-m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \varepsilon_{N-1}^2 & \ldots & \varepsilon_N^2 & h_{N-1} & \ldots & h_{N-m}
\end{bmatrix}. \]

Therefore, the parameters of the GARCH model that require estimation are the model index \( N_j \), for
$j \in \{1, \ldots, 10\}$, which takes into account the ARCH and GARCH orders $l$ and $m$, respectively, and the coefficient vector $\alpha = [\alpha_0, \alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_m]^T$.

We now recall the necessary conditions for stationarity of the GARCH model (Bollerslev 1986):

$$\sum_{i=1}^{l} \alpha_i + \sum_{j=1}^{m} \beta_j < 1$$

$$\alpha_i > 0, \beta_j > 0 \text{ for } i \in \{1, \ldots, l\} \text{ and } j \in \{1, \ldots, m\}.$$  \hspace{1cm} (3.2.19)

### 3.2.1 Prior Distribution

The prior distributions to be used in the analysis are

$$p(N_j) \propto \frac{1}{(l+1)^{\tau_l} (m+1)^{\tau_m}}$$

$$p(\alpha|N_j) \propto \mathbb{I}_A(\alpha).$$  \hspace{1cm} (3.2.20)

In the function $\mathbb{I}_A(\alpha)$, $\mathcal{A}$ represents the stationarity conditions as outlined in equation (3.2.19). This function is an indicator function, which takes the value one if the stationarity conditions are satisfied for the given value of $\alpha$, and otherwise takes the value zero. This prior distribution, previously used by Chen and So (2006) and Vermaak et al. (2004), ensures that the parameter estimates result in a stationary fitted model. When the enforcement of stationarity is not required, it is relatively simple to implement a change in this prior distribution.

The prior distribution for the model index is a discrete distribution designed to allow for some flexibility in determining prior knowledge about the model choice. The parameters $\tau_l$ and $\tau_m$ are used to refine the shape of the prior distribution. Setting $\tau_l = \tau_m = 0$ yields the discrete uniform distribution as a special case. Positive values of $\tau_l$ and $\tau_m$ allow for models of greater complexity to have lower prior probabilities.

### 3.2.2 Likelihood Function

A given model index $N_j$ has model orders $l$ and $m$ as shown in Table 3.1. The coefficient vector has the form, $\alpha = [\alpha_0, \alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_m]$. Under the assumption that $\eta_t \sim \mathcal{N}(0,1)$, and given the model index, the likelihood function is

$$p(x|\alpha, N) = [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \varepsilon^T H^{-1} \varepsilon \right],$$  \hspace{1cm} (3.2.21)

where $H$ is a diagonal matrix of the form $H = \text{diag}(h)$. 

---

50
3.2.3 Posterior Distribution

By combining the prior distribution from equation (3.2.20) with the likelihood function from equation (3.2.21), we obtain the following expression for the joint posterior distribution:

\[
p(\alpha, N_j|x) \propto p(x|\alpha, N_j) p(\alpha|N_j) p(N_j) \propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \mathbf{e}^T H^{-1} \mathbf{e} \right] \mathbb{I}_{A}(\alpha) \frac{1}{(l+1)^m (m+1)^m}. \tag{3.2.22}
\]

The posterior distribution shown in equation (3.2.22) is of a non-standard form: the model index \(N_j\) impacts upon the dimensionality of the estimation problem. Therefore, an RJMCMC algorithm is required for the estimation procedure.

When running the algorithm, the candidate model \(N_j^*\) and proposed coefficient parameter vector \(\alpha^*\) are rejected more often than they are accepted. If the algorithm is set up so that the only step performed is to simulate from \(p(\alpha, N_j|x)\), the chain will be highly correlated, with new values of \(\alpha\) only generated when model moves are accepted or if the candidate model is the current model. With this in mind, when moves to a new model index are rejected, the next step in the algorithm will be to simulate from the full conditional posterior distribution for \(\alpha\), that is \(p(\alpha|x, N_j)\). Therefore, the conditional posterior distributions that will be employed in the estimation procedure are:

\[
p(\alpha, N_j|x) \propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \mathbf{e}^T H^{-1} \mathbf{e} \right] \mathbb{I}_{A}(\alpha) \frac{1}{(l+1)^m (m+1)^m},
\]

\[
p(\alpha|x, N_j) \propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \mathbf{e}^T H^{-1} \mathbf{e} \right] \mathbb{I}_{A}(\alpha). \tag{3.2.23}
\]

3.2.4 Posterior Simulator

In order to simulate from the joint posterior distribution in equation (3.2.22), we have developed a custom MCMC algorithm as outlined below. It is constructed as a Gibbs sampler style of algorithm with an RJMCMC step (RJ\(N_j\) Step) for the joint distribution for \(N_j\) and \(\alpha\), and a Metropolis-Hastings step (MH\(\alpha\) Step) for the coefficient parameter vector.

GARCH - Main Algorithm

1. Initialise \((N_j^{(0)}, \alpha^{(0)})\) either randomly or deterministically and set \(i = 1\).

2. At the \(i\)th iteration, execute the following steps employing the posterior distributions shown in (3.2.23):
   (a) (i) RJ\(N_j\) Step - Simulate the GARCH model index and coefficient parameter vector with
the conditional posterior distribution using

\[
\left( \mathcal{N}^{(i)}_j, \alpha^{(i)} \right) \sim p(\mathcal{N}_j, \alpha|x).
\]

(ii) **MHα Step** - If the proposal in step 2(a)(i) is rejected, simulate the coefficient parameter vector from the full conditional posterior distribution

\[
\alpha^{(i)} \sim p(\alpha|x, \mathcal{N}^{(i)}_j).
\]

One of the key difficulties in implementing the above algorithm lies in choosing the proposal distribution for the coefficient parameter \( \alpha \). In the STAR algorithm, when attempting to move to a candidate model, the data can be used to find a suitable location parameter for the proposal distribution of the mean coefficient vector. Unfortunately, we are unable to use the data to determine a suitable location parameter for the proposal distribution of the variance coefficient vector in the GARCH model.

One remedy for this problem is to simply prescribe a gamma distribution for \( \alpha_0 \) and uniform distributions over the range \((0,1)\) for \( \alpha_i \) and \( \beta_j \) for \( i \in \{1, \ldots, l\} \) and \( j \in \{1, \ldots, m\} \), respectively. This method was initially implemented, but, as expected, performed poorly. The reason was that proposals were often made in areas of low posterior probabilities, leading to very low acceptance rates and poor mixing.

Another option, suggested by Vrontos, Dellaportas, and Politis (2000) for use in an RJMCMC algorithm for GARCH models, is to set up a pilot run to determine a set of reasonable location parameters. We shall follow this approach. Our pilot run will simply apply the GARCH model algorithm with fixed GARCH model indices for each of the models that we wish to include in our search. The initial values for the pilot run can be randomly selected, but, for each of the models considered, we shall use the MLE of the respective coefficient parameter vector as the set of initial values. This choice ensures a reasonable starting location and reasonably quick convergence. One advantage of using a pilot run is that the variance of the posterior distribution for each model in the pilot run provides some guidance as to what the variance of the proposal distributions should be for the full RJMCMC run of the algorithm.

**Reversible Jump - RJN Step**

As for the STAR model, we require an RJMCMC algorithm to handle the change in dimensionality resulting from jumps between models. If the main algorithm is currently at the beginning of the \( i \)th iteration, the most recently simulated values for the model index and coefficient parameter will be \( \mathcal{N}^{(i-1)}_j \) and \( \alpha^{(i-1)} \), respectively.
Recall that in Section 2.2.4, we simplified the full acceptance probability formula as reported in Green (1995). The formula given in equation (2.2.18) is rewritten here:

\[
\gamma = \min \left( 1, \frac{\pi(\theta^*, m^*, m) q(\theta|\theta^*, m, m^*)}{\pi(\theta, m|m) q(\theta^*|\theta, m, m^*)} \right). \tag{3.2.24}
\]

Applying the notation of the GARCH model’s parameters to the acceptance probability shown in equation (3.2.24) yields

\[
\gamma_{(N_j, \omega)} = \min \left( 1, \frac{p(\alpha^*, N_j^*|x) q(\alpha|\alpha^*, N_j, N_j^*)}{p(\alpha, N_j|x) q(\alpha^*|\alpha, N_j, N_j^*)} \right). \tag{3.2.25}
\]

The indexation of the GARCH models can be thought of as a kind of model nesting. Models with indices close to the current model will be of a similar level of complexity. Those models with a lower model index will tend to be simpler, whereas models with a higher model index will tend to be more complicated. With this in mind, the discretised Laplacian distribution given in equation (2.2.17) will be used as the jumping density. That is,

\[
\gamma_{(N_j^*, \omega)} \propto \exp \left[ -\Delta N_j |j^* - j| \right] \tag{3.2.26}
\]

for \( N_j^* \in \{1, \ldots, 10\} \). This choice of jumping distribution ensures that the jumping densities can be removed from the acceptance probability for making the model move, thereby maintaining the form of the acceptance probability shown in equation (3.2.25).

As previously discussed, the location parameters for the coefficient parameters will be determined by executing a pilot run of the algorithm before implementing the main algorithm. The proposal distribution for the coefficient parameter will therefore be

\[
q(\alpha^*|\alpha, N_j, N_j^*) \sim T_{\nu + m^*} \left( \hat{\alpha}_{N_j^*}, \nu, \Sigma_{\alpha_{N_j^*}} \right), \tag{3.2.27}
\]

where the means \( \hat{\alpha}_{N_j^*} \) of the proposal distributions are determined from the mean of the pilot run. The degrees of freedom parameter \( \nu \) is set to some fixed amount. Setting \( \nu = 5 \) lead to acceptable results. The scale matrix \( \Sigma_{\alpha_{N_j^*}} \) is determined by creating a diagonal matrix in which the diagonal entries correspond to the variances of the posterior distributions for the pilot run. The proposal parameters \( \hat{\alpha}_{N_j^*} \) and \( \Sigma_{\alpha_{N_j^*}} \) both have the subscript \( N_j^* \) to indicate that these parameters are different for each of the candidate models considered by the estimation algorithm.

Therefore, by taking the form of the joint posterior distribution shown in equation (3.2.22), together with the multivariate \( t \)-distribution defined in equation (3.2.27) as the proposal distribution, the ac-
The acceptance probability expression can be simplified to

\[
\begin{align*}
    r(N_j, UG) &= \min \left( 1, \left[ |H^*| |H|^{-\frac{1}{2}} \left( \left| \Sigma_{\alpha_{N_j}} \right| \left| \Sigma_{\alpha_{N_i}} \right|^{-1} \right) \right]^\frac{1}{2} \left( \frac{m + 1}{m^* + 1} \right)^{\frac{m}{2}} \times \right. \\
    & \quad \times \frac{\left[ 1 + \frac{1}{\nu} (\alpha^* - \hat{\alpha}_{N_j})^T \Sigma^{-1}_{\alpha_{N_j}} (\alpha^* - \hat{\alpha}_{N_j}) \right]^{\frac{\nu + 1 + l + m^*}{2}}}{\left[ 1 + \frac{1}{\nu} (\alpha - \hat{\alpha}_{N_j})^T \Sigma^{-1}_{\alpha_{N_j}} (\alpha - \hat{\alpha}_{N_j}) \right]^{\frac{\nu + 1 + l + m^*}{2}}} \\
    & \quad \times \left( \frac{\Gamma \left( \frac{\nu + 1 + l + m^*}{2} \right)}{\Gamma \left( \frac{\nu + 1 + l + m^*}{2} \right)} \right) \frac{1}{2^{\frac{\nu + l + m}{2}}} \exp \left[ -\frac{1}{2} \varepsilon^T \left( H^* - H \right) \frac{1}{\varepsilon} \right] \\
    & \quad \times \left( \frac{1}{2^{\frac{\nu + l + m}{2}}} \right). 
\end{align*}
\]

(3.2.28)

The details of the simplification of the acceptance probability shown in equation (3.2.28) are presented in Appendix B.1.1. We now outline the steps for the RJ\(_N\) Algorithm.

**RJ\(_N\) Algorithm**

1. The current state is \( (N_j^{(i)}, \alpha^{(i)}) \).

2. Propose a candidate model index \( N_j^{*} \) from \( j(N_j^*, N_j) \) using the discretised Laplacian shown in equation (3.2.26). Propose a new coefficient parameter \( \alpha^* \) from \( q(\alpha^* | \alpha, N_j, N_j^*) \) using the proposal distribution shown in equation (3.2.27).

3. Calculate the acceptance probability \( r(N_j, UG) \) per equation (3.2.28).

4. Simulate \( u \sim U(0, 1) \):

   (a) If \( r(N_j, UG) > u \), then:

   (i) Accept the candidate model index \( N_j^* \) and set \( N_j^{(i)} = N_j^* \).

   (ii) Accept the coefficient parameter vector \( \alpha^* \) and set \( \alpha^{(i)} = \alpha^* \).

   (b) If \( r(N_j, UG) < u \), then:

   (i) Reject the candidate model index \( N_j^* \) and proposed variance coefficient parameter \( \alpha^* \). Set \( N_j^{(i)} = N_j^{(i-1)} \) and \( \alpha^{(i)} = \alpha^{(i-1)} \).

   (ii) Using the distribution in equation (3.2.23), simulate the conditional variance coefficient parameter

   \[ \alpha^{(i)} \sim p(\alpha | x, N_j^{(i)}) \]

   using the Metropolis-Hastings algorithm described below (MH\(_\alpha\) Step).
Section 3.2: Generalised Autoregressive Conditional Heteroskedasticity Models

Metropolis-Hastings - MH\(\alpha\) Step

When the proposal to move from model \(N_j\) to \(N_{j'}\) is rejected, we simulate from the full conditional distribution for \(\alpha\), that is, \(p(\alpha|x,N_j)\). This distribution is not of a standard form and requires simulation using an embedded Metropolis-Hastings algorithm. Recall from Chapter 2 that the acceptance probability for a Metropolis-Hastings algorithm shown in equation (2.2.12) is

\[
r = \min \left( 1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^{(i-1)}|\theta^*)} \right).
\]

Employing the notation used for the GARCH model in the acceptance formula from equation (3.2.29) yields

\[
r_{(\alpha,UG)} = \min \left( 1, \frac{p(\alpha^*|x,N_{j'}) q(\alpha^{(i-1)}|\alpha^*,N_j)}{p(\alpha^{(i-1)}|x,N_j) q(\alpha^*|\alpha^{(i-1)},N_{j'})} \right). \tag{3.2.30}
\]

The proposal distribution for the coefficient parameter in equation (3.2.30) is defined similarly to the proposal distribution for the Reversible Jump step. However, rather than using the location parameter from the pilot run, the location parameter will be set to the most recently simulated value. Thus,

\[
q(\alpha^*|\alpha,N_j) \sim T_{l+m}(\alpha,\nu,\Sigma_{\alpha N_j}), \tag{3.2.31}
\]

where \(\alpha\) is the most recently simulated value, that is, \(\alpha = \alpha^{(i-1)}\). The degrees of freedom \(\nu\) and the scale matrix \(\Sigma_{\alpha N_j}\) will be defined following the proposal distribution for the Reversible Jump step.

Since the multivariate \(t\)-distribution prior given in equation (3.2.31) is a symmetric distribution, in that \(q(\alpha^*|\alpha,N_j) = q(\alpha|\alpha^*,N_j)\), the ratio of the proposal distributions will be one. This means that the acceptance probability in equation (3.2.30) will simplify to give the ratio of the full conditional posterior distributions shown in (3.2.23). The simplification procedure is shown in Appendix B.1.2, but we give the final expression for the acceptance probability here:

\[
r_{(\alpha,UG)} = \min \left( 1, \left[ |H^*||H^{-1}| \right]^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \varepsilon^T \left[ H^{*^{-1}} - H^{-1} \right] \varepsilon \right] \right). \tag{3.2.32}
\]

Now we have completely determined the acceptance probability, we can outline the steps of the Metropolis-Hastings algorithm, assuming that we are currently at the \(i\)th iteration.

**MH\(\alpha\) Algorithm**

1. The current state is \((N_j^{(i-1)}, \alpha^{(i-1)})\).

2. Propose a new coefficient parameter \(\alpha^*\) from \(q(\alpha^*|\alpha,N_j)\) using the proposal distribution shown
3. Calculate the acceptance probability $r_{\alpha, \mathcal{U}}$ per equation (3.2.32).

4. Simulate $u \sim \mathcal{U}(0, 1)$:

(a) If $r_{\alpha, \mathcal{U}} > u$, accept the proposed coefficient parameter $\alpha^*$ and set $\alpha^{(i)} = \alpha^*$.

(b) If $r_{\alpha, \mathcal{U}} < u$, reject the proposed coefficient parameter $\alpha^*$ and set $\alpha^{(i)} = \alpha^{(i-1)}$.

Other methods of proposing the coefficient parameter for both the Reversible Jump step and Metropolis-Hastings step were explored. One example that worked well was to split the proposal of the coefficient parameter up as $q(\alpha^* | \cdot) = q(\alpha^*_0 | \cdot) q(\alpha^*_{-\alpha_0} | \cdot)$. The proposal of $\alpha^*_0$ was performed using a gamma distribution, as this parameter is greater than zero and not bounded above. The vector $\alpha^*_{-\alpha_0}$ of the remaining elements had length $l^* + m^*$. The proposal of $\alpha^*_{-\alpha_0}$ was then performed using a Dirichlet distribution with $l^* + m^* + 1$ categories.

The randomly generated value from the additional category was ignored. This can be justified by thinking of the definition of the conditional covariance equation $h_t$ as having an extra term on the end made up of the extra parameter multiplied by zero. We used the Dirichlet distribution since the proportions summed to one. Therefore, ignoring the last value in the randomly generated vector ensured that $\alpha^*$ always met the stationarity conditions. While this method for proposing the variance coefficient parameter vector worked well, it failed to outperform the much simpler multivariate $t$-distribution that was finally settled upon.

### 3.3 STAR-GARCH Models

The STAR-GARCH model is a combination of the two models described in the previous two sections. This model is used to account for changes in the conditional mean and conditional variance. The model fitted for the conditional mean is denoted by $\mathcal{M}_k$, and the model for the conditional variance is denoted by $\mathcal{N}_j$. The conditional mean models are indexed directly by the model order $k$. However, the GARCH model orders are again combined into a single model index. This model index is shown in Table 3.2, as in Section 3.2.

Recall from equation (2.1.5), that the STAR($k$)-GARCH($l, m$) model may be represented as follows:

\[
\begin{align*}
    x_t &= \phi_{1,0} + \phi_{1,1}x_{t-1} + \cdots + \phi_{1,k}x_{t-k} + \\
         &\quad (\phi_{2,0} + \phi_{2,1}x_{t-1} + \cdots + \phi_{2,k}x_{t-k})F_t + \varepsilon_t \\
    \varepsilon_t &= \sqrt{h_t} \eta_t \\
    h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_l \varepsilon_{t-l}^2 + \beta_1 h_{t-1} + \cdots + \beta_m h_{t-m}.
    \end{align*}
\]  

(3.3.33)
In equation (3.3.33), $\eta_t \sim D(0, 1)$ is an error term and $D(0, 1)$ is a distribution with mean 0 and variance 1. The transition function $F_t$ may be any smooth function bounded by 0 and 1. This is often prescribed to be the logistic function or an exponential function.

The transition function used here will be the scale free logistic function of Gerlach and Chen (2008), earlier defined in equation (2.1.3). The form of the transition function used is

$$F_t(s_t, d, \gamma, c) = \frac{1}{1 + \exp \left[-\frac{\gamma}{S(s_t)}(s_t - d - c)\right]}.$$  

(3.3.34)

The parameters in (3.3.34) are the transition variable $s_t - d \in \mathbb{R}$, which is sometimes set as a lagged endogenous variable $x_{t-d}$, where $d \in D = \{1, \ldots, d_{\text{max}}\}$ is referred to as the delay parameter; the standard deviation of the transition variable $S(s_t)$; the smoothing parameter $\gamma \in \mathbb{R}^+$; and the location parameter $c \in \mathbb{R}$.

As in the previous sections, the model shown in equation (3.3.33) may be rewritten in vector matrix notation as follows:

$$\begin{align*}
x &= Z_1 \phi + \varepsilon, \\
\varepsilon &= h^{\frac{1}{2}} \odot \eta \\
h &= Z_2 \alpha,
\end{align*}$$

where $Z_1 \in \mathbb{R}^{n \times 2(k+1)}$ is the mean design matrix, $\phi \in \mathbb{R}^{2(k+1)}$ is the mean coefficient vector, $\varepsilon \in \mathbb{R}^n$ is a vector of error terms, $h \in \mathbb{R}^n$ is a vector of conditional variance terms, $\alpha \in \mathbb{R}^{1+m+l}$, and $Z_2 \in \mathbb{R}^{n \times 1+m+l}$ is the variance design matrix. These matrices and vectors have the following forms:

$$Z_1 = \begin{bmatrix}
1 & x_s & \ldots & x_{s+1-k} & F_{s+1} & x_sF_{s+1} & \ldots & x_{s+1-k}F_{s+1} \\
1 & x_{s+1} & \ldots & x_{s+2-k} & F_{s+2} & x_{s+1}F_{s+2} & \ldots & x_{s+2-k}F_{s+2} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{N-1} & \ldots & x_{N-k} & F_N & x_{N-1}F_N & \ldots & x_{N-k}F_N
\end{bmatrix}$$

$$\phi = [\phi_0, \phi_1, \ldots, \phi_{1,k}, \phi_{2,0}, \phi_{2,1}, \ldots, \phi_{2,k}]^T$$

$$\varepsilon = [\varepsilon_s, \varepsilon_{s+1}, \ldots, \varepsilon_N]^T$$

$$h = [h_s, h_{s+1}, \ldots, h_N]^T$$

$$h^{\frac{1}{2}} = [h^{\frac{1}{2}}_s, h^{\frac{1}{2}}_{s+1}, \ldots, h^{\frac{1}{2}}_N]^T$$

$$\alpha = [\alpha_0, \alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m]^T$$

$$Z_2 = \begin{bmatrix}
1 & \varepsilon_2^2 & \ldots & \varepsilon_{s+1-l}^2 & h_s & \ldots & h_{s-m} \\
1 & \varepsilon_{s+1}^2 & \ldots & \varepsilon_{s+2-l}^2 & h_{s+1} & \ldots & h_{s+1-m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \varepsilon_{N-1}^2 & \ldots & \varepsilon_{N-l}^2 & h_{N-1} & \ldots & h_{N-m}
\end{bmatrix}.$$
In summary, let the collection of parameters that we wish to estimate be $\Theta$. The elements of $\Theta$ that are of most interest are the conditional mean model index $M_k$; the mean coefficient parameter vector $\phi$; the transition function parameters $d$, $\gamma$, and $c$; the conditional variance model index $N_j$, which encompasses the ARCH and GARCH model order parameters $l$ and $m$; and the conditional variance coefficient vector $\alpha$. Therefore, $\Theta = (M_k, \phi, d, \gamma, c, N_j, \alpha)$.

The indexation of the conditional variance model orders is the same as presented in Section 3.2 and is repeated in Table 3.2 below.

<table>
<thead>
<tr>
<th>Model</th>
<th>$N_1$</th>
<th>$N_2$</th>
<th>$N_3$</th>
<th>$N_4$</th>
<th>$N_5$</th>
<th>$N_6$</th>
<th>$N_7$</th>
<th>$N_8$</th>
<th>$N_9$</th>
<th>$N_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3.2: Index of GARCH($l, m$) models.

### 3.3.1 Prior Distribution

The joint prior distribution $p(\Theta) = p(M_k, \phi, d, \gamma, c, N_j, \alpha)$ can be broken down into a hierarchical structure as follows:

$$p(\Theta) = p(M_k, \phi, d, \gamma, c, N_j, \alpha)$$
$$= p(M_k, \phi) p(d, \gamma, c) p(N_j, \alpha)$$
$$= p(\phi|M_k) p(M_k) p(d) p(\gamma) p(c) p(\alpha|N_j) p(N_j).$$

In the derivation of the hierarchical structure of the prior distribution in equation (3.3.36), it has been assumed that the transition function parameters $(d, \gamma, c)$, the conditional mean parameters $(M_k, \phi)$ and the conditional variance parameters $(N_j, \alpha)$ are independent of each other.

In an effort to ensure that the posterior distribution is in a tractable form, and to allow some flexibility in setting our prior beliefs about each parameter, the prior distribution from equation (3.3.36) will have the following individual prior distributions assigned to each of its parameters:

$$p(\phi|M_k) \sim N_{2k+2} (0, \Sigma_\phi)$$
$$p(M_k) \propto \left[ \frac{\Lambda_k}{k!} \right]^{\tau_k} \mathbb{I}_K (k)$$
$$p(d) \propto \mathbb{I}_D (d)$$
$$p(\gamma) \sim G (\alpha_\gamma, \beta_\gamma)$$
$$p(c) \sim N (\mu_c, \sigma_c^2)$$
$$p(\alpha|N_j) \propto \mathbb{I}_A (\alpha)$$
$$p(N_j) \propto \frac{1}{(l+1)^l (m+1)^m}.$$

58
In the above function definition, \( I_J(j) \) denotes an indicator function that takes the value 1 if \( j \in J \), and 0 otherwise. In the case of \( \alpha, A \) represents the required conditions on the variance equation parameters to ensure stationarity as given in equation (3.2.19). For the prior distributions in equation (3.3.37), the quantities \( (\Lambda, \alpha, \beta, \mu_c, \sigma_c^2, \pi_k, \tau_l, \tau_m) \), and the prior covariance matrix \( \Sigma_\phi \) are assumed fixed and known. These quantities are chosen to reflect the user’s prior beliefs about the parameters.

### 3.3.2 Likelihood Function

The likelihood function \( p(x|\Theta) \) for the model arises from a random variable transformation from \( ε \) to \( x \). By letting \( \mu_t = \phi_{1,0} + \phi_{1,1}x_{t-1} + \cdots + \phi_{1,k}x_{t-k} + (\phi_{2,0} + \phi_{2,1}x_{t-1} + \cdots + \phi_{2,k}x_{t-k})F_t \) be the conditional mean of the model, we now have \( x_t = \mu_t + \varepsilon_t \). As \( \varepsilon_t = \sqrt{\eta_t} \), this can be rewritten as \( x_t = \mu_t + \sqrt{\eta_t} \varepsilon_t \). For the case where \( \eta_t \sim \mathcal{N}(0,1) \), this leads to \( x_t \sim \mathcal{N}(\mu_t, h_t) \). The product of \( n \) normally distributed variables results in the following likelihood function:

\[
p(x|\Theta) = [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right], \tag{3.3.38}
\]

where \( H \) is the \( n \times n \) diagonal matrix with diagonal elements corresponding to \( h \) as defined in (3.3.35).

### 3.3.3 Posterior Distribution

Combining the prior distributions from equation (3.3.37) with the likelihood function from equation (3.3.38), we obtain the following joint posterior distribution:

\[
p(\Theta|x) \propto p(x|M_k, \phi, d, \gamma, c, N_j, \alpha) p(M_k, \phi, d, \gamma, c, N_j, \alpha) 
\propto p(x|M_k, \phi, d, \gamma, c, N_j, \alpha) p(\phi|M_k) p(d) p(\gamma) p(c) p(\alpha|N_j) p(N_j)
\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right] \times 
[2\pi]^{-\frac{k(k+1)}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \phi^T \Sigma_\phi^{-1} \phi \right] \left[ \frac{A_k}{k!} \right] I_{\mathcal{K}}(k) \times 
I_D(d) \Gamma(\alpha_c) \beta_{\gamma_c}^{\alpha_c-1} \exp \left[ -\beta_\gamma \left[ 2\pi \sigma_c^2 \right]^{-\frac{1}{2}} \exp \left[ -\frac{(c - \mu_c)^2}{2\sigma_c^2} \right] \right] I_A(\alpha) \times 1 
\frac{1}{(l+1)^{m}(m+1)^{m}} .
\]

The expression for the joint posterior is of non-standard form, and so cannot be sampled from directly. In order to simulate samples from the above joint posterior distribution, a custom RJMCMC algorithm is required. It is set up as a Gibbs sampler with separate Reversible Jump steps for the STAR model order, the GARCH model index and their respective coefficient parameter vectors. The implicit
parameters $\gamma$ and $c$ will be simulated using an embedded Metropolis-Hastings step within the Gibbs sampler.

Due to the complexity of the GARCH model, there are no stages in the Gibbs sampler for which the conditional posterior distribution takes a standard form. Other than the discrete conditional posterior distribution for the delay parameter $d$, each parameter must be simulated using embedded RJMCMC or Metropolis-Hastings algorithms. Therefore, minimal manipulation of the conditional posterior distributions will be performed. This yields:

\[
\begin{align*}
p(M_k, \phi|x, \Theta_{-(M_k, \phi)}) &\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right] \times \\
&\times [2\pi]^{-\frac{2(n-1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \phi^T \Sigma_{\phi}^{-1} \phi \right] I_n(k) \\
p(\phi|x, \Theta_{-\phi}) &\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right] \\
&\times [2\pi]^{-\frac{2(n-1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \phi^T \Sigma_{\phi}^{-1} \phi \right] \\
p(d|x, \Theta_{-d}) &\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right] I_n(d) \\
p(\gamma,c|x, \Theta_{-(\gamma,c)}) &\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right] \\
&\times \frac{\beta^2}{\Gamma(\alpha\gamma)} \gamma^{\alpha\gamma-1} \exp \left[ -\beta \gamma \right] \left[ 2\pi \sigma^2 \right]^{-\frac{1}{2}} \exp \left[ -\frac{(c - \mu)^2}{2\sigma^2} \right] \\
p(\alpha_j|x, \Theta_{-(\alpha_j)}) &\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right] \\
&\times \frac{1}{(l+1)^m (m+1)^m} I_m(\alpha) \\
p(\alpha|x, \Theta_{-(\alpha)}) &\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right] I_m(\alpha).
\end{align*}
\]

\[ (3.3.39) \]

### 3.3.4 Posterior Simulator

In order to simulate from the joint posterior distribution, a custom MCMC algorithm is required. We shall use a Gibbs style algorithm, employing the conditional distribution outlines given in equation (3.3.39). The main algorithm can be outlined as follows:

**STAR-GARCH - Main Algorithm**

1. Initialise \( (M_k^{(0)}, \phi^{(0)}, d^{(0)}, \gamma^{(0)}, c^{(0)}, \alpha_j^{(0)}, \alpha^{(0)}) \) either randomly or deterministically and set \( i = 1 \).

2. At the \( i \)th iteration, execute the following steps employing the conditional posterior distributions obtained in equation (3.3.39):
Section 3.3: STAR-GARCH Models

(a) (i) **RJ\(M\) Step** - Simulate the STAR model order and coefficient parameter vector using
\[ (M_k^{(i)}, \phi^{(i)}) \sim p(M_k, \phi|x, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)}, N_j^{(i-1)}, \alpha^{(i-1)}) \].

(ii) **MH\(\phi\) Step** - If the proposal in step 2(a)(i) is rejected, simulate the mean coefficient parameter vector using
\[ \phi^{(i)} \sim p(\phi|x, M_k^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)}, N_j^{(i-1)}, \alpha^{(i-1)}) \].

(b) Simulate the delay parameter \(d\) using
\[ d^{(i)} \sim p(d|x, M_k^{(i)}, \phi^{(i)}, \gamma^{(i-1)}, c^{(i-1)}, N_j^{(i-1)}, \alpha^{(i-1)}) \].

(c) **MH\((\gamma,c)\) Step** - Simulate the smoothing and location parameters for the transition function using
\[ (\gamma^{(i)}, c^{(i)}) \sim p(\gamma, c|x, M_k^{(i)}, \phi^{(i)}, d^{(i)}, N_j^{(i-1)}, \alpha^{(i-1)}) \].

(d) (i) **RJ\(N\) Step** - Simulate the GARCH model index and coefficient parameter vector using
\[ (N_j^{(i)}, \alpha^{(i)}) \sim p(N_j, \alpha|x, M_k^{(i)}, \phi^{(i)}, d^{(i)}, \gamma^{(i)}, N_j^{(i)}, \alpha^{(i)}) \].

(ii) **MH\(\alpha\) Step** - If the proposal in step 2(d)(i) is rejected, simulate the variance coefficient parameter vector from the full conditional posterior distribution:
\[ \alpha^{(i)} \sim p(\alpha|x, M_k^{(i)}, \phi^{(i)}, d^{(i)}, \gamma^{(i)}, c^{(i)}, N_j^{(i)}) \].

3. Set \(i = i + 1\). If all the MCMC chains have converged, STOP, else, move to step 2.

As shown in Section 3.1, the simulation of the implicit parameters \(\gamma\) and \(c\) for the transition function can be performed using an embedded MTM algorithm. For the STAR algorithm, this technique was found to improve the acceptance rate at this step. However, it comes at the cost of additional computational time when making proposals and calculating the acceptance probability during the MCMC run. The additional complexity of the STAR-GARCH model results in a more complicated simulation algorithm. With this in mind, the simpler Metropolis-Hastings algorithm can be successfully used to simulate from the joint conditional posterior distribution for \(\gamma\) and \(c\).

In the main algorithm above, there are sub-algorithms that require explanation. These are the Reversible Jump steps at 2(a)(i) and 2(d)(i), and the Metropolis-Hastings steps at 2(a)(ii), 2(d)(ii) and 2(c). They are outlined in detail below.
Reversible Jump Steps

The Reversible Jump steps are designed similarly to those in Sections 3.1.4 and 3.2.4. That is, upon failure to move to the candidate model, the relevant coefficient parameter vector will be simulated from the full conditional posterior distribution. This is to ensure that, within a particular model space, there is adequate mixing of the MCMC chains.

Recall from equation (2.2.18) that the general acceptance probability for an RJMCMC algorithm in which the full coefficient parameter is proposed without a transformation and a symmetric jumping distribution is employed has the formula

$$r = \min \left( 1, \frac{\pi(\theta^*, m^* | x) q(\theta | \theta^*, m, m^*)}{\pi(\theta, m | x) q(\theta^* | \theta, m, m^*)} \right).$$

(3.3.40)

As previously noted, the discretised Laplacian distribution will be used as the jumping distribution for the following Reversible Jump steps. For a model index of $m$, the distribution is

$$j(m^* | m) \propto \exp \left[ -\Delta_m | m^* - m | \right].$$

Reversible Jump - RJM Step

Using the form of the acceptance probability from equation (3.3.40), and applying the parameters of the conditional mean yields

$$r = \min \left( 1, \frac{\pi(M_{k^*}, \phi^* | x, \Theta - (M_k, \phi)) q(\phi | \phi^*, M_k, M_{k^*})}{\pi(M_k, \phi | x, \Theta - (M_k, \phi)) q(\phi^* | \phi, M_k, M_{k^*})} \right).$$

(3.3.41)

To determine a suitable proposal distribution for $\phi$, we need to investigate the conditional distribution for $\phi$. This is similar to the derivation of the conditional posterior distribution for $\phi$ for the STAR model shown in Appendix A.1.2. For the STAR-GARCH model, the derivation of the conditional posterior distribution for $\phi$ is performed in Appendix C.1.1 and the end result is

$$p(\phi | x, \Theta - \phi) \propto [2\pi]^{-\frac{1}{2}} |C_\phi|^{-\frac{1}{2}} |H|^{-\frac{1}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ x^T H^{-1} x - \phi^T C_\phi^{-1} \phi \right] \right] \times$$

$$[2\pi]^{-\frac{2(k+1)}{2}} |C_\phi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ (\phi - \hat{\phi})^T C_\phi^{-1} (\phi - \hat{\phi}) \right] \right].$$

(3.3.42)

When the conditional variance of the model is constant, in other words, when the diagonal entries of the matrix $H$ are all equal, the conditional posterior density in equation (3.3.42) is proportional to a multivariate normal distribution with mean vector $\hat{\phi}$ and covariance matrix $C_\phi$. Assuming that the
Section 3.3: STAR-GARCH Models

diagonal entries are all equal provides a suitable approximation for the target distribution in which
$H$, in fact, has non-constant diagonal entries. With this in mind, the proposal density for $\phi$ in the
$\text{RJ}_M$ Step will be a multivariate normal density with location and covariance parameters $\hat{\phi}$ and $\Delta_\phi \hat{C}_\phi$, respectively.

Therefore, after some manipulation, the proposal distribution for $\phi$ is

$$ q(\phi^*|\phi,M_k,M_{k^*}) \sim \mathcal{N}_{2k+2}(\hat{\phi},\Delta_\phi \hat{C}_\phi),$$

where

$$\hat{\phi} = \hat{C}_\phi Z_1^T \hat{H}^{-1} x$$

$$\hat{C}_\phi^{-1} = Z_1^T \hat{H}^{-1} Z_1 + \Sigma_\phi^{-1}.$$ (3.3.44)

The parameter $\Delta_\phi$ in equation (3.3.44) is a tuning parameter that is adjusted in order to achieve an
acceptable acceptance rate. The matrix $\hat{H}$ is a diagonal matrix similar to $H$, but with diagonal entries
replaced by the sample variance of the data.

The acceptance probability for making the move from model $M_k$, with coefficient parameter vector
$\phi$, to model $M_{k^*}$, with coefficient parameter vector $\phi^*$, is calculated in Appendix C.2.1. The formula
obtained for the acceptance probability is:

$$ r_{(M_k,SG)} = \min \left( 1, \frac{1}{|H^*|} \frac{|H|}{|H^*|} \frac{|2\pi|}{|2\pi|^k} \frac{|\Sigma_{\phi^*}|}{|\Sigma_{\phi}|} \frac{1}{2} \left[ \frac{A^{k+1}}{A^k} \right]^{\frac{T}{2}} \frac{1}{2} |\Delta_\phi \hat{C}_\phi|^{-\frac{1}{2}} |\Delta_\phi \hat{C}_{\phi^*}|^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2} \left( (x - Z_1^* \phi^*)^T H^{-1} (x - Z_1^* \phi^*) + (\phi^*)^T \Sigma_{\phi^*}^{-1} \phi^* - (x - Z_1 \phi)^T \hat{H}^{-1} (x - Z_1 \phi) - \phi^T \Sigma_\phi^{-1} \phi + \frac{1}{2} \left( \phi - \hat{\phi} \right)^T \left[ \Delta_\phi \hat{C}_\phi \right]^{-1} \left( \phi - \hat{\phi} \right) - \frac{1}{2} \left( \phi^* - \hat{\phi} \right)^T \left[ \Delta_\phi \hat{C}_{\phi^*} \right]^{-1} \left( \phi^* - \hat{\phi} \right) \right] \right) \right).$$ (3.3.45)

In order to maintain consistency with the findings of Section 5.1, the calculations used to obtain
equation (3.3.45) employ the same data points in the calculation of the likelihood function for the
current parameters and the proposed parameters. The first $s^* = \max(k^*,k,d,l,m)$ data points are
allocated as initial conditions. The parameters of the proposal distribution for $\phi$ shown in equation
(3.3.44) are also calculated, conditional on the first $s^*$ data points. The structure of the Reversible
Jump step for the conditional mean can be outlined as follows:

**RJ$_M$ Algorithm**

1. The current state is $(M_k^{(i-1)},\phi^{(i-1)},d^{(i-1)},\gamma^{(i-1)},e^{(i-1)},N_j^{(i-1)},\alpha^{(i-1)})$. 

63
2. Propose a candidate model index $M_{k^*}$ from $\mathcal{M}_{k^*} \mid M_{k}$ using the discretised Laplacian shown in equation (2.2.17). Propose a new conditional mean coefficient parameter vector $\phi$ using $q(\phi^* \mid \phi, M_k, M_{k^*})$ from equation (3.3.43).

3. Calculate the acceptance probability $r_{(M_{k}, SG)}$ per equation (3.3.45).

4. Simulate $u \sim U(0, 1)$:
   
   (a) If $r_{(M_{k}, SG)} > u$, then:
   
   (i) Accept the candidate model index $M_{k^*}$ and set $M_k^{(i)} = M_{k^*}$.
   
   (ii) Accept the candidate conditional mean coefficient parameter $\phi^*$ and set $\phi^{(i)} = \phi^*$.

   (b) If $r_{(M_{k}, SG)} < u$, then:
   
   (i) Reject the candidate model index $M_{k^*}$ and proposed mean coefficient parameter $\phi^*$.
   
   Set $M_k^{(i)} = M_k^{(i-1)}$ and $\phi^{(i)} = \phi^{(i-1)}$.

   (ii) Using the distribution from equation (3.3.39), simulate the conditional mean coefficient parameter
   
   $$\phi^{(i)} \sim p(\phi \mid x, M_k^{(i)}, \alpha^{(i)}, \gamma^{(i)}, c^{(i)}, N_j^{(i)}, \alpha^{(i-1)})$$
   
   using the Metropolis-Hastings algorithm described below (MH$\phi$ Step).

**Reversible Jump - RJ$\lambda$ Step**

Applying the parameters of the conditional variance to the general Reversible Jump acceptance probability from equation (3.3.40) yields

$$r_{(N_j, SG)} = \min \left( 1, \frac{\pi(N_j^*, \alpha^* \mid x, \Theta_{-(N_j, \alpha)}) q(\alpha^* \mid \alpha, N_j^*, N_j^*)}{\pi(N_j, \alpha \mid x, \Theta_{-(N_j, \alpha)}) q(\alpha^* \mid \alpha, N_j, N_j^*)} \right).$$

As in the pure GARCH model estimation procedure, we encounter the dilemma of how to determine the location parameters of the proposal distribution for the conditional variance coefficient parameter $\alpha$. For the GARCH model, the MLE’s were a very good starting point for the pilot and were often close to the location parameters determined from the pilot run.

However, for the STAR-GARCH model, we have the added complexity of a non-constant conditional mean. Initially, the location parameters for the proposal distributions for $\alpha$ were set to the MLE’s for each of the GARCH model indices included in the search. This assumed that the conditional mean was constant. Repeated trials using simulated data from a true STAR-GARCH process showed that this was rather ineffective and led to poor performance. Ultimately, the misspecification of the conditional mean led to a poor specification of the conditional variance location parameters. This meant that the
location parameters were not in the areas of high posterior probability and, consequently, the algorithm rarely jumped between models when the full MCMC algorithm was executed.

Experience from fitting univariate STAR and univariate GARCH models in this thesis to real data would suggest that the data tends to contain more information pertaining to the parameters of the conditional mean than pertaining to the parameters in the conditional variance equation. The evidence, as represented by the posterior probabilities, for a particular conditional mean model order tends to be significantly stronger when compared to the conditional variance model index. This can be seen by examining the model order posterior probabilities shown in Chapter 5. With this in mind, two pilot runs will be executed for the STAR-GARCH algorithm. The first run will be used to specify the conditional mean, assuming constant variance. The conditional mean parameters from this pilot run will then be used as the conditional mean parameters for the second pilot run. The second pilot run will determine location and variance parameters for the coefficient parameter vectors of the GARCH models being considered.

The proposal distribution for the conditional variance coefficient parameter vector is defined similarly to the one described for the pure GARCH model in Section 3.2. That is,

$$ q(\alpha^*|\alpha, N_j, N_{j^*}) \sim T_{l^* + m^*} \left( \hat{\alpha}_{N_{j^*}}, \nu, \Sigma_{\alpha_{N_{j^*}}} \right). \quad (3.3.47) $$

The location parameters for the proposal distributions in equation (3.3.47), $\hat{\alpha}_{N_{j^*}}$, will be determined from the results of the pilot runs for each GARCH model under consideration. The degrees of freedom parameter $\nu$ will be set to some fixed amount. Setting $\nu = 5$ lead to acceptable results. The scale matrix $\Sigma_{\alpha_{N_{j^*}}}$ will be found by creating a diagonal matrix whose diagonal entries correspond to the variance of the respective posterior distribution from the pilot run. Both of the proposal parameters $\hat{\alpha}_{N_{j^*}}$ and $\Sigma_{\alpha_{N_{j^*}}}$ are given the subscript $N_{j^*}$ to indicate that these parameters are different for each of the candidate model moves.

Using the joint conditional posterior for $N_j$ and $\alpha$, and the proposal distribution for $\alpha$ in equations (3.3.39) and (3.3.47), respectively, we can now determine an expression for the acceptance probability calculation using equation (3.3.46). The details of this substitution and simplification are given in Appendix C.2.2, and the resulting expression is:

$$ r_{(N_j, \alpha)} = \min \left( 1, |H^*|^{-\frac{1}{2}} |H|^{\frac{1}{2}} \left( \frac{l + 1}{l^* + 1} \right)^{\frac{m}{2}} \left( \frac{m + 1}{m^* + 1} \right)^{\frac{m}{2}} \times \left( x - Z_1 \phi \right)^T H^{-1} (x - Z_1 \phi) \right), \quad (3.3.48) $$
where the most recent values from the MCMC chain are $N_j = N_j^{(i-1)}$, and, therefore, $l = l^{(i-1)}$, $m = m^{(i-1)}$ and $\alpha = \alpha^{(i-1)}$. The $s$ is placed on the $H$ matrix to remind the reader that it is a function of $\alpha$, $l$ and $m$.

Now that the form of the acceptance probability for a model move is complete, the structure of the conditional variance Reversible Jump step can be outlined.

**RJ$_N$ Algorithm**

1. The current state is $(M_k^{(i)}, \phi^{(i)}, d^{(i)}, \gamma^{(i)}, c^{(i)}, N_j^{(i-1)}, \alpha^{(i-1)})$.

2. Propose a candidate model index $N_{j^*}$ using $j(N_{j^*}|N_j)$, the discretised Laplacian from equation (2.2.17). Propose a new conditional variance coefficient parameter $\alpha$ using $q(\alpha^*|\alpha, N_j, N_{j^*})$ in equation (3.3.47).

3. Calculate the acceptance probability $r(N_{j^*}, S_G)$ as in equation (3.3.48).

4. Simulate $u \sim U(0, 1)$:

   (a) If $r(N_{j^*}, S_G) > u$, then:

      (i) Accept the candidate model index $N_{j^*}$ and set $N_j^{(i)} = N_{j^*}$.

      (ii) Accept the proposed conditional variance coefficient parameter $\alpha^*$ and set $\alpha^{(i)} = \alpha^*$.

   (b) If $r(N_{j^*}, S_G) < u$, then:

      (i) Reject the candidate model index $N_{j^*}$ and proposed variance coefficient parameter $\alpha^*$. Set $N_j^{(i)} = N_j^{(i-1)}$ and $\alpha^{(i)} = \alpha^{(i-1)}$.

      (ii) Using the distribution from equation (3.3.39), simulate the conditional variance coefficient parameter

      $$\alpha^{(i)} \sim p(\alpha|\mathbf{x}, M_k^{(i)}, \phi^{(i)}, d^{(i)}, \gamma^{(i)}, c^{(i)}, N_j^{(i)})$$

      using a Metropolis-Hastings algorithm as described below (**MH$_\alpha$ Step**).

Both of the Reversible Jump steps outlined above require the simulation of the relevant coefficient parameter when a move to a candidate model is rejected. The Metropolis-Hastings algorithms for simulating these coefficient parameters are outlined below.

**Metropolis-Hastings Steps**

When the proposed move to candidate model $M_k^*$ or $N_{j^*}$ is rejected, the corresponding conditional mean or conditional variance coefficient parameter is also rejected. As the proposed move to a candidate model is rejected more often than not, the coefficient parameter is rarely updated. If no simulation of
the coefficient parameter from the full conditional posterior distribution occurred, the algorithm would converge slowly and exhibit poor mixing of the MCMC chains.

Therefore, to improve mixing of the MCMC chains and the speed of convergence, simulation from the full conditional posterior distribution of the conditional mean and the conditional variance coefficient parameter is performed using a Metropolis-Hastings algorithm. These algorithms, and the Metropolis-Hastings algorithm for the smoothing and location parameters of the transition function, are outlined in the following sections.

The general form of the acceptance probability originally noted in equation (2.2.12) is recalled here:

$$r = \min \left( 1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^*|\theta^{(i-1)})} \right). \quad (3.3.49)$$

### Metropolis-Hastings - MH\(\phi\) Step

When simulating the conditional mean coefficient parameter \(\phi\) using an Metropolis-Hastings algorithm, a similar proposal distribution to the one defined in equation (3.3.43) is used. The key difference is that the location parameter is simply defined to be the most recently simulated value. That is,

$$q(\phi^*|\phi, M_k) \sim N_{2k+2} \left( \phi, \Delta_\phi \hat{C}_\phi \right), \quad (3.3.50)$$

where \(\hat{C}_\phi^{-1} = Z_1^T H^{-1} Z_1 + \Sigma_\phi^{-1}\), \(\phi = \phi^{(i-1)}\) and \(\Delta_\phi\) is a tuning parameter which adjusts the acceptance rate. The diagonal matrix \(H\) is similar to \(H\), but its diagonal entries are replaced with the sample variance of the data. This proposal distribution is symmetric, meaning that the ratio of the proposal distributions is one, and so it is cancelled out in the acceptance probability calculation.

Using the terms relevant to the conditional mean of the STAR-GARCH model, the acceptance probability is derived in Appendix C.2.3 and shown below in equation (3.3.51). The most recent realisation for \(\phi\) is denoted by \(\phi = \phi^{(i-1)}\). Thus,

$$r_{(\phi,SG)} = \min \left( 1, \left[ \left| H^* \right| H^{-1} \right]^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( (x - Z_1 \phi^*)^T H^{-1} (x - Z_1 \phi^*) + \phi^* T \Sigma_\phi^{-1} \phi^* - (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) - \phi^T \Sigma_\phi^{-1} \phi \right) \right] \right). \quad (3.3.51)$$

### MH\(\phi\) Algorithm

1. The current state is \(\left( M_k^{(i)}, \phi^{(i-1)}, \gamma^{(i-1)}, \epsilon^{(i-1)}, N_j^{(i-1)}, \alpha^{(i-1)} \right)\).
2. Propose a new conditional mean coefficient parameter $\phi$ using $q(\phi^*|\phi,M_k)$ from equation (3.3.50).

3. Calculate the acceptance probability $r_{(\phi,SG)}$ using equation (3.3.51).

4. Simulate $u \sim U(0,1)$:
   
   (a) If $r_{(\phi,SG)} > u$, accept the proposed conditional mean coefficient parameter and set $\phi^{(i)} = \phi^*$.
   
   (b) If $r_{(\phi,SG)} < u$, reject the proposed conditional mean coefficient parameter and set $\phi^{(i)} = \phi^{(i-1)}$.

**Metropolis-Hastings - MH$_\alpha$ Step**

The proposal distribution for the Metropolis-Hastings algorithm to simulate the conditional variance coefficient parameter $\alpha$ will be prescribed similarly to the one in equation (3.3.47). The difference is that the location parameter is simply defined to be the most recent simulated value. That is,

$$q(\alpha^*|\alpha,N_j) \sim T_{n+m}^1(\alpha, \nu, \Sigma_{\alpha,N_j})^*.$$  

(3.3.52)

Therefore, the proposal distribution is again symmetric and will not feature in the acceptance probability calculation. After applying the terms for the conditional variance of the model, the simplification of the acceptance probability is performed in Appendix C.2.4. Its formula is given below:

$$r_{(\alpha,SG)} = \min\left(1, \left[|H^*||H|^{-1}\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left((x - Z_1\phi)^T (H^* - H^{-1}) (x - Z_1\phi)\right)\right]\right).$$  

(3.3.53)

**MH$_\alpha$ Algorithm**

1. The current state is $(M_k^{(i)}, \phi^{(i)}, d^{(i)}, \gamma^{(i)}, c^{(i)}, N_j^{(i)}, \alpha^{(i-1)})$.

2. Propose a new conditional variance coefficient parameter $\alpha$ using $q(\alpha^*|\alpha,N_j)$ from equation (3.3.52).

3. Calculate the acceptance probability $r_{(\alpha,SG)}$ using equation (3.3.53).

4. Simulate $u \sim U(0,1)$:
   
   (a) If $r_{(\alpha,SG)} > u$, accept the proposed conditional variance coefficient parameter and set $\alpha^{(i)} = \alpha^*$.
   
   (b) If $r_{(\alpha,SG)} < u$, reject the proposed conditional variance coefficient parameter and set $\alpha^{(i)} = \alpha^{(i-1)}$. 

---

68
Metropolis-Hastings - MH\(_{(\gamma,c)}\) Step

The proposal distributions for the smoothing and location parameters of the transition function are the same as those used in Section 3.1.4. These were originally defined for the MTM algorithm in equation (3.1.12), and are recalled here:

\[
q\left(\gamma^* | \gamma(i-1)\right) \sim \mathcal{G}\left(\frac{(\gamma(i-1))^2}{\Delta \gamma}, \frac{\gamma(i-1)}{\Delta \gamma}\right)
\]

\[
q\left(c^* | c(i-1)\right) \sim \mathcal{N}\left(c(i-1), \Delta c\right).
\]

(3.3.54)

For the proposal distributions specified in equation (3.3.54), \(\Delta \gamma\) and \(\Delta c\) are the variances and are used to tune the acceptance rate of the embedded algorithm. The tuning parameters should be set with the aim of achieving an acceptance rate of somewhere between 10% and 50%.

The joint conditional posterior distribution shown in equation (3.3.39) is used in conjunction with the proposal distributions outlined in equation (3.3.54) to determine the acceptance probability of jointly accepting the candidate smoothing and location parameters. The derivation of this acceptance probability is performed in Appendix C.2.5, and the end result is given below:

\[
r_{(\gamma,c,SG)} = \min \left(1, \left[\frac{[\gamma^* \gamma^{-1}]^{\alpha \gamma - 1} \Gamma\left(\frac{\gamma^*}{\Delta \gamma}\right)}{\Gamma\left(\frac{\gamma}{\Delta \gamma}\right)} \frac{\gamma^*}{\Delta \gamma} \gamma^{\gamma^* - \gamma - \frac{1}{2}} \exp\left[-\frac{1}{2} (x - Z_1^\phi)^T H^{-1} (x - Z_1^\phi) - \frac{1}{\sigma_c^2} (c^* - \mu_c)^2 - (c - \mu_c)^2 \right]\right] \right).
\]

(3.3.55)

We are now able to outline the MH\(_{(\gamma,c)}\) Algorithm.

MH\(_{(\gamma,c)}\) Algorithm

1. The current state is \((M_k^{(i)}, \phi^{(i)}, d^{(i)}, \gamma^{(i-1)}, c^{(i-1)}, \alpha^{(i-1)})\).

2. Propose new smoothing and location parameters from the proposal distributions \(q(\gamma^* | \gamma)\) and \(q(c^* | c)\), respectively, as presented in equation (3.3.54).

3. Calculate the acceptance probability \(r_{(\gamma,c,SG)}\) using equation (3.3.55).

4. Simulate \(u \sim \mathcal{U}(0,1)\):
(a) If \( r_{(\gamma,c,SG)} > u \), accept the proposed smoothing and location parameters. Set \( \gamma^{(i)} = \gamma^* \) and \( c^{(i)} = c^* \).

(b) If \( r_{(\gamma,c,SG)} < u \), reject the proposed smoothing and location parameters. Set \( \gamma^{(i)} = \gamma^{(i-1)} \) and \( c^{(i)} = c^{(i-1)} \).

### 3.4 Conclusion

In this chapter, posterior simulation algorithms for univariate STAR, GARCH and STAR-GARCH models have been developed. These algorithms employ a variety of MCMC algorithms in order to incorporate model selection and parameter estimation into one task. Posterior simulation algorithms for the multivariate extensions of these models are presented in the following chapter.
Chapter 4

Multivariate Models

Rather than beginning our discussion with the multivariate STAR model, it may be easier to first review the multivariate autoregressive model. This model is commonly referred to as a Vector Autoregressive (VAR) model, and has a similar construction to the multivariate STAR model.

4.1 Vector Autoregressive

In the multivariate setting, each \( x_t \) is no longer an individual data point that represents a single value from a univariate time series. Instead, each data point \( x_t \) is a \( p \times 1 \) vector whose entries represent the values of \( p \) separate time series at time \( t \).

Our goal is to fit a VAR model to the given data. Let \( \mathcal{M}_k \), where \( k \in \mathcal{K} = \{1, \ldots, k_{\text{max}}\} \) denote the collection of candidate models. Without loss of generality, we may define the VAR(\( k \)) model with no intercept term in a similar fashion to the AR model as follows:

\[
x_t = \Phi_1 x_{t-1} + \cdots + \Phi_k x_{t-k} + \varepsilon_t
\]

(4.1.1)

Here, \( \Phi_i \) are \( p \times p \) coefficient matrices for \( i \in \{1, \ldots, k\} \). The error of the process \( \varepsilon_t \) is a \( p \times 1 \) vector such that

\[
\varepsilon_t \sim \mathcal{N}(0_{p \times 1}, \Sigma_{p \times p}).
\]

To simplify the model, we can collect the autoregressive terms in (4.1.1) into a single vector \( z_t \) and
Chapter 4: Multivariate Models

collect the $k$ coefficient matrices in (4.1.1) into a single matrix $\Phi$ of coefficients as follows:

$$z_t^T = [x_{t-1}^T, \ldots, x_{t-k}^T]$$

$$\Phi^T = [\Phi_1, \ldots, \Phi_k].$$

The form of the model can now be simplified to

$$x_t = \Phi^T z_t + \varepsilon_t.$$

Another form of the model that is useful for mathematical manipulation is the matrix form of the model shown below:

$$X = Z_1 \Phi + E,$$

where

$$\begin{bmatrix}
x_{k+1}^T \\
x_{k+2}^T \\
\vdots \\
x_N^T \\
z_{k+1}^T \\
z_{k+2}^T \\
\vdots \\
z_N^T \\
x_{k-1}^T & x_{k-2}^T & \ldots & x_1^T \\
x_{k-1}^T & x_{k-2}^T & \ldots & x_2^T \\
\vdots & \vdots & \vdots & \vdots \\
x_{N-1}^T & x_{N-2}^T & \ldots & x_{N-k}^T
\end{bmatrix}$$

The matrices in (4.1.2) have the following dimensions: $X \in \mathbb{R}^{n \times p}$, $E \in \mathbb{R}^{n \times p}$, $\Phi \in \mathbb{R}^{kp \times p}$ and $Z_1 \in \mathbb{R}^{n \times kp}$.

4.1.1 Likelihood Function

Under the assumption that the vector $\varepsilon_t$ is distributed from a multivariate normal distribution, the likelihood function will be the product of $n$ multivariate normal densities, with mean $\Phi^T z_t$ and covariance matrix $\Sigma$. If $k$ is the model order, the likelihood will be calculated using the first $k$ data points as initial values. The useful data is therefore represented by data points $k+1$ to $N$, and so has length $n = N - k$. The conditional likelihood is calculated as follows:
Section 4.1: Vector Autoregressive

\[
p (X|M_k, \Phi, \Sigma) = \prod_{t=(k+1)}^{N} p (x_t|M_k, \Phi, \Sigma)
\]

\[
= \prod_{t=(k+1)}^{N} (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( x_t - \Phi^T z_t \right)^T \Sigma^{-1} \left( x_t - \Phi^T z_t \right) \right]
\]

\[
= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp \left[ -\frac{1}{2} \sum_{t=(k+1)}^{N} \left( x_t - \Phi^T z_t \right)^T \Sigma^{-1} \left( x_t - \Phi^T z_t \right) \right].
\]

By focusing on the sum in the exponent in (4.1.3), substituting \( \varepsilon_t = x_t - \Phi^T z_t \) for convenience and performing some matrix manipulations, we obtain:

\[
\sum_{t=(k+1)}^{N} \left( x_t - \Phi^T z_t \right)^T \Sigma^{-1} \left( x_t - \Phi^T z_t \right)
\]

\[
= \sum_{t=(k+1)}^{N} \varepsilon_t^T \Sigma^{-1} \varepsilon_t
\]

\[
= \varepsilon_{k+1}^T \Sigma^{-1} \varepsilon_{k+1} + \varepsilon_{k+2}^T \Sigma^{-1} \varepsilon_{k+2} + \ldots + \varepsilon_N^T \Sigma^{-1} \varepsilon_N
\]

\[
= \text{tr} (\varepsilon_{k+1}^T \Sigma^{-1} \varepsilon_{k+1}) + \text{tr} (\varepsilon_{k+2}^T \Sigma^{-1} \varepsilon_{k+2}) + \ldots + \text{tr} (\varepsilon_N^T \Sigma^{-1} \varepsilon_N)
\]

\[
= \text{tr} (\varepsilon_{k+1}^T \Sigma^{-1} \varepsilon_{k+1}) + \text{tr} (\varepsilon_{k+2}^T \Sigma^{-1} \varepsilon_{k+2}) + \ldots + \text{tr} (\varepsilon_N^T \Sigma^{-1} \varepsilon_N)
\]

\[
= \text{tr} \left( \left( \varepsilon_{k+1}^T \varepsilon_{k+1} + \varepsilon_{k+2}^T \varepsilon_{k+2} + \ldots + \varepsilon_N^T \varepsilon_N \right) \Sigma^{-1} \right)
\]

\[
= \text{tr} \left( \left[ \begin{bmatrix} k+1 \\ k+2 \\ \vdots \\ N \end{bmatrix} \Sigma^{-1} \right] \right)
\]

\[
= \text{tr} \left( \left[ X - Z_1 \Phi \right]^T \left[ X - Z_1 \Phi \right] \Sigma^{-1} \right).
\]

The likelihood function can now be simplified to yield

\[
p (X|M_k, \Phi, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( \left[ X - Z_1 \Phi \right]^T \left[ X - Z_1 \Phi \right] \Sigma^{-1} \right) \right].
\]

Now that the derivation of the likelihood function in (4.1.5) is complete, we shall discuss the prior distribution.

4.1.2 Prior Distribution

A prior distribution similar to the one used by Geweke, Koop, and Van Dijk (2011) in the analysis of VAR models in econometrics will be employed. This is a simple and uninformative prior distribution.
with expression:
\[ p(M_k, \Phi, \Sigma) \propto \left[ \frac{\Lambda_k}{k!} \right]^{r_k} I_K(k) |\Sigma|^{-\frac{n+p+1}{2}}. \]  
(4.1.6)

Other acceptable options for the prior distribution are the conjugate prior distributions from the Normal family and the Minnesota Prior (Doan, Litterman, and Sims (1984)). The Minnesota Prior is particularly popular when forecasting is the primary objective of the user. We have chosen the simple prior distribution in (4.1.6), bearing in mind that we shall need to analyse the significantly more complicated Multivariate STAR-GARCH model.

### 4.1.3 Posterior Distribution

Combining the likelihood and the prior distribution results in the following posterior distribution:

\[ p(M_k, \Phi, \Sigma | X) \propto p(X | M_k, \Phi, \Sigma) p(M_k, \Phi, \Sigma) \]
\[ \propto (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{n}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{\Lambda_k}{k!} \right]^{r_k} I_K(k) |\Sigma|^{-\frac{n+p+1}{2}} \]  
(4.1.7)

Now, concentrating on the trace in the exponent, we can apply a similar procedure to the one used in (A.1.2) for the STAR model to obtain

\[
\text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T - \Phi^T Z_1^T] [X - Z_1 \Phi] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T Z_1^T \Phi - \Phi^T Z_1^T X + \Phi^T Z_1^T Z_1^T \Phi ] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T Z_1^T \Phi - \Phi^T Z_1^T X - C_\Phi^{-1} (\Phi - C Z_1^T X) ] \Sigma^{-1} \right). 
\]

Letting \( C_\Phi^{-1} = Z_1^T Z_1 \), we now have

\[
\text{tr} \left( [X^T Z_1^T \Phi - \Phi^T Z_1^T X - C_\Phi^{-1} (\Phi - C Z_1^T X) ] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T Z_1^T \Phi - \Phi^T (C_\Phi^{-1} \Phi - Z_1^T X) ] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T Z_1^T \Phi - \Phi^T (C_\Phi^{-1} \Phi - C Z_1^T X) ] \Sigma^{-1} \right). 
\]

After setting \( \Phi = C_\Phi Z_1^T X \), we can simplify the exponent as shown in Appendix D.1. Substituting the
simplified exponent back into the posterior distribution yields
\[
p(M_k, \Phi, \Sigma | X) \propto (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n+p+1}{2}} \exp \left[ -\frac{1}{2} \left( \text{tr} \left( (\Phi - \hat{\Phi})^T C_\Phi^{-1} (\Phi - \hat{\Phi}) \Sigma^{-1} \right) + \text{tr} \left( \left[ X^T X - \hat{\Phi}^T C_\Phi^{-1} \hat{\Phi} \right] \Sigma^{-1} \right) \right] \right] \Lambda_k \tau_k \mathbb{I}_K (k). \tag{4.1.8}
\]

In order to derive the conditional posterior distributions for a particular parameter, we assume that the other parameters are constants and look at the proportional relationship. The joint conditional posterior distribution for \( M_k \) and \( \Phi \) is
\[
p(M_k, \Phi | X, \Sigma) \propto \exp \left[ -\frac{1}{2} \left( \text{tr} \left( (\Phi - \hat{\Phi})^T C_\Phi^{-1} (\Phi - \hat{\Phi}) \Sigma^{-1} \right) + \text{tr} \left( \left[ X^T X - \hat{\Phi}^T C_\Phi^{-1} \hat{\Phi} \right] \Sigma^{-1} \right) \right] \right] \Lambda_k \tau_k \mathbb{I}_K (k). \tag{4.1.9}
\]

In addition to the joint conditional posterior distribution in (4.1.8), we require the full conditional posterior distribution for the coefficient matrix \( \Phi \) as shown below:
\[
p(\Phi | M_k, X, \Sigma) \propto (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n+p+1}{2}} \exp \left[ -\frac{1}{2} \left( \text{tr} \left( (\Phi - \hat{\Phi})^T C_\Phi^{-1} (\Phi - \hat{\Phi}) \Sigma^{-1} \right) + \text{tr} \left( \left[ X^T X - \hat{\Phi}^T C_\Phi^{-1} \hat{\Phi} \right] \Sigma^{-1} \right) \right] \right] \Lambda_k \tau_k \mathbb{I}_K (k). \tag{4.1.10}
\]

The conditional posterior distribution shown in (4.1.9) is proportional to a Matrix Normal distribution.

