COMPLETE BAYESIAN ANALYSIS OF SOME MIXTURE TIME SERIES MODELS

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2012

By
A.B.M. Shahadat Hossain
School of Mathematics
Contents

Abstract 18

Declaration 19

Copyright 20

Acknowledgements 23

1 Introduction 24

2 Background Theory 27

2.1 Introduction ................................................................. 27

2.2 The Bayesian approach .................................................. 27

2.2.1 Bayes theorem and statistical inference ......................... 28

2.2.2 Conjugate prior .......................................................... 29

2.3 MCMC methods ............................................................. 30

2.3.1 Markov chain ............................................................ 31

2.3.2 The MCMC algorithm .................................................. 32

2.4 Convergence diagnostic tests .......................................... 33

2.4.1 Brooks, Gelman & Rubin convergence diagnostic ............. 33

2.4.2 Geweke convergence diagnostic ................................... 34
2.4.3 Heidelberger and Welch convergence diagnostic ............ 35
2.4.4 Raftery and Lewis’s diagnostic ............................ 35
2.5 Gibbs sampling ............................................. 36
2.6 The Metropolis-Hastings (M-H) algorithm ..................... 37
2.7 Bayesian model selection .................................... 37
   2.7.1 Reversible Jump MCMC ................................. 38
   2.7.2 Carlin and Chib method ............................... 41
2.8 Within-model simulation ...................................... 42
   2.8.1 Marginal likelihood from the Gibbs output ............ 42
   2.8.2 Marginal likelihood from the M-H output ............ 44
2.9 Finite mixture models ....................................... 46
   2.9.1 Missing data formulation and the likelihood .......... 46
2.10 Identifiability ................................................. 48
   2.10.1 The permutation sampler ............................... 49
   2.10.2 Relabelling ............................................. 49
   2.10.3 A new reordering scheme ............................. 49
2.11 Review of some distributions ................................ 50
   2.11.1 Gamma distribution .................................. 50
   2.11.2 The Inverted-Gamma distribution .................... 51
   2.11.3 Beta distribution ..................................... 52
   2.11.4 Generalized beta distribution ....................... 53
   2.11.5 The Normal distribution .............................. 53
   2.11.6 Truncated Normal distribution ....................... 54
   2.11.7 The Student-\(t\) distribution ....................... 55
3 Mixture of Gaussian AR components

3.1 Introduction ................................................. 58
3.2 Background ................................................. 59
3.3 The $\mathcal{M}_{\text{AR}}$ model ............................... 60
   3.3.1 Mean and variance of the $\mathcal{M}_{\text{AR}}$ model ..... 61
   3.3.2 Hypothesis for the $\mathcal{M}_{\text{AR}}$ model .......... 62
3.4 Missing data formulation and the likelihood function .......... 62
3.5 Priors setup ................................................. 63
   3.5.1 Formulae for hyperparameters .......................... 65
3.6 MCMC moves for parameter estimation ........................ 65
   3.6.1 Move I: updating $z$ .................................. 66
   3.6.2 Move II: updating $\omega$ .............................. 67
   3.6.3 Move III: updating $\mu$ .............................. 68
   3.6.4 Move IV: updating $\phi$ .............................. 71
   3.6.5 Move V: updating $\lambda$ and $\tau$ .................. 73
3.7 MCMC moves for model determination ........................ 74
   3.7.1 Move VI: updating of $p$ ............................. 75
   3.7.2 Move VII: updating of $g$ ............................ 79
3.8 Examples .................................................. 82
   3.8.1 Experiment with simulated data sets ................... 82
   3.8.2 Experiment with real datasets ......................... 90
4 Mixture of Gaussian ARMA components

4.1 Introduction .......................................................... 108
4.2 Background ........................................................... 109
4.3 The $\mathcal{M}_{\text{ARMA}}$ model ................................. 110
  4.3.1 Hypotheses of the model ........................................ 111
4.4 Missing data formulation and the likelihood function .......... 111
4.5 Priors setup ............................................................ 112
  4.5.1 Priors of $\omega$, $\mu$ and $\tau$ ............................. 113
  4.5.2 Priors for $\phi$ and $\theta$ .................................... 113
  4.5.3 Priors for $g$, $p$ and $v$ .................................... 114
  4.5.4 Formulae for hyperparameters .............................. 114
4.6 MCMC moves for parameter estimation ......................... 114
  4.6.1 Move I: updating $\omega$ ..................................... 115
  4.6.2 Move II: updating $\mu$ ..................................... 115
  4.6.3 Move III: updating $\phi$ ..................................... 117
  4.6.4 Move IV: updating $\theta$ ..................................... 119
  4.6.5 Move V: updating $\lambda$ and $\tau$ ......................... 121
  4.6.6 Move VI: updating $z$ ....................................... 121
4.7 MCMC moves for model determination .......................... 122
  4.7.1 Move VII: updating of $p$ .................................... 122
  4.7.2 Move VIII: updating of $q$ .................................. 123
  4.7.3 Move IX: updating of $g$ .................................... 125
4.8 Examples .............................................................. 128
  4.8.1 Experiment with simulated data sets ....................... 128
4.8.2 Experiment with real datasets ..................................... 129

5 Mixture of Gaussian ARMA-GARCH components ................. 139

5.1 Introduction ....................................................... 139
5.2 Background ....................................................... 140
5.3 The $\mathcal{M}_{\text{ARMA-GARCH}}$ model .......................... 141
5.3.1 Hypotheses for the model ..................................... 142
5.4 Missing data formulation and the likelihood function .......... 143
5.5 Priors setup ....................................................... 145
5.5.1 Priors of $\omega$ and $\mu$ ................................. 145
5.5.2 Prior for $\phi$ and $\theta$ ..................................... 146
5.5.3 Priors for $\alpha$ and $\beta$ ................................. 146
5.5.4 Priors for $\Lambda$ ......................................... 147
5.5.5 Formulae for hyperparameters .............................. 148
5.6 MCMC moves for parameter estimation .......................... 148
5.6.1 Move I: updating $z$ ....................................... 148
5.6.2 Move II: updating $\omega$ .................................... 150
5.6.3 Move III: updating $\mu$ .................................... 150
5.6.4 Move IV-A: updating $\phi$ ............................... 152
5.6.5 Move IV-B: updating $\theta$ ............................... 155
5.6.6 Move V-A: updating $\alpha$ ............................... 156
5.6.7 Move V-B: updating $\beta$ ............................... 158
5.7 MCMC moves for model determination ......................... 159
5.7.1 Move VI-A: updating of $p$ ............................... 159
5.7.2 Move VI-B: updating of $q$ ............................... 160
5.7.3 Move VII-A: updating of $r$ ................................. 161
5.7.4 Move VII-B: updating of $s$ ................................. 164
5.7.5 Move VIII: updating of $g$ ................................. 165
5.8 Examples ......................................................... 166

6 Mixture of Student-\(t\) AR components ........................................... 177
6.1 Introduction ......................................................... 177
6.2 The $\mathcal{M}_{AR}^t$ model .......................................... 178
  6.2.1 Hypothesis for the $\mathcal{M}_{AR}^t$ model ................. 178
6.3 Missing data formulation and the likelihood function ....... 179
6.4 Priors setup ......................................................... 180
  6.4.1 Priors for $\omega$, $\mu$, and $\tau$ ......................... 180
  6.4.2 Priors for $g$, $p$ and $v$ .............................. 181
  6.4.3 Formulae for hyperparameters ....................... 181
6.5 MCMC moves for parameter estimation .......................... 182
  6.5.1 Move I: updating $\omega$ .............................. 182
  6.5.2 Move II: updating $\mu$ .................................. 183
  6.5.3 Move III: updating $\phi$ ................................. 186
  6.5.4 Move IV: updating $\lambda$ and $\tau$ .................... 187
  6.5.5 Move V: updating $z_t$ .................................. 188
  6.5.6 Move VI: updating $\xi$ .................................. 189
6.6 MCMC moves for model determination ............................ 190
  6.6.1 Move VII: updating of $p$ .......................... 190
  6.6.2 Move VIII: updating of $v$ .......................... 191
  6.6.3 Move IX: updating of $g$ .......................... 194
7 Mixture of Student-t ARMA components

7.1 Introduction ......................................................... 204
7.2 The $\mathcal{M}_{\text{ARMA}}^t$ model ................................. 205
7.3 Missing data formulation and the likelihood function ............. 205
  7.3.1 Hypotheses of the model ....................................... 207
7.4 Priors setup ............................................................. 207
  7.4.1 Priors for $\omega$, $\mu$, and $\tau$ .......................... 208
  7.4.2 Priors for $\phi$ and $\theta$ .................................. 208
  7.4.3 Priors for $g$, $p$ and $v$ .................................. 209
  7.4.4 Formulae for hyperparameters ............................... 209
7.5 MCMC moves for parameter estimation .............................. 209
  7.5.1 Move I: updating $z$ ........................................ 210
  7.5.2 Move II: updating $\omega$ .................................. 211
  7.5.3 Move III: updating $\mu$ .................................. 211
  7.5.4 Move IV-A: updating $\phi$ ................................. 213
  7.5.5 Move IV-B: updating $\theta$ ................................. 215
  7.5.6 Move V: updating $\lambda$ and $\tau$ ....................... 216
  7.5.7 Move VI: updating $z_t$ .................................. 217
  7.5.8 Move VII: updating $\xi$ .................................. 218
7.6 MCMC moves for model determination .............................. 218
  7.6.1 Move VIII-A: updating of $p$ ............................. 218
  7.6.2 Move VIII-B: updating of $q$ ............................. 219
  7.6.3 Move VIII-C: updating of $v$ ............................. 221
7.6.4 Move IX: updating of $g$ ................................. 222

7.7 Examples .................................................... 223

8 Mixture of Student-$t$ ARMA-GARCH components 229

8.1 Introduction ................................................ 229

8.2 The $M^t_{\text{ARMA-GARCH}}$ model .......................... 230

8.2.1 Hypotheses for the model ............................... 231

8.3 Missing data formulation and the likelihood function .... 232

8.4 Priors setup ................................................. 234

8.4.1 Priors of $\omega, \mu$ .................................. 234

8.4.2 Priors for $\phi, \theta, \alpha$, and $\beta$ ...................... 234

8.4.3 Priors for $\Lambda$ ......................................... 235

8.4.4 Formulae for hyperparameters .......................... 236

8.5 MCMC moves for parameter estimation ..................... 236

8.5.1 Move I: updating $z$ ..................................... 238

8.5.2 Move II: updating $\omega$ ................................ 238

8.5.3 Move III: updating $\mu$ .................................. 238

8.5.4 Move IV-A: updating $\phi$ ............................... 241

8.5.5 Move IV-B: updating $\theta$ ............................... 243

8.5.6 Move V-A: updating $\alpha$ ............................... 244

8.5.7 Move V-B: updating $\beta$ ............................... 245

8.5.8 Move VI: updating $\xi$ .................................. 246

8.6 MCMC moves for model determination ..................... 247

8.6.1 Move VII-A: updating of $p$ ............................ 248

8.6.2 Move VII-B: updating of $q$ ............................ 249
8.6.3 Move VIII-A: updating of $r$ ........................................... 250
8.6.4 Move VIII-B: updating of $s$ ........................................... 250
8.6.5 Move VIII-C: updating of $v$ ........................................... 250
8.6.6 Move IX: updating of $g$ ................................................. 252
8.7 Examples ................................................................. 253

9 Conclusion ............................................................... 261

A MCMC output ......................................................... 262
List of Tables

3.1 Raftery and Lewis’s diagnostic of samples of parameter when fitting the $\mathcal{M}_{\text{AR}}$ model to the $y_{\text{MoI}}$ series 84
3.2 Acceptance Rate for M-H chains of AR coefficients when fitting the $\mathcal{M}_{\text{AR}}$ model to the $y_{\text{MoI}}$ series 85
3.3 Heidelberger and Welch’s convergence diagnostic of parameters when fitting the $\mathcal{M}_{\text{AR}}$ model to the $y_{\text{MoI}}$ series 85
3.4 Estimation results of parameters when fitting the $\mathcal{M}_{\text{AR}}$ model to the $y_{\text{MoI}}$ series 86
3.5 Estimation results of parameters when fitting the $\mathcal{M}_{\text{AR}}$ model with Sampietro’s setup to the $y_{\text{MoI}}$ series 87
3.6 Estimation results of parameters when fitting the $\mathcal{M}_{\text{AR}}$ model with Sampietro’s setup to the $y_{\text{MoI}}$ series by using Monte Carlo technique 88
3.7 Estimation results of parameters when fitting the $\mathcal{M}_{\text{AR}}$ model to the $y_{\text{MoII}}$ series 98
3.8 Estimation results of parameters when fitting the $\mathcal{M}_{\text{AR}}$ model to the $y_{\text{IBM}}$ data 102
4.1 Acceptance Rate for M-H chains of $\phi$ when fitting the $\mathcal{M}_{\text{ARMA}}$ model to the $y_{\text{MoI}}$ series 129
4.2 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA}}$ model to the $y_{\text{MoI}}$ series 130
4.3 Convergence diagnostics of MCMC sampled values of $\omega$, $\mu$ and $\sigma$
when fitting the $\mathcal{M}_{\text{ARMA}}$ model to the $y\text{MoII}$ series

4.4 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA}}$ model
to the $y\text{MoII}$ series

4.5 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA}}$ model
to the $y\text{IBM}$ data

4.6 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA}}$ model
to the $y\text{DGS3}$ data

5.1 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}$
model to the $y\text{MoIII}$ series

5.2 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}$
model to the $y\text{DGS3}$ data

5.3 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}$
model to the $y\text{DGS3}$ data

6.1 Estimation results of parameters while fitting the $\mathcal{M}_{\text{AR}}^t$ model to
the $y\text{MoI}$ series

6.2 Convergence diagnostics of sampled values of parameters when fit-
ting the $\mathcal{M}_{\text{AR}}^t$ model to the $y\text{DGS3}$ data

6.3 Estimation results of parameters while fitting the $\mathcal{M}_{\text{AR}}^t$ model to
the $y\text{DGS3}$ data

7.1 Estimation results of parameters while fitting the $\mathcal{M}_{\text{AR}}^t$ model to
the $y\text{MoI}$ series

8.1 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}$
model to the $y\text{MoIII}$ series