We can derive the following expression for the posterior distribution for \( \Sigma \):
\[
p(\Sigma | M_k, X, \Phi) \propto (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n+p+1}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( (X - Z_1 \Phi)^T (X - Z_1 \Phi) \Sigma^{-1} \right) \right]. \tag{4.1.10}
\]

The distribution in (4.1.10) is proportional to an Inverted Wishart distribution. In summary, the conditional posterior distributions that are of standard form are
\[
p(\Phi | X, \Sigma) \sim N(k, p) \left( \Phi, C_\Phi, \Sigma \right) \tag{4.1.11}
p(\Sigma | X, \Phi) \sim IW(\Psi, n),
\]

where
\[
\Phi = C_\Phi Z_1^T X \\
C_\Phi^{-1} = Z_1^T Z_1 \\
\Psi = [X - Z_1 \Phi]^T [X - Z_1 \Phi].
\]
We are now ready to design an MCMC algorithm to estimate the parameters for the VAR model.

### 4.1.4 Posterior Simulator

The inclusion of the VAR model order as a parameter for estimation necessitates the employment of an RJMCMC scheme. Although this algorithm will not be implemented, the structure of the RJMCMC step will be similar to that in the algorithm for the Multivariate STAR model. The algorithm design incorporates a Gibbs sampler algorithm with an embedded RJMCMC step.

**VAR - Main Algorithm**

1. Initialise \((M_k^{(0)}, \Phi^{(0)}, \Sigma^{(0)})\) either randomly or deterministically and set \(i = 1\).

2. At the \(i\)th iteration, use the conditional posterior distributions shown in (4.1.11) to perform the following steps:

   (a) Simulate the covariance matrix of the error vector using

   \[ \Sigma^{(i)} \sim p(\Sigma|X, M_k^{(i-1)}, \Phi^{(i-1)}) . \]

   (b) **RJ Step** - Simulate the STAR model order and coefficient parameter matrix using

   \[ \left( M_k^{(i)}, \Phi^{(i)} \right) \sim p(M_k, \Phi|X, \Sigma^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)}) . \]

   (ii) If the proposal in step 2(b)(i) is rejected, simulate the coefficient parameter matrix using

   \[ \Phi^{(i)} \sim p(\Phi|X, M_k^{(i)}, \Sigma^{(i)}) . \]

3. Set \(i = i + 1\). If all the MCMC chains have converged, STOP; else, move to step 2.

In order to implement this algorithm, we shall use an embedded RJMCMC algorithm to simulate the joint conditional posterior distribution \(p(M_k, \Phi|X, \Sigma)\).

**Reversible Jump - RJ Step**

Recall the joint conditional posterior distribution \(p(M_k, \Phi|X, \Sigma)\) in (4.1.8). We shall simulate from this distribution using the RJMCMC algorithm. Note that a change in the model order \(k\) leads to changes in the dimension of the parameter space.
Recall from Section 2.2.4, the form of the acceptance probability used throughout this thesis for Reversible Jump steps is

\[ r = \min \left( 1, \frac{\pi(\theta^*|m, m^*|X) \cdot q(\theta|\theta^*, m, m^*)}{\pi(\theta|m, m|X) \cdot q(\theta^*|\theta, m, m^*)} \right). \]  

(4.1.12)

Using the notation for the VAR model with the acceptance probability formula in (4.1.12), the acceptance probability becomes

\[ r_{(M_k, \Phi)} = \min \left( 1, \frac{p(M_k, \Phi^*|X) \cdot q(\Phi^*|\Phi, M_k, M_{k*})}{p(M_k, \Phi|X) \cdot q(\Phi^*|\Phi, M_k, M_{k*})} \right). \]  

(4.1.13)

As in Troughton and Godsill (1997), Lopes and Salazar (2006), and for the STAR model in Chapter 3, the Candidate’s Identity (Besag 1989) can be used here to greatly simplify the acceptance probability formula through a careful selection of the proposal distribution and some mathematical manipulation.

The manipulation relating to the STAR model is performed in Appendix A.2.1.

First, the proposal distribution for the coefficient matrix, \( \Phi \) will be defined as the full conditional posterior distribution of \( \Phi \) shown in equation (4.1.9). That is,

\[ q(\Phi^*|\Phi, M_k, M_{k*}) \sim p(\Phi^*|X, M_{k*}, \Sigma). \]  

(4.1.14)

Second, we can mathematically manipulate the Candidate’s Identity using the notation of the VAR model as follows:

\[ \frac{p(M_k, \Phi|X, \Sigma)}{p(\Phi|X, M_k, \Sigma)} = \frac{p(X|M_k, \Sigma) \cdot p(M_k|\Sigma)}{p(X|\Sigma)} \]  

(4.1.15)

p (M_k, \Phi|X, \Sigma) = \frac{p(X|M_k, \Sigma) \cdot p(M_k|\Sigma) \cdot p(\Phi|X, M_k, \Sigma)}{p(X|\Sigma)}.

Using the expression in (4.1.15), the ratio of the target distributions becomes

\[ \frac{p(M_k^*, \Phi^*|X, \Sigma)}{p(M_k, \Phi^*|X, \Sigma)} = \frac{p(X|M_k^*, \Sigma) \cdot p(M_k^*|\Sigma) \cdot p(\Phi^*|X, M_k^*, \Sigma)}{p(X|\Sigma)} \cdot \frac{p(X|M_k, \Sigma) \cdot p(M_k|\Sigma) \cdot p(\Phi|X, M_k, \Sigma)}{p(X|\Sigma)}. \]  

(4.1.16)

Now, by substituting the result in (4.1.16), together with the proposal distribution for the coefficient matrix defined in (4.1.14), into the acceptance probability formula in (4.1.13), we obtain the following
form of the acceptance probability:

\[
    r_{(M_k, \nu)} = \min \left( 1, \frac{p(X|M_k, \nu, \Sigma) p(M_k, \nu, \Sigma) p(\Phi^*|X, M_k, \Sigma)}{p(X|\nu, M_k) p(M_k|\nu) p(\Phi^*|X, M_k, \Sigma)} \right)
\]

\[
= \min \left( 1, \frac{p(X|M_k, \nu) p(M_k|\nu)}{p(X|\nu, M_k) p(M_k|\nu)} \right).
\]

(4.1.17)

In (4.1.17), \(p(M_k|\Sigma)\) is simply a prior distribution for the model index, given the covariance matrix of the error vector. The other term, \(p(X|M_k, \Sigma)\) is a marginal likelihood and is derived as follows:

\[
p(X|M_k, \Sigma) \propto \int p(X|\nu, M_k, \Phi, \Sigma) p(\Phi|M_k, \Sigma) d\Phi. \tag{4.1.18}
\]

The terms in (4.1.18) are obtained by multiplying the likelihood function by the conditional prior distribution for \(\Phi\). This has already been performed in (4.1.7), although \(k\) is not treated as a variable there. However, the derivation does not assume that \(k\) is a constant. Therefore, the marginal likelihood is

\[
p(X|M_k, \Sigma) \propto \int p(X|\nu, M_k, \Phi, \Sigma) p(\Phi|M_k, \Sigma) d\Phi
\]

\[
\propto \int (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{np+1}{2}} \exp \left[ -\frac{1}{2} \left( \left( \Phi - \Phi^* \right)^T C_{\Phi^{-1}} \left( \Phi - \Phi^* \right) \right) \right] \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \left( X^T X - \Phi^* C_{\Phi^{-1}} \Phi \right) \Sigma^{-1} \right) \right] d\Phi
\]

\[
\propto \int (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{np+1}{2}} (2\pi)^{\frac{np^2}{2}} |\Sigma|^{\frac{np}{2}} |C_{\Phi}|^{\frac{np}{2}} \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \left( X^T X - \Phi^* C_{\Phi^{-1}} \Phi \right) \Sigma^{-1} \right) \right] d\Phi
\]

\[
\propto (2\pi)^{\frac{np^2}{2}} |\Sigma|^{\frac{np}{2}} |C_{\Phi}|^{\frac{np}{2}} \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \left( X^T X - \Phi^* C_{\Phi^{-1}} \Phi \right) \Sigma^{-1} \right) \right] \times
\]

\[
\int \exp \left[ -\frac{1}{2} \left( \left( \Phi - \Phi^* \right)^T C_{\Phi^{-1}} \left( \Phi - \Phi^* \right) \right) \right] d\Phi
\]

\[
\propto (2\pi)^{\frac{np^2}{2}} |\Sigma|^{\frac{np}{2}} |C_{\Phi}|^{\frac{np}{2}} \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \left( X^T X - \Phi^* C_{\Phi^{-1}} \Phi \right) \Sigma^{-1} \right) \right].
\]
We use the prior model order probability as defined below

\[
p(M_k|\Sigma) \propto \left[ \frac{\Lambda_k}{k!} \right]^{\tau_k} \mathbb{I}_K(k).
\]  

(4.1.20)

If we substitute the prior model probability from (4.1.20) and the simplified marginal likelihood from (4.1.19) into the quotient on the right hand side of the acceptance probability, we obtain

\[
\frac{p(X|M_k, \Sigma)p(M_k|\Sigma)}{p(X|\Sigma)p(M_k)}
\]

\[
= (2\pi)^{-\frac{n+p+k^2}{2}} \left| \Sigma \right|^{-\frac{n+p+k^2}{2}} |C_\Phi|^\frac{k}{2} \exp \left[ -\frac{1}{2} \text{tr} \left( \left[ X^TX - \phi^* C_\Phi^{-1} \phi^* \right] \Sigma^{-1} \right) \right] \times
\]

\[
\left[ \frac{\Lambda_k}{k!} \right]^{\tau_k} \mathbb{I}_K(k^*)
\]

\[
= (2\pi)^{-\frac{n+p+k^2}{2}} \left| \Sigma \right|^{-\frac{n+p+k^2}{2}} |C_\Phi|^\frac{k}{2} \left[ \frac{\Lambda_k}{k!} \right]^{\tau_k} \times
\]

\[
\exp \left[ -\frac{1}{2} \text{tr} \left( \left[ X^T X - \phi^* C_\Phi^{-1} \phi^* \right] \Sigma^{-1} \right) \right] \cdot
\]

\[
= (2\pi)^{-\frac{n+p+k^2}{2}} \left| \Sigma \right|^{-\frac{n+p+k^2}{2}} |C_\Phi|^\frac{k}{2} \left[ \frac{\Lambda_k}{k!} \right]^{\tau_k} \times
\]

\[
\exp \left[ -\frac{1}{2} \text{tr} \left( \left( -[\phi^*]^T C_\Phi^{-1} \phi^* + \phi^* C_\Phi^{-1} \phi^* \right) \Sigma^{-1} \right) \right].
\]

Therefore, the acceptance probability for making a model move to candidate model $M_{k^*}$ has the formula,

\[
\alpha_{(M_k, \phi)} = \min \left( 1, \left( 2\pi \right)^{-\frac{n+p+k^2}{2}} \left| \Sigma \right|^{-\frac{n+p+k^2}{2}} |C_\Phi|^\frac{k}{2} \left[ \frac{\Lambda_k}{k!} \right]^{\tau_k} \times
\]

\[
\exp \left[ -\frac{1}{2} \text{tr} \left( \left( -[\phi^*]^T C_\Phi^{-1} \phi^* + \phi^* C_\Phi^{-1} \phi^* \right) \Sigma^{-1} \right) \right] \right).
\]

(4.1.21)

The acceptance probability shown in (4.1.21) is for $p(\phi, M_k|X, \Sigma)$. As with the univariate case, the acceptance probability in (4.1.21) is not a function of $\phi$, and therefore does not require the simulation of a random parameter matrix $\Phi^*$. This saves computational resources and, therefore, time when implementing the algorithm. However, once a model move is either accepted or rejected, we shall need to simulate a coefficient matrix from the full conditional distribution for $\Phi$, conditional on the current model order. Therefore, the RJMCMC algorithm at 2(b) will have the form outlined below.

**RJ$_M$ Algorithm**

1. The current state is $(M_k^{(i-1)}, \Phi^{(i-1)}, \Sigma^{(i)})$. 

79
2. Propose a candidate model index, $M_{k^*}$ from $j(M_{k^*}|M_k)$ using the discretised Laplacian shown in (2.2.17).

3. Calculate the acceptance probability, $r(M_{k^*},V)$, per (4.1.21).

4. Simulate $u \sim U(0,1)$:
   
   (a) If $r(M_{k^*}) > u$, then:
       
       (i) Accept the candidate model index $M_{k^*}$ and set $M_k^{(i)} = M_{k^*}$.
       
       (ii) Simulate the coefficient parameter matrix using the full conditional posterior distribution,
           
           $\Phi^{(i)} \sim p(\Phi|X, M_k^{(i)}, \Sigma^{(i)})$.
           
   (b) If $r(M_{k^*}) < u$, then:
       
       (i) Reject the candidate model order $M_{k^*}$ and set $M_k^{(i)} = M_k^{(i-1)}$.
       
       (ii) Simulate the coefficient parameter matrix using the full conditional posterior distribution per step 2(b)(ii) of the VAR - Main Algorithm.

Now that we have found the structure of the VAR model and designed an RJMCMC algorithm for simulation from the model’s posterior distribution, it is relatively straightforward to extend this model to form the multivariate STAR model. The parameter estimation scheme will be structured similarly to the one for the VAR model, and is outlined in the following section.

### 4.2 Multivariate STAR

When there is the potential for a group of time series to exhibit regime changes in the conditional mean, a model suited to capturing such behaviour is required. One such model is the multivariate extension of the STAR model. The multivariate extension of the autoregressive model is generally referred to as the vector AR model. Similarly, the multivariate extension of the STAR model is often referred to as the vector STAR model. However, in order to ensure consistency with the nomenclature surrounding multivariate GARCH models, we shall refer to it as the Multivariate STAR (M-STAR) model throughout this thesis.

As with the VAR model in Section 4.1, each data point $x_t$ is a $p \times 1$ vector, which represents the values of $p$ different time series at time $t$. The following two regime parametrisation of the M-STAR model was introduced in (2.1.6):

$$x_t = \left[ \tilde{B}_1^T + \tilde{F}_t \tilde{B}_2^T \right] \tilde{z}_t + \varepsilon_t,$$
where
\[ \tilde{B}_1^T = [\Phi_{1,1}, \ldots, \Phi_{1,k}] \]
\[ \tilde{B}_2^T = [\Phi_{2,1}, \ldots, \Phi_{2,k}] \]
\[ \tilde{F}_t = \begin{bmatrix} F_{1,t}(s_{1,t-d}, \gamma_1, c_1) & 0 \\ \vdots & \ddots \\ 0 & F_{p,t}(s_{p,t-d_p}, \gamma_{p_p}, c_{p_p}) \end{bmatrix} \]
\[ \tilde{z}_t = [x_{t-1}^T, \ldots, x_{t-k}^T]^T. \]

The matrix \( \tilde{F}_t \) is a diagonal matrix whose diagonal entries are transition functions for each time series within the multivariate data set. A common simplification of the model is to assume a common transition function across the \( p \) time series (Weise 1999; Dijk, Teräsvirta, and Franses 2002; Camacho 2004). That is,
\[ F_t(s_{t-d}, \gamma, c) = F_{1,t}(s_{1,t-d}, \gamma_1, c_1) \equiv \cdots \equiv F_{p,t}(s_{p,t-d_p}, \gamma_{p_p}, c_{p_p}). \]

With this in mind, the form of the model may be manipulated as follows:
\[ x_t = \left[ [\Phi_{1,1}, \ldots, \Phi_{1,k}] + \tilde{F}_t [\Phi_{2,1}, \ldots, \Phi_{2,k}] \right] \tilde{z}_t + \varepsilon_t \]
\[ = \left[ [\Phi_{1,1}, \ldots, \Phi_{1,k}] + F_t(s_{t-d}, \gamma, c) I_p [\Phi_{2,1}, \ldots, \Phi_{2,k}] \right] \tilde{z}_t + \varepsilon_t \]
\[ = [\Phi_{1,1}, \ldots, \Phi_{1,k}] \tilde{z}_t + \tilde{F}_t(s_{t-d}, \gamma, c) I_p [\Phi_{2,1}, \ldots, \Phi_{2,k}] \tilde{z}_t + \varepsilon_t \]
\[ = [\Phi_{1,1}, \ldots, \Phi_{1,k}] [\Phi_{2,1}, \ldots, \Phi_{2,k}] \tilde{z}_t, \tilde{F}_t(s_{t-d}, \gamma, c) + \varepsilon_t \]
\[ = [\Phi_{1,1}, \ldots, \Phi_{1,k}] [\Phi_{2,1}, \ldots, \Phi_{2,k}] \tilde{z}_t, \tilde{F}_t(s_{t-d}, \gamma, c) + \varepsilon_t \]
\[ = [\Phi_{1,1}, \ldots, \Phi_{1,k}] F_t(s_{t-d}, \gamma, c) + \varepsilon_t \]
\[ = \Phi^T \tilde{z}_t + \varepsilon_t. \]

Other parametrisations of the simplified M-STAR model exist, but a simple redefinition of the coefficient parameters will convert them to the form shown in (4.2.22). The scalar \( F_t \) in (4.2.22) is a single common transition function at time \( t \). This function was previously defined in Section 3.1 as
\[ F_t(s_{t-d}, \gamma, c) = \frac{1}{1 + \exp \left[ -\gamma \frac{1}{S(s_{t-d})} (s_{t-d} - c) \right]}. \]

It is assumed that the error \( \varepsilon_t \) for the process is a \( p \times 1 \) vector that it distributed via a multivariate normal distribution, that is, \( \varepsilon_t \sim \mathcal{N}(0_p, \Sigma_p) \).

The parameters of the transition function are defined in a similar manner to those for the univariate STAR model in Section 3.1. The transition variable \( s_{t-d} \) can either be an endogenous or an exogenous
variable, lagged by the variable $d \in D = \{1, \ldots, d_{\max}\}$. The smoothing and location parameters are \( \gamma \in \mathbb{R}^+ \) and \( c \in \mathbb{R} \), respectively.

Assuming that the transition functions are equal across the \( p \) time series simplifies the estimation procedure by reducing the number of parameters that must be estimated. It also allows the manipulation of the model equation (4.2.22) into a form consistent with the one used for the STAR model in Section 3.1.

Therefore, in matrix form, the M-STAR model with a single transition function can be expressed as follows:

\[
X = Z_1 \Phi + E,
\]

where

\[
X = \begin{bmatrix}
x_{s+1}^T \\
x_{s+2}^T \\
\vdots \\
x_N^T
\end{bmatrix}, \quad E = \begin{bmatrix}
epsilon_{s+1}^T \\
epsilon_{s+2}^T \\
\vdots \\
epsilon_N^T
\end{bmatrix}, \quad \Phi = \begin{bmatrix}
\Phi_{1,1}^T \\
\Phi_{1,2}^T \\
\vdots \\
\Phi_{2,k}^T
\end{bmatrix}, \quad Z_1 = \begin{bmatrix}
z_{s+1}^T \\
z_{s+2}^T \\
\vdots \\
z_N^T
\end{bmatrix} = \begin{bmatrix}
x_s^T & \ldots & x_1^T & x_1^T F_{s+1} & \ldots & x_1^T F_{s+1} \\
x_{s+1}^T & \ldots & x_2^T & x_2^T F_{s+2} & \ldots & x_2^T F_{s+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{N-1}^T & \ldots & x_{N-k}^T & x_{N-k}^T F_N & \ldots & x_{N-k}^T F_N
\end{bmatrix}.
\]

The matrices in (4.2.23) have the following dimensions: \( X \in \mathbb{R}^{n \times p} \), \( E \in \mathbb{R}^{n \times p} \), \( \Phi \in \mathbb{R}^{2kp \times p} \) and \( Z_1 \in \mathbb{R}^{n \times 2kp} \). The length of the useful data is \( n = N - s \), where \( s = \max(k, d) \).

### 4.2.1 Likelihood Function

The derivation of the likelihood is similar to the one performed in Section 4.1.1. However, we now have a new definition of \( z_t \) and, hence, a new definition of the design matrix \( Z_1 \). To derive the likelihood function, we use the fact that the vector \( \varepsilon_t \) is distributed via a multivariate normal distribution. The likelihood function will therefore be the product of \( n \) multivariate normal densities, with mean \( \Phi^T z_t \) and covariance matrix \( \Sigma \). If we let \( s = \max(k, d) \), the likelihood can be calculated using the first \( s \) data points as initial values. The useful data for this calculation therefore has indices \( s + 1 \) to \( N \) and length \( n = N - s \).
The likelihood, conditional on the first $s$ data points, is
\[
p(X|M_k, \Phi, \Sigma, d, \gamma, c) = \prod_{t=(s+1)}^N p(x_t|M_k, \Phi, \Sigma, d, \gamma, c)
= \prod_{t=(s+1)}^N \left(2\pi\right)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left(x_t - \Phi^T z_t\right)^T \Sigma^{-1} \left(x_t - \Phi^T z_t\right)\right]
= \left(2\pi\right)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left[\frac{1}{2} \sum_{t=(s+1)}^N \left(x_t - \Phi^T z_t\right)^T \Sigma^{-1} \left(x_t - \Phi^T z_t\right)\right].
\] (4.2.24)

The sum in the exponent in (4.2.24) can be manipulated to form
\[
\sum_{t=(s+1)}^N \left(x_t - \Phi^T z_t\right)^T \Sigma^{-1} \left(x_t - \Phi^T z_t\right) = \text{tr} \left( [X - Z_1\Phi]^T [X - Z_1\Phi] \Sigma^{-1} \right).
\] (4.2.25)

The details of this manipulation are similar to those presented in (4.1.4), with the only difference being that we now have new definitions of $Z_1$ and $\Phi$. The likelihood function is therefore
\[
p(X|M_k, \Phi, \Sigma, d, \gamma, c) = \left(2\pi\right)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} \text{tr} \left( [X - Z_1\Phi]^T [X - Z_1\Phi] \Sigma^{-1} \right)\right].
\] (4.2.26)

### 4.2.2 Prior Distribution

The joint prior distribution used for the coefficient parameter matrix and the covariance matrix of the error term will be the same as that used in Section 4.1. We shall decompose it as follows:
\[
p(M_k, \Phi, \Sigma, d, \gamma, c) = p(M_k, \Phi, \Sigma|d, \gamma, c) p(d, \gamma, c)
= p(M_k, \Phi, \Sigma|d, \gamma, c) p(d) p(\gamma) p(c).
\] (4.2.27)

The joint prior distribution for the model index, coefficient parameter matrix and covariance of the error shown on the right hand side of (4.2.27) is defined as for the VAR models in Section 4.1.2. In addition, the prior distributions for the implicit parameters are the same as those used for the univariate STAR model. The individual prior distributions are
\[
p(M_k, \Phi, \Sigma|d, \gamma, c) \propto \left[\frac{A_k}{k!}\right]^{\gamma_k} I_K(k) |\Sigma|^{-\frac{n+1}{2}}
p(d) \propto I_D(d)
p(\gamma) \sim G(\alpha_\gamma, \beta_\gamma)
p(c) \sim N(\mu_c, \sigma_c^2).
\] (4.2.28)
4.2.3 Posterior Distribution

Combining the likelihood function from (4.2.26) with the prior distributions in (4.2.28) results in the following posterior distribution:

\[
p(M_k, \Phi, \Sigma, d, \gamma, c | X) \\
\propto p(M_k | M_k, \Phi, \Sigma, d, \gamma, c) p(M_k, \Phi, \Sigma) d (\gamma) p(c) *
\]

\[
\propto (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k} I_k (k) \times \]

\[
|\Sigma|^{-\frac{(a+1)}{2}} \frac{\beta_d^a}{\Gamma(\alpha_d)} \gamma^{(a_d, -1)} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k} \times \]

\[
\frac{\beta_k^d}{\Gamma(\alpha_k)} \gamma^{(a_k, -1)} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k}.
\]

Restricting the possible values that \( k \) and \( d \) can take when coding the MCMC algorithm allows us to ignore the indicator functions in the joint posterior distribution.

If we assume that all other parameters are constants, we can derive the following joint conditional posterior distribution, \( p(M_k, \Phi | X, \Sigma, d, \gamma, c) \), for the model index and coefficient parameter matrix:

\[
p(M_k, \Phi | X, \Sigma, d, \gamma, c) \\
\propto (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{(a+1)}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k} \times \]

\[
\frac{\beta_d^a}{\Gamma(\alpha_d)} \gamma^{(a_d, -1)} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k} \times \]

\[
\frac{\beta_k^d}{\Gamma(\alpha_k)} \gamma^{(a_k, -1)} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k}.
\]

The derivation of the full conditional posterior of the coefficient parameter matrix follows the same lines as the derivation performed in Section 4.1.3. Hence,

\[
p(\Phi | X, M_k, \Sigma, d, \gamma, c) \\
\propto (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{(a+1)}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k} \times \]

\[
\frac{\beta_d^a}{\Gamma(\alpha_d)} \gamma^{(a_d, -1)} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k} \times \]

\[
\frac{\beta_k^d}{\Gamma(\alpha_k)} \gamma^{(a_k, -1)} \exp \left[ -\frac{1}{2} \text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{A_k^2}{k!} \right]^{T_k}. \tag{4.2.29}
\]
In order to simplify the posterior distribution, we focus on the trace in the exponent:

\[
\text{tr} \left( [X - Z_1 \Phi]^T [X - Z_1 \Phi] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T - \Phi^T Z_1^T] [X - Z_1 \Phi] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T X - X^T Z_1 \Phi - \Phi^T Z_1^T X + \Phi^T Z_1^T \Phi] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T X + \Phi^T Z_1^T Z_1 \Phi - \Phi^T Z_1^T X - X^T Z_1 \Phi] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T X + \Phi^T (Z_1^T Z_1 \Phi - Z_1^T X) - X^T Z_1 \Phi] \Sigma^{-1} \right).
\]

Setting \( C_{\Phi}^{-1} = Z_1^T Z_1 \) yields

\[
\text{tr} \left( [X^T X + \Phi^T (Z_1^T Z_1 \Phi - Z_1^T X) - X^T Z_1 \Phi] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T X + \Phi^T (C_{\Phi}^{-1} \Phi - Z_1^T X) - X^T Z_1 \Phi] \Sigma^{-1} \right) \\
= \text{tr} \left( [X^T X + \Phi^T C_{\Phi}^{-1} (\Phi - C_\phi Z_1^T X) - X^T Z_1 \Phi] \Sigma^{-1} \right).
\]

Moreover, the trace in the exponent may be simplified further by setting \( \hat{\Phi} = C_\phi Z_1^T X \). The subsequent manipulation is symbolically the same as the one performed for VAR models in Appendix D.1. The only difference here is the redefinition of the design and coefficient parameter matrices.

Substituting the result of the manipulation from Appendix D.1 into the conditional posterior distribution for \( \Phi \) yields

\[
p(\Phi | X, M, k, \Sigma, d, \gamma, c) \\
\propto (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{np+1}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( (\Phi - \hat{\Phi})^T C_{\Phi}^{-1} (\Phi - \hat{\Phi}) \Sigma^{-1} \right) \right] + \text{exp} \left[ -\frac{1}{2} \text{tr} \left( (\Phi - \hat{\Phi})^T C_{\Phi}^{-1} (\Phi - \hat{\Phi}) \Sigma^{-1} \right) \right] \\
\propto (2\pi)^{\frac{2np}{2}} |\Sigma|^{\frac{2np}{2}} |C_{\Phi}|^{\frac{np}{2}}.
\]

The conditional posterior distribution for \( \Phi \) shown in (4.2.30) is proportional to a Matrix Normal distribution:

\[
p(\Phi | X, M, k, \Sigma, d, \gamma, c) \sim N_{2k,p} \left( \Phi, C_{\Phi}, \Sigma \right),
\]

where

\[\Phi = C_\phi Z_1^T X \]

\[C_{\Phi}^{-1} = Z_1^T Z_1.\]
The conditional posterior distribution for $\Sigma$ is derived as follows:

$$p(\Sigma | \mathbf{X}, \mathcal{M}_k, \Phi, d, \gamma, c)$$

$$\propto (2\pi)^{-\frac{n_p}{2}} |\Sigma|^{-\frac{n_p + 1}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( [\mathbf{X} - \mathbf{Z}_1 \Phi]^T [\mathbf{X} - \mathbf{Z}_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{\Lambda_k^2}{k!} \right]^{r_k} \times$$

$$\frac{\beta_{\gamma}^{\alpha \gamma}}{\Gamma(\alpha \gamma)} \gamma^{(\alpha \gamma - 1)} \exp \left[ -\beta_{\gamma} \gamma \right] \frac{2 \sigma_c^2 \pi}{2} \exp \left[ -\frac{1}{2 \sigma_c^2} (c - \mu_c)^2 \right] (4.2.31)$$

The distribution in (4.2.31) is proportional to an Inverted Wishart distribution. The conditional posterior distributions for the implicit parameters $d$, $\gamma$, and $c$ are all of non-standard forms, while the posterior for the delay parameter is simply a discrete distribution that is proportional to the likelihood function. We shall use an embedded Metropolis-Hastings step to simulate the smoothing and location parameters from their joint conditional posterior distribution $p(\gamma, c | \mathbf{X}, \mathcal{M}_k, \Phi, \Sigma, d)$. The joint conditional posterior distribution is given by

$$p(\gamma, c | \mathbf{X}, \mathcal{M}_k, \Phi, \Sigma, d)$$

$$\propto (2\pi)^{-\frac{n_p}{2}} |\Sigma|^{-\frac{n_p + 1}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( [\mathbf{X} - \mathbf{Z}_1 \Phi]^T [\mathbf{X} - \mathbf{Z}_1 \Phi] \Sigma^{-1} \right) \right] \left[ \frac{\Lambda_k^2}{k!} \right]^{r_k} \times$$

$$\frac{\beta_{\gamma}^{\alpha \gamma}}{\Gamma(\alpha \gamma)} \gamma^{(\alpha \gamma - 1)} \exp \left[ -\beta_{\gamma} \gamma \right] \frac{2 \sigma_c^2 \pi}{2} \exp \left[ -\frac{1}{2 \sigma_c^2} (c - \mu_c)^2 \right] \times$$

$$\frac{\beta_{\gamma}^{\alpha \gamma}}{\Gamma(\alpha \gamma)} \gamma^{(\alpha \gamma - 1)} \exp \left[ -\beta_{\gamma} \gamma \right] \frac{2 \sigma_c^2 \pi}{2} \exp \left[ -\frac{1}{2 \sigma_c^2} (c - \mu_c)^2 \right].$$

Therefore, in summary, the conditional posterior distributions that will be employed in the estimation algorithm are as follows:

$$p(\Phi | \mathbf{X}, \mathcal{M}_k, \Sigma, d, \gamma, c) \sim \mathcal{MN} \left( \Phi, C_{\Phi}, \Sigma \right)$$

$$p(\Sigma | \mathbf{X}, \mathcal{M}_k, \Phi, d, \gamma, c) \sim \mathcal{IW} (\Psi, n)$$

$$p(d | \mathbf{X}, \mathcal{M}_k, \Phi, \Sigma, \gamma, c) \propto \exp \left[ -\frac{1}{2} \text{tr} \left( [\mathbf{X} - \mathbf{Z}_1 \Phi]^T [\mathbf{X} - \mathbf{Z}_1 \Phi] \right) \right] (4.2.32)$$

$$p(\gamma, c | \mathbf{X}, \mathcal{M}_k, \Phi, \Sigma, d) \propto \exp \left[ -\frac{1}{2} \text{tr} \left( [\mathbf{X} - \mathbf{Z}_1 \Phi]^T [\mathbf{X} - \mathbf{Z}_1 \Phi] \Sigma^{-1} \right) \right] \frac{\beta_{\gamma}^{\alpha \gamma}}{\Gamma(\alpha \gamma)} \times$$

$$\gamma^{(\alpha \gamma - 1)} \exp \left[ -\beta_{\gamma} \gamma \right] \frac{2 \sigma_c^2 \pi}{2} \exp \left[ -\frac{1}{2 \sigma_c^2} (c - \mu_c)^2 \right],$$

where

$$\Phi = C_{\Phi} \mathbf{Z}_1^T \mathbf{X}$$

$$C_{\Phi}^{-1} = \mathbf{Z}_1^T \mathbf{Z}_1$$

$$\Psi = [\mathbf{X} - \mathbf{Z}_1 \Phi]^T [\mathbf{X} - \mathbf{Z}_1 \Phi].$$
4.2.4 Posterior Simulator

The construction of the M-STAR estimation algorithm is similar to that for the VAR model. There are some additional steps in the algorithm to account for the parameters unique to the M-STAR model. Again, the inclusion of the M-STAR model order as a parameter for estimation leads to the need for an RJMCMC scheme. The structure of the main algorithm is a Gibbs sampler, as shown below.

M-STAR - Main Algorithm

1. Initialise \((M_k^{(0)}, \Phi^{(0)}, \Sigma^{(0)}, d^{(0)}, \gamma^{(0)}, c^{(0)})\) either randomly or deterministically and set \(i = 1\).

2. At the \(i\)th iteration, execute the following steps employing the conditional posterior distributions shown in (4.2.32):

   (a) Simulate the covariance matrix of the error vector using 
   
   \[
   \Sigma^{(i)} \sim p \left( \Sigma | X, M_k^{(i-1)}, \Phi^{(i-1)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)} \right).
   \]

   (b) (i) RJ Step - Simulate the STAR model order and coefficient parameter matrix using 
   
   \[
   \left( M_k^{(i)}, \Phi^{(i)} \right) \sim p \left( M_k, \Phi | X, \Sigma^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)} \right).
   \]

   (ii) If the proposal in step 2(b)(i) is rejected, simulate the coefficient parameter matrix using 
   
   \[
   \Phi^{(i)} \sim p \left( \Phi | X, M_k^{(i)}, \Sigma^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)} \right).
   \]

   (c) Simulate the delay parameter \(d\) of the transition function using 
   
   \[
   d^{(i)} \sim p \left( d | X, M_k^{(i)}, \Phi^{(i)}, \Sigma^{(i)}, \gamma^{(i-1)}, c^{(i-1)} \right).
   \]

   (d) MH\((\gamma, c)\) Step - Simulate the smoothing and location parameters of the transition function using 
   
   \[
   \left( \gamma^{(i)}, c^{(i)} \right) \sim p \left( \gamma, c | X, M_k^{(i)}, \Phi^{(i)}, \Sigma^{(i)}, d^{(i)} \right).
   \]

3. Set \(i = i + 1\). If all the MCMC chains have converged, STOP, else, move to step 2.

The embedded Reversible Jump step in 2(b)(i) and the embedded Metropolis-Hastings step in 2(d) are outlined below.
Reversible Jump - RJ\textsubscript{M} Step

Recall the joint conditional posterior distribution $p(\mathcal{M}_k, \Phi | X, \Sigma, d, \gamma, c)$ from (4.2.32). Given that a change in the model index $\mathcal{M}_k$ results in a change in the dimension of the coefficient matrix $\Phi$, we shall require an RJMCMC algorithm here.

Recall from Section 2.2.4 that the form of the acceptance probability used throughout this thesis for a Reversible Jump step is

$$r = \min \left( 1, \frac{\pi(\theta^*_m, m^* | x) q(\theta^*, m, m^*)}{\pi(\theta_m, m | x) q(\theta^* | \theta, m, m^*)} \right).$$

Using the notation for the STAR model, the acceptance probability for the M-STAR algorithm has the following form:

$$r_{(\mathcal{M}_k, \Sigma)} = \min \left( 1, \frac{p(\mathcal{M}_k, \Phi^* | X, \Sigma, d, \gamma, c) q(\Phi^*, \mathcal{M}_k, \mathcal{M}_k^*)}{p(\mathcal{M}_k, \Phi | X, \Sigma, d, \gamma, c) q(\Phi^* | \Phi, \mathcal{M}_k, \mathcal{M}_k^*)} \right).$$

(4.2.33)

We can again use the Candidate’s Identity to simplify the acceptance probability. This simplification is similar to the one performed in Section 4.1.4. Beginning with the Candidate’s Identity, and employing the notation for the target distribution for the M-STAR model, we obtain

$$p(\mathcal{M}_k, \Phi | X, \Sigma, d, \gamma, c) = \frac{p(\mathcal{M}_k, \Phi | X, \Sigma, d, \gamma, c) p(\mathcal{M}_k | \Sigma, d, \gamma, c)}{p(\mathcal{M}_k | \Sigma, d, \gamma, c)}. \quad (4.2.34)$$

Using the expression in (4.2.34), the ratio of the target distributions becomes

$$\frac{p(\mathcal{M}_k, \Phi^* | X, \Sigma, d, \gamma, c)}{p(\mathcal{M}_k, \Phi | X, \Sigma, d, \gamma, c)} = \frac{p(X | \mathcal{M}_k^*, \Sigma, d, \gamma, c) p(\mathcal{M}_k^* | \Sigma, d, \gamma, c) p(\Phi^* | X, \mathcal{M}_k^*, \Sigma, d, \gamma, c)}{p(X | \mathcal{M}_k, \Sigma, d, \gamma, c) p(\mathcal{M}_k | \Sigma, d, \gamma, c) p(\Phi^* | X, \mathcal{M}_k^*, \Sigma, d, \gamma, c)} \quad \hat{=} \quad (4.2.35)$$

Now we define the proposal distribution for the coefficient matrix $\Phi$ to be the full conditional posterior distribution of $\Phi$ shown in equation (4.2.30). That is,

$$q(\Phi^* | \Phi, \mathcal{M}_k, \mathcal{M}_k^*) \sim p(\Phi^* | X, \mathcal{M}_k^*, \Sigma, d, \gamma, c).$$
Substituting the definition of the proposal distribution in (4.2.36), together with the ratio of the target distributions from (4.2.35), into the acceptance probability formula in (4.2.33) yields the following formula for the acceptance probability:

\[
\begin{align*}
\pi(x|\theta) & = \min \left( \min \left( \frac{p(\theta|x,\theta_{\text{old}})}{\pi(\theta|\theta_{\text{old}})} \right) \right)
\end{align*}
\]

(4.2.37)

In (4.2.37), \( p(M_k|\Sigma, d, \gamma, c) \) is simply a prior distribution for the model index, given the other parameters shown. The other term, \( p(X|M_k, \Sigma, d, \gamma, c) \), is a marginal likelihood and is derived as follows:

\[
p(X|M_k, \Sigma, d, \gamma, c) \propto \int p(X|M_k, \Phi, \Sigma, d, \gamma, c) p(\Phi|M_k, \Sigma, d, \gamma, c) d\Phi.
\]

(4.2.38)

The terms in (4.2.38) are the product of the likelihood function and the conditional prior distribution for \( \Phi \). This has already been calculated as shown in (4.2.30). As no part of the simplification procedure in that derivation assumed that \( k \) was constant, the marginal likelihood may be derived as follows:

\[
p(X|M_k, \Sigma, d, \gamma, c) \propto \int p(X|M_k, \Phi, \Sigma, d, \gamma, c) p(\Phi|M_k, \Sigma, d, \gamma, c) d\Phi.
\]

(4.2.39)
The acceptance probability in (4.2.41) is not a function of \( \Phi \). Substituting the prior model probability from (4.2.40) and the simplified marginal likelihood from (4.2.39) into the quotient on the right hand side of the acceptance probability yields

\[
p(M_k|\Sigma, d, \gamma, c) \propto \left[ \frac{\Lambda^k \pi_k}{k!} \right] I_k(k).
\]

Substituting the prior model probability from (4.2.40) and the simplified marginal likelihood from (4.2.39) into the quotient on the right hand side of the acceptance probability yields

\[
p(X|M_{k*}, \Sigma, d, \gamma, c) p(M_{k*}|\Sigma, d, \gamma, c)
\]

\[
= (2\pi)^{-\frac{np+2k^*+\beta}{2}} |\Sigma|^{-\frac{np+1+2k+p}{2}} |C_\Phi| \left[ \frac{2}{\pi} \right] \left[ \frac{\Lambda^{k*}}{k^{*+1}} \right]^{\tau_k} I_k(k)
\]

\[
= (2\pi)^{-\frac{2p^2(k^*-k)}{2}} |\Sigma|^{-\frac{p}{2}(k^*-k)} \left[ C_\Phi^* |C_\Phi|^{-1} \right]^{\frac{2}{\pi}} \left[ \frac{\Lambda^{k^*-k}}{k^{*+1}} \right]^{\tau_k} \times
\]

\[
\exp \left[ -\frac{1}{2} \text{tr} \left( \left[ X^T X - \hat{\Phi}^* C_\Phi^{-1} \Phi^* \right] - \left[ X^T X - \hat{\Phi} C_\Phi^{-1} \Phi \right] \right) \right]
\]

\[
= (2\pi)^{-\frac{2p^2(k^*-k)}{2}} |\Sigma|^{-\frac{p}{2}(k^*-k)} \left[ C_\Phi^* |C_\Phi|^{-1} \right]^{\frac{2}{\pi}} \left[ \frac{\Lambda^{k^*-k}}{k^{*+1}} \right]^{\tau_k} \times
\]

\[
\exp \left[ -\frac{1}{2} \text{tr} \left( \left[ X^T X - \hat{\Phi}^* C_\Phi^{-1} \Phi^* \right] - \left[ X^T X - \hat{\Phi} C_\Phi^{-1} \Phi \right] \right) \right].
\]

Therefore, we have the following acceptance probability formula for making a model move to candidate model \( M_{k*} \):

\[
r_{(M_k, M_{k*})} = \min \left( 1, \frac{1}{(2\pi)^{\frac{2p^2(k^*-k)}{2}} |\Sigma|^{-\frac{p}{2}(k^*-k)} \left[ C_\Phi^* |C_\Phi|^{-1} \right]^{\frac{2}{\pi}} \left[ \frac{\Lambda^{k^*-k}}{k^{*+1}} \right]^{\tau_k} \times
\]

\[
\exp \left[ -\frac{1}{2} \text{tr} \left( \left[ X^T X - \hat{\Phi}^* C_\Phi^{-1} \Phi^* \right] - \left[ X^T X - \hat{\Phi} C_\Phi^{-1} \Phi \right] \right) \right] \right).
\]

The acceptance probability shown in (4.2.41) is for \( p(\Phi, M_k|X, \Sigma, d, \gamma, c) \). As with the univariate case, the acceptance probability in (4.2.41) is not a function of \( \Phi \), and therefore does not require the simulation of a random \( \Phi \). This saves computational resources, and therefore time, in the implementation.
of the algorithm. However, after a model move is either accepted or rejected, we must simulate a coefficient matrix from the full conditional distribution for $\Phi$, conditional on the current model order. Therefore, our RJMCMC algorithm at 2(b) will be as outlined below.

**RJM Algorithm**

1. The current state is $(\mathcal{M}_k^{(i-1)}, \Phi^{(i-1)}, \Sigma^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)})$.

2. Propose a candidate model index, $\mathcal{M}_{k^*}$ from $\mathcal{J}(\mathcal{M}_{k^*}|\mathcal{M}_k)$ using the discretised Laplacian shown in (2.2.17).

3. Calculate the acceptance probability, $r(\mathcal{M}_k, \mathcal{M}_{k^*})$, per (4.2.41).

4. Simulate $u \sim \mathcal{U}(0,1)$:
   
   (a) If $r(\mathcal{M}_k, \mathcal{M}_{k^*}) > u$, then:
      
      (i) Accept the candidate model index $\mathcal{M}_{k^*}$ and set $\mathcal{M}_k^{(i)} = \mathcal{M}_{k^*}$.
      
      (ii) Simulate the coefficient parameter matrix using the full conditional posterior distribution
      $$\Phi^{(i)} \sim p\left(\Phi|X, \mathcal{M}_k^{(i)}, \Sigma^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)}\right).$$
   
   (b) If $r(\mathcal{M}_k, \mathcal{M}_{k^*}) < u$, then:
      
      (i) Reject the candidate model order $\mathcal{M}_{k^*}$ and set $\mathcal{M}_k^{(i)} = \mathcal{M}_k^{(i-1)}$.
      
      (ii) Simulate the coefficient parameter matrix using the full conditional posterior distribution as in step 2(b)(ii) of the M-STAR - Main Algorithm.

**Metropolis-Hastings - MH$(_{(\gamma,c)}$) Step**

For the formulation of the M-STAR model used here, the simulation of the smoothing and location parameters for the transition function is similar to that performed in the univariate case. The key difference is that the likelihood function is defined on a multivariate data set. Below, we outline the Metropolis-Hastings step for the simulation of the smoothing and location parameters.

The proposal distributions for this step are the same as those used in Section 3.1. That is,\footnote{4.2.42}

$$q(\gamma^*|\gamma) \sim \mathcal{G}\left(\frac{\gamma^2}{\Delta\gamma}, \frac{\gamma}{\Delta\gamma}\right)$$

$$q\left(c^*|c\right) \sim \mathcal{N}\left(c, \Delta_c\right),$$

where the most recently simulated values are $\gamma = \gamma^{(i-1)}$ and $c = c^{(i-1)}$. 

---

91
Recall from Section 2.2.2 that the general form of the acceptance probability for a Metropolis-Hastings algorithm is as follows:

$$r = \min \left( 1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^*|\theta^{(i-1)})} \right)$$

(4.2.43)

Substituting the target distribution from (4.2.32) and the density of the proposal distributions from (4.2.42) into the quotient on the right hand side of the acceptance probability in (4.2.43) results in the following calculation, required for the Metropolis-Hastings step:

$$\frac{p(\gamma^*, c^*|X, M_k, \Phi, \Sigma, d) q(c|c^*) q(\gamma|\gamma^*)}{p(\gamma, c|X, M_k, \Phi, \Sigma, d) q(c|c) q(\gamma|\gamma^*)} = \frac{\exp \left[ -\frac{1}{2} \text{tr} \left( (X - Z_1 \Phi)^T (X - Z_1 \Phi) \Sigma^{-1} \right) \right] \gamma^*(\alpha, \gamma) \exp \left[ -\frac{1}{2\sigma_c^2} (c^* - \mu_c)^2 \right]}{\exp \left[ -\frac{1}{2} \text{tr} \left( (X - Z_1 \Phi)^T (X - Z_1 \Phi) \Sigma^{-1} \right) \right] \gamma^*(\alpha, \gamma) \exp \left[ -\frac{1}{2\sigma_c^2} (c - \mu_c)^2 \right]}$$

$$= \frac{\left( \frac{2^*}{\gamma} \right)^{(\alpha, \gamma)} \left[ \frac{\gamma^{\frac{2}{2}}}{\Gamma \left( \frac{2}{2} \right)} \gamma^* \left( \gamma^{\frac{2}{2}} - 1 \right) \right]^{-1}}{\left( \frac{\gamma^{2}}{\alpha^{2}} \right)^{\frac{2}{2}} \left[ \frac{\gamma^* \gamma^{\frac{2}{2}}}{\Gamma \left( \frac{2}{2} \right)} \right]} \times \exp \left[ -\frac{1}{2} \text{tr} \left( (X - Z_1 \Phi)^T (X - Z_1 \Phi) \Sigma^{-1} \right) \right] \exp \left[ -\beta_\gamma (\gamma^* - \gamma) \right] \exp \left[ -\frac{1}{2\sigma_c^2} (c^* - \mu_c)^2 + \frac{1}{2\sigma_c^2} (c - \mu_c)^2 \right]$$

$$= \frac{\left( \frac{2^*}{\gamma} \right)^{(\alpha, \gamma)} \left[ \frac{\gamma^{\frac{2}{2}}}{\Gamma \left( \frac{2}{2} \right)} \gamma^* \left( \gamma^{\frac{2}{2}} - 1 \right) \right]^{-1}}{\left( \frac{\gamma^{2}}{\alpha^{2}} \right)^{\frac{2}{2}} \left[ \frac{\gamma^* \gamma^{\frac{2}{2}}}{\Gamma \left( \frac{2}{2} \right)} \right]} \times \exp \left[ -\frac{1}{2} \text{tr} \left( (X - Z_1 \Phi)^T (X - Z_1 \Phi) \Sigma^{-1} \right) - \frac{1}{2} \text{tr} \left( (X - Z_1 \Phi)^T (X - Z_1 \Phi) \Sigma^{-1} \right) - 2\beta_\gamma (\gamma^* - \gamma) + \frac{1}{\sigma_c^2} \left[ (c^* - \mu_c)^2 - (c - \mu_c)^2 \right] \right].$$
The acceptance probability is

\[
r_{(\gamma,c,MS)} = \min \left( 1, \left[ \frac{\gamma^{\gamma}}{\gamma} \right] (\alpha,\gamma-1) \left[ \frac{(\gamma^2)^{\frac{\gamma^2}{2\tau}}}{\gamma} \Gamma \left( \frac{\gamma^2}{2\tau} \right) \gamma^{\left( \frac{\gamma^2}{2\tau} \right)} \right] \times \exp \left[ -\frac{1}{2} \left( \text{tr} \left( [X - Z_1\Phi]^T [X - Z_1\Phi] \Sigma^{-1} - [X - Z_1\Phi]^T [X - Z_1\Phi] \Sigma^{-1} \right) + 2\beta_\gamma (\gamma^* - \gamma) + \frac{1}{\sigma_c^2} \left( \left( c^* - \mu_c \right)^2 - \left( c - \mu_c \right)^2 \right) \right] \right) \right).\]

Now that the Metropolis-Hastings step is finalised, we can complete the outline of the algorithm.

**MH\((\gamma,c)\) Algorithm**

1. The current state is \( (M_{k(i)}, \Phi(i), \Sigma(i), d(i), \gamma^{(i-1)}, c^{(i-1)}) \).

2. Propose new smoothing and location parameters \( \gamma^* \) and \( c^* \) from their respective proposal distributions as shown in (4.2.42).

3. Calculate the acceptance probability \( r_{(\gamma,c,MS)} \) per (4.2.45).

4. Simulate \( u \sim \mathcal{U}(0,1) \):
   
   (a) If \( r_{(\gamma,c,MS)} > u \), accept the proposed smoothing and location parameters. Set \( \gamma^{(i)} = \gamma^* \) and \( c^{(i)} = c^* \).
   
   (b) If \( r_{(\gamma,c,MS)} < u \), reject the proposed smoothing and location parameters. Set \( \gamma^{(i)} = \gamma^{(i-1)} \) and \( c^{(i)} = c^{(i-1)} \).

Now that we have completed the details of the estimation scheme for an M-STAR model, we turn our attention to M-GARCH models. The next section develops the machinery needed to create an estimation algorithm that includes model selection for some M-GARCH models.

### 4.3 Multivariate GARCH

There are several extensions of the univariate GARCH model to the multivariate setting, differentiated by the formulations of their conditional covariance equations. Each has its own strengths and weaknesses. The two classes of M-GARCH models that we shall discuss are the Vech and BEKK
formulations. Before looking at specific formulations of the covariance equations, we shall present the basic structure of M-GARCH models.

As for the M-STAR model, our aim is to fit an M-GARCH model $N_j$ to a set of data, $X$. We shall use a $p \times 1$ vector, denoted by $x_t$, to represent the values of $p$ different time series at time $t$. The following structure of an M-GARCH model with a conditional mean term added is similar to the representation introduced in (2.1.7):

$$x_t = \mu_t + \varepsilon_t$$

$$\varepsilon_t = H_t^{\frac{1}{2}} \eta_t.$$  

The conditional mean of the process is $\mu_t$ which, for now, will be assumed to be zero across the $p$ time series. We let $\eta_t$ denote an independent identically distributed random vector of length $p$ such that $E[\eta_t] = 0$ and $E[\eta_t \eta_t^T] = I_p$. Furthermore, $H_t \in \mathbb{R}^{p \times p}$ is the conditional covariance matrix at time $t$ and is required to be positive definite for all $t$, and the $p \times 1$ vector $\varepsilon_t$ is a vector of errors for the process. Under the assumption that $\eta_t$ is distributed via a multivariate normal distribution, the errors of the process follow a multivariate normal distribution. That is, $\varepsilon_t \sim N_p(0, H_t)$.

To derive the likelihood function and posterior distributions, it is best to write the model in matrix form as follows:

$$X = U + E$$

$$\text{vec}(E^T) = H^{\frac{1}{2}} \eta,$$

where

$$X = \begin{bmatrix} x_{s+1}^T \\ x_{s+2}^T \\ \vdots \\ x_{N}^T \end{bmatrix}, \quad U = \begin{bmatrix} 0^T \\ 0^T \\ \vdots \\ 0^T \end{bmatrix}, \quad E = \begin{bmatrix} \varepsilon_{s+1}^T \\ \varepsilon_{s+2}^T \\ \vdots \\ \varepsilon_{N}^T \end{bmatrix}.$$  

(4.3.46)

The matrix $H \in \mathbb{R}^{np \times np}$ in (4.3.46) is a block diagonal matrix in which each $H_i$ is calculated using the conditional covariance equation relevant to the M-GARCH model being employed. The other matrices in (4.3.46) have the following dimensions: $X \in \mathbb{R}^{n \times p}$ and $E \in \mathbb{R}^{n \times p}$. In this section, the relevant model for the conditional covariance equation will either be a VECH or a BEKK model. These two formulations are described below.
4.3.1 VECH Conditional Covariance Equation

The formulation of the VECH model is an extension of the univariate GARCH model to a multivariate setting. Recall from (2.1.8) that

\[
\text{vech} (H_t) = \bar{c} + \sum_{i=1}^{l} A_i \text{vech} (\varepsilon_{t-i} \varepsilon_{t-i}^T) + \sum_{j=1}^{m} B_j \text{vech} (H_{t-j}),
\]

where \(A_i, B_j\) are the \(p(p+1)/2 \times p(p+1)/2\) coefficient parameter matrices, and \(\bar{c}\) is an intercept vector with length \(p(p+1)/2\). The parameters \(l\) and \(m\) are the model orders for the ARCH and GARCH parts of the model, respectively.

As in the univariate GARCH model, the covariance stationarity conditions of the model are met when all eigenvalues of the matrix

\[
\sum_{i=1}^{l} A_i + \sum_{j=1}^{m} B_j,
\]

have modulus less than one.

The advantages of using the VECH formulation are that the model is sufficiently general, and the interpretation of the parameters \(A, B\) and \(\bar{c}\) is fairly straightforward. However, the number of parameters requiring estimation is \((l + m) \left(\frac{p(p+1)}{2}\right)^2 + \frac{p(p+1)}{2}\), which is large, even for small \(p, l\) and \(m\). A high level of computational power is required to estimate all parameters for the model.

Hudson and Gerlach (2008) proposed a novel approach to restrict the number of parameters requiring estimation and to ensure positive definiteness in the estimation process. In their approach, only the following parameters are non zero:

\[
a_{ii}, b_{ii}, \quad \text{for } i = 1, \ldots, p(p+1)/2
\]

\[
a_{ij}, b_{ij}, \quad \text{for } i \neq j = 1, p + 1, 2p, 3p - 2, 4p - 5, \ldots, p(p+1)/2.
\]

The sequence for \(a_{ij}\) and \(b_{ij}\) for \(i \neq j\) above is given by the following recurrence relationship

\[
r_1 = 1
\]

\[
r_n = r_{n-1} + p + 2 - n,
\]

which has closed form

\[
r_n = (n - 1)p - \frac{1}{2} (n - 3) n,
\]

where \(r_n\) is the \(n\)th term in the sequence.
For example, when $p = 3$, the ARCH coefficient matrices $A_1$ will have the form

$$A_1 = \begin{bmatrix}
    a_{11} & 0 & 0 & a_{14} & 0 & a_{16} \\
    0 & a_{22} & 0 & 0 & 0 & 0 \\
    0 & 0 & a_{33} & 0 & 0 & 0 \\
    a_{41} & 0 & 0 & a_{44} & 0 & a_{46} \\
    0 & 0 & 0 & 0 & a_{55} & 0 \\
    a_{61} & 0 & 0 & a_{64} & 0 & a_{66}
\end{bmatrix}.$$ 

For a VECH model with $l = m = 1$, the parametrisation of Hudson and Gerlach (2008) reduces the number of parameters requiring estimation to \( \frac{3}{2} p(p + 1) + 2p(p - 1) \). This amount is always less than the number of parameters of the full VECH model, but is larger than the number of parameters required for the BEKK model. However, this amount is usually much closer to the number of parameters for the BEKK model than to the number for the full VECH model. The conditional covariance at time $t$ takes into account the errors in the relevant series at time $(t - 1)$, while the equation for the conditional variance at time $t$ includes extra terms that allow squared shock spillovers between series and capture the volatility within the series.

In addition to the parameter restriction, Hudson and Gerlach (2008) used a prior distribution on the coefficient matrices that was equal to one when the resulting conditional covariance matrices $H_t$ were positive definite for all $t$. Otherwise, their prior distribution took the value of zero. This prior distribution enabled them to enforce positive definiteness of the covariance matrices for the model during the estimation scheme.

We created an estimation scheme similar, but not identical to, the one implemented in Hudson and Gerlach (2008), and tested it in a simulation study. The results suggested that the estimation algorithm was able to successfully find reasonable point estimates for the parameters. Unfortunately, the estimation procedure failed to estimate the parameters correctly too often for it to be considered effective.

We discuss the application of our VECH estimation scheme in Chapter 5 and provide some potential reasons for its failure. In summary, the key differences between the schemes arise from the use of a non-constant conditional mean, the number of steps in one iteration of the Gibbs sampler and the proposal distributions used.

Although it would be possible to address these key differences and deal with the problems that they cause, the VECH formulation does not guarantee positive definiteness of the covariance matrices. This dramatically reduces the speed of the algorithm, even for relatively small $p$. As both $p$ and $N$ become
large, the chances of rejecting parameters due to their lack of positive definiteness is greatly increased, further decreasing the speed of the algorithm. With this in mind, it was decided to explore another form of M-GARCH model, namely the BEKK formulation of the covariance equation.

### 4.3.2 BEKK Conditional Covariance Equation

Recall the conditional covariance equation of the BEKK formulation originally presented in (2.1.9):

\[
H_t = C_0 C_0^T + \sum_{q=1}^{Q} \sum_{i=1}^{l} A_{qi} \varepsilon_{t-i} \varepsilon_{t-i}^T A_{qi}^T + \sum_{q=1}^{Q} \sum_{j=1}^{m} B_{qj} H_{t-j} B_{qj}^T,
\]

(4.3.47)

where \( A_{qi}, B_{qj}, \) and \( C_0 \) are \( p \times p \) parameter matrices, with \( C_0 \) being lower triangular. The parameter \( Q \) is included in the model to allow for more general representations of the conditional covariance equation.

Silvennoinen and Teräsvirta (2009b) suggested that, for practical purposes, it is often assumed that \( l = m = Q = 1 \). When this is not the case, numerical complications commonly arise during parameter estimation. With this in mind, we shall restrict our model orders to those satisfying \( l = m = Q = 1 \).

In addition, (4.3.47) shall be redefined so that the intercept matrix is defined directly to be a positive definite matrix, \( C \in \mathbb{R}^{p \times p} \). In repeated trials, we found that it was very difficult to estimate the intercept matrix \( C_0 \) due to an identifiability issue caused by the calculation of \( C_0 C_0^T \). We shall, consequently, employ the following form of the BEKK conditional covariance equation in this thesis:

\[
H_t = C + A \varepsilon_{t-1} \varepsilon_{t-1}^T A^T + B H_{t-1} B^T.
\]

(4.3.48)

One advantage of the general BEKK conditional covariance equation shown in (4.3.48) is that it forces the conditional covariance matrix \( H_t \) to be positive definite. Another advantage is that the number of parameters requiring estimation is smaller than that for either the full VECH model or the parameter restricted VECH model, both of which were outlined in Section 4.3.1.

The number of parameters requiring estimation for the full BEKK model shown in (4.3.47) is \((l + m) p^2 Q + \frac{p}{2} (p + 1)\). While this is fewer than for a VECH model, even for small \( p \), it is still a considerable number. The diagonal BEKK (DBEKK) model is a simpler form of the full BEKK model in which the coefficient parameter matrices \( A \) and \( B \) are diagonal. This form of the BEKK model has \((l + m) p Q + \frac{p}{2} (p + 1)\) parameters that require estimation. For some sets of data, for example those in which there is little transmission of volatility between series, the DBEKK model will be sufficient to account for the conditional covariance.

Within the univariate setting of Section 3.2, the algorithm was designed to search over GARCH models...
with differing model orders, \( l \) and \( m \). As previously noted, for practical reasons, BEKK models are often restricted to the case where \( l = m = Q = 1 \). Therefore, in the multivariate setting, rather than searching over BEKK models of differing model orders \( l \) and \( m \), we shall only search over a constant conditional covariance model, a DBEKK model and a full BEKK model, as shown in Table 4.1.

The purely constant conditional covariance model is not an interesting model to fit to time series data. However, it has been included here to allow the incorporation of a pure M-STAR model into the scheme for fitting multivariate STAR-GARCH models in the following section.

The parameters that require estimation for the DBEKK and BEKK models are the intercept matrix \( C \) and the coefficient matrices \( A \) and \( B \). For the constant conditional covariance model, only the intercept matrix \( C \) is estimated. The prior distributions are outlined in the following section.

### 4.3.3 Prior Distribution

For the purposes of estimation, the coefficient matrices will be combined into a single vector of matrices, \( A = [A, B]^T \). We shall decompose the prior distribution for the M-GARCH estimation algorithm as follows:

\[
\]

The individual prior distributions are

\[
p(A, C|\mathcal{N}_j) \propto \mathbb{1}_A(A)
\]

\[
p(\mathcal{N}_j) \propto \frac{1}{j^{\gamma}}.
\]

(4.3.49)

As for the univariate GARCH model in Section 3.2.1, the stationarity will be enforced through the prior distribution of the coefficient matrices. This is achieved using the indicator function, \( \mathbb{1}_A(A) \). This function takes the value of one when the conditions for covariance stationarity are met, and zero otherwise. The BEKK model will be covariance stationary when the eigenvalues of the matrix

\[
\sum_{q=1}^{Q} \sum_{i=1}^{l} A_{qi} \otimes A_{qi} + \sum_{q=1}^{Q} \sum_{j=1}^{m} B_{qj} \otimes B_{qj},
\]

all have modulus less than one (Engle and Kroner 1995).
The prior distribution for the model index is defined in a way that allows simpler models with lower indices to have greater prior probabilities. Setting the parameter $\tau_N = 0$ incorporates the case where all prior model probabilities are proportional to one into the prior. In practice, the prior distribution can be set to any vector of prior probabilities that the user decides upon.

Now that we have considered the prior distributions, the next step is to derive the likelihood function for the M-GARCH model.

### 4.3.4 Likelihood Function

The derivation of the likelihood function hinges upon the distribution of $\eta_t$. When $\eta_t$ is distributed from a multivariate normal distribution, the likelihood function is the product of $n$ multivariate normal densities with zero mean vector and covariance matrix $H_t$.