8.2 Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}$
model to the $y\text{DGS3}$ data
A.1 Acceptance Rate for M-H chains of $\phi$ when fitting the $\mathcal{M}_{AR}^t$ model to the yMoI series ........................................ 262
A.2 The Geweke diagnostic results of the $\mathcal{M}_{ARMA}$ model fitted with the yMoI series ........................................ 263
A.3 Estimation results of parameters while fitting the $\mathcal{M}_{AR}$ model to the yDGS3 data ........................................ 264
A.4 The Heidelberger-Welch diagnostic results of the parameter while fitting the $\mathcal{M}_{ARMA}$ model to the yMoI series ........................................ 267
A.5 The Raftery-Lewis diagnostic results of parameters while fitting the $\mathcal{M}_{ARMA}$ model to the yMoI series ........................................ 268
A.6 Diagnostic results of the parameter $\sigma$ while fitting the $\mathcal{M}_{AR}^t$ model to the simulated yMoI series ......................... 269
A.7 Diagnostic results of the parameter $\gamma$ while fitting the $\mathcal{M}_{ARMA}$ model to the simulated yMoI series ......................... 270
A.8 Diagnostic results of the parameter $\sigma$ while fitting the $\mathcal{M}_{ARMA}$ model to the simulated yMoI series ......................... 271
A.9 Diagnostic results of the parameter $\theta$ while fitting $\mathcal{M}_{ARMA}$ to the simulated series yMoI ................................. 272
A.10 Diagnostic results of the parameter $\phi$ while fitting $\mathcal{M}_{AR}^t$ to the simulated series yMoI ................................. 273
A.11 Diagnostic results of the parameter $\omega$ while fitting $\mathcal{M}_{ARMA}^t$ to the simulated yMoI series ................................. 274
A.12 Diagnostic results of the parameter $\gamma$ while fitting $\mathcal{M}_{ARMA}^t$ model to the simulated yMoI series ................................. 275
A.13 Diagnostic results of the parameter $\sigma$ while fitting $\mathcal{M}_{ARMA}^t$ model to the yMoI series ................................. 276
A.14 Diagnostic results of the parameter $\phi$ while fitting $\mathcal{M}_{AR}^t$ to the yMoI series ................................. 277
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Directed Acyclic Graph of the $\mathcal{M}_\text{AR}$ model</td>
<td>66</td>
</tr>
<tr>
<td>3.2</td>
<td>Graph of the first-proposed probability functions for updating AR order</td>
<td>76</td>
</tr>
<tr>
<td>3.3</td>
<td>Graph of the second-proposed probability functions for updating AR order</td>
<td>77</td>
</tr>
<tr>
<td>3.4</td>
<td>Graphical analysis of the simulated series $\text{yMoI}$</td>
<td>92</td>
</tr>
<tr>
<td>3.5</td>
<td>MCMC output of all parameters when fitting the $\mathcal{M}_\text{AR}$ model to the $\text{yMoI}$ series without label-switched.</td>
<td>93</td>
</tr>
<tr>
<td>3.6</td>
<td>MCMC output of all parameters when fitting the $\mathcal{M}_\text{AR}$ model to the $\text{yMoI}$ series with label-switched.</td>
<td>94</td>
</tr>
<tr>
<td>3.7</td>
<td>MCMC output of parameters when fitting the $\mathcal{M}_\text{AR}$ model to the $\text{yMoI}$ series without label-switched.</td>
<td>95</td>
</tr>
<tr>
<td>3.8</td>
<td>MCMC output of parameters when fitting the $\mathcal{M}_\text{AR}$ model to the $\text{yMoI}$ series with label-switched.</td>
<td>96</td>
</tr>
<tr>
<td>3.9</td>
<td>Graphical analysis of the simulated series $\text{yMoII}$</td>
<td>97</td>
</tr>
<tr>
<td>3.10</td>
<td>MCMC output of parameters when fitting the $\mathcal{M}_\text{AR}$ model to the $\text{yMoII}$ series</td>
<td>99</td>
</tr>
<tr>
<td>3.11</td>
<td>MCMC output of parameters when fitting the $\mathcal{M}_\text{AR}$ model to the $\text{yMoII}$ series with label-switched.</td>
<td>100</td>
</tr>
<tr>
<td>3.12</td>
<td>Properties of the $\text{yIBM}$ series</td>
<td>101</td>
</tr>
</tbody>
</table>
3.13 MCMC output of parameters when fitting the $\mathcal{M}_{AR}$ model to the $y_{IBM}$ data without label-switched ........................................ 103
3.14 MCMC output of parameters when fitting the $\mathcal{M}_{AR}$ model to the $y_{IBM}$ data without label-switched ........................................ 104
3.15 Properties of the $y_{DGS3}$ series ............................................. 105
3.16 MCMC output of parameters $\omega$, $\mu$, $\gamma$ and $\sigma$ when fitting the $\mathcal{M}_{AR}$ model to the $y_{DGS3}$ data without label-switched ........................................ 106
3.17 MCMC output of parameters $\omega$, $\mu$, $\gamma$ and $\sigma$ when fitting the $\mathcal{M}_{AR}$ model to the $y_{DGS3}$ data with label-switched ........................................ 107

4.1 Directed Acyclic Graph of the $\mathcal{M}_{ARMA}$ model ............................................. 115
4.2 MCMC output of parameters $\omega$, $\gamma$ and $\sigma$ when fitting the $\mathcal{M}_{ARMA}$ to the $y_{MoII}$ series ............................................. 133
4.3 MCMC output of $\phi$ and $\theta$ when fitting the $\mathcal{M}_{ARMA}$ model to the $y_{MoII}$ series ............................................. 134
4.4 MCMC output of parameters when fitting the $\mathcal{M}_{ARMA}$ model to the $y_{IBM}$ data ............................................. 137
4.5 MCMC output of parameters while fitting the $\mathcal{M}_{ARMA}$ model to the $y_{DGS3}$ data ............................................. 138

5.1 Directed Acyclic Graph of the $\mathcal{M}_{ARMA-GARCH}$ model ............................................. 149
5.2 Graphical analysis of the simulated series $y_{MoIII}$ ............................................. 171
5.3 MCMC output of parameters when fitting the $\mathcal{M}_{ARMA-GARCH}$ model to the $y_{MoIII}$ series ............................................. 172
5.4 Comparison between the MCMC output of $\omega$ and $\mu$ when fitting the $\mathcal{M}_{ARMA-GARCH}$ model to the $y_{DGS3}$ data with different $N$ ............................................. 173
5.5 Comparison between the MCMC output of $\phi$ when fitting the $\mathcal{M}_{ARMA-GARCH}$ model to the $y_{DGS3}$ data with different $N$ ............................................. 174
5.6 Comparison between the MCMC out of $\theta$ when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}$ model to the $y_{DGS3}$ data with different $N$ .................... 175

5.7 Comparison between the MCMC outputs of $\alpha$ and $\beta$ coefficients when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}$ model to the $y_{DGS3}$ data with different $N$ ................................. 176

6.1 Directed Acyclic Graph of a $\mathcal{M}^t_{\text{AR}}$ model ..................... 182

6.2 MCMC output of parameters while fitting the $\mathcal{M}^t_{\text{AR}}$ model to the $y_{MoI}$ series ............................. 201

6.3 MCMC output of parameters while fitting the $\mathcal{M}^t_{\text{AR}}$ model to the $y_{DGS3}$ data ............................. 202

6.4 MCMC output of $\phi$ while fitting the $\mathcal{M}^t_{\text{AR}}$ model to the $y_{DGS3}$ data ............................. 203

7.1 Directed Acyclic Graph of the $\mathcal{M}^t_{\text{ARMA}}$ model ..................... 210

7.2 MCMC output of parameters while fitting the $\mathcal{M}^t_{\text{ARMA}}$ model to the $y_{MoI}$ series ............................. 225

7.3 MCMC output of $\omega$, $\mu$, and $\sigma$ while fitting the $\mathcal{M}^t_{\text{ARMA}}$ model to the $y_{DGS3}$ data without label-switched and with label-switched ............................. 226

7.4 MCMC output of AR coefficients while fitting the $\mathcal{M}^t_{\text{ARMA}}$ model to the $y_{DGS3}$ data without label-switched and with label-switched ............................. 227

7.5 MCMC output of $\theta$ while fitting the $\mathcal{M}^t_{\text{ARMA}}$ model to the $y_{DGS3}$ data without label-switched and with label-switched ............................. 228

8.1 Directed Acyclic Graph of the $\mathcal{M}^t_{\text{ARMA-GARCH}}$ model ..................... 237

8.2 MCMC output of sampled values of $\varphi$ when fitting the model $\mathcal{M}^t_{\text{ARMA-GARCH}}$ to the $y_{MoIII}$ series ............................. 256

8.3 MCMC output of $\omega$ and $\mu$ when fitting $\mathcal{M}_{\text{ARMA-GARCH}}$ and $\mathcal{M}^t_{\text{ARMA-GARCH}}$ models to the $y_{DGS3}$ data ............................. 257
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.4</td>
<td>MCMC output of $\phi$ when fitting $M_{\text{ARMA-GARCH}}$ and $M_{\text{ARMA-GARCH}}^t$ models to the yDGS3 data</td>
<td>258</td>
</tr>
<tr>
<td>8.5</td>
<td>MCMC output of $\theta$ when fitting $M_{\text{ARMA-GARCH}}$ and $M_{\text{ARMA-GARCH}}^t$ models to the yDGS3 data</td>
<td>259</td>
</tr>
<tr>
<td>8.6</td>
<td>MCMC output of $\alpha$ and $\beta$ coefficients when fitting $M_{\text{ARMA-GARCH}}$ and $M_{\text{ARMA-GARCH}}^t$ models to the yDGS3 data</td>
<td>260</td>
</tr>
<tr>
<td>A.1</td>
<td>MCMC output of parameters while fitting the $M_{\text{AR}}$ model to the yDGS3 data</td>
<td>265</td>
</tr>
<tr>
<td>A.2</td>
<td>MCMC output of $\varphi$ while fitting the $M_{\text{ARMA}}$ model to the simulated yMoI series</td>
<td>266</td>
</tr>
</tbody>
</table>
Abstract
In this thesis we consider some finite mixture time series models in which each component is following a well-known process, e.g. AR, ARMA or ARMA-GARCH process, with either normal-type errors or Student-$t$ type errors. We develop MCMC methods and use them in the Bayesian analysis of these mixture models. We introduce some new models such as mixture of Student-$t$ ARMA components and mixture of Student-$t$ ARMA-GARCH components with complete Bayesian treatments. Moreover, we use component precision (instead of variance) with an additional hierarchical level which makes our model more consistent with the MCMC moves. We have implemented the proposed methods in R and give examples with real and simulated data.
Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
Notations

The following is a list of notations which we are going to use throughout our thesis.

$\propto$ : proportional to

$\min(a, b)$ : the minimum of $a$ and $b$

acf : autocorrelation function

pdf : probability density function

pacf : partial autocorrelation function

$\mathbb{Z}_+$ : the set of all positive integers

$\mathbb{R}$ : the set of all real numbers

$\mathbb{R}_+$ : the set of all non-negative real numbers

$y$ : observed data (or time series)

$L$ : the likelihood function

$\mathcal{G}_t$ : the past information up to time $t$ and all other variables

$\pi(\varphi)$ : the pdf of the prior distribution of $\varphi$

$\pi(\varphi \mid \cdot)$ : the pdf of the posterior distribution of $\varphi$
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>Autoregressive</td>
</tr>
<tr>
<td>ARMA</td>
<td>Autoregressive Moving average</td>
</tr>
<tr>
<td>ARCH</td>
<td>Autoregressive Conditional Heteroskedastic</td>
</tr>
<tr>
<td>GARCH</td>
<td>Generalized Autoregressive Conditional Heteroskedastic</td>
</tr>
<tr>
<td>$\mathcal{M}_{\text{AR}}$</td>
<td>mixture of Gaussian AR components</td>
</tr>
<tr>
<td>$\mathcal{M}'_{\text{AR}}$</td>
<td>mixture of Student-$t$ AR components</td>
</tr>
<tr>
<td>$\mathcal{M}_{\text{ARMA}}$</td>
<td>mixture of Gaussian ARMA components</td>
</tr>
<tr>
<td>$\mathcal{M}'_{\text{ARMA}}$</td>
<td>mixture of Student-$t$ ARMA components</td>
</tr>
<tr>
<td>$\mathcal{M}_{\text{ARMA-GARCH}}$</td>
<td>mixture of Gaussian ARMA-GARCH components</td>
</tr>
<tr>
<td>$\mathcal{M}'_{\text{ARMA-GARCH}}$</td>
<td>mixture of Student-$t$ ARMA-GARCH components</td>
</tr>
</tbody>
</table>
Acknowledgements

Praise be to Allah, the Most Generous, who created man and provided him with the tools for acquiring knowledge, namely hearing, sight and wisdom. He taught man that which he knew not. With His will I am able to present this doctoral dissertation.

At first, I would like to express my intense gratitude to my supervisor, Dr. Georgi Boshnakov, for his admirable guidance and skillful supervision throughout my 3-year PhD study. His motivating suggestions, ideas and accessibility helped me a lot on the way to the development of this work.

Special thanks go to the Overseas Research Students Awards Scheme (ORS) and School of Mathematics, The University of Manchester for providing me the financial support during my study at The University of Manchester.

I would also like to thank all of my friends and colleagues at the School of mathematics for their encouragements and companionship along the way, particularly Sadek, Zaheer, Sergio, Xue, Mary and Anuar.

Finally, I am indebted to my loving family (particularly to my parents Mohammad Sadar Ali and Mrs. Sajeda Begum and to my kids Abrar Arifin and Samiha Shahadat) for their endless love, support and constant encouragement. Their understanding and patience during my PhD is highly appreciable. This thesis is dedicated to them.
Chapter 1

Introduction

The idea of a mixture distribution comes in a natural way when observations made up of some subgroups, mixed at random in proportion to the relative group sizes but whose individual groups are not defined. The main objective of using mixture distributions is to include a finite or infinite number of components that can describe different characteristic of data. Thus these distributions ease much more careful description of complex systems. Now-a-days we are really in the comfortable position of being able to more closely describe, estimate, predict and infer about such systems. Very good explanations and treatments about finite mixture model can be found in the books written by McLachlan and Peel (2000) and Titterington et al. (1985).

We can find an endless benchmark for assessing new techniques, from the EM algorithm to reversible jump methodology. Due to the formidable opportunity provided by new computational technologies like Markov Chain Monte Carlo (MCMC) algorithms, Bayesian approaches to mixture modelling have attracted great interest among them. In the Bayesian paradigm (see Berger (1985)), unknown parameters have been considered random variables and therefore, probability statements have been made directly about unknown parameters, prior (or expert opinion) to be included in the analysis, and hierarchical descriptions of both local-scale and global features of the model. Moreover, the Bayesian framework also allows the complicated mixture model to be decomposed into a set of simpler structures through the use of auxiliary or latent variables. When the number of components is unknown, it can well be argued that the Bayesian paradigm is the only sensible approach to its estimation (Green and Richardson (1997)).