In general, if $l$ and $m$ are model orders for the conditional covariance equation, the likelihood function will be calculated using the first $s = \max(l, m)$ data points as initial values. The data available for the likelihood calculation is therefore indexed from $s + 1$ to $N$ and the useful data has length $n = N - s$.

The conditional likelihood is calculated as follows:

$$p(X|\mathcal{F}_j, A, C) = \prod_{t=(s+1)}^{N} p(x_t|\mathcal{F}_j, A, C)$$

$$= \prod_{t=(s+1)}^{N} (2\pi)^{-\frac{n}{2}} |H_t|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} x_t^T H_t^{-1} x_t \right]$$

$$= (2\pi)^{-\frac{n}{2}} \prod_{t=(s+1)}^{N} |H_t|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \sum_{t=(s+1)}^{N} x_t^T H_t^{-1} x_t \right].$$

Focusing on the sum in the exponent in (4.3.50), we calculate

$$\sum_{t=(s+1)}^{N} x_t^T H_t^{-1} x_t = x_{s+1}^T H_{s+1}^{-1} x_{s+1} + x_{s+2}^T H_{s+2}^{-1} x_{s+2} + \cdots + x_N^T H_N^{-1} x_N$$

$$= [x_{s+1}^T, x_{s+2}^T, \ldots, x_N^T] \begin{bmatrix} \left|H_1^{-1}\right| & [x_{s+1}, x_{s+2}, \ldots, x_N] & \left|H_1^{-1}\right] \\
\end{bmatrix}$$

$$= \left[\text{vec} (X^T)\right]^T \left[ H^{-1} \right] \left[ \text{vec} (X^T) \right].$$

The matrix $H$ in (4.3.51) is a block diagonal matrix in which the $i$th diagonal entry corresponds to $H_i$.

Substituting the expression for the sum from equation (4.3.51) back into the likelihood function gives us the final likelihood function for the M-GARCH model:

$$p(X|\mathcal{F}_j, A, C) = (2\pi)^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec} (X^T) \right] \right].$$
4.3.5 Posterior Distribution

Combining the likelihood function in (4.3.52) with the prior distributions in (4.3.49) results in the following joint posterior distribution:

\[
p(N_j, A, C | X) \propto p(X|N_j, A, C) p(N_j, A, C) \propto (2\pi)^{-\frac{np^2}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T [H^{-1}] \left[ \text{vec}(X^T) \right] \right] \frac{I_A(A)}{j^N}. \tag{4.3.53}
\]

The parameters that are of interest are the model index, \(N_j\), the collection of coefficient parameter matrices \(A\) and the intercept matrix \(C\). As we move from one model index to another, the dimension of the parameter space changes. Therefore, an RJMCMC algorithm is required for parameter estimation.

The RJMCMC algorithm will be a Gibbs-style algorithm. In addition to the full posterior distribution in (4.3.53), Gibbs-style algorithms require the conditional posterior distribution for the coefficient matrix and intercept matrix. Therefore, the following posterior distributions will be employed in the MCMC estimation procedure:

\[
p(N_j, A, C | X) \propto (2\pi)^{-\frac{np^2}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T [H^{-1}] \left[ \text{vec}(X^T) \right] \right] \frac{I_A(A)}{j^N}
\]

\[
p(A, C | X, N_j) \propto (2\pi)^{-\frac{np^2}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T [H^{-1}] \left[ \text{vec}(X^T) \right] \right]. \tag{4.3.54}
\]

We can now design the required RJMCMC algorithm.

4.3.6 Posterior Simulator

Outlined below is the RJMCMC scheme for estimating the parameters of the BEKK model. The models to be included in the search are a constant conditional covariance model, a diagonal BEKK model and a full BEKK model.

M-GARCH - Main Algorithm

1. Initialise \((N_j^{(0)}, A^{(0)}, C^{(0)})\) either randomly or deterministically and set \(i = 1\).

2. At the \(i\)th iteration, execute the following steps employing the posterior distributions shown in (4.3.54):

(a) (i) **RJ\(N\)** Step - Simulate the GARCH model index, the intercept matrix and coefficient
Section 4.3: Multivariate GARCH

parameter matrix using

\[
\begin{pmatrix} N_j^{(i)} \ A^{(i)} \ C^{(i)} \end{pmatrix} \sim p(N_j, A, C | X).
\]

(ii) **DR\_A Step** - If the proposal in step 2(a)(i) is rejected, simulate the intercept and coefficient parameter matrices from the full conditional posterior distribution:

\[
\begin{pmatrix} A^{(i)} \ C^{(i)} \end{pmatrix} \sim p(A, C | X, N_j^{(i)}).
\]

Next, we expand the steps from the main algorithm that employ the RJMCMC and DRMH algorithms.

**Reversible Jump - RJ\_N Step**

Step 2(a) in the main algorithm must be a Reversible Jump step as the move to another model index changes the dimension of the parameter space. The acceptance probability for a Reversible Jump step outlined in Section 2.2.4 is

\[
r = \min \left( 1, \frac{\pi(\theta^{m*}, m* | X) q(\theta^{m*} | \theta^m, m^*)}{\pi(\theta^m, m | X) q(\theta^m | \theta^{m*}, m^*)} \right).
\]

Using the notation for the M-GARCH-BEKK model with the form of the acceptance probability in (4.3.55) results in

\[
r_{(N_j \rightarrow \star, \text{MG})} = \min \left( 1, \frac{p(N_j^*, A^*, C^* | X) q(A^*, C^* | A^*, C^*, N_j^*, N_j^*)}{p(N_j, A, C | X) q(A^* | A, C, N_j, N_j^*)} \right).
\]

The proposals for the candidate intercept and coefficient matrices will be obtained separately. If we assume that \( A \) and \( C \) are independent, the proposal distributions will have the following structures:

\[
q(A^* | A, C, N_j, N_j^*) = q(A^* | A, N_j, N_j^*) q(C^* | C, N_j, N_j^*).
\]

As the intercept matrix must be positive definite, the proposal distribution for \( C \) will be a Wishart distribution. There are two options for the coefficient matrices. The first is to simply use separate Matrix Normal distributions for each \( A_i \) and \( B_j \). While this is perfectly acceptable, it results in the multiplication of \( l + m \) densities when models of orders \( l, m > 1 \) are allowed. This will add complexity to the coding of the algorithm.

Therefore, for the full BEKK coefficient matrix, the complete matrix \( A^* \) will be proposed from a Matrix Normal distribution. The dimension of this Matrix Normal distribution is \( p \times p (l + m) \). The derivation
of the mathematical form of the proposal distribution for the coefficient matrix is provided in Appendix E.1 and its result is given in (E.1.5). The structure of this derivation indicates that proposing the full coefficient parameter matrix is equivalent to proposing the parameter matrices separately, assuming the among row covariance is equal for each parameter matrix, that is \( U_1 = U_2 = \cdots = U_r \).

For the proposal of the elements of the DBEKK model coefficient parameter matrix, the diagonal elements are simply combined into a single vector and a Multivariate Normal distribution is used.

As with the univariate GARCH estimation procedure, determining the location parameters of the proposal distributions is difficult, but is essential to the success of the algorithm. A simple and effective option is to perform a short pilot run to find reasonable location parameters that will allow the full algorithm to make model moves under the Reversible Jump scheme.

In general, the proposal distributions for the M-GARCH Reversible Jump step will be

\[
q(A^* | A, N_j^*, N_j^*) = 1 \quad \text{(Constant)}
\]

\[
q(A^* | A, N_j^*, N_j^*) \sim N_{l+m} (m_{N_j^*}, V_{N_j^*}) \quad \text{(DBEKK)}
\]

\[
q(A^* | A, N_j^*, N_j^*) \sim N_{p,p(\pm m)} (M_{N_j^*}, U_{N_j^*}, V_{N_j^*}) \quad \text{(BEKK)}
\]

\[
q(C^* | C, N_j^*, N_j^*) \sim W_{p} \left( \Delta_{C_{N_j^*}}^{-1}, C_{N_j^*}, \Delta_{C_{N_j^*}} \right).
\]

The Wishart distribution in (4.3.57) has a location parameter of \( C_{N_j^*} \). This location parameter, along with the location parameters for the proposal distribution of the conditional covariance coefficient parameter matrix, \( M_{N_j^*} \) or \( m_{N_j^*} \), is determined during the pilot run. We have given the parameters the subscript \( N_j^* \) to remind the reader that these parameters have the potential to be different for each model index. The variance parameters \( U_{N_j^*}, V_{N_j^*}, \) and \( \Delta_{C_{N_j^*}} \) can roughly be determined during the pilot run, but may require some adjustment to ensure an adequate acceptance rate for the algorithm.

The proposal distributions shown in (4.3.57) for the conditional covariance coefficient parameter matrix pertain to the general case in which the RJMCMC scheme allows not only jumps from different model types, but also jumps to differing model orders within a particular model. For practical reasons, the estimation algorithm implemented will only consider the models shown in Table 4.1. In particular, the model orders \( l \) and \( m \) will be fixed.

The most complicated acceptance probability that involves a change in either or both model orders is one that involves a move within the BEKK model. The acceptance probability for such a move has been derived in Appendix E.2, and the formula is given in (E.2.6).

As mentioned above, all of the models considered here are listed in Table 4.1. When a move to a candidate model is proposed, the proposal distributions for \( A \) for the candidate and current models
will differ. This means that at each Reversible Jump, the acceptance probability can take a total of
6 different mathematical forms, depending on the indices of the current and candidate models. Each
of these acceptance probabilities is calculated in Appendix E.2, and the results are summarised in
equations (E.2.7), (E.2.8), (E.2.9), (E.2.10), (E.2.11), and (E.2.12).

If the calculations described in Appendix E.2 are performed directly, computational problems associated
with either overflow or underflow are likely to occur, especially for large data sets. Moreover, when we
used the inbuilt function for the Wishart distribution to calculate the ratio of the proposal distribution
for the intercept matrix, we encountered problems. The solution involved taking the logarithm of the
formula and calculating as much as possible before taking the exponential of the result to yield the
final probability.

The algorithm for the RJ\(N\) Step is shown below.

**RJ\(N\) Algorithm**

1. The current state is \((N_j^{(i-1)}, A^{(i-1)}, C^{(i-1)})\).

2. Propose a candidate model index \(N_j^*\), from \(j (N_j^* | N_j)\) using the discretised Laplacian shown in
   \((2.2.17)\). Propose a candidate intercept and coefficient matrices using the proposal distributions
   in \((4.3.57)\).

3. Determine the appropriate acceptance probability, based on \(N_j^*\) and \(N_j\), and calculate the ac-
   ceptance probability, \(r_{(N_j, j \rightarrow j^*, \text{MG})}\), using the relevant equation from \((E.2.7)\), \((E.2.8)\), \((E.2.9)\),
   \((E.2.10)\), \((E.2.11)\) and \((E.2.12)\).

4. Simulate \(u \sim U(0, 1)\):
   
   (a) If \(r_{(N_j, j \rightarrow j^*, \text{MG})} > u\), then:
   
   (i) Accept the candidate model index \(N_j^*\) and set \(N_j^{(i)} = N_j^*\).
   
   (ii) Accept the intercept and coefficient parameter matrices. Set \(C^{(i)} = C^*\) and \(A^{(i)} = A^*\).
   
   (b) If \(r_{(N_j, j \rightarrow j^*, \text{MG})} < u\), then:

   (i) Reject the candidate model index \(N_j^*\) and proposed covariance parameter matrices.
   
   Set \(N_j^{(i)} = N_j^{(i-1)}\), \(C^{(i)} = C^{(i-1)}\) and \(A^{(i)} = A^{(i-1)}\).

   (ii) Using the distribution in \((4.3.54)\), simulate the intercept and coefficient parameter ma-
   trices,

   \[
   (A, C)^{(i)} \sim p \left( A, C | X, N_j^{(i)} \right),
   \]

   using the DRMH algorithm described below (DR\(\text{A}\) Step).
Delayed Rejection Metropolis-Hastings - DR\textsubscript{A} Step

When the proposed jump to a new model index is rejected, the intercept and coefficient matrices need
to be simulated from their joint conditional posterior distribution. If this did not occur, the intercept
and coefficient matrices would only be updated upon model moves. When the data presents significant
evidence for one model over the others included in the search, it is possible that model moves will not
often take place. If the intercept and coefficient matrices are not separately updated, the algorithm
will exhibit poor mixing and slow convergence.

Slow convergence of the main algorithm is a significant problem faced by the user. Within the univariate
GARCH setting, it was possible to use MLE’s as starting values for the algorithm. However, in the
multivariate setting, finding the MLE’s is not a simple task in itself and thus it is a reasonably difficult
problem to identify suitable starting values quickly. Given that our starting values are likely to be
relatively poor, we need to find techniques that allow our algorithms to move quickly towards the area
of convergence. One possible approach would be the MTM algorithm, but we have had more success
with the DRMH algorithm. The DRMH algorithm allows the use of relatively large variances for the
proposal distributions so that initially the algorithm moves quickly towards the area of convergence,
but also allows small movements if a poor proposal is made in the wrong direction, or even past the
area of convergence. The DRMH algorithm is detailed in Section 2.2.3.

Here we shall implement a two stage DRMH algorithm. This involves the prescription of proposal
distributions for both stages of the algorithm. The proposal distributions used in the first stage will
similar to those used for the Reversible Jump step above, but they will be centred on the most recent
value in the MCMC chain. The proposal distributions employed are

\begin{align}
q_{1A}(A^*|A) &\sim N_{p,l+m}(A, U_{N_j}, V_{N_j}) \\
q_{1C}(C^*|C) &\sim W_{p}\left(\Delta_{C_{N_j}}^{-1} C, \Delta_{C_{N_j}}\right),
\end{align}

where $\Delta_{2C}$ and $\Delta_{2A}$ are parameters included to reduce the variance of the proposal in the second
stage. Setting these values in the range of 0.50 to 0.75 yielded successful results during testing of the
The target distribution for this step is the joint conditional posterior distribution shown in (4.3.54). That is,

\[ q(A, C|X, N_j) \propto (2\pi)^{-n_p/2} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( X^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( X^T \right) \right] \right]. \]

Since there are three different situations in which this algorithm could be implemented, namely for the constant conditional covariance model, the DBEKK model and the BEKK model, we only present the most complicated form of the acceptance probability. The difference between the possible forms ultimately lies in the ratio of the proposal distributions for the conditional covariance coefficient parameter matrices.

The acceptance probability for the first stage of the DRMH algorithm is relatively simple. It is simply a Metropolis-Hastings acceptance probability, which can be written using the notation for the M-GARCH model as follows:

\[ r_{(A,1,MG)}(A^*, C^*, A, C) = \min \left( 1, \frac{p(A^*, C^*|X, N_j) q_{llA}(A|A^*) q_{llC}(C|C^*)}{p(A, C|X, N_j) q_{llA}(A^*|A) q_{llC}(C^*|C)} \right). \]  \hspace{1cm} (4.3.60)

Therefore, starting with the quotient in the acceptance probability in (4.3.60), we may calculate

\[
\frac{p(A^*, C^*|X, N_j) q_{llA}(A|A^*) q_{llC}(C|C^*)}{p(A, C|X, N_j) q_{llA}(A^*|A) q_{llC}(C^*|C)} = \frac{(2\pi)^{-n_p/2} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( X^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( X^T \right) \right] \right]}{(2\pi)^{-n_p/2} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( X^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( X^T \right) \right] \right]}.
\]

\[
= |C|^{\frac{1}{2}} \binom{\Delta_{C^j}}{C_j} \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_j} C_j^{-1} C \right) \right] \frac{2^\frac{n_p}{2} \binom{\Delta_{C^j}}{C_j} \binom{\Delta_{C^j}}{C_j}}{2^\frac{n_p}{2} \binom{\Delta_{C^j}}{C_j} \binom{\Delta_{C^j}}{C_j}}.
\]

\[
= |H|^{-\frac{1}{2}} \binom{C_j}{C_j} \frac{\binom{\Delta_{C^j}}{C_j} \Delta_{C^j}^{-p-1}}{\binom{\Delta_{C^j}}{C_j}} \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C^j} C_j^{-1} C \right) \right] \frac{2^\frac{n_p}{2} \binom{\Delta_{C^j}}{C_j} \binom{\Delta_{C^j}}{C_j}}{2^\frac{n_p}{2} \binom{\Delta_{C^j}}{C_j} \binom{\Delta_{C^j}}{C_j}}.
\]

\[
= |H|^{-\frac{1}{2}} \binom{C_j}{C_j} \frac{\binom{\Delta_{C^j}}{C_j} \Delta_{C^j}^{-p-1}}{\binom{\Delta_{C^j}}{C_j}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( X^T \right) \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( X^T \right) \right] - \left[ \text{vec} \left( X^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( X^T \right) \right] + \text{tr} \left( \left[ \Delta_{C^j} C_j^{-1} C \right] - \left[ \Delta_{C^j} C_j^{-1} C \right] \right) \right].
\]
This gives the following formula for the acceptance probability for the first stage of the DR\(_A\) Algorithm:

\[
\rho_{(A,1,MG)}(A^*, C^*, A, C) = \min \left(1, \frac{1}{|H|^\frac{3}{2}} \left| C^\star \right|^{-\frac{1}{2}} \left( \Delta_{CN_j}^{-p-1} \right) \frac{\Delta_{CN_j}}{2} \left| C^\star \right|^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2} \left( \text{vec} \left( X^T \right)^T \left[ H^{-1} \right] \text{vec} \left( X^T \right) \right) \right] + \right. \\
\left. \text{tr} \left( \Delta_{CN_j}^{-1} C^\star - \left[ \Delta_{CN_j}^{-1} C \right]^{-1} C^\star \right) \right) 
\]

In cases where the initial proposed parameters are rejected, the DR\(_A\) Algorithm enters the second stage. This stage requires a slightly more complicated acceptance probability to ensure convergence to the correct stationary distribution. The required acceptance probability has the following form:

\[
\rho_{(A,2,MG)}(A^*, C^*, A^*, C^*, A, C) = \min \left(1, \frac{p \left( A^*, C^* \middle| X, N_j \right) \rho_{1A} \left( A^* \middle| A^* \right) \rho_{1C} \left( C^* \middle| C^* \right)}{p \left( A, C \middle| X, N_j \right) \rho_{1A} \left( A^* \middle| A \right) \rho_{1C} \left( C^* \middle| C \right)} \times \right. \\
\left. \rho_{2A} \left( A \middle| A^*, A^* \right) \rho_{2C} \left( C \middle| C^*, C^* \right) \left( 1 - \rho_{(A,1,MG)} \left( A^*, C^*, A^*, C^* \right) \right) \times \right. \\
\left. \rho_{2A} \left( A^* \middle| A, A^* \right) \rho_{2C} \left( C^* \middle| C^*, C^* \right) \left( 1 - \rho_{(A,1,MG)} \left( A^*, C^*, A, C \right) \right) \right).
\]

Focusing on the quotient on the right hand side of the acceptance probability shown in (4.3.62), we may calculate

\[
\frac{p \left( A^*, C^* \middle| X, N_j \right) \rho_{1A} \left( A^* \middle| A^* \right) \rho_{1C} \left( C^* \middle| C^* \right)}{p \left( A, C \middle| X, N_j \right) \rho_{1A} \left( A^* \middle| A \right) \rho_{1C} \left( C^* \middle| C \right)} \times \\
\frac{\rho_{2A} \left( A \middle| A^*, A^* \right) \rho_{2C} \left( C \middle| C^*, C^* \right) \left( 1 - \rho_{(A,1,MG)} \left( A^*, C^*, A^*, C^* \right) \right)}{\rho_{2A} \left( A^* \middle| A, A^* \right) \rho_{2C} \left( C^* \middle| C^*, C^* \right) \left( 1 - \rho_{(A,1,MG)} \left( A^*, C^*, A, C \right) \right)}
\]

\[
= \frac{\exp \left[ -\frac{1}{2} \text{tr} \left( \frac{V_{N_j}}{2} (A^* - A)^T U_{N_j}^{-1} (A^* - A) \right) \right] \left| C^\star \right|^{\frac{1}{2}} \left( \Delta_{CN_j}^{-p-1} \right) \exp \left[ -\frac{1}{2} \text{tr} \left( \frac{\Delta_{CN_j}^{-1}}{2} \right) \right]}{(2\pi)^{\frac{1}{2}} |H|^{1/2} \exp \left[ -\frac{1}{2} \text{tr} \left( \frac{\text{vec} \left( X^T \right)^T H^{-1} \text{vec} \left( X^T \right) \right) \right]}
\]

\[
= \left( \frac{\left| V_{N_j} \right|^{1/2}}{\left| V_{N_j} \right|^{1/2}} \right)^{\frac{p \Delta_{CN_j}}{2}} \frac{2 \Delta_{CN_j}^{-1} C^\star \Gamma_p \left( \Delta_{CN_j}^{-1} \right)}{(2\pi)^{\frac{1}{2}} |H|^{1/2} \exp \left[ -\frac{1}{2} \text{tr} \left( \frac{\text{vec} \left( X^T \right)^T H^{-1} \text{vec} \left( X^T \right) \right) \right]}
\]

\[
\quad \times \left( \frac{\left| V_{N_j} \right|^{1/2}}{\left| V_{N_j} \right|^{1/2}} \right)^{\frac{p \Delta_{CN_j}}{2}} \frac{2 \Delta_{CN_j}^{-1} C^\star \Gamma_p \left( \Delta_{CN_j}^{-1} \right)}{(2\pi)^{\frac{1}{2}} |H|^{1/2} \exp \left[ -\frac{1}{2} \text{tr} \left( \frac{\text{vec} \left( X^T \right)^T H^{-1} \text{vec} \left( X^T \right) \right) \right]}
\]
\[
\begin{align*}
&\exp\left[-\frac{1}{2}\text{tr}\left(\Delta_2 A V_{N_j}^T (A - \frac{1}{2} (A^* + A^*))^T U_{N_j}^{-1} (A - \frac{1}{2} (A^* + A^*))\right)\right] \\
&\times \frac{\exp\left[-\frac{1}{2}\text{tr}\left(\Delta_2 A V_{N_j}^T (A' - \frac{1}{2} (A + A^*))^T U_{N_j}^{-1} (A' - \frac{1}{2} (A + A^*))\right)\right]}{\exp\left[-\frac{1}{2}\text{tr}\left(\Delta_2 A V_{N_j}^T (A' - \frac{1}{2} (A + A^*))^T U_{N_j}^{-1} (A' - \frac{1}{2} (A + A^*))\right)\right]} \\
&\times \left[C\right]^{-\frac{1}{2}} \left(\Delta_{C, N_j}^{1 - \frac{1}{2}}\frac{\Delta_{C, N_j}^{\frac{1}{2}}}{2}\right)^{\frac{1}{2}} \times \\
&\exp\left[-\frac{1}{2} \left[V_{N_j}^{-1} (A' - A)^T U_{N_j}^{-1} (A' - A)\right]\right] \times \\
&\exp\left[-\frac{1}{2} \left[V_{N_j}^{-1} (A^* - A)^T U_{N_j}^{-1} (A^* - A)\right]\right] \times \\
&\exp\left[-\frac{1}{2} \text{tr}\left(\Delta_2 A V_{N_j}^T (A - \frac{1}{2} (A^* + A^*))^T U_{N_j}^{-1} (A - \frac{1}{2} (A^* + A^*))\right)\right] \\
&\times \left(1 - \frac{r_{(A,1, MG)}}{r_{(A,1, MG)}}\left(A^*, C^*, A', C'\right)\right) \times \\
&\left(1 - \frac{r_{(A,1, MG)}}{r_{(A,1, MG)}}\left(A^*, C^*, A, C\right)\right) \\
&\left[C\right]^{-\frac{1}{2}} \left(\Delta_{C, N_j}^{1 - \frac{1}{2}}\frac{\Delta_{C, N_j}^{\frac{1}{2}}}{2}\right)^{\frac{1}{2}} \times \\
&\exp\left[-\frac{1}{2} \left[V_{N_j}^{-1} (A' - A)^T U_{N_j}^{-1} (A' - A)\right]\right] \times \\
&\exp\left[-\frac{1}{2} \left[V_{N_j}^{-1} (A^* - A)^T U_{N_j}^{-1} (A^* - A)\right]\right] \times \\
&\exp\left[-\frac{1}{2} \text{tr}\left(\Delta_2 A V_{N_j}^T (A - \frac{1}{2} (A^* + A^*))^T U_{N_j}^{-1} (A - \frac{1}{2} (A^* + A^*))\right)\right] \\
&\times \left(1 - \frac{r_{(A,1, MG)}}{r_{(A,1, MG)}}\left(A^*, C^*, A', C'\right)\right) \times \\
&\left(1 - \frac{r_{(A,1, MG)}}{r_{(A,1, MG)}}\left(A^*, C^*, A, C\right)\right) \\
\end{align*}
\]
\[
\exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{2A} V_{N_j}^{-1} \left( A - \frac{1}{2} \left( A' + A^* \right) \right)^T U_{N_j}^{-1} \left( A - \frac{1}{2} \left( A' + A^* \right) \right) \right) \right] \times \\
\exp \left[ \frac{1}{2} \text{tr} \left( \Delta_{2A} V_{N_j}^{-1} \left( A' - \frac{1}{2} \left( A + A^* \right) \right)^T U_{N_j}^{-1} \left( A' - \frac{1}{2} \left( A + A^* \right) \right) \right) \right] \times \\
\exp \left[ -\frac{1}{2} \text{tr} \left( \left[ \left[ 2 \Delta_{Cj} \Delta_{2C} \right]^{-1} \left( C + C^* \right) \right]^{-1} C \right) \right] \times \\
\exp \left[ \frac{1}{2} \text{tr} \left( \left[ \left[ 2 \Delta_{Cj} \Delta_{2C} \right]^{-1} \left( C + C^* \right) \right]^{-1} C' \right) \right] \left( \frac{1 - r_1 \left( A^*, C^*, A', C' \right)}{1 - r_1 \left( A^*, C^*, A, C \right)} \right)
\]

\[
= |H|^{-\frac{1}{2}} |H|^{\frac{1}{2}} \left[ |C | |C'^{-1} \left( \frac{\Delta_{Cj} \Delta_{2C}}{2} \right)^{\frac{\Delta_{Cj} \Delta_{2C}}{2}} \times \\
\exp \left[ -\frac{1}{2} \left( [\text{vec}(X^T)]^T \left[ H^{-1} \right] [\text{vec}(X^T)] \right) \right] - \\
\frac{1}{2} \text{tr} \left( \left[ \left[ 2 \Delta_{Cj} \Delta_{2C} \right]^{-1} \left( C + C^* \right) \right]^{-1} C \right) + \\
\frac{1}{2} \text{tr} \left( \left[ \left[ 2 \Delta_{Cj} \Delta_{2C} \right]^{-1} \left( C + C^* \right) \right]^{-1} C' \right) \left( \frac{1 - r_1(A_{1,\text{MG}}) \left( A^*, C^*, A', C' \right)}{1 - r_1(A_{1,\text{MG}}) \left( A^*, C^*, A, C \right)} \right)
\]

\[
= |H|^{-\frac{1}{2}} |H|^{\frac{1}{2}} \left[ |C | |C'^{-1} \left( \frac{\Delta_{Cj} \Delta_{2C}}{2} \right)^{\frac{\Delta_{Cj} \Delta_{2C}}{2}} \times \\
\exp \left[ -\frac{1}{2} \left( [\text{vec}(X^T)]^T \left[ H^{-1} \right] [\text{vec}(X^T)] \right) \right] - \\
\frac{1}{2} \text{tr} \left( \left[ \left[ 2 \Delta_{Cj} \Delta_{2C} \right]^{-1} \left( C + C^* \right) \right]^{-1} C \right) + \\
\frac{1}{2} \text{tr} \left( \left[ \left[ 2 \Delta_{Cj} \Delta_{2C} \right]^{-1} \left( C + C^* \right) \right]^{-1} C' \right) \left( \frac{1 - r_1(A_{1,\text{MG}}) \left( A^*, C^*, A', C' \right)}{1 - r_1(A_{1,\text{MG}}) \left( A^*, C^*, A, C \right)} \right)
\]
The acceptance probability for the second stage of the DR Algorithm is found by calculating the probabilities in (4.3.63) and (4.3.61) and then substituting them into the following acceptance probability:

\[
\begin{align*}
   r_{(A,2, MG)} & \left( A^*, C^*, A', C' \right) \\
   & = \min \left( 1, |H|^{-\frac{1}{2}} |H|^\frac{1}{2} \left[ C^* \right]^{-1} \right) \frac{\Delta_{C_{N_j}} \Delta_{2C} - p - 1}{\Delta_{C_{N_j}}} \times \\
   & \exp \left[ -\frac{1}{2} \left( \text{vec} (X^T) \right)^T \left( [H^{-1}] - [H^{-1}] \right) \text{vec} (X^T) \right] + \\
   & \text{tr} \left( V_{N_j}^{-1} \left( A^* - A' \right)^T U_{N_j}^{-1} \left( A^* - A' \right) \right) - \text{tr} \left( [\Delta_{C_{N_j}} C^*]^{-1} C^* \right) - \\
   & \text{tr} \left( \Delta_{2A} V_{N_j}^{-1} \left( A - \frac{1}{2} (A + A') \right)^T U_{N_j}^{-1} \left( A - \frac{1}{2} (A + A') \right) \right) - \\
   & \text{tr} \left( \Delta_{2A} V_{N_j}^{-1} \left( A' - \frac{1}{2} (A + A') \right)^T U_{N_j}^{-1} \left( A' - \frac{1}{2} (A + A') \right) \right) + \\
   & \text{tr} \left( \left[ 2\Delta_{C_{N_j}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right)^{-1} C \right) - \\
   & \end{align*}
\]
\[
\text{tr} \left( \left[ 2\Delta_{CN} \Delta_{2C} \right]^{-1} (C + C^*)^{-1} C' \right) \left[ \frac{1 - r_{(A,1,MG)}(A^*, C^*, A', C')} {1 - r_{(A,1,MG)}(A^*, C^*, A, C)} \right].
\]

The layout of the DR\textsubscript{A} Algorithm is outlined below.

**DR\textsubscript{A} Algorithm**

1. The current state is \( \left( \mathcal{X}^{(i)}, A^{(i-1)}, C^{(i-1)} \right) \).

2. Propose new intercept and coefficient parameter matrices \( C^* \) and \( A^* \) from the proposal densities, \( q_{1C}(C^* | C) \) and \( q_{1A}(A^* | A) \), respectively which are given in (4.3.58).

3. Calculate the acceptance probability \( r_{(A,1,MG)}(A^*, C^*, A, C) \) using the formula in (4.3.61).

4. Simulate \( u_1 \sim \mathcal{U}(0, 1) \):
   
   (a) If \( r_{(A,1,MG)}(A^*, C^*, A, C) > u_1 \), accept the proposed intercept and coefficient parameter matrices \( C^* \) and \( A^* \). Set \( C^{(i)} = C^* \) and \( A^{(i)} = A^* \).
   
   (b) If \( r_{(A,1,MG)}(A^*, C^*, A, C) < u_1 \), then:
      
      (i) Generate new proposed intercept and coefficient parameter matrices, \( C' \) and \( A' \) from \( q_2C(C' | C^*, C) \) and \( q_2A(A' | A^*, A) \), respectively, using the proposal distributions in (4.3.59).
      
      (ii) Calculate the acceptance probability, \( r_{(A,2,MG)}(A', C', A^*, C^*, A, C) \), using the formulas in (4.3.63), (4.3.61) and (4.3.64).
      
      (iii) Simulate \( u_2 \sim \mathcal{U}(0, 1) \):
          
          (A) If \( r_{(A,2,MG)}(A', C', A^*, C^*, A, C) > u_2 \), accept the new proposed intercept and coefficient parameter matrices. Set \( C^{(i)} = C' \) and \( A^{(i)} = A' \).
          
          (B) If \( r_{(A,2,MG)}(A', C', A^*, C^*, A, C) < u_2 \), reject the new proposed intercept and coefficient parameter matrices. Set \( C^{(i)} = C^{(i-1)} \) and \( A^{(i)} = A^{(i-1)} \).

It is relatively straight forward to extend the DRMH algorithm above to include more stages. The acceptance probability shown in (2.2.15) is the general acceptance probability that should be used for later stages of the algorithm.

The following section will combine the work of Sections 4.2 and 4.3 to create an MCMC estimation algorithm for the Multivariate STAR-GARCH model.
4.4 Multivariate STAR-GARCH

Combining the conditional mean of the M-STAR model outlined in Section 4.2 with the conditional covariance of the M-GARCH model in Section 4.3 results in the Multivariate STAR-GARCH model (M-STAR-GARCH). For convenience, the form of the model analysed here is restricted to assume a common transition function and two regimes with no intercept terms in the conditional mean equation.

Assuming an individual data point \( x_t \) is a \( p \times 1 \) vector representing \( p \) different time series at time \( t \), the M-STAR-GARCH model can be defined as follows:

\[
\begin{align*}
\mathbf{x}_t &= \mu_t + \mathbf{\varepsilon}_t \\
\mu_t &= \Phi_{1,1} \mathbf{x}_{t-1} + \cdots + \Phi_{1,k} \mathbf{x}_{t-k} + (\Phi_{2,1} \mathbf{x}_{t-1} + \cdots + \Phi_{2,k} \mathbf{x}_{t-k}) \mathbf{F}_t \\
\mathbf{\varepsilon}_t &= \mathbf{H}_t^{\frac{1}{2}} \mathbf{\eta}_t.
\end{align*}
\]

The conditional mean \( \mu_t \) can be manipulated using a similar approach to the one applied for the M-STAR model in Section 4.2. That is,

\[
\begin{align*}
\mu_t &= \Phi_{1,1} \mathbf{x}_{t-1} + \cdots + \Phi_{1,k} \mathbf{x}_{t-k} + (\Phi_{2,1} \mathbf{x}_{t-1} + \cdots + \Phi_{2,k} \mathbf{x}_{t-k}) \mathbf{F}_t \\
&= \Phi_{1,1} \mathbf{x}_{t-1} + \cdots + \Phi_{1,k} \mathbf{x}_{t-k} + \Phi_{2,1} \mathbf{x}_{t-1} \mathbf{F}_t + \cdots + \Phi_{2,k} \mathbf{x}_{t-k} \mathbf{F}_t \\
&= [\Phi_{1,1}, \ldots, \Phi_{1,k}, \Phi_{2,1}, \ldots, \Phi_{2,k}] \begin{bmatrix} \mathbf{x}_{t-1}^T, & \ldots, & \mathbf{x}_{t-k}^T, & \mathbf{F}_t, & \ldots, & \mathbf{x}_{t-k}^T \mathbf{F}_t \end{bmatrix}^T.
\end{align*}
\]

The form of the model in (4.4.65) is then simplified by collecting the terms in the mean equation. Let the collection of terms be

\[
\Phi^T = [\Phi_{1,1}, \ldots, \Phi_{1,k}, \Phi_{2,1}, \ldots, \Phi_{2,k}]
\]

\[
z_t^T = [\mathbf{x}_{t-1}^T, \ldots, \mathbf{x}_{t-k}^T, \mathbf{F}_t, \ldots, \mathbf{x}_{t-k}^T \mathbf{F}_t].
\]

The conditional mean equation is now expressed as

\[
\mu_t = \Phi^T z_t.
\]

The transition function \( \mathbf{F}_t \), shown in (4.4.65) is the logistic function used in Section 4.2. This function has formula,

\[
\mathbf{F}_t (s_t, d, \gamma, c) = \frac{1}{1 + \exp \left[ -\frac{\gamma}{s_t} (s_t - d - c) \right]}.
\]
the smoothing and location parameters $\gamma \in \mathbb{R}^+$ and $c \in \mathbb{R}$, respectively.

For the conditional covariance, $\eta_t$ is an independent identically distributed random vector of length $p$ such that $E[\eta_t] = 0$ and $E[\eta_t \eta_t^T] = I_p$. The $p \times p$ conditional covariance matrix at time $t$ is given by $H_t$. As discussed in Section 4.3, there are several forms of the covariance equation that may be used to calculate $H_t$, but we shall employ the BEKK formulation for the M-STAR-GARCH model. The general form of its covariance equation is

$$H_t = C + \sum_{q=1}^{Q} \sum_{i=1}^{l} A_{qi} \epsilon_{t-i} \epsilon_{t-i}^T A_{qj}^T + \sum_{q=1}^{Q} \sum_{j=1}^{m} B_{qj} H_{t-j} B_{qj}^T.$$  

The parameters of the covariance equation are the intercept matrix and the coefficient matrix. In Section 4.3, it was assumed that the maximum model orders to be considered for the conditional covariance equation were $Q = l = m = 1$. This simplifies the conditional covariance equation and, consequently, the estimation procedure. This assumption is one that is often made in practice when applying the M-GARCH model (Silvennoinen and Teräsvirta 2009b). Therefore, the conditional covariance equation that will be used for the estimation procedure will be

$$H_t = C + A \epsilon_{t-1} \epsilon_{t-1}^T A^T + B H_{t-1} B^T.$$  \hspace{1cm} (4.4.67)

It is required that the intercept matrix $C$ be positive definite. For simplicity, we shall collect the coefficient matrices into one coefficient matrix, $A = [A,B]$.

As in the previous sections, the presentation of the model in matrix notation can make the mathematical manipulation a little easier to follow. The matrix representation of the model is

$$X = Z_1 \Phi + E$$

$$\text{vec} \left( E^T \right) = H^\frac{1}{2} \eta,$$

where

$$X = \begin{bmatrix} x_{s+1}^T \\ x_{s+2}^T \\ \vdots \\ x_N^T \end{bmatrix}, \quad E = \begin{bmatrix} \epsilon_{s+1}^T \\ \epsilon_{s+2}^T \\ \vdots \\ \epsilon_N^T \end{bmatrix}, \quad \Phi = \begin{bmatrix} \Phi_{1,1} \\ \vdots \\ \Phi_{1,k} \\ \Phi_{2,1} \\ \vdots \\ \Phi_{2,k} \end{bmatrix}, \quad H = \begin{bmatrix} H_s & 0 & \ldots & 0 \\ 0 & H_{s+1} & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \ldots & 0 & H_N \end{bmatrix}.$$
The above matrices have the following dimensions: $X \in \mathbb{R}^{n \times p}$, $E \in \mathbb{R}^{n \times p}$, $\Phi \in \mathbb{R}^{2kp \times p}$, $Z_1 \in \mathbb{R}^{n \times 2kp}$ and $H \in \mathbb{R}^{np \times np}$. The length of the useful data is $n = N - s$, where $s = \max(k, d, l, m)$.

The model orders for the conditional mean and the model types for the conditional covariance equation will be indexed using $M_k$ and $N_j$, respectively. The possible models for the conditional covariance equation will be a constant conditional covariance model, a Diagonal BEKK model and a full BEKK model. The next section presents the definition of the prior distributions for the parameters of interest.

### 4.4.1 Prior Distribution

The joint prior distribution used for the coefficient parameter matrix and the covariance matrix of the error term will be the same as that used in Section 4.1. It will be decomposed as follows:

$$
p(M_k, \Phi, d, \gamma, c, N_j, C, A) = p(M_k, \Phi, d, \gamma, c) p(N_j, C, A) \tag{4.4.68}
$$

$$
= p(M_k, \Phi | d, \gamma, c) p(d, \gamma, c) p(C, A | N_j) p(N_j) \tag{4.4.69}
$$

The separate prior distributions on the right hand side of (4.4.68) are defined using the distributions presented below:

$$
p(M_k, \Phi | d, \gamma, c) \propto \left[ \frac{\Lambda_k}{k!} \right]^{\gamma_k} I_{K}(k)
$$

$$
p(d) \propto I_D(d)
$$

$$
p(\gamma) \sim \mathcal{G} (\alpha_\gamma, \beta_\gamma)
$$

$$
p(c) \sim \mathcal{N} (\mu_c, \sigma_c^2)
$$

$$
p(A, C | N_j) \propto I_A(A)
$$

$$
p(N_j) \propto \frac{1}{j^7 \gamma_j}
$$

The formulation of the prior distributions for the transition function smoothing parameter $\gamma$ and location parameter $c$ will be as given in Section 4.2. The prior distribution for the GARCH coefficient and intercept matrices is chosen in order to enforce stationarity for the conditional covariance of the
model. This is done through the indicator function, \( I_A(A) \), which takes the value of one when the conditions of covariance stationarity are met, and otherwise zero. When the matrix

\[
\sum_{q=1}^{Q} \sum_{i=1}^{l} A_{qi} \otimes A_{qi} + \sum_{q=1}^{Q} \sum_{j=1}^{m} B_{qj} \otimes B_{qj},
\]

has all eigenvalues of modulus less than one, the BEKK model will be covariance stationary (Engle and Kroner 1995).

The prior distribution for the model index is defined in such a way as to allow the simpler models with lower indices to have greater prior probabilities. By setting the parameter \( \tau_N = 0 \), the prior includes the case in which all prior model probabilities are proportional to one. In practice, this prior distribution can be set to any vector of prior probabilities that the user decides upon.

We shall derive the likelihood function for the M-STAR-GARCH model in the following section.

### 4.4.2 Likelihood Function

The form the likelihood function takes depends on the distribution assumed for \( \eta_t \). When a multivariate normal distribution is assumed, the likelihood function is the product of \( n \) multivariate normal densities with conditional mean vector \( \Phi^T z_t \) and covariance matrix \( H_t \).

In general, if \( l \) and \( m \) are model orders for the conditional covariance equation in a full BEKK model, the likelihood function will be calculated using the first \( s = \max(k, d, l, m) \) data points as initial values. Therefore, the data available for the likelihood function will have indices from \( s + 1 \) to \( N \) and the useful data will have length \( n = N - s \). The conditional likelihood will then be calculated as follows:

\[
p(X|M_{k,\Phi,d,\gamma,c,N_j,A,C}) = \prod_{t=(s+1)}^{N} p(x_t|M_{k,\Phi,d,\gamma,c,N_j,A,C})
\]

\[
= \prod_{t=(s+1)}^{N} (2\pi)^{-\frac{np}{2}} |H_t|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x_t - \Phi^T z_t)^T H_t^{-1} (x_t - \Phi^T z_t) \right]
\]

\[(4.4.70)\]

Focusing on the sum in the exponent in (4.4.70) yields

\[
\sum_{t=(s+1)}^{N} (x_t - \Phi^T z_t)^T H_t^{-1} (x_t - \Phi^T z_t)
\]
Therefore, substituting the expression in (4.4.71) back into the likelihood function in (4.4.70) yields

\[ H \]

The matrix \( H \) in (4.4.71) is a block diagonal matrix whose \( i \)th diagonal entry corresponds to \( H_i \). Therefore, substituting the expression in (4.4.71) back into the likelihood function in (4.4.70) yields

\[
p(X|M_k, \Phi, d, \gamma, c, N_j, A, C) = (2\pi)^{-\frac{N}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( X - Z_1 \Phi^T \right)^T \left( H^{-1} \right) \left[ \text{vec} \left( X - Z_1 \Phi^T \right) \right] \right] \right].
\]  

(4.4.72)

By combining the likelihood function in (4.4.72) with the prior distributions in (4.4.69), we can derive the joint posterior distribution. The details are given in the following section.

### 4.4.3 Posterior Distribution

The joint posterior distribution is obtained as follows:

\[
p(M_k, \Phi, d, \gamma, c, N_j, A, C | X) \propto p(X|M_k, \Phi, d, \gamma, c, N_j, A, C) p(M_k, \Phi|d, \gamma, c) p(\gamma) p(c) p(C, A|N_j) p(N_j)
\]

\[
\propto (2\pi)^{-\frac{N}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( X - Z_1 \Phi^T \right)^T \left( H^{-1} \right) \left[ \text{vec} \left( X - Z_1 \Phi^T \right) \right] \right] \right] \times
\]

\[
\left[ \frac{\Gamma_k}{k!} \right]^{\frac{\gamma_0}{2}} I_k(k) I_D(d) \frac{\beta_0^\gamma}{\Gamma(\alpha_0^\gamma)} \gamma^\alpha \exp \left[ \frac{-\gamma \beta_0}{2\sigma_0^2} \right] \times
\]

\[
\left[ \frac{\Gamma_k}{k!} \right]^{\frac{\gamma_0}{2}} I_k(k) I_D(d) \frac{\beta_0^\gamma}{\Gamma(\alpha_0^\gamma)} \gamma^\alpha \exp \left[ -\frac{1}{2\sigma_0^2} (c - \mu_c)^2 \right]
\]

(4.4.73)

The distribution in (4.4.73) is in an intractable form. In order to simulate samples from it, we require expressions for several conditional posterior distributions so that it can be incorporated into an MCMC algorithm. We shall employ the conditional posterior distributions outlined below:

\[
p(M_k, \Phi|X, d, \gamma, c, N_j, A, C)
\]

\[
\propto (2\pi)^{-\frac{N}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( X - Z_1 \Phi^T \right)^T \left( H^{-1} \right) \left[ \text{vec} \left( X - Z_1 \Phi^T \right) \right] \right] \right] \times
\]

\[
\left[ \frac{\Gamma_k}{k!} \right]^{\frac{\gamma_0}{2}} I_k(k) I_D(d) \frac{\beta_0^\gamma}{\Gamma(\alpha_0^\gamma)} \gamma^\alpha \exp \left[ -\frac{1}{2\sigma_0^2} (c - \mu_c)^2 \right]
\]

\[
p(\Phi|X, M_k, d, \gamma, c, N_j, A, C)
\]
\[ p(d|X, M_k, \Phi, \gamma, c, N_j, A, C) \]
\[ p(\gamma, c|X, M_k, \Phi, d, N_j, A, C) \]
\[ p(N_j, A, C|X, M_k, \Phi, d, c) \]
\[ p(A, C|X, M_k, \Phi, d, c, N_j) \]

Several indicator functions that take the value of one when their conditions are met, and otherwise zero, are included in the expression for the joint posterior distribution in (4.4.73). These functions have been omitted from the conditional posterior distributions shown in (4.4.74). This is because, when coding the algorithm, only values that meet the required conditions will be proposed.

All of the posterior distributions in (4.4.74) are of non-standard forms, except for the posterior distribution for the delay parameter \(d\), which is simply a discrete distribution. This implies that none of the stages in the Gibbs Sampler employs a conjugate posterior distribution, and therefore a Metropolis-Hastings, an RJMCMC or a DRMH algorithm will be used to simulate the parameters. The posterior simulation algorithm is explained in the following section.

### 4.4.4 Posterior Simulator

The posterior simulator for the M-STAR-GARCH model will first simulate from the conditional posterior distributions for the conditional mean parameters and then from the distributions for the conditional covariance parameters. Reversible Jump steps will be used for the conditional posterior distributions of each model index and their respective coefficient parameters.

The estimation scheme for the parameters of the M-STAR-GARCH model using the BEKK covariance equation is outlined below.
M-STAR-GARCH - Main Algorithm

1. Set the initial values, \( \left( M_k^{(0)}, \Phi^{(0)}, d^{(0)}, \gamma^{(0)}, \epsilon^{(0)}, N_j^{(0)}, A^{(0)}, C^{(0)} \right) \), either randomly or deterministically and set \( i = 1 \).

2. At the \( i \)th iteration, execute the following steps employing the posterior distributions shown in (4.4.74):

(a) (i) **RJ**\(_M\) **Step** - Simulate the STAR model order and conditional mean coefficient parameter matrix using

\[
(M_k^{(i)}, \Phi^{(i)}) \sim p(M_k, \Phi | X, d^{(i-1)}, \gamma^{(i-1)}, \epsilon^{(i-1)}, N_j^{(i-1)}, A^{(i-1)}, C^{(i-1)}).
\]

(ii) **MH**\(_\Phi\) **Step** - If the proposal in step 2(a)(i) is rejected, simulate the conditional mean coefficient parameter matrix using

\[
\Phi^{(i)} \sim p(\Phi | X, M_k^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, \epsilon^{(i-1)}, N_j^{(i-1)}, A^{(i-1)}, C^{(i-1)}).
\]

(b) Simulate the delay parameter of the transition function using

\[
d^{(i)} \sim p(d | X, M_k^{(i)}, \Phi^{(i)}, \gamma^{(i-1)}, \epsilon^{(i-1)}, N_j^{(i-1)}, A^{(i-1)}, C^{(i-1)}).
\]

(c) **MH\(_{\gamma, \epsilon}\) Step** - Simulate the smoothing and location parameters of the transition function using

\[
(\gamma^{(i)}, \epsilon^{(i)}) \sim p(\gamma, \epsilon | X, M_k^{(i)}, \Phi^{(i)}, d^{(i)}, N_j^{(i-1)}, A^{(i-1)}, C^{(i-1)}).
\]

(d) (i) **RJ**\(_A\) **Step** - Simulate the GARCH model index and coefficient parameter matrix using

\[
(N_j^{(i)}, A^{(i)}, C^{(i)}) \sim p(N_j, A, C | X, M_k^{(i)}, \Phi^{(i)}, d^{(i)}, \gamma^{(i)}, \epsilon^{(i)}).
\]

(ii) **DR**\(_A\) **Step** - If the proposal in step 2(d)(i) is rejected, simulate the coefficient parameter matrix from the full conditional posterior distribution:

\[
(A^{(i)}, C^{(i)}) \sim p(A, C | X, M_k^{(i)}, \Phi^{(i)}, d^{(i)}, \gamma^{(i)}, \epsilon^{(i)}, N_j^{(i)}).
\]

3. Set \( i = i + 1 \). If all the MCMC chains have converged, STOP, else, move to step 2.

Within the main algorithm above, there are several sub-algorithms requiring further explanation. The Reversible Jump algorithms for the conditional mean model index, coefficient parameter matrix, the
conditional covariance model index and the coefficient parameter matrix will be explained in the following sections.

**Reversible Jump Steps**

The Reversible Jump steps shown in 2(a) and 2(d) in the main algorithm above will be implemented similarly to the ones in Sections 4.2.4 and 4.3.6, respectively. If the candidate model index is rejected at the Reversible Jump step, then the relevant coefficient parameter will be simulated from its full conditional posterior distribution. This is to ensure that there is good mixing of the MCMC chains within a particular model space.

The general acceptance probability given in (2.2.18) for an RJMCMC algorithm in which the full coefficient parameter is proposed without a transformation and a symmetric jumping distribution is employed is

$$r = \min \left(1, \frac{\pi(\theta^*_{m^*}, m^*|x) q(\theta^*|\theta^*, m, m^*)}{\pi(\theta_m, m|x) q(\theta^*|\theta, m, m^*)} \right).$$

(4.4.75)

The general form of the jumping distribution used for the following RJMCMC algorithms is the discretised Laplacian distribution that has been used throughout this thesis. This distribution is given by

$$j(m^*|m) \propto \exp \left[-\Delta_m|m^* - m|\right].$$

(4.4.76)

The following sections will elaborate on each of the Reversible Jump steps.

**Reversible Jump - RJ$_M$ Step**

The joint conditional posterior distribution $p(M_k, \Phi|X, d, \gamma, c, N_j, A, C)$ is shown in (4.4.74). As for the other estimation algorithms, given that jumping to a new model index $k^*$ results in a change in dimension of the coefficient parameter matrix $\Phi$, an RJMCMC algorithm will be required.

Substituting the required target distribution for this step into the acceptance probability from (4.4.75), and using the required notation for the proposal distribution yields the following form for the acceptance probability:

$$r_{(M_k, MSG)} = \min \left(1, \frac{p(M_{k^*}, \Phi^*|X, d, \gamma, c, N_j, A, C) q(\Phi^*|\Phi^*, M_k, M_{k^*})}{p(M_k, \Phi|X, d, \gamma, c, N_j, A, C) q(\Phi^*|\Phi, M_k, M_{k^*})} \right).$$

(4.4.77)

Unfortunately, we are unable to use the Candidate’s Identity to simplify the acceptance probability here as the integration of the marginal likelihood is not tractable. Therefore, the acceptance of a
model move needs to be performed by not only proposing a candidate model index, but also proposing a candidate conditional mean coefficient matrix and testing them both using the acceptance probability formula in (4.4.77).

Making sensible proposals for the conditional mean coefficient matrix will be the key to this algorithm. In order to find a suitable proposal distribution, we shall use the full conditional posterior distribution for the parameter coefficient matrix, but make the assumption that the conditional covariance is constant over time. That is, we substitute a new matrix \( \hat{H} \) for \( H \), where \( \hat{H} \) is assumed to be a block diagonal matrix in which each of the blocks on the diagonal is set to the unconditional covariance of the data set. Therefore, the full conditional posterior for \( \Phi \) is given by

\[
p(\Phi | X, M_k, d, \gamma, c, N_\gamma, A, C) \propto (2\pi)^{-\frac{mk}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T \left( H^{-1} \right) \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) \right]
\]

(4.4.78)

where \( \hat{H} \) represents the unconditional covariance of the data.

The distribution in (4.4.78) is similar to the one in equation (4.2.29) from Section 4.2.3. However, rather than including the covariance parameter \( \Sigma \), we include the sample covariance of the data, \( \hat{H} \). Equation (4.4.78) may be manipulated in the same way as in Section 4.2.3, resulting in a distribution that is proportional to a Matrix Normal distribution. This Matrix Normal distribution is a good starting point for a proposal distribution for the conditional mean coefficient matrix. The only addition to the distribution will be a tuning parameter \( \Delta_\Phi \) for the acceptance probability. Therefore, the proposal distribution will be

\[
q(\Phi^* | \Phi, M_k, M_k^*) = \frac{\exp \left[ -\frac{1}{2} \text{tr} \left( \left( \Phi^* - \hat{\Phi} \right)^T C_\Phi^{-1} \left( \Phi^* - \hat{\Phi} \right) \left( \Delta_\Phi \hat{H} \right)^{-1} \right) \right]}{(2\pi)^{\frac{mk^2}{2}} \left| \Delta_\Phi \hat{H} \right|^{\frac{mk^2}{2}} |C_\Phi^*|^{\frac{1}{2}}} \quad (4.4.79)
\]

\[
q(\Phi^* | \Phi, M_k, M_k^*) \sim N_{2p^* \times p} \left( \hat{\Phi}^*, C_\Phi^*, \Delta_\Phi \hat{H} \right),
\]

where

\[
\hat{\Phi}^* = C_\Phi^* Z_1^T X
\]

\[
C_\Phi^{-1} = Z_1^T Z_1.
\]
To derive the required formula for the Reversible Jump step, we first simplify the quotient on the right hand side of the acceptance probability in (4.4.77):

\[
p(M_k, \Phi^* | X, d, \gamma, c, N, A, C) q(\Phi | \Phi^*, M_k, M_{k+}) p(M_k, \Phi | X, d, \gamma, c, N, A, C) q(\Phi^* | \Phi, M_k, M_{k+}) = \frac{(2\pi)^{-n/2} |H^*|^{-1/2} \exp \left[ -\frac{1}{2} \left( X^T (X - Z_1 \Phi^*) \right)^T \left( X^T (X - Z_1 \Phi^*) \right) \right]}{(2\pi)^{-n/2} |H|^{-1/2} \exp \left[ -\frac{1}{2} \left( X^T (X - Z_1 \Phi) \right)^T \left( X^T (X - Z_1 \Phi) \right) \right]} \times \\
\left( M_k^* \right)^{\tau_k} \exp \left[ -\frac{1}{2} \text{tr} \left( \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) [\Delta_\Phi H]^{-1} \right) \right] \times \\
\left( M_{k+} \right)^{\tau_k} \exp \left[ -\frac{1}{2} \text{tr} \left( (\Phi^* - \hat{\Phi}^*)^T C_\phi^{-1} (\Phi^* - \hat{\Phi}^*) [\Delta_\Phi H]^{-1} \right) \right] \\
= (2\pi)^{k^p} (2\pi)^{-k^p} \left| \Delta_\Phi H \right|^{k^p} \left| \Delta_\Phi H \right|^{-k^p} \left| C_\phi \right|^{\frac{\tau_k}{2}} \left| C_\phi \right|^{-\frac{\tau_k}{2}} |H^*|^{-1/2} |H|^{-1/2} \frac{\left( \Lambda_k^{k^p - k^p} \right)^{\tau_k}}{\left( \Lambda_k^{k^p + k^p} \right)^{\tau_k}} \times \\
\exp \left[ -\frac{1}{2} \left( \left( X^T (X - Z_1 \Phi^*) \right)^T \left( X^T (X - Z_1 \Phi^*) \right) \right) \right] - \\
\exp \left[ -\frac{1}{2} \left( \left( X^T (X - Z_1 \Phi) \right)^T \left( X^T (X - Z_1 \Phi) \right) \right) \right] + \\
\text{tr} \left( \left( \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) - (\Phi^* - \hat{\Phi}^*)^T C_\phi^{-1} (\Phi^* - \hat{\Phi}^*) \right) [\Delta_\Phi H]^{-1} \right) \right].
\]

The acceptance probability for the Reversible Jump step for the conditional mean parameters is therefore

\[ r_{(M_k, M_{k+})} = \text{min} \left( 1, (2\pi)^{2(k^p - k)} \left| \Delta_\Phi H \right|^{2(k^p - k)} \left| C_\phi \right| \left| C_\phi \right|^{-1} \right)^{\frac{\tau_k}{2}} \left| H^* \right|^{-1/2} \left| H \right|^{-1/2} \frac{\left( \Lambda_k^{k^p - k^p} k^{k^p} \right)^{\tau_k}}{\left( \Lambda_k^{k^p + k^p} k^{k^p} \right)^{\tau_k}} \times \\
\exp \left[ -\frac{1}{2} \left( \left( X^T (X - Z_1 \Phi^*) \right)^T \left( X^T (X - Z_1 \Phi^*) \right) \right) \right] - \\
\exp \left[ -\frac{1}{2} \left( \left( X^T (X - Z_1 \Phi) \right)^T \left( X^T (X - Z_1 \Phi) \right) \right) \right] + \\
\text{tr} \left( \left( \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) - (\Phi^* - \hat{\Phi}^*)^T C_\phi^{-1} (\Phi^* - \hat{\Phi}^*) \right) [\Delta_\Phi H]^{-1} \right) \right). \]
The structure of the RJMCMC algorithm for step 2(a)(i) for the conditional mean model index and conditional mean coefficient parameter matrix is outlined below.

**RJ Algorithm**

1. The current state is \( (\mathcal{M}_k^{(i-1)}, \Phi^{(i-1)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)}, N_j^{(i-1)}, A^{(i-1)}, C^{(i-1)}) \).

2. Propose a candidate model, \( \mathcal{M}_k^* \) from \( j(\mathcal{M}_k^*|\mathcal{M}_k) \) using the discretised Laplacian shown in (4.4.76). Propose a new conditional mean coefficient parameter matrix using the proposal distribution given in (4.4.79).

3. Calculate the acceptance probability, \( r(\mathcal{M}_k, MSG) \), per (4.4.80).

4. Simulate \( u \sim U(0, 1) \):
   
   (a) If \( r(\mathcal{M}_k, MSG) > u \), then:
   
   (i) Accept the candidate model index \( \mathcal{M}_k^* \) and set \( \mathcal{M}_k^{(i)} = \mathcal{M}_k^* \).
   
   (ii) Accept the proposed conditional mean coefficient parameter matrix \( \Phi^* \) and set \( \Phi^{(i)} = \Phi^* \).

   (b) If \( r(\mathcal{M}_k, MSG) < u \), then:
   
   (i) Reject the candidate model index \( \mathcal{M}_k^* \) and conditional mean coefficient parameter matrix \( \Phi^* \). Set \( \mathcal{M}_k^{(i)} = \mathcal{M}_k^{(i-1)} \) and \( \Phi^{(i)} = \Phi^{(i-1)} \).

   (ii) Using the distribution in (4.4.74), simulate the conditional mean coefficient parameter matrix,

   \[
   \Phi^{(i)} \sim \frac{p(\Phi|\mathcal{M}_k^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)}, N_j^{(i-1)}, A^{(i-1)}, C^{(i-1)})}{p(\Phi^{(i)}, \mathcal{M}_k^{(i)}, d^{(i-1)}, \gamma^{(i-1)}, c^{(i-1)}, N_j^{(i-1)}, A^{(i-1)}, C^{(i-1)})},
   \]

   using the Metropolis-Hastings algorithm described below (\( \text{MH}_\Phi \) Step).

**Reversible Jump - RJ\( \mathcal{N} \) Step**

Step 2(d)(ii) of the main algorithm is a Reversible Jump algorithm for the conditional covariance model index and coefficient parameter matrix. Using the notation relevant to the conditional covariance of the M-STAR-GARCH model and substituting it into the general form of a Reversible Jump acceptance probability shown in (4.4.75) yields:

\[
\begin{align*}
    r(N_j, j \rightarrow j^*, MSG) &= \min \left( 1, \frac{p(N_j^*, A^*, C^*, N_j^*, \Phi, d, \gamma, c|\mathcal{M}_k, A, C, \mathcal{M}_k, \Phi, d, \gamma, c)}{p(N_j, A, C|\mathcal{M}_k, \Phi, d, \gamma, c)} \right) \left( \frac{q(A^*, C^*|A, C, N_j, N_j^*)}{q(A, C|A^*, C^*, N_j, N_j^*)} \right) \right),
\end{align*}
\]  

(4.4.81)

We will assume that the matrices \( A \) and \( C \) are independent, and that their proposal distribution can
be decomposed as follows:

\[
q(A^*, C^* \mid A, C, N_j, N_j^*) = q(A^* \mid A, N_j, N_j^*) q(C^* \mid C, N_j, N_j^*).
\]

As with the pure GARCH model shown in Section 4.3.6, the proposals of the intercept matrix \( C \) and coefficient matrix \( A \) will be performed separately. Because the intercept matrix \( C \) is required to be positive definite, it will be proposed from a Wishart distribution. The proposal for the coefficient matrix \( A \) of the BEKK model will be taken from a matrix normal distribution as in Section 4.3. When a proposal is made for a move to the DBEKK model, a multivariate normal distribution will be used to determine the diagonal entries of each of the individual coefficient matrices.

For the pure M-GARCH model, determining suitable location parameters for the proposal distributions is essential to the success of the algorithm. A pilot run was used to obtain reasonable estimates that will allow the full algorithm to make model moves under the Reversible Jump scheme. For the M-STAR-GARCH model, the additional complexity of a changing conditional mean makes this task even more difficult.

Within the univariate setting, it was found that, for the STAR-GARCH processes, simply using the MLE’s of the GARCH parameters assuming a constant conditional mean was ineffective due to the misspecification of the conditional mean. This would lead to poor specification of the conditional variance location parameters and, consequently, to poor mixing and the failure of the algorithm to converge on the true model in any reasonable amount of time. For the algorithm to be successful, the location parameters need to be specified in areas of high posterior probability so that model moves are relatively likely to be accepted.

With this in mind, we shall employ a slightly different strategy from the one used for the univariate STAR-GARCH model. Two pilot runs were employed to estimate parameters for the univariate STAR-GARCH model. The first specified temporary parameters for the conditional mean. These parameters were then used in a second pilot run to determine the location parameters for the intercept and coefficient matrices for the conditional covariance part of the model. We initially tried this strategy on the M-STAR-GARCH model, but found it to be less successful than the single pilot run strategy eventually settled upon.

To estimate parameters for the M-STAR-GARCH model, we will fix a pilot run in which the model order of the conditional mean is able to jump between models, but the conditional covariance of the model is fixed. This will be done to determine the location parameters of the proposal distributions for each of the models shown in Table 4.2.
The proposal distributions that will be used in the RJ\textsubscript{N} Step are

\[
q(\mathbf{A}|\mathbf{A}, \mathcal{N}_{j}, \mathcal{N}_{j^*}) = 1 \quad \text{(Constant)}
\]

\[
q(\mathbf{A}|\mathbf{A}, \mathcal{N}_{j}, \mathcal{N}_{j^*}) \sim \mathcal{N}_{l+m}(m_{\mathcal{N}_{j^*}}, V_{\mathcal{N}_{j^*}}) \quad \text{(DBEKK)}
\]

\[
q(\mathbf{A}|\mathbf{A}, \mathcal{N}_{j}, \mathcal{N}_{j^*}) \sim \mathcal{N}_{p,p}(l+m)(M_{\mathcal{N}_{j^*}}, U_{\mathcal{N}_{j^*}}, V_{\mathcal{N}_{j^*}}) \quad \text{(BEKK)}
\]

\[
q(C^*|C, \mathcal{N}_{j}, \mathcal{N}_{j^*}) \sim W_{p}
\left(\Delta_{C_{\mathcal{N}_{j^*}}}, C_{\mathcal{N}_{j^*}}, \Delta_{C_{\mathcal{N}_{j^*}}}
\right).
\]

The Wishart distribution in (4.4.82) has a location parameter of $C_{\mathcal{N}_{j^*}}$. This location parameter, along with the location parameters for the proposal distribution of the conditional covariance coefficient parameter matrix, $M_{\mathcal{N}_{j^*}}$ or $m_{\mathcal{N}_{j^*}}$, are determined from the pilot run. The parameters have the subscript $\mathcal{N}_{j^*}$ to remind the reader that they may potentially be different for each model index. The variance parameters $U_{\mathcal{N}_{j^*}}$, $V_{\mathcal{N}_{j^*}}$ and $\Delta_{C_{\mathcal{N}_{j^*}}}$ can roughly be determined from the pilot run, but may require some adjustment to ensure an adequate acceptance rate for the algorithm.

The proposal distributions shown in (4.4.82) for the conditional covariance coefficient parameter matrix are used in the general case where the RJMCMC scheme allows not only jumps from different model types, but also jumps to differing model orders within a particular model. For practical reasons, the estimation algorithm implemented will only consider the models shown in Table 4.2. That is, the model orders $l$ and $m$ will be fixed.

The most complicated acceptance probability expression allows a change in either or both model orders, and is the one that involves a move within the BEKK model. The acceptance probability for such a move has been determined in Appendix F.1, and its equation is given in (F.1.1).

As mentioned above, the models considered here are shown in Table 4.2. When there is the proposal to move to a candidate model, the proposal distributions for $\mathbf{A}$ will be different for the candidate model and the current model. This means that a Reversible Jump step has one of six possible acceptance probabilities of differing mathematical forms, depending on the current and the candidate model. Each of these acceptance probabilities is determined in Appendix F.1, and the resulting formulas are shown in (F.1.2), (F.1.3), (F.1.4), (F.1.5), (F.1.6) and (F.1.7).

If the calculations described in Appendix F.1 are performed directly, computational problems associated with overflow and underflow are likely to occur, especially for large data sets. Using the built in function for the Wishart distribution to calculate the ratio of the proposal distribution for the intercept matrix
Chapter 4: Multivariate Models

also results in problems. To avoid these problems, the logarithm of the acceptance probability formula is taken, and then as much of the expression as possible is calculated before applying the exponential to produce the final probability expression.

We can now outline the RJN Step.

RJN Algorithm

1. The current state is \( \left( M_k^{(i)}, \Phi^{(i)}, d^{(i)}, \gamma^{(i)}, c^{(i)}_k, N_{i-1}^{(i)} \right) \).

2. Propose a candidate model index, \( N_j^* \) from \( N_j \) using the discretised Laplacian shown in (4.4.76). Propose candidate conditional covariance intercept and coefficient parameter matrices using the proposal distributions in (4.4.82).

3. Determine the appropriate acceptance probability to be used, based on \( N_j \) and \( N_j^* \), and calculate the acceptance probability \( r(N_j, N_j^* \rightarrow MMSG) \) using one of (F.1.2), (F.1.3), (F.1.4), (F.1.5), (F.1.6) and (F.1.7).

4. Simulate \( u \sim U(0, 1) \):

   (a) If \( r(N_j, N_j^* \rightarrow MMSG) > u \), then:

      (i) Accept the candidate model index \( N_j^* \) and set \( N_j^{(i)} = N_j^* \).

      (ii) Accept the proposed conditional covariance intercept and coefficient parameter matrices. Set \( C^{(i)} = C^* \) and \( A^{(i)} = A^* \).

   (b) If \( r(N_j, N_j^* \rightarrow MMSG) < u \), then:

      (i) Reject the candidate model index \( N_j^* \), proposed conditional covariance intercept and coefficient parameter matrix. Set \( N_j^{(i)} = N_j^{(i-1)} \), \( C^{(i)} = C^{(i-1)} \) and \( A^{(i)} = A^{(i-1)} \).

      (ii) Simulate the conditional covariance intercept and coefficient parameter matrices using

      \[
      (A, C)^{(i)} \sim p \left( A, C | X, M_k^{(i)}, \Phi^{(i)}, d^{(i)}, \gamma^{(i)}, c^{(i)}_k, N_j^{(i)} \right)
      \]

      and employing a DRM algorithm (DR\( A \) Step).

Metropolis-Hastings Steps

The simple Metropolis-Hastings algorithm will be employed for estimation of the conditional mean coefficient parameter matrix in the event of the proposed move to model \( M_k^* \) being rejected. The same algorithm will also be used for the joint conditional posterior distribution of the smoothing and location parameters of the transition functions. Given the additional difficulty in simulating the conditional covariance coefficient parameter matrix, we shall employ a DRM algorithm for this step.
As a reminder, the general form of the acceptance probability originally noted in (2.2.12) is
\[ r = \min \left( 1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^*|\theta^{(i-1)})} \right). \] (4.4.83)

We shall first discuss the algorithm for \( \Phi \) and follow this with the algorithm for \( (\gamma, c) \).

**Metropolis-Hastings - MH\( \Phi \) Step**

For the pure M-STAR model in Section 4.2, the conditional posterior distribution for \( \Phi \) is of a standard form that is relatively easy to sample from. The conditional posterior for the M-STAR-GARCH model is complicated by the non-constant conditional covariance and is of a non-standard form. Consequently, we shall use the Metropolis-Hastings algorithm to sample the conditional posterior distribution. While other more complicated algorithms such as the DRMH algorithm could be used here, testing indicated that the simple Metropolis-Hastings algorithm was more than adequate and relatively fast in its implementation.

The proposal distribution used here will be similar to the one used for the RJ\( M_k \) Step discussed above. The key difference is that the location parameter for the proposal distribution will be the most recent conditional mean coefficient parameter matrix in the MCMC chain. Therefore, the proposal distribution shall be a Matrix Normal distribution such that
\[ q(\Phi^*|\Phi) \sim N_{2pk,p}(\Phi, C_\Phi, \Delta \bar{H}), \] (4.4.84)

where
\[ \Phi = \Phi^{(i-1)} \]
\[ C_\Phi^{-1} = Z_1^T Z_1. \]

Let the most recent simulated parameter in the MCMC chain be \( \theta = \theta^{(i-1)} \). We begin by manipulating the quotient on the right hand side of the acceptance probability in (4.4.83). The first step is to substitute in the expressions for the target distribution and proposal distribution in (4.4.74) and (4.4.84), respectively, as follows:
\[
\frac{\pi(\theta^*)}{\pi(\theta)} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^*|\theta^{(i-1)})} = \frac{p(\Phi^*|X, M_k, d, \gamma, c, N_j, A, C)}{p(\Phi|X, M_k, d, \gamma, c, N_j, A, C)} \frac{q(\Phi^*|\Phi)}{q(\Phi^*|\Phi)}
\]
(2\pi)^{-\frac{n}{2}} |H^*|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \left[ \text{vec} \left( [X - Z_1 \Phi^*]^T \right) \right]^T [H^*]^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi^*]^T \right) \right] \right) 

\times \frac{\exp \left(-\frac{1}{2} \text{tr} \left( (\Phi - \Phi^*)^T C_{\Phi}^{-1} (\Phi - \Phi^*) \right) (H - \bar{H})^{-1} \right)}{(2\pi)^{\frac{n}{2}} |\Delta_\Phi|^{\frac{n}{2}} |C_{\Phi}|^{\frac{n}{2}}}

\exp \left(-\frac{1}{2} \text{tr} \left( (\Phi^* - \Phi)^T C_{\Phi}^{-1} (\Phi^* - \Phi) \right) (\Delta_\Phi) \right)

(2\pi)^{-\frac{n}{2}} |H^{-1}|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T [H]^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] \right)

\left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T [H^{-1}] \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] - \left[ \text{vec} \left( [X - Z_1 \Phi^*]^T \right) \right]^T [H^*]^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi^*]^T \right) \right] 

Therefore, the acceptance probability for the MH_\Phi Algorithm is

\[ r(\Phi, \text{MSG}) \]

= \min \left( 1, \exp \left(-\frac{1}{2} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T [H]^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] \right) - \left[ \text{vec} \left( [X - Z_1 \Phi^*]^T \right) \right]^T [H^*]^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi^*]^T \right) \right] \right) \right)^{\frac{1}{2}} \right)

\[ (4.4.85) \]

The outline of the MH_\Phi Algorithm is given below.

**MH_\Phi Algorithm**

1. The current state is \( \left( M_k^{(i)}, \Phi^{(i-1)}, d^{(i-1)}, \gamma^{(i-1)}, \epsilon^{(i-1)}, N_j^{(i-1)}, \alpha^{(i-1)} \right) \).

2. Propose a new conditional mean coefficient parameter \( \Phi \) using \( q(\Phi^* | \Phi) \) in (3.3.50).

3. Calculate the acceptance probability \( r(\Phi, \text{MSG}) \) in (4.4.85).

4. Simulate \( u \sim U(0, 1) \):

   (a) If \( r(\Phi, \text{MSG}) > u \), accept the proposed conditional mean coefficient parameter matrix and set \( \Phi^{(i)} = \Phi^* \).

   (b) If \( r(\Phi, \text{MSG}) < u \), reject the proposed conditional mean coefficient parameter matrix and set \( \Phi^{(i)} = \Phi^{(i-1)} \).

---

126
Metropolis-Hastings - MH\((\gamma,c)\) Step

The proposal distributions for this step will be the same as those discussed in Section 4.2. They are a Gamma distribution and a Normal distribution for the smoothing and location parameters, respectively. That is,

\[
q(\gamma^*|\gamma) \sim \mathcal{G} \left( \frac{\gamma^2}{\Delta_\gamma}, \frac{\gamma}{\Delta_\gamma} \right) \\
q(c^*|c) \sim \mathcal{N}(c, \Delta_c),
\]

where the most recently simulated values are \(\gamma = \gamma^{(i-1)}\) and \(c = c^{(i-1)}\).

The target distribution for the algorithm is the joint conditional posterior distribution given in (4.4.74). Formulas for the densities of the proposal distributions are given in (4.4.86). We can use these to determine the form of the quotient on the right hand side of the acceptance probability in (4.4.83). The symmetry of the distribution allows us to ignore the ratio of the proposal for \(c\). We calculate

\[
p(\gamma^*, c^*|X, M_k, \Phi, \Sigma, d, N_j, A, C) q(c^*|c) q(\gamma^*|\gamma) \\
P(\gamma, c|X, M_k, \Phi, \Sigma, d, N_j, A, C) q(c^*|c) q(\gamma^*|\gamma)
\]

\[
= \frac{(2\pi)^{-\frac{n}{2} - \frac{d}{2}} |H^*|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \right)^T \left( \begin{bmatrix} H^* & \Sigma \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \end{bmatrix} \right) \right]}{(2\pi)^{-\frac{n}{2} - \frac{d}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \right)^T \left( \begin{bmatrix} H & \Sigma \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \end{bmatrix} \right) \right]} \times
\]

\[
\frac{\beta^{\gamma^*}_\gamma \gamma^{\alpha^* - 1} \exp \left[ -\gamma^* \beta \gamma \right] [2\pi \sigma^2]^{-\frac{d}{2}} \exp \left[ -\frac{1}{2\alpha} (c^* - \mu_c)^2 \right]}{\beta^{\gamma}_\gamma \gamma^{\alpha - 1} \exp \left[ -\gamma \beta \gamma \right] [2\pi \sigma^2]^{-\frac{d}{2}} \exp \left[ -\frac{1}{2\alpha} (c - \mu_c)^2 \right]}
\]

\[
= \left| H^* \right|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \right)^T \left( \begin{bmatrix} H^* & \Sigma \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \end{bmatrix} \right) \right] \times
\]

\[
\left| H \right|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \right)^T \left( \begin{bmatrix} H & \Sigma \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \end{bmatrix} \right) \right]
\]

\[
\gamma^{\alpha^* - 1} \exp \left[ -\gamma^* \beta \gamma \right] \exp \left[ -\frac{1}{2\alpha} (c^* - \mu_c)^2 \right] \frac{\gamma^2}{\gamma^2 + \frac{2\pi}{\alpha}} \gamma^{\alpha^* - 1}
\]

\[
\gamma^{\alpha - 1} \exp \left[ -\gamma \beta \gamma \right] \exp \left[ -\frac{1}{2\alpha} (c - \mu_c)^2 \right] \frac{\gamma^2}{\gamma^2 + \frac{2\pi}{\alpha}} \gamma^{\alpha - 1}
\]

\[
= \left| H^* \right| \left| H \right|^{-\frac{1}{2}} \left[ \frac{\gamma^*}{\gamma} \right]^{\alpha^* - 1} \frac{\gamma^2}{\gamma^2 + \frac{2\pi}{\alpha}} \gamma^{\alpha^* - 1} \times
\]

\[
\frac{\gamma^2}{\gamma^2 + \frac{2\pi}{\alpha}} \gamma^{\alpha^* - 1} \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \right)^T \left( \begin{bmatrix} H^* & \Sigma \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \end{bmatrix} \right) \right] \times
\]

\[
\exp \left[ \frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \right)^T \left( \begin{bmatrix} H & \Sigma \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} \text{vec} \left( \left[ X - Z_i^\dagger \Phi \right]^T \right) \end{bmatrix} \right) \right]
\]
The full form of the acceptance probability for the Metropolis-Hastings step is shown below:

\[
\begin{align*}
\exp[-\gamma^* \beta_N] \exp \left[-\frac{1}{2\sigma_N^2} (c^* - \mu_N)^2\right] \exp [\gamma^i \beta_N] \exp \left[\frac{1}{2\sigma_N^2} (c - \mu_N)^2\right] \\
= \left[|H^*| |H|^{-1}\right]^{-\frac{1}{2}} \left[\frac{\gamma^*}{\gamma}\right]^{\alpha_{i,\gamma} - 1} \frac{\Gamma \left(\frac{\alpha_{i,\gamma}}{2}\right) \frac{\gamma^*}{\gamma} \frac{\gamma^*}{\gamma}}{\Gamma \left(\frac{\alpha_{i,\gamma}}{2}\right) \frac{\gamma^*}{\gamma} \frac{\gamma^*}{\gamma}} \times \\
\exp \left[-\frac{1}{2} \left(\text{vec} \left((X - Z_i \Phi)^T\right)^T \left[|H^*| \left[\text{vec} \left((X - Z_i \Phi)^T\right)\right]\right]\right]ight. + \\
\left.\frac{1}{2} \left(\text{vec} \left((X - Z_i \Phi)^T\right)^T \left[|H^{-1}| \left[\text{vec} \left((X - Z_i \Phi)^T\right)\right]\right]\right) - \\
\gamma^* \beta_N - \frac{1}{2\sigma_N^2} (c^* - \mu_N)^2 + \gamma^* \beta_N + \frac{1}{2\sigma_N^2} (c - \mu_N)^2 \\
= \left[|H^*| |H|^{-1}\right]^{-\frac{1}{2}} \left[\frac{\gamma^*}{\gamma}\right]^{\alpha_{i,\gamma} - 1} \frac{\Gamma \left(\frac{\alpha_{i,\gamma}}{2}\right) \frac{\gamma^*}{\gamma} \frac{\gamma^*}{\gamma}}{\Gamma \left(\frac{\alpha_{i,\gamma}}{2}\right) \frac{\gamma^*}{\gamma} \frac{\gamma^*}{\gamma}} \times \\
\exp \left[-\frac{1}{2} \left(\text{vec} \left((X - Z_i \Phi)^T\right)^T \left[|H^*| \left[\text{vec} \left((X - Z_i \Phi)^T\right)\right]\right]\right]ight. - \\
\left[\text{vec} \left((X - Z_i \Phi)^T\right)^T \left[|H^{-1}| \left[\text{vec} \left((X - Z_i \Phi)^T\right)\right]\right]\right] + \\
2\beta_N (\gamma^* - \gamma) + \frac{1}{\sigma_N^2} (c^* - \mu_N)^2 - (c - \mu_N)^2 \right). 
\end{align*}
\]

The above result is required to determine the acceptance of the proposed parameters \(\gamma^*\) and \(c^*\). The full form of the acceptance probability for the Metropolis-Hastings step is shown below:

\[
r_{(\gamma,c,MSG)} = \min \left(1, \left[|H^*| |H|^{-1}\right]^{-\frac{1}{2}} \left[\frac{\gamma^*}{\gamma}\right]^{\alpha_{i,\gamma} - 1} \frac{\Gamma \left(\frac{\alpha_{i,\gamma}}{2}\right) \frac{\gamma^*}{\gamma} \frac{\gamma^*}{\gamma}}{\Gamma \left(\frac{\alpha_{i,\gamma}}{2}\right) \frac{\gamma^*}{\gamma} \frac{\gamma^*}{\gamma}} \times \\
\exp \left[-\frac{1}{2} \left(\text{vec} \left((X - Z_i \Phi)^T\right)^T \left[|H^*| \left[\text{vec} \left((X - Z_i \Phi)^T\right)\right]\right]\right]ight. - \\
\left[\text{vec} \left((X - Z_i \Phi)^T\right)^T \left[|H^{-1}| \left[\text{vec} \left((X - Z_i \Phi)^T\right)\right]\right]\right] + \\
2\beta_N (\gamma^* - \gamma) + \frac{1}{\sigma_N^2} (c^* - \mu_N)^2 - (c - \mu_N)^2 \right). 
\]

We are now able to outline the algorithm for the Metropolis-Hastings step.

**MH\(_{(\gamma,c)}\)** Algorithm

1. The current state is \(\left(\mathcal{M}^{(i)}_k, \Phi^{(i)}, d^{(i)}, \gamma^{(i-1)}, c^{(i-1)}, \mathcal{M}^{(i-1)}_j, \mathbf{A}^{(i-1)}, C^{(i-1)}\right)\).

2. Propose new smoothing and location parameters \(\gamma^*\) and \(c^*\) from their respective proposal distributions shown in (4.4.86).

3. Calculate the acceptance probability \(r_{(\gamma,c,MSG)}\) per (4.4.87).

4. Simulate \(u \sim \mathcal{U}(0,1)\):
(a) If $r_{\gamma,c,MSG} > u$, accept the proposed smoothing and location parameters. Set $\gamma_i = \gamma^*$ and $c_i = c^*$.

(b) If $r_{\gamma,c,MSG} < u$, reject the proposed smoothing and location parameters. Set $\gamma_i = \gamma_{i-1}$ and $c_i = c_{i-1}$.

### Delayed Rejection Metropolis-Hastings - DR\(_A\) Step

When the move to the proposed model index is rejected in the RJ\(_M\) Step, the proposed conditional covariance coefficient parameter matrix and intercept matrix are also rejected. These must be simulated from their joint conditional posterior distribution, for otherwise, the intercept and coefficient matrices will only be updated upon model moves. If the posterior distribution shows significant evidence for one particular model index over the others, the algorithm may barely move from that model, so that the coefficient and intercept matrices will rarely be updated, leading to poor mixing and slow convergence.

Given the number of parameters, the amount of data that is required to estimate them and the difficulty in simulating the conditional covariance coefficient parameter matrix and intercept matrix, the DRMH algorithm will be used at this stage to help improve the speed of convergence.

As with the algorithm for the M-GARCH model, the DRMH algorithm allows the use of relatively large variances for the proposal distributions so that initially the algorithm moves quickly towards the area of convergence. It also allows small movements on the second iteration if the first proposal is made outside of the area of convergence. The DRMH algorithm is detailed in Section 2.2.3.

A two stage DRMH algorithm will be implemented here. Upon rejection of the first candidate values, a second stage is entered to test new candidate values. These new values can be proposed using information from the first stage candidate values. For the first stage of the algorithm, the proposal distributions will be similar to those used for the Reversible Jump step above, but will be centred on the most recently simulated value. The proposal distributions will be:

\[
q_{1A}(A^*|A) \sim N_{p,p(l+m)}(A, U_{N_j}^j, V_{N_j}^j) \\
q_{1C}(C^*|C) \sim W_p(\Delta_{C_{N_j}^j}^{-1} C, \Delta_{C_{N_j}^j} C_{N_j}).
\]

(4.4.88)

For the second stage of the proposal, we wish to use the information within the most recently simulated parameter as well as the candidate parameter that was just rejected in the first stage. The proposal distributions will be:

\[
q_{2A}(A^*|A^*, A) \sim N_{p,l+m}\left(\frac{1}{2} (A^* + A), U_{N_j}^j, V_{N_j}^j\right) \\
q_{2C}(C^*|C^*, C) \sim W_p\left([2\Delta_{C_{N_j}^j} \Delta_{2C}]^{-1} (C^* + C), \Delta_{C_{N_j}^j} \Delta_{2C}\right).
\]

(4.4.89)
where the parameters $\Delta_{2C}$ and $\Delta_{2A}$ are included to reduce the variance of the proposal in the second stage. Setting these values in the range of 0.50 to 0.75 yielded successful results during testing.

The target distribution for this step is the following joint conditional posterior distribution, which is also given in (4.4.74):

$$p(A, C|X, M_k, \Phi, d, \gamma, c, X')$$

$$\propto (2\pi)^{-\frac{np}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T H^{-1} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) \right].$$

Since there are three different situations in which this algorithm could be implemented, namely for the constant conditional covariance model, the DBEKK model and the BEKK model, only the most complicated calculation of the acceptance probability is presented. The difference between the presentations ultimately lies in the ratio of the proposal distributions for the conditional covariance coefficient parameter matrix.

The initial stage of the DRMH algorithm is simply an application of the usual Metropolis-Hastings algorithm. The form of the acceptance probability that takes into account the parameters of the M-STAR-GARCH model is

$$r_{(A, 1, \text{MSG})} (A^*, C^*, A, C)$$

$$= \min \left( 1, \frac{p(A^*, C^*|X, M_k, \Phi, d, \gamma, c, X') q_{1A}(A|A^*) q_{1C}(C|C^*)}{p(A, C|X, M_k, \Phi, d, \gamma, c, X') q_{1A}(A^*|A) q_{1C}(C^*|C)} \right).$$

Therefore, starting with the quotient on the right hand side in the acceptance probability in (4.4.90), we have

$$\frac{p(A^*, C^*|X, M_k, \Phi, d, \gamma, c, X') q_{1A}(A|A^*) q_{1C}(C|C^*)}{p(A, C|X, M_k, \Phi, d, \gamma, c, X') q_{1A}(A^*|A) q_{1C}(C^*|C)}$$

$$= (2\pi)^{-\frac{np}{2}} |H^*|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T H^*^{-1} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) \right]$$

$$\times \frac{\frac{1}{2} \left( \Delta_{C_N I}^{-p-1} \right) \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_N I}^{-1} C \right) \right]}{\frac{1}{2} \left( \Delta_{C_N I}^{-p-1} \right) \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_N I}^{-1} C^* \right) \right]}$$

$$\times \frac{\frac{1}{2} \left( \Delta_{C_N I}^{-p-1} \right) \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_N I}^{-1} C \right) \right]}{\frac{1}{2} \left( \Delta_{C_N I}^{-p-1} \right) \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_N I}^{-1} C^* \right) \right]}$$

$$= |H^*|^{-\frac{1}{2}} |H|^{\frac{1}{2}} \left| C \right| \left| C^* \right|^{-\frac{1}{2}} \frac{\Delta_{C_N I}^{-p-1} |C|^{-\frac{1}{2}} \Delta_{C_N I} |C|^{-\frac{1}{2}}}{\Delta_{C_N I}^{-p-1} |C|^{-\frac{1}{2}} \Delta_{C_N I} |C|^{-\frac{1}{2}}} \times$$

$$\exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T H^{-1} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) -$$

130
Section 4.4: Multivariate STAR-GARCH

\[
\left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right] + \text{tr} \left( \left[ \Delta_{C_{N_j}} C^* \right]^{-1} C - \left[ \Delta_{C_{N_j}} C \right]^{-1} C^* \right) \]

The acceptance probability for the first stage of the DRMH algorithm is calculated as follows:

\[
r_{(A,1,MSG)} (A^*, C^*, A, C) = \min \left( 1, |H^*|^{-\frac{1}{2}} |H|^{\frac{1}{2}} \left[ |C^*|^{-1} \right]^{-\frac{1}{2}} \left[ \Delta_{C_{N_j}} C^* \right] - \frac{\Delta_{C_{N_j}} p^{-1}}{2} \right) \left[ C^* \right]^{-\frac{1}{2}} \left[ \Delta_{C_{N_j}} C \right]^{-\frac{1}{2}} \times 
\]

\[
\exp \left[ -\frac{1}{2} \left( \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right] - \text{tr} \left( \left[ \Delta_{C_{N_j}} C^* \right]^{-1} C - \left[ \Delta_{C_{N_j}} C \right]^{-1} C^* \right) \right) \right] - (4.4.91)
\]

If the proposed parameters in the first stage of the algorithm are rejected, the algorithm continues into the second stage. The form of the acceptance probability for the second stage of the algorithm is a little more complicated. Using the notation of the M-STARD-GARCH model, the form of the acceptance probability becomes

\[
r_{(A,2,MSG)} (A', C', A^*, C^*, A, C) = \min \left( 1, \frac{p \left( A', C' | X, M_k, \Phi, d, \gamma, c, N_j \right) \cdot q_{1A} (A^* | A') \cdot q_{1C} (C^* | C') \cdot q_{2A} (A | A', A^*) \cdot q_{2C} (C | C, C^*) \right) \times \left( 1 - r_{(A,1,MSG)} (A^*, C^*, A', C') \right) \times \left( 1 - r_{(A,1,MSG)} (A^*, C^*, A, C) \right)
\]

(4.4.92)

Focusing on the quotient on the right hand side of the acceptance probability shown in (4.4.92), we calculate

\[
p \left( A', C' | X, M_k, \Phi, d, \gamma, c, N_j \right) \cdot q_{1A} (A^* | A') \cdot q_{1C} (C^* | C') \times 
\frac{q_{2A} (A | A', A^*) \cdot q_{2C} (C | C, C^*) \cdot \left( 1 - r_{(A,1,MSG)} (A^*, C^*, A', C') \right)}{q_{2A} (A | A, A^*) \cdot q_{2C} (C | C, C^*) \cdot \left( 1 - r_{(A,1,MSG)} (A^*, C^*, A, C) \right)}
\]

\[
= \frac{(2\pi)^{-\frac{np}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right] \right]}{(2\pi)^{-\frac{np}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right] \right]}
\]
\[
\begin{align*}
\exp \left[-\frac{1}{2} \text{tr} \left( V_{N_j}' (A^* - A')^T U_{N_j}^{-1} (A^* - A') \right) \right] |C|^{\frac{1}{2}} \left( \Delta_{CN_j} A_{2c-p-1} \right) \exp \left[-\frac{1}{2} \text{tr} \left( \Delta_{CN_j} C^{-1} C' \right) \right] \\
\left( 1 - r_{(\mathbf{A}, \text{MSG})} (A^*, C', A', C') \right) \\
\left( 1 - r_{(\mathbf{A}, \text{MSG})} (A^*, C^*, A', C') \right)
\end{align*}
\]

\[
= |H|^{-\frac{1}{2}} |H|^{\frac{1}{2}} \left[ \Delta_{CN_j} C \right]^{-\frac{1}{2}} \left[ \Delta_{CN_j} C' \right]^{-\frac{1}{2}} \times
\]

\[
\exp \left[-\frac{1}{2} \text{vec} \left( (X - Z_l \Phi)^T \right)^T H^{-1} \text{vec} \left( (X - Z_l \Phi)^T \right) \right] \times
\]

\[
\exp \left[-\frac{1}{2} \text{vec} \left( (X - Z_1 \Phi)^T \right)^T H^{-1} \text{vec} \left( (X - Z_1 \Phi)^T \right) \right] \times
\]

\[
\exp \left[-\frac{1}{2} \text{tr} \left( V_{N_j}' (A^* - A')^T U_{N_j}^{-1} (A^* - A') \right) \right] \times
\]

\[
\exp \left[-\frac{1}{2} \text{tr} \left( \Delta_{CN_j} C^{-1} C' \right) \right] \\
\left( 1 - r_{(\mathbf{A}, \text{MSG})} (A^*, C', A', C') \right)
\left( 1 - r_{(\mathbf{A}, \text{MSG})} (A^*, C^*, A', C') \right)
\]

\[
= |H|^{-\frac{1}{2}} |H|^{\frac{1}{2}} \left[ C \right] |C'|^{-\frac{1}{2}} \left[ C \right] |C'|^{-\frac{1}{2}} \left( \Delta_{CN_j} A_{2c-p-1} \right) \times
\]

\[
\exp \left[-\frac{1}{2} \text{tr} \left( \Delta_{CN_j} C^{-1} C' \right) \right] \\
\left( 1 - r_{(\mathbf{A}, \text{MSG})} (A^*, C', A', C') \right)
\left( 1 - r_{(\mathbf{A}, \text{MSG})} (A^*, C^*, A', C') \right)
\]
\[\left[ C + C^* \right]^{-\frac{\Delta}{2}} \times \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( (X - Z_1 \Phi)^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( (X - Z_1 \Phi)^T \right) \right] \right] \times \exp \left[ \frac{1}{2} \left[ \text{vec} \left( (X - Z_1 \Phi)^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( (X - Z_1 \Phi)^T \right) \right] \right] \times \exp \left[ -\frac{1}{2} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) - \frac{1}{2} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] \times \exp \left[ \frac{1}{2} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) + \frac{1}{2} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] \times \exp \left[ -\frac{1}{2} \left[ \left[ \Delta_{C_{N_1}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right) \right] \right] \times \exp \left[ \frac{1}{2} \left[ \left[ \Delta_{C_{N_1}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right) \right] \right] \left( 1 - r_{(A_{1,MSG})} \left( A^*, C^*, A', C' \right) \right) \right] \times \frac{1}{1 - r_{(A_{1,MSG})} \left( A^*, C^*, A, C \right)} \right] = \left[ H \right]^{-\frac{1}{2}} \left[ H \right]^{\frac{1}{2}} \left[ C \right]^{-\frac{\Delta}{2}} \times \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( (X - Z_1 \Phi)^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( (X - Z_1 \Phi)^T \right) \right] \right] \times \exp \left[ \frac{1}{2} \left[ \text{vec} \left( (X - Z_1 \Phi)^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( (X - Z_1 \Phi)^T \right) \right] \right] \times \exp \left[ -\frac{1}{2} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) - \frac{1}{2} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] \times \exp \left[ \frac{1}{2} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) + \frac{1}{2} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] \times \exp \left[ -\frac{1}{2} \left[ \left[ \Delta_{C_{N_1}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right) \right] \right] \times \exp \left[ \frac{1}{2} \left[ \left[ \Delta_{C_{N_1}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right) \right] \right] \left( 1 - r_{(A_{1,MSG})} \left( A^*, C^*, A', C' \right) \right) \right] \times \frac{1}{1 - r_{(A_{1,MSG})} \left( A^*, C^*, A, C \right)} \right] = \left[ H \right]^{-\frac{1}{2}} \left[ H \right]^{\frac{1}{2}} \left[ C \right]^{-\frac{\Delta}{2}} \times \exp \left[ -\frac{1}{2} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) + \text{tr} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] - \text{tr} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) + \text{tr} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] \times \exp \left[ \frac{1}{2} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) + \frac{1}{2} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] \times \exp \left[ -\frac{1}{2} \left[ \left[ \Delta_{C_{N_1}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right) \right] \right] \times \exp \left[ \frac{1}{2} \left[ \left[ \Delta_{C_{N_1}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right) \right] \right] \left( 1 - r_{(A_{1,MSG})} \left( A^*, C^*, A', C' \right) \right) \right] \times \frac{1}{1 - r_{(A_{1,MSG})} \left( A^*, C^*, A, C \right)} \right] = \left[ H \right]^{-\frac{1}{2}} \left[ H \right]^{\frac{1}{2}} \left[ C \right]^{-\frac{\Delta}{2}} \times \exp \left[ -\frac{1}{2} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) + \text{tr} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] - \text{tr} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) + \text{tr} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] \times \exp \left[ \frac{1}{2} \left( V_{N_1}^{-1} \left( A^* - A' \right)^T U_{N_1}^{-1} \left( A^* - A' \right) \right) + \frac{1}{2} \left( \left[ \Delta_{C_{N_1}}^{-1} C \right]^{-1} C^* \right) \right] \times \exp \left[ -\frac{1}{2} \left[ \left[ \Delta_{C_{N_1}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right) \right] \right] \times \exp \left[ \frac{1}{2} \left[ \left[ \Delta_{C_{N_1}} \Delta_{2C} \right]^{-1} \left( C' + C^* \right) \right] \right] \left( 1 - r_{(A_{1,MSG})} \left( A^*, C^*, A', C' \right) \right) \right] \times \frac{1}{1 - r_{(A_{1,MSG})} \left( A^*, C^*, A, C \right)} \right]
\[
\begin{align*}
&\text{tr} \left( V_{X_j}^{-1} (A^* - A)^T U_{X_j}^{-1} (A^* - A) \right) - \text{tr} \left( \left[ \Delta_{C_{X_j}}^{-1} C \right]^{-1} C^* \right) + \\
&\text{tr} \left( \Delta_{2A} V_{X_j}^{-1} \left( A - \frac{1}{2} (A' + A^*) \right)^T U_{X_j}^{-1} \left( A - \frac{1}{2} (A' + A^*) \right) \right) - \\
&\text{tr} \left( \Delta_{2A} V_{X_j}^{-1} \left( A' - \frac{1}{2} (A + A^*) \right)^T U_{X_j}^{-1} \left( A' - \frac{1}{2} (A + A^*) \right) \right) + \\
&\text{tr} \left( \left[ 2\Delta_{C_{X_j}} \Delta_{2C} \right]^{-1} (C' + C^*)^{-1} C \right) - \\
&\text{tr} \left( \left[ 2\Delta_{C_{X_j}} \Delta_{2C} \right]^{-1} (C + C^*)^{-1} C' \right) \right) \left( \frac{1 - r_{(A,1,MSG)} \left( A^*, C^*, A', C' \right)}{1 - r_{(A,1,MSG)} \left( A^*, C^*, A, C \right)} \right)
\end{align*}
\]

Therefore the acceptance probability for the second stage of the DRMH algorithm is

\[
r_{(A,2,MSG)} \left( A', C', A^*, C^*, A, C \right) =
\min \left( 1, \left| H \right|^{-\frac{1}{2}} \left| H \right|^{\frac{1}{2}} \left| C \right| \left| C' \right|^{-1} \right) \frac{\Delta_{C_{X_j}} \Delta_{2C} \Delta_{X_j}^{-1} \Delta_{2C}^{-1}}{\Delta_{2C}} \times \\
\exp \left[ -\frac{1}{2} \left( \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] \right) - \\
\left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] + \\
\text{tr} \left( V_{X_j}^{-1} (A^* - A')^T U_{X_j}^{-1} (A^* - A') \right) + \text{tr} \left( \left[ \Delta_{C_{X_j}}^{-1} C' \right]^{-1} C^* \right) - \\
\text{tr} \left( V_{X_j}^{-1} (A^* - A)^T U_{X_j}^{-1} (A^* - A) \right) - \text{tr} \left( \left[ \Delta_{C_{X_j}}^{-1} C \right]^{-1} C^* \right) + \\
\text{tr} \left( \Delta_{2A} V_{X_j}^{-1} \left( A - \frac{1}{2} (A' + A^*) \right)^T U_{X_j}^{-1} \left( A - \frac{1}{2} (A' + A^*) \right) \right) - \\
\text{tr} \left( \Delta_{2A} V_{X_j}^{-1} \left( A' - \frac{1}{2} (A + A^*) \right)^T U_{X_j}^{-1} \left( A' - \frac{1}{2} (A + A^*) \right) \right) + \\
\text{tr} \left( \left[ 2\Delta_{C_{X_j}} \Delta_{2C} \right]^{-1} (C' + C^*)^{-1} C \right) - \\
\text{tr} \left( \left[ 2\Delta_{C_{X_j}} \Delta_{2C} \right]^{-1} (C + C^*)^{-1} C' \right) \right) \times \\
\left( \frac{1 - r_{(A,1,MSG)} \left( A^*, C^*, A', C' \right)}{1 - r_{(A,1,MSG)} \left( A^*, C^*, A, C \right)} \right).
\]

The term \( r_{(A,1,MSG)} \left( A^*, C^*, A, C \right) \) in the denominator of the last expression in (4.4.93) will already have been calculated during stage one of the algorithm. This leaves \( r_{(A,1,MSG)} \left( A^*, C^*, A', C' \right) \) to be calculated. This is simply done using the same function as in stage one, but with \( A' \) and \( C' \) used in
place of \( A \) and \( C \). We calculate

\[
\begin{align*}
  r_{(A,1,MSG)}(A^*, C^*, A', C') &= \min \left( 1, \left| H^* \right|^{-\frac{1}{2}} \left| H \right|^{-\frac{1}{2}} \left| C \right| \left| C^* \right|^{-1} \right) \frac{(\Delta^*_{CN_j} - p^{-1})}{\left| C^* \right|^{-\frac{1}{2}} \left| C \right|^{-\frac{1}{2}}} \\
  &= \exp \left( -\frac{1}{2} \left[ \left( \left| X - Z_1 \Phi \right| T \right) \right] T \left( H^* \right)^{-1} \left[ \left( \left| X - Z_1 \Phi \right| T \right) \right] + \right. \right. \\
  &\left. \left. \text{tr} \left( \left( \Delta^*_{CN_j} C^* \right)^{-1} C' - \left( \Delta^*_{CN_j} C \right)^{-1} C^* \right) \right) \right).
\end{align*}
\]

(4.4.94)

Therefore, to calculate the acceptance probability for the second stage of the algorithm, we use the formulas detailed in (4.4.93) and (4.4.94) together with the value already calculated in stage one of the algorithm. It is relatively straightforward to extend the DRMH algorithm above to include more stages. The acceptance probability shown in (2.2.15) shows the general acceptance probability that should be used for later stages of the algorithm.

The algorithm may now be outlined as shown below.

**DR\( _A \) Algorithm**

1. The current state is \( \left( M^{(i)}, \Phi^{(i)}, d^{(i)}, \gamma^{(i)}, c^{(i)}, N^{(i)}, A^{(i-1)}, C^{(i-1)} \right) \).

2. Propose new intercept and coefficient parameter matrices \( C^* \) and \( A^* \) from the proposal densities \( q_{1C}(C^*|C) \) and \( q_{1A}(A^*|A) \) respectively, outlined in (4.4.88).

3. Calculate the acceptance probability \( r_{(A,1,MSG)}(A^*, C^*, A, C) \) per (4.4.91).

4. Simulate \( u_1 \sim \mathcal{U}(0,1) \):

   (a) If \( r_{(A,1,MSG)}(A^*, C^*, A, C) > u_1 \), accept the proposed intercept and coefficient parameter matrices \( C^* \) and \( A^* \). Set \( C^{(i)} = C^* \) and \( A^{(i)} = A^* \).

   (b) If \( r_{(A,1,MSG)}(A^*, C^*, A, C) < u_1 \), then:

      (i) Generate new proposed intercept and coefficient parameter matrices \( C' \) and \( A' \) from \( q_{2C}(C'|C^*, C) \) and \( q_{2A}(A'|A^*, A) \), respectively, using the proposal distributions in (4.4.89).

      (ii) Calculate the acceptance probability \( r_{(A,2,MSG)}(A', C', A^*, C^*, A, C) \) using the formulas detailed in (4.4.93) and (4.4.94), together with the probability calculated at step 3 of this algorithm.

      (iii) Simulate \( u_2 \sim \mathcal{U}(0,1) \):
(A) If $r_{(A,2,MSG)}\left(A', C', A^*, C^*, A, C\right) > u_2$, accept the new proposed intercept and coefficient parameter matrices. Set $C^{(i)} = C'$ and $A^{(i)} = A'$.

(B) If $r_{(A,2,MSG)}\left(A', C', A^*, C^*, A, C\right) < u_2$, reject the new proposed intercept and coefficient parameter matrices. Set $C^{(i)} = C^{(i-1)}$ and $A^{(i)} = A^{(i-1)}$.

4.5 Conclusion

In this chapter, algorithms for simulating from the posterior distributions of the M-STAR, M-GARCH and M-STAR-GARCH models have been developed. All of these algorithms incorporate model selection and parameter estimation into a single task.

The algorithms developed for the multivariate models in this chapter and the algorithms for univariate models developed in chapter 3 are verified using simulation studies in chapter 5. Chapter 5 also provides applications of these algorithms to real world data.
Chapter 5

Empirical Studies

This thesis has developed a number of algorithms for the estimation of the parameters and model orders in the univariate and multivariate settings for STAR, GARCH and STAR-GARCH models. To validate these algorithms, it is crucial to conduct extensive simulation studies. After the successful application of the algorithms to these simulation studies, we present empirical studies for each model. We begin with a simulation study for the investigation of the best method of conditioning the likelihood function in an RJMCMC algorithm within a time series setting.

Each algorithm was coded in R (R Development Core Team 2012) and implemented using parallel computing on The University of Newcastle Research Computer Grid (RCG). This facility allows a significant number of independent applications of an algorithm to run simultaneously in parallel.

For each model, several simulation studies are conducted. In each simulation study, a true model is used to create 1,000 independent simulated data sets. We use three series, that is, $p = 3$, for the multivariate simulation studies. The algorithms are subsequently applied to each data set. Given that the estimates for some parameters vary dramatically with the model index, the results of the algorithms are calculated, based on the number of runs in which the correct model index was identified.

The results shown are the means of the point estimates for each parameter, conditional on the correct model index being identified. The means of the lower and upper limits of the 95% credible intervals (CrI) for each parameter is also calculated as an indication of a typical credible interval. The coverage probability is also calculated for each parameter. The coverage probability is determined by looking at the proportion of times the true value lies within the 95% credible interval.

For simulation studies within this chapter, the number of iterations that each algorithm is run for as well as the number of iterations to burn is determined through pilot runs of each algorithm. These values are overestimated to ensure convergence of the MCMC chains. For both simulation and em-
empirical studies, the MCMC convergence is confirmed using Heidelberger and Welch’s, and Geweke’s convergence diagnostic tests (Cowles and Carlin 1996), in addition to visual inspection of the MCMC time plots. The tuning parameters are initially set within a range found to be typical from experience. The tuning parameters are then tuned automatically up to a predefined iteration of the algorithm. While the acceptance rates are within an acceptable range, the automated tuning methods employed are simple and crude. Overall, each algorithm would benefit from a more sophisticated automated tuning technique.

5.1 Time Series Data Length and the Reversible Jump Algorithm: A Simulation Study

For time series models, the order of the model often impacts upon how much data is available to calculate the likelihood function. In a Reversible Jump algorithm, the amount of data available for the calculation of each likelihood function may be different, depending on the orders of the current and candidate models. The question then arises as to whether we should use the maximum amount of data for each likelihood function, or condition on the same initial values and use the same data in each calculation. This problem is discussed in more detail in Section 2.4.

In order to determine the best method for calculating the likelihood function, a simple simulation study was implemented. The performances of the likelihood estimation methods were then compared. The estimation performance was measured by finding the proportion of runs in which the true model was identified. We shall call our two methods for calculating the likelihood function Methods A and B. Method A uses the maximum amount of data available for each likelihood estimation, whereas Method B uses the same data for both calculations. A total of 40,000 independent data sets were simulated from a true model. Method A was applied to the first 20,000 data sets, and Method B was applied to the remaining 20,000 using the RCG.

For this study, the data was simulated from a true AR(20) model of lengths $N = 150$ and $N = 225$. The true model is shown in (5.1.1), and uses the same coefficient parameter as that used in Troughton and Godsill (1997). The error term $\varepsilon_t$ is such that $\varepsilon_t \sim \mathcal{N}(0, 0.01)$. The true model has the expression:
Section 5.1: Time Series Data Length and Reversible Jump

\[ x_t = -0.5078x_{t-1} + 4.5564x_{t-2} + 1.9504x_{t-3} - 11.2203x_{t-4} - \\
3.5378x_{t-5} + 19.1868x_{t-6} + 3.8193x_{t-7} - 24.8657x_{t-8} - \\
2.4029x_{t-9} + 25.0465x_{t-10} + 0.2687x_{t-11} - 19.7237x_{t-12} + \\
1.1703x_{t-13} + 12.0275x_{t-14} - 1.3091x_{t-15} - 5.5202x_{t-16} + \\
0.6804x_{t-17} + 1.7487x_{t-18} - 0.1543x_{t-19} - 0.2984x_{t-20} + \varepsilon_t. \]  

(5.1.1)

To analyse the posterior proportions, a uniform prior distribution was used. This equates to a Beta(1,1) prior distribution. Due to the large sample size, and the expectation that the number of successes would be greater than zero, the influence of the prior distribution was insignificant. Plots of the posterior distributions are shown in Figure 5.1. The solid line represents method A and the dotted line represents method B.

\[ \begin{array}{ccc}
\text{Density} & \text{Density} & \text{Density} \\
\text{Density} & \text{Density} & \text{Density} \\
\text{Density} & \text{Density} & \text{Density} \\
\end{array} \]

Figure 5.1: Posterior probability distributions for \( N = 150 \) (top left) and \( N = 225 \) (top right), and the posterior probability distribution for the difference in proportions where \( N = 225 \) (bottom).

Evidently, when \( N = 150 \), there is a statistically significant difference between the proportions, with method B performing better. When \( N = 225 \), the proportions are much closer. The posterior distribution for the difference in proportions when \( N = 225 \) is shown in Figure 5.1. This posterior was obtained via a simple Monte Carlo simulation.

As \( N \) becomes larger, the two proportions approach one another and the difference between the two eventually becomes insignificant. Having said this, the empirical evidence above would suggest that
the more effective method for calculating the likelihood function as part of the RJMCMC acceptance probability is method B. This is where the likelihood function is calculated, conditional on the same initial starting values, and the same data is used in the calculation of the acceptance probability formula. Therefore, throughout this thesis, Method B has been implemented for the calculation of acceptance probabilities involving an RJMCMC step.

## 5.2 STAR Algorithm

### 5.2.1 Simulation Study

Two simulation studies were performed on the univariate STAR estimation algorithm in order to demonstrate its ability to identify and estimate the true simulated model. For the two studies S-I and S-II, 10,000 iterations of the MCMC estimation algorithm were executed. The first 1,000 iterations were treated as a burn in period to remove the effects of the chosen initial values and discarded.

The true model for S-I was

$$x_t = 0.02 + 0.55x_{t-1} - 0.68x_{t-2} + (0.40x_{t-1} - 0.85x_{t-2})F_t + \varepsilon_t,$$  \hfill (5.2.2)

where \(\varepsilon_t \sim \mathcal{N}(0, \sigma^2 = 0.75^2)\). The transition function in (5.2.2) was defined as follows:

$$F_t(s_t, d, \gamma, c) = \frac{1}{1 + \exp \left[ -\frac{s_t - 3 - 0.2}{s_t - 3 - 0.2} \right]},$$  \hfill (5.2.3)

The transition variable \(s_t\) in (5.2.3) is an exogenous process that was simulated from a GARCH(1,1) process.

A time plot of a simulated time series from S-I is shown in Figure 5.2 (top), together with a time plot for the transition variable \(s_t\) (bottom). The transition variable has been plotted in black with the transition function \(F_t\) plotted in grey. The transition function has been rescaled so that a value of zero and one in the transition function corresponds to the minimum and maximum values respectively of the transition variable. The blue line represents the value of the location parameter \(c\) for the transition function. From these time plots, it can be seen that the length of each data set in S-I was \(N = 600\).

Figure 5.3 contains an example set of posterior distributions from one of the estimation runs. The red vertical line in each plot indicates the position of the true parameter value. The posterior distribution for the model index \(M_k\) and the delay parameter \(d\) are omitted as the posterior probability was 100% for both of those parameters.
Section 5.2: STAR Algorithm

Figure 5.2: Time plot of simulated STAR data $x_t$ from S-I (top), and time plot of the simulated transition variable over the transition function $F_t$ (grey) with the horizontal line at the true value of the location parameter $c$ (bottom).

Figure 5.3: An example set of parameter posterior distributions from S-I. The red lines indicate the positions of the true values for each parameter.
Table 5.1: Summary statistics for the simulated STAR model of S-I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{1,0}$</td>
<td>0.02</td>
<td>0.0195</td>
<td>(-0.0681, 0.1076)</td>
<td>95.85%</td>
</tr>
<tr>
<td>$\phi_{1,1}$</td>
<td>0.55</td>
<td>0.5492</td>
<td>(0.4522, 0.6420)</td>
<td>94.53%</td>
</tr>
<tr>
<td>$\phi_{1,2}$</td>
<td>-0.68</td>
<td>-0.6785</td>
<td>(-0.7899, -0.5771)</td>
<td>95.24%</td>
</tr>
<tr>
<td>$\phi_{2,0}$</td>
<td>0.00</td>
<td>0.0007</td>
<td>(-0.1507, 0.1519)</td>
<td>95.34%</td>
</tr>
<tr>
<td>$\phi_{2,1}$</td>
<td>0.40</td>
<td>0.3965</td>
<td>(0.2373, 0.5653)</td>
<td>94.43%</td>
</tr>
<tr>
<td>$\phi_{2,2}$</td>
<td>0.85</td>
<td>0.8411</td>
<td>(0.6677, 1.0366)</td>
<td>95.65%</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.5625</td>
<td>0.5644</td>
<td>(0.5032, 0.6339)</td>
<td>96.26%</td>
</tr>
<tr>
<td>$d$</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>99.40%</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>5.00</td>
<td>5.87</td>
<td>(3.543, 9.405)</td>
<td>93.83%</td>
</tr>
<tr>
<td>$c$</td>
<td>0.20</td>
<td>0.1977</td>
<td>(0.0762, 0.3188)</td>
<td>94.84%</td>
</tr>
<tr>
<td>$M_k$</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>99.80%</td>
</tr>
</tbody>
</table>

The results from S-I are shown in Table 5.1. These indicate that the estimation algorithm worked successfully. The means of the point estimates for each of the parameters were very close to the true values. The performance of the credible intervals shows that the true value of each parameter was within the estimated credible interval close to 95% of the time, with the worst performing parameter being the smoothing parameter $\gamma$. The point estimate for this parameter is also slightly too high. It appears as though not much information about this parameter was contained in the data sets created with this particular mix of true parameter values. The choice of prior distribution has slightly increased the mean of our estimates.

One thing to realise is that, despite the results showing that the true model index of 3 was correctly identified 99.80% of the time, this was not the proportion of times that the true model was identified. For certain values of the transition function shape and location parameters, the logistic STAR model may result in a linear AR model.

The estimation of $c$ in eleven of the 1,000 runs was incorrect and resulted in the specification of what is effectively a linear AR model. In two runs, the model index $M_k$ was incorrectly identified. In six runs, the delay parameter $d$ was incorrectly estimated. Only one of these runs was separate from the eleven in which $c$ was incorrectly estimated. Therefore, there were a total of twelve runs in which the wrong model was identified.

For the second simulation study, S-II, the true model was defined as

$$x_t = 0.01 + 1.55x_{t-1} - 0.88x_{t-2} + 0.25x_{t-3} + (0.03 - 1.05x_{t-1} + 0.35x_{t-2} + 0.35x_{t-3})F_t + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \sigma^2 = 0.5^2)$. The transition function in (5.2.4) was defined as follows:

$$F_t(s_t, d, \gamma, c) = \frac{1}{1 + \exp \left[-12\|s_t\| (s_t - 0.15)\right]}.$$  

(5.2.5)
Again, the transition variable \( s_t \) in (5.2.5) was simulated from an exogenous GARCH(1,1) process.

An example time plot of a simulated data set under S-II is shown in Figure 5.4. This figure also presents a time plot of the transition variable \( s_t \) in black, together with a plot of the transition function \( F_t \). The transition function has been rescaled so that a value of zero and one in the transition function corresponds to the minimum and maximum values respectively of the transition variable. Under S-II, the length of the data in each simulation was defined to be \( N = 1,000 \).

The results for S-II outlined in Table 5.2 indicate that the estimation algorithm again worked success-

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_{1,0} )</td>
<td>0.01</td>
<td>0.0093</td>
<td>(-0.0415, 0.0602)</td>
<td>93.87%</td>
</tr>
<tr>
<td>( \phi_{1,1} )</td>
<td>1.55</td>
<td>1.5539</td>
<td>(1.4771, 1.6330)</td>
<td>94.89%</td>
</tr>
<tr>
<td>( \phi_{1,2} )</td>
<td>-0.88</td>
<td>-0.8828</td>
<td>(-0.9697, -0.7966)</td>
<td>95.71%</td>
</tr>
<tr>
<td>( \phi_{1,3} )</td>
<td>0.25</td>
<td>0.2478</td>
<td>(0.1773, 0.3176)</td>
<td>95.19%</td>
</tr>
<tr>
<td>( \phi_{2,0} )</td>
<td>0.03</td>
<td>0.0306</td>
<td>(-0.0397, 0.1010)</td>
<td>95.09%</td>
</tr>
<tr>
<td>( \phi_{2,1} )</td>
<td>-1.05</td>
<td>-1.0610</td>
<td>(-1.1707, -0.9561)</td>
<td>94.79%</td>
</tr>
<tr>
<td>( \phi_{2,2} )</td>
<td>0.35</td>
<td>0.3565</td>
<td>(0.2375, 0.4774)</td>
<td>98.67%</td>
</tr>
<tr>
<td>( \phi_{2,3} )</td>
<td>0.35</td>
<td>0.3510</td>
<td>(0.2544, 0.4491)</td>
<td>97.65%</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.25</td>
<td>0.2505</td>
<td>(0.2292, 0.2745)</td>
<td>94.79%</td>
</tr>
<tr>
<td>( d )</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>98.20%</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>12.00</td>
<td>10.90</td>
<td>(7.60, 15.34)</td>
<td>92.13%</td>
</tr>
<tr>
<td>( c )</td>
<td>-0.15</td>
<td>-0.1504</td>
<td>(-0.1973, -0.1018)</td>
<td>95.40%</td>
</tr>
<tr>
<td>( M_k )</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 5.2: Summary statistics for the simulated STAR model of S-II.
fully. The estimation of $\gamma$ was again the least successful: its mean underestimated the true value. The posterior estimates appear to have been influenced by the prior distribution somewhat. The credible interval coverage probability for this parameter is the worst as its true parameter value only lies within the 95\% interval 92.13\% of the time. The rest of the parameters have coverage probabilities that are relatively close to the notional coverage of the intervals.

As for S-I, the fact that the model index was identified correctly 100\% of the time needs to be evaluated against the success rates for estimating $d$, $\gamma$, and $c$. In seventeen instances, the estimation procedure failed to estimate $c$ within the neighbourhood of the true parameter value. It also failed to identify the true parameter value for the delay parameter $d$ eighteen times. Four of these runs were separate from the runs that failed to estimate $c$ correctly. Therefore, the number of runs that identified the data as a linear AR model as opposed to a LSTAR model was 22. This means that the true model was identified in only 97.80\% of the runs.

Figure 5.5: An example set of parameter posterior distributions from S-II. The red lines indicate the positions of the true values for each parameter.
The plots in Figure 5.5 display an example set of posterior distributions from one of the estimation runs. The posterior distribution for the model index $M_k$ and the delay parameter $d$ are omitted as the posterior probability is 100% for both of those parameters.

The estimation scheme implemented in Livingston Jr and Nur (2016a) was slightly different from the one employed here. Most notably, the prior distribution structure was somewhat different, as was the definition of the transition function. Despite these subtle differences, we were able to confirm the results of Livingston Jr and Nur (2016a) that indicated the insensitivity of the prior distributions for $\gamma$ and $c$, and the recommendation of a large initial value for $\gamma$.

5.2.2 Empirical Data

Following the simulation studies, a case study on the monthly US unemployment rate from 1960 to 2004 for civilian males over 20 years old was implemented. This data set, in one form or another, was also analysed by Dijk, Teräsvirta, and Franses (2002) and Deschamps (2008), among others. The data is displayed in Figure 5.6.

Figure 5.6: Monthly US unemployment rate from 1960 to 2004 time series plots: the original rate $U_t$ (top); the transformed rate $x_t$ (middle); and the seasonally differenced rate $s_t$ (bottom).
The data set contains the original unemployment rate data set $u_t$ that was originally analysed by Dijk, Teräsvirta, and Franses (2002). The transformed data set $x_t$ analysed here was also analysed by Deschamps (2008). Both of these papers used a lagged seasonally differenced transition variable $s_t = u_{t-1} - u_{t-13}$ whose values closely resemble the business cycle, with low values of $s_t$ corresponding to expansions and high values to contractions. We also employed this transition variable.

The algorithm was executed for 10,000 iterations, of which the first 1,000 were discarded as burn in iterations. The parameter estimates are presented in Table 5.3. The most commonly visited model was the one in which $k = 2$ and $d = 1$. The estimates shown in Table 5.3 are based on only the iterations in which this model was the current one. These results are consistent with the results in Livingston Jr and Nur (2016a), but are not directly comparable as the transition function used in Livingston Jr and Nur (2016a) did not take into account the standard deviation of the transition variable. In addition, a different prior distribution structure was employed.

The posterior histogram for $\gamma$ shown in Figure 5.7 highlights the lack of information regarding the value of the smoothing parameter contained within the data. The posterior distribution for $\gamma$ has only slightly moved to the right of the prior distribution. On the other hand, the posterior distribution for $c$ is quite different from the prior distribution.

The fitted values $\hat{x}_t$ using the parameter estimates shown in Table 5.3 and the rescaled fitted transition function $F_t$ are presented in Figure 5.8. A transition function that is closer to one represents periods in which the unemployment rate is growing. Conversely, when the transition function is closer to zero, periods of economic growth and a falling unemployment rate are represented. This is to be expected as the transition function is driven by the transition variable $s_t$ which closely resembles the business cycle.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{1,0}$</td>
<td>-0.0482</td>
<td>(-0.0863, -0.0089)</td>
</tr>
<tr>
<td>$\phi_{1,1}$</td>
<td>0.6243</td>
<td>(0.5145, 0.7363)</td>
</tr>
<tr>
<td>$\phi_{1,2}$</td>
<td>0.3642</td>
<td>(0.2514, 0.4744)</td>
</tr>
<tr>
<td>$\phi_{2,0}$</td>
<td>-0.0523</td>
<td>(-0.1294, 0.0219)</td>
</tr>
<tr>
<td>$\phi_{2,1}$</td>
<td>0.6659</td>
<td>(0.4698, 0.8618)</td>
</tr>
<tr>
<td>$\phi_{2,2}$</td>
<td>-0.6938</td>
<td>(-0.8824, -0.5009)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.00153</td>
<td>(0.00136, 0.00172)</td>
</tr>
<tr>
<td>$d$</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>8.59</td>
<td>(4.9007, 14.1828)</td>
</tr>
<tr>
<td>$c$</td>
<td>0.1937</td>
<td>(0.0557, 0.3207)</td>
</tr>
<tr>
<td>$M_k$</td>
<td>2</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.3: Summary statistics for the US unemployment data.
Section 5.3: GARCH Algorithm

5.3.1 Simulation Study

Two simulation studies were performed for the GARCH estimation algorithm, denoted by G-I and G-II. They were used to examine the effectiveness of the algorithm as outlined in Section 3.2.4. The algorithm searched over five candidate models: the first was a constant conditional variance model, and the other four were GARCH(1,1), GARCH(2,1), GARCH(1,2) and GARCH(2,2) models.

In order to effectively run the Reversible Jump algorithm, reasonable location parameters for the

The form of the fitted model, with the parameters that contain zero in their credible intervals set to zero, is

\[x_t = -0.05 + 0.62x_{t-1} + 0.36x_{t-2} + (0.67x_{t-1} - 0.69x_{t-2})F_t + \varepsilon_t\]

\[\varepsilon_t \sim \mathcal{N}(0, \sigma^2 = 0.0015)\]

\[F_t(s_t, d, \gamma, c) = \frac{1}{1 + \exp\left(-\frac{s_{0.59}}{s_{0.59}}(s_{t-1} - 0.1937)\right)}.

5.3 GARCH Algorithm

Figure 5.7: Posterior histogram for \(\gamma\) (left) and \(c\) (right) showing the density of the respective prior distributions (dashed line) for the US unemployment data.

Figure 5.8: US unemployment rate \(x_t\) with fitted values \(\hat{x}_t\) (red), together with the corresponding rescaled fitted transition function \(\hat{F}_t\) (grey).
proposal distribution are required. The location and covariance parameters were determined from an initial pilot run. For both G-I and G-II, the pilot run was executed for 6,000 iterations. The first 2,000 of these iterations were discarded as burn in. The full algorithm, allowing jumps between candidate models, was run for 15,000 iterations, of which the first 5,000 were discarded as burn in.

For G-I, data were created with \( N = 1,800 \) from the GARCH(1,1) model as shown below:

\[
\begin{align*}
  x_t &= \varepsilon_t \\
  \varepsilon_t &= \sqrt{h_t} \eta_t \\
  h_t &= 0.05 + 0.75 a_{t-1}^2 + 0.10 h_{t-1}.
\end{align*}
\]

An example time plot from one of the data sets in G-I is shown in Figure 5.9. The parameter estimates and 95% credible intervals for G-I are presented in Table 5.4. The posterior mean estimates were close to the true values for each parameter. The coverage probabilities of the parameters were slightly higher than the notional interval values.

![Figure 5.9: Time plot of simulated GARCH data \( x_t \) from G-I.](image)

![Figure 5.10: An example set of parameter posterior distributions from G-I. The red lines indicate the positions of the true values for each parameter.](image)
An example set of posterior distributions from one of the estimation runs is shown in Figure 5.10, together with a time plot of the model index MCMC chain.

The second simulation study, G-II, was performed using the true GARCH(1,2) model shown below:

\[ x_t = \varepsilon_t \]

\[ \varepsilon_t = \sqrt{h_t} \eta_t \]

\[ h_t = 1.5 + 0.56 \varepsilon_{t-1}^2 + 0.15 h_{t-1} + 0.22 h_{t-2}. \]

A time plot from one of the data sets simulated in G-II is shown in Figure 5.11. The compiled results for G-II are shown in Table 5.5.

![Time plot of simulated GARCH data](image)

Figure 5.11: Time plot of simulated GARCH data \( x_t \) from G-II.

The algorithm successfully identified the true model 995 times out of the 1,000 replications. Interestingly, the mean posterior probability for \( N_4 \) was 84.31\%, indicating that there was not extremely strong evidence for that particular model. An example set of posterior distributions from one of the runs is shown in Figure 5.12, along with the MCMC chain for the model index. The MCMC chain for the model index shows how the algorithm jumped between model indices, with the vast majority of

### Table 5.5: Summary statistics for the simulated GARCH model of G-II.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_0 )</td>
<td>0.56</td>
<td>0.5413</td>
<td>(0.4328, 0.6539)</td>
<td>96.08%</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.22</td>
<td>0.2041</td>
<td>(0.0922, 0.3163)</td>
<td>95.98%</td>
</tr>
<tr>
<td>( N_j )</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>99.50%</td>
</tr>
</tbody>
</table>
visits being to $N_4$ and $N_5$.

![Figure 5.12: An example set of parameter posterior distributions from G-II. The red lines indicate the positions of the true values for each parameter.](image)

The means of the point estimates over the 995 series that correctly identified the true model were reasonably close to the true values. The value of $\alpha_0$ was slightly overestimated. A histogram showing the distribution of individual point estimates for $\alpha_0$ from the 1,000 implementations of the algorithm is shown in Figure 5.13. The histogram indicates that there was significant variation in the values of the point estimates. It appears that the coverage probability was again slightly over the notional interval value of 95%.

### 5.3.2 Empirical Data

The estimation procedure for univariate GARCH models was also applied to daily observations of the Deutsch-mark vs British Pound (DEM/GBP) foreign exchange log-returns. This particular data was applied in Ardia (2008), where it was noted that it is “an informal benchmark for GARCH time series software validation”. The data set was also analysed in Bollerslev and Ghysels (1996).

![Figure 5.13: Distribution of point estimates for $\alpha_0$ from G-II.](image)
A time plot of the full data set is shown in Figure 5.14, together with a plot of the conditional variance as calculated using a rolling window of ten data points. It consists of 1,974 data points from 3 January 1985 to 31 December 1991. Very large volatilities during the periods from 1986 to 1987 and 1990 to 1991 are evident from the time plot.