Mixture of autoregressive model is a generalisation of the standard mixture of normal distributions whose mean are linear functions of the past values of the observed variable. i.e. its component densities can be represented as the conditional distributions of different Gaussian autoregressive models. Wong and Li
(2000) first introduced this model for the modelling of nonlinear time series. They derived stationary conditions and autocorrelation, estimated the parameters by using the EM algorithm and addressed the model selection criterion. Boshnakov (2009) demonstrated that the multi-step conditional distributions of predictors in mixture autoregressive models are tractable analytically and are also mixtures. A characteristic function contains the entire distributional information and the conditional characteristic functions are the natural tools for calculations in this type of model. Mixture periodic (or periodically correlated) autoregressive (MPAR) models are introduced by Shao (2006). The aim of this model is to capture the periodicity feature exhibited by the autocovariance structure of many encountered financial time series with eventual multimodal distributions. There are other mixture models considered in the literature such as mixture autoregressive conditional heteroscedastic model by Wong and Li (2001), Mixtures of ARMA models by Xiong and Yeung (2002), mixture of ARMA-GARCH model by Tang et al. (2003), mixture GARCH model by Xu (1998) and so on.

On the other hand, Bayesian analysis of a mixture autoregressive model has been discussed recently by Sampietro (2006) where they consider the number of mixture components and the order of each AR component as stochastic variable. Also they have considered the stationary conditions on the model parameters and used MCMC method to estimate parameters. Also, Nakatsuma (2000) developed Bayesian analysis of ARMA-GARCH models while Ausín and Galeano (2007) discussed Bayesian estimation of the Gaussian mixture model.

Based on our investigation about mixture models and their associated methodologies, we can conclude that mixture models deserve more Bayesian treatment. Therefore, the main objective of this thesis is to develop Bayesian methods of some (finite) mixture time series models, including the case where the number of component is unknown.

The rest of the thesis is structured as follows:

Chapter 2 provides a brief overview of methods of analysing mixture models in the Bayesian context. It includes an introduction to Bayesian statistics along with MCMC methods, Reversible Jump MCMC methods and their use in a Bayesian treatment of mixture models.

Chapter 3 describes a complete MCMC method of performing a Bayesian analysis of mixture autoregressive models, considering global stationarity instead of local stationarity with some new prior settings (see Boshnakov (2011)). While updating the order of the autoregressive coefficients, we propose two functions of it which are very simple in form and efficient for the job. We also illustrate our techniques with some simulated series as well as with real data.
Chapter 4 is about complete Bayesian analysis of mixture autoregressive moving average component with normal errors. We use partial autocorrelations as a tool to generate autoregressive and moving average coefficients in the prior settings so that they are randomly chosen from the stationarity and invertibility region.

Chapter 5 is devoted to Bayesian analysis of mixture of Gaussian ARMA-GARCH components. A big challenge has been faced here while setting this model for Bayesian analysis. Ultimately, we estimate the GARCH part of each component of the model by converting it into an ARMA process by using the techniques described in Bollerslev (1986). The priors and posteriors are updated accordingly.

Chapter 6 - 8 also describes complete Bayesian analysis of mixture of Student-t AR, mixture of Student-t ARMA (\( \mathcal{M}_{\text{ARMA}}^t \)) and mixture of Student-t ARMA-GARCH (\( \mathcal{M}_{\text{ARMA-GARCH}}^t \)) respectively. We use the same simulated series and real data to justify these three new class of mixture models.

Chapter 9 concludes how all these mixture models fit against the simulated series and the real market data and states further scopes to be explored.

Main contributions

We studied several mixture type time series models with different error distributions and GARCH-type effects. The \( \mathcal{M}_{\text{ARMA}}^t \) and \( \mathcal{M}_{\text{ARMA-GARCH}}^t \) models have not been considered before.

We developed complete MCMC method for all models discussed in this dissertation. To the best of our knowledge, the MCMC approach has not been considered for these models before, except for the mixture of Gaussian AR components (\( \mathcal{M}_{\text{AR}} \)) model.

We improved the complete MCMC method of Sampietro (2006) for the \( \mathcal{M}_{\text{AR}} \) model to use the complete stationarity region of the parameters and considered precision (instead of variance) with an additional hierarchical level.

We proposed a new function of component’s AR order for updating the AR order (increasing or decreasing the order) of the mixture AR models. We used it for the remaining models as well due to its simplicity and effectiveness.

We are preparing the results to be considered for publication.
Chapter 2

Background Theory

2.1 Introduction

Bayesian methods for the analysis of complex statistical models, for example, finite mixture models, become more popular now-a-days due to the simultaneous rapid development of stochastic integration methodology, particularly MCMC techniques such as the Gibbs sampler and Metropolis-Hasting sampler.

In this chapter we will present a brief overview of mixture models as well as some concepts of Bayesian inference. For more details the reader is referred to Bernardo et al. (1994), Gelman et al. (2004), Robert (2007).

2.2 The Bayesian approach

In Bayesian theory, parameters are treated as random quantities and point estimates for parameters are replaced by distributions on the parameter space which represents our knowledge or belief about the value of the parameters. Parameters are estimated by averaging the quantities of interest over the parameter space, weighting by the posterior distribution of the parameters. In general all aspects of the posterior distribution of the parameters are valid quantities for inference. On the other hand, in classical statistics given a parametric model, we assume that the parameters are fixed. We can then use the data to make inference about the parameters, often use the maximum likelihood (ML) estimation. Comparisons between the Bayesian approach and the ML approach can be argued but there are practical advantages to the Bayesian approach (see Smith and Naylor (1987), Douady et al. (2003)).

To find the expectations of some functions in Bayesian methods is sometimes challenging if the joint posterior distribution involves too many parameters. This
is true because the integrals form of the expectations are not generally analytically tractable. So standard numerical approximation methods for integration cannot be applied to give accurate results due to high dimensionality of the parameter space. It is often hard to find the posterior distribution of a model (e.g. MAR) parameters analytically.

2.2.1 Bayes theorem and statistical inference

We consider a data set, denoted by \( y = \{y_1, y_2, \ldots, y_T\} \), which is described by a model (in our case, with a finite mixture model) that contains unknown parameters \( \varphi \). The Bayesian approach of statistical inference assumes that the parameter \( \varphi \) determining the distribution of \( y \) is unknown and uncertain. Thus, we need two probabilistic models:

1. a data model specifying the likelihood: \( \mathcal{L}(\varphi; y) = \prod_{i=1}^{T} \mathcal{L}(y_i \mid \varphi) \) which is regarded as a function of \( \varphi \) and can also be denoted by \( \mathcal{L}(\varphi) \). (Assuming independent data).

2. a prior model specifying the prior distribution: \( \pi(\varphi) \), where the parameters of the prior are called hyperparameters and they are initially assumed to be known.

Our goal is to derive statistical information about \( \varphi \) on the data \( y \). Once the data \( y \) are observed, the distribution of \( \varphi \) is revised and the resultant is sought in the form of the probability density \( \pi(\varphi \mid y) \), which is called the posterior (joint) density of \( \varphi \). From this density we will be then able to calculate the mean, variance and other statistical properties of \( \varphi \).

According to Bayes’ theorem, we have

\[
\pi(\varphi \mid y) = \frac{\pi(\varphi) \mathcal{L}(\varphi; y)}{\int \pi(\varphi) \mathcal{L}(\varphi; y) \, d\varphi}
\]

\[
\propto \pi(\varphi) \mathcal{L}(\varphi; y)
\]

i.e. Posterior \( \propto \) Prior \( \times \) Likelihood

(2.1)

(2.2)

(2.3)

Note that the term \( 1/ \int \pi(\varphi) \mathcal{L}(y \mid \varphi) \, d\varphi \) in (2.1) is called a normalizing constant and (2.3) represents the more compact form which is more acceptable as the analytical calculation of the normalizing constant is not always possible.
In the case of $\phi$ being multivariate parameter, say $\phi = (\phi_1, \ldots, \phi_g)$, it is easy to get the marginal and conditional posterior distributions from the joint posterior density.

(i) The marginal posterior density of $\phi_k$ is:

$$\pi(\phi_k \mid y) = \int \pi(\phi_1, \ldots, \phi_g \mid y) d\phi_{-k},$$

where $\phi_{-k} = (\phi_1, \ldots, \phi_{k-1}, \phi_{k+1}, \ldots, \phi_g)$.

(ii) The full conditional posterior distribution of $\phi_k$ is defined as:

$$\pi(\phi_k \mid \phi_{-k}, y) = \frac{\pi(\phi_1, \ldots, \phi_g \mid y)}{\pi(\phi_{-k} \mid y)} \propto \pi(\phi_1, \ldots, \phi_g \mid y).$$

The name suggests that the distribution is conditional on the all remaining components of the model.

Bayesian inference has another important ingredient which is called the predictive distribution. If $y$ denotes an observed sample and assume $\tilde{y}$ be a unknown observable variable, then the distribution of $\tilde{y}$ conditional on $y$ is called the predictive distribution or posterior predictive distribution and it is equal to:

$$\pi(\tilde{y} \mid y) = \int \pi(\tilde{y}, \phi) d\phi = \int \pi(\tilde{y} \mid \phi, y) \pi(\phi \mid y) d\phi.$$

### 2.2.2 Conjugate prior

In Bayesian probability theory, we denote a statistical model with following notation:

$$y \mid \phi \sim \mathcal{L}(\phi; y)$$

$$\phi \sim \pi(\phi)$$

and then generate the posterior density $\pi(\phi \mid y)$ by using the Bayes’ theorem as discussed in the previous section.

If the posterior distributions $\pi(\phi \mid y)$ are in the same family as the prior probability distribution $\pi(\phi)$, the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood. For example,
the Gaussian family is conjugate to itself (or self-conjugate) with respect to a Gaussian likelihood function: if the likelihood function is Gaussian, choosing a Gaussian prior over the mean will ensure that the posterior distribution is also Gaussian. The concept, as well as the term “conjugate prior”, were first introduced by Diaconis and Ylvisaker (1979).

A conjugate prior is an algebraic convenience, giving a closed-form expression for the posterior, otherwise a difficult numerical integration may be necessary. Further, conjugate priors may give intuition, by more transparently showing how a likelihood function updates a distribution. All members of the exponential family have conjugate priors.

2.3 MCMC methods

The Markov chain Monte Carlo (MCMC) methods are widely used in Bayesian inference at present. In quantitative sciences, often we face the problem of evaluation of integrals of the type

\[ I = \int g(x)dx. \]

One of the solution methods is based on generating random samples from a known distribution and then obtaining the integral shown above by its statistical unbiased estimate, the sample mean. So, the integral can be expressed as

\[ I = \int \left( \frac{g(x)}{f(x)} \right) f(x)dx = \int g^*(x)f(x)dx, \]

where \( f(x) \), the density of a random variable, enable us to easily generate a random sample. The two steps, by which the integral \( I \) can be efficiently estimated, are:

1. Generate \( x^{(1)}, \ldots, x^{(T)} \) from the target distribution with probability density function (p.d.f) \( f(x) \).
2. Calculate the sample the mean

\[ \hat{I} = \frac{1}{T} \sum_{i=1}^{T} \left[ \frac{g(x^{(i)})}{f(x^{(i)})} \right]. \quad (2.4) \]

The above concept was originally adopted by the research team of Metropolis in Los Alamos (Anderson, 1986; Metropolis and Ulam, 1949). For a suitable large generated sample (say, e.g., \( T = 10,000 \)), this approach is very accurate (Gamerman and Lopes, 2006, p. 96).
The above described method is directly applicable to many problems in Bayesian
inference, e.g. for every function of the parameter of interest \( g(\varphi) \), we can calcu-
late the posterior mean and variance by the similar steps as follows:

1. Generate a sample of size \( T : \varphi^{(1)}, \ldots, \varphi^{(T)} \) from the posterior distribution
   \( \pi(\varphi \mid y) \).

2. Calculate the sample mean of \( g(\varphi) \) by
   \[
   \hat{I} = \frac{1}{T} \sum_{i=1}^{T} g(\varphi^{(i)}).
   \]

The main problem in the above procedure is how to generate from the posterior
density \( \pi(\varphi \mid y) \) which, in most cases, is not straightforward.

The method described above refers mostly to univariate distributions and cannot
be used to get samples from any posterior distribution of interest. Simulation
techniques based on Markov chains overcome such problems because of their
generality and flexibility.

Markov chain Monte Carlo techniques enabled quantitative researchers to use
highly complicated models and estimate the corresponding posterior distributions
with accuracy. In this way, MCMC methods have greatly contributed to the
development and propagation of Bayesian theory. Extensive details of the use of
MCMC methods can be found in Gilks et al. (1996).

MCMC techniques are based on the construction of a Markov chain that eventu-
ally converges to the target distribution (called stationary or equilibrium) which,
in our case, is the posterior distribution \( \pi(\varphi \mid y) \). This is the main way to distin-
guish MCMC algorithms from direct simulation methods, which provide samples
directly from the target - posterior distribution. Therefore, MCMC output is a
dependent sample since it is generated from a Markov chain, in contrast to the
output of direct methods, which is an independent sample.

MCMC methods incorporate the notion of an iterative procedure (for this reason
they are frequently called iterative methods) since in every step they produce
values depending on the previous one.

### 2.3.1 Markov chain

A Markov Chain \( \{X_n, n \geq 0\} \) is a stochastic process which satisfies the following
property:
Pr(X_{n+1} \in A \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = Pr(X_{n+1} \in A \mid X_n = x_n),

where $X_n$ denotes the state of the process after $n$ steps. Clearly, the future is independent of the past given the present state of the process. The game *Snakes and Ladder* can be considered as a simple example of a Markov chain. The reader is referred to Grimmett and Stirzaker (1992), Section 6 for more details about the theory of Markov Chains.

**Remark 2.3.1.** In most examples of MCMC, we assume that the state space (set of possible values) of the parameters, $\varphi$, will be $\mathbb{R}$ which is uncountable. When the Markov chain is irreducible, aperiodic, and recurrent, as $t \to \infty$ the distribution of $\varphi^{(t)}$ converges to its equilibrium distribution, which is independent of the initial values of the chain $\varphi^{(0)}$, is the posterior distribution of our interest.

### 2.3.2 The MCMC algorithm

In order to generate a sample from $\pi(\varphi \mid y)$, we must construct a Markov chain with two desired properties:

- $\pi \left( \varphi^{(t+1)} \mid \varphi^{(t)} \right)$ should be “easy to generate from”, and
- the equilibrium distribution of the selected Markov chain must be the posterior distribution of interest $\pi(\varphi \mid y)$.

Assuming that we have constructed a Markov chain with these requirements, we then

1. Select an initial value $\varphi^{(0)}$.
2. Generate $N$ MCMC runs.
3. Monitor the convergence of the algorithm using convergence diagnostics. If convergence diagnostics fail, we then generate more MCMC runs.
4. Cut off the first $B$ observations as burn-in.
5. Consider $\left\{ \varphi^{(B+1)}, \ldots, \varphi^{(T)} \right\}$ as the sample for the posterior analysis.
6. Plot the posterior distribution (usually focus is on the univariate marginal distributions).
7. Finally, obtain summaries of the posterior distribution (mean, median, standard deviation, quantiles, correlations).
Convergence of the MCMC algorithm

This term refers to whether the algorithm has reached its equilibrium (target) distribution. If this is true, then the generated sample comes from the correct target distribution. Hence, monitoring the convergence of the algorithm is essential for producing results from the posterior distribution of interest.

There are many ways to monitor convergence.

- The simplest way is to monitor the Monte Carlo (MC) error (calculated as described above) since small values of this error will indicate that we have calculated the quantity of interest with precision. Monitoring autocorrelations is also very useful since low or high values indicate fast or slow convergence, respectively.

- A second way is to monitor the trace plots: the plots of the iterations versus the generated values. If all values are within a zone without strong periodicities and (especially) tendencies, then we can assume convergence.

- Another useful plot is produced by depicting the evolution of the ergodic mean of a quantity over the number of iterations. The term ergodic mean refers to the mean value until the current iteration. If the ergodic mean stabilizes after some iterations, then this is an indication of the convergence of the algorithm.

- Another tactic, which is very efficient in practice, is to run multiple chains with different starting points. When we observe that the lines of different chains mix or cross in trace and (or) the ergodic mean plots, then convergence is ensured.

2.4 Convergence diagnostic tests

In this section, we discuss the most commonly used methods used to assess the convergence of MCMC output. A brief explanation of each approach is given in the following sections. Readers are referred to the work of Brooks and Roberts (1998) and Cowles and Carlin (1996) for a more in-depth review and comparison of these methods.

2.4.1 Brooks, Gelman & Rubin convergence diagnostic

The Brooks, Gelman and Rubin convergence diagnostic is appropriate for the analysis of two or more parallel chains, each with different starting values which
are over dispersed with respect to the target distribution. Several methods for generating starting values for the MCMC samplers have been proposed (Gelman and Rubin, 1992; Applegate et al., 1990; Jennison, 1993).

The diagnostic originally proposed by Gelman and Rubin (1992) is based on a comparison of the within and between chain variance for each variable. This comparison is used to estimate the potential scale reduction factor (PSRF) - the multiplicative factor by which the sampling-based estimate of the scale parameter of the marginal posterior distribution might be reduced if the chains were run to infinity. To adjust for the sampling variability in the variance estimates, the correction proposed by Brooks and Gelman (1998) is applied to the PSRF to produce the corrected scale reduction factor (CSRF). If the estimates are approximately equal to one (or, as a rule of thumb, the 0.975 quantile is \( \leq 1.2 \)), the samples may be considered to have arisen from the stationary distribution. In this case, descriptive statistics may be calculated for the combined latter 50% of iterations from all of the chains.

### 2.4.2 Geweke convergence diagnostic

The Geweke’s convergence diagnostic is appropriate for the analysis of individual chains when convergence of the mean of some function of the sampled parameters is of interest. Geweke (1992) proposed a technique for MCMC chains based on a test for equality of the means of the first, usually 10% of a chain and last, 50%, of that chain. There should be a sufficient number of iterations between the two windows to reasonably assume that the two means are approximately independent. In addition, both the means are equal when the samples are drawn from the stationary distribution of the chain and in this circumstances, Geweke’s statistics has an asymptotically standard normal distribution.

The test statistic is a standard Z-score which is the difference between the two means divided by its standard standard error of their difference, where the variance is determined by spectral density at zero. So it takes into account any autocorrelation. As the number of iterations approaches infinity, the Z statistic approaches the \( \mathcal{N}(0,1) \) if the chain has converged. Z values which fall in the extreme tails of the \( \mathcal{N}(0,1) \) suggest that the chain in the first window had not fully converged. It is common practice to conclude that there is evidence against convergence when the p-value is less than 0.05. Otherwise, it can be said that the results of this test do not provide any evidence against convergence. This does not, however, prove that the chain has converged.
2.4.3 Heidelberger and Welch convergence diagnostic

The Heidelberger and Welch convergence diagnostic is appropriate for the analysis of individual chains. Heidelberger and Welch’s (1983) stationarity test is based on Brownian bridge theory and uses the Cramer-von-Mises statistic. If there is evidence of non-stationarity, the test is repeated after discarding the first 10% of the iterations. This process continues until the resulting chain passes the test or more than 50% of the iterations have been discarded. Failure of the chain to pass this test indicates that a longer run of the MCMC sampler is needed in order to achieve convergence. A halfwidth test is performed on the portion of the chain that passes the stationarity test for each variable. Spectral density estimation is used to compute the asymptotic standard error of the mean. If the halfwidth of the confidence interval for the mean is less than a specified fraction (accuracy) of this mean, say 0.1, the halfwidth test indicates that the mean is estimated with acceptable accuracy.

2.4.4 Raftery and Lewis’s diagnostic

The Raftery and Lewis convergence diagnostic is appropriate for the analysis of individual chains. The diagnostic proposed by Raftery and Lewis (1992) tests for convergence to the stationary distribution and estimates the run-lengths (say, $N$) needed to accurately estimate the quantile, say $q$, to within an accuracy of $\pm r$ with probability $p$ is calculated. And it performs separate calculations for each variable within each chain. It prints an error message indicating the minimum length (say $N_{\text{min}}$), of pilot run when the number of iterations in data is too small. The minimum length is actually the required sample size for a MCMC chain with no correlation between consecutive samples. The dependence factor (say $I$), measures the multiplicative increase in the number of iterations needed to reach convergence due to within-chain correlation. Values of $I$ greater than 5.0 indicate strong autocorrelation as well as indicate convergence failure. This situation could appear due to a poor choice of starting value, high posterior correlations or ‘stickiness’ of the MCMC algorithm. In such situations a reparameterization of the model may help. Finally, the diagnostic also calculates the number of ‘burn-in’ iterations to be discarded at the beginning of the chain.

In the next two sections we are going to discuss two most popular MCMC methods such as Metropolis-Hastings algorithm (Metropolis et al. (1953); Hastings (1970)) and Gibbs sampling (Geman et al. (1984)).
2.5 Gibbs sampling

With Gibbs sampling, we can generate sample (random number or draws) from all of the full conditional distributions. This is easy if we can infer that each of these distributions is one of the well-known distributions and we have a random number generator for it.

Gibbs sampling is a special case of Metropolis-Hasting algorithm which is introduced by Geman et al. (1984). Gibbs sampling is a technique to sample the posterior joint distribution. The sample is then used to infer the form of the marginal densities.

1. First, choose initial values for the parameters $\varphi$; that is, set $\varphi^{(0)} = (\varphi_1^{(0)}, \varphi_2^{(0)}, \ldots)$.

2. In the next steps, draw from each full conditional distribution in turn, using the recent updated values obtained from the previous and the current steps.

   i.e. For step $i = 1, 2, \ldots$, simulate

   (a) draw $\varphi_1^{(i)} \sim \pi \left( \varphi_1 \mid \varphi_2^{(i-1)}, \varphi_3^{(i-1)}, \ldots, y \right)$

   (b) draw $\varphi_2^{(i)} \sim \pi \left( \varphi_2 \mid \varphi_1^{(i)}, \varphi_3^{(i-1)}, \ldots, y \right)$

     \[ \vdots \]

   (c) draw $\varphi_k^{(i)} \sim \pi \left( \varphi_k \mid \varphi_1^{(i)}, \varphi_2^{(i)}, \ldots, \varphi_{k-1}^{(i)}, \varphi_{k+1}^{(i-1)}, \ldots, y \right)$

     \[ \vdots \]

3. Repeat step 2 until you have a random number from all full conditional distributions and go to next step.

Remark 2.5.1. The values of the parameters obtained from Gibbs sampling steps form a Markov chain. Under mild regularity conditions, it can be shown that the Markov chain converges to the stationary (joint posterior) distribution $\pi(\varphi_1, \varphi_2, \ldots)$ is guaranteed. In practice, a sample from the posterior joint distribution can be obtained by deleting some early draws and considering the remaining draws to be draws (approximately) from the posterior joint distribution.
2.6 The Metropolis-Hastings (M-H) algorithm

Metropolis et al. (1953) originally formulated the Metropolis algorithm, by introducing the Markov-chain-based simulation methods used in science. Later, Hastings (1970) generalized the original method in what is known as the Metropolis-Hastings algorithm. The latter is considered to be the general formulation of all MCMC methods. Green (1995) further generalized the Metropolis-Hastings algorithm by introducing reversible jump Metropolis-Hastings algorithms for sampling from parameter spaces with different dimensions.

Assume that $\pi(\varphi \mid y)$ is a (target) density function from which we wish to simulate the vector $\varphi$. An MCMC algorithm, called M-H algorithm that generates a sample of draws from $\pi(\cdot)$ proceeds as follows:

- Suppose that $\varphi^{(i)}$ is the current draw in the chain.
- To obtain the next draw $\varphi^{(i+1)}$, first draw a candidate $\varphi^*$ from a suitable density $q(\varphi^{(i)}, \varphi^*)$ which is called the proposal density or the candidate-generating density.
- The candidate draw is now subjected to a further randomization and is accepted with probability

$$
\lambda(\varphi^{(i)}, \varphi^*) = \min \left\{ 1, \frac{\pi(\varphi^*)/q(\varphi^{(i)}, \varphi^*)}{\pi(\varphi^{(i)})/q(\varphi^*, \varphi^{(i)})} \right\}.
$$

Note that successful implementation of the M-H algorithm, with a high acceptance rate of candidate draws, requires a suitable proposal density.

- If $\varphi^*$ is rejected, $\varphi^{(i+1)}$ is set equal to $\varphi^{(i)}$. This process is iterated.
- It should be noted that this procedure does not require the normalizing constant of $\pi(\varphi)$.

Remark 2.6.1. The Metropolis-Hastings algorithm will converge to its equilibrium distribution regardless of what proposal distribution $q$ is selected. Nevertheless, in practice, choice of the proposal is important since poor choices will considerably delay convergence towards the equilibrium distribution.

2.7 Bayesian model selection

In this section, we will discuss the Bayesian techniques of choosing between two models which is mainly based on posterior model probabilities. More precisely, suppose that the competing models are represented by the set $M = \{M_1, M_2, \ldots\}$ which is enumerable and can be indexed by $m \in \Lambda$. 
CHAPTER 2. BACKGROUND THEORY

The posterior distribution of the parameter $\varphi_m$ conditional to the given model $M_m$ is

$$
\pi(\varphi_m \mid y, m) = \frac{\mathcal{L}(\varphi_m, m) \pi(\varphi_m \mid m)}{\int \mathcal{L}(\varphi_m, m) \pi(\varphi_m \mid m) d\varphi_m},
$$

(2.5)

where $\mathcal{L}(\varphi_m, m)$ and $\pi(\varphi_m \mid m)$ denote the likelihood or probability model and the prior distribution of the parameters of model $M_m$ respectively.

The marginal posterior distribution of $m$, used naturally to compare between models, is derived by Bayes’ theorem:

$$
\pi(m \mid y) = \frac{f(y \mid m) \pi(m)}{\sum_{m=1}^{M} f(y \mid m) \pi(m)},
$$

(2.6)

where $\pi(m)$ is a discrete prior for the models and $f(y \mid m)$ is the marginal likelihood defined as

$$
f(y \mid m) = \int \mathcal{L}(\varphi_m, m) \pi(\varphi_m \mid m) d\varphi_m,
$$

and it is noticed that the calculation of the normalizing constant of Equation (2.6) poses the usual problem in terms of analytical tractability, especially with a high number of possible models. It lies at the heart of the model selection and comparison problem since they are used to compute the Bayes factors and, consequently, the posterior odds ratios,

$$
\frac{\pi(m \mid y)}{\pi(m^* \mid y)} = \frac{f(y \mid m)}{f(y \mid m^*)} \times \frac{\pi(m)}{\pi(m^*)}.
$$

Posterior Odds  Bayes Factor  Prior Odds

For more details about Bayesian model selection, see Dellaportas et al. (2002).

2.7.1 Reversible Jump MCMC

Here we present the Reversible Jump Markov Chain Monte Carlo (RJMCMC) introduced by Green (1995). MCMC technique for Bayesian computation moves around the parameter space of one model, while the RJMCMC method moves around the parameter space of a collection of finite number of models.
The algorithm

Let us have a collection of candidate models $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots\}$ which is enumerable and can be indexed by $m \in \Lambda$. Then the posterior distribution of the parameters given the model $\mathcal{M}_m$ is

$$\pi(\varphi_m \mid y, m) \propto \mathcal{L}(\varphi_m, m) \pi(\varphi_m \mid m), \quad (2.7)$$

where $\mathcal{L}(\varphi_m, m)$ is the likelihood of the model and $\pi(\varphi_m \mid m)$ is the prior distribution of the parameters. Now the joint posterior distribution of the parameter $\varphi_m$ and the model index $m$ can be written as

$$\pi(\varphi_m, m \mid y) \propto \pi(m) \pi(\varphi_m \mid m, y). \quad (2.8)$$

Basically, the RJMCMC techniques use Metropolis-Hastings type algorithms that move a simulation analysis between models defined by $(m, \varphi_m)$ to $(m^*, \varphi_{m^*})$ with different defining dimensions $m$ and $m^*$. In addition, it is designed to be reversible so as to maintain detailed balance of a irreducible and aperiodic chain that converges to the correct target measure. (For more details see Green (1995)).

For $t = 1, \ldots, L$, repeat the following steps $1 - 4$ for generating a sample for the model incators, $\{m_i : i = 1, \ldots, L\}$:

(i) If the current model is $\mathcal{M}_m$ to start in i.e. the current state of the Markov chain is $(m, \varphi_m)$, then propose to move to a new model $\mathcal{M}_{m^*}$ with probability $q(m^* \mid m)$.

(ii) Draw a sample $u$ from a proposal density $q(u \mid \varphi_m, m, m^*)$.

(iii) Set $(\varphi_{m^*}, u^*) = g_{m,m^*}(\varphi_m, u)$, where $g_{m,m^*}(\cdot)$ is a bijection between $(\varphi_m, u)$ and $(\varphi_{m^*}, u^*)$, where $u$ and $u^*$ play the role of matching the dimensions of both vectors.