The application in Ardia (2008) looked at the first 750 data points for estimation purposes. With this in mind, the estimation procedure was first applied to the truncated data set, and then to the full data set.

The pilot run was set for 6,000 iterations, while the full algorithm was executed for 50,000 iterations. The estimation results are shown in Table 5.6. The posterior distributions for $\mathcal{N}_j$ and the coefficient parameters $\alpha_0$, $\alpha_1$, and $\beta_1$ are shown in Figure 5.15, together with a time plot of the MCMC chain for the model index $\mathcal{N}_j$. The time plot of the model index indicates that the model moves were often accepted.

The posterior distribution for $\mathcal{N}_j$ indicated fairly strong evidence for a GARCH(1,1) model with a posterior probability of 62.90%. The next most likely model was the GARCH(1,2) model with a

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>0.6555</td>
<td>(0.0293, 0.0922)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.2459</td>
<td>(0.1496, 0.3655)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.5943</td>
<td>(0.4207, 0.7477)</td>
</tr>
<tr>
<td>$\mathcal{N}_j$</td>
<td>2</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.6: Summary statistics for the DEM/GBP exchange log returns for the first $N = 750$ observations.
Figure 5.15: Posterior distributions for $N_j$ and the coefficient parameter vector $\alpha$ for the first $N = 750$ observations.

posterior probability of 23.47%. The point estimates for the coefficient parameter vector were reasonably consistent with the results obtained in Ardia (2008). The estimates in Ardia (2008) were $\hat{\alpha} = [0.048, 0.226, 0.636]^T$.

For the complete data set, where $N = 1,974$, the estimation procedure was also applied. In order to guarantee convergence, the full algorithm was run for a total of 150,000 iterations. The posterior probability for the GARCH(1,2) model was the highest at 77.95%. The closest model to this was the GARCH(1,1) model with a posterior probability of 15.94%. Plots of the MCMC chain for $N_j$ as well as the posterior distributions for $N_j$ and $\alpha$ are shown in Figure 5.16. The time plot of the MCMC chain for the model index again indicates that model moves were often accepted.

Figure 5.16: Posterior distributions for $N_j$ and the coefficient parameter vector $\alpha$ for the whole $N = 1,974$ observations.
Table 5.7: Summary statistics for the DEM/GBP exchange log returns for the whole $N = 1,974$ observations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>CI</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>0.0129</td>
<td>(0.0072, 0.0203)</td>
<td>0.0113</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.1800</td>
<td>(0.1289, 0.2398)</td>
<td>0.1693</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.4690</td>
<td>(0.2450, 0.7089)</td>
<td>0.4840</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.3011</td>
<td>(0.0787, 0.5306)</td>
<td>0.3020</td>
</tr>
<tr>
<td>$\gamma_j$</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The results of the estimation procedure are shown in Table 5.7. The MLE is also included, as calculated using the fGARCH package in R. The point estimates from the Bayesian estimation scheme were reasonably close to the MLE values.

The mathematical form of the fitted model is

\[
x_t = \varepsilon_t \\
\varepsilon_t = \sqrt{h_t} \eta_t \\
h_t = 0.013 + 0.18 \varepsilon_{t-1}^2 + 0.47 h_{t-1} + 0.30 h_{t-2}.
\]

Using the above fitted model, the conditional variance for $N = 1,974$ was calculated for the DEM/GBP data. This plot is presented in Figure 5.17. The shape of the plot mirrors the plot of the calculated conditional variance from a rolling window as shown in Figure 5.14.

![Figure 5.17: Fitted conditional variance for the DEM/GBP exchange log returns using the whole $N = 1,974$ observations.](image)

5.4 STAR-GARCH Algorithm

5.4.1 Simulation Study

In order to verify the MCMC procedure outlined in Section 3.3.4, three simulation studies were conducted in which independent time series were simulated from a particular true model. We then applied the estimation algorithm to these series. The results are presented below. The STAR model or-
orders included in the search were \( K \in \{1, \ldots, 5\} \). The orders considered for GARCH models were \( N \in \{1, \ldots, 5\} \).

For the first study, SG-I, the time series were generated from the STAR(1)-GARCH(2,2) model shown below:

\[
x_t = 0.35x_{t-1} + (-0.03 - 0.85x_{t-1}) F_t (s_t, \gamma, c) + \varepsilon_t
\]

\[
F_t (s_t, \gamma, c) = \frac{1}{1 + \exp \left[ -\frac{15}{s_t^2} (s_t - 0.09) \right]}
\]

\[
\varepsilon_t = \sqrt{h_t} \eta_t
\]

\[
h_t = 0.50 + 0.22a_{t-1}^2 + 0.30a_{t-2}^2 + 0.10h_{t-1} + 0.32h_{t-2}
\]

(5.4.6)

The exogenous variable \( s_t \) in the transition function was randomly generated from a simple GARCH(1,1) process with \( \alpha_s = [0.01, 0.2, 0.5]^T \). We followed the same procedure for all three studies.

An example of a series from SG-I is shown in Figure 5.18. The plot shows periods of higher volatility. The initial pilot for the conditional mean specification was run for 1,000 iterations. A second pilot, for the proposal parameters of the conditional variance equation, was executed for 10,000 iterations on each GARCH model. The full algorithm, allowing model moves, was then run for 5,000 iterations.

Out of the 1,000 replications, the algorithm successfully identified the true STAR model order 993 times and the true GARCH model index 937 times. A GARCH(2,1) model index was identified 62 times and a GARCH(1,1) model index was identified once. The true STAR(1)-GARCH(2,2) model was identified correctly 931 times. On all 931 occasions, the conditional mean was identified as an LSTAR model. In one of the 69 cases in which the algorithm failed to identify the true STAR(1)-GARCH(2,2) model, the conditional mean was instead identified as an almost linear AR(4) model.

It appears from the results in Table 5.8, that the means of the estimates were reasonably close to the true values. The parameters estimated poorly were \( \gamma \) and \( \beta_1 \). The true value for \( \gamma \) was reasonably high when compared to the prior distribution that was used in the estimation procedure. The prior distribution for \( \gamma \) had a mean of 8 and a variance of 10. For this distribution, the density at \( \gamma = 15 \).
was reasonably low, and \(P(\gamma > 15) = 0.0287\). This explains the poor estimate obtained for \(\gamma\). Despite the strong influence of the prior distribution, the true value was still within the 95% credible interval 83.50% of the time.

For the parameter \(\beta_1\), the poor result may be due to the poor estimation of \(\gamma\). The intervals for \(\beta_1\) were fairly wide, indicating that the data did not contain much information pertaining to this parameter. This particular mix of parameters in the true model may have also caused difficulties in estimating \(\beta_1\). The 95% credible interval contained the true value 98.08% of the time. The estimates for the other parameters were fairly close to their true values, and their respective credible intervals contained their true values almost 95% of the time.

An example set of posterior distributions from one of the estimation runs is shown in Figure 5.19. The posterior distributions for the discrete parameters \(M_k\), \(N_j\) and \(d\) are not included as their posterior probabilities were 100%, 98.9% and 100%, respectively.

For the second study, SG-II, data sets of length \(N = 3,000\) were simulated from the following true STAR(3)-GARCH(2,1) model:

\[
x_t = 0.02 + 0.76x_{t-1} - 0.30x_{t-2} + 0.50x_{t-3} + \]
\[
(0.03 - 0.45x_{t-1} + 0.40x_{t-2} - 0.30x_{t-3}) F_t(s_t, \gamma, c) + \epsilon_t
\]

\[
F_t(s_t, \gamma, c) = \frac{1}{1 + \exp\left(-\frac{6}{\pi^{\gamma}}(s_{t-1} + 0.05)\right)}
\]

\[
\epsilon_t = \sqrt{h_t \eta_t}
\]

\[
h_t = 0.90 + 0.12a^2_{t-1} + 0.22a^2_{t-2} + 0.60h_{t-1}
\]

A time plot of a series from SG-II is shown in Figure 5.20.
As with SG-I, the initial pilot for the conditional mean specification was run for 1,000 iterations, whereas the second pilot, which was used to estimate the proposal parameters for the conditional variance equation, was run for 10,000 iterations on each GARCH model. The full algorithm, allowing jumps between model indices, was run for 5,000 iterations.
Out of the 1,000 simulations, the algorithm successfully identified the true STAR model order 998 times. The true GARCH model index was identified 917 times. The other 83 simulations identified a GARCH(2,2) model with parameter estimates for $\alpha_0$, $\alpha_1$, and $\alpha_2$ close to the true values. The true STAR(3)-GARCH(2,1) model was correctly identified 915 times out of the 1,000 simulations. An example set of posterior distributions from one of the estimation runs is shown in Figure 5.21. The posterior probabilities for $M_k$ and $d$ were 100%, and so we have excluded their plots.

The estimation results for SG-II are shown in Table 5.9. The coverage probability was quite close to the notional value of 95%. For $\phi$, the complete conditional mean coefficient vector, the mean coverage
the implicit parameters $(\gamma, c)$ had a mean coverage probability of 94.70%, whereas the complete coefficient parameter for the conditional variance equation $\alpha$ had a mean coverage probability of 94.14%.

Overall, the means of the estimates were quite close to the true values. The worst performing parameters were $\alpha_0$ and $\gamma$. For the smoothing parameter $\gamma$, it appears as though the influence of the prior distribution inflated the mean of the estimates towards its own mean. A possible explanation for the poor performance in estimating $\alpha_0$ could arise from the perturbation of the MCMC chains that occurs when a model move is accepted. This also applies to the estimation of $\beta_1$ in SG-I.

If the location parameters for the variance coefficient parameter proposal distribution are slightly different from the true values, the accepted variance coefficient parameter is likely to lie slightly outside the area of convergence when moves to a candidate model order are accepted. It may take several iterations for the chains to converge towards the location of the true values and away from the neighbourhood of the location parameter.

To overcome this problem, we could have implemented the strategy of discarding iterations that occur directly after a model move is made, incorporating a sort of “mini burn in period” within the full algorithm. Even without implementing this strategy, the parameter estimates were reasonably close to the true values.

The third study, SG-III, was performed on a pure STAR model with a data length of $N = 3,000$. This model had a constant conditional variance. In other words, the true model was effectively a

---

### Table 5.9: Summary statistics for the simulated STAR-GARCH model of SG-II.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{1,0}$</td>
<td>0.02</td>
<td>0.0211 (-0.1274, 0.1692)</td>
<td>96.59%</td>
<td></td>
</tr>
<tr>
<td>$\phi_{1,1}$</td>
<td>0.76</td>
<td>0.7640 (0.6972, 0.8364)</td>
<td>94.39%</td>
<td></td>
</tr>
<tr>
<td>$\phi_{1,2}$</td>
<td>-0.30</td>
<td>-0.3042 (-0.3830, -0.2306)</td>
<td>94.69%</td>
<td></td>
</tr>
<tr>
<td>$\phi_{1,3}$</td>
<td>0.50</td>
<td>0.5020 (0.4395, 0.5682)</td>
<td>95.09%</td>
<td></td>
</tr>
<tr>
<td>$\phi_{2,0}$</td>
<td>0.03</td>
<td>0.0274 (-0.1721, 0.2269)</td>
<td>96.39%</td>
<td></td>
</tr>
<tr>
<td>$\phi_{2,1}$</td>
<td>-0.45</td>
<td>-0.4564 (-0.5515, -0.3693)</td>
<td>95.49%</td>
<td></td>
</tr>
<tr>
<td>$\phi_{2,2}$</td>
<td>0.40</td>
<td>0.4040 (0.3065, 0.5084)</td>
<td>94.99%</td>
<td></td>
</tr>
<tr>
<td>$\phi_{2,3}$</td>
<td>-0.30</td>
<td>-0.3026 (-0.3912, -0.2192)</td>
<td>95.49%</td>
<td></td>
</tr>
<tr>
<td>$d$</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>100.0%</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>6</td>
<td>6.7972 (3.8682, 11.3322)</td>
<td>97.20%</td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td>-0.05</td>
<td>-0.0514 (-0.0742, -0.0296)</td>
<td>92.20%</td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.90</td>
<td>0.9627 (0.7113, 1.2533)</td>
<td>95.31%</td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.12</td>
<td>0.1221 (0.0766, 0.1719)</td>
<td>95.31%</td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.22</td>
<td>0.2253 (0.1578, 0.2940)</td>
<td>92.91%</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.60</td>
<td>0.5883 (0.5252, 0.6473)</td>
<td>93.02%</td>
<td></td>
</tr>
<tr>
<td>$M_k$</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>99.80%</td>
</tr>
<tr>
<td>$N_j$</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>91.70%</td>
</tr>
</tbody>
</table>
STAR(2)-GARCH(0,0) model. This model has the following form:

\[
x_t = -0.02 + 0.70x_{t-1} - 0.55x_{t-2} + \\
(-0.65x_{t-1} + 0.85x_{t-2}) F_t (s_t, \gamma, \zeta) + \epsilon_t
\]

\[
F_t (s_t, \gamma, \zeta) = \frac{1}{1 + \exp \left(-\frac{4}{s(t-2)-0.05}\right)}
\]

\[
\epsilon_t \sim \mathcal{N}(0, \sigma^2 = 2.25).
\]

A sample time plot from SG-III is shown in Figure 5.22. It is evident that this time plot exhibits constant conditional variance when it is compared to the example time plots for SG-I and SG-II in Figures 5.18 and 5.20, respectively.

The results from SG-III are shown in Table 5.10. The algorithm successfully identified the true STAR model order for all 1,000 series. The true GARCH model index was identified 993 times out of the 1,000 simulations. The means of the point estimates were again close to the true values. In addition, the coverage probability was close to the notional interval percentage of 95%. For \( \phi \), the complete conditional mean coefficient vector, the mean coverage probability was 94.58%. For the implicit parameters \((\gamma, \zeta)\), the mean coverage probability was 92.35%.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrtI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_{1,0} )</td>
<td>-0.02</td>
<td>-0.0218</td>
<td>(-0.1027, 0.0591)</td>
<td>95.90%</td>
</tr>
<tr>
<td>( \phi_{1,1} )</td>
<td>0.70</td>
<td>0.6967</td>
<td>(0.6365, 0.7617)</td>
<td>93.90%</td>
</tr>
<tr>
<td>( \phi_{1,2} )</td>
<td>-0.55</td>
<td>-0.5441</td>
<td>(-0.6157, -0.4780)</td>
<td>94.50%</td>
</tr>
<tr>
<td>( \phi_{2,0} )</td>
<td>0.00</td>
<td>0.0022</td>
<td>(-0.1320, 0.1360)</td>
<td>95.20%</td>
</tr>
<tr>
<td>( \phi_{2,1} )</td>
<td>-0.65</td>
<td>-0.6430</td>
<td>(-0.7518, -0.5439)</td>
<td>94.20%</td>
</tr>
<tr>
<td>( \phi_{2,2} )</td>
<td>0.85</td>
<td>0.8364</td>
<td>(0.7278, 0.9578)</td>
<td>93.80%</td>
</tr>
<tr>
<td>( d )</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>100.00%</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>4</td>
<td>4.5482</td>
<td>(3.1097, 6.5928)</td>
<td>91.40%</td>
</tr>
<tr>
<td>( \zeta )</td>
<td>0.05</td>
<td>0.0493</td>
<td>(0.0044, 0.0953)</td>
<td>93.30%</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>2.25</td>
<td>2.2514</td>
<td>(2.1401, 2.3682)</td>
<td>93.27%</td>
</tr>
<tr>
<td>( M_k )</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>100.00%</td>
</tr>
<tr>
<td>( N_j )</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>99.30%</td>
</tr>
</tbody>
</table>

Table 5.10: Summary statistics for the simulated STAR-GARCH model of SG-III.
An example set of posterior distributions from one of the estimation runs is shown in Figure 5.23. The posterior distributions for $M_k$ and $d$ are excluded as their posterior probabilities were 100%.

![Figure 5.23: An example set of parameter posterior distributions from SG-III. The red lines indicate the positions of the true values for each parameter.](image)

### 5.4.2 Empirical Data

The estimation algorithm for the univariate STAR-GARCH model was applied to the Southern Oscillation Index (SOI). This index is an important gauge for measuring the strength of El Niño and La Niña events, and their potential impacts on the Australian region. Its calculation is based on differences in the average monthly surface air pressure between Tahiti and Darwin, and takes into account the long term average and standard deviation of these differences.

We analysed data arising from runs from January 1876 to December 2015, giving a total of $N = 1,680$ data points. This data is available at [http://www.bom.gov.au/climate/current/soihtm1.shtml](http://www.bom.gov.au/climate/current/soihtm1.shtml). A time plot of the data is shown in Figure 5.24. The transition variable was set such that $s_{t-d} = x_{t-d}$.
Figure 5.24: Time plot of the monthly SOI from January 1876 to December 2015 (top) and the conditional variance calculated on a rolling window (bottom).

The variance on a rolling window of five data points was calculated to give an idea of the conditional variance within the data set. This is shown at Figure 5.24. There were several periods in which the relative variance was relatively high, but for the vast majority of data points, the variance was less than 100.

The first pilot run for determining the conditional mean parameters was run for 10,000 iterations, of which the first 2,000 were discarded as burn in. This resulted in a conditional mean model order of

Figure 5.25: Transition function from pilot run on the conditional mean for SOI data.
$k = 1$. The fitted values for $\gamma$ and $c$ led to a STAR model for the conditional mean. The histogram and time plot of its transition function are shown in Figure 5.25.

The second pilot run, for determining the parameters of the proposal distributions for the conditional variance coefficient parameter, was run for 30,000 iterations, of which the first 5,000 were discarded as burn in. The full algorithm was then run for 100,000 iterations, of which the first 5,000 were disregarded as burn in. The final parameter estimates are shown in Table 5.11.

An interesting observation regarding the parameter estimates for $\phi_{2,1}$ and $\phi_{2,2}$ arises from a review of the results in Table 5.11. These estimates are relatively large, and have relatively wide interval estimates. At first this seems rather odd, but reviewing the parameter estimates for $\gamma$ and $c$ reveals the cause.

The posterior distributions for $\gamma$ and $c$ are shown in Figure 5.26, together with the prior densities, and their estimates result in values of the transition function $F_t$ relatively close to zero. The part of the

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{1,0}$</td>
<td>0.0028</td>
<td>(-0.3247, 0.3291)</td>
</tr>
<tr>
<td>$\phi_{1,1}$</td>
<td>0.4967</td>
<td>(0.4484, 0.5460)</td>
</tr>
<tr>
<td>$\phi_{1,2}$</td>
<td>0.2229</td>
<td>(0.1736, 0.2717)</td>
</tr>
<tr>
<td>$\phi_{2,0}$</td>
<td>0.0054</td>
<td>(-0.6130, 0.6118)</td>
</tr>
<tr>
<td>$\phi_{2,1}$</td>
<td>5.54e3</td>
<td>(-1.81e5, 1.92e5)</td>
</tr>
<tr>
<td>$\phi_{2,2}$</td>
<td>3.78e3</td>
<td>(-1.80e5, 1.81e5)</td>
</tr>
<tr>
<td>$d$</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>11.3355</td>
<td>(6.1332, 18.6561)</td>
</tr>
<tr>
<td>$c$</td>
<td>50.3769</td>
<td>(42.6272, 62.9914)</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>23.8886</td>
<td>(8.9421, 43.9503)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.6827</td>
<td>(0.0363, 0.1445)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.0022</td>
<td>(0.0083, 0.5965)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.4372</td>
<td>(0.0438, 0.6878)</td>
</tr>
<tr>
<td>$M_k$</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>$N_j$</td>
<td>4</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.11: Summary statistics for the SOI data.
model that involves the transition function is shown below

$$(\phi_{2,0} + \phi_{2,1}x_{t-1} + \ldots + \phi_{2,k}x_{t-k}) F_t(s_t, \gamma, c). \tag{5.4.7}$$

Using the parameter estimates for $\gamma$, $c$, $d$, and the relevant elements of $\phi$ shown in Table 5.11, we calculated (5.4.7) for all $t$. The maximum value obtained was 0.0151, with a 99% quantile of 1.8036e-08. This indicated that the portion of the model shown in (5.4.7) was close enough to zero to have no impact on the model. That is, the fitted conditional mean was actually an AR(2) model. Quickly fitting an AR(2) model by MLE in R resulted in very similar estimates for the coefficient parameters.

The remaining posterior distributions are shown in Figure 5.27. There appears to be strong evidence for a conditional mean model order of $k = 2$, with a posterior probability of 86.75%. The posterior distribution for the conditional variance model index was not conclusive. The GARCH(1,2) model had the highest posterior probability of 47.05%. The conditional variance model with the second highest

![Figure 5.27: Posterior distributions for the monthly SOI data.](image)
posterior probability was the constant variance model with 33.71%.

Another interesting feature of the posterior distributions shown in Figure 5.27 occurs for $\beta_1$. The posterior distribution is located towards zero. This could indicate that the following conditional variance equation is more appropriate:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-2}.$$ 

While this configuration of the conditional variance equation is not included among the models shown in Table 3.2, the estimation methodology allows such models to be added and included in the search, albeit with some adjustment to the coding of the algorithm.

Therefore, setting the parameters with zero in their credible intervals equal to zero, the fitted model was the following AR(2)-GARCH(1,2) model:

$$x_t = 0.50 x_{t-1} + 0.22 x_{t-2} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} \eta_t$$

$$h_t = 23.89 + 0.08 x_{t-1}^2 + 0.19 h_{t-1} + 0.35 h_{t-2}.$$ 

The fitted values were calculated, employing the fitted model in (5.4.8). Figure 5.28 shows the fitted values for the model in red, together with the rescaled transition function in grey. The rescaled transition function appears to be close to zero for $t$. Also included in Figure 5.28 is a plot of the fitted conditional variance. The periods of high conditional variance are mirrored by those shown in

![Figure 5.28: Time plot of the monthly SOI data $x_t$ with fitted values $\hat{x}_t$ (red), together with the corresponding rescaled fitted transition function $\hat{F}_t$ (grey), (top); and fitted conditional variance $\hat{h}_t$ (bottom).](image)
the approximation to the conditional variance in Figure 5.24. Conditional on the diagnostic testing, it appears as though the fitted AR(2)-GARCH(1,2) is a reasonable model for the monthly SOI data.

5.5 Multivariate STAR Algorithm

5.5.1 Simulation Study

Two simulation studies were performed and denoted by MS-I and MS-II. For the first study, the length of each series in the data sets was $N = 2,500$. The true model for MS-I was

$$x_t = \begin{bmatrix} 0.60 & 0.05 & -0.15 \\ 0.02 & -0.20 & 0.02 \\ 0.03 & -0.15 & 0.30 \end{bmatrix} x_{t-1} + \begin{bmatrix} 0.30 & 0.10 & 0.10 \\ 0.05 & 0.20 & 0.10 \\ 0.05 & 0.10 & 0.00 \end{bmatrix} x_{t-2} + \begin{bmatrix} -0.50 & 0.34 & -0.03 \\ -0.20 & 0.14 & -0.38 \\ 0.06 & 0.24 & 0.10 \end{bmatrix} x_{t-1} + \begin{bmatrix} -0.13 & -0.12 & 0.03 \\ -0.11 & 0.21 & 0.06 \\ -0.02 & 0.16 & 0.14 \end{bmatrix} x_{t-2} F_t + \varepsilon_t,$$

where

$$\varepsilon_t \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.050 & 0.003 & -0.004 \\ 0.003 & 0.006 & 0.005 \\ -0.004 & 0.005 & 0.030 \end{bmatrix} \right).$$

The true transition function $F_t$ for this model is

$$F_t = F_t (s_t, \gamma, c) = \frac{1}{1 + \exp \left( -6 \frac{6}{\gamma \sigma^2} (s_{t-2} + 0.04) \right)}.$$

(5.5.9)

The exogenous transition variable $s_t$ in equation (5.5.9) is simply a set of independently and identically distributed random variables from a Normal distribution. That is, $s_t \sim \mathcal{N}(0,1)$. An example of a simulated data set is shown in Figure 5.29. For both MS-I and MS-II, the algorithm was run for 30,000 iterations of which the first 10,000 were used as a burn in period.

A set of posterior distributions from a sample estimation run is shown in Figures 5.30, 5.31, and 5.32. The posterior distributions for $M_k$ and $d$ are excluded as their posterior probabilities were 100% at the true values.

The results in Table 5.12 suggest that the algorithm worked well in terms of point estimation. The true conditional mean model order was identified 933 times out of the 1,000 replications. The posterior
Figure 5.29: Time plot of simulated M-STAR data $x_t$ from MS-I.

Figure 5.30: An example set of parameter posterior distributions from MS-I. The red lines indicate the positions of the true values for each parameter.
Section 5.5: Multivariate STAR Algorithm

Figure 5.31: An example set of parameter posterior distributions from MS-I (continued). The red lines indicate the positions of the true values for each parameter.
mean estimates of all parameters were very close to their true values. The discrete delay parameter \( d \) was estimated well, with 984 simulations runs identifying the correct model parameter. The only times this parameter was incorrectly identified were when the conditional mean model order was incorrectly identified. A review of the coverage probabilities suggests that the credible intervals were estimated reasonably well: the true parameter value was within the 95% credible interval close to 95% of the time.

The second simulation study, MS-II, used data of length \( N = 1,800 \). The true model is outlined as follows:

\[
x_t = \begin{bmatrix} 0.17 & 0.00 & 0.08 \\ -0.07 & -0.18 & 0.39 \\ 0.24 & -0.19 & 0.04 \end{bmatrix} x_{t-1} + \begin{bmatrix} 0.33 & -0.18 & -0.13 \\ 0.16 & 0.01 & -0.26 \\ 0.18 & -0.16 & 0.03 \end{bmatrix} x_{t-2} + \begin{bmatrix} 0.03 & -0.08 & -0.09 \\ -0.10 & -0.10 & -0.19 \\ 0.08 & -0.09 & 0.04 \end{bmatrix} x_{t-3} + \begin{bmatrix} 0.21 & -0.05 & -0.11 \\ -0.34 & -0.17 & 0.13 \\ -0.04 & 0.10 & -0.16 \end{bmatrix} x_{t-4} + \begin{bmatrix} 0.01 & 0.40 & 0.28 \\ -0.15 & -0.01 & -0.18 \\ 0.13 & -0.40 & -0.06 \end{bmatrix} F_t + \varepsilon_t
\]

\( \varepsilon_t \sim \mathcal{N} \left( 0, \begin{bmatrix} 0.50 & 0.02 & -0.08 \\ 0.02 & 0.6 & 0.05 \\ -0.08 & 0.05 & 0.03 \end{bmatrix} \right) \)

\( F_t = F_t(s_t, \gamma, c) = \frac{1}{1 + \exp \left( -\frac{1}{s_t} (s_{t-1} + 0.12) \right)} \).

The exogenous transition variable \( s_t \) in the transition function shown in (5.5.10) was simulated from the following GARCH(1,1) process:

\[
s_t = \sqrt{h_t} \eta_t, \quad h_t = 0.10 + 0.25 s_{t-1}^2 + 0.60 h_{t-1}, \quad \eta_t \sim \mathcal{N}(0, 1).
\]
### Table 5.12: Summary statistics for the simulated M-STAR model of MS-I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Phi_{1,1} ) [1, 1]</td>
<td>0.60</td>
<td>0.5997</td>
<td>(0.5386, 0.6616)</td>
<td>95.07%</td>
</tr>
<tr>
<td>( \Phi_{1,1} ) [2, 1]</td>
<td>0.02</td>
<td>0.0205</td>
<td>(-0.0009, 0.0421)</td>
<td>94.21%</td>
</tr>
<tr>
<td>( \Phi_{1,1} ) [3, 1]</td>
<td>0.03</td>
<td>0.0296</td>
<td>(-0.0166, 0.0758)</td>
<td>94.64%</td>
</tr>
<tr>
<td>( \Phi_{1,1} ) [1, 2]</td>
<td>0.05</td>
<td>0.0489</td>
<td>(-0.1052, 0.2024)</td>
<td>96.03%</td>
</tr>
<tr>
<td>( \Phi_{1,1} ) [2, 2]</td>
<td>-0.20</td>
<td>-0.2010</td>
<td>(-0.2544, -0.1478)</td>
<td>95.93%</td>
</tr>
<tr>
<td>( \Phi_{1,1} ) [3, 2]</td>
<td>-0.30</td>
<td>-0.3008</td>
<td>(-0.4198, -0.1821)</td>
<td>96.25%</td>
</tr>
<tr>
<td>( \Phi_{1,1} ) [1, 3]</td>
<td>-0.15</td>
<td>-0.1494</td>
<td>(-0.2330, -0.0657)</td>
<td>93.89%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [2, 3]</td>
<td>0.02</td>
<td>0.0211</td>
<td>(-0.0096, 0.0524)</td>
<td>94.11%</td>
</tr>
<tr>
<td>( \Phi_{1,3} ) [1, 2]</td>
<td>0.30</td>
<td>0.2981</td>
<td>(0.2333, 0.3628)</td>
<td>94.53%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [1, 1]</td>
<td>0.30</td>
<td>0.2982</td>
<td>(0.2359, 0.3606)</td>
<td>93.78%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [2, 1]</td>
<td>0.05</td>
<td>0.0499</td>
<td>(0.0281, 0.0717)</td>
<td>95.18%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [3, 1]</td>
<td>0.05</td>
<td>0.0508</td>
<td>(0.0026, 0.0989)</td>
<td>95.18%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [1, 2]</td>
<td>0.10</td>
<td>0.1002</td>
<td>(-0.0349, 0.2354)</td>
<td>95.71%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [2, 2]</td>
<td>0.20</td>
<td>0.1996</td>
<td>(0.1522, 0.2466)</td>
<td>95.30%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [3, 2]</td>
<td>0.10</td>
<td>0.0989</td>
<td>(-0.0058, 0.2035)</td>
<td>94.21%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [1, 3]</td>
<td>0.10</td>
<td>0.1004</td>
<td>(0.0159, 0.1849)</td>
<td>95.28%</td>
</tr>
<tr>
<td>( \Phi_{1,2} ) [2, 3]</td>
<td>0.10</td>
<td>0.1000</td>
<td>(0.0706, 0.1292)</td>
<td>94.86%</td>
</tr>
<tr>
<td>( \Phi_{1,3} ) [3, 3]</td>
<td>0.00</td>
<td>-0.0001</td>
<td>(-0.0657, 0.0653)</td>
<td>94.43%</td>
</tr>
<tr>
<td>( \Phi_{2,1} ) [1, 1]</td>
<td>-0.50</td>
<td>-0.5010</td>
<td>(-0.5932, -0.4103)</td>
<td>94.00%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [2, 1]</td>
<td>-0.20</td>
<td>-0.2008</td>
<td>(-0.2330, -0.1691)</td>
<td>95.39%</td>
</tr>
<tr>
<td>( \Phi_{2,1} ) [3, 1]</td>
<td>0.06</td>
<td>0.0612</td>
<td>(-0.0077, 0.1302)</td>
<td>94.64%</td>
</tr>
<tr>
<td>( \Phi_{2,1} ) [1, 2]</td>
<td>0.34</td>
<td>0.3415</td>
<td>(0.1125, 0.5715)</td>
<td>95.82%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [2, 2]</td>
<td>0.14</td>
<td>0.1409</td>
<td>(0.0616, 0.2205)</td>
<td>95.50%</td>
</tr>
<tr>
<td>( \Phi_{2,1} ) [3, 2]</td>
<td>0.24</td>
<td>0.2399</td>
<td>(0.0628, 0.4177)</td>
<td>94.53%</td>
</tr>
<tr>
<td>( \Phi_{2,1} ) [1, 3]</td>
<td>-0.03</td>
<td>-0.0299</td>
<td>(-0.1549, 0.0949)</td>
<td>95.07%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [2, 3]</td>
<td>-0.38</td>
<td>-0.3812</td>
<td>(-0.4278, -0.3358)</td>
<td>93.68%</td>
</tr>
<tr>
<td>( \Phi_{2,3} ) [1, 3]</td>
<td>0.10</td>
<td>0.1032</td>
<td>(0.0067, 0.2000)</td>
<td>94.32%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [1, 2]</td>
<td>-0.13</td>
<td>-0.1307</td>
<td>(-0.2239, -0.0378)</td>
<td>94.86%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [2, 1]</td>
<td>-0.11</td>
<td>-0.1104</td>
<td>(-0.1430, -0.0780)</td>
<td>95.18%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [3, 1]</td>
<td>-0.02</td>
<td>-0.0208</td>
<td>(-0.0927, 0.0511)</td>
<td>94.96%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [1, 2]</td>
<td>-0.12</td>
<td>-0.1201</td>
<td>(-0.3219, 0.0815)</td>
<td>95.39%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [2, 2]</td>
<td>0.21</td>
<td>0.2105</td>
<td>(0.1404, 0.2811)</td>
<td>94.21%</td>
</tr>
<tr>
<td>( \Phi_{2,3} ) [2, 2]</td>
<td>0.16</td>
<td>0.1622</td>
<td>(0.0063, 0.3185)</td>
<td>95.82%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [1, 3]</td>
<td>0.03</td>
<td>0.0280</td>
<td>(-0.0982, 0.1543)</td>
<td>93.57%</td>
</tr>
<tr>
<td>( \Phi_{2,2} ) [2, 3]</td>
<td>0.06</td>
<td>0.0593</td>
<td>(0.0156, 0.1031)</td>
<td>94.21%</td>
</tr>
<tr>
<td>( \Phi_{2,3} ) [2, 3]</td>
<td>0.14</td>
<td>0.1391</td>
<td>(0.0415, 0.2371)</td>
<td>94.43%</td>
</tr>
<tr>
<td>( \Sigma_{1,1} )</td>
<td>0.050</td>
<td>0.0503</td>
<td>(0.0474, 0.0530)</td>
<td>96.14%</td>
</tr>
<tr>
<td>( \Sigma_{2,1} )</td>
<td>0.003</td>
<td>0.0031</td>
<td>(0.0023, 0.0037)</td>
<td>95.61%</td>
</tr>
<tr>
<td>( \Sigma_{3,1} )</td>
<td>-0.004</td>
<td>-0.0040</td>
<td>(-0.0055, -0.0025)</td>
<td>94.86%</td>
</tr>
<tr>
<td>( \Sigma_{2,2} )</td>
<td>0.006</td>
<td>0.0061</td>
<td>(0.0057, 0.0064)</td>
<td>94.64%</td>
</tr>
<tr>
<td>( \Sigma_{3,2} )</td>
<td>0.005</td>
<td>0.0050</td>
<td>(0.0044, 0.0056)</td>
<td>94.96%</td>
</tr>
<tr>
<td>( \Sigma_{3,3} )</td>
<td>0.030</td>
<td>0.0300</td>
<td>(0.0284, 0.0317)</td>
<td>94.53%</td>
</tr>
<tr>
<td>( d )</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>98.40%</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>6</td>
<td>6.0682</td>
<td>(4.6736, 7.8550)</td>
<td>95.61%</td>
</tr>
<tr>
<td>( c )</td>
<td>-0.04</td>
<td>-0.0390</td>
<td>(-0.0893, 0.0112)</td>
<td>94.00%</td>
</tr>
<tr>
<td>( M_k )</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>93.30%</td>
</tr>
</tbody>
</table>
An example time plot from one of the data sets used in MS-II is shown in Figure 5.33. When compared to the time plots in Figure 5.29, there appears to be a more noticeable relationship between $x_{1,t}$ and $x_{3,t}$. Both series seem to follow each other in short periods of expansion, followed by short periods of contraction.

The results of the estimation are shown in Tables 5.13 and 5.14. The posterior mean point estimates indicate that the algorithm worked well. The credible intervals appear to contain the true parameter value close to 95% of the time. The only exception is the smoothing parameter of the transition function, which has a coverage probability of 88.1%. An example set of posterior distributions from one of the estimation runs is shown in Figures 5.34, 5.35 and 5.36.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi_{1,1} [1,1]$</td>
<td>0.17</td>
<td>0.1702</td>
<td>(0.0848, 0.2555)</td>
<td>95.50%</td>
</tr>
<tr>
<td>$\Phi_{1,1} [2,1]$</td>
<td>-0.07</td>
<td>-0.0690</td>
<td>(-0.1626, 0.0246)</td>
<td>95.20%</td>
</tr>
<tr>
<td>$\Phi_{1,1} [3,1]$</td>
<td>0.24</td>
<td>0.2400</td>
<td>(0.2191, 0.2609)</td>
<td>95.20%</td>
</tr>
<tr>
<td>$\Phi_{1,1} [1,2]$</td>
<td>0.00</td>
<td>0.0010</td>
<td>(-0.0659, 0.0679)</td>
<td>96.20%</td>
</tr>
<tr>
<td>$\Phi_{1,1} [2,2]$</td>
<td>-0.18</td>
<td>-0.1783</td>
<td>(-0.2516, -0.0150)</td>
<td>94.20%</td>
</tr>
<tr>
<td>$\Phi_{1,1} [3,2]$</td>
<td>-0.19</td>
<td>-0.1903</td>
<td>(-0.2067, -0.1739)</td>
<td>96.00%</td>
</tr>
<tr>
<td>$\Phi_{1,1} [1,3]$</td>
<td>0.08</td>
<td>0.0836</td>
<td>(-0.1239, 0.2911)</td>
<td>95.60%</td>
</tr>
<tr>
<td>$\Phi_{1,1} [2,3]$</td>
<td>0.39</td>
<td>0.3898</td>
<td>(0.1622, 0.6172)</td>
<td>92.60%</td>
</tr>
</tbody>
</table>

**Table 5.13:** Summary statistics for the simulated M-STAR model of MS-II.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi_{1,1}$</td>
<td>0.04</td>
<td>0.0393</td>
<td>(-0.0116, 0.0902)</td>
<td>94.60%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>0.33</td>
<td>0.3273</td>
<td>(0.2277, 0.4268)</td>
<td>96.50%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>0.16</td>
<td>0.1588</td>
<td>(0.0497, 0.2679)</td>
<td>94.20%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>0.18</td>
<td>0.1805</td>
<td>(0.1560, 0.2049)</td>
<td>95.20%</td>
</tr>
<tr>
<td>$\Phi_{1,1}$</td>
<td>-0.18</td>
<td>-0.1810</td>
<td>(-0.2513, -0.1107)</td>
<td>95.40%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>0.01</td>
<td>0.0090</td>
<td>(-0.0678, 0.0859)</td>
<td>94.00%</td>
</tr>
<tr>
<td>$\Phi_{1,3}$</td>
<td>-0.16</td>
<td>-0.1598</td>
<td>(-0.1774, -0.1422)</td>
<td>95.40%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>-0.13</td>
<td>-0.1342</td>
<td>(-0.2613, -0.0071)</td>
<td>95.60%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>-0.26</td>
<td>-0.2589</td>
<td>(-0.3981, -0.1195)</td>
<td>96.10%</td>
</tr>
<tr>
<td>$\Phi_{1,3}$</td>
<td>0.03</td>
<td>0.0309</td>
<td>(-0.0002, 0.0621)</td>
<td>95.50%</td>
</tr>
<tr>
<td>$\Phi_{1,1}$</td>
<td>0.03</td>
<td>0.0308</td>
<td>(-0.0823, 0.1439)</td>
<td>94.50%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>-0.10</td>
<td>-0.1018</td>
<td>(-0.2259, 0.0222)</td>
<td>93.90%</td>
</tr>
<tr>
<td>$\Phi_{1,3}$</td>
<td>0.08</td>
<td>0.0799</td>
<td>(0.0520, 0.1078)</td>
<td>96.70%</td>
</tr>
<tr>
<td>$\Phi_{1,1}$</td>
<td>-0.08</td>
<td>-0.0802</td>
<td>(-0.1723, 0.0121)</td>
<td>95.00%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>-10</td>
<td>-0.1010</td>
<td>(-0.2020, 0.0001)</td>
<td>94.00%</td>
</tr>
<tr>
<td>$\Phi_{1,1}$</td>
<td>-0.09</td>
<td>-0.0904</td>
<td>(-0.1130, -0.0767)</td>
<td>94.80%</td>
</tr>
<tr>
<td>$\Phi_{1,3}$</td>
<td>-0.09</td>
<td>-0.0874</td>
<td>(-0.1911, 0.0163)</td>
<td>95.00%</td>
</tr>
<tr>
<td>$\Phi_{1,3}$</td>
<td>-0.19</td>
<td>-0.1852</td>
<td>(-0.2989, -0.0716)</td>
<td>94.60%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>0.04</td>
<td>0.0398</td>
<td>(0.0143, 0.0652)</td>
<td>95.80%</td>
</tr>
<tr>
<td>$\Phi_{1,1}$</td>
<td>0.21</td>
<td>0.2111</td>
<td>(0.0934, 0.3291)</td>
<td>94.90%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>-0.34</td>
<td>-0.3395</td>
<td>(-0.4688, -0.2104)</td>
<td>95.10%</td>
</tr>
<tr>
<td>$\Phi_{1,1}$</td>
<td>-0.04</td>
<td>-0.0398</td>
<td>(-0.0687, -0.0110)</td>
<td>95.40%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>-0.05</td>
<td>-0.0516</td>
<td>(-0.1441, 0.0409)</td>
<td>94.70%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>-0.17</td>
<td>-0.1711</td>
<td>(-0.2724, -0.0698)</td>
<td>95.40%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>0.10</td>
<td>0.1006</td>
<td>(0.0780, 0.1234)</td>
<td>95.90%</td>
</tr>
<tr>
<td>$\Phi_{1,1}$</td>
<td>-0.11</td>
<td>-0.1117</td>
<td>(-0.3983, 0.1749)</td>
<td>94.90%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>0.13</td>
<td>0.1325</td>
<td>(-0.1816, 0.4468)</td>
<td>94.70%</td>
</tr>
<tr>
<td>$\Phi_{1,3}$</td>
<td>-0.16</td>
<td>-0.1597</td>
<td>(-0.2300, -0.0894)</td>
<td>95.70%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>0.01</td>
<td>0.0103</td>
<td>(-0.1273, 0.1478)</td>
<td>96.70%</td>
</tr>
<tr>
<td>$\Phi_{2,2}$</td>
<td>-0.19</td>
<td>-0.1494</td>
<td>(-0.3002, 0.0014)</td>
<td>94.40%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>0.13</td>
<td>0.1299</td>
<td>(0.0962, 0.1637)</td>
<td>94.20%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>0.40</td>
<td>0.4037</td>
<td>(0.3066, 0.5007)</td>
<td>95.20%</td>
</tr>
<tr>
<td>$\Phi_{2,2}$</td>
<td>-0.01</td>
<td>-0.0089</td>
<td>(-0.1152, 0.0973)</td>
<td>94.20%</td>
</tr>
<tr>
<td>$\Phi_{2,3}$</td>
<td>-0.40</td>
<td>-0.4006</td>
<td>(-0.4249, -0.3764)</td>
<td>95.60%</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>0.28</td>
<td>0.2812</td>
<td>(0.1057, 0.4568)</td>
<td>94.40%</td>
</tr>
<tr>
<td>$\Phi_{2,2}$</td>
<td>-0.18</td>
<td>-0.1751</td>
<td>(-0.3676, 0.0174)</td>
<td>96.20%</td>
</tr>
<tr>
<td>$\Phi_{2,3}$</td>
<td>-0.06</td>
<td>-0.0605</td>
<td>(-0.1036, -0.0175)</td>
<td>94.90%</td>
</tr>
<tr>
<td>$\Phi_{3,1}$</td>
<td>0.02</td>
<td>0.0191</td>
<td>(-0.1370, 0.1753)</td>
<td>94.20%</td>
</tr>
<tr>
<td>$\Phi_{3,1}$</td>
<td>-0.09</td>
<td>-0.0917</td>
<td>(-0.2629, 0.0795)</td>
<td>95.00%</td>
</tr>
<tr>
<td>$\Phi_{3,2}$</td>
<td>-0.32</td>
<td>-0.3201</td>
<td>(-0.3586, -0.2817)</td>
<td>95.40%</td>
</tr>
<tr>
<td>$\Phi_{3,2}$</td>
<td>-0.14</td>
<td>-0.1396</td>
<td>(-0.2669, -0.0123)</td>
<td>94.70%</td>
</tr>
<tr>
<td>$\Phi_{3,1}$</td>
<td>-0.02</td>
<td>-0.0172</td>
<td>(-0.1568, 0.1224)</td>
<td>95.50%</td>
</tr>
<tr>
<td>$\Phi_{3,2}$</td>
<td>0.18</td>
<td>0.1806</td>
<td>(0.1494, 0.2119)</td>
<td>94.60%</td>
</tr>
<tr>
<td>$\Phi_{2,3}$</td>
<td>-0.30</td>
<td>-0.2812</td>
<td>(-0.4476, -0.1610)</td>
<td>95.00%</td>
</tr>
<tr>
<td>$\Phi_{3,2}$</td>
<td>0.23</td>
<td>0.2220</td>
<td>(0.0650, 0.3790)</td>
<td>95.00%</td>
</tr>
<tr>
<td>$\Sigma [1,1]$</td>
<td>0.20</td>
<td>0.1999</td>
<td>(0.1648, 0.2351)</td>
<td>95.00%</td>
</tr>
<tr>
<td>$\Sigma [2,1]$</td>
<td>0.50</td>
<td>0.5006</td>
<td>(0.4688, 0.5347)</td>
<td>94.60%</td>
</tr>
<tr>
<td>$\Sigma [3,1]$</td>
<td>0.02</td>
<td>0.0197</td>
<td>(-0.0058, 0.0543)</td>
<td>95.30%</td>
</tr>
<tr>
<td>$\Sigma [2,2]$</td>
<td>-0.08</td>
<td>-0.0801</td>
<td>(-0.0871, -0.0735)</td>
<td>95.00%</td>
</tr>
<tr>
<td>$\Sigma [3,2]$</td>
<td>0.60</td>
<td>0.6014</td>
<td>(0.5631, 0.6422)</td>
<td>95.00%</td>
</tr>
<tr>
<td>$\Sigma [3,3]$</td>
<td>0.05</td>
<td>0.0501</td>
<td>(0.0436, 0.0569)</td>
<td>94.50%</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>11</td>
<td>10.8020</td>
<td>(9.5325, 12.2513)</td>
<td>88.10%</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-0.12</td>
<td>-0.1203</td>
<td>(-0.1312, -0.1093)</td>
<td>95.90%</td>
</tr>
</tbody>
</table>

Table 5.14: Summary statistics for the simulated M-STAR model of MS-II (continued).
The effectiveness of the estimation procedure was confirmed by the above simulation studies. Next, the algorithm was applied to real data. The results are detailed in the next section.

**Figure 5.34:** An example set of parameter posterior distributions from MS-II. The red lines indicate the positions of the true values for each parameter.
Figure 5.35: An example set of parameter posterior distributions from MS-II (continued). The red lines indicate the positions of the true values for each parameter.
Figure 5.36: An example set of parameter posterior distributions from MS-II (continued). The red lines indicate the positions of the true values for each parameter.

Figure 5.37: Time plots for the Jökulsá river flow (top); Vatnsdalsá river flow (middle); and the precipitation (bottom).
5.5.2 Empirical Data

The estimation procedure was applied to the Icelandic river flow data for the years 1972-1974. The data measures river flows in cubic meters per second for two rivers, the Jökulsá and the Vatnsdalsá. It is also accompanied by time series of the temperatures and precipitation over the same period. The data was compiled by the Hydrological Survey of the National Energy Authority of Iceland.

The data was first analysed by Tong, Thanoon, and Gudmundsson (1985) using a univariate threshold model. The two rivers were analysed jointly by Tsay (1998), who applied a bivariate threshold model. Another multivariate application for this data may be found in Teräsvirta and Yang (2014b), where a multivariate STAR model was applied to the data.

The Jökulsá and Vatnsdalsá river flows are denoted by \( x_{1,t} \) and \( x_{2,t} \), respectively. The transition variable \( s_t \) is set to the precipitation over the same period. Plots of these time series are shown in Figure 5.37.

The models fitted in Tong, Thanoon, and Gudmundsson (1985) and Teräsvirta and Yang (2014b) were considerably different from the one applied here. For example, the model in Tong, Thanoon, and Gudmundsson (1985) was a threshold model which included the exogenous variables of precipitation and temperature in the mean equation. The model fitted in Teräsvirta and Yang (2014b) was a three regime M-STAR model with the exogenous variables of precipitation and temperature included as the transition variable in each of the transition functions used. In addition, the latter model used a non-standardised transition function and incorporated a seasonal component into the mean equation.

Both models fitted in Tong, Thanoon, and Gudmundsson (1985) and Teräsvirta and Yang (2014b) used model orders of up to 15. With this in mind, the maximum model order was set to \( k_{\text{max}} = 20 \) to allow this possibility. The maximum value for the delay parameter was set to \( d_{\text{max}} = 20 \). After running a quick pilot run of the algorithm, it was noticed that the posterior for the smoothing parameter \( \gamma \) was slightly bimodal with peaks in the distribution around the prior mean and in an area of low posterior probability to the right of the prior distribution. It was decided that the range of the prior distribution should be increased in order to increase the ability of the MCMC chain to explore larger values of \( \gamma \).

The prior parameters for \( \gamma \) were set so that \( \alpha_{\gamma} = \frac{25^2}{\text{dof}} \) and \( \beta_{\gamma} = \frac{25}{\text{dof}} \).

The algorithm was run for 100,000 iterations. Generally, the best guide for checking convergence is to review the plots for \( \gamma \) and \( c \). It is usual that, when these parameters have converged, the same is true of the other parameters of the model. The plots for these parameters, together with the plot for \( d \) from the first iteration to the 15,000th iteration are shown at Figure 5.38.

It appears that the MCMC chains shown in Figure 5.38 converged well before the 5,000th iteration.
Figure 5.38: Time plots of the MCMC chains for $\gamma$ (top); $c$ (middle); and $d$ (bottom).

Therefore, as the MCMC chain was sufficiently large, the first 10,000 iterations were discarded as burn in for estimation purposes.

The posterior distributions for the smoothing and location parameters are shown in Figure 5.39. The posterior distribution for $\gamma$ sits on the right hand tail of the prior distribution. The posterior distribution for $c$ is located in an area of low prior probability in the right hand tail. The other posterior distributions are shown in Figure 5.40. The posterior for the delay parameter $d$ is not shown, as the posterior probability for that parameter was 100% at $d = 4$.

Figure 5.39: Posterior distributions for the parameters $\gamma$ and $c$ for the river flow data. The prior densities are shown as a dashed line.
Figure 5.40: Posterior distributions for the M-STAR model fitted to the Icelandic river flow data.
The posterior distribution for the conditional mean model order is shown in Figure 5.40. The model with highest posterior probability was the one in which \( k = 3 \) and \( d = 4 \), with a posterior probability of 81.16%. Virtually identical results were obtained when the algorithm was executed with differing initial values, including \( M_k^{(0)} = 15 \). The algorithm converged to the same stationary distribution each time.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Phi_{11} ) [1, 1]</td>
<td>1.1871 (1.1147, 1.2591)</td>
<td>( \Phi_{21} ) [1, 1]</td>
<td>-0.1213 (-0.2635, 0.0211)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{12} ) [2, 1]</td>
<td>-0.0057 (-0.0277, 0.0164)</td>
<td>( \Phi_{22} ) [2, 1]</td>
<td>0.0715 (0.0280, 0.1154)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{11} ) [1, 2]</td>
<td>-0.9500 (0.7520, 1.1467)</td>
<td>( \Phi_{21} ) [1, 2]</td>
<td>-1.3590 (-2.0349, -0.6811)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{12} ) [2, 2]</td>
<td>1.3360 (1.2752, 1.3963)</td>
<td>( \Phi_{22} ) [2, 2]</td>
<td>-0.6096 (-0.8180, -0.4009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{11} ) [1, 1]</td>
<td>-0.3810 (-0.4820, -0.2809)</td>
<td>( \Phi_{22} ) [1, 1]</td>
<td>0.1688 (-0.1664, 0.5038)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{12} ) [2, 1]</td>
<td>0.0126 (-0.0182, 0.0437)</td>
<td>( \Phi_{22} ) [2, 1]</td>
<td>-0.1658 (-0.2699, -0.0628)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{11} ) [1, 2]</td>
<td>-0.9336 (-1.2341, -0.6324)</td>
<td>( \Phi_{22} ) [1, 2]</td>
<td>0.5953 (-0.3295, 1.5264)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{12} ) [2, 2]</td>
<td>-0.5393 (-0.6318, -0.4462)</td>
<td>( \Phi_{22} ) [2, 2]</td>
<td>0.1218 (-0.1625, 0.4045)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{11} ) [1, 1]</td>
<td>0.1479 (0.0862, 0.2098)</td>
<td>( \Phi_{22} ) [1, 2]</td>
<td>-0.0301 (-0.2551, 0.1951)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{12} ) [2, 1]</td>
<td>-0.0052 (-0.0244, 0.0140)</td>
<td>( \Phi_{22} ) [2, 2]</td>
<td>0.1045 (0.0348, 0.1742)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{11} ) [1, 2]</td>
<td>0.1829 (-0.0290, 0.3964)</td>
<td>( \Phi_{22} ) [1, 2]</td>
<td>0.5611 (0.0325, 1.0855)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Phi_{12} ) [2, 2]</td>
<td>0.1870 (0.1217, 0.2526)</td>
<td>( \Phi_{22} ) [2, 2]</td>
<td>0.3340 (0.1704, 0.4965)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d )</td>
<td>4</td>
<td>-</td>
<td>( \Sigma ) [1, 1]</td>
<td>41.4695 (38.0815, 45.1361)</td>
<td></td>
</tr>
<tr>
<td>( \gamma )</td>
<td>47.7036 (32.4171, 64.3173)</td>
<td>( \Sigma ) [2, 2]</td>
<td>3.6382 (2.8658, 4.4527)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c )</td>
<td>10.3946 (10.2065, 10.5029)</td>
<td>( \Sigma ) [2, 2]</td>
<td>3.9394 (3.6187, 4.2901)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M_k )</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.15: Parameter estimates for the M-STAR model fitted to the Icelandic river flow data.

The parameter estimates are shown in Table 5.15. After setting the parameters that contain zero in their 95% credible intervals to zero, the form of the fitted model was

\[
x_t = \begin{bmatrix} 1.19 & -0.95 \\ 0.00 & 1.34 \end{bmatrix} x_{t-1} + \begin{bmatrix} -0.38 & -0.93 \\ 0.00 & -0.54 \end{bmatrix} x_{t-2} + \begin{bmatrix} 0.15 & 0.00 \\ 0.00 & 0.19 \end{bmatrix} x_{t-3} + \begin{bmatrix} 0.00 & -1.36 \\ 0.07 & -0.61 \end{bmatrix} \varepsilon_{t-1} + \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \varepsilon_{t-2} + \begin{bmatrix} 0.00 & 0.56 \\ 0.10 & 0.33 \end{bmatrix} \varepsilon_{t-3} \]

\[F_t = \varepsilon_t \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 41.46 & 3.64 \\ 3.64 & 3.94 \end{bmatrix}\right)\]

\[F_t(s_t, \gamma, c) = \left[1 + \exp\left(- \frac{41.46}{S(s_t)} (s_t - 10.39)\right)\right]^{-1}.

Time plots of the observed and fitted values are shown in black and red, respectively in Figure 5.41. The fitted M-STAR model seems to be a reasonably good model for the data, even though there are a few over and under estimates. The plots of the transition function in Figure 5.42 show that it usually takes on the values zero or one, and suggest an M-TAR model. The plot of the transition function against the transition variable shows that the curve almost has a step shape. This is to be expected, given the point estimate for \( \gamma \).
5.6 Multivariate GARCH Algorithm

5.6.1 VECH Model

A multivariate GARCH model can be made to take on many forms through the reformulation of its covariance matrix equation. We initially investigated the VECH formulation and designed an estimation procedure that employed a restriction on the parameters for estimation and enforced the positive definiteness of the covariance matrices through the prior distribution. These techniques emulated those developed in Hudson and Gerlach (2008).

As mentioned in Section 4.3.1, the procedure created failed to effectively estimate the parameters for the model so often that it could not be considered a success. We provide an example of such a failure below, where we present some results from testing the MCMC procedure. Data was created from an
Figure 5.43: Time plot of simulated data $\mathbf{x}_t$ from the bivariate VECH model in (5.6.11).

The model described in (5.6.11) is a bivariate model, that is, $p = 2$. The length of the data is $N = 1,000$. A time plot of a data set simulated from the model in (5.6.11) is displayed in Figure 5.43.

The algorithm was run for a total of 50,000 iterations, the first 10,000 of which were discarded as burn in. A typical successful application of the algorithm resulted in a set of posterior distributions as shown in Figure 5.44. The red vertical lines on each of the histograms are located at the true values of their corresponding parameters. Each of the MCMC chains is located in the neighbourhood of the true value.

However, as previously noted, the successful result shown in Figure 5.44 was not achieved consistently enough to call the estimation algorithm a success. A typical unsuccessful application of the algorithm is shown in Figure 5.45. The time plots of each of the MCMC chains ignore the first 10,000 iterations, as these are discarded as burn in. From the start of the plots, the MCMC chains are reasonably close to the true values of each of their respective parameters. However, around the 30,000th iteration (20,000th on the plots), several of the chains move away from the location of the true parameter value.
The reason why the procedure was only partially successful by contrast to the implementation in Hudson and Gerlach (2008) is related to key differences between the two schemes. First, the Hudson and Gerlach (2008) scheme included a non-constant conditional mean, whereas our implementation was for a pure M-GARCH model. Under the univariate setting, the inclusion of a non-constant conditional mean, such as in the STAR-GARCH model, appeared to make it easier to estimate the conditional variance parameters than in the pure univariate GARCH model. Therefore, the inclusion of a non-constant conditional mean in the multivariate setting may also ease estimation of the conditional covariance parameters.

Second, in Hudson and Gerlach (2008), the authors chose to partition the parameter space, based on the parameters that appeared in each of the \( \frac{p(p+1)}{2} \) equations for the covariance matrix. Ignoring the estimation of the conditional mean, their procedure contained \( \frac{p(p+1)}{2} \) steps within a Gibbs sampler algorithm. This means, during one iteration of the MCMC algorithm, the conditional covariance matrices would need to be inverted at least \( \frac{p(p+1)}{2} \) times. This requires considerable computational
resources, even for small \( p \), especially as \( N \) becomes large. With this in mind, it was decided that our estimation procedure would propose the intercept matrix and coefficient matrices separately, but would test them together in one Metropolis-Hastings step. This drastically reduced the number of times the covariance matrices needed to be inverted per iteration, leading to a faster implementation of the algorithm.

Third, an important difference between the estimation schemes is related to the proposal distributions used. In the estimation procedure here, all parameters were proposed, conditional on the most recently simulated value, using a random walk Metropolis-Hastings step. Hudson and Gerlach (2008) chose to use a random walk Metropolis-Hastings algorithm only for the diagonal elements of the coefficient matrices \( A \) and \( B \). During the burn in period, the off diagonal entries of those matrices had proposal means equal to zero and relatively small variances. This meant that the results for those parameters were restricted to the area around zero. The algorithm was then unable to explore areas of high posterior probability that were not located near zero.

---

**Figure 5.45:** A typical set of MCMC chains for the VECH algorithm showing a failure of the chains to converge.
The restriction on the proposal distribution for the off diagonal elements was not initially used in our implementation. Interestingly, the problems that occurred in an application of the algorithm were caused by these parameters. In an attempt to remedy this, the off diagonal elements were prescribed proposal means located at zero during the burn in period. Unfortunately, this resulted in the algorithm operating very slowly as the proposed values were too often rejected due to the lack of positive definiteness of at least one conditional covariance matrix.

The key differences and difficulties previously discussed here could have been addressed, but the required changes would have compromised some of the goals of the estimation procedure. For example, it is not unrealistic that an M-GARCH model without a conditional mean may need to be fitted. Furthermore, choosing a proposal distribution for the burn in period that does not allow full exploration of the parameter space is essentially equivalent to making an assumption about the location of that parameter. As we were trying to keep the estimation procedure as general as possible, such an assumption was not deemed desirable.

In addition, breaking up the Gibbs Sampler into several Metropolis-Hastings steps will result in more inversions of the conditional covariance matrices in each iteration of the algorithm, leading to a compromise on the speed of the estimation algorithm. This is compounded when moving to data sets where \( p \geq 3 \). This is not only due to the additional time required to invert larger conditional covariance matrices, but also arises from an increase in the number of rejections caused by the non-positive definiteness of those matrices.

It was decided that, due to the shortcomings of the estimation procedure, ultimately related to the speed of the algorithm, the BEKK formulation would instead be investigated. The BEKK formulation guarantees the positive definiteness of the conditional covariance matrices, and the simulation study for the algorithm described in Section 4.3.6 is presented in the following section.

### 5.6.2 BEKK Model - Simulation Study

In contrast to the univariate GARCH estimation procedure, which searched over models of varying model orders \( l \) and \( m \) through the model index \( \mathcal{N}_j \), the multivariate GARCH procedure searches over three types of models with fixed orders \( l \) and \( m \). The three models are the fixed conditional covariance model, the DBEKK and the full BEKK formulations of the conditional covariance equation in an M-GARCH model. These are indexed using \( \mathcal{N}_j \), as outlined in Table 5.16.

<table>
<thead>
<tr>
<th>Index</th>
<th>( \mathcal{N}_1 )</th>
<th>( \mathcal{N}_2 )</th>
<th>( \mathcal{N}_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Constant: ( l = 0,m = 0 )</td>
<td>DBEKK: ( l = 1,m = 1 )</td>
<td>BEKK: ( l = 1,m = 1 )</td>
</tr>
</tbody>
</table>

**Table 5.16**: Index of M-GARCH models.
The first study, MG-I, with \( N = 1,500 \), used the following true DBEKK model:

\[
x_t = \varepsilon_t \\
\varepsilon_t = H_t^{1/2} \eta_t \\
H_t = \begin{bmatrix}
0.20 & -0.05 & 0.25 \\
-0.05 & 0.30 & 0.00 \\
0.25 & 0.00 & 0.60 \\
0.55 & 0.00 & 0.00 \\
0.00 & 0.65 & 0.00 \\
0.00 & 0.00 & 0.45 \\
\end{bmatrix} + \begin{bmatrix}
0.70 & 0.00 & 0.00 \\
0.00 & 0.50 & 0.00 \\
0.00 & 0.00 & 0.75 \\
0.70 & 0.00 & 0.00 \\
0.00 & 0.50 & 0.00 \\
0.00 & 0.00 & 0.75 \\
\end{bmatrix} T_t \varepsilon_{t-1} + \varepsilon_t^{T_t} H_{t-1} \varepsilon_{t-1} T_t.
\]

An example of simulated data from MG-I is shown in Figure 5.46. The pilot runs for each of the three model indices were executed for differing numbers of iterations. The simpler models, being the constant conditional covariance and the DBEKK model, generally converged to the area of high posterior probability fairly quickly, when compared with the full BEKK model. With this in mind, the lengths of the pilot runs were set to 1,000, 4,000 and 12,000 iterations for the constant, DBEKK and BEKK models, respectively.

Figure 5.46: Time plot of simulated M-GARCH data \( x_t \) from MG-I.
Table 5.17: Summary statistics for the simulated M-GARCH model of MG-I.

The main algorithm that allows jumps between models was run for 50,000 iterations, of which the first 10,000 iterations were discarded as a burn in. The results shown in Table 5.17 indicate that the true model was correctly identified 977 times out of the 1,000 replications. In the other 23 runs, a full BEKK model was identified. The means of the parameter estimates from the runs that incorrectly identified the BEKK model were similar to those for the true DBEKK model.

The results shown in Table 5.17 indicate that the algorithm was successful. This was because the posterior mean estimates were very close to the true values. In addition, the coverage probability was
close to the notional interval value of 95%. A set of posterior distributions from one of the estimation runs is shown in Figure 5.47. The posterior distribution for the model order is omitted as the posterior probability was 100% at the true value for this application of the estimation scheme.