(iv) Calculate the acceptance probability of the new model $\mathcal{M}_{m^*}$ which is $\min\{1, A\}$, where
CHAPTER 2. BACKGROUND THEORY

$$\Omega = \frac{\mathcal{L}(y \mid \varphi_{m^*}, m^*)}{\mathcal{L}(y \mid \varphi_{m}, m)} \times \frac{\pi(\varphi_{m^*} \mid m^*) \pi(m^*)}{\pi(\varphi_{m} \mid m) \pi(m)} \times \frac{q(m \mid m^*) q(u^* \mid \varphi_{m^*}, m^*, m)}{q(m^* \mid m) q(u \mid \varphi_{m}, m, m^*)} \times \left| \frac{\partial g_{m,m^*}(\varphi_{m}, m)}{\partial (\varphi_{m}, u)} \right|. \quad (2.9)$$

Finally \( \pi(m \mid y) \) can be estimated by

$$\hat{\pi}(m \mid y) = \frac{1}{L} \sum_{i=1}^{L} 1_{m}(m_{i}), \quad (2.10)$$

where \( 1_{m}(m_{i}) = 1 \) if \( m = m_{i} \) and 0 otherwise.

**Remark 2.7.1.** One should be cautious while choosing the model proposal probabilities, \( q(m^* \mid m) \), and the proposal densities, \( q(u \mid m, \varphi_{m}, m^*) \), especially in highly parameterized problems.

1. If all parameters of the proposed model are generated from the proposal distribution, then \((u^*, \varphi_{m^*}) = (u, \varphi_{m})\) and the Jacobian in (2.9) is one.

2. If the proposed model \( \mathcal{M}_{m^*} \) is the same as the current model \( \mathcal{M}_{m} \), the RJMCMC algorithm corresponds to the traditional Metropolis-Hastings algorithm (Hastings (1970)).

3. If \( \pi(\theta_{m} \mid y, m) \) is available in closed form for each model \( \mathcal{M}_{m} \), then \( q(u^* \mid \varphi_{m^*}, m^*, m) = \pi(\theta_{m} \mid y, m) \) and then the equation (2.9) reduces to

$$\mathcal{A} = \frac{\mathcal{L}(y \mid m^*)}{\mathcal{L}(y \mid m)} \times \frac{\pi(m^*) \pi(m)}{\pi(m) \pi(m)} \times \frac{q(m \mid m^*)}{q(m^* \mid m)} \times \left| \frac{\partial g_{m,m^*}(\varphi_{m}, m)}{\partial (\varphi_{m}, u)} \right|. \quad (2.11)$$

using the fact that \( f(y \mid \varphi_{m}, m)\pi(\varphi_{m})\pi(m) = \pi(\varphi_{m}, m)f(y \mid m) \). Again, the Jacobian equals one. Moreover, if \( q(m \mid m^*) = q(m^* \mid m) \), the acceptance probability is the minimum between one and the posterior odds ratio from model \( \mathcal{M}_{m^*} \) to the model \( \mathcal{M}_{m} \), i.e. the move is automatically accepted when model \( \mathcal{M}_{m^*} \) has higher posterior probability than model \( \mathcal{M}_{m} \); otherwise the posterior odds ratio determines how likely is to move to a lower posterior probability model.
2.7.2 Carlin and Chib method

Carlin and Chib (1995) introduced an idea for Bayesian model choice, along with an MCMC algorithm that does not suffer from convergence difficulties. We describe their idea briefly below.

Let $\varphi = (\varphi_m, \varphi_{-m})$ be the vector containing all the parameters of all competing model, where $\varphi_{-m}$ contains all the model parameters except the model $M_m$. Then the joint distribution of all random quantities is:

$$
\pi(\varphi, m \mid y) \propto f(y \mid \varphi, m) \pi(\varphi \mid m)\pi(m),
$$

where we assume that

$$
\pi(\varphi \mid m) = \prod_{j=1}^{M} \pi(\varphi_j \mid m),
$$

where $\varphi_m$ are conditionally independent given $m$. The prior $\pi(\varphi_i \mid m)$, for $i \neq m$, is called pseudo prior and it specifies the distribution of the parameters of model $i$ given another model $m$. With these settings the joint distribution becomes:

$$
\pi(\varphi, m \mid y) = f(y \mid \varphi_m, m) \pi(m) \prod_{j=1}^{M} \pi(\varphi_j \mid m).
$$

Carlin and Chib propose a Gibbs sampler where the full conditional posterior distributions are

$$
\pi(\varphi_m \mid \varphi_{-m}, i, y) \propto \left\{ \begin{array}{ll}
f(y \mid \varphi_m, m) \pi(\varphi_m \mid m), & i = m \\
\pi(\varphi_m \mid i), & i \neq m
\end{array} \right. \quad (2.12)
$$

and

$$
\pi(m \mid \varphi, y) \propto f(y \mid \varphi_m, m)\pi(m) \prod_{j=1}^{M} \pi(\varphi_j \mid m). \quad (2.13)
$$

It is noticed that the main difficulty with Carlin and Chib’s Gibbs sampler is the need of evaluating and drawing from the pseudo-prior distributions at each iteration of the MCMC scheme. Also, it may not be possible to sample directly for some of the $\varphi_m$; in such cases, a M-H move may be used. Moreover, the algorithm shows sensitivity with respect to the model prior specifications and for
certain priors the chain does not seem to move between models.

## 2.8 Within-model simulation

Within-model simulation techniques calculate the posterior distribution of the model \( \pi(m \mid y) \) through the estimation of the marginal likelihood \( f(y \mid m) \) for all \( m \) using a sample from the posterior \( \pi(\varphi_m \mid y, m) \). Here we discuss the two important methods to estimate the marginal likelihood using the MCMC outputs, one from Gibbs moves and other from M-H moves.

### 2.8.1 Marginal likelihood from the Gibbs output

Here we present a method, proposed by Chib (1995), to estimate the marginal likelihood \( f(y \mid m) \) using the MCMC output generated by a Gibbs sampler. Since the marginal likelihood is the normalizing constant of the posterior density expression, we can write:

\[
f(y \mid m) = \frac{\mathcal{L}(\varphi_m, m)}{\pi(\varphi_m \mid y, m)} \pi(\varphi_m).
\]  

(2.14)

The equation (2.14) can be regarded as basic marginal likelihood identity. Therefore, we can estimate the marginal likelihood by finding an estimate of the posterior ordinate \( \pi(\varphi^* \mid y, m) \) in a single point \( \varphi^* \) (For estimation efficiency, we choose the point \( \varphi^* \) as a high-density point in the support of the posterior). Hence, we can estimate the marginal likelihood by using logarithm scale which is computationally convenient as follows:

\[
\ln \hat{f}(y \mid m) = \ell(\varphi^*) + \ln \pi(\varphi^* \mid m) - \ln \hat{\pi}(\varphi^* \mid y),
\]  

(2.15)

where \( \hat{\pi}(\varphi^* \mid y) \) is the estimate of \( \pi(\varphi^* \mid y) \) and \( \ell(\varphi^*) \) is the loglikelihood. So, it is clear from equation (2.15) that it does not suffer from any instability problem and it requires evaluation of the loglikelihood function and the prior at \( \varphi^* \).

Suppose the parameter \( \varphi \) is split into \( B \) blocks \( \varphi = (\varphi_1, \ldots, \varphi_B) \). Then the complete set of full conditional is:

\[
\pi(\varphi_k \mid \varphi_{-k}, y, m), \quad k = 1, \ldots, B,
\]  

(2.16)

where \( \varphi_{-k} = (\varphi_1, \ldots, \varphi_{k-1}, \varphi_{k+1}, \ldots, \varphi_B) \).
Let $\mathbf{u}_{k-1} = (\varphi_1, \ldots, \varphi_{k-1})$ and $\mathbf{u}^{k+1} = (\varphi_{k+1}, \ldots, \varphi_B)$. Then the posterior density $\pi(\varphi^* \mid y, m)$ can be factorized as:

$$
\pi(\varphi^* \mid y, m) = \prod_{k=1}^B \pi(\varphi_k^* \mid \mathbf{u}_{k-1}^*, y, m). \quad (2.17)
$$

It is noticed that each term of the equation (2.17) is:

$$
\pi(\varphi_k^* \mid \mathbf{u}_{k-1}^*, y, m) = \int \pi(\varphi_k^*, \mathbf{u}^{k+1} \mid \mathbf{u}_{k-1}^*, y, m) d\mathbf{u}^{k+1}
$$

$$
= \int \pi(\varphi_k^* \mid \mathbf{u}^{k+1}, \mathbf{u}_{k-1}^*, y, m) \pi(\mathbf{u}^{k+1} \mid \mathbf{u}_{k-1}^*, y, m) d\mathbf{u}^{k+1}
$$

and the corresponding Monte Carlo estimator is:

$$
\hat{\pi}(\varphi_k^* \mid \mathbf{u}_{k-1}^*, y, m) = \frac{1}{N} \sum_{i=1}^N \pi(\varphi_k^* \mid \mathbf{u}^{k+1,(i)}, \mathbf{u}_{k-1}^*, y, m), \quad (2.18)
$$

where $\{\mathbf{u}^{k+1}\}$, for $i = 1, \ldots, N$, are samples from $\pi(\mathbf{u}^{k+1} \mid \mathbf{u}_{k-1}^*, y, m)$.

**The algorithm**

The method mentioned above estimates each term of (2.17) by successive Gibbs samplers, i.e. with a decreasing number of full conditionals. The steps are as follows:

1. Set $k = 1$ and sample $\{\mathbf{u}^{k+1}\}$, for $i = 1, \ldots, N$ from $\pi(\mathbf{u}^{k+1} \mid \mathbf{u}_{k-1}^*, y, m)$.

2. Evaluate:

$$
\hat{\pi}(\varphi_k^* \mid \mathbf{u}_{k-1}^*, y, m) = \frac{1}{N} \sum_{i=1}^N \pi(\varphi_k^* \mid \mathbf{u}^{k+1,(i)}, \mathbf{u}_{k-1}^*, y, m).
$$

3. Update $k = k + 1$ and repeat step (1) until $k = B$.

Finally, we have the estimate of the loglikelihood which is:

$$
\ln \hat{f}(y \mid m) = \ell(\varphi^*, m) + \ln \pi(\varphi^* \mid m) - \ln \hat{\pi}(\varphi^* \mid \mathbf{u}_{k-1}^*, y, m). \quad (2.19)
$$
2.8.2 Marginal likelihood from the M-H output

Here, we present a method of estimating the marginal likelihood using the M-H output which is proposed by Chib and Jeliazkov (2001).

First we consider the simple case in which the posterior density is sampled in one block by the M-H algorithm. Let the produced sample is \( \varphi^{(1)}, \ldots, \varphi^{(N)} \), where \( \varphi^{(i)} = (\varphi_1^{(i)}, \ldots, \varphi_k^{(i)}) \), for \( i = 1, \ldots, N \). If we draw a candidate \( \varphi^* \) from a proposal density \( q(\varphi, \varphi^*) \), then the acceptance probability is defined as

\[
A(\varphi, \varphi^*) = \min \left( 1, \frac{\mathcal{L}(\varphi^*, m) \pi(\varphi^*) q(\varphi^*, \varphi)}{\mathcal{L}(\varphi, m) \pi(\varphi) q(\varphi, \varphi^*)} \right).
\]

Since the M-H technique satisfies the detailed balance, so we can write:

\[
A(\varphi, \varphi^*) q(\varphi, \varphi^*) \pi(\varphi | y) = A(\varphi^*, \varphi) q(\varphi^*, \varphi) \pi(\varphi^* | y) \tag{2.20}
\]

for any point \( \varphi^* \). Now integrating both sides of (2.20) with respect to \( \varphi \), we have

\[
\pi(\varphi^* | y) = \frac{\int A(\varphi, \varphi^*) q(\varphi, \varphi^*) \pi(\varphi | y) d\varphi}{\int A(\varphi^*, \varphi) q(\varphi^*, \varphi) d\varphi} = \frac{E_1 [A(\varphi, \varphi^*) q(\varphi, \varphi^*)]}{E_2 [A(\varphi^*, \varphi)]},
\]

where the expectation \( E_1 \) is evaluated with respect to \( \pi(\varphi | y) \) and the expectation \( E_2 \) is with respect to the product \( q(\varphi^*, \varphi) \). The posterior ordinate is then estimated by the Monte Carlo estimator:

\[
\tilde{\pi}(\varphi^* | y) = \frac{1}{N} \sum_{i=1}^{N} A \left( \varphi^{(i)}, \varphi^* \right) q \left( \varphi^{(i)}, \varphi^* \right),
\]

\[
\frac{1}{M} \sum_{j=1}^{M} A \left( \varphi^*, \tilde{\varphi}^{(j)} \right)
\]

where \( \varphi^{(i)} \), for \( i = 1, \ldots, N \), are the samples from the posterior and \( \tilde{\varphi}^{(j)} \), for \( j = 1, \ldots, M \) from \( q(\varphi^*, \varphi) \), given the fixed value \( \varphi^* \).

The marginal likelihood is then estimated by

\[
\ln \hat{f}(y | m) = \ell(\varphi^*) + \ln \pi(\varphi^*) - \ln \tilde{\pi}(\varphi^* | y).
\]
Now we develop the general case in which $\varphi$ is split into $B$ blocks $\varphi = (\varphi_1, \ldots, \varphi_B)$ as in the previous section. Suppose a component-wise M-H algorithm is available and each block is updated with a probability of the form:

$$A(\varphi_k, \varphi^*) = \min \left( 1, \frac{\pi(\varphi^* | \varphi_{-k}, y) q(\varphi^*, \varphi_k)}{\pi(\varphi_k | \varphi_{-k}, y) q(\varphi_k, \varphi^*)} \right)$$

$$= \min \left( 1, \frac{\mathcal{L}(\varphi^*, \varphi_{-k}) \pi(\varphi^* | \varphi_{-k}) q(\varphi^*, \varphi_k)}{\mathcal{L}(\varphi_k, \varphi_{-k}) \pi(\varphi_k | \varphi_{-k}) q(\varphi_k, \varphi^*)} \right).$$

Let $\tilde{U}_{k-1} = (\varphi_1, \ldots, \varphi_{k-1})$ and $\tilde{U}_{k+1} = (\varphi_{k+1}, \ldots, \varphi_B)$. Then the posterior ordinate at a given point $\varphi^*$ can also be decomposed as in (2.17) and each term of it equal to

$$\pi(\varphi_k^* | \tilde{U}_{k-1}, y) = \frac{E_1 [A(\varphi_k, \varphi_k^*) q(\varphi_k, \varphi_k^*)]}{E_2 [A(\varphi_k^*, \varphi_k)]}, \quad (2.21)$$

where the expectation $E_1$ is evaluated with respect to $\pi(\varphi_k, \tilde{U}_{k+1} | \varphi_{k-1}^*, y)$ and the expectation $E_2$ is with respect to the product $\pi(\tilde{U}_{k+1} | \varphi_{k-1}^*, y) q(\varphi_k, \varphi_k^*)$.

Suppose we samples $\{\tilde{U}_{1,(i)}\}$, for $i = 1, \ldots, N_1$, from the available component-wise M-H algorithm. Now to estimate the two integrals in (2.21), the following steps are performed:

**The algorithm**

1. Set $k = 1$ and $\tilde{U}_{k}^* = (\tilde{U}_{k-1}^*, \varphi_k^*)$ and sample $\{\tilde{U}_{k+1,(i)}\}$, for $i = 1, \ldots, N_{k+1}$, from the reduced M-H algorithm with distribution of interest $\pi(\tilde{U}_{k+1} | \tilde{U}_{k}^*, y)$. Also draw $\tilde{\varphi}_k^{(i)}$ from $q(\varphi_k^*, \varphi_k)$.

2. Evaluate the Monte Carlo estimator:

$$\hat{\pi}(\varphi_{k-1}^*, y) = \frac{1}{N_{k+1}} \sum_{i=1}^{N_k} A \left( \varphi_k^{(i)}, \varphi_k^* \right) q \left( \varphi_k^{(i)}, \varphi_k^* \right)$$

$$= \frac{1}{N_{k+1}} \sum_{i=1}^{N+k} A \left( \varphi_k^*, \tilde{\varphi}_k^{(i)} \right)$$

where samples $\varphi_k^{(i)}$ are from $\pi(\varphi_k, \tilde{U}_{k+1} | \tilde{U}_{k-1}^*, y)$ and $\tilde{\varphi}_k^{(i)}$ are from $\pi(\tilde{U}_{k+1} | \tilde{U}_{k}^*, y) q(\varphi_k^*, \varphi_k)$.

3. Set $\tilde{U}_{k+1,(i)} = \tilde{U}_{k+1,(i)}$, for $i = 1, \ldots, N_{k+1}$. 

(4) Update \( k = k + 1 \) and repeat steps (1) to (3) until \( k = B \).

## 2.9 Finite mixture models

Mixture model is such a model in which data or observations \( y = \{y_t\}_{t=1}^T \) are assumed to arise from the mixture distribution. Mathematically, we can represent a mixture of \( g \)-components (\( g \) possibly unknown but finite) model as below:

\[
\begin{align*}
  y_t \mid \mathcal{G}_{t-1} &\sim \sum_{k=1}^{g} \omega_k f(y_t; \varphi_k), \quad g \geq 1 \\
  \omega_k &> 0, \quad (k = 1, \ldots, g); \quad \sum_{k=1}^{g} \omega_k = 1
\end{align*}
\]

where \( \mathcal{G}_t \) is the information set up to time \( t \); \( \omega_k \) is the \( k \)-th mixing weights; \( f \) denotes a distribution chosen from a parametric family of distributions with unknown parameters \( \varphi_k \) and each observation is assumed to be taken from one of \( g \) groups. In particular, if the distributions \( f(y; \varphi) \)'s are normal distributions, then \( \varphi = (\mu, \tau^{-1}) \), where the \( \mu \) is the unknown mean and \( \tau \) is the unknown precision of the data.

**Remark 2.9.1.** Two important points should be noted here:

1. The tail behaviour of a mixture model is always described by one or two of its components and that it therefore reflects the choice of the parametric family \( f(\cdot; \varphi) \).

2. The representation of mixtures as convex combinations of distributions implies that the moments of (2.23) are also convex combinations of the \( m \)-moments of \( f \)'s:

\[
E[y_t^m] = \sum_{k=1}^{g} \omega_k E^f[y_t^m].
\]

### 2.9.1 Missing data formulation and the likelihood

Mixture models demand *missing data formulation* as each observation \( y_k \) is assumed to arise from a specific but unknown (that is, missing) component of the mixture. Therefore, we need to introduce the auxiliary variables \( z = (z_1, \ldots, z_T) \) which identify to which component the observations \( y = (y_1, \ldots, y_T) \) belong.
Mathematically, $z_t$ is assumed to be independent and identically distributed discrete random variables such that

$$\Pr(z_t = k \mid \mathcal{G}_{t-1}) = \omega_k \quad (t = 1, \ldots, T; \ k = 1, \ldots, g).$$

We call the missing data $z$ the **latent variables** and the pair $(y, z)$ the **completed data**. Therefore, the model (2.23) can be rewritten with the missing data formulation:

$$\begin{align*}
y_t & \mid z_t = k, \mathcal{G}_{t-1} \sim f(y_t \mid \varphi_k), \quad (t = 1, \ldots, T; \ k = 1, \ldots, g) \\
\Pr(z_t = k \mid \mathcal{G}_{t-1}) & = \omega_k; \ \omega_k > 0, \quad (k = 1, \ldots, g); \ \sum_{k=1}^{g} \omega_k = 1
\end{align*} \quad (2.25)$$

In other words, observations $y$ conditional on $z$ are assumed to be independent from the model densities:

$$\pi(y_t \mid z_t = k, \mathcal{G}_{t-1}) = f(y_t; \varphi_k) \quad (t = 1, \ldots, T).$$

Integrating out the missing data $z_1, \ldots, z_T$ then yields the model (2.25):

$$\pi(y_t \mid \mathcal{G}_{t-1}) = \sum_{k=1}^{g} \Pr(z_t = k \mid \mathcal{G}_{t-1}) \pi(y_t \mid z_t = k, \mathcal{G}_{t-1})$$

$$= \sum_{k=1}^{g} \omega_k f(y_t; \varphi_k). \quad (2.26)$$

On the other hand, the probability of $z_t = k$ conditional on the corresponding observation $y_t$ is:

$$\pi(z_t = k \mid y_t, \mathcal{G}_{t-1}) \propto \Pr(z_t = k \mid \mathcal{G}_{t-1}) \pi(y_t \mid z_t = k, \mathcal{G}_{t-1})$$

$$\propto \omega_k f(y_t; \varphi_k) \quad (2.27)$$

which gives
\[ \pi(z_t = k \mid y_t, G_{t-1}) = \frac{\omega_k f(y_t \mid \varphi_k)}{\sum_{s=1}^{g} \omega_s f(y_t \mid \varphi_s)}. \] (2.28)

With these settings, the likelihood can be written as:

\[ \mathcal{L}(\varphi \mid y) = \prod_{t=1}^{T} \sum_{k=1}^{g} \omega_k f(y_t \mid \varphi_k) \cdot 1_{(z_t = k)} \]

\[ = \prod_{k=1}^{g} \prod_{t: z_t = k} f(y_t \mid \varphi_k). \] (2.29)

The above form of likelihood of mixture model plays the important role in Bayesian context\(^1\).

### 2.10 Identifiability

The parameter of a distribution, is said to be **identifiable** if the distribution of the data conditional on that parameter determines the value of the parameter. More precisely, two such distributions \( y \mid \varphi_1 \) and \( y \mid \varphi_2 \) are said to be identifiable if \( y \mid \varphi_1 \sim y \mid \varphi_2 \Rightarrow \varphi_1 = \varphi_2 \), where \( y \mid \varphi \) denotes the conditional distribution of the observed data \( y \) given some parameter vector \( \varphi \). Conversely, \( \varphi \) is not identifiable if there exists \( \varphi_1 \neq \varphi_2 \) with \( y \mid \varphi_1 \sim y \mid \varphi_2 \).

Let \( \varphi = (\omega, \theta) \) denote all the parameters of the mixture model defined by (2.23). Also let \( s \) be any permutation of \( 1, \ldots, g \). For a permutation \( s \in S_g \), set of all permutations of \( \{1, \ldots, g\} \), we denote by:

\[ s(\varphi) = \{ (\omega_{s(1)}, \ldots, \omega_{s(g)}), (\theta_{s(1)}, \ldots, \theta_{s(g)}) \}. \] (2.30)

Now it is clear from the structure of mixture model that it is invariant under permutation of the indices of the components. In other words, we can say that the likelihood (2.29) is the same for all permutations of \( \varphi \). This means that we cannot distinguish component 1 from component 2 and so on, because they are exchangeable. But the identifiability feature is very crucial for both Bayesian inference and computational issues. This problem is also known as **label switching problem**.

\(^1\)In particular, it allows to get full form of posterior distributions in Gibbs moves.
In Bayesian analysis, if there is no prior information that distinguishes between the components of the mixture (that is, the joint prior distribution is the same for all permutations of $\varphi$), then the posterior distribution will be similarly symmetric. As a consequence, posterior shows artificial multi-modality, which poses obvious problems in terms of parameter estimations.

A simple technique discussed in the literature to overcome this problem, is to impose an *identifiability constraint* on the parameter, by ordering the mixing weights or component means in a mixture (see Richardson and Green (1997), section 5.4.3). This constraint breaks the symmetry of the prior. In such cases, most mixture models are identifiable (for details, see Titterington et al. (1985) and McLachlan and Peel (2000)).

### 2.10.1 The permutation sampler

A very good technique has been proposed by Frühwirth-Schnatter (2001) involving MCMC estimation of models affected by the label switching problem based on a permutation sampler. The main idea is: each iteration of the MCMC algorithm is concluded by selecting a permutation of the current labelling of the states in such a way that the identifiability constraint is fulfilled. That means, the algorithm doesn’t reject forthwith parameter values when they don’t satisfy the constraint, but it jumps between the various labelling subspaces in a balanced fashion. Since this method is more efficient with respect to the simple rejection, we used it in our analysis.

### 2.10.2 Relabelling

Another solution to overcome the so-called label switching problem proposed by Stephens (2000) which is more straightforward to implement: after the simulations have been completed, impose a reordering constraint which does not face the difficulties liked to a forced ordering of the mixing weight or other quantities discussed in the previous section. Once the simulation output has been reordered, the posterior mean is approximated by the empirical average.

### 2.10.3 A new reordering scheme

A new reordering technique, proposed and used by Marin et al. (2005), is both straightforward and very efficient. The scheme is mainly based on the simulated sample size, say $N$ and the steps are as follows:
(i) Calculate the pivot $\varphi^{(i^*)}$ such that

$$i^* = \arg \max_{i=1,\ldots,N} \pi(\varphi^{(i)} \mid y),$$

that is, a Monte Carlo approximation of the Maximum a Posteriori (MAP) estimator of $\varphi$.

(ii) For $i = 1, \ldots, N$

(a) Compute

$$s_i = \arg \min_{s \in S} \langle s(\varphi^{(i)}), \varphi^{(i^*)} \rangle_{\kappa+1}\gamma,$$

where $\kappa$ is the dimension of the $\varphi$'s and $\langle \cdot, \cdot \rangle_l$ denotes the canonical scalar product of $\mathbb{R}^l$.

(b) Set $\varphi^{(i)} = s_i(\varphi^{(i)})$.

Note that the step (ii) selects the reordering that is the closest to the approximate MAP estimator and thus solves the identifiability problem without requiring a preliminary and most likely unnatural ordering on one of the parameters of the model.

### 2.11 Review of some distributions

#### 2.11.1 Gamma distribution

The gamma distribution is a two-parameter family of continuous probability distributions. It has a shape parameter $\alpha$ and an inverse scale parameter $\beta$. The gamma distribution is frequently a probability model for waiting times; for instance, in life testing, the waiting time until death is a random variable that is frequently modelled with a gamma distribution.

**Definition 2.11.1.** A continuous random quantity $x$ has a gamma distribution with parameters $\alpha$ and $\beta$ ($\alpha > 0, \beta > 0$) if its density function $G(x; \alpha, \beta)$ is

$$G(x; \alpha, \beta) = c \, x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

where

$$c = \frac{\beta^\alpha}{\Gamma(\alpha)}.$$
• Systematic application of the gamma integral

\[
\int_0^\infty x^{\alpha-1}e^{-\beta x} = \frac{\beta^\alpha}{\Gamma(\alpha)}
\]

gives the mean, variance (second central moment) of a gamma distribution:

\[
E[x] = \frac{\alpha}{\beta}, \quad \text{and} \quad V[x] = \frac{\alpha}{\beta^2}.
\]

• If \(\alpha > 1\), there is a unique mode at \((\alpha - 1)/\beta\); if \(\alpha < 1\) there are no modes (the density is unbounded).

• If \(\alpha = 1\), \(x\) is said to have an exponential \(E_x(x; \beta)\) distribution with parameter \(\beta\) and density

\[
E_x(x; \beta) = \beta e^{-\beta x}, \quad x \geq 0.
\]

• By considering the transformed random quantities \(y = a + x\) or \(z = b - x\), where \(x\) has a \(G(x; \alpha, \beta)\) density, the gamma distribution can be generalised to the ranges \((a, \infty)\) or \((-\infty, b)\).

• The sum of \(k\) independent gamma random quantities with parameters \((\alpha_i, \beta)\), \(i = 1, \ldots, k\), is a gamma random quantity with parameters \(\alpha_1 + \cdots + \alpha_k\) and \(\beta\).

2.11.2 The Inverted-Gamma distribution

**Definition 2.11.2.** A continuous random quantity \(x\) has an inverted-gamma distribution with parameters \(\alpha\) and \(\beta\) (\(\alpha > 0, \ \beta > 0\)) if its density function \(G^{-1}(x; \alpha, \beta)\) is

\[
G^{-1}(x; \alpha, \beta) = c x^{-(\alpha+1)}e^{-\beta/x}, \quad x > 0,
\]

where

\[
c = \frac{\beta^\alpha}{\Gamma(\alpha)}.
\]

• Systematic application of the gamma integral

\[
\int_0^\infty x^{-(\alpha+1)}e^{-\beta/x} = \frac{\beta^\alpha}{\Gamma(\alpha)}
\]
yields:

\[ E[x] = \frac{\beta}{\alpha - 1}, \quad \alpha > 1 \quad \text{and} \quad V[x] = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}, \quad \alpha > 2. \]

\begin{itemize}
  \item There is a unique mode at \( \frac{\beta}{\alpha + 1} \).
  \item The name of this distribution derives from the fact that if \( x \) has a \( G(x; \alpha, \beta) \) density then \( y = x^{-1} \) has an \( G^{-1}(x; \alpha, \beta) \).
  \item A continuous random quantity \( y \) has a square-root inverted-gamma density, \( G^{-1}(\beta/2)(y; \alpha, \beta) \), if \( x = y^{-2} \) has a \( G(x; \alpha, \beta) \) density.
\end{itemize}

### 2.11.3 Beta distribution

The beta distribution is a family of continuous probability distributions defined on the interval \((0, 1)\) parameterized by two positive shape parameters, typically denoted by \( \alpha \) and \( \beta \). The beta distribution can be suited to the statistical modeling of proportions in applications where values of proportions equal to 0 or 1 do not occur.