For the second study, MG-II, the data length was set to \( N = 1,500 \). The example of simulated data shown in Figure 5.48 was created from the following true BEKK model:

\[
x_t = \varepsilon_t \quad \varepsilon_t = H_t^{1/2} \eta_t,
\]

with conditional covariance equation

\[
H_t = \left[ \begin{array}{ccc} 0.56 & 0.19 & 0.08 \\ 0.19 & 0.47 & 0.21 \\ 0.08 & 0.21 & 0.42 \end{array} \right] + \left[ \begin{array}{ccc} 0.68 & 0.18 & 0.40 \\ 0.35 & 0.50 & 0.00 \\ -0.25 & 0.12 & 0.35 \end{array} \right] \varepsilon_{t-1} \varepsilon_{t-1}^T + \left[ \begin{array}{ccc} 0.55 & 0.20 & -0.20 \\ -0.15 & 0.60 & 0.30 \\ 0.25 & -0.40 & 0.65 \end{array} \right] \eta_{t-1}^T.
\]
Section 5.6: Multivariate GARCH Algorithm

Comparing the plots in Figures 5.46 and 5.48, it is evident that the full BEKK formulation of the conditional covariance equation allows for shocks to spill over to other series. Areas of high volatility are shared by each of the plots.

The pilot runs for each of the three types of models were the same as for MG-I. The main algorithm, however, was run for 150,000 iterations, of which the first 10,000 were discarded as a burn in. A typical set of posterior distributions from an individual run of the algorithm is shown in Figure 5.49.

The results shown in Table 5.18 indicate that the means of the parameter point estimates were very close to their true values at two decimal places. The coverage probabilities indicate that the credible intervals underperformed: most coverage probabilities were less than 95%. There were a few parameters for which the coverage probability was less than 90%. However, the mean coverage probability over all parameters except for the model index was 92%, which is not substantially different from the notional value of 95%.

The application of the M-GARCH estimation scheme to real world data is discussed in the following section.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A[1,1]</td>
<td>0.68</td>
<td>0.6820</td>
<td>(0.6306, 0.7350)</td>
<td>95.0%</td>
</tr>
<tr>
<td>A[2,1]</td>
<td>0.35</td>
<td>0.3541</td>
<td>(0.3169, 0.3924)</td>
<td>94.5%</td>
</tr>
<tr>
<td>A[3,1]</td>
<td>-0.25</td>
<td>-0.2510</td>
<td>(-0.2857, -0.2170)</td>
<td>92.6%</td>
</tr>
<tr>
<td>A[1,2]</td>
<td>0.18</td>
<td>0.1843</td>
<td>(0.1257, 0.2433)</td>
<td>93.4%</td>
</tr>
<tr>
<td>A[2,2]</td>
<td>0.50</td>
<td>0.4964</td>
<td>(0.4463, 0.5480)</td>
<td>95.1%</td>
</tr>
<tr>
<td>A[3,2]</td>
<td>0.12</td>
<td>0.1192</td>
<td>(0.0750, 0.1639)</td>
<td>93.3%</td>
</tr>
<tr>
<td>A[1,3]</td>
<td>0.40</td>
<td>0.4008</td>
<td>(0.3415, 0.4611)</td>
<td>92.2%</td>
</tr>
<tr>
<td>A[2,3]</td>
<td>0.00</td>
<td>0.0019</td>
<td>(-0.0454, 0.0492)</td>
<td>93.8%</td>
</tr>
<tr>
<td>A[3,3]</td>
<td>0.35</td>
<td>0.3474</td>
<td>(0.2993, 0.3963)</td>
<td>94.0%</td>
</tr>
<tr>
<td>B[1,1]</td>
<td>0.55</td>
<td>0.5431</td>
<td>(0.4971, 0.5872)</td>
<td>91.0%</td>
</tr>
<tr>
<td>B[2,1]</td>
<td>-0.15</td>
<td>-0.1511</td>
<td>(-0.1901, -0.1117)</td>
<td>92.3%</td>
</tr>
<tr>
<td>B[3,1]</td>
<td>0.25</td>
<td>0.2445</td>
<td>(0.2061, 0.2836)</td>
<td>90.6%</td>
</tr>
<tr>
<td>B[1,2]</td>
<td>0.20</td>
<td>0.1968</td>
<td>(0.1175, 0.2771)</td>
<td>91.0%</td>
</tr>
<tr>
<td>B[2,2]</td>
<td>0.60</td>
<td>0.6000</td>
<td>(0.5413, 0.6562)</td>
<td>91.0%</td>
</tr>
<tr>
<td>B[3,2]</td>
<td>-0.40</td>
<td>-0.3978</td>
<td>(-0.4614, -0.3347)</td>
<td>88.9%</td>
</tr>
<tr>
<td>B[1,3]</td>
<td>-0.20</td>
<td>-0.1949</td>
<td>(-0.2641, -0.1280)</td>
<td>93.3%</td>
</tr>
<tr>
<td>B[2,3]</td>
<td>0.30</td>
<td>0.2909</td>
<td>(0.2287, 0.3555)</td>
<td>91.2%</td>
</tr>
<tr>
<td>B[3,3]</td>
<td>0.65</td>
<td>0.6394</td>
<td>(0.5749, 0.7025)</td>
<td>89.7%</td>
</tr>
<tr>
<td>C[1,1]</td>
<td>0.56</td>
<td>0.5704</td>
<td>(0.4079, 0.7468)</td>
<td>94.5%</td>
</tr>
<tr>
<td>C[2,1]</td>
<td>0.19</td>
<td>0.1906</td>
<td>(0.0565, 0.3320)</td>
<td>91.0%</td>
</tr>
<tr>
<td>C[3,1]</td>
<td>0.08</td>
<td>0.0933</td>
<td>(-0.0353, 0.2212)</td>
<td>91.2%</td>
</tr>
<tr>
<td>C[2,2]</td>
<td>0.47</td>
<td>0.4690</td>
<td>(0.2997, 0.6469)</td>
<td>91.1%</td>
</tr>
<tr>
<td>C[2,3]</td>
<td>0.21</td>
<td>0.2321</td>
<td>(0.0914, 0.3618)</td>
<td>87.9%</td>
</tr>
<tr>
<td>C[3,3]</td>
<td>0.42</td>
<td>0.4202</td>
<td>(0.2660, 0.5829)</td>
<td>89.5%</td>
</tr>
</tbody>
</table>

$N_j$ | 3 | 3 | - | 100.00% |

Table 5.18: Summary statistics for the simulated M-GARCH model of MG-II.
Figure 5.49: An example set of parameter posterior distributions from MG-II. The red lines indicate the positions of the true values for each parameter.
Section 5.6: Multivariate GARCH Algorithm

5.6.3 Empirical Data

An illustration of the M-GARCH estimation algorithm was implemented using bivariate data derived from the daily share prices for BHP Billiton Limited and Rio Tinto Limited on the ASX over the period 20 January 2010 to 11 April 2016, inclusive. The daily share prices are plotted in Figure 5.50. From the plots it is evident that the two series are highly correlated. Both plots follow similar paths with increases and decreases occurring at the same times. This is somewhat expected, given the similarities between the two companies: both are large multinational organisations in the mining sector, traded on the Australian Stock Exchange.

Figure 5.50: Daily share prices for BHP Limited and Rio Tinto Limited.

Figure 5.51: Time plots of the log returns $x_{1,t}$ and $x_{2,t}$, for BHP Limited and Rio Tinto Limited respectively.
Rather than analysing this series directly, the data items were transformed into their log returns as follows:

\[ x_{1,t} = 100 \log (\text{BHP}_t - \text{BHP}_{t-1}) \]
\[ x_{2,t} = 100 \log (\text{RIO}_t - \text{RIO}_{t-1}) . \]

A time plot of the data that was analysed is shown in Figure 5.51. Each of the individual series appeared to exhibit non-constant conditional variance at around the same time periods.

The sample variances of a rolling window of length ten were calculated for both series and are depicted in Figure 5.52. There are periods where both plots exhibit increases in the sample variances of the rolling window at the same time. A multivariate GARCH model should be well suited for modelling this bivariate data.

The pilot runs of the algorithm were executed for 20,000 iterations for each of the three models under consideration. The first 10,000 iterations were used as burn in. The full algorithm was then executed for 300,000 iterations, again using the first 10,000 iterations as a burn in period.

The posterior evidence supporting the full BEKK formulation of the conditional covariance equation was extremely strong. The algorithm never departed from that model type after the first few iterations. Plots of the posterior distributions for each of the parameters are shown in Figure 5.53 and the estimation results are displayed in Table 5.19.

The fitted conditional covariance and correlations over the time period are shown in Figure 5.54. The correlation plot shows that, over time, the correlation varied between a minimum of 0.66 and a maximum of 0.95. The middle 95% ranged from 0.72 to 0.93, with a mean correlation of 0.83.

Figure 5.52: Sample variances of a rolling window of length 10, for BHP Limited and Rio Tinto Limited.
Section 5.6: Multivariate GARCH Algorithm

Setting the insignificant parameters to zero, the mathematical form of the fitted model becomes

\[
x_t = \varepsilon_t
\]

\[
\varepsilon_t = H_t^{\frac{1}{2}} \eta_t
\]

\[
H_t = \begin{bmatrix}
0.03 & 0.04 \\
0.04 & 0.10
\end{bmatrix} + \begin{bmatrix}
0.19 & 0.07 \\
0.09 & 0.22
\end{bmatrix} \varepsilon_{t-1} \varepsilon_{t-1}^T \begin{bmatrix}
0.19 & 0.07 \\
0.09 & 0.22
\end{bmatrix} + \begin{bmatrix}
0.99 & -0.03 \\
0.00 & 0.93
\end{bmatrix} H_{t-1} \begin{bmatrix}
0.99 & -0.03 \\
0.00 & 0.93
\end{bmatrix}.
\]

Figure 5.53: Posterior distributions for the M-GARCH model fitted to the BHP Limited and Rio Tinto Limited data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A[1,1])</td>
<td>0.1860</td>
<td>(0.1223, 0.2480)</td>
<td>(B[1,1])</td>
<td>0.9928</td>
<td>(0.9692, 1.0092)</td>
</tr>
<tr>
<td>(A[2,1])</td>
<td>0.0883</td>
<td>(0.0037, 0.1680)</td>
<td>(B[2,1])</td>
<td>0.0100</td>
<td>(-0.0155, 0.0343)</td>
</tr>
<tr>
<td>(A[1,2])</td>
<td>0.0653</td>
<td>(0.0020, 0.1293)</td>
<td>(B[1,2])</td>
<td>-0.0331</td>
<td>(-0.0579, -0.0026)</td>
</tr>
<tr>
<td>(A[2,2])</td>
<td>0.2217</td>
<td>(0.1450, 0.3036)</td>
<td>(B[2,2])</td>
<td>0.9281</td>
<td>(0.8914, 0.9587)</td>
</tr>
<tr>
<td>(C[1,1])</td>
<td>0.0283</td>
<td>(0.0109, 0.0532)</td>
<td>(C[2,1])</td>
<td>0.0522</td>
<td>(0.0224, 0.0907)</td>
</tr>
<tr>
<td>(C[2,2])</td>
<td>0.1050</td>
<td>(0.0513, 0.1745)</td>
<td>(N_j)</td>
<td>3</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.19: Point estimates and credible intervals for the BHP Limited and Rio Tinto Limited data.
5.7 Multivariate STAR-GARCH Algorithm

5.7.1 Simulation Study

For the M-STAR-GARCH estimation scheme, two simulation studies were performed. The first study, MSG-I, was performed using the following M-STAR-GARCH model with a DBEKK conditional covariance equation:

\[ x_t = \mu_t + \varepsilon_t \]
\[ \varepsilon_t = H_t^\frac{1}{2} \eta_t, \]

where

\[
\mu_t = \begin{bmatrix}
0.10 & 0.05 & -0.15 \\
0.02 & -0.20 & 0.02 \\
0.03 & -0.16 & 0.30 \\
0.40 & -0.1 & -0.30 \\
0.50 & -0.1 & -0.20 \\
0.12 & -0.4 & 0.15 \\
-0.13 & -0.12 & 0.03 \\
-0.11 & 0.21 & 0.06 \\
-0.02 & 0.16 & 0.14 \\
\end{bmatrix}
\]

\[
F_t = F_t(s_t, \gamma, c) = \frac{1}{1 + \exp \left[ -\frac{6.5}{S(s_t)} (s_t - 0.75) \right]},
\]

Figure 5.54: Estimated conditional covariances (top and bottom left) and correlation (bottom right) over the time period, for the BHP Limited and Rio Tinto Limited data.
The conditional covariance equation $H_t$ in (5.7.12) is defined as follows:

$$H_t = \begin{bmatrix}
0.56 & 0.15 & 0.22 \\
0.15 & 0.30 & 0.07 \\
0.22 & 0.07 & 0.51 \\
\end{bmatrix} + \begin{bmatrix}
0.68 & 0.00 & 0.00 \\
0.00 & 0.46 & 0.00 \\
0.00 & 0.00 & 0.53 \\
\end{bmatrix} \varepsilon_t \varepsilon_t^T - \begin{bmatrix}
0.68 & 0.00 & 0.00 \\
0.00 & 0.46 & 0.00 \\
0.00 & 0.00 & 0.53 \\
\end{bmatrix} + \begin{bmatrix}
0.31 & 0.00 & 0.0 \\
0.00 & 0.62 & 0.0 \\
0.00 & 0.00 & 0.7 \\
\end{bmatrix}^{T} H_{t-1} \begin{bmatrix}
0.31 & 0.00 & 0.0 \\
0.00 & 0.62 & 0.0 \\
0.00 & 0.00 & 0.7 \\
\end{bmatrix}^T.$$

The length of the data for MSG-I was set to $N = 2,500$. An example of the time plots from MSG-I is given in Figure 5.55. The plots indicate that the conditional means of the three time series are related, with some form of dependence between the series. Any dependence between the series relating to the conditional covariance is less evident. This is to be expected when looking at data simulated from an M-STAR-GARCH process with a DBEKK conditional covariance equation.

The pilots for the three types of conditional covariance models were run with differing numbers of iterations as in Section 5.6. For both MSG-I and MSG-II, the constant conditional covariance, DBEKK and full BEKK model pilot runs were executed for 1,000, 2,000 and 3,000 iterations, respectively. The
main algorithms allowing jumps between models were run for 100,000 iterations, with the first 10,000 iterations being discarded as burn in. An example set of posterior distributions from one application of the algorithm in MSG-I is shown in Figures 5.56, 5.57, and 5.58.

Figure 5.56: An example set of parameter posterior distributions from MSG-I. The red lines indicate the positions of the true values for each parameter.
Figure 5.57: An example set of parameter posterior distributions from MSG-I (continued). The red lines indicate the positions of the true values for each parameter.
Figure 5.58: An example set of parameter posterior distributions from MSG-I (continued). The red lines indicate the positions of the true values for each parameter.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi_{1,1}$</td>
<td>$0 \cdot 10$</td>
<td>$0 \cdot 0992$</td>
<td>(0.0602, 0.1383)</td>
<td>95.07%</td>
</tr>
<tr>
<td>$\Phi_{1,2}$</td>
<td>$0 \cdot 02$</td>
<td>$0 \cdot 0194$</td>
<td>(-0.0099, 0.0487)</td>
<td>94.68%</td>
</tr>
<tr>
<td>$\Phi_{1,3}$</td>
<td>$0 \cdot 03$</td>
<td>$0 \cdot 0300$</td>
<td>(-0.0162, 0.0761)</td>
<td>95.09%</td>
</tr>
<tr>
<td>$\Phi_{1,4}$</td>
<td>$0 \cdot 05$</td>
<td>$0 \cdot 0503$</td>
<td>(0.0124, 0.0882)</td>
<td>94.88%</td>
</tr>
<tr>
<td>$\Phi_{1,5}$</td>
<td>$-0 \cdot 20$</td>
<td>$-0 \cdot 1988$</td>
<td>(-0.2345, -0.1632)</td>
<td>94.58%</td>
</tr>
<tr>
<td>$\Phi_{1,6}$</td>
<td>$-0 \cdot 15$</td>
<td>$-0 \cdot 1498$</td>
<td>(-0.1733, -0.1264)</td>
<td>93.48%</td>
</tr>
<tr>
<td>$\Phi_{1,7}$</td>
<td>$0 \cdot 02$</td>
<td>$0 \cdot 0205$</td>
<td>(-0.0008, 0.0418)</td>
<td>95.39%</td>
</tr>
<tr>
<td>$\Phi_{1,8}$</td>
<td>$0 \cdot 30$</td>
<td>$0 \cdot 2991$</td>
<td>(0.2598, 0.3383)</td>
<td>94.38%</td>
</tr>
<tr>
<td>$\Phi_{1,9}$</td>
<td>$0 \cdot 08$</td>
<td>$0 \cdot 0798$</td>
<td>(0.0441, 0.1154)</td>
<td>94.88%</td>
</tr>
<tr>
<td>$\Phi_{1,10}$</td>
<td>$0 \cdot 05$</td>
<td>$0 \cdot 0512$</td>
<td>(0.0218, 0.0807)</td>
<td>95.19%</td>
</tr>
<tr>
<td>$\Phi_{1,11}$</td>
<td>$0 \cdot 05$</td>
<td>$0 \cdot 0507$</td>
<td>(0.0041, 0.0974)</td>
<td>95.29%</td>
</tr>
<tr>
<td>$\Phi_{1,12}$</td>
<td>$0 \cdot 10$</td>
<td>$0 \cdot 1005$</td>
<td>(0.0638, 0.1373)</td>
<td>95.49%</td>
</tr>
<tr>
<td>$\Phi_{1,13}$</td>
<td>$0 \cdot 20$</td>
<td>$0 \cdot 1988$</td>
<td>(0.1650, 0.2326)</td>
<td>94.28%</td>
</tr>
<tr>
<td>$\Phi_{1,14}$</td>
<td>$0 \cdot 10$</td>
<td>$0 \cdot 0995$</td>
<td>(0.0479, 0.1510)</td>
<td>96.99%</td>
</tr>
<tr>
<td>$\Phi_{1,15}$</td>
<td>$0 \cdot 10$</td>
<td>$0 \cdot 1001$</td>
<td>(0.0745, 0.1256)</td>
<td>95.09%</td>
</tr>
</tbody>
</table>

Table 5.20: Summary statistics for the simulated M-STAR-GARCH model of MSG-I.
### Table 5.21: Summary statistics for the simulated M-STAR-GARCH model of MSG-I (continued)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>1</td>
<td>1</td>
<td></td>
<td>100.00%</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>6.5</td>
<td>6.32</td>
<td>(4.816, 8.265)</td>
<td>91.40%</td>
</tr>
<tr>
<td>$c$</td>
<td>0.75</td>
<td>0.7624</td>
<td>(0.7184, 0.8097)</td>
<td>92.10%</td>
</tr>
<tr>
<td>$A_{1,1}$</td>
<td>0.68</td>
<td>0.6783</td>
<td>(0.6299, 0.7277)</td>
<td>95.79%</td>
</tr>
<tr>
<td>$A_{2,2}$</td>
<td>0.46</td>
<td>0.4609</td>
<td>(0.4116, 0.5110)</td>
<td>96.60%</td>
</tr>
<tr>
<td>$A_{3,3}$</td>
<td>0.53</td>
<td>0.5319</td>
<td>(0.4850, 0.5801)</td>
<td>96.59%</td>
</tr>
<tr>
<td>$B_{1,1}$</td>
<td>0.31</td>
<td>0.3002</td>
<td>(0.2126, 0.3840)</td>
<td>96.19%</td>
</tr>
<tr>
<td>$B_{2,2}$</td>
<td>0.62</td>
<td>0.6077</td>
<td>(0.5041, 0.7000)</td>
<td>96.89%</td>
</tr>
<tr>
<td>$B_{3,3}$</td>
<td>0.70</td>
<td>0.6925</td>
<td>(0.6280, 0.7499)</td>
<td>96.59%</td>
</tr>
<tr>
<td>$C_{1,1}$</td>
<td>0.56</td>
<td>0.5659</td>
<td>(0.5022, 0.6312)</td>
<td>95.49%</td>
</tr>
<tr>
<td>$C_{2,1}$</td>
<td>0.15</td>
<td>0.1537</td>
<td>(0.1241, 0.1853)</td>
<td>96.60%</td>
</tr>
<tr>
<td>$C_{3,1}$</td>
<td>0.22</td>
<td>0.2252</td>
<td>(0.1784, 0.2756)</td>
<td>95.89%</td>
</tr>
<tr>
<td>$C_{2,2}$</td>
<td>0.30</td>
<td>0.3078</td>
<td>(0.2344, 0.3854)</td>
<td>97.49%</td>
</tr>
<tr>
<td>$C_{2,3}$</td>
<td>0.07</td>
<td>0.0730</td>
<td>(0.0462, 0.1025)</td>
<td>96.60%</td>
</tr>
<tr>
<td>$C_{3,3}$</td>
<td>0.51</td>
<td>0.5257</td>
<td>(0.4001, 0.6675)</td>
<td>95.89%</td>
</tr>
<tr>
<td>$M_k$</td>
<td>3</td>
<td>3</td>
<td></td>
<td>99.70%</td>
</tr>
<tr>
<td>$N_j$</td>
<td>2</td>
<td>2</td>
<td></td>
<td>99.80%</td>
</tr>
</tbody>
</table>
The compiled results for the simulation study are shown in Tables 5.20 and 5.21. The true models for the conditional mean and conditional covariance were identified 997 and 998 times, respectively, out of the 1,000 runs. Both the true conditional mean and covariance model were correctly identified 995 times.

The estimation of $\gamma$ and $c$ was performed reasonably well. None of the 1,000 simulation runs identified an approximation to an VAR or TVAR model for the conditional mean. The credible interval coverage probabilities were slightly below the notional credible interval percentages for $\gamma$ and $c$. The other parameters were all reasonably close to 95% coverage probability: the average coverage over all the parameters, excluding the model indexes and the delay parameter $d$, was 94.97%.

The second study, MSG-II, was performed using an M-STAR-GARCH model with a BEKK conditional covariance equation. The true model was

$$
\begin{align*}
\mathbf{x}_t &= \mathbf{\mu}_t + \mathbf{\varepsilon}_t \\
\mathbf{\varepsilon}_t &= H_t^{\frac{1}{2}} \mathbf{\eta}_t,
\end{align*}
$$

where

$$
\begin{align*}
\mathbf{\mu}_t &= 
\begin{bmatrix}
0.40 & -0.05 & -0.15 \\
0.02 & -0.20 & 0.02 \\
0.03 & -0.15 & 0.25
\end{bmatrix}
\mathbf{x}_{t-1} + 
\begin{bmatrix}
0.13 & 0.10 & 0.10 \\
1.05 & -0.11 & 0.10 \\
0.05 & -0.10 & 0.12
\end{bmatrix}
\mathbf{x}_{t-2} + \\
&\left( 
\begin{bmatrix}
-0.50 & 0.34 & -0.03 \\
-0.20 & 0.14 & -0.38 \\
0.06 & 0.24 & -0.10
\end{bmatrix}
\mathbf{x}_{t-1} + 
\begin{bmatrix}
-0.13 & 0.12 & 0.03 \\
-1.11 & 0.21 & -0.35 \\
-0.02 & 0.16 & 0.14
\end{bmatrix}
\mathbf{x}_{t-2}
\right)
\mathbf{F}_t
\end{align*}
$$

$$
\mathbf{F}_t = F_t(s_t, \gamma, c) = \frac{1}{1 + \exp\left(-\frac{9}{s_t(s_t-0.15)}\right)}.
$$

The conditional covariance equation $H_t$ in (5.7.13) was defined as follows:

$$
\begin{align*}
H_t &= 
\begin{bmatrix}
0.16 & 0.09 & 0.08 \\
0.09 & 0.17 & 0.01 \\
0.08 & 0.01 & 0.12
\end{bmatrix}
+ 
\begin{bmatrix}
0.68 & 0.35 & -0.25 \\
0.18 & 0.50 & 0.12 \\
0.40 & 0.00 & 0.35
\end{bmatrix}
\mathbf{\varepsilon}_{t-1}\mathbf{\varepsilon}_{t-1}^T
\begin{bmatrix}
0.68 & 0.35 & -0.25 \\
0.18 & 0.50 & 0.12 \\
0.40 & 0.00 & 0.35
\end{bmatrix}
+ \\
\begin{bmatrix}
0.55 & -0.15 & 0.25 \\
0.20 & 0.60 & -0.40 \\
-0.20 & 0.30 & 0.65
\end{bmatrix}
H_{t-1}
\begin{bmatrix}
0.55 & -0.15 & 0.25 \\
0.20 & 0.60 & -0.40 \\
-0.20 & 0.30 & 0.65
\end{bmatrix}^T
+ \\
\begin{bmatrix}
0.20 & 0.60 & -0.40 \\
-0.20 & 0.30 & 0.65
\end{bmatrix}
H_{t-2}
\begin{bmatrix}
0.55 & -0.15 & 0.25 \\
0.20 & 0.60 & -0.40 \\
-0.20 & 0.30 & 0.65
\end{bmatrix}^T
\end{align*}
$$

Time plots of a simulated data set from MSG-II are shown in Figure 5.59, and indicate that $N = 2,500$. When this figure is compared with the one from MSG-I in Figure 5.55, differences between
the true models become evident. The plot in Figure 5.59 indicates that the true model’s conditional covariance equation allowed the volatility of one series to impact upon the volatility of other series. The conditional mean plots also indicate some slight dependence between the series.

The results for MSG-II shown in Tables 5.22 and 5.23 indicate that the algorithm worked quite well. The correct conditional mean model order was identified 927 times out of the 1,000 simulations. The delay parameter and conditional covariance model index were both correctly identified on all runs of the algorithm. The posterior means for all parameters were close to the true values, with the worst performing parameters found in $B$.

Across all parameters, the coverage probability appeared to be close to the notional interval value. The worst performing collections of parameters were those for the conditional covariance models $A$, $B$ and $C$. There were several parameters with coverage probabilities below 90%, with one as low as 85.22%. Despite this, the mean coverage probability over these parameters was still 91.64%. The mean coverage probability for $\Phi$ was 94.79%. A typical set of posterior distributions from one of the runs of the algorithm is shown in Figures 5.60, 5.61, and 5.62.
Table 5.22: Summary statistics for the simulated M-STAR-GARCH model of MSG-II.
### Table 5.23: Summary statistics for the simulated M-STAR-GARCH model of MSG-II (continued).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Estimate</th>
<th>CrI</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{[1,1]}$</td>
<td>0.16</td>
<td>0.1553</td>
<td>(0.1152, 0.1998)</td>
<td>92.34%</td>
</tr>
<tr>
<td>$C_{[2,1]}$</td>
<td>0.09</td>
<td>0.0690</td>
<td>(0.0615, 0.1305)</td>
<td>91.69%</td>
</tr>
<tr>
<td>$C_{[3,1]}$</td>
<td>0.08</td>
<td>0.0801</td>
<td>(0.0491, 0.1125)</td>
<td>93.74%</td>
</tr>
<tr>
<td>$C_{[2,2]}$</td>
<td>0.17</td>
<td>0.1729</td>
<td>(0.1263, 0.2222)</td>
<td>90.18%</td>
</tr>
<tr>
<td>$C_{[2,3]}$</td>
<td>0.01</td>
<td>0.0076</td>
<td>(-0.0230, 0.0416)</td>
<td>85.22%</td>
</tr>
<tr>
<td>$C_{[3,3]}$</td>
<td>0.12</td>
<td>0.1208</td>
<td>(0.0829, 0.1600)</td>
<td>89.43%</td>
</tr>
<tr>
<td>$M_k$</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>92.70%</td>
</tr>
<tr>
<td>$N_j$</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

**Figure 5.60:** An example set of parameter posterior distributions from MSG-II. The red lines indicate the positions of the true values for each parameter.
Figure 5.61: An example set of parameter posterior distributions from MSG-II (continued). The red lines indicate the positions of the true values for each parameter.
5.7.2 Empirical Data

We applied the M-STAR-GARCH algorithm to the Icelandic river flow data previously analysed in Section 5.5.2. The Jökulsá and Vatnsdalsá river flows are denoted by $x_{1,t}$ and $x_{2,t}$, respectively. The transition variable $s_t$, was set to precipitation over the same period as the river flows. Recall the original time plot from Figure 5.37, shown below in Figure 5.63.

A plot of the calculated variance of the Jökulsá and Vatnsdalsá river flows over a rolling window of

Figure 5.63: Time plots for the Jökulsá river flow (top); Vatnsdalsá river flow (middle); and the precipitation (bottom).
length ten is shown in Figure 5.64. On each of the plots, there are three periods of significantly high volatility. These correspond roughly to the northern hemisphere’s summer months, during which the data plot shows higher than normal river flow each year.

The application of the pure M-STAR estimation scheme resulted in a fitted model with a conditional mean order of \( k = 3 \) and delay parameter of \( d = 4 \). To remain consistent with the M-STAR algorithm, the prior distribution for the smoothing parameter \( \gamma \) was set up with the same mean and variance. The pilot was run for 5,000 iterations for the constant conditional covariance model, and 50,000 iterations for the DBEKK and full BEKK models. A review of the pilot results for the constant conditional covariance model matrix reveals extremely similar results to those given in Section 5.5.2. The full algorithm, allowing model moves, was then run for 500,000 iterations, of which the first 50,000 iterations were discarded as a burn in.

The results for the estimation procedure are shown in Table 5.24. There was an expectation before running the algorithm that the model identified would be effectively the same as that identified in Section 5.5.2. The fitted model from the full M-STAR-GARCH estimation scheme had the same conditional mean model order, that is, \( k = 3 \). However, the delay parameter \( d \) was identified as \( d = 1 \). In addition, the model for the conditional covariance equation was identified as being of DBEKK form. The estimates for the smoothing and location parameters \((\gamma, c)\) in the transition function also differed. Under the pure M-STAR model, they were estimated to be 47.70 and 10.39, respectively, whereas under the M-STAR-GARCH model they were both substantially lower at 33.33 and 5.20.

Posterior plots for the parameters are shown in Figures 5.65 and 5.66. The posterior plots for the
model indices and delay parameter are omitted as their posterior probabilities were 100% at the point estimates. Interestingly, the posterior plot for \( \Phi_{2,3} [2,1] \) indicated some degree of bi-modality in the posterior distribution. The point estimate and credible interval in Table 5.24 were calculated without taking into account the bi-modality of the distribution. The bi-modality may have arisen from the nature of the smoothing parameter \( \gamma \). The posterior distribution for \( \gamma \) is fairly wide, with values in the left tail having considerably different effects on the conditional mean coefficient parameter matrix.

The effects on the other parameters in the conditional mean coefficient parameter matrix may present as skewness in the individual posterior distributions.

**Figure 5.65**: Posterior distributions for the M-STAR-GARCH model fitted to the Icelandic river flow data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
<th>Parameter</th>
<th>Estimate</th>
<th>CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Phi_{1,1} [1,1] )</td>
<td>1.5220</td>
<td>(1.4365, 1.6060)</td>
<td>( \Phi_{2,1} [1,1] )</td>
<td>-0.5471</td>
<td>(-0.6800, -0.4105)</td>
</tr>
<tr>
<td>( \Phi_{1,1} [2,1] )</td>
<td>0.0052</td>
<td>(-0.0026, 0.0132)</td>
<td>( \Phi_{2,1} [2,1] )</td>
<td>-0.0032</td>
<td>(-0.0213, 0.0117)</td>
</tr>
<tr>
<td>( \Phi_{1,1} [1,2] )</td>
<td>0.2925</td>
<td>(0.1653, 0.4176)</td>
<td>( \Phi_{2,1} [1,2] )</td>
<td>-0.1736</td>
<td>(-0.3376, -0.0065)</td>
</tr>
<tr>
<td>( \Phi_{1,1} [2,2] )</td>
<td>1.3204</td>
<td>(1.2189, 1.4261)</td>
<td>( \Phi_{2,1} [2,2] )</td>
<td>-0.7927</td>
<td>(-0.9403, -0.6141)</td>
</tr>
<tr>
<td>( \Phi_{1,2} [1,1] )</td>
<td>-0.5933</td>
<td>(-0.7308, -0.4702)</td>
<td>( \Phi_{2,2} [1,1] )</td>
<td>0.1784</td>
<td>(-0.0484, 0.4017)</td>
</tr>
<tr>
<td>( \Phi_{1,2} [2,1] )</td>
<td>-0.0044</td>
<td>(-0.0173, 0.0080)</td>
<td>( \Phi_{2,2} [2,1] )</td>
<td>0.0089</td>
<td>(-0.0219, 0.0321)</td>
</tr>
<tr>
<td>( \Phi_{1,2} [1,2] )</td>
<td>-0.4071</td>
<td>(-0.6174, -0.1884)</td>
<td>( \Phi_{2,2} [1,2] )</td>
<td>0.2430</td>
<td>(-0.0328, 0.5105)</td>
</tr>
<tr>
<td>( \Phi_{1,2} [2,2] )</td>
<td>-0.4360</td>
<td>(-0.5954, -0.2854)</td>
<td>( \Phi_{2,2} [2,2] )</td>
<td>0.4968</td>
<td>(0.2766, 0.7325)</td>
</tr>
<tr>
<td>( \Phi_{1,3} [1,1] )</td>
<td>0.0673</td>
<td>(-0.0225, 0.1577)</td>
<td>( \Phi_{2,3} [1,1] )</td>
<td>0.3558</td>
<td>(0.1826, 0.5287)</td>
</tr>
<tr>
<td>( \Phi_{1,3} [2,1] )</td>
<td>0.0013</td>
<td>(-0.0063, 0.0091)</td>
<td>( \Phi_{2,3} [2,1] )</td>
<td>-0.0077</td>
<td>(-0.0248, 0.0259)</td>
</tr>
<tr>
<td>( \Phi_{1,3} [1,2] )</td>
<td>0.1532</td>
<td>(0.0179, 0.2871)</td>
<td>( \Phi_{2,3} [1,2] )</td>
<td>-0.0397</td>
<td>(-0.2660, 0.1870)</td>
</tr>
<tr>
<td>( \Phi_{1,3} [2,2] )</td>
<td>0.0990</td>
<td>(0.0044, 0.2167)</td>
<td>( \Phi_{2,3} [2,2] )</td>
<td>0.2959</td>
<td>(0.0817, 0.4881)</td>
</tr>
<tr>
<td>( d )</td>
<td>1</td>
<td>-</td>
<td>( C [1,1] )</td>
<td>0.1676</td>
<td>(0.1193, 0.2273)</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>33-3285</td>
<td>(21.2718, 49.8227)</td>
<td>( C [2,1] )</td>
<td>0.1199</td>
<td>(0.0834, 0.1609)</td>
</tr>
<tr>
<td>( c )</td>
<td>5-2040</td>
<td>(4.8852, 5.4735)</td>
<td>( C [2,2] )</td>
<td>0.1311</td>
<td>(0.1009, 0.1717)</td>
</tr>
<tr>
<td>( A [1,1] )</td>
<td>0.3816</td>
<td>(0.3644, 0.3987)</td>
<td>( B [1,1] )</td>
<td>0.9239</td>
<td>(0.9167, 0.9308)</td>
</tr>
<tr>
<td>( A [2,2] )</td>
<td>0.7831</td>
<td>(0.7348, 0.8410)</td>
<td>( B [2,2] )</td>
<td>0.6164</td>
<td>(0.5354, 0.6738)</td>
</tr>
<tr>
<td>( M_k )</td>
<td>3</td>
<td>-</td>
<td>( N_j )</td>
<td>2</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 5.24**: Parameter estimates for the M-STAR-GARCH model fitted to the Icelandic river flow data.
Figure 5.66: Posterior distributions for the M-STAR-GARCH model fitted to the Icelandic river flow data (continued).
Section 5.7: Multivariate STAR-GARCH Algorithm

Setting the parameters whose 95% credible intervals contained zero equal to zero, the form of the fitted model becomes

\[
\begin{align*}
    x_t &= \begin{bmatrix} 1.52 & 0.29 \\ 0.00 & 1.32 \end{bmatrix} x_{t-1} + \begin{bmatrix} 0.60 & 0.41 \\ 0.00 & 0.44 \end{bmatrix} x_{t-2} + \begin{bmatrix} 0.00 & 0.15 \\ 0.00 & 0.10 \end{bmatrix} x_{t-3} + \\
    &+ \begin{bmatrix} 0.55 & 0.17 \\ 0.00 & 0.79 \end{bmatrix} x_{t-1} + \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.50 \end{bmatrix} x_{t-2} + \\
    &+ \begin{bmatrix} 0.36 & 0.00 \\ 0.00 & 0.30 \end{bmatrix} x_{t-3} + F_t + \epsilon_t \\

    F_t(s_t, \gamma, c) &= \left[ 1 + \exp \left( -33.33 \frac{S(s_t)}{S(s_{t-1})} (s_t - 5.20) \right) \right]^{-1} \\

    \epsilon_t &= H_t^{1/2} \eta_t
\end{align*}
\]

\[
H_t = \begin{bmatrix} 0.17 & 0.12 \\ 0.12 & 0.13 \end{bmatrix} + \begin{bmatrix} 0.38 & 0.00 \\ 0.00 & 0.78 \end{bmatrix} \epsilon_{t-1} \epsilon_{t-1}^T \begin{bmatrix} 0.38 & 0.00 \\ 0.00 & 0.78 \end{bmatrix}^T + \\

\begin{bmatrix} 0.92 & 0.00 \\ 0.00 & 0.62 \end{bmatrix} H_{t-1} \begin{bmatrix} 0.92 & 0.00 \\ 0.00 & 0.62 \end{bmatrix}^T.
\]

A plot of the fitted values is shown in red over the original data in Figure 5.67. While the majority of fitted values appear reasonable, some fitted values were underestimated in around 1973 to 1974. The transition function in Figure 5.68 is as expected, given the relatively large smoothing parameter. The function most commonly takes the value of zero or one, with a moderate proportion of values on the transition.

![Figure 5.67: Time plots for the actual (black) and the fitted (red) values for Jökulsá (top) and Vatnsdalsá (bottom) river flow.](image-url)
Applying the M-STAR-GARCH algorithm to the Icelandic river flow data was found to be quite challenging. Initially, the acceptance rate for the simulation of the conditional covariance parameters was very low. Even after significantly lowering the variance of the proposal distributions $\Delta C$ and $\Delta A$ in the first stage of the DRMH algorithm, the acceptance rate of the algorithm was still low. In order to achieve reasonable acceptance rates, the variances $\Delta^2 C$ and $\Delta^2 A$ of the proposal distributions in the second stage of the algorithm had to be lowered significantly. This resulted in higher than ideal autocorrelation in the MCMC chains.

Further investigation into why the acceptance rate was very low revealed that the conditional covariance parameters bordered upon the stationarity condition. That is, proposals were often made that did not satisfy the condition, leading to zero posterior densities due to the influence of the prior distribution. Testing the fitted values for the covariance equation parameters to see if they met the stationarity conditions gave the following results:

$$\hat{A}[1, 1]^2 + \hat{B}[1, 1]^2 = 0.9992$$
$$\hat{A}[2, 2]^2 + \hat{B}[2, 2]^2 = 0.9932.$$ 

Given that $\hat{A}[i, i]^2 + \hat{B}[i, i]^2$ for $i \in \{1, 2\}$ were only slightly less than one, the MCMC chains were located in the neighbourhood of the non-stationarity threshold. Following the work of Vermaak et al. 2004, in which estimation schemes for AR models with and without enforcing stationarity were developed, our estimation scheme was reapplied to the data with the stationarity condition removed. That is, the prior distribution was modified to remove the indicator function testing the stationarity condition.

The unconstrained algorithm was run for the same number of iterations as the constrained case, with the same burn in periods. The unconstrained algorithm resulted in significantly better MCMC results. The tuning parameters for the conditional covariance parameters were able to be set at larger values,
resulting in better mixing of the chains and lower autocorrelation. As expected, the parameter chains for the conditional covariance parameters moved into the area of non-stationarity. The parameter estimates are presented in Table 5.25.

Table 5.26 allows the comparison of the point estimates for the parameters in the scenario where

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Constrained</th>
<th>Unconstrained</th>
<th>Parameter</th>
<th>Constrained</th>
<th>Unconstrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi_{1,1} [1, 1]$</td>
<td>1.5220</td>
<td>1.3197</td>
<td>$\Phi_{2,1} [1, 1]$</td>
<td>0.4968</td>
<td>-0.4597</td>
</tr>
<tr>
<td>$\Phi_{1,1} [1, 2]$</td>
<td>0.0052</td>
<td>-0.0007</td>
<td>$\Phi_{2,1} [2, 2]$</td>
<td>0.3585</td>
<td>1.0747</td>
</tr>
<tr>
<td>$\Phi_{1,1} [2, 1]$</td>
<td>0.2925</td>
<td>0.3060</td>
<td>$\Phi_{2,2} [1, 1]$</td>
<td>-0.0077</td>
<td>0.0065</td>
</tr>
<tr>
<td>$\Phi_{1,1} [2, 2]$</td>
<td>1.3204</td>
<td>1.3244</td>
<td>$\Phi_{2,2} [1, 2]$</td>
<td>-0.0397</td>
<td>-0.2644</td>
</tr>
<tr>
<td>$\Phi_{1,2} [1, 1]$</td>
<td>-0.5993</td>
<td>-0.1734</td>
<td>$\Phi_{2,3} [2, 2]$</td>
<td>0.2959</td>
<td>0.5737</td>
</tr>
<tr>
<td>$\Phi_{1,2} [2, 1]$</td>
<td>-0.0044</td>
<td>0.0066</td>
<td>$\gamma$</td>
<td>33.3285</td>
<td>28.9649</td>
</tr>
<tr>
<td>$\Phi_{1,2} [2, 2]$</td>
<td>-0.4071</td>
<td>-0.5460</td>
<td>$c$</td>
<td>5.2040</td>
<td>4.7750</td>
</tr>
<tr>
<td>$\Phi_{2,1} [1, 1]$</td>
<td>-0.4360</td>
<td>-0.3878</td>
<td>$A [1, 1]$</td>
<td>0.1676</td>
<td>0.0471</td>
</tr>
<tr>
<td>$\Phi_{2,1} [1, 2]$</td>
<td>0.0673</td>
<td>-0.1585</td>
<td>$C [1, 1]$</td>
<td>0.1199</td>
<td>0.0242</td>
</tr>
<tr>
<td>$\Phi_{2,2} [1, 1]$</td>
<td>0.0013</td>
<td>-0.0033</td>
<td>$C [2, 2]$</td>
<td>0.1311</td>
<td>0.1050</td>
</tr>
<tr>
<td>$\Phi_{2,2} [1, 2]$</td>
<td>0.1532</td>
<td>0.2741</td>
<td>$A [1, 2]$</td>
<td>0.3816</td>
<td>1.5657</td>
</tr>
<tr>
<td>$\Phi_{2,2} [2, 2]$</td>
<td>0.0990</td>
<td>0.0417</td>
<td>$A [2, 1]$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\Phi_{2,3} [1, 1]$</td>
<td>-0.5471</td>
<td>-0.3234</td>
<td>$A [2, 2]$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\Phi_{2,3} [1, 2]$</td>
<td>-0.0032</td>
<td>-0.0036</td>
<td>$B [1, 1]$</td>
<td>0.7831</td>
<td>1.2624</td>
</tr>
<tr>
<td>$\Phi_{2,3} [2, 1]$</td>
<td>-0.1736</td>
<td>-0.3041</td>
<td>$B [1, 2]$</td>
<td>0.9239</td>
<td>0.7458</td>
</tr>
<tr>
<td>$\Phi_{2,3} [2, 2]$</td>
<td>-0.7927</td>
<td>-0.1669</td>
<td>$B [2, 2]$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\Phi_{2,4} [1, 1]$</td>
<td>-0.1784</td>
<td>-0.7303</td>
<td>$B [2, 2]$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\Phi_{2,4} [2, 1]$</td>
<td>0.0089</td>
<td>-0.0007</td>
<td>$B [2, 2]$</td>
<td>0.6164</td>
<td>0.5858</td>
</tr>
<tr>
<td>$\Phi_{2,4} [2, 2]$</td>
<td>0.2430</td>
<td>0.5107</td>
<td>$B [2, 2]$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

Table 5.26: Comparison of point estimates for the contained and unconstrained M-STAR-GARCH models.
stationarity is enforced, and where it is not. Some parameters have very similar estimates, while others do not. The key difference between the two cases is that, for the unconstrained case, the estimation algorithm resulted in a BEKK model for the conditional covariance, whereas a DBEKK model arose in the constrained scenario. This difference, no doubt, contributed to the differences in estimates for some parameters.

Posterior plots for the unconstrained case are shown in Figures 5.69 and 5.70. Again, the posterior plots for the model indices and delay parameter are omitted as their posterior probabilities were 100% at the point estimates, indicating extremely strong evidence in support of these model indices. It is

![Posterior distributions](image)

**Figure 5.69:** Posterior distributions for the unconstrained M-STAR-GARCH model fitted to the Icelandic river flow data.
worth noting that the posterior distribution for $\Phi_{2,3} [2, 1]$ is now unimodal.

The plots of the transition function in Figure 5.71 show a similar result to the constrained case. This is expected, given the similarities between the point estimates of the smoothing and location parameters, $\gamma$ and $c$. The original data is plotted with the fitted values from the unconstrained model overlaid in Figure 5.72. As in the constrained case, some peaks were underestimated during the period from 1973 to 1974.

After setting the parameters with 95% credible intervals containing zero equal to zero, the form of the
The fitted model for the unconstrained M-STAR-GARCH algorithm becomes

\[
x_t = \begin{bmatrix} 1.32 & 0.31 \\ 0.00 & 1.32 \end{bmatrix} x_{t-1} + \begin{bmatrix} -0.17 & -0.55 \\ 0.00 & -0.39 \end{bmatrix} x_{t-2} + \begin{bmatrix} -0.16 & 0.27 \\ 0.00 & 0.00 \end{bmatrix} x_{t-3} +
\]

\[
\begin{pmatrix} -0.32 & -0.30 \\ 0.00 & -0.17 \end{pmatrix} x_{t-1} + \begin{pmatrix} -0.73 & 0.51 \\ 0.00 & -0.43 \end{pmatrix} x_{t-2} +
\begin{pmatrix} 1.07 & -0.26 \\ 0.00 & 0.57 \end{pmatrix} x_{t-3} \right) F_t + \varepsilon_t,
\]

\[
F_t(s_t, \gamma, c) = \left[ 1 + \exp \left( -\frac{28.96}{S(s_t)} (s_{t-1} - 4.78) \right) \right]^{-1}
\]

\[
\varepsilon_t = H_t^{\frac{1}{2}} \eta_t
\]

\[
H_t = \begin{bmatrix} 0.05 & 0.00 \\ 0.00 & 0.11 \end{bmatrix} + \begin{bmatrix} 1.06 & -0.51 \\ 0.00 & 1.23 \end{bmatrix} \varepsilon_{t-1}^{T} \varepsilon_{t-1} + \begin{bmatrix} 1.06 & -0.51 \\ 0.00 & 1.23 \end{bmatrix}^{T} +
\]

\[
\begin{bmatrix} 0.75 & 0.28 \\ 0.01 & 0.59 \end{bmatrix} H_{t-1} \begin{bmatrix} 0.75 & 0.28 \\ 0.01 & 0.59 \end{bmatrix}^{T}.
\]

Conditional on the diagnostics performance of each model, the DIC was calculated for the three estimation schemes applied to this data, namely, the pure M-STAR algorithm and the M-STAR-GARCH algorithm, with and without enforcing the stationarity of the conditional covariance. The DIC scores were 11,738.38, 9,106.26 and 8,782.89, respectively. Given that the full M-STAR-GARCH algorithms did not identify a pure M-STAR model for the data, it was expected that the DIC score for the pure M-STAR model would be highest. Also, as the constrained run of the algorithm appeared to be restricted by the stationarity constraint, it was also expected that the model resulting from the unconstrained application of the algorithm would result in a lower DIC score.

The fitted conditional covariance and correlation plots for the M-STAR-GARCH model from the un-
constrained run of the algorithm are shown in Figure 5.73. The conditional variance plots mirror the shapes of those given in Figure 5.64. The extremely large fitted conditional variance before June 1974 in each plot hides the shape somewhat, but the plots indicate three clusters that correspond to the northern hemisphere’s Summer months.

**Figure 5.72:** Time plots for actual (black) and fitted (red) values for Jökulsá (top) and Vatnsdalsá (bottom) river flow for the unconstrained case.

**Figure 5.73:** Fitted conditional covariance and correlation over the time period for the unconstrained case.
Chapter 6

Conclusion and Future Work

The ultimate objective and original contribution of this thesis was the development of a set of algorithms that combine the tasks of parameter estimation and model selection for univariate and multivariate STAR-GARCH models into a single process. Algorithms were developed to achieve this aim for subsets of the univariate and multivariate STAR-GARCH models. The model index for the conditional mean and conditional covariance were both successfully included as parameters for estimation in these models.

The theoretical components of this work for univariate and multivariate models are explored in chapters three and four, respectively. These chapters and corresponding appendices contain detailed derivations of the joint posterior distributions. Given that the joint posterior distributions are of non-standard forms, MCMC algorithms must be used to simulate from them. Consequently, the required conditional posterior distributions are also derived in detail. The design and outline of all MCMC estimation schemes are presented in chapters three and four, which also provide the details of any sub-algorithms that they may require.

The performance of each estimation scheme is investigated in chapter five through the implementation of extensive simulation studies. These studies address not only the point estimate performance, but also the credible interval performance of each scheme. Practical applications of the estimation schemes to real world data are provided for each model type.

For each of the six models considered, STAR, GARCH, STAR-GARCH, M-STAR, M-GARCH, and M-STAR-GARCH we observe that:

(i) The design of posterior simulators using a combination of MCMC algorithms which include the Gibbs sampler, Metropolis-Hastings, MTM, DRMH and RJMCMC algorithms is computationally
feasible and can be implemented in R.

(ii) The proposed methods provide reliable estimates, based upon the performance of the point estimates and coverage probabilities. This was determined through extensive simulation studies for each of the estimation schemes.

(iii) Uncertainties surrounding all parameters including (where applicable) the model order of the conditional mean and the model index for the conditional covariance are incorporated into the estimation procedures.

(iv) The task of model selection can be embedded into the proposed methods using a Bayesian RJMCMC approach.

The following are some specific findings for each of the estimation procedures developed.

We found that the algorithm for the STAR model was faster and more effective when a large initial value for the smoothing parameter $\gamma$ was provided. The posterior estimates obtained under different prior specifications were insensitive to the specification of the prior distributions imposed on the implicit smoothing and location parameters $\gamma$ and $c$, respectively. Both of these findings are consistent with the work of Livingston Jr, Nur, and Hudson (2013) and Livingston Jr and Nur (2016a). It was also found that the successful estimation of the smoothing parameter was important for correctly estimating the other parameters for the model.

For the GARCH estimation scheme, one of the simulation studies showed that the variability of the point estimates for $\alpha_0$ from replication to replication was quite large, despite the posterior mean estimate being reasonably close to the true value. This could have been due to the mix of true parameter values chosen for the study as it was not repeated in the other GARCH simulation study. The application of the GARCH estimation scheme to real data yielded results that were relatively consistent with those found in Ardia (2008). In Ardia (2008), the results were obtained by fixing the model index and estimating the parameters conditional on the fixed model orders, rather than using the RJMCMC scheme employed here.

Through the application of the STAR-GARCH estimation scheme (Livingston Jr and Nur 2014) to real world data, we were able to demonstrate the ways in which the algorithm allows for the identification and estimation of subsets of the STAR-GARCH class of models. When applied to the SOI data, the final model was an AR(2)-GARCH(1,2) model. In addition, as part of the simulation study, the algorithm was shown to successfully identify a pure STAR model, that is, an AR($k$)-GARCH(0,0) model.

The estimation scheme for the M-STAR model (Livingston Jr and Nur 2015) highlighted that, as for the univariate STAR model, the estimation of the smoothing parameter $\gamma$ can be difficult and
impacts on how successfully the other parameters for the model are estimated. The application of the estimation procedure to the bivariate Icelandic river flow data resulted in a conditional mean model order different from those identified by previous authors who analysed the data. There were, however, significant differences between the models applied and the estimation methodology employed.

We observed that, for an M-GARCH model estimation scheme, the design of a posterior simulator under the VECH formulation of the conditional covariance equation using an estimation scheme similar to that discussed in Hudson and Gerlach (2008) led to inconsistent results when implemented. The scheme was also relatively slow, even for bivariate series. When the final BEKK estimation scheme developed was applied to real world data, it was critical for the success of the algorithm to obtain good estimates for the proposal location parameters. This necessitated running the pilots for significantly longer than was required for the simulation studies. A useful outcome of applying the M-GARCH estimation scheme was that the results could be used to determine how the correlation between series changed over time (Livingston Jr and Nur 2016b).

It was possible to identify specific conditional covariance equation formulations for the M-STAR-GARCH estimation scheme for the constant, DBEKK and BEKK forms. The estimation procedure was applied to the same Icelandic river flow data as the M-STAR model. While the same conditional mean model order was identified in both cases, the presence of a non-constant conditional covariance significantly changed the point estimates of the entries of the conditional mean coefficient parameter matrix. This indicates that correctly identifying the appropriate model for the conditional covariance equation is critical when estimating the parameters for the conditional mean. Interestingly, for the application to the river flow data, when the constraint based on the stationarity condition was lifted, the conditional mean parameters were again affected as the final model for the conditional covariance was then identified as a full BEKK model.

As in the M-GARCH model, it was found to be critically important to obtain good proposal distribution location parameters for the covariance coefficient parameter matrix for the M-STAR-GARCH model. If the pilot runs were not long enough, suitable location parameters were not always obtained, and the algorithm was sometimes unable to jump to the most suitable model for the conditional covariance.

Due to the irregular shapes of the posterior distributions for the M-STAR-GARCH and M-GARCH models, there was a risk that some MCMC chains would become trapped within local maxima. In order to reduce the chance of this occurring, the proposal variances needed to be set as large as possible, while still trying to maintain reasonable acceptance rates. These algorithms also required a large number of iterations in order to ensure convergence of the MCMC chains, which in turn meant the run time of the algorithm was quite large. Another reason for needing to run the algorithms for many iterations relates to the fact that the MCMC chains for the GARCH coefficient parameter matrices are highly
correlated. For the Icelandic river flow data analysed in Section 5.7.2, the algorithm was run for 500,000 iterations, taking several days to complete. Even with 500,000 iterations, the effective sample size for the MCMC chains was still only around 2,000.

Overall, the aims of the thesis were achieved, but some limitations and possible areas for future extensions were identified. First, the computation algorithms were implemented in the statistical software package R without taking measures to optimise the speed of the code. While the algorithms were shown to work effectively, they were not as efficient as expected. As previously mentioned, the algorithm took a very long time to run. Implementation of the code in a faster programming language such as C++ would yield significant improvements in the speed of the algorithms.

When estimating an AR model through the structure of a STAR model, the MCMC chains for the smoothing and location parameters, $\gamma$ and $c$, respectively, are sometimes able to “walk” around the parameter space. This can lead to computational issues with the algorithm, which ultimately cause failure. While this identifiability issue was contained though the specification of the prior distributions for $\gamma$ and $c$, the prior distributions employed were relatively unsophisticated. There is scope for improvement through the use of a prior distribution structure that better takes into account the magnitude of $\gamma$.

The model index for the conditional mean and conditional covariance employed in this thesis did not allow particular parameters to be set equal to zero. For example, in the real world application of the STAR-GARCH estimation scheme, the fitted model for the conditional variance was a GARCH(1,2) model. A review of the posterior distribution for $\beta_1$ showed that the density was highest towards zero, possibly indicating that this parameter estimate should be zero, while still allowing a non zero estimate for $\beta_2$. One way to approach this would be to add different combinations of parameters, some of which are set to zero, while others are allowed to vary, to the index of models for searching. Another option is to apply a spike and slab type prior distribution as discussed in Mitchell and Beauchamp (1988).

With the increased availability of parallel computing, the advantage of having a single scheme for parameter estimation and model selection is somewhat diminished. However, the value of a scheme that combines the two aspects of modelling is still evident. The fact that the combination is possible is potentially a move towards a fully automated “black box” time series modelling tool. A fully automated modelling procedure would be advantageous to users with limited experience in time series modelling and/or those who lack the skills to code an estimation scheme for more complicated models.

To facilitate this aim, there are several ways in which the work of this thesis could be extended. While some unsophisticated adaptive procedures were implemented when coding the algorithms, there are other, more effective adaptive techniques that would allow automatic tuning of the acceptance rates within the algorithms. This would eliminate the need for the user to control the acceptance rates by
specifying the tuning parameters of the algorithms.

Another extension would be to include more classes of models in the searches. There are a plethora of time series models that could easily be incorporated into the model indices. For example, other transition functions such as the exponential function, giving rise to an ESTAR model, could be included within the conditional mean model index. The MSAR model could be added as an alternative regime switching formulation for the conditional mean. Asymmetric covariance equations could be added to the index of possible models for the conditional covariance.

The error distributions used throughout this thesis were the univariate Normal and multivariate Normal distributions for univariate and multivariate models, respectively. Changing these to t-distributions for which the degrees of freedom are a parameter to be estimated would generalise the process even further. Some financial time series would be better suited to models that incorporate a fatter tailed error distribution.

While the goal of creating a relatively automatic time series modelling procedure is a highly ambitious task, it is not an impossible one. The abovementioned areas of future work are realistic steps upon the path towards achieving this.
Appendix A

Univariate Smooth Transition Autoregressive Models

A.1 Posterior Distribution Derivations

For the univariate STAR model, several conditional posterior distributions are required for the MCMC posterior simulator. These derivations are performed in the following subsections.

A.1.1 Conditional Posterior for \((M_k, \phi)\)

By assuming that all the parameters in \(\Theta_{-(M_k, \phi)}\) are constant and starting with the joint posterior distribution in (3.1.7), we are able to derive the joint conditional posterior distribution for \(M_k\) and \(\phi\) as follows:

\[
\begin{align*}
p(M_k, \phi | x, \Theta_{-(M_k, \phi)}) & \propto [2\pi]^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T (\sigma^2 I_n)^{-1} (x - Z_1 \phi) \right] \times \\
& \times [2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \phi^T \Sigma_{\phi}^{-1} \phi \right] \times \\
& \times \frac{\beta_0^{a_0}}{\Gamma(a_0)} (\sigma^2)^{-a_0 - 1} \exp \left[ -\frac{\beta_0}{(\sigma^2)} \right] \frac{\Lambda_k^{k}}{k!} \mathbb{I}_k(k) \mathbb{I}_D(d) \times \\
& \times \frac{\beta_{\gamma}^{c_\gamma}}{\Gamma(c_\gamma)} \gamma^{c_\gamma - 1} \exp \left[ -\beta_{\gamma} \right] [2\pi \sigma_c^2]^{-\frac{1}{2}} \exp \left[ -\frac{(c - \mu_c)^2}{2 \sigma_c^2} \right] \\
& \propto [2\pi]^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T (\sigma^2 I_n)^{-1} (x - Z_1 \phi) \right] \times \\
& \times [2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \phi^T \Sigma_{\phi}^{-1} \phi \right] \frac{\Lambda_k^{k}}{k!} \mathbb{I}_k(k).
\end{align*}
\]
Chapter A: Univariate Smooth Transition Autoregressive Models

A.1.2 Conditional Posterior for $\phi$

Beginning with the joint posterior distribution in (3.1.7) and assuming that all parameters in $\Theta_{-\phi}$ are constant, we can calculate:

$$p(\phi|x, \Theta_{-\phi}) \propto [2\pi]^{-\frac{d}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (x - Z_1 \phi)^T (\sigma^2 I_n)^{-1} (x - Z_1 \phi) \right] \times$$

$$[2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \phi^T \Sigma_{\phi}^{-1} \phi \right] \times$$

$$\frac{\beta_0^\alpha}{\Gamma(\alpha_0)} (\sigma^2)^{-\alpha_0 - 1} \exp \left[-\frac{\beta_0}{(\sigma^2)} \right] \frac{1}{k!} I_k(k) I_0(d) \times$$

$$\frac{\beta_1^\gamma}{\Gamma(\gamma_0)} \gamma^{-\gamma - 1} \exp \left[-\beta_1 \gamma \right] [2\pi \sigma_c^2]^{-\frac{d}{2}} \exp \left[- \frac{(c - \mu_c)^2}{2\sigma_c^2} \right]$$

$$\propto [2\pi]^{-\frac{d}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (x - Z_1 \phi)^T (\sigma^2 I_n)^{-1} (x - Z_1 \phi) \right] \times$$

$$[2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \times$$

$$\exp \left[-\frac{1}{2\sigma^2} (x - Z_1 \phi)^T (x - Z_1 \phi) - \frac{1}{2} \phi^T \Sigma_{\phi}^{-1} \phi \right]$$

$$\propto [2\pi \sigma^2]^{-\frac{d}{2}} [2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \times$$

$$\exp \left[-\frac{1}{2 \sigma^2} \left( x^T - \phi^T Z_1^T \right) (x - Z_1 \phi) + \phi^T \Sigma_{\phi}^{-1} \phi \right]$$

$$\propto [2\pi \sigma^2]^{-\frac{d}{2}} [2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \times$$

$$\exp \left[-\frac{1}{2} \left( \frac{x^T}{\sigma^2} - \phi^T Z_1^T \right) (x - Z_1 \phi) + \phi^T \Sigma_{\phi}^{-1} \phi \right]$$

$$\propto [2\pi \sigma^2]^{-\frac{d}{2}} [2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \times$$

$$\exp \left[-\frac{1}{2} \left( \frac{x^T}{\sigma^2} + \frac{\phi^T Z_1^T}{\sigma^2} + \frac{\phi^T \Sigma_{\phi}^{-1} \phi}{\sigma^2} - \frac{\phi^T Z_1^T x}{\sigma^2} - \frac{x^T Z_1 \phi}{\sigma^2} \right) \right]$$

$$\propto [2\pi \sigma^2]^{-\frac{d}{2}} [2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} \times$$

$$\exp \left[-\frac{1}{2} \left( \frac{x^T}{\sigma^2} + \phi^T \left( \frac{Z_1^T Z_1}{\sigma^2} + \Sigma_{\phi}^{-1} \phi - \frac{Z_1^T Z_1 x}{\sigma^2} \right) - \frac{x^T Z_1 \phi}{\sigma^2} \right) \right]$$

Letting $C_{\phi}^{-1} = \frac{Z_1^T Z_1}{\sigma^2} + \Sigma_{\phi}^{-1}$ allows for further simplification of the conditional posterior distribution.
Section A.1: Posterior Distribution Derivations

for $\phi$. Continuing with the derivation, we now have

$$p(\phi | x, \Theta, \sigma) \propto \left[ 2\pi \sigma^2 \right]^{-\frac{n}{2}} \left[ 2\pi \right]^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2} \left( \frac{x^T x}{\sigma^2} + \phi^T C^{-1}_\phi \left( \phi - \hat{\phi} \right) \right) \right]$$

$$\propto \left[ 2\pi \sigma^2 \right]^{-\frac{n}{2}} \left[ 2\pi \right]^{-\frac{k+1}{2}} |\Sigma|^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2} \left( \frac{x^T x}{\sigma^2} + \phi^T C^{-1}_\phi \left( \phi - \hat{\phi} \right) \right) \right]$$

$$\propto \left[ 2\pi \sigma^2 \right]^{-\frac{n}{2}} \left[ 2\pi \right]^{-\frac{k+1}{2}} |\Sigma|^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2} \left( \frac{x^T x}{\sigma^2} + \phi^T C^{-1}_\phi \left( \phi - \hat{\phi} \right) - \phi^T C^{-1}_\phi \hat{\phi} \right) \right]$$

To further simplify the above expression, let $\hat{\phi} = \frac{C_{\phi} Z^T x}{\sigma^2}$. Then

$$p(\phi | x, \Theta, \sigma) \propto \left[ 2\pi \sigma^2 \right]^{-\frac{n}{2}} \left[ 2\pi \right]^{-\frac{k+1}{2}} |\Sigma|^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2} \left( \frac{x^T x}{\sigma^2} + \phi^T C^{-1}_\phi \left( \phi - \hat{\phi} \right) - \phi^T C^{-1}_\phi \hat{\phi} \right) \right]$$

Breaking up the exponent in the above results in a conditional posterior distribution that is proportional
to a multivariate normal distribution as follows:

\[
p(\phi|\mathbf{x}, \Theta, \phi) \propto \left[2\pi\sigma^2\right]^{-\frac{n}{2}} |\Sigma_{\phi}|^{-\frac{1}{2}} |C_{\phi}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\frac{\mathbf{x}^T \mathbf{x}}{\sigma^2} - \phi^T C_{\phi}^{-1} \phi\right)\right] \times [2\pi]^{-\frac{2(k+1)}{2}} |C_{\phi}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left((\phi - \hat{\phi})^T C_{\phi}^{-1} (\phi - \hat{\phi})\right)\right].
\]  

(A.1.1)

The conditional posterior distribution in (A.1.1) is proportional to a multivariate normal distribution with mean vector \(\hat{\phi}\) and covariance matrix \(C_{\phi}\). That is,

\[
p(\phi|\mathbf{x}, \Theta, \phi) \sim \mathcal{N}\left(\hat{\phi}, C_{\phi}\right),
\]

where

\[
\hat{\phi} = \frac{C_{\phi} \mathbf{Z}_1^T \mathbf{x}}{\sigma^2},
\]

\[
C_{\phi} = \left[\frac{\mathbf{Z}_1^T \mathbf{Z}_1}{\sigma^2} + \Sigma_{\phi}^{-1}\right]^{-1}.
\]

(A.1.2)

### A.1.3 Conditional Posterior for \(\sigma^2\)

Starting with the joint posterior distribution in (3.1.7) and assuming that all of the parameters in \(\Theta_{-\sigma^2}\) are constant, we can calculate

\[
p(\sigma^2|\mathbf{x}, \Theta_{-\sigma^2}) \propto \left[2\pi\right]^{-\frac{n}{2}} |\Sigma|^{-\frac{n}{2}} |C|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\mathbf{x}^T \mathbf{x} - \mathbf{Z}_1 \phi^T C^{-1} \mathbf{Z}_1 \phi\right)\right] \times [2\pi]^{-\frac{2(k+1)}{2}} |C|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left((\phi - \hat{\phi})^T C^{-1} (\phi - \hat{\phi})\right)\right] 
\]

\[
\times \left[\frac{\beta_0}{\Gamma(\alpha_0)} (\sigma^2)^{-\alpha_0 - 1} \exp \left[-\frac{\beta_0}{\sigma^2}\right]\right] \left[\frac{\beta_0}{\Gamma(\alpha)} (\sigma^2)^{-\alpha - 1} \exp \left[-\frac{\beta_0}{\sigma^2}\right]\right] \left[\frac{\beta_0}{\Gamma(\alpha_0)} (\sigma^2)^{-\alpha_0 - 1}\right] \left[\frac{\beta_0}{\Gamma(\alpha_0)} (\sigma^2)^{-\alpha_0 - 1}\right] \times \left(\frac{\beta_0}{\Gamma(\alpha)} (\sigma^2)^{-\alpha - 1}\right) \left(\frac{\beta_0}{\Gamma(\alpha_0)} (\sigma^2)^{-\alpha_0 - 1}\right) \times 
\]

\[
\exp \left[-\frac{\beta_0}{\sigma^2} - \frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{Z}_1 \phi)^T (\mathbf{x} - \mathbf{Z}_1 \phi)\right] 
\]

\[
\propto (\sigma^2)^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \left(\frac{\beta_0}{\sigma^2} + \frac{1}{2}(\mathbf{x} - \mathbf{Z}_1 \phi)^T (\mathbf{x} - \mathbf{Z}_1 \phi)\right)\right] 
\]

\[
\propto (\sigma^2)^{-\alpha_0 - \frac{n}{2} - 1} \exp \left[-\frac{1}{2} \left(\frac{\beta_0}{\sigma^2} + \frac{1}{2}(\mathbf{x} - \mathbf{Z}_1 \phi)^T (\mathbf{x} - \mathbf{Z}_1 \phi)\right)\right] 
\]

\[
\propto (\sigma^2)^{-\alpha_0 - \frac{n}{2} - 1} \exp \left[-\frac{1}{2} \left(\frac{\beta_0}{\sigma^2} + \frac{1}{2}(\mathbf{x} - \mathbf{Z}_1 \phi)^T (\mathbf{x} - \mathbf{Z}_1 \phi)\right)\right].
\]

(A.1.3)
The expression in (A.1.3) is proportional to an inverted gamma density. That is,
\[
p(\sigma^2|x, \Theta_{-\sigma^2}) \sim IG(\alpha_k, \beta_k),
\]
where
\[
\alpha_k = \alpha_0 + \frac{n}{2},
\]
\[
\beta_k = \beta_0 + \frac{1}{2}(x - Z_1\phi)^T(x - Z_1\phi).
\]

### A.1.4 Conditional Posterior for \(d\)

Using the joint posterior distribution in (3.1.7) and assuming that the parameters in \(\Theta_{-d}\) are constant, we arrive at the full conditional posterior distribution below:

\[
p(d|x, \Theta_{-d}) \propto \frac{n}{2}\pi\left|\sigma^2I_n\right|^{-\frac{1}{2}} \exp \left[-\frac{1}{2}(x - Z_1\phi)^T(\sigma^2I_n)^{-1}(x - Z_1\phi)\right] \mathbb{1}_D(d)
\]

\[
\propto \exp \left[-\frac{1}{2\sigma^2}(x - Z_1\phi)^T(x - Z_1\phi)\right] \mathbb{1}_D(d).
\]

The conditional posterior for \(d\) is a discrete distribution in which the posterior probabilities are simply proportional to the likelihood values for each \(d \in D\).