We list the density and second order properties of beta distribution as follows:

**Definition 2.11.3.** A continuous random quantity \( x \) has a beta distribution with parameters \( \alpha \) and \( \beta \) \((\alpha > 0, \beta > 0)\) if its density function \( B(x; \alpha, \beta) \) is

\[ B(x; \alpha, \beta) = c \ x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1 \]

where

\[ c = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \]

and \( \Gamma(z) \) stands for the Gamma function is defined as:

\[ \Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt. \]

*Integer and half-integer values of the gamma function are easily found from the recursive relation \( \Gamma(n+1) = n\Gamma(n) \), and the values \( \Gamma(1) = 1 \) and \( \Gamma(1/2) = \sqrt{\pi} \approx 1.7725 \).*

- Systematic application of the beta integral,

\[ \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \]
gives
\[ E[x] = \frac{\alpha}{\alpha + \beta} \text{ and } V[x] = \frac{\alpha \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \]

- If \( \alpha > 1 \) and \( \beta > 1 \), there is a unique mode at \((\alpha - 1)/(\alpha + \beta - 2)\).
- If \( x \) has a \( B(x; \alpha, \beta) \) density, then \( y = 1 - x \) has a \( B(x; \beta, \alpha) \) density.
- If \( \alpha = \beta = 1 \), \( x \) is said to have a uniform distribution \( U(x; 0, 1) \) on \((0, 1)\).

### 2.11.4 Generalized beta distribution

If a random variable \( X \) has a \( B(x; \alpha, \beta) \) density, it can be generalized to any finite interval \((a, b)\) by considering the following transformation:

\[ y = a + (b - a)x. \]

Alternatively, a continuous random quantity \( x \) has a generalised beta distribution on \((a, b)\) if it’s density function \( B_{(a,b)}(x; \alpha, \beta) \) is:

\[ B_{(a,b)}(x; \alpha, \beta) = c(x - a)^{\alpha - 1}(b - x)^{\beta - 1}, \quad a < x < b \]

where

\[ c = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)(b - a)^{\alpha + \beta - 2}}. \]

Therefore, beta distribution would be suitable for the statistical modelling of proportions in applications where values of proportions equal to 0, \( a \) or \( b \) do not occur.

### 2.11.5 The Normal distribution

A continuous random quantity has a normal distribution with parameters \( \mu \) (mean) and \( \tau \) (precision) \((\mu \in \mathbb{R}, \tau > 0)\) if its density function is

\[ \mathcal{N}(x; \mu, \tau^{-1}) = \frac{\tau^{1/2}}{\sqrt{2\pi}} \exp \left[ -\frac{\tau}{2} (x - \mu)^2 \right], \quad x \in \mathbb{R}. \]
CHAPTER 2. BACKGROUND THEORY

The distribution is symmetrical about \( x = \mu \). The mean and mode are \( E[x] = M[x] = \mu \) and the variance is \( V[x] = \tau^{-1} \) of the distribution.

- If \( \mu = 0, \tau = 1, x \) is said to have a standard normal distribution, with distribution function \( \phi \) given by
  \[
  \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left\{ -\frac{1}{2} t^2 \right\} dt.
  \]

- If \( y = \tau^{1/2}(x - \mu) = (x - \mu)/\sigma \), where \( x \) has a normal density \( \mathcal{N}(x; \mu, \tau) \), then \( y \) has a \( \mathcal{N}(y; 0, 1) \) (standard) density.

- In general, if \( y = a + \sum_{i=1}^{k} b_i x_i \), where the \( x_i \) are independent with \( \mathcal{N}(x_i; \mu_i, \tau_i) \) densities, then \( y \) has a normal density, \( \mathcal{N}(y; a + \sum_{i=1}^{k} b_i \mu_i, \tau) \) where \( \tau = \left( \sum_{i=1}^{k} b_i/\tau_i \right)^{-1} \), a weighted harmonic mean of the individual precisions.

- If \( x_1, \ldots, x_k \) are mutually independent standard normal random quantities, then \( z = \sum_{i=1}^{k} x_i^2 \) has a (central) \( \chi^2_k \) distribution.

2.11.6 Truncated Normal distribution

In general, if \( x \) is a continuous random variable with pdf \( \pi(x) \), then the density of \( x \) can be truncated between two specified values \( l \) and \( u \) by:

\[
\pi(x; l < x < u) = \frac{\pi(x)}{\Pr(l < x < u)}.
\]

In particular, the density of the truncated normal distribution with mean \( \mu \) and precision \( \tau \) can be denoted and defined by:

\[
\mathcal{N}_{(a,b)}(x; \mu, \tau) = \frac{\mathcal{N}(x; \mu, \tau)}{\Phi_{\mathcal{N}}(b; \mu, \tau) - \Phi_{\mathcal{N}}(a; \mu, \tau)},
\]

where \( \Phi_{\mathcal{N}} \) is the cumulative normal distribution function.
2.11.7 The Student-\textit{t} distribution

A continuous random quantity \( x \) has a Student-\textit{t} distribution with parameters \( \mu (\in \mathbb{R}), \tau (> 0) \) and \( \nu (> 0) \) if its density \( S_t(x; \mu, \tau^{-1}, \nu) \) is defined as:

\[
S_t(x; \mu, \tau^{-1}, \nu) = c \left[ 1 + \frac{\tau}{\nu} (x - \mu)^2 \right]^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}
\]  

(2.31)

where

\[
c = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right) \left( \frac{\tau}{\nu} \right)^{1/2}}.
\]

The distribution is symmetrical about \( x = \mu \), and has a unique mode \( M[x] = \mu \).

The mean and variance are

\[
E[x] = \mu, \quad \text{if} \ \nu > 1,
\]

\[
V[x] = \frac{1}{\tau (\nu - 2)}, \quad \text{if} \ \nu > 2.
\]

The parameter \( \tau \) and \( \nu \) are usually referred to as the precision and the degrees of freedom of the distribution respectively.

- The distribution is generated by the mixture

\[
S_t(x; \mu, \tau^{-1}, \nu) = \int_{0}^{\infty} N \left( \mu, (\tau \xi)^{-1} \right) G \left( \xi; \frac{\nu}{2}, \frac{\nu}{2} \right) d\xi,
\]

and includes the normal distribution as a limiting case, since

\[
N \left( x; \mu, \tau^{-1} \right) = \lim_{\nu \to \infty} S_t(x; \mu, \tau^{-1}, \nu).
\]

- If \( y = \tau^{1/2}(x - \mu) \), where \( x \) has a \( S_t(x; \mu, \tau^{-1}, \nu) \) density, then \( y \) has a student density \( S_t(y; 0, 1, \nu) \). If \( \nu = 1 \), \( x \) is said to have a Cauchy distribution, with density \( Ca(x; \mu, \tau^{-1}) \).

- If \( x \) has a normal density \( N(x; 0, 1) \) and \( y \) has a \( \chi^2(\nu) \), and \( x \) and \( y \) are mutually independent, then

\[
z = \frac{x}{(y/\nu)^{1/2}}
\]

has a standard Student density \( S_t(z; 0, 1, \nu) \).
2.11.8 Multinomial distribution

If the random variable $X$ follows the binomial distribution with parameters $n$ and $p$, then the probability of getting exactly $k$ successes in $n$ trials is given by the probability mass function:

$$
\text{Bi}(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \text{ for } k = 0, 1, 2, \ldots, n,
$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is called the binomial coefficient.

The multinomial distribution is a generalization of the binomial distribution. Let a discrete random vector $\mathbf{x} = (x_1, \ldots, x_k)$ follows a multinomial distribution with parameters $n$ and $\mathbf{p} = (p_1, \ldots, p_k)$, where $n > 0$, $p_i \geq 0$ for $i = 1, \ldots, k$ and $\sum p_i = 1$; $x_i$ denotes the number of times outcome number $i$ was observed over the $n$ trials with probability $p_i$. We denote this distribution by $\mathcal{M}_k(p_1, \ldots, p_k)$ and its probability mass function is defined as:

$$
\mathcal{M}_k(p_1, \ldots, p_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}, \text{ when } \sum_{i=1}^k x_i = 1
$$

for non-negative integers $x_1, \ldots, x_k$.

2.11.9 The Dirichlet distribution

The Dirichlet distribution is a family of continuous multivariate probability distributions parametrized by a vector of positive reals. It is the multivariate generalization of the beta distribution. Dirichlet distributions are very often used as prior distributions in Bayesian statistics, and in fact the Dirichlet distribution is the conjugate prior of the multinomial distribution.

**Definition 2.11.4.** A continuous random vector $\mathbf{x} = (x_1, \ldots, x_k)$ has a Dirichlet distribution of dimension $k$, with parameters $\mathbf{\delta} = (\delta_1, \ldots, \delta_{k+1}) \quad (\delta_i > 0, i = 1, \ldots, k+1)$ if its probability density $\mathcal{D}_k^i(\mathbf{x}; \alpha), 0 < x_i < 1, \text{ and } x_1 + \cdots + x_k < 1$, is

$$
\mathcal{D}_k^i(\mathbf{x}; \mathbf{\delta}) = c \prod_{i=1}^k x_i^{\delta_i - 1} \left(1 - \sum_{i=1}^k x_i\right)^{\delta_{k+1} - 1},
$$

where
CHAPTER 2. BACKGROUND THEORY

\[ c = \frac{\Gamma \left( \sum_{i=1}^{k+1} \delta_i \right)}{\prod_{i=1}^{k+1} \Gamma (\delta_i)}. \]

- In particular case, if \( k = 1, \mathcal{D}_k^i(x; \delta) \) reduces to the beta density \( \mathcal{B}(x; \delta_1, \delta_2). \)

- The mean vector and covariance matrix are given by

\[
E[x_i] = \frac{\delta_i}{\sum_{j=1}^{k+1} \delta_j}; \quad V[x_i] = \frac{E[x_i](1 - E[x_i])}{1 + \sum_{j=1}^{k+1} \delta_j}; \quad \text{Cov}[x_i, x_j] = \frac{-E[x_i]E[x_j]}{1 + \sum_{j=1}^{k+1} \delta_j}.
\]

- If \( \delta_i > 1, i = 1, \ldots, k; \) there is a mode given by

\[
M[x_i] = \frac{\delta_i - 1}{\sum_{j=1}^{k+1} \delta_j - k - 1}.
\]

- The marginal distribution of \( x^m = (x_1, \ldots, x_m), m < k, \) is the Dirichlet

\[
p(x^{(m)}) = \mathcal{D}_m^i \left( x^{(m)}; \delta_1, \ldots, \delta_m, \sum_{j=m+1}^{k+1} \delta_j \right).
\]
Chapter 3

Mixture of Gaussian AR components

3.1 Introduction

In this chapter, we represent a Bayesian analysis of a finite mixture of autoregressive (AR) components with normal errors. This model is known as “mixture autoregressive model” in the literature and was first introduced by Wong and Li (2000). We prefer to call this model “mixture of Gaussian AR component” and denote this class of model by $\mathcal{M}_{\text{AR}}$ hereafter. The analysis is complete in Bayesian sense because all the parameters of the model, including the number of mixture components and the autoregressive order of each component have been considered as stochastic variables. We use the MCMC and Reversible Jump MCMC (RJMCMC) techniques for parameter estimation and model selection.

Wong and Li (2000) estimated the $\mathcal{M}_{\text{AR}}$ model parameters by performing a numerical maximum likelihood estimation (MLE) method. Very recently, Sampietro (2006) developed a fully Bayesian analysis of the $\mathcal{M}_{\text{AR}}$ model in which parameter estimation and model selection were performed using MCMC.

Our approach is similar to Sampietro (2006) but we consider component precision with an additional hierarchical level, which is considered as the natural choice in the Bayesian literature, instead of component variances of the mixture. Sampietro (2006) considered a sufficient stationarity condition for the $\mathcal{M}_{\text{AR}}$ model while we consider a necessary and sufficient stationarity condition for that (see Boshnakov (2011)). More precisely, we consider a direct prior for the AR coefficients of the $\mathcal{M}_{\text{AR}}$ model while Sampietro (2006) proposed indirect prior based on the partial autocorrelations for that of the model. This setting allows us to choose a mixture of a nonstationary AR component and a stationary AR component which results in overall stationary process (see Wong and Li (2000)). Therefore, the priors and
the posterior distributions have been set and developed correspondingly. Moreover, we propose two specific forms of the birth function for updating AR orders of the model with illustrations in Bayesian context.

Finally, the label switching problem is solved by imposing an identifiability constraint on the parameter space. We illustrate our methods with the simulated series as well as with the real data.

3.2 Background

An autoregressive (AR) process \( \{y_t\} \) of order \( p \) is usually denoted by \( \text{AR}(p) \) and is defined as:

\[
y_t = \mu + \sum_{j=1}^{p} \phi_j(y_{t-j} - \mu) + \varepsilon_t, \quad \varepsilon_t \sim \text{iid} \mathcal{N}(0, \tau^{-1}),
\]

where \( \tau \) is the precision of the error term; \( \phi_j \)s are the AR coefficients and \( \mu \) is the mean of the model.

Let \( G(L) = \sum_{i=1}^{p} \phi_i L^i \), where \( L \) is the lag operator. The stationarity of AR process depends on the autoregressive (AR) part of the model. To ensure stationarity of the process we have to impose the following constraint in the AR model:

\((C_S)\): all roots of \( 1 - G(L) = 0 \) are outside the unit circle.

If we denote:

\[
\nu_t = \mu + \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu)
\]

the conditional distribution of \( y_t \) can be written as

\[
y_t \mid G_{t-1} \sim \mathcal{N}(y_t; \nu_t, \tau^{-1}),
\]

where “ \( G_{t-1} \)” is used to denote conditioning on the past observations and on all other variables.
CHAPTER 3. MIXTURE OF GAUSSIAN AR COMPONENTS

3.3 The $\mathcal{M}_{AR}$ model

A process $\{y_t\}$ is said to be a $\mathcal{M}_{AR}$ process if it is a mixture of the following form:

$$
\begin{align*}
\left\{ y_t \mid G_{t-1} \right\} & \sim \sum_{k=1}^{g} \omega_k \mathcal{N}(y_t; \nu_{k,t}, \tau_k^{-1}), \quad (t = 1, \ldots, T), \\
\nu_{k,t} & = \mu_k + \sum_{j=1}^{p_k} \phi_{kj}(y_{t-j} - \mu_k), \quad (k = 1, \ldots, g), \\
\omega_k & > 0, \quad (k = 1, \ldots, g); \quad \sum_{k=1}^{g} \omega_k = 1,
\end{align*}
\tag{3.4}
$$

where " $\mid G_{t-1}$ " is used to denote conditioning on the past observations and on all other variables; $g(\in \mathbb{Z}_+)$ denotes the number of components in the model; $\nu_{k,t}$ indicates the conditional mean of the $k$-component; $\mu_k$ denotes the mean of the $k$th component; $\omega_k$ denotes the mixing weight of the $k$-th component; $\mathcal{N}(x; a, b^{-1})$ denotes the univariate normal density function of $x$ with mean $a$ and precision $b$. For details, see Section 2.11.5.