### A.1.5 Conditional Posterior for \((\gamma, c)\)

Assuming that all parameters in \(\Theta_{-d}\) are constant and using the joint posterior distribution in (3.1.7), we arrive at the full conditional posterior distribution for \(\gamma\) and \(c\) shown below:

\[
p(\gamma, c|x, \Theta_{-(\gamma, c)})
\]

\[
\propto \exp \left[-\frac{1}{2}(x - Z_1\phi)^T(\sigma^2I_n)^{-1}(x - Z_1\phi)\right] \gamma^{\alpha - 1} \exp \left[-\beta_\gamma \cdot \exp \left[-\frac{(c - \mu_c)^2}{2\sigma_c^2}\right]\right] \tag{A.1.4}
\]

\[
\propto \gamma^{\alpha - 1} \exp \left[-\frac{1}{2\sigma^2}(x - Z_1\phi)^T(x - Z_1\phi) - \beta_\gamma - \frac{(c - \mu_c)^2}{2\sigma_c^2}\right].
\]

The conditional posterior distribution in (A.1.4) is of a non-standard form. Consequently, an embedded MTM step is used for the simulation of these parameters. We chose this algorithm in an attempt to improve the acceptance rates for this stage of the algorithm.
A.2 Acceptance Probabilities

The RJMCMC and MTM acceptance probabilities used in the univariate STAR posterior simulator are derived in the following subsections.

A.2.1 Reversible Jump Acceptance Probability

The key ingredient required to simplify the acceptance probability shown in (3.1.10) is the Candidate’s Identity (Besag 1989). The Candidate's Identity may be written in the notation relevant for this thesis as:

\[
p \left( \frac{M_k^*, \phi^* | x, \Theta_{-(M_k, \phi)}}{\phi^* | M_k^*, x, \Theta_{-(M_k, \phi)}} \right) = \frac{p(x | M_k^*, \Theta_{-(M_k, \phi)}) p(M_k^* | \Theta_{-(M_k, \phi)})}{p(x | \Theta_{-(M_k, \phi)})}.
\]  

We can rearrange the Candidate’s Identity in (A.2.5) to make \( p(M_k^*, \phi^* | x, \Theta_{-(M_k, \phi)}) \) the subject of the formula. Similarly, an expression for \( p(M_k, \phi | x, \Theta_{-(M_k, \phi)}) \) can be found through the use of the Candidate’s Identity.

Focusing on the quotient shown on the right hand side of the acceptance probability from (3.1.10), and substituting in the expressions found using the Candidate’s Identity results in the following expression for this quotient:

\[
\pi \left( \frac{M_k^*, \phi^* | x, \Theta_{-(M_k, \phi)}}{M_k, \phi | x, \Theta_{-(M_k, \phi)}} \right) \frac{q(\phi | \phi^*, M_k, M_k^*)}{q(\phi^* | \phi, M_k, M_k^*)} = \frac{p(x | M_k^*, \Theta_{-(M_k, \phi)}) p(M_k^* | \Theta_{-(M_k, \phi)}) p(M_k | \Theta_{-(M_k, \phi)}) p(\phi | M_k, x, \Theta_{-(M_k, \phi)}) q(\phi^* | \phi, M_k, M_k^*)}{p(x | M_k, \Theta_{-(M_k, \phi)}) p(M_k | \Theta_{-(M_k, \phi)}) p(\phi | M_k, x, \Theta_{-(M_k, \phi)}) q(\phi^* | \phi, M_k, M_k^*)}.
\]

The proposal distribution for the coefficient parameter \( \phi \) is defined to be the full conditional posterior distribution, that is \( q(\phi^* | \phi, m, m^*) \sim p(\phi^* | M_k^*, x, \Theta_{-(M_k, \phi)}) \). This distribution was derived in Appendix A.1.2. The acceptance probability therefore simplifies even further to yield

\[
p \left( \frac{x | M_k^*, \Theta_{-(M_k, \phi)}}{x | M_k, \Theta_{-(M_k, \phi)}} \right) p(M_k | \Theta_{-(M_k, \phi)}) p(M_k^* | \Theta_{-(M_k, \phi)}) = \int p(x | M_k, \phi, \Theta_{-(M_k, \phi)}) p(\phi | M_k, \Theta_{-(M_k, \phi)}) d\phi.
\]

In the expression shown in (A.2.6), \( p(M_k | \Theta_{-(M_k, \phi)}) \) is simply the density of the prior distribution for the model order as assigned in (3.1.5). The other element is the marginal likelihood \( p(x | M_k, \Theta_{-(M_k, \phi)}) \) which may be derived using the following expression:

\[
p(x | M_k, \Theta_{-(M_k, \phi)}) \propto \int p(x | M_k, \phi, \Theta_{-(M_k, \phi)}) p(\phi | M_k, \Theta_{-(M_k, \phi)}) d\phi.
\]
The terms in the integrand in (A.2.7) are the likelihood function and a conditional prior distribution for \( \phi \), both of which are multivariate normal. We have already calculated their product in the derivation of the full conditional posterior distribution for \( \phi \), the expression for which is shown in (A.1.2). Substituting this into (A.2.7) yields

\[
\frac{p(\mathbf{x}|\mathcal{M}_k, \Theta_{-(\mathcal{M}_k)})}{p(\mathbf{x}|\mathcal{M}_k, \Theta_{-(\mathcal{M}_k)})} \propto \int [2\pi \sigma^2]^{-\frac{3}{2}} |\Sigma_\phi|^{-\frac{1}{2}} |C_\phi|^{\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \frac{\mathbf{x}^T \mathbf{x}}{\sigma^2} - \hat{\phi}^T C_\phi^{-1} \hat{\phi} \right) \right] 
\]

Substitution of the simplified marginal likelihood shown in (A.2.8), together with the prior distribution for the model order \( \mathcal{M}_k \), into the acceptance probability in (A.2.6) gives the following simplified expression for the acceptance probability for the Reversible Jump step:

\[
p(\mathbf{x}|\mathcal{M}_k, \Theta_{-(\mathcal{M}_k, \phi)}) \propto \int [2\pi \sigma^2]^{-\frac{3}{2}} |\Sigma_\phi|^{-\frac{1}{2}} |C_\phi|^{\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \frac{\mathbf{x}^T \mathbf{x}}{\sigma^2} - \hat{\phi}^T C_\phi^{-1} \hat{\phi} \right) \right] \left[ \frac{\Lambda_k}{k!} \right]^{\tau_k} \delta_{\mathcal{K}}(k^*) \quad (A.2.9)
\]

In practice, the indicator function in (A.2.9) may be ignored as the proposal distribution for \( k \) will only allow proposals such that \( k^* \in \mathcal{K} \). Continuing our derivation, we obtain:

\[
p(\mathbf{x}|\mathcal{M}_k, \Theta_{-(\mathcal{M}_k, \phi)}) \propto \int [2\pi \sigma^2]^{-\frac{3}{2}} |\Sigma_\phi|^{-\frac{1}{2}} |C_\phi|^{\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \frac{\mathbf{x}^T \mathbf{x}}{\sigma^2} - \hat{\phi}^T C_\phi^{-1} \hat{\phi} \right) \right] \left[ \frac{\Lambda_k}{k!} \right]^{\tau_k} \delta_{\mathcal{K}}(k^*) \quad (A.2.10)
\]

In (A.2.10), \( \hat{\phi} \) and \( C_\phi \) are calculated using the formulas in (A.1.2). The superscript \( * \) in terms such as \( \hat{\phi}^* \) or \( C_\phi^* \) is used to indicate that these values are calculated, conditional on the candidate model order \( \mathcal{M}_{k^*} \). When the probability of each model order is proportional to one, that is, \( \tau_k = 0 \), the expression in (A.2.10) will simplify further.
A.2.2 Multiple-Try Metropolis Acceptance Probability

Substitution of the target distribution and proposal distribution into the simplified form of the MTM acceptance probability shown in (3.1.16) results in the following form for the acceptance probability:

\[
\frac{\sum_{j=1}^{n_{\text{MTM}}} \pi(\theta^*_j) q(\theta^{(i-1)}|\theta^*_j)}{\sum_{j=1}^{n_{\text{MTM}}} \pi(\psi^*_j) q(\theta^*_j|\psi^*_j)} = \frac{\sum_{j=1}^{n_{\text{MTM}}} p(\gamma^*_j, c_j^*|\mathbf{x}, M_k^{(1)}, \phi^{(1)}, \sigma^{2(1)}, d^{(1)}) q(\gamma^{(i-1)}|\gamma^*_j) q(c^{(i-1)}|c_j^*)}{\sum_{j=1}^{n_{\text{MTM}}} p(\gamma_j, c_j'|\mathbf{x}, M_k^{(1)}, \phi^{(1)}, \sigma^{2(1)}, d^{(1)}) q(\gamma^*_j|\gamma_j) q(c^*|c_j')}
\]

\[
= \left[ \sum_{j=1}^{n_{\text{MTM}}} (\gamma^*_j)^{\alpha_j - 1} \exp \left[ -\frac{1}{2\sigma^2} (x - Z_{i(j)}^* \phi)^T (x - Z_{i(j)}^* \phi) - \beta_j \gamma^*_j - \frac{(c^*_j - \mu_c)^2}{2\sigma_c^2} \right] \right] \times
\]

\[
\left[ \frac{(\gamma^*_j)^{\gamma_j^2}}{\Gamma(\gamma_j^2)} \gamma_j^{(i-1)\gamma_j^2 - 1} \exp \left[ -\frac{\gamma_j^2}{\Delta^*} \gamma_j^{(i-1)} \right] \frac{1}{\sqrt{2\pi\Delta^*_c}} \exp \left[ -\frac{1}{2\Delta^*_c} (c^{(i-1)} - c_j^*)^2 \right] \right]
\]

\[
= \left[ \sum_{j=1}^{n_{\text{MTM}}} (\gamma^*_j)^{\alpha_j - 1} \exp \left[ -\frac{1}{2\sigma^2} (x - Z_{i(j)}^* \phi)^T (x - Z_{i(j)}^* \phi) - \beta_j \gamma^*_j - \frac{(c^*_j - \mu_c)^2}{2\sigma_c^2} \right] \times
\]

\[
\left[ \frac{(\gamma_j)^\gamma}{\Gamma(\gamma^2)} \gamma^\gamma - 1 \exp \left[ -\frac{\gamma_j^2}{\Delta^*} \gamma \right] \frac{1}{\sqrt{2\pi\Delta^*_c}} \exp \left[ -\frac{1}{2\Delta^*_c} (c^* - c_j^*)^2 \right] \right]
\]

\[
= \left[ \sum_{j=1}^{n_{\text{MTM}}} (\gamma^*_j)^{\alpha_j - 1} \exp \left[ -\frac{1}{2\sigma^2} (x - Z_{i(j)}^* \phi)^T (x - Z_{i(j)}^* \phi) - \beta_j \gamma_j - \frac{(c^*_j - \mu_c)^2}{2\sigma_c^2} \right] \times
\]

\[
\left[ \frac{(\gamma_j)^\gamma}{\Gamma(\gamma^2)} \gamma^\gamma - 1 \exp \left[ -\frac{\gamma_j^2}{\Delta^*} \gamma \right] \frac{1}{\sqrt{2\pi\Delta^*_c}} \exp \left[ -\frac{1}{2\Delta^*_c} (c^* - c_j^*)^2 \right] \right]
\]

\[
= \left[ \sum_{j=1}^{n_{\text{MTM}}} (\gamma^*_j)^{\alpha_j - 1} \exp \left[ -\frac{1}{2\sigma^2} (x - Z_{i(j)}^* \phi)^T (x - Z_{i(j)}^* \phi) - \beta_j \gamma_j - \frac{(c^*_j - \mu_c)^2}{2\sigma_c^2} \right] \times
\]

\[
\left[ \frac{(\gamma_j)^\gamma}{\Gamma(\gamma^2)} \gamma^\gamma - 1 \exp \left[ -\frac{\gamma_j^2}{\Delta^*} \gamma \right] \frac{1}{\sqrt{2\pi\Delta^*_c}} \exp \left[ -\frac{1}{2\Delta^*_c} (c^* - c_j^*)^2 \right] \right].
\]
Appendix B

Univariate Generalised Autoregressive Conditional Heteroskedasticity Models

B.1 Acceptance Probabilities

Acceptance probabilities are required for the Reversible Jump and Metropolis-Hastings steps of the univariate GARCH posterior simulator. The calculation of these are presented in the following subsections.

B.1.1 Reversible Jump Acceptance Probability

The simplification of the acceptance probability for the RJ \( N \) Algorithm uses the quotient on the right hand side of the acceptance probability shown in (3.2.25) as follows:

\[
\frac{p(\alpha^*, N_l | x) q(\alpha | \alpha^*, N_l, N_{l^*})}{p(\alpha, N_l | x) q(\alpha^* | \alpha, N_l, N_{l^*})} = \frac{[2\pi]^{-\frac{d}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \epsilon^T H^{-1} \epsilon \right] \mathcal{I}_A(\alpha^*) [l^* + 1] \mathcal{I}_A(\alpha) [l + 1]^{-1}}{[2\pi]^{-\frac{d}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \epsilon^T H^{-1} \epsilon \right] \mathcal{I}_A(\alpha) [l + 1] \mathcal{I}_A(\alpha^*) [l^* + 1]^{-1}} \times \frac{\Gamma\left(\frac{n^* + 1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{\Gamma\left(\frac{n^* + 1 + m^*}{2}\right)}{\Gamma\left(\frac{n + 1 + m}{2}\right)} \frac{\left| \Sigma_{\alpha_{N_l}} \right|^{\frac{1}{2}} \left[ 1 + \frac{1}{2} (\alpha - \alpha_{N_l})^T \Sigma_{\alpha_{N_l}}^{-1} (\alpha - \alpha_{N_l}) \right]^{n^* + 1}}{\left| \Sigma_{\alpha_{N_{l^*}}} \right|^{\frac{1}{2}} \left[ 1 + \frac{1}{2} (\alpha^* - \alpha_{N_{l^*}})^T \Sigma_{\alpha_{N_{l^*}}}^{-1} (\alpha^* - \alpha_{N_{l^*}}) \right]^{n + 1}}}
\]
The acceptance probability for the Metropolis-Hastings step simplifies significantly when one notices:

\[ \frac{\Gamma\left(\frac{\nu+1+\nu\nu}{2}\right)}{\Gamma\left(\frac{(\nu+1+t^1)^2}{2}\right)} \times \]

\[ \frac{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}}{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ 1 \frac{1}{\nu} \left( \alpha - \tilde{\alpha}_{N_i} \right) \right]^{T} \left[ \sum_{\alpha N_i}^{-1} \right] \left( \alpha - \tilde{\alpha}_{N_i} \right) \]

\[ \frac{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}}{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ 1 \frac{1}{\nu} \left( \alpha - \tilde{\alpha}_{N_i} \right) \right]^{T} \left[ \sum_{\alpha N_i}^{-1} \right] \left( \alpha - \tilde{\alpha}_{N_i} \right) \]

\[ \frac{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}}{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ 1 \frac{1}{\nu} \left( \alpha - \tilde{\alpha}_{N_i} \right) \right]^{T} \left[ \sum_{\alpha N_i}^{-1} \right] \left( \alpha - \tilde{\alpha}_{N_i} \right) \]

\[ \frac{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}}{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ 1 \frac{1}{\nu} \left( \alpha - \tilde{\alpha}_{N_i} \right) \right]^{T} \left[ \sum_{\alpha N_i}^{-1} \right] \left( \alpha - \tilde{\alpha}_{N_i} \right) \]

\[ \frac{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}}{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ 1 \frac{1}{\nu} \left( \alpha - \tilde{\alpha}_{N_i} \right) \right]^{T} \left[ \sum_{\alpha N_i}^{-1} \right] \left( \alpha - \tilde{\alpha}_{N_i} \right) \]

The indicator function from the prior distribution for \( \alpha \) is removed from the acceptance probability in (B.1.1) as only candidate coefficient parameters that meet the stationarity conditions will be tested for acceptance. Therefore, the indicator functions will always take the value one when testing acceptance.

### B.1.2 Metropolis-Hastings Acceptance Probability

The acceptance probability for the Metropolis-Hastings step simplifies significantly when one notices the symmetry of the proposal distribution. Starting with the quotient on the right hand side of (3.2.30), we may calculate:

\[ \frac{p(\alpha^*|x, N_i)}{p(\alpha^{(t-1)}|x, N_i)} \frac{q(\alpha^{(t-1)}|\alpha^*, N_i)}{q(\alpha^*|\alpha^{(t-1)}, N_i)} \]

\[ \frac{[2\pi]^{-\frac{\nu^2}{2}} \left| H^* \right|^{-\frac{\nu}{2}} \exp \left[ -\frac{1}{2} \epsilon^T \left( H^* \right)^{-1} \epsilon \right] \| \sum_{\alpha N_i} \left( \alpha^* \right) \|_2 \left( \nu \nu \right)^{\frac{\nu}{2}}}{[2\pi]^{-\frac{\nu^2}{2}} \left| H \right|^{-\frac{\nu}{2}} \exp \left[ -\frac{1}{2} \epsilon^T \left( H^{-1} \right) \epsilon \right] \| \sum_{\alpha N_i} \left( \alpha \right) \|_2 \left( \nu \nu \right)^{\frac{\nu}{2}}} \]

\[ \frac{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}}{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ 1 \frac{1}{\nu} \left( \alpha - \tilde{\alpha}_{N_i} \right) \right]^{T} \left[ \sum_{\alpha N_i}^{-1} \right] \left( \alpha - \tilde{\alpha}_{N_i} \right) \]

\[ \frac{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}}{\nu^{\frac{\nu+\nu^2}{2}} \pi^{\frac{\nu+\nu^2}{2}}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ \sum_{\alpha N_i} \right]^{\frac{\nu+\nu^2}{2}} \left[ 1 \frac{1}{\nu} \left( \alpha - \tilde{\alpha}_{N_i} \right) \right]^{T} \left[ \sum_{\alpha N_i}^{-1} \right] \left( \alpha - \tilde{\alpha}_{N_i} \right) \]
Section B.1: Acceptance Probabilities

The ratio of the proposal distributions in (B.1.2) is equal to one. Continuing yields:

$$\frac{p(\alpha^*|x, n_{t})}{p(\alpha^{(i-1)}|x, n_{t})} \cdot \frac{q(\alpha^{(i-1)}|\alpha^*, n_{t})}{q(\alpha^*|\alpha^{(i-1)}, n_{t})} = \frac{|H^*|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \epsilon^T H^{-1} \epsilon \right]}{|H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \epsilon^T H^{-1} \epsilon \right]}$$

$$= \left[ |H^*| |H|^{-1} \right]^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \epsilon^T H^{-1} \epsilon + \frac{1}{2} \epsilon^T H^{-1} \epsilon \right]$$

$$= \left[ |H^*| |H|^{-1} \right]^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \epsilon^T H^{-1} \epsilon - \epsilon^T H^{-1} \epsilon \right) \right]$$

$$= \left[ |H^*| |H|^{-1} \right]^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \epsilon^T \left( H^{-1} - H^{-1} \right) \epsilon \right].$$

The expression in (B.1.3) is the major calculation required for the acceptance probability. It is essentially the ratio of the likelihood functions.
Appendix C

Univariate STAR-GARCH Models

C.1 Posterior Distribution Derivations

For the STAR-GARCH MCMC algorithm, several conditional posterior distributions are required. They do not require significant manipulation as they are all of non-standard form. The conditional posterior distribution for \( \phi \) provides some information as to a suitable proposal distribution for that parameter, and the derivation is given in the following section.

C.1.1 Conditional Posterior for \( \phi \)

An investigation into the conditional posterior distribution for \( \phi \) is useful for determining a suitable proposal distribution for \( \phi \). Starting with the full conditional posterior distribution for \( \phi \) in (3.3.39), we calculate:

\[
p(\phi | x, \Theta_{-\phi}) \propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2}(x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right] \times \\
\quad [2\pi]^{-\frac{k+1}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \phi^T \Sigma_\phi^{-1} \phi \right] \\
\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} [2\pi]^{-\frac{k+1}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \times \\
\exp \left[ -\frac{1}{2}(x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) - \frac{1}{2} \phi^T \Sigma_\phi^{-1} \phi \right] \\
\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} [2\pi]^{-\frac{k+11}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \times \\
\exp \left[ -\frac{1}{2} \left( x^T - \phi^T Z_1^T \right) H^{-1} (x - Z_1 \phi) - \frac{1}{2} \phi^T \Sigma_\phi^{-1} \phi \right] \\
\propto [2\pi]^{-\frac{n}{2}} |H|^{-\frac{1}{2}} [2\pi]^{-\frac{k+1}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \times \\
\exp \left[ -\frac{1}{2} \left( x^T H^{-1} - \phi^T Z_1^T H^{-1} \right) (x - Z_1 \phi) + \phi^T \Sigma_\phi^{-1} \phi \right]
\]
Continuing with the derivation, we now have

\[
\phi \propto \exp \left[ -\frac{1}{2} J^\top H^{-1} J + \phi^T \left( Z_1^T H^{-1} Z_1 \phi + \Sigma^{-1}_\phi \phi - Z_1^T H^{-1} J \right) - x^T H^{-1} J \phi \right].
\]

To simplify the derivation of the conditional posterior distribution for \( \phi \), let \( C_\phi^{-1} = Z_1^T H^{-1} Z_1 + \Sigma^{-1}_\phi \).

Continuing with the derivation, we now have

\[
p(\phi | x, \Theta_\phi) \propto [2\pi]^{-\frac{\tilde{d}}{2}} |H|^{-\frac{1}{2}} [2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \times 
\exp \left[ -\frac{1}{2} J^\top H^{-1} J + \phi^T \left( C_\phi^{-1} \phi - Z_1^T H^{-1} J \phi \right) - x^T H^{-1} J \phi \right].
\]

To further simplify the above expression, let \( \hat{\phi} = C_\phi Z_1^T H^{-1} x \). Then

\[
p(\phi | x, \Theta_\phi) \propto [2\pi]^{-\frac{\tilde{d}}{2}} |H|^{-\frac{1}{2}} [2\pi]^{-\frac{2(k+1)}{2}} |\Sigma_\phi|^{-\frac{1}{2}} \times 
\exp \left[ -\frac{1}{2} J^\top H^{-1} J + \phi^T \left( C_\phi^{-1} \phi - Z_1^T H^{-1} J \phi \right) - x^T H^{-1} J \phi \right].
\]
The proposal distribution for \( \phi \) is shown in (3.3.43), and the ratio is simplified below:

\[
\exp \left[ -\frac{1}{2} \left( x^T H^{-1} x - \hat{\phi}^T C_{\phi}^{-1} \hat{\phi} + \phi^T C_{\phi}^{-1} \phi \right) \right]
\]

This is a very similar distribution to the one found for the STAR model. Assuming that the covariance matrix \( H \) is diagonal with equal values on the diagonal, this is a good candidate for the proposal distribution. This is explained in Section 3.3.4.

### C.2 Acceptance Probabilities

In the following sections, we derive the required acceptance probabilities for the univariate STAR-GARCH model.

#### C.2.1 Reversible Jump Acceptance Probability - RJ\(_M\)

The acceptance probability calculation is performed using the quotient on the right hand side of the acceptance probability shown in (3.3.41). This expression is rather large, so to aid in simplification, the expression will first be split up into the ratio of the target distributions and the ratio of the proposal distributions.

The proposal distribution for \( \phi \) is shown in (3.3.43), and the ratio is simplified below:

\[
\frac{q(\phi^* | \phi, \mathcal{M}_k, \mathcal{M}_{k^*})}{q(\phi^* | \phi, \mathcal{M}_k, \mathcal{M}_{k^*})} = \frac{[2\pi]^{-2(k+1)/2} |\Delta_C C_{\phi}^*|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left( \phi^* - \hat{\phi} \right)^T \Delta_C C_{\phi}^* \left( \phi^* - \hat{\phi} \right) \right]}{[2\pi]^{-2(k+1)/2} |\Delta_C C_{\phi}^*|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left( \phi - \hat{\phi} \right)^T \Delta_C C_{\phi}^* \left( \phi - \hat{\phi} \right) \right]}
\]

\[
= \frac{|\Delta_C C_{\phi}^*|^{-\frac{1}{2}} |\Delta_C C_{\phi}^*|^{-\frac{1}{2}} \times \left( \phi^* - \hat{\phi} \right)^T \Delta_C C_{\phi}^* \left( \phi^* - \hat{\phi} \right)}{\left( \phi - \hat{\phi} \right)^T \Delta_C C_{\phi}^* \left( \phi - \hat{\phi} \right)}
\]

\[
= \Delta_C C_{\phi}^* \times \frac{1}{\Delta_C C_{\phi}^*} \times \left( \phi^* - \hat{\phi} \right)^T \Delta_C C_{\phi}^* \left( \phi^* - \hat{\phi} \right) \]

(C.2.1)
Section C.2: Acceptance Probabilities

\[
\exp \left[ -\frac{1}{2} (\phi - \tilde{\phi})^T \left[ \Delta_\phi \tilde{C}_\phi \right]^{-1} (\phi - \tilde{\phi}) + \frac{1}{2} (\phi^* - \tilde{\phi})^T \left[ \Delta_\phi \tilde{C}_\phi^* \right]^{-1} (\phi^* - \tilde{\phi}) \right].
\]

Now, looking at the ratio of the joint conditional posterior distributions for $M_k$ and $\phi$, we choose the distribution shown in (3.3.39). The simplification of the ratio of the target distributions is as follows:

\[
\frac{\pi (M_k^*, \phi^* | x, \Theta_{-(M_k, \phi)})}{\pi (M_k, \phi | x, \Theta_{-(M_k, \phi)})} = \frac{\left[2\pi\right]^{-\frac{1}{2} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1^* \phi^*)^T H^{-1} (x - Z_1^* \phi^*) \right]} \times \frac{\left[2\pi\right]^{-\frac{1}{2} |\Sigma_{\phi^*}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\phi^*)^T \Sigma_{\phi^*}^{-1} \phi^* \right]}{\left[2\pi\right]^{-\frac{1}{2} |\Sigma_{\phi^*}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\phi^*)^T \Sigma_{\phi^*}^{-1} \phi^* \right]} \times \left[ \Lambda^{k^* |1|}_{k^*+1} \right]^{T_k}} \times \exp \left[ -\frac{1}{2} (x - Z_1^* \phi^*)^T H^{-1} (x - Z_1^* \phi^*) \right]
\]

(C.2.2)

Combining the ratio of the proposal distributions in (C.2.1) with the ratio of the target distributions in (C.2.2) gives the formula for the acceptance probability. We calculate

\[
\frac{\pi (M_k^*, \phi^* | x, \Theta_{-(M_k, \phi)}) \cdot q(\phi^* | M_k, M_k^*)}{\pi (M_k, \phi | x, \Theta_{-(M_k, \phi)}) \cdot q(\phi^* | M_k, M_k^*)} = \frac{\left[2\pi\right]^{-\frac{1}{2} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1^* \phi^*)^T H^{-1} (x - Z_1^* \phi^*) \right]} \times \frac{\left[2\pi\right]^{-\frac{1}{2} |\Sigma_{\phi^*}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\phi^*)^T \Sigma_{\phi^*}^{-1} \phi^* \right]}{\left[2\pi\right]^{-\frac{1}{2} |\Sigma_{\phi^*}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\phi^*)^T \Sigma_{\phi^*}^{-1} \phi^* \right]} \times \left[ \Lambda^{k^* |1|}_{k^*+1} \right]^{T_k}} \times \exp \left[ -\frac{1}{2} (x - Z_1^* \phi^*)^T H^{-1} (x - Z_1^* \phi^*) \right]
\]

233
\[
\exp \left[ \frac{1}{2}(x - Z_1^* \phi)^T \Sigma^{-1}_z (x - Z_1^* \phi) \right]
\times
\exp \left[ \frac{1}{2}(x - Z_1^* \phi)^T \Phi^{-1} (x - Z_1^* \phi) + \frac{1}{2} \phi^T \Phi^{-1} \phi \right] 
\times
\exp \left[ -\frac{1}{2} \left( \phi - \hat{\phi} \right)^T \Sigma^{-1}_z \left( \phi - \hat{\phi} \right) + \frac{1}{2} \left( \phi^* - \hat{\phi} \right)^T \Sigma^{-1}_z \left( \phi^* - \hat{\phi} \right) \right] 
\]

(C.2.3)

The acceptance probability in (C.2.3) is rather lengthy and one may argue that its expression is more complicated than the original. That may be the case, but the reason for expressing it in this form is that it makes taking the logarithm of the expression easier, and the subsequent simplification more straightforward. Without combining the exponentials and reordering the calculation of certain elements of the acceptance probability, computational issues would arise in this calculation. This can be avoided by cancelling certain terms before the computer software rounds their values to infinity or zero.

Considerable care has been taken in the coding of the algorithm to ensure that these computational issues are avoided.

C.2.2 Reversible Jump Acceptance Probability - RJ\(\lambda\)

The substitution and simplification of the calculation required to determine if a jump is to be made from the current model index to a new candidate model index for the conditional variance is detailed below:

\[
\frac{\pi(N_{j^*}, \alpha^* | x, \Theta_{\lambda(N_{j^*}, \alpha)}) \pi(\alpha | \alpha^*, N_{j^*}, N_{j^*})}{\pi(N_{j^*}, \alpha | x, \Theta_{-(N_{j^*}, \alpha)}) \pi(\alpha^* | \alpha, N_{j^*}, N_{j^*})} 
= \frac{[2\pi]^{-\frac{1}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right]}{[2\pi]^{-\frac{1}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - Z_1 \phi)^T H^{-1} (x - Z_1 \phi) \right]} \times \frac{\gamma_{\lambda}^{\alpha^*}(\alpha^*)}{\gamma_{\lambda}\alpha(\alpha)} 
\]
Section C.2: Acceptance Probabilities

The acceptance probability for the proposal distributions is equal to one as these distributions are symmetric, and therefore the

\[
\hat{\alpha}_{C.2.4} \text{ Metropolis-Hastings Acceptance Probability - MH } \\
\]

For consistency with the rest of this thesis, the Reversible Jump acceptance probability calculations are made, conditional on the first \( n^* = N - s^* \) data points, where \( s^* = \max\{k, d, l, m, l^*, m^*\} \).

C.2.3 Metropolis-Hastings Acceptance Probability - MH\(_\phi\)

We begin our calculation of the Metropolis-Hastings acceptance probability for the full conditional posterior distribution for \( \phi \) by simplifying the quotient on the right hand side of (3.3.49):

\[
\frac{p(\phi^*|x, \Theta_{-\phi})}{p(\phi|x, \Theta_{-\phi})} = \frac{2\pi^{-\frac{n}{2}}|H^*|^{-\frac{1}{2}} \exp\left[\frac{1}{2}(x - Z_1\phi^*)^T H^*^{-1} (x - Z_1\phi^*)\right]}{2\pi^{-\frac{n}{2}}|H|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x - Z_1\phi)^T H^{-1} (x - Z_1\phi)\right]} \times \\
\frac{2\pi^{-\frac{2k+1}{2}}|\Sigma_\phi|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\phi^T \Sigma^{-1}_\phi \phi^*\right]}{2\pi^{-\frac{2k+1}{2}}|\Sigma^*_\phi|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\phi^* T \Sigma^{-1}_\phi \phi^*\right]} = \\
|H^*|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x - Z_1\phi^*)^T H^*^{-1} (x - Z_1\phi^*)\right] \exp\left[-\frac{1}{2}\phi^* T \Sigma^{-1}_\phi \phi^*\right] \\
|H|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x - Z_1\phi)^T H^{-1} (x - Z_1\phi)\right] \exp\left[-\frac{1}{2}\phi^T \Sigma^{-1}_\phi \phi\right] = \\
|H^*|^{-\frac{1}{2}} |H|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x - Z_1\phi^*)^T H^*^{-1} (x - Z_1\phi^*) - \frac{1}{2}\phi^* T \Sigma^{-1}_\phi \phi^* - (x - Z_1\phi)^T H^{-1} (x - Z_1\phi) - \phi^T \Sigma^{-1}_\phi \phi\right] \\
= \left[|H^*| |H|^{-1}\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x - Z_1\phi^*)^T H^*^{-1} (x - Z_1\phi^*) + \phi^* T \Sigma^{-1}_\phi \phi^* - (x - Z_1\phi)^T H^{-1} (x - Z_1\phi) - \phi^T \Sigma^{-1}_\phi \phi\right].
\]

C.2.4 Metropolis-Hastings Acceptance Probability - MH\(_\alpha\)

Below is the simplification of the calculation required to determine the acceptance probability for the Metropolis-Hastings step used in simulating the conditional variance coefficient parameter \( \alpha \). The ratio of the proposal distributions is equal to one as these distributions are symmetric, and therefore the
Focusing on the quotient on the right hand side of (C.2.5), we calculate:

\[
p\big(\alpha^*|x, \Theta_{-(\alpha)}\big) = \frac{\left[2\pi\right]^{-\frac{n}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - \mathbf{Z}_1 \phi)^T \mathbf{H}^{-1} (x - \mathbf{Z}_1 \phi) \right]}{p\big(\alpha|x, \Theta_{-(\alpha)}\big) \left[2\pi\right]^{-\frac{n}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - \mathbf{Z}_1 \phi)^T \mathbf{H}^{-1} (x - \mathbf{Z}_1 \phi) \right] I_d(\alpha^*)} = \left[\mathbf{H}^*|\mathbf{H}^{-1}\right]^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( (x - \mathbf{Z}_1 \phi)^T \mathbf{H}^{-1} (x - \mathbf{Z}_1 \phi) \right) - \frac{1}{2} \left( (x - \mathbf{Z}_1 \phi)^T (\mathbf{H}^{-1} - \mathbf{H}^{-1}) (x - \mathbf{Z}_1 \phi) \right) \right].
\]

(C.2.4)

The first \(s^* = \max(k,d,l,m)\) data points are used as initial conditions for the calculation of the acceptance probability in (C.2.4). This makes the factorisation in the exponent possible as the matrices \(\mathbf{H}^*\) and \(\mathbf{H}\) will have the same dimensions.

C.2.5 Metropolis-Hastings Acceptance Probability - MH\(_{(\gamma,c)}\)

Starting with the general form of the Metropolis-Hastings acceptance probability in (3.3.49) and applying the terms relevant for the smoothing and location parameters of the transition function, we obtain

\[
1_{(g,\delta,g)} = \min\left( 1, \frac{p\big(\gamma^*, c^*|x, \Theta_{-(\gamma,c)}\big) q(\gamma|\gamma^*) \left( c|c^* \right)}{p\big(\gamma|c,x, \Theta_{-(\gamma,c)}\big) q(\gamma^*|\gamma) \left( c^*|c \right)} \right) = \min\left( 1, \frac{p\big(\gamma^*, c^*|x, \Theta_{-(\gamma,c)}\big) q(\gamma|\gamma^*)}{p\big(\gamma|c,x, \Theta_{-(\gamma,c)}\big) q(\gamma^*|\gamma)} \right).
\]

(C.2.5)

Focusing on the quotient on the right hand side of (C.2.5), we calculate:

\[
p\big(\gamma^*, c^*|x, \Theta_{-(\gamma,c)}\big) q(\gamma|\gamma^*) \frac{\beta_\gamma}{\Gamma(\alpha_\gamma)} \exp \left[ -\beta_\gamma \gamma^* \right] \frac{\beta_c}{\Gamma(\alpha_c)} \exp \left[ -\beta_c c^* \right] \times \frac{\left[2\pi\right]^{-\frac{n}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - \mathbf{Z}_1 \phi)^T \mathbf{H}^{-1} (x - \mathbf{Z}_1 \phi) \right]}{\left[2\pi\right]^{-\frac{n}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - \mathbf{Z}_1 \phi)^T \mathbf{H}^{-1} (x - \mathbf{Z}_1 \phi) \right]} \\
\times \frac{\gamma^*^{\alpha_\gamma - 1} \exp \left[ -\beta_\gamma \gamma^* \right] \left[2\pi\sigma^2\right]^{-\frac{1}{2}} \exp \left[ -\frac{(c - \mu)^2}{2\sigma^2} \right]}{\gamma^{\alpha_\gamma - 1} \exp \left[ -\beta_\gamma \gamma \right] \left[2\pi\sigma^2\right]^{-\frac{1}{2}} \exp \left[ -\frac{(c - \mu)^2}{2\sigma^2} \right]} \\
\times \frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]}
\]

\[
= \frac{\left[2\pi\right]^{-\frac{n}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - \mathbf{Z}_1 \phi)^T \mathbf{H}^{-1} (x - \mathbf{Z}_1 \phi) \right]}{\left[2\pi\right]^{-\frac{n}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - \mathbf{Z}_1 \phi)^T \mathbf{H}^{-1} (x - \mathbf{Z}_1 \phi) \right]} \\
\times \frac{\gamma^*^{\alpha_\gamma - 1} \exp \left[ -\beta_\gamma \gamma^* \right] \left[2\pi\sigma^2\right]^{-\frac{1}{2}} \exp \left[ -\frac{(c - \mu)^2}{2\sigma^2} \right]}{\gamma^{\alpha_\gamma - 1} \exp \left[ -\beta_\gamma \gamma \right] \left[2\pi\sigma^2\right]^{-\frac{1}{2}} \exp \left[ -\frac{(c - \mu)^2}{2\sigma^2} \right]} \\
\times \frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]}
\]

\[
\frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]} = \frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]} = \frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]}
\]

\[
\frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]} = \frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]}
\]

\[
\frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]} = \frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]}
\]

\[
\frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]} = \frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]}
\]

\[
\frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]} = \frac{\gamma^*^{\alpha_c - 1} \exp \left[ -\beta_c c^* \right]}{\gamma^{\alpha_c - 1} \exp \left[ -\beta_c c \right]}
\]
Therefore, the acceptance probability is

\[
\begin{align*}
    r_{(\phi,\betaG)} &= \min \left( 1, \left[ |H^*||H|^{-\frac{1}{2}} \right] \left[ (\gamma^* - \gamma)^{-\frac{1}{2}} \right] \left[ \Gamma \left( \frac{\gamma^*}{\Delta \gamma} \right) \right]^{-1} \left[ \frac{\gamma^*}{\Delta \gamma} \right]^{-\frac{1}{2}} \left[ \gamma \right]^{-\frac{1}{2}} \times \gamma \times \gamma^* - \frac{1}{2} \exp \left[ -\frac{1}{2} \right] \left( x - Z_1^* \phi \right)^T \left( H^* \right)^{-1} \left( x - Z_1^* \phi \right) - \frac{1}{2} \left( \left( x - Z_1^* \phi \right)^T H^* \right)^{-1} \left( x - Z_1^* \phi \right) \right] \right),
\end{align*}
\]
Appendix D

VAR Models

D.1 Posterior Distribution Derivation

We simplify the exponent for the likelihood function after setting $\hat{\Phi} = C_\phi Z_i^T X$:

$$\text{tr} \left( \left[ X^T X + \Phi^T C_\phi^{-1} (\Phi - C_\phi Z_i^T X) - X^T Z_i \Phi \right] \Sigma^{-1} \right)$$

$$= \text{tr} \left( \left[ X^T X + \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) \right] \Sigma^{-1} - \left[ X^T Z_i \Phi \right] \Sigma^{-1} \right)$$

$$= \text{tr} \left( \left[ X^T X + \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) \right] \Sigma^{-1} \right) - \text{tr} \left( \left[ X^T Z_i \Phi \right] \Sigma^{-1} \right)$$

$$= \text{tr} \left( \left[ X^T X + \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) \right] \Sigma^{-1} \right) - \text{tr} \left( \left[ X^T Z_i \Phi \right] \Sigma^{-1} \right)$$

$$= \text{tr} \left( \left[ X^T X + \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) \right] \Sigma^{-1} \right) - \text{tr} \left( \left[ X^T Z_i \Phi \right] \Sigma^{-1} \right)$$

$$= \text{tr} \left( \left[ X^T X + \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) \right] \Sigma^{-1} \right) - \text{tr} \left( \left[ X^T Z_i \Phi \right] \Sigma^{-1} \right)$$

$$= \text{tr} \left( \left[ X^T X + \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) \right] \Sigma^{-1} \right) - \text{tr} \left( \left[ X^T Z_i \Phi \right] \Sigma^{-1} \right)$$

$$= \text{tr} \left( \left[ X^T X + \Phi^T C_\phi^{-1} (\Phi - \hat{\Phi}) \right] \Sigma^{-1} \right) - \text{tr} \left( \left[ X^T Z_i \Phi \right] \Sigma^{-1} \right)$$
\[
\text{tr} \left( \Phi^T C_\Phi^{-1} \Phi \right) \Sigma^{-1} \\
= \text{tr} \left( X^T X + \Phi^T C_\Phi^{-1} \left( \Phi - \hat{\Phi} \right) \right) \Sigma^{-1} - \text{tr} \left( \Sigma^{-1} \hat{\Phi}^T C_\Phi^{-1} \left( \Phi - \hat{\Phi} \right) \right) - \\
\text{tr} \left( \Phi^T C_\Phi^{-1} \Phi \right) \Sigma^{-1} \\
= \text{tr} \left( X^T X + \Phi^T C_\Phi^{-1} \left( \Phi - \hat{\Phi} \right) \right) \Sigma^{-1} - \text{tr} \left( \Phi^T C_\Phi^{-1} \left( \Phi - \hat{\Phi} \right) \Sigma^{-1} \right) - \\
\text{tr} \left( \Phi^T C_\Phi^{-1} \Phi \right) \Sigma^{-1} \\
= \text{tr} \left( X^T X + \Phi^T C_\Phi^{-1} \left( \Phi - \hat{\Phi} \right) \Phi - \hat{\Phi} C_\Phi^{-1} \Phi \left( \Phi - \hat{\Phi} \right) \right) \Sigma^{-1} \\
= \text{tr} \left( X^T X + \left( \Phi^T - \hat{\Phi}^T \right) C_\Phi^{-1} \left( \Phi - \hat{\Phi} \right) - \hat{\Phi} C_\Phi^{-1} \Phi \Sigma^{-1} \right) \\
= \text{tr} \left( \left( \Phi - \hat{\Phi} \right)^T C_\Phi^{-1} \left( \Phi - \hat{\Phi} \right) + X^T X - \hat{\Phi} C_\Phi^{-1} \Phi \Sigma^{-1} \right) \\
= \text{tr} \left( \left( \Phi - \hat{\Phi} \right)^T C_\Phi^{-1} \left( \Phi - \hat{\Phi} \right) \Sigma^{-1} \right) + \text{tr} \left( X^T X - \hat{\Phi} C_\Phi^{-1} \Phi \Sigma^{-1} \right).
\]
Appendix E

Multivariate GARCH Models

E.1 Proposal Distribution of the Coefficient Matrix

Candidate coefficient matrices will be proposed from a single Matrix Normal distribution. Under the univariate setting, the parameters \( \alpha_i \) and \( \beta_j \), \( i \in \{1, \ldots, l\} \) and \( j \in \{1, \ldots, m\} \), are all combined to form the coefficient vector \( \mathbf{\alpha} \). The full vector is then proposed from a single multivariate normal distribution.

The same thinking is applied here for the general case. We seek to combine the parameter matrices \( A_i \) and \( B_j \), for \( i \in \{1, \ldots, l\} \) and \( j \in \{1, \ldots, m\} \), into the complete coefficient parameter matrix \( \mathbf{A} = [A_1, \ldots, A_l, B_1, \ldots, B_m] \).

The general form of a Matrix Normal distribution is

\[
p(X|M,U,V) = \frac{\exp\left[ -\frac{1}{2} \text{tr} \left( V^{-1} (X - M)^T U^{-1} (X - M) \right) \right]}{(2\pi)^{mn} |V|^{\frac{m}{2}} |U|^{\frac{n}{2}}},
\]

where \( M \in \mathbb{R}^{m \times n} \) is the location matrix, and \( U \in \mathbb{R}^{m \times m} \) and \( V \in \mathbb{R}^{n \times n} \) are the among row and among column covariance matrices, respectively.

Suppose we have \( r \) matrix normal variables combined into a larger matrix, where \( r = l + m \). The variables will be denoted by \( X_i \), with mean and covariance parameters \( M_i, U_i \) and \( V_i \), respectively for \( i = \{1, \ldots, r\} \).

If we stack the variables from left to right so that \( \mathbf{X} = [X_1, \ldots, X_r] \), the mean matrix will then be \( \mathbf{M} = [M_1, \ldots, M_r] \). The matrices have been stacked left to right rather than top to bottom so that if the vec(\( \Theta \)) operator is used, the first portion of that vector corresponds to the first matrix and the
Section E.1: Proposal Distribution of the Coefficient Matrix

Now, focusing on the expression in the exponent on (E.1.2), we have

\[
p(X|M, U, V_1, \ldots, V_r) = p(X_1|M_1, U, V_1) p(X_2|M_2, U, V_2) \ldots p(X_r|M_r, U, V_r)
\]

\[
= \exp \left[ -\frac{1}{2} \text{tr} \left( V_1^{-1} (X_1 - M_1)^T U^{-1} (X_1 - M_1) \right) \right] \times \frac{1}{(2\pi)^{\frac{m^2}{2}} |V_1|^{\frac{m}{2}} |U|^{\frac{m}{2}}}
\]

\[
\times \exp \left[ -\frac{1}{2} \text{tr} \left( V_2^{-1} (X_2 - M_2)^T U^{-1} (X_2 - M_2) \right) \right] \times \ldots \times \frac{1}{(2\pi)^{\frac{m^2}{2}} |V_1|^{\frac{m}{2}} |U|^{\frac{m}{2}}}
\]

\[
\times \exp \left[ -\frac{1}{2} \text{tr} \left( V_r^{-1} (X_r - M_r)^T U^{-1} (X_r - M_r) \right) \right] \times \ldots
\]

\[
= \exp \left[ -\frac{1}{2} \text{tr} \left( V_1^{-1} E_1^T U^{-1} E_1 \right) \right] \exp \left[ -\frac{1}{2} \text{tr} \left( V_2^{-1} E_2^T U^{-1} E_2 \right) \right] \times \ldots \times \frac{1}{(2\pi)^{\frac{m^2}{2}} |V_1|^{\frac{m}{2}} |U|^{\frac{m}{2}}}
\]

\[
= \exp \left[ -\frac{1}{2} \text{tr} \left( \sum_{i=1}^{r} E_i^T U^{-1} E_i \right) \right] \times \frac{1}{(2\pi)^{\frac{m^2}{2}} |V_1|^{\frac{m}{2}} |U|^{\frac{m}{2}}}
\]

\[
\times \exp \left[ -\frac{1}{2} \text{tr} \left( \sum_{i=1}^{r} E_i^T U^{-1} E_i \right) \right] \times \ldots \times \frac{1}{(2\pi)^{\frac{m^2}{2}} |V_1|^{\frac{m}{2}} |U|^{\frac{m}{2}}}
\]

\[
= \exp \left[ -\frac{1}{2} \text{vec} (E_1)^T \text{vec} \left( U^{-1} E_1 V_1^{-1} \right) + \ldots + \text{vec} (E_r)^T \text{vec} \left( U^{-1} E_r V_r^{-1} \right) \right]
\]

\[
= \exp \left[ -\frac{1}{2} \text{vec} (E_1)^T \text{vec} \left( U^{-1} E_1 V_1^{-1} \right) \right] \times \ldots \times \frac{1}{(2\pi)^{\frac{m^2}{2}} |V_1 \otimes U|^{\frac{m}{2}}}
\]

\[
= \exp \left[ -\frac{1}{2} \text{vec} (E_1)^T \text{vec} \left( U^{-1} E_1 V_1^{-1} \right) \right] \times \ldots \times \frac{1}{(2\pi)^{\frac{m^2}{2}} |V_1 \otimes U|^{\frac{m}{2}}}
\]

\[
= \exp \left[ -\frac{1}{2} \text{vec} (E_1)^T \left( V_1 \otimes U \right)^{-1} \text{vec} (E_1) + \ldots + \text{vec} (E_r)^T \left( V_r \otimes U \right)^{-1} \text{vec} (E_r) \right]
\]

Now, focusing on the expression in the exponent on (E.1.2), we have

\[
\text{vec} (E_1)^T \left( V_1 \otimes U \right)^{-1} \text{vec} (E_1) + \ldots + \text{vec} (E_r)^T \left( V_r \otimes U \right)^{-1} \text{vec} (E_r)
\]
\[
\begin{bmatrix}
(V_1 \otimes U)^{-1} & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & (V_r \otimes U)^{-1}
\end{bmatrix}
\begin{bmatrix}
\vec{(E_1)} \\
\vdots \\
\vec{(E_r)}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
(V_1 \otimes U)^{-1} & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & (V_r \otimes U)^{-1}
\end{bmatrix}
\begin{bmatrix}
\vec{(E)}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
(V_1 \otimes U)^{-1} & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & (V_r \otimes U)
\end{bmatrix}^{-1}
\begin{bmatrix}
\vec{(E)}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
(V \otimes U)^{-1} \vec{(E)}
\end{bmatrix}
\]

where \( V \) is defined to be a block diagonal matrix made up of the individual among column covariance matrices and has the form:

\[
V = \begin{bmatrix}
V_1 & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & V_r
\end{bmatrix}
\]

Therefore, substituting the result in (E.1.3) back into the expression for the joint density in (E.1.2), we obtain

\[
p(X|\mathbf{M}, U, V) = \exp \left[ -\frac{1}{2} \begin{bmatrix}
\vec{(E)}^T
\end{bmatrix} \begin{bmatrix}
V \otimes U
\end{bmatrix}^{-1} \begin{bmatrix}
\vec{(E)}
\end{bmatrix}
\right]
\]

The expression in (E.1.4) is in the form of a matrix normal distribution with location matrix \( \mathbf{M} \in \mathbb{R}^{m \times r} \), and covariance matrices \( U \in \mathbb{R}^{m \times m} \) and \( V \in \mathbb{R}^{r \times r} \) which are the among row and among column covariance matrices, respectively.

For the M-GARCH model, we have \( r = l + m \), and the dimension of each individual parameter matrix is \( p \times p \), where \( p \) is the number of series in the data set. Therefore, the general form of the proposal
distribution will be:

\[
p(A|M, U, V) = \frac{\exp \left[ -\frac{1}{2} \text{tr} \left( V^{-1} (A - M)^T U^{-1} (A - M) \right) \right]}{(2\pi)^{(l+m)p+q} \left| V^\frac{1}{2} \right| \left| U^{(l+m)p} \right|}.
\]  
(E.1.5)

In practice, when including model orders greater than 1 in the search, the mean matrix will be critical to the success of the algorithm. As in the univariate case, a pilot run will be required in the multivariate setting to come up with reasonable location matrices, so that when running the full RJMCMC algorithm, proposals of the coefficient matrix are made in areas of reasonable posterior probability. Separate location matrices are required for each of the models included in the search.

### E.2 RJN - Acceptance Probabilities

The Reversible Jump step involves the movement from the current model index to the candidate model index. A total of three model indices for the M-GARCH model result in six forms that the acceptance probability can take. These acceptance probabilities are outlined in the following subsections.

#### E.2.1 RJN - BEKK(l, m) to BEKK(l*, m*)

The most complicated within model move acceptance probability is one that involves a change in either or both model orders of the BEKK model. Starting with the acceptance probability in (4.3.56) and using the full posterior distribution shown in (4.3.53), together with the proposal distributions shown in (4.3.57), the acceptance probability is determined as follows:

\[
\frac{p(N_j, A^*, C^* | X) q(A, C | A^*, C^*, N_j, N_j^*)}{p(N_j, A, C | X) q(A^*, C^* | A, C, N_j, N_j^*)} = \frac{p(N_j, A^*, C^* | X) q(A|A^*, N_j, N_j^*) q(C|C^*, N_j, N_j^*)}{p(N_j, A, C | X) q(A^*|A, N_j, N_j^*) q(C^*|C, N_j, N_j^*)} \\
= \frac{(2\pi)^{-\frac{n+2}{2}} |H^*|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left[ H^*^{-1} \right] \left[ \text{vec} (X^T) \right] \right]}{(2\pi)^{-\frac{n+2}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec} (X^T) \right] \right]} \frac{|\Lambda(A^*)|}{|\Lambda(A)|} \\
\times \frac{\exp \left[ -\frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A - M_{N_j})^T U_{N_j}^{-1} (A - M_{N_j}) \right) \right]}{(2\pi)^{\frac{n+1}{2}p} |V_{N_j}|^{-\frac{1}{2}} |U_{N_j}|^{-\frac{1}{2}} |U_{N_j}|^{-\frac{1}{2}}} \\
\times \frac{\exp \left[ -\frac{1}{2} \text{tr} \left( V_{N_j^*}^{-1} (A^* - M_{N_j^*})^T U_{N_j^*}^{-1} (A^* - M_{N_j^*}) \right) \right]}{(2\pi)^{\frac{n+1}{2}p} |V_{N_j^*}|^{-\frac{1}{2}} |U_{N_j^*}|^{-\frac{1}{2}} |U_{N_j^*}|^{-\frac{1}{2}}} \\
= \frac{j^{2N} j_{m}^{m} |H^*|^{-\frac{1}{2}} |H|^{\frac{1}{2}} \left[ 2 \pi \right]^{\frac{n+2}{2} (l^* + m^* - l - m)} |U_{N_j^*}|^{-\frac{1}{2}} |U_{N_j}|^{-\frac{1}{2}}}
When moving from a constant conditional covariance model to a DBEKK model, the relevant proposal

\[
q(A^*|A, N_j, \mathcal{N}_j) = 1
\]  

(Constant)
Therefore, the major calculation for the acceptance probability for such a move is

$$\begin{align*}
p(N_j^*, A^*, C^*|X) q(A, C|A^*, C^*, N_j, N_j^*) \\
p(N_j, A, C|X) q(A^*, C^*|A, C, N_j, N_j^*) \\
= \left(2\pi\right)^{-\frac{n+k}{2}} |H|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec} (X^T) \right] \right] \frac{I_{A}^{(A^*)}}{\Gamma_p \left(\frac{N_j^*}{2}\right)} \times \frac{1}{|C|^\frac{1}{2} |C^*|^\frac{1}{2} \left(\Delta_{C_{N_j^*}}^{-1} C_{N_j^*} \right)^{p-1} \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right)} \\
\exp \left[-\frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec} (X^T) \right] \right] + \frac{1}{2} \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right) + \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right)^{p-1} \left(\Delta_{C_{N_j^*}}^{-1} C_{N_j^*} \right) \\
= \left[H^* \right]^{-\frac{1}{2}} \left[C\right]^{-\frac{1}{2}} \left[C^*\right]^{-\frac{1}{2}} \left(\Delta_{C_{N_j^*}}^{-1} C_{N_j^*} \right)^{p-1} \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right) \frac{\Delta_{C_{N_j^*}}}{\Gamma_p \left(\frac{N_j^*}{2}\right)} \times \frac{\Delta_{C_{N_j}}}{\Gamma_p \left(\frac{N_j}{2}\right)} \\
\exp \left[-\frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec} (X^T) \right] \right] - \frac{1}{2} \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right) + \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right)^{p-1} \left(\Delta_{C_{N_j^*}}^{-1} C_{N_j^*} \right) \\
= \left[H^* \right]^{-\frac{1}{2}} \left[C\right]^{-\frac{1}{2}} \left[C^*\right]^{-\frac{1}{2}} \left(\Delta_{C_{N_j^*}}^{-1} C_{N_j^*} \right)^{p-1} \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right) \frac{\Delta_{C_{N_j^*}}}{\Gamma_p \left(\frac{N_j^*}{2}\right)} \times \frac{\Delta_{C_{N_j}}}{\Gamma_p \left(\frac{N_j}{2}\right)} \\
\exp \left[-\frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec} (X^T) \right] \right] - \frac{1}{2} \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right) + \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right)^{p-1} \left(\Delta_{C_{N_j^*}}^{-1} C_{N_j^*} \right) \\
= \left[H^* \right]^{-\frac{1}{2}} \left[C\right]^{-\frac{1}{2}} \left[C^*\right]^{-\frac{1}{2}} \left(\Delta_{C_{N_j^*}}^{-1} C_{N_j^*} \right)^{p-1} \left(\Delta_{C_{N_j}}^{-1} C_{N_j} \right) \frac{\Delta_{C_{N_j^*}}}{\Gamma_p \left(\frac{N_j^*}{2}\right)} \times \frac{\Delta_{C_{N_j}}}{\Gamma_p \left(\frac{N_j}{2}\right)}.
\end{align*}$$
where \( a^* \) is a vector formed by combining the diagonal entries of the proposed matrix \( A^* \) into a single vector.

Therefore, the acceptance probability for a move from a constant conditional covariance model to a DBEKK model is

\[
\begin{align*}
    r_{(N_j,1\rightarrow 2,\text{MG})} & = \min \left( 1, \left[ |H^*||H|^{-1} \right]^{-\frac{1}{2}} \left[ C \left( \Delta_{C_{N_j}, \cdot}^{-p-1} \right) |C^*|^{-\frac{1}{2}} \right] \right| \left( \Delta_{C_{N_j}, \cdot}^{-p-1} \right) \right| \left| V_{N_j} \right|^{-\frac{1}{2}} \right) \\
    & \times 2^{\frac{r_{\Delta C_{N_j} \cdot}}{2} - \frac{r_{\Delta C_{N_j} \cdot}}{2}} |\Delta_{C_{N_j}, \cdot}^{-1} C_{N_j}|^{-\frac{1}{2}} \left[ \Delta_{C_{N_j}, \cdot}^{-1} C_{N_j} \right] \left[ \Delta_{C_{N_j}, \cdot}^{-1} C_{N_j} \right]^{\frac{1}{2}} \frac{\Gamma_p \left( \frac{\Delta_{C_{N_j}, \cdot}}{2} \right)}{\Gamma_p \left( \frac{\Delta_{C_{N_j}, \cdot}}{2} \right)} \\
    & \times \exp \left[ -\frac{1}{2} \left( \left[ \text{vec}(X^T) \right]^T \left[ |H^*||H|^{-1} \right] \left[ \text{vec}(X^T) \right] - (a^* - m_{N_j, \cdot})^T V_{N_j}^{-1} (a^* - m_{N_j, \cdot}) + \text{tr} \left( \left[ \Delta_{C_{N_j}, \cdot}^{-1} C_{N_j} \right]^{-1} C \right) - \text{tr} \left( \left[ \Delta_{C_{N_j}, \cdot}^{-1} C_{N_j} \right]^{-1} C^* \right) \right) \right] \\
    & \times \frac{j^{2N_j}}{j^{2m_j}} (2\pi)^{\frac{p+m_j}{2}} \end{align*}
\]

(E.2.7)

### E.2.3 R\( J_\mathcal{N} \) - DBEKK(\( l, m \)) to Constant

When moving from a constant conditional covariance model to a DBEKK model, the relevant proposal distributions from (4.3.57) are the following:

\[
\begin{align*}
    q (A^*|A, N_j, N_j) & = 1 \quad \text{(Constant)} \\
    q (A^*|A, N_j, N_j) \sim N_{i+m} (m_{N_j, \cdot}, V_{N_j}) \quad \text{(DBEKK)} \\
    q (C^*|C, N_j, N_j) \sim W_p \left( \Delta_{C_{N_j}, \cdot}^{-1} C_{N_j}, \Delta_{C_{N_j}, \cdot} \right).
\end{align*}
\]

Therefore, the major calculation for the acceptance probability for such a move is

\[
\begin{align*}
    p (N_j, A^*, C^*|X) q (A, C|A^*, C^*, N_j, N_j) \\
    & \times p (N_j, A, C|X) q (A^*|A, C^*, N_j, N_j) q (C^*|C, N_j, N_j) \\
    & = \frac{p (N_j, A^*, C^*|X) q (A^*|A, C^*, N_j, N_j) q (C^*|C, N_j, N_j)}{p (N_j, A, C|X)} \\
    & = \frac{(2\pi)^{-\frac{p+m_j}{2}} |H^*|^{\frac{1}{2}} \text{exp} \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T \left[ H^*^{-1} \right] \left[ \text{vec}(X^T) \right] \right] \frac{\text{tr}(A^*)}{j^{2N_j}} \times \frac{\text{tr}(A)}{j^{2N_j}} \\
    & \times \frac{1}{2} \text{exp} \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec}(X^T) \right] \right] \frac{\text{tr}(A)}{j^{2N_j}} \\
    & \times \frac{1}{2} \text{exp} \left[ -\frac{1}{2} \left( a - m_{N_j, \cdot} \right)^T V_{N_j}^{-1} (a - m_{N_j, \cdot}) \right] \times \end{align*}
\]
Therefore, the acceptance probability for a move to a constant conditional covariance model from a 
diagonal model is

\[
\begin{align*}
\tau_{(N, p, -1, MG)} &= \min\left(1, \left[\mathbf{H}^* \mid \mathbf{H}^{-1}\right]^{-\frac{1}{2}} \left[\mathbf{C}^{\frac{1}{2}} \left(\frac{\mathbf{C}^{-1}}{N} \mathbf{C}_{N_j} \right)^{-1} \mathbf{C}^{*\frac{1}{2}} \right]^{-\frac{1}{2}} \left[\mathbf{V}_{N_j} \right]^{-\frac{1}{2}} \times \right. \\
&\quad \times \left(2 - \frac{\mathbf{a} \cdot \mathbf{m}_{N_j}}{\mathbf{V}_{N_j}^{-1}} \mathbf{a} - \mathbf{m}_{N_j} \right) + \text{tr} \left(\left[\frac{\mathbf{C}^{-1}}{N} \mathbf{C}_{N_j} \right]^{-1} \mathbf{C}^* \right) - \\
&\quad \text{tr} \left(\left[\frac{\mathbf{C}^{-1}}{N} \mathbf{C}_{N_j} \right]^{-1} \mathbf{C}^* \right) \right).
\end{align*}
\]

where \(\mathbf{a}\) is a vector formed by combining the diagonal entries of the matrix \(\mathbf{A}\) into a single vector.

Therefore, the acceptance probability for a move to a constant conditional covariance model from a 
DBEKK model is

\[
\begin{align*}
\tau_{(N, 2, -1, MG)} &= \min\left(1, \left[\mathbf{H}^* \mid \mathbf{H}^{-1}\right]^{-\frac{1}{2}} \left[\mathbf{C}^{\frac{1}{2}} \left(\frac{\mathbf{C}^{-1}}{N} \mathbf{C}_{N_j} \right)^{-1} \mathbf{C}^{*\frac{1}{2}} \right]^{-\frac{1}{2}} \left[\mathbf{V}_{N_j} \right]^{-\frac{1}{2}} \times \\
&\quad \times \left(2 - \frac{\mathbf{a} \cdot \mathbf{m}_{N_j}}{\mathbf{V}_{N_j}^{-1}} \mathbf{a} - \mathbf{m}_{N_j} \right) + \text{tr} \left(\left[\frac{\mathbf{C}^{-1}}{N} \mathbf{C}_{N_j} \right]^{-1} \mathbf{C}^* \right) - \\
&\quad \text{tr} \left(\left[\frac{\mathbf{C}^{-1}}{N} \mathbf{C}_{N_j} \right]^{-1} \mathbf{C}^* \right) \right).
\end{align*}
\]

(E.2.8)
When moving from a constant conditional covariance model to a BEKK model, the relevant proposal distributions from (4.3.57) are as follows:

\[
\begin{align*}
q(A^*|A, N_j, N_j^*) &= 1 \quad \text{(Constant)} \\
q(A^*|A, N_j, N_j^*) &\sim N_p, p(t+m) \left( M_{N_j}, U_{N_j}, V_{N_j} \right) \quad \text{(BEKK)} \\
q(C^*|C, N_j, N_j^*) &\sim W_p \left( \Delta_{C_N}^{-1}, C_{N_j^*}, \Delta_{C_N} \right) .
\end{align*}
\]

Therefore, the major calculation for the acceptance probability for such a move is

\[
\begin{align*}
p(N_j^*, A^*, C^*|X) q(A, C|A^*, C^*, N_j, N_j^*) \\
p(N_j, A, C|X) q(A^*, C^*|A, N_j, N_j^*)
\end{align*}
\]

\[
= \frac{p(N_j^*, A^*, C^*|X) q(A, C|A^*, C^*, N_j, N_j^*)}{p(N_j, A, C|X) q(A^*, C^*|A, N_j, N_j^*)} \times \frac{1}{(2\pi)^{\frac{nu}{2}}} |H|^{-\frac{nu}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left( H^{-1} \right) \left[ \text{vec} (X^T) \right] \right] \frac{1_{4d}(A^*)}{j^{\frac{nu}{2}}} \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \frac{U_{N_j^*}}{\left( U_{N_j^*} \right)^{\frac{1}{2}}} \right) \left( A^* - M_{N_j^*} \right)^T \frac{U_{N_j}}{\left( U_{N_j} \right)^{\frac{1}{2}}} \left( A^* - M_{N_j^*} \right) \right]
\]

\[
\frac{|C| \left( \frac{1}{2} \left( \Delta_{C}^{-1} \right)^{-p-1} \right) \exp \left[ -\frac{1}{2} \text{tr} \left( \left( \Delta_{C}^{-1} \right)^{-1} C \right) \right]}{2^{-\frac{1}{2}} |C_{N_j}^*|^{\frac{1}{2}} \left( \frac{\Delta_{C_{N_j}^*}}{2} \right)^{-p-1} \Gamma_p \left( \frac{\Delta_{C_{N_j}^*}}{2} \right) \exp \left[ -\frac{1}{2} \text{tr} \left( \left( \Delta_{C_{N_j}^*}^{-1} \right)^{-1} C \right) \right]} \times
\]

\[
\frac{2^{-\frac{1}{2}} |\Delta_{C_{N_j}}^{-1} C_{N_j}|^{\frac{1}{2}} \Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)}{\Delta_{C_{N_j}}^{\frac{1}{2}} C_{N_j}} \exp \left[ -\frac{1}{2} \text{tr} \left( \left( \Delta_{C_{N_j}}^{-1} \right)^{-1} C \right) \right]
\]

\[
= \left| H \right|^{-\frac{nu}{2}} \left| C \right|^{\frac{1}{2}} \left( \Delta_{C}^{-1} \right)^{-p-1} \left( \frac{1}{2} \left( \frac{\Delta_{C_{N_j}}}{2} \right)^{-p-1} \right) \left( 2\pi \right)^{\frac{1}{2}} \Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right) \times
\]

\[
\frac{\left| U_{N_j^*} \right|^{\frac{1}{2}} \left| V_{N_j} \right|^{\frac{1}{2}}}{2^{-\frac{1}{2}} \left( \frac{\Delta_{C_{N_j^*}}}{2} \right)^{-p-1} \Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)} \exp \left[ -\frac{1}{2} \text{tr} \left( \left( \Delta_{C_{N_j}}^{-1} \right)^{-1} C \right) \right]
\]

\[
\frac{\left| \Delta_{C_{N_j}} C_{N_j} \right|^{\frac{1}{2}} \Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)}{\Delta_{C_{N_j}}^{\frac{1}{2}} C_{N_j}} \exp \left[ -\frac{1}{2} \text{tr} \left( \left( \Delta_{C_{N_j}}^{-1} \right)^{-1} C \right) \right] + \frac{1}{2} \left[ \text{vec} (X^T) \right]^T \left( H^{-1} \right) \left[ \text{vec} (X^T) \right] +
\]

248
The acceptance probability for a move to a BEKK model from a constant conditional covariance model is

\[
\tau_{(N_t,1\rightarrow 3, MG)} = \min \left( 1, \left| H^* \right|^{-\frac{1}{2}} |H|^{\frac{3}{2}} \left| C \right| \frac{1}{2} \left( \Delta_{C_{N_t}}^{-p-1} \right) \left| C^* \right|^{-\frac{1}{2}} \left( \Delta_{C_{N_t}^*}^{-p-1} \right) (2\pi)^{\frac{(r+m)^2}{2}} j^{N_t} \times \frac{\left| U_{N_t} \right|}{\left| V_{N_t} \right|} \frac{\left| V_{N_t} \right|}{\left| U_{N_t} \right|} \left| \Delta_{C_{N_t}}^{-1} \right| \left| \Delta_{C_{N_t}^*}^{-1} \right| \Gamma_p \left( \frac{\Delta_{C_{N_t}}^{-1}}{2} \right) \times \Gamma_p \left( \frac{\Delta_{C_{N_t}^*}^{-1}}{2} \right) \right),
\]

where

\[
\frac{1}{2} \text{tr} \left( V_{N_t}^{-1} (A^* - M_{N_t})^T U_{N_t}^{-1} (A^* - M_{N_t}) \right) - \frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_t}}^{-1} \right] C_{N_t} \right) + \frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_t}^*}^{-1} \right] C_{N_t}^* \right)
\]

\[
= |H^*|^{-\frac{1}{2}} |H|^{\frac{3}{2}} \left| C \right| \frac{1}{2} \left( \Delta_{C_{N_t}}^{-p-1} \right) \left| C^* \right|^{-\frac{1}{2}} \left( \Delta_{C_{N_t}^*}^{-p-1} \right) (2\pi)^{\frac{(r+m)^2}{2}} j^{N_t} \times \frac{\left| U_{N_t} \right|}{\left| V_{N_t} \right|} \left| V_{N_t} \right| \left| U_{N_t} \right| \left| \Delta_{C_{N_t}}^{-1} \right| \left| \Delta_{C_{N_t}^*}^{-1} \right| \Gamma_p \left( \frac{\Delta_{C_{N_t}}^{-1}}{2} \right) \times \Gamma_p \left( \frac{\Delta_{C_{N_t}^*}^{-1}}{2} \right) \exp \left[ -\frac{1}{2} \left( \text{vec} \left( X^T \right) \right)^T \left( \left[ H^{-1} \right] - \left[ H^{-1} \right] \right) \left( \text{vec} \left( X^T \right) \right) - \text{tr} \left( V_{N_t}^{-1} (A^* - M_{N_t})^T U_{N_t}^{-1} (A^* - M_{N_t}) \right) + \text{tr} \left( \left[ \Delta_{C_{N_t}}^{-1} \right] C_{N_t} \right) - \text{tr} \left( \left[ \Delta_{C_{N_t}^*}^{-1} \right] C_{N_t}^* \right) \right].
\]
E.2.5 RJ\(_N\) - BEKK\((l, m)\) to Constant

When moving to a constant conditional covariance model from a BEKK model, the relevant proposal distributions from (4.3.57) are the following:

\[
\begin{align*}
q(A^* | A, N_j, N_j^\prime) & = 1 \quad \text{(Constant)} \\
q(A^* | A, N_j, N_j^\prime) & \sim \mathcal{N}_{p, p(l+m)}(M_{N_j^*}, U_{N_j^*}, V_{N_j^*}) \quad \text{(BEKK)} \\
q(C^* | C, N_j, N_j^\prime) & \sim \mathcal{W}_p\left(\Delta_{C_{N_j^*}}^{-1}, C_{N_j^*}, \Delta_{C_{N_j^*}}\right).
\end{align*}
\]

Therefore, the major calculation for the acceptance probability for such a move is

\[
\begin{align*}
p(N_j^*, A^*, C^* | X) & = \frac{p(N_j^*, A^*, C^* | X) q(A^* | A, C, N_j, N_j^\prime) q(C^* | C, N_j, N_j^\prime)}{p(N_j, A, C | X) q(A^* | A, C, N_j, N_j^\prime) q(C^* | C, N_j, N_j^\prime)} \\
& \times \exp \left[ -\frac{1}{2} \text{tr} \left( V_{N_j^*}^{-\frac{1}{2}} (A - M_{N_j^*})^T U_{N_j^*}^{-\frac{1}{2}} (A - M_{N_j^*}) \right) \right] \\
& \times \frac{|C|^\frac{1}{4} \left( \Delta_{C_{N_j^*}}^{-p-1} \right)^{\frac{1}{2}} \left( \Delta_{C_{N_j}} \Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right) \right)}{2 \frac{\Delta_{C_{N_j}}}{2}} \\
& \times \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}^{-1}, C_{N_j^*}^{-1}} \right) \right] \\
& \times \frac{\Gamma_p \left( \frac{\Delta_{C_{N_j^*}}}{2} \right)}{\Gamma_p \left( \frac{\Delta_{C_{N_j}}} {2} \right)} \\
& \times \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}^{-1}, C_{N_j^*}^{-1}} \right) \right] \\
& \times \frac{\Gamma_p \left( \frac{\Delta_{C_{N_j^*}}}{2} \right)}{\Gamma_p \left( \frac{\Delta_{C_{N_j}}} {2} \right)} \\
& \times \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}^{-1}, C_{N_j^*}^{-1}} \right) \right] \\
& \times \frac{\Gamma_p \left( \frac{\Delta_{C_{N_j^*}}}{2} \right)}{\Gamma_p \left( \frac{\Delta_{C_{N_j}}} {2} \right)} \\
& \times \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}^{-1}, C_{N_j^*}^{-1}} \right) \right]
\end{align*}
\]

\[
\begin{align*}
= |H|^\frac{1}{2} |H|^\frac{1}{2} \left| C \right| \frac{1}{2} \left( \Delta_{C_{N_j}^{-p-1}} \right) \left( \Delta_{C_{N_j}^{-1}} \right) \left( 2\pi \right)^{-\frac{(l+m)^2}{2}} \frac{j^{TN}}{j^{TN} \times}
\end{align*}
\]
The move from a DBEKK model to a BEKK model requires the following proposal distributions from E.2.6:

\[
|U_{N_j}|^{\frac{\tau_{(l,m)\min}}{2}} |V_{N_j}|^{-\frac{1}{2}} 2^{-\frac{p \Delta C_{N_j}^*}{2}} 2^{-\frac{\Delta C_{N_j}^*}{2}} \left| \Delta_{C_{N_j}^*}^{-1} C_{N_j}^* \right|^{\frac{\Delta C_{N_j}^*}{2}} \times
\]

\[
|\Delta_{C_{N_j}^*}^{-1} C_{N_j}^*|^{\frac{\Delta C_{N_j}^*}{2}} \Gamma_p \left( \frac{\Delta C_{N_j}^*}{2} \right) \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \text{vec} (X^T) \right)^T [H^{-1} - H^{-1}] \text{vec} (X^T) \right] +
\]

\[
\text{tr} \left( V_{N_j}^{-1} (A^* - M_{N_j})^T U_{N_j}^{-1} (A^* - M_{N_j}) \right) + \text{tr} \left( \left[ \Delta_{C_{N_j}^*}^{-1} C_{N_j}^* \right]^{-1} C^* \right) -
\]

\[
\text{tr} \left( \left[ \Delta_{C_{N_j}^*}^{-1} C_{N_j}^* \right]^{-1} C^* \right) \left] \right).
\]

The acceptance probability for a move to a BEKK model from a constant conditional covariance model is

\[
r_{(N_j,3\rightarrow1,MG)} = \min \left( 1, |H^*|^{-\frac{1}{2}} |H|^{-\frac{1}{2}} |C|^{-\frac{1}{2}} \left( \Delta_{C_{N_j}^*}^{-p-1} \right) |C^*|^{-\frac{1}{2}} \left( \Delta_{C_{N_j}^*}^{-p-1} \right) \right) \left( \frac{p \pi^{p/2}}{2} \right)^{\frac{\tau_{(l,m)\min}}{2}} \frac{j^\tau}{j^\tau} \times
\]

\[
|U_{N_j}|^{\frac{\tau_{(l,m)\min}}{2}} |V_{N_j}|^{-\frac{1}{2}} 2^{-\frac{p \Delta C_{N_j}^*}{2}} 2^{-\frac{\Delta C_{N_j}^*}{2}} \left| \Delta_{C_{N_j}^*}^{-1} C_{N_j}^* \right|^{\frac{\Delta C_{N_j}^*}{2}} \times
\]

\[
|\Delta_{C_{N_j}^*}^{-1} C_{N_j}^*|^{\frac{\Delta C_{N_j}^*}{2}} \Gamma_p \left( \frac{\Delta C_{N_j}^*}{2} \right) \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \text{vec} (X^T) \right)^T [H^{-1} - H^{-1}] \text{vec} (X^T) \right] +
\]

\[
\text{tr} \left( V_{N_j}^{-1} (A^* - M_{N_j})^T U_{N_j}^{-1} (A^* - M_{N_j}) \right) + \text{tr} \left( \left[ \Delta_{C_{N_j}^*}^{-1} C_{N_j}^* \right]^{-1} C^* \right) -
\]

\[
\text{tr} \left( \left[ \Delta_{C_{N_j}^*}^{-1} C_{N_j}^* \right]^{-1} C^* \right) \left] \right).
\]

### E.2.6 RJ\(\mathcal{N}\) - DBEK\(\mathcal{N}\)(\(l, m\)) to BEK\(\mathcal{N}\)(\(l^*, m^*\))

The move from a DBEKK model to a BEKK model requires the following proposal distributions from (4.3.57):

\[
q(A^* | A, N_j, N_j) \sim \mathcal{N}_{l+m}(m_{N_j^*}, V_{N_j^*}) \quad \text{(DBEKK)}
\]

\[
q(A^* | A, N_j, N_j) \sim \mathcal{N}_{p,p(l+m)}(M_{N_j^*}, U_{N_j^*}, V_{N_j^*}) \quad \text{(BEKK)}
\]

\[
q(C^* | C, N_j, N_j) \sim W_p \left( \Delta_{C_{N_j}^*}^{-1} C_{N_j^*}, \Delta_{C_{N_j}^*} \right).
\]
Therefore, the major calculation for the acceptance probability for such a move is

\[
p(N_j^*, A^*, C^* | X) q(A, C | A^*, C^*, N_j, N_j^*)
\]

\[
= p(N_j^*, A^*, C^* | X) q(A | A^*, C^*, N_j, N_j^*) q(C | C^*, N_j, N_j^*)
\]

\[
= (2\pi)^{-\frac{m}{2}} |H^{*}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T \left[ H^{*-1} \right] \left[ \text{vec}(X^T) \right] \right] \frac{\text{det}(A^*)}{\text{det}(A)} \times
\]

\[
(2\pi)^{-\frac{m}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec}(X^T) \right] \right] \frac{\text{det}(A)}{\text{det}(A^*)} \times
\]

\[
(2\pi)^{-\frac{m}{2}} |N_j|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (a - m_{N_j})^T V_{N_j}^{-1} (a - m_{N_j}) \right] \times
\]

\[
\exp \left[ -\frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A^* - M_{N_j})^T U_{N_j}^{-1} (A^* - M_{N_j}) \right) \right]
\]

\[
= |H^{*}|^{-\frac{1}{2}} |H|^{-\frac{1}{2}} |C|^{-\frac{1}{2}} \left( \Delta_{C_{N_j}}^{-p-1} \right) C^*^{-\frac{1}{2}} \left( \Delta_{C_{N_j}}^{-p-1} \right) (2\pi)^{\frac{(1+m)p}{2}} \left( 2\pi \right)^{-\frac{m}{2}} \times
\]

\[
\frac{j^{\frac{N}{2}}}{j^{\frac{N}{2}}} |N_j|^{-\frac{1}{2}} |U_{N_j}|^{-\frac{1}{2}} |V_{N_j}|^{-\frac{1}{2}} \times
\]

\[
\exp \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T \left[ H^{*-1} \right] \left[ \text{vec}(X^T) \right] \right] +
\]

\[
\frac{1}{2} \left[ \text{vec}(X^T) \right]^T \left[ H^{-1} \right] \left[ \text{vec}(X^T) \right] \right] - \frac{1}{2} (a - m_{N_j})^T V_{N_j}^{-1} (a - m_{N_j})
\]

\[
+ \frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A^* - M_{N_j})^T U_{N_j}^{-1} (A^* - M_{N_j}) \right) - \frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_j}}^{-1} \right] C^* \right) +
\]

\[
= |H^{*}|^{-\frac{1}{2}} |H|^{-\frac{1}{2}} |C|^{-\frac{1}{2}} \left( \Delta_{C_{N_j}}^{-p-1} \right) C^*^{-\frac{1}{2}} \left( \Delta_{C_{N_j}}^{-p-1} \right) (2\pi)^{\frac{(1+m)p}{2}} \left( 2\pi \right)^{-\frac{m}{2}} \times
\]

\[
\frac{j^{\frac{N}{2}}}{j^{\frac{N}{2}}} |N_j|^{-\frac{1}{2}} |U_{N_j}|^{-\frac{1}{2}} |V_{N_j}|^{-\frac{1}{2}} \times
\]

\[
\exp \left[ -\frac{1}{2} \left[ \text{vec}(X^T) \right]^T \left( \left[ H^{*-1} \right] - \left[ H^{-1} \right] \right) \left[ \text{vec}(X^T) \right] \right] +
\]

\[
\frac{1}{2} (a - m_{N_j})^T V_{N_j}^{-1} (a - m_{N_j})
\]

\[
- \text{tr} \left( V_{N_j}^{-1} (A^* - M_{N_j})^T U_{N_j}^{-1} (A^* - M_{N_j}) \right) +
\]

252
The major calculation that is required to determine the acceptance probability for such a move is:

\[
\text{tr} \left( \left[ \Delta_{C_{N_j}}^{-1} C_{N_j} \right]^{-1} C \right) - \text{tr} \left( \left[ \Delta_{C_{N_j}}^{-1} C_{N_j} \right]^{-1} C^* \right) \].
\]

The acceptance probability for a move to a BEKK model from a DBEKK model has the formula

\[
r_{(N_j, 2\sim 3, RJ)} = \min \left( 1, \left| \mathbf{H}^* \right|^{-\frac{1}{2}} \left| \mathbf{H} \right|^{\frac{1}{2}} \left| \mathbf{C} \right|^2 \left( \Delta_{C_{N_j}}^{-p-1} \right) \left| \mathbf{C}^* \right|^{-\frac{1}{2}} \left( \Delta_{C_{N_j}}^{-p-1} \right) (2\pi)^{\frac{(l^*+m^*)p^2-1-m}{2}} \times \right.
\]

\[
\left| V_{N_j} \right|^{-\frac{1}{2}} \left| U_{N_j} \right|^{-\frac{1}{2}} \left| \mathbf{C} \right|^{2} \left[ \frac{2}{2} \left( \frac{\Delta_{C_{N_j}}^{-1}}{2} \right) \Gamma_p \left( \frac{\Delta_{C_{N_j}}^{-1}}{2} \right) \right]^{j^T N_j} \times \right.
\]

\[
\exp \left[ -\frac{1}{2} \left( \text{vec} (\mathbf{X}^T) \right)^T \left( \left[ \mathbf{H}^{-1} \right] - \left[ \mathbf{H}^{-1} \right] \right) \left[ \text{vec} (\mathbf{X}^T) \right] + \right.
\]

\[
a - m_{N_j} \right)^T V_{N_j}^{-1} \left( a - m_{N_j} \right) - \right.
\]

\[
\left. \text{tr} \left( V_{N_j}^{-1} \left( \mathbf{A}^* - M_{N_j} \right)^T U_{N_j}^{-1} \left( \mathbf{A}^* - M_{N_j} \right) \right) \right] + \right.
\]

\[
\left. \text{tr} \left( \left[ \Delta_{C_{N_j}}^{-1} C_{N_j} \right]^{-1} C \right) - \text{tr} \left( \left[ \Delta_{C_{N_j}}^{-1} C_{N_j} \right]^{-1} C^* \right) \right). \]
\]

**E.2.7 RJ\textsubscript{N} - BEKK(l, m) to DBEKK(l\textsuperscript{*}, m\textsuperscript{*})**

The move from a BEKK model to a DBEKK model requires the following proposal distributions from (4.3.57):

\[
q(\mathbf{A}^*|\mathbf{A}, N_j, N_j) \sim N_{l+m} (m_{N_j}, V_{N_j}) \quad \text{(DBEKK)}
\]

\[
q(\mathbf{A}^*|\mathbf{A}, N_j, N_j) \sim N_{p,p(l+m)} (M_{N_j}, U_{N_j}, V_{N_j}) \quad \text{(BEKK)}
\]

\[
q(C^*|C, N_j, N_j) \sim \mathcal{W}_p \left( \Delta_{C_{N_j}}^{-1} C_{N_j}, \Delta_{C_{N_j}} \right).
\]

The major calculation that is required to determine the acceptance probability for such a move is

\[
p(N_j, \mathbf{A}^*, C^*|\mathbf{X}) q(\mathbf{A}, C|\mathbf{A}^*, C^*, N_j, N_j) \]

\[
p(N_j, \mathbf{A}, C|\mathbf{X}) q(\mathbf{A}^*, C^*|\mathbf{A}, C, N_j, N_j)
\]

\[
= \frac{p(N_j, \mathbf{A}^*, C^*|\mathbf{X}) q(\mathbf{A}^*, C^*|\mathbf{A}, C, N_j, N_j)}{p(N_j, \mathbf{A}, C|\mathbf{X}) q(\mathbf{A}^*, C^*|\mathbf{A}, C, N_j, N_j)}
\]

\[
= \frac{(2\pi)^{-\frac{l^*+m^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} (\mathbf{X}^T) \right]^T \left[ \mathbf{H}^{-1} \right] \left[ \text{vec} (\mathbf{X}^T) \right] \right] \frac{1}{2} \frac{\left| \mathbf{A}^* \right|^2}{\left| \mathbf{X}^T \right|^2}}{(2\pi)^{-\frac{l^*+m^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} (\mathbf{X}^T) \right]^T \left[ \mathbf{H}^{-1} \right] \left[ \text{vec} (\mathbf{X}^T) \right] \right] \frac{1}{2} \frac{\left| \mathbf{A} \right|^2}{\left| \mathbf{X}^T \right|^2}} \times
\]
\[
\frac{\exp\left[-\frac{1}{2} \text{tr} \left(V_N^{-1}_j \left(A-M_N \right)^T U_N^{-1}_j \left(A-M_N \right) \right)\right]}{(2\pi)^{\frac{(l+m)^2}{2}} |V_N_j|^{\frac{p}{2}} |U_N_j|^{\frac{(l+m)p}{2}}} \times \\
(2\pi)^{-\frac{p+m^2}{2}} |V_N_j|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\mathbf{a}^* - \mathbf{m}_N \right)^T V_N^{-1}_j \left(\mathbf{a}^* - \mathbf{m}_N \right)\right] \\
|C|^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{\frac{1}{2}} \exp \left[-\frac{1}{2} \text{tr} \left(\left[\Delta C_{N_j}^{-1} C_{N_j}\right]^{-1} C\right)\right] \\
\frac{p \Delta C_{N_j} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right)}{2 \frac{p \Delta C_{N_j} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right)}{2} \frac{p \Delta C_{N_j} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right)}{2}} \\
|\Delta^{-1}_C C_{N_j}|^{-\frac{1}{2}} \frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \exp \left[-\frac{1}{2} \text{vec} (X^T)^T [H^{-1}] \text{vec} (X^T) + \frac{1}{2} \text{tr} (V_{N_j}^{-1} \left(A-M_N \right)^T U_{N_j}^{-1} \left(A-M_N \right))\right] + \\
\frac{1}{2} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \\
\left[\Delta^{-1}_C C_{N_j}\right]^{-1} C \left[\Delta^{-1}_C C_{N_j}\right]^{-1} C \\
= |H^{*}|^{-\frac{1}{2}} |H|^{\frac{1}{2}} |C|^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \left(\Delta C_{N_j}^{t-p-1}\right)^{-\frac{1}{2}} \\
\frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \\
|\Delta^{-1}_C C_{N_j}|^{-\frac{1}{2}} \frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \frac{\Delta C_{N_j}}{2} \Gamma_p \left(\frac{\Delta C_{N_j}}{2}\right) \\
\exp \left[-\frac{1}{2} \text{vec} (X^T)^T \left(\left[\Delta^{-1}_C C_{N_j}\right]^{-1} C \right) \right] + \\
\text{tr} \left(\text{vec}(X^T)^T \left[\Delta^{-1}_C C_{N_j}\right]^{-1} C \right) - \text{tr} \left(\left[\Delta^{-1}_C C_{N_j}\right]^{-1} C \right) \\
\text{tr} \left(\left[\Delta^{-1}_C C_{N_j}\right]^{-1} C \right) - \text{tr} \left(\left[\Delta^{-1}_C C_{N_j}\right]^{-1} C \right) \
\\]
Therefore, the acceptance probability for a move to a BEKK model from a DBEKK model is

\[ r_{(N_j, 3 \rightarrow 2, MG)} = \min \left( 1, |H^*|^{-\frac{1}{2}} |H|^{\frac{1}{2}} |C|^{\frac{1}{2}} \left( \Delta_{CN_j} - p - 1 \right) |C^*|^{-\frac{1}{2}} \left( \Delta_{CN_j} - p - 1 \right) (2\pi)^{\frac{(l^* + m^*)p^2 + 1 + m}{2}} \right. \]

\[ \times \left( \left| V_{N_j} \right|^{-\frac{1}{2}} \left| U_{N_j} \right|^{\frac{1}{2}} \left| V_{N_j}^* \right|^{\frac{3}{2}} 2^{\frac{p \Delta_{CN_j}}{2}} 2^{-\frac{p \Delta_{CN_j}}{2}} \left| \Delta_{CN_j} C_{N_j} \right| \right)^{\frac{\Delta_{CN_j}}{2}} \left( \frac{\Delta_{CN_j}}{2} \right)^{j_{1N}} \times \frac{\Gamma_p \left( \frac{\Delta_{CN_j}}{2} \right)}{\Gamma_p \left( \frac{\Delta_{CN_j}}{2} \right)} \]

\[ \exp \left[ -\frac{1}{2} \left( \text{vec}(X^T) \right)^T \left( \left[ H^* - 1 \right] - \left[ H^{-1} \right] \right) \left[ \text{vec}(X^T) \right] + \left( a - m_{N_j} \right)^T V_{N_j}^{-1} \left( a - m_{N_j} \right) - \text{tr} \left( V_{N_j}^{-1} \left( A^* - M_{N_j} \right)^T U_{N_j}^{-1} \left( A^* - M_{N_j} \right) \right) + \text{tr} \left( \left[ \Delta_{CN_j}^{-1} C_{N_j} \right]^{-1} C - \text{tr} \left( \left[ \Delta_{CN_j}^{-1} C_{N_j} \right]^{-1} C^* \right) \right) \right]. \]

(E.2.12)
Appendix F

Multivariate STAR-GARCH Models

F.1 RJ \(N\) - Acceptance Probabilities

For the M-STAR-GARCH model, the Reversible Jump step for the conditional covariance model involves the movement from the current model index to a candidate model index. A total of three model indices result in six forms that the acceptance probability can take. These acceptance probabilities are outlined in the following subsections.

F.1.1 RJ \(N\) - BEKK \((l, m)\) to BEKK \((l^*, m^*)\)

The most complicated within model move is a move from a BEKK conditional covariance model to another BEKK model in which either of the model orders differ. The acceptance probability for such a move is presented below. Starting with the acceptance probability in (4.4.81) and using the full posterior distribution shown in (4.4.74), together with the proposal distributions shown in (4.4.82), the acceptance probability is determined as follows:

\[
p \left( N_j^*, A^*, C^* | X, M_k, \Phi, d, \gamma, c \right) q \left( A, C | A^*, C^*, N_j, N_j^* \right) \\
= \frac{p \left( N_j^*, A^*, C^* | X, M_k, \Phi, d, \gamma, c \right) q \left( A^*, C^* | A, C, N_j, N_j^* \right)}{p \left( N_j, A, C | X, M_k, \Phi, d, \gamma, c \right) q \left( A^* | A, N_j, N_j^* \right) q \left( C^* | C, N_j, N_j^* \right)} \\
= \frac{(2\pi)^{-n/2} |H^*|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left[ \text{vec} \left( [X - Z_l \Phi]^T \right) \right]^T \left[ H^{*-1} \right] \left[ \text{vec} \left( [X - Z_l \Phi]^T \right) \right] \right]}{(2\pi)^{-n/2} |H|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left[ \text{vec} \left( [X - Z_l \Phi]^T \right) \right]^T \left[ H^{-1} \right] \left[ \text{vec} \left( [X - Z_l \Phi]^T \right) \right] \right]} \\
\times \frac{L_{A}(A^*)}{L_{A}(A)} \times \frac{L_{A}(A^*)}{L_{A}(A)}
\]
\[
\frac{1}{(2\pi)^{1/2}} |\mathbf{V}_{N_j}^{1/2} \mathbf{A} - \mathbf{M}_{N_j}^{} |^{-1/2} (\mathbf{A} - \mathbf{M}_{N_j}^{})^T U_{N_j}^{-1} (\mathbf{A} - \mathbf{M}_{N_j}^{}) \]
\]

As are all calculations of acceptance probabilities for the Reversible Jump steps, this calculation is made, conditional on the first \(s^* = \max (l, m, l^*, m^*)\) data points for the numerator and the denominator. The most recently simulated values for the intercept and coefficient matrices are \(C = C^{(i-1)}\) and \(\mathbf{A} = \mathbf{A}^{(i-1)}\), respectively. Therefore, the acceptance probability is

\[
r_{(N_j, 3 \rightarrow 3, \text{MSG})} = \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]

\[= \min \left( \frac{j_{N_j}^{C^{(i-1)}}}{j_{N_j}^{C^{(i-1)}}} |\mathbf{H}^*|^{-1/2} |\mathbf{H}^{1/2} | (2\pi)^{1/2} |\mathbf{V}_{N_j}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} |\mathbf{C}^{1/2} |^{-1/2} \right) \]
F.1.2 \( \mathcal{R} \mathcal{J}_N \) - Constant to DBEKK(\( l^*, m^* \))

When moving from a constant conditional covariance model to a DBEKK model, the relevant proposal distributions from (4.3.57) are the following:

\[
q(A^*|A, N_j, N_j) = 1 \quad \text{(Constant)}
\]

\[
q(A^*|A, N_j, N_j, \sim A_{i+m}(m_{N_j}, V_{N_j}) \quad \text{(DBEKK)}
\]

\[
q(C^*|C, N_j, N_j) \sim W_p \left( \Delta^{-1}_{C_{N_j}}, C_{N_j}, \Delta_{C_{N_j}} \right).
\]

Therefore, the major calculation for the acceptance probability for such a move is

\[
p(N_j, A^*, C^*|X, M_k, \Phi, d, \gamma, c) q(A, C|A^*, C^*, N_j, N_j)
\]

\[
p(N_j, A, C|X, M_k, \Phi, d, \gamma, c) q(A^*, C^*|A, C, N_j, N_j)
\]

\[
= (2\pi)^{-\frac{n}{2}} |H^*|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( X - Z_i \Phi^T \right) \right)^T H^{-1} \left( \text{vec} \left( X - Z_i \Phi^T \right) \right) \right] \times
\]

\[
(2\pi)^{-\frac{n}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec} \left( X - Z_i \Phi^T \right) \right)^T H^{-1} \left( \text{vec} \left( X - Z_i \Phi^T \right) \right) \right] \times
\]

\[
\frac{1}{2} \left( \frac{\Delta_{C_{N_j}} - p^{-1}}{\Delta_{C_{N_j}} - p^{-1}} \right) \Delta_{C_{N_j}}^{-1} C_{N_j} \times
\]

\[
\frac{1}{2} \left( \frac{\Delta_{C_{N_j}} - p^{-1}}{\Delta_{C_{N_j}} - p^{-1}} \right) \Delta_{C_{N_j}}^{-1} C_{N_j} \times
\]

\[
\exp \left[ -\frac{1}{2} \left( \text{vec} \left( X - Z_i \Phi^T \right) \right)^T H^{-1} \left( \text{vec} \left( X - Z_i \Phi^T \right) \right) \right] +
\]

\[
\frac{1}{2} \left( \text{vec} \left( X - Z_i \Phi^T \right) \right)^T H^{-1} \left( \text{vec} \left( X - Z_i \Phi^T \right) \right) +
\]

\[
\frac{1}{2} \left( a^* - m_{N_j} \right)^T V_{N_j}^{-1} \left( a^* - m_{N_j} \right) - \frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right) \times
\]

\[
\frac{1}{2} \left( a^* - m_{N_j} \right)^T V_{N_j}^{-1} \left( a^* - m_{N_j} \right) - \frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right) \times
\]

\[
\left( 2\pi \right)^{\frac{\nu + m^*}{2}} \left| V_{N_j} \right|^\frac{1}{2} \times
\]

258
When moving from a constant conditional covariance model to a DBEKK model, the relevant proposal distribution from (4.3.57) are

\[
\begin{align*}
q(A^*|A, N_j, N_{j*}) &= 1 \quad \text{(Constant)} \\
q(A^*|A, N_j, N_{j*}) &\sim N_{l+m}(m_{N_{j*}}, V_{N_{j*}}) \quad \text{(DBEKK)} \\
q(C^*|C, N_j, N_{j*}) &\sim W_p\left(\Delta_{N_{j*}}^{-1}, C_{N_{j*}}, \Delta_{C_{N_{j*}}}\right).
\end{align*}
\]

Therefore, the major calculation for the acceptance probability for such a move is

\[
\frac{p(N_{j*}, A^*, C^*|X, M_k, \Phi, d, \gamma, c) q(A, C|A^*, C^*, N_j, N_{j*})}{p(N_{j}, A, C|X, M_k, \Phi, d, \gamma, c) q(A^*, C^*|A, C, N_j, N_{j*})}.
\]
Therefore, the acceptance probability for a move to a constant conditional covariance model from a

\[
p (N^*, A^*, C^* | X, M_k, \Phi, d, \gamma, c) q (A | A^*, N_j, N_j^*) q (C | C^*, N_j, N_j^*)
\]

\[
\frac{1}{p (N^*, A, C | X, M_k, \Phi, d, \gamma, c)} q (A | A^*, N_j, N_j^*) q (C | C^*, N_j, N_j^*)
\]

\[
(2\pi)^{-\frac{n_d - 1}{2}} |H^*|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T H^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] \right]
\]

\[
(2\pi)^{-\frac{n_d - 1}{2}} |H|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T H^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] \right]
\]

\[
\frac{1}{j^{dN}} \left( \frac{2\pi}{|C|} \left[ \Delta_{C N_j}^{-p-1} \right] |C^*|^{-\frac{1}{2}} \left( \Delta_{C N_j, C_j}^{-1} \right) \right) \frac{1}{j^{dN}} |V_{N_j}^{-1} |^{-\frac{1}{2}} \times
\]

\[
\exp \left[-\frac{1}{2} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T H^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] \right] +
\]

\[
\frac{1}{2} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T H^{-1} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] -
\]

\[
\frac{1}{2} (a - m_{N_j})^T V_{N_j}^{-1} (a - m_{N_j}) - \frac{1}{2} \text{tr} \left[ \left( \Delta_{C N_j}^{-1} \right) C \right] +
\]

\[
\frac{1}{2} \text{tr} \left( \left( \Delta_{C N_j, C_j}^{-1} \right) C \right)
\]

\[
\exp \left[-\frac{1}{2} \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right]^T \left( [H^{-1}] - [H^{-1}] \right) \left[ \text{vec} \left( [X - Z_1 \Phi]^T \right) \right] \right] +
\]

\[
(a - m_{N_j})^T V_{N_j}^{-1} (a - m_{N_j}) + \text{tr} \left( \left( \Delta_{C N_j}^{-1} \right) C \right) -
\]

\[
\text{tr} \left( \left( \Delta_{C N_j, C_j}^{-1} \right) C \right)
\]

where \( a \) is a vector formed by combining the diagonal elements of the matrix \( A \) into a single vector.

Therefore, the acceptance probability for a move to a constant conditional covariance model from a
DBEKK model is

\[
\begin{align*}
\rho_{(N_j, 2\rightarrow 1, \text{MSG})} & = \min \left( 1, \left[ |H^*| \right]^{-\frac{1}{2}} \left[ |C| - \left( \Delta C_{N_j} - p^{-1} \right) \right] \right)^{\frac{1}{2}} \left[ V_{N_j} \right]^{-\frac{1}{2}} \times \left[ \begin{array}{c}
2^{-\frac{p \Delta C_{N_j}}{2}} - \frac{\rho_{\Delta C_{N_j}}}{2} \left( |\Delta C_{N_j}| C_{N_j} - \frac{\Delta C_{N_j}}{2} \right) \Gamma_p \left( \frac{\Delta C_{N_j}}{2} \right) \times \exp \left[ -\frac{1}{2} \left( \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right]^T \left[ \Delta C_{N_j} - |H|^{-1} \right] \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right] + \left( a - m_{N_j} \right)^T V_{N_j}^{-1} \left( a - m_{N_j} \right) + \text{tr} \left( \left[ \Delta C_{N_j} C_{N_j} \right]^{-1} C \right) - \text{tr} \left( \left[ \Delta C_{N_j} C_{N_j} \right]^{-1} C \right) \right] \right] \right) \frac{(2\pi)^{-\frac{1}{2}}}{\sqrt{\det(2\pi)}}.
\end{align*}
\]

(F.1.3)

\[\text{F.1.4 RJ}_N - \text{Constant to BEKK}(l^*, m^*)\]

When moving from a constant conditional covariance model to a BEKK model, the relevant proposal distributions from (4.3.57) are

\[
\begin{align*}
q(A^* | A, N_j, N_{j'}) & = 1 \quad \text{(Constant)} \\
q(A^* | A, N_j, N_{j'}) & \sim N_{p,p(t+m)} \left( M_{N_j}, U_{N_{j'}}, V_{N_{j'}} \right) \quad \text{(BEKK)} \\
q(C^* | C, N_j, N_{j'}) & \sim W_p \left( \Delta C_{N_j}^{-1} C_{N_j}, \Delta C_{N_j} \right).
\end{align*}
\]

Therefore, the major calculation for the acceptance probability for such a move is

\[
\begin{align*}
p(N_j, A^*, C^* | X, M_k, \Phi, d, \gamma, c) \times q(A, C | A^*, C^*, N_j, N_{j'}) & = p(N_j, A^*, C^* | X, M_k, \Phi, d, \gamma, c) \times q(A^* | A, N_j, N_{j'}) \times q(C^* | C, N_j, N_{j'})
\end{align*}
\]

\[
\begin{align*}
& = \frac{(2\pi)^{-\frac{n_p}{2}}}{\sqrt{\det(2\pi)^{-\frac{n_p}{2}}}} \left| H \right|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right]^T \left[ \Delta C_{N_j} - |H|^{-1} \right] \left[ \text{vec} \left( [X - Z_i \Phi]^T \right) \right] \right] \times \frac{1}{\Gamma_p \left( \frac{\Delta C_{N_j}}{2} \right)} \times \exp \left[ -\text{tr} \left( \Delta C_{N_j} \left( A^* - M_{N_j} \right)^T U_{N_{j'}} \left( A^* - M_{N_{j'}} \right) \right) \right] \times \frac{1}{\Gamma_p \left( \frac{\Delta C_{N_j}}{2} \right)}\times \frac{1}{\Gamma_p \left( \frac{\Delta C_{N_j}}{2} \right)}
\end{align*}
\]
The acceptance probability for a move to a BEKK model from a constant conditional covariance model is

\[
\begin{align*}
&= |H^*|^{-\frac{1}{2}} |H| |C|^{-\frac{1}{2}} \left( \Delta C_{N_j, -p-1} \right) |C^*|^{-\frac{1}{2}} \left( \Delta C_{N_j, -p} \right) \frac{(\nu + m)^{m/2}}{2} \frac{j^{\gamma N}}{j^{\gamma N}} \\
&\quad \times \left| U_{N_j} \right|^{\frac{\nu (\nu + m)}{2}} \left| V_{N_j} \right|^{\frac{\nu (\nu + m - 1)}{2}} \left| U_{N_j} \right|^{\frac{\nu (\nu + m - 1)}{2}} \left| V_{N_j} \right|^{\frac{\nu (\nu + m - 1)}{2}} \\
&\quad \times \exp \left[ -\frac{1}{2} \left( \text{vec} \left( (X - Z_1 \Phi)^T \right) \right)^T \left( \text{vec} \left( (X - Z_1 \Phi)^T \right) \right) \right] \\
&\quad + \frac{1}{2} \text{vec} \left( (X - Z_1 \Phi)^T \right)^T \left( H^{-1} \right) \left( \text{vec} \left( (X - Z_1 \Phi)^T \right) \right) \\
&\quad + \frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A^* - M_{N_j})^T U_{N_j}^{-1} (A^* - M_{N_j}) \right) - \frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_j}} C_{N_j} \right]^{-1} C \right) \\
&\quad - \frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_j}} C_{N_j} \right]^{-1} C \right) \right].
\end{align*}
\]

The acceptance probability for a move to a BEKK model from a constant conditional covariance model is then

\[
\begin{align*}
&= \min \left( 1, |H^*|^{-\frac{1}{2}} |H| |C|^{-\frac{1}{2}} \left( \Delta C_{N_j, -p-1} \right) |C^*|^{-\frac{1}{2}} \left( \Delta C_{N_j, -p} \right) \frac{(\nu + m)^{m/2}}{2} \frac{j^{\gamma N}}{j^{\gamma N}} \\
&\quad \times \left| U_{N_j} \right|^{\frac{\nu (\nu + m)}{2}} \left| V_{N_j} \right|^{\frac{\nu (\nu + m - 1)}{2}} \left| U_{N_j} \right|^{\frac{\nu (\nu + m - 1)}{2}} \left| V_{N_j} \right|^{\frac{\nu (\nu + m - 1)}{2}} \\
&\quad \times \exp \left[ -\frac{1}{2} \left( \text{vec} \left( (X - Z_1 \Phi)^T \right) \right)^T \left( \text{vec} \left( (X - Z_1 \Phi)^T \right) \right) \right] \\
&\quad + \frac{1}{2} \text{vec} \left( (X - Z_1 \Phi)^T \right)^T \left( H^{-1} \right) \left( \text{vec} \left( (X - Z_1 \Phi)^T \right) \right) \\
&\quad + \frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A^* - M_{N_j})^T U_{N_j}^{-1} (A^* - M_{N_j}) \right) - \frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_j}} C_{N_j} \right]^{-1} C \right) \\
&\quad - \frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_j}} C_{N_j} \right]^{-1} C \right) \right) \right].
\end{align*}
\]
F.1.5 RJ_\mathcal{N} - BEKK(l, m) to Constant

When moving to a constant conditional covariance model from a BEKK model, the relevant proposal distributions from (4.3.57) are

\[
q(A^*|A, N_j, N_{ij}) = 1 \quad \text{ (Constant)}
\]

\[
q(A^*|A, N_j, N_{ij}) \sim N_{p,p(t+m)}(M_{N_j}, U_{N_j}, V_{N_j}) \quad \text{(BEKK)}
\]

\[
q(C^*|C, N_j, N_{ij}) \sim W_p \left( \Delta_{C_{N_j}}^{-1}, C_{N_j}, \Delta_{C_{N_j}} \right).
\]

Therefore, the major calculation for the acceptance probability for such a move is

\[
p(N_j, A^*, C^*|X, M_k, \Phi, d, \gamma, c) q(A, C|A^*, C^*, N_j, N_{ij})
\]

\[
= \frac{p(N_j, A^*, C^*|X, M_k, \Phi, d, \gamma, c) q(A^*|A, C, N_j, N_{ij}) q(C^*|C, N_j, N_{ij})}{p(N_j, A, C|X, M_k, \Phi, d, \gamma, c) q(A^*|A, C, N_j, N_{ij}) q(C^*|C, N_j, N_{ij})}
\]

\[
= \frac{(2\pi)^{-\frac{n+p}{2}} |H^*|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec}(X - Z_i \Phi)^T \right)^T H^{-1} \left( \text{vec}(X - Z_i \Phi)^T \right) \right]}{(2\pi)^{-\frac{n+p}{2}} |H|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \text{vec}(X - Z_i \Phi)^T \right)^T H^{-1} \left( \text{vec}(X - Z_i \Phi)^T \right) \right]}
\]

\[
\times \frac{\exp \left[ -\frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A - M_{X_j})^T U_{N_j}^{-1} (A - M_{X_j}) \right) \right]}{\exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right) \right]}
\]

\[
\times \frac{\exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right) \right]}{\exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right) \right]}
\]

\[
= |H^*|^{-\frac{1}{2}} |H|^{-\frac{1}{2}} |C| \left( \Delta_{C_{N_j}}^{-1} \right) \left( \Delta_{C_{N_j}}^{-1} \right) \left( 2\pi \right)^{-\frac{(l+m)p}{2}} j^{\frac{1}{2}N} \times
\]

\[
\left| U_{N_j} \right|^{-\frac{(l+m)p}{2}} \left| V_{N_j} \right|^{-\frac{n+p}{2}} 2^{-\frac{n+p}{2}} \left| \Delta_{C_{N_j}}^{-1} C_{N_j} \right| \left| \Delta_{C_{N_j}}^{-1} C_{N_j} \right|
\]

\[
\times \exp \left[ -\frac{1}{2} \left( \text{vec}(X - Z_i \Phi)^T \right)^T H^{-1} \left( \text{vec}(X - Z_i \Phi)^T \right) \right] + \frac{1}{2} \left( \text{vec}(X - Z_i \Phi)^T \right)^T H^{-1} \left( \text{vec}(X - Z_i \Phi)^T \right) - \frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A^* - M_{X_j})^T U_{N_j}^{-1} (A^* - M_{X_j}) \right) - \frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right) + \frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right).
\]
The move from a DBEKK model to a BEKK model requires the following proposal distributions from \( \mathbb{N} \).

\[
\begin{align*}
|U_{N_j}| & \sim \frac{\Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)}{\Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)} \times \\
|V_{N_j}| & \sim \frac{\Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)}{\Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)} \times \\
\Delta_{C_{N_j}} \mathcal{C}_{N_j} & \sim \frac{\Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)}{\Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)} \times \\
\exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T \left( \left[ H^{-1} \right] - \left[ H^{-1} \right] \right) \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) \right] + \\
\text{tr} \left( V_{N_j}^T \left( A^* - M_{N_j} \right) U_{N_j} \left( A^* - M_{N_j} \right) \right) + \text{tr} \left( \left[ \Delta_{C_{N_j}} \mathcal{C}_{N_j} \right]^{-1} C \right) - \\
\text{tr} \left( \left[ \Delta_{C_{N_j}} \mathcal{C}_{N_j} \right]^{-1} \left[ \Delta_{C_{N_j}} \mathcal{C}_{N_j} \right]^{-1} C \right) \right].
\end{align*}
\]

Therefore, the acceptance probability for a move to a BEKK model from a constant conditional covariance model is

\[
r_{(N_j \rightarrow 1, \text{MSG})} = \min \left( 1, \frac{|H^*|^{-\frac{1}{2}} |H|^{-\frac{1}{2}} |C|^{-\frac{1}{2}} \left( \Delta_{C_{N_j}} \mathcal{C}_{N_j} \right)^{-\frac{1}{2}} \left( \Delta_{C_{N_j}} \mathcal{C}_{N_j} \right)^{-\frac{1}{2}}} \right) \times \frac{\left( 2\pi \right)^{-\frac{(l+m)^2}{2}}}{\left( J^{T N} \right)^{\frac{1}{2}}} \times \\
\frac{|U_{N_j}|^{\frac{\Delta_{C_{N_j}}}{2}} |V_{N_j}|^{\frac{\Delta_{C_{N_j}}}{2}} \Delta_{C_{N_j}} \mathcal{C}_{N_j}^{\frac{\Delta_{C_{N_j}}}{2}}}{\left( \Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right) \right)^{\frac{\Delta_{C_{N_j}}}{2}}} \times \\
\exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T \left( \left[ H^{-1} \right] - \left[ H^{-1} \right] \right) \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) \right] + \\
\text{tr} \left( V_{N_j}^T \left( A^* - M_{N_j} \right) U_{N_j} \left( A^* - M_{N_j} \right) \right) + \text{tr} \left( \left[ \Delta_{C_{N_j}} \mathcal{C}_{N_j} \right]^{-1} C \right) - \\
\text{tr} \left( \left[ \Delta_{C_{N_j}} \mathcal{C}_{N_j} \right]^{-1} \left[ \Delta_{C_{N_j}} \mathcal{C}_{N_j} \right]^{-1} C \right) \right].
\]

\[
(F.1.5)
\]

\section*{F.1.6 RJ_{N_j} - DBEKK(l, m) to BEKK(l^*, m^*)}

The move from a DBEKK model to a BEKK model requires the following proposal distributions from (4.3.57):

\[
q \left( A^* | A, N_j, \mathcal{N}_j \right) \sim N_{l+m} \left( m_{N_j}, V_{N_j} \right) \quad \text{(DBEKK)}
\]

\[
q \left( A^* | A, N_j, \mathcal{N}_j \right) \sim N_{p,p(l+m)} \left( M_{N_j}, U_{N_j}, V_{N_j} \right) \quad \text{(BEKK)}
\]

\[
q \left( C^* | C, N_j, \mathcal{N}_j \right) \sim W_p \left( \Delta_{C_{N_j}} \mathcal{C}_{N_j}, \Delta_{C_{N_j}} \right) \quad.
\]
Therefore, the major calculation for the acceptance probability for such a move is

\[
p \left( N_{j', A^*, C^*} | X, M_k, \Phi, d, \gamma, c \right) q \left( A, C | A^*, C^*, N_{j', N_{j'}} \right) = p \left( N_{j}, A^*, C^* | X, M_k, \Phi, d, \gamma, c \right) q \left( A^* | A^* C^* N_{j}, N_{j'} \right) q \left( C^* | C^* N_{j}, N_{j'} \right)
\]

\[
\approx (2\pi)^{-n-j} |H|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right) \right] \right] \right\} \times
\]

\[
\left( 2\pi \right)^{-n-j} |H|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right) \right] \right] \right\} \times
\]

\[
\exp \left\{ -\frac{1}{2} \left( V_{N_{j'}} \left( A^* - M_{N_{j'}} \right)^T U_{N_{j'}}^{-1} \left( A^* - M_{N_{j'}} \right) \right) \right\} \times
\]

\[
\frac{1}{C_i} \left( \Delta_{C_{N_{j'}}, -p-1} \right) \exp \left\{ -\frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_{j'}}, -1} \right] C \right) \right\} \times
\]

\[
\frac{\Delta_{C_{N_{j'}}, -1} \exp \left\{ -\frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_{j'}}, -1} \right] C \right) \right\}}{\Delta_{C_{N_{j'}}, -1} \exp \left\{ -\frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_{j'}}, -1} \right] C \right) \right\}} \times
\]

\[
= \left| H \right|^{-\frac{1}{2}} \left| H \right| \left| C \right| \left( \Delta_{C_{N_{j'}}, -p-1} \right) \left( C^* \right) \left( \Delta_{C_{N_{j'}}, -p-1} \right) \left( 2\pi \right)^{\frac{(n+m)p^2}{2}} \left( 2\pi \right)^{-n-j} \frac{j_{N_{j}}}{j_{N_{j'}}} \times
\]

\[
\exp \left\{ -\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right) \right] \right) \right\} +
\]

\[
\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right) \right] \right) - \frac{1}{2} \left( a - m_{N_{j}} \right)^T V_{N_{j}}^{-1} \left( a - m_{N_{j}} \right) +
\]

\[
\frac{1}{2} \text{tr} \left( V_{N_{j}} \left( A^* - M_{N_{j}} \right)^T U_{N_{j}}^{-1} \left( A^* - M_{N_{j}} \right) \right) - \frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_{j'}}, -1} \right] C \right) +
\]

\[
\frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_{j'}}, -1} \right] C \right) \times
\]

\[
\left| V_{N_{j'}} \right|^{-\frac{1}{2}} \left| U_{N_{j'}} \right| \frac{\left( 2\pi \right)^{\frac{(n+m)p^2}{2}}}{\left( 2\pi \right)^{-n-j} \frac{j_{N_{j}}}{j_{N_{j'}}}} \times
\]

\[
\exp \left\{ -\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right) \right] \right) \right\} +
\]

\[
\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right) \right] \right) - \frac{1}{2} \left( a - m_{N_{j}} \right)^T V_{N_{j}}^{-1} \left( a - m_{N_{j}} \right) +
\]

\[
\frac{1}{2} \text{tr} \left( \left[ \Delta_{C_{N_{j'}}, -1} \right] C \right) \times
\]

\[
\left| \Delta_{C_{N_{j'}}, -1} \right| \frac{\left( 2\pi \right)^{\frac{(n+m)p^2}{2}}}{\left( 2\pi \right)^{-n-j} \frac{j_{N_{j}}}{j_{N_{j'}}}} \times
\]

\[
\exp \left\{ -\frac{1}{2} \left( \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right)^T \left[ H^{-1} \right] \left[ \text{vec} \left( \left[ X - Z_i \Phi \right]^T \right) \right] \right) \right\} +
\]

\[
\left( a - m_{N_{j}} \right)^T V_{N_{j}}^{-1} \left( a - m_{N_{j}} \right) -
\]

\[
\frac{1}{2} \text{tr} \left( V_{N_{j}} \left( A^* - M_{N_{j}} \right)^T U_{N_{j}}^{-1} \left( A^* - M_{N_{j}} \right) \right) +
\]

\[
\]
The major calculation that is required to determine the acceptance probability for such a move is

\[
\text{tr} \left( \left[ \Delta_{C,N_j}^{-1} C_{N_j} \right]^{-1} C \right) - \text{tr} \left( \left[ \Delta_{q,N_j}^{-1} C_{N_j} \right]^{-1} C^* \right).
\]

Therefore, the acceptance probability for a move to a BEKK model from a DBEKK model is

\[
t_{(N_j, 2 \to 3, \text{MSG})} = \min \left( 1, |H^*|^{-\frac{1}{2}} |H|^{-\frac{1}{2}} \left| C \right|^{-\frac{1}{2}} \left( \Delta_{C,N_j}^{-p-1} C_{N_j} \right) \left( \Delta_{q,N_j}^{-p-1} C_{N_j} \right)^* \right)
\]

\[
\times \left( 2\pi \right)^{\frac{(l^* + m^*)^2 - l + m}{4}} \times
\]

\[
\times \exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_l \Phi]^T \right)^T \left( [H^* - \text{vec} \left( [X - Z_l \Phi]^T \right)] \right) \right]
\]

\[
- \frac{1}{2} \left( \text{vec} \left( [X - Z_l \Phi]^T \right)^T \left( [H^* - \text{vec} \left( [X - Z_l \Phi]^T \right)] \right) \right) +
\]

\[
\left( a - m_{N_j} \right)^T V_{N_j}^{-1} \left( a - m_{N_j} \right) -
\]

\[
\text{tr} \left( V_{N_j}^{-1} \left( A^* - M_{N_j} \right)^T U_{N_j}^{-1} \left( A^* - M_{N_j} \right) \right) +
\]

\[
\text{tr} \left( \left[ \Delta_{C,N_j}^{-1} C_{N_j} \right]^{-1} C \right) - \text{tr} \left( \left[ \Delta_{q,N_j}^{-1} C_{N_j} \right]^{-1} C^* \right) \right).
\]

\textbf{F.1.7 RJ}_N \text{- BEKK}(l, m) \text{ to DBEKK}(l^*, m^*)

The move from a BEKK model to a DBEKK model requires the following proposal distributions from (4.3.57):

\[
q \left( A^* | A, N_j, N_j^* \right) \sim N_{l+m} \left( m_{N_j}, V_{N_j} \right) \quad \text{(DBEKK)}
\]

\[
q \left( A^* | A, N_j, N_j^* \right) \sim N_{p,p(l+m)} \left( m_{N_j}, U_{N_j}, V_{N_j} \right) \quad \text{(BEKK)}
\]

\[
q \left( C^* | C, N_j, N_j^* \right) \sim W_p \left( \Delta_{C,N_j}^{-1} C_{N_j}, \Delta_{C,N_j} \right).
\]
\[ \frac{1}{(2\pi)^{(k+m)\frac{N}{2}}} |V_{N_j}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A - M_{N_j})^T U_{N_j}^{-1} (A - M_{N_j}) \right) \right] \times \]

\[ \frac{1}{(2\pi)^{(k+m)\frac{N}{2}}} |V_{N_j}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (a^* - m_{N_j})^T V_{N_j}^{-1} (a^* - m_{N_j}) \right] \times \]

\[ |C|^\frac{1}{2} (\Delta_{C_{N_j}}^{-p-1}) \exp \left[ -\frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right)^{-1} C \right] \]

\[ \frac{2^{\frac{pA_{C_{N_j}}}{2}} |\Delta_{C_{N_j}}^{-1} C_{N_j}|^{-\frac{2}{2}} \Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)}{2^{\frac{pA_{C_{N_j}}}{2}} |\Delta_{C_{N_j}}^{-1} C_{N_j}|^{-\frac{2}{2}} \Gamma_p \left( \frac{\Delta_{C_{N_j}}}{2} \right)} \]

\[ \frac{j^{2\Delta_{C_{N_j}}}}{j^{2\Delta_{C_{N_j}}}} |V_{N_j}|^{-\frac{1}{2}} |U_{N_j}|^{-\frac{1}{2}} |V_{N_j}|^{-\frac{1}{2}} 2^{pA_{C_{N_j}} - 2} |\Delta_{C_{N_j}}^{-1} C_{N_j}|^{-\frac{2}{2}} \times \]

\[ \exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T \left( H^{-1} \right) \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) \right] + \]

\[ \frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T \left( H^{-1} \right) \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) - \]

\[ \frac{1}{2} \text{tr} \left( V_{N_j}^{-1} (A - M_{N_j})^T U_{N_j}^{-1} (A - M_{N_j}) \right) + \]

\[ \frac{1}{2} (a^* - m_{N_j})^T V_{N_j}^{-1} (a^* - m_{N_j}) - \frac{1}{2} \text{tr} \left( \Delta_{C_{N_j}}^{-1} C_{N_j} \right)^{-1} C \]
Therefore, the acceptance probability for a move to a BEKK model from a DBEKK model is

\[
\begin{align*}
\tau(N_j, 3 \rightarrow 2, MSG) &= \min \left( 1, |H^*|^{-\frac{1}{2}} |H|^\frac{1}{2} |C|^{-\frac{1}{2}} (\Delta_{cN_j} - p - 1)^{-\frac{1}{2}} (\Delta_{cN_j} - p - 1)^{-\frac{1}{2}} \right) \left( 2\pi \right)^{(l^* + m^*)^2 - n_m} \times \\
&\quad \left| V_{N_j} \right|^{-\frac{1}{2}} \left| U_{N_j^*} \right|^{-\frac{1}{2}} \left| V_{N_j^*} \right|^{-\frac{1}{2}} \left| U_{N_j} \right|^{-\frac{1}{2}} \Gamma_p \left( \frac{\Delta_{cN_j^*}}{2} \right) \frac{j^{\tau N_j}}{j^{\tau N_j^*}} \times \\
&\quad \left| \Delta_{cN_j} \right|^{-\frac{1}{2}} \left| \Delta_{cN_j^*} \right|^{-\frac{1}{2}} \Gamma_p \left( \frac{\Delta_{cN_j^*}}{2} \right) \frac{j^{\tau N_j}}{j^{\tau N_j^*}} \times \\
&\quad \exp \left[ -\frac{1}{2} \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right)^T \left( [H^* - I] - [H^{-1}] \right) \left( \text{vec} \left( [X - Z_1 \Phi]^T \right) \right) + \right. \\
&\quad \left. (a - m_{N_j})^T V_{N_j}^{-1} (a - m_{N_j}) - ight. \\
&\quad \left. \text{tr} \left( V_{N_j}^{-1} (A^* - M_{N_j})^T U_{N_j}^{-1} (A^* - M_{N_j}) \right) + ight. \\
&\quad \left. \text{tr} \left( \left[ \Delta_{cN_j} \right]^{-1} C \right) - \text{tr} \left( \left[ \Delta_{cN_j^*} \right]^{-1} C^* \right) \right].
\end{align*}
\]
Bibliography