Note that we can also express $\nu_{k,t}$ as

$$
\nu_{k,t} = \gamma_k + \sum_{j=1}^{p_k} \phi_{kj}y_{t-j}, \quad (k = 1, \ldots, g),
$$

which is equivalent to expression in (3.4) as the mixture shifts $\gamma_k$ is equal to $\mu_k(1 - \phi_{k1} \cdots - \phi_{kp_k})$. So, there is a one-to-one relation between $\mu_k$ and $\gamma_k$ and clearly, $\gamma_k = 0$ when $\mu_k = 0$.

We denote this model by $\mathcal{M}_{AR}(g; p_1, \ldots, p_g)$. To make the $\mathcal{M}_{AR}$ model notationally simple, we can assume $p = \max\{p_1, \ldots, p_g\}$ by setting $\phi_{kj} = 0$, if $j > p_k$ for $k = 1, \ldots, g$.

The following vectors will be used throughout.

- $\omega = (\omega_1, \ldots, \omega_g)$, denotes the mixing weight vector of the model;
- $\mu = (\mu_1, \ldots, \mu_g)$, denotes the mean vector of the model;
- $\tau = (\tau_1, \ldots, \tau_g)$, denotes the precision vector of the error term of the model;
- $\phi = (\phi_1, \ldots, \phi_g)$ where $\phi_k = (\phi_{k,1}, \ldots, \phi_{k,p_k})$, for $k = 1, \ldots, g$, denotes the AR coefficients of $k$-th component;
- $p = (p_1, \ldots, p_g)$, as the vector of orders of the AR part of the model.

The vector of all parameters will be denoted by $\varphi$, i.e. $\varphi = (\omega, \mu, \tau, \phi)$. 
3.3.1 Mean and variance of the $\mathcal{M}_{\text{Ar}}$ model

The conditional expectation and conditional variance of $y_t$ given the past information has been derived by Wong and Li (2000). We discuss these in details as follows:

Firstly, we have from the $\mathcal{M}_{\text{Ar}}$ model (3.4):

$$E(y_t \mid G_{t-1}) = \sum_{k=1}^{g} \omega_k \nu_{k,t},$$

(3.5)

which depends on the past values of the time series.

And, secondly we have

$$\text{Var}(y_t \mid G_{t-1}) = \sum_{k=1}^{g} \omega_k \tau_{k}^{-1} + \sum_{k=1}^{g} \omega_k \left( \nu_{k,t} - \sum_{s=1}^{g} \omega_s \nu_{s,t} \right)^2.$$

(3.6)

But

$$\sum_{k=1}^{g} \omega_k \left( \nu_{k,t} - \sum_{s=1}^{g} \omega_s \nu_{s,t} \right)^2$$

$$= \sum_{k=1}^{g} \omega_k \nu_{k,t}^2 + \left( \sum_{k=1}^{g} \omega_k \right) \left( \sum_{s=1}^{g} \omega_s \nu_{s,t} \right)^2 - 2 \left( \sum_{k=1}^{g} \omega_k \nu_{k,t} \right) \sum_{s=1}^{g} \omega_s \nu_{s,t}$$

$$= \sum_{k=1}^{g} \omega_k \nu_{k,t}^2 + \left( \sum_{s=1}^{g} \omega_s \nu_{s,t} \right)^2 - 2 \left( \sum_{s=1}^{g} \omega_s \nu_{s,t} \right)^2,$$

since $\sum_{k=1}^{g} \omega_k = 1$

$$= \sum_{k=1}^{g} \omega_k \nu_{k,t}^2 - \left( \sum_{k=1}^{g} \omega_k \nu_{k,t} \right)^2.$$

Therefore,

$$\text{Var}(y_t \mid G_{t-1}) = \sum_{k=1}^{g} \omega_k \tau_{k}^{-1} + \sum_{k=1}^{g} \omega_k \nu_{k,t}^2 - \left( \sum_{k=1}^{g} \omega_k \nu_{k,t} \right)^2.$$

(3.7)

It is clear from (3.5) and (3.7) that $E(\cdot)$ and $\text{Var}(\cdot)$ are implicitly conditional on $\phi$ and $\tau$ and the shape of the conditional distributions may change from unimodal to multi-modal. Therefore, the conditional expectation of $y_t$ may not be the best predictor of the future values. On the other hand, the conditional variance depends on both the conditional means and variance of the components.
as in equation (3.7). The first term summarises the conditional variance on the past “errors” and the second and the third terms characterise the change of the conditional variance resulting from the differences in the conditional means of the components.

### 3.3.2 Hypothesis for the $M_{AR}$ model

Let $G_k(L) = \sum_{j=1}^{p_k} \phi_{kj} L^j$, $k = 1, \ldots, g$, where $L$ is the lag operator. Since the stationarity of AR process depends only on the autoregressive AR coefficients of the model and so does the $M_{AR}$. To ensure stationarity of $M_{AR}$ model we impose the following constraint:

$(C_S) : \text{all roots of } 1 - G_k(L) = 0, \quad (k = 1, \ldots, g) \text{ are outside the unit circle.}
$

The constraint $(C_S)$ mentioned above is a sufficient condition for stationarity of the $M_{AR}$ model and it has been considered by Sampietro (2006). But we consider the less restrictive necessary and sufficient condition for first and second order stationarity of the $M_{AR}$ model proposed by Boshnakov (2011) which is as follows:

$(C_3^*) : \text{the largest eigenvalue of } \sum_{k=1}^{g} \omega_k A_k \otimes A_k \text{ is smaller than 1,}$

where

$$A_k = \begin{pmatrix} \phi_{k1} & \phi_{k2} & \cdots & \phi_{kp-1} & \phi_{kp} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad (k = 1, \ldots, g).$$

Then the $M_{AR}$ model is said to be *Stable* if the constraint $(C_3^*)$ is satisfied. For more details see Theorem 3 in Boshnakov (2011).

### 3.4 Missing data formulation and the likelihood function

The likelihood function takes a complicated form when time series $\{y_t\}_{t=1}^T$ is given by the $M_{AR}$ model (3.4). However, it can be simplified by introducing the usual
missing data formulation for mixture setups (see e.g. Diebolt and Robert (1994)), where \( \{ z_t \}_{t=1}^T \) is the set of allocation variables such that

\[
z_t = \begin{cases} 
1, & \text{with probability } \omega_1, \\
\vdots & \\
k, & \text{with probability } \omega_k, \\
\vdots & \\
g, & \text{with probability } \omega_g, \\
\end{cases} \quad (t = 1, \ldots, T). 
\]  

(3.8)

That is,

\[
\Pr(z_t = k \mid g, \omega_1, \ldots, \omega_g) = \omega_k, \quad (k = 1, \ldots, g).
\]

Consequently, our observed series \( y = \{ y_t \}_{t=1}^T \) are being completed with a missing data set \( z = \{ z_t \}_{t=1}^T \). More specifically, conditional on \( z_t \)'s, the observations \( y_t \)'s are drawn from their respective individual subpopulation, i.e.

\[
y_t \mid \mathcal{F}_{t-1}, z_t \sim \begin{cases} 
\mathcal{N}(y_t; \nu_{1,t}, \tau_1^{-1}) & \text{if } z_t = 1, \\
\vdots & \\
\mathcal{N}(y_t; \nu_{k,t}, \tau_k^{-1}) & \text{if } z_t = k, \\
\vdots & \\
\mathcal{N}(y_t; \nu_{g,t}, \tau_g^{-1}) & \text{if } z_t = g. \\
\end{cases} 
\]

(3.9)

As a consequence, the full conditional likelihood function separates into \( g \) parts where each one concerning the data assigned to each of the \( g \) mixture components,

\[
\mathcal{L}(\varphi \mid g, y, z) = \prod_{k=1}^g \prod_{t: z_t = k} \mathcal{N}(y_t; \nu_{k,t}, \tau_k^{-1}) \\
= \prod_{k=1}^g \prod_{t: z_t = k} \frac{\sqrt{\tau_k}}{\sqrt{2\pi}} \exp \left[ -\frac{\tau_k}{2} (y_t - \nu_{k,t})^2 \right]. 
\]

(3.10)

3.5 Priors setup

In this section, we specify the prior distribution of the parameters of the \( \mathcal{M}_{AR} \) model by considering the hierarchical approach. We follow the rules proposed by Richardson and Green (1997) and Sampietro (2006) which are based on a careful analysis of the sensitivity of the results to the choice of prior on some real datasets.

When \( g \), the number of components, is fixed, we will have a specific \( \mathcal{M}_{AR} \) model
for which the prior distributions, $\pi(\varphi \mid g)$, of the model parameters are chosen as follows:

- prior for the mixing weight:
  \[
  \omega \mid g \sim \mathcal{D}(\omega; \delta_1, \ldots, \delta_g),
  \]  
  \[
  \text{(3.11)}
  \]
  where $\mathcal{D}(x; \delta_1, \ldots, \delta_g)$ denotes the Dirichlet distribution for $x$ on the simplex:
  \[
  \{(\omega_1, \ldots, \omega_{g-1}, 1 - \omega_1 - \cdots - \omega_{g-1}) : \omega_1 + \cdots + \omega_{g-1} \leq 1, \omega_i > 0\}
  \]
  with density proportional to
  \[
  \omega_1^{\delta_1-1} \times \cdots \times \omega_{g-1}^{\delta_{g-1}-1} \times (1 - \omega_1 - \cdots - \omega_{g-1})^{\delta_g-1}.
  \]

- prior for mean of the $k$-th component:
  \[
  \mu_k \mid g \sim \mathcal{N}(\mu_k; \zeta, \kappa^{-1}),
  \]
  \[
  \text{(3.12)}
  \]
  where $\mathcal{N}(\cdot; \zeta, \kappa^{-1})$ denotes the normal distribution with fixed mean $\zeta$ and fixed precision $\kappa$.

- prior for the precision of the $k$-th component with an additional hierarchical level (see Green and Richardson (1997)):
  \[
  \tau_k \mid \lambda \sim \mathcal{G}(\tau_k; c, \lambda), (k = 1, \ldots, g),
  \]
  \[
  \text{(3.13)}
  \]
  \[
  \lambda \sim \mathcal{G}(\lambda; a, b),
  \]
  where $\mathcal{G}(\cdot; a, b)$ denotes the Gamma distribution for $x$ with mean $a/b$ and variance $a/b^2$.

- we can set the following proper prior for $\phi_k = (\phi_{k1}, \ldots, \phi_{kp_k})$, the AR coefficients of the $k$-th component of the model
  \[
  \phi_{ki} \sim \mathcal{N}(\phi_{ki}; \mu_\phi, \tau_\phi^{-1}) \mathbf{1}_{\mathcal{M}_{AR}}(\text{Stable}), (k = 1, \ldots, g; i = 1, \ldots, p_k),
  \]
  where $\mathbf{1}_{\mathcal{M}_{AR}}(\text{Stable})$ is the indicator function which takes unity if the $\mathcal{M}_{AR}$ model is stable (i.e. it satisfies the constraint $(C^*_S)$ discussed in Section 3.3.2); otherwise zero, and $\mathcal{N}(\cdot)\mathbf{1}_{\mathcal{M}_{AR}}(\text{Stable})$ represents that the normal distribution $\mathcal{N}(\cdot)$ is truncated at the boundary of the support of the indicator function $\mathbf{1}_{\mathcal{M}_{AR}}(\text{Stable})$. In practice, we may choose $\mu_\phi = 0$ and $\tau_\phi^{-1} = 1$. 

so that the value of $\phi_{ki}$ lie between -1.96 and +1.96 and we check the stability of the model by implementing the indicator function $1_{M_{AR}}(\text{Stable})$ which takes unity if the model is stable (see Boshnakov (2011)).

For complete Bayesian analysis of our parameters, we stress that $g$ and $p$ are stochastic also. The priors for these can be set as:

$$g \sim U(g; 1, g_{\text{max}}),$$

$$p_k | g \sim \text{iid} U(p_k; 0, p_{\text{max}}), \quad (k = 1, \ldots, g),$$

where $U(x; a, b)$ denotes the discrete uniform distribution for $x = a, a + 1, \ldots, b$ and $p_{\text{max}}$ and $g_{\text{max}}$ are fixed.

For details of the distributions mentioned above, see Section 2.11.

### 3.5.1 Formulae for hyperparameters

In the above prior settings, $\lambda$ is a hyperparameter, and $\delta, \zeta, \kappa, c, a, b$ are constants. The values of these are to be chosen from the following formulae, which are based on the observed interval of variation of the data, and the length of this interval, say $R_y$ (see Richardson and Green (1997), Stephens (2000)):

$$\delta_1 = \cdots = \delta_g = 1,$$

$$R_y = \max(y) - \min(y), \quad \zeta = \min(y) + \frac{R_y}{2}, \quad \kappa = \frac{1}{R_y},$$

$$c = 2, \quad a = 0.2, \quad b = \frac{100 a}{c R_y^2}.$$

Note that the Dirichlet distribution is the natural conjugate distribution for the mixing weights (or mixture proportions), and choosing $\delta = 1$ gives a uniform prior over the space $\omega_1 + \cdots + \omega_g = 1$. The values chosen above for $\zeta$ and $\kappa$ ensure a prior of $\mu_k$ which is fairly flat over the range of the data, $R_y$. The choice of $c$ with a relatively flat hyperprior on the hyperparameter $\lambda$ express the belief that the $\tau_k$ are similar, without being very informative about their absolute size. This is less restrictive than constraining all the precisions to be equal.

### 3.6 MCMC moves for parameter estimation

We set $\Lambda = (g, p)$ as the model index. Given the priors of the parameter $\varphi$ specified above and the data, we apply MCMC component-wise techniques to
calculate the full conditional posterior distribution $\pi(\varphi \mid \Lambda, y)$ in which each of the parameter will update separately. For example, given an initial value $\varphi^{(0)} = (\omega^{(0)}, \gamma^{(0)}, \tau^{(0)}, \phi^{(0)})$ we can generate a Markov chain $\{\varphi^{(i)} = (\omega^{(i)}, \gamma^{(i)}, \tau^{(i)}, \phi^{(i)}) : i = 1, \ldots, N\}$ by using Gibbs and Metropolis-Hasting (M-H) sampling techniques, which has equilibrium distribution $\pi(\varphi \mid \Lambda, y)$. Both these sampling are produced by cycling repeatedly through draws of each parameter conditional on the remaining parameters. Moreover, we need to update the latent variable $z$ as well. Therefore, we have a set of SEVEN moves in total which is discussed below.

### 3.6.1 Move I: updating $z$

This is a Gibbs-typed move and is updated according to the Bayes Theorem. From (3.8) and (3.9), we have the full conditional posterior density for $z \mid \varphi, y$ is:

$$
\pi(z \mid \varphi, y) \propto \mathcal{L}(\varphi \mid y, z)\pi(z \mid \omega)
$$

$$
\propto \prod_{t=1}^{T} \sum_{k=1}^{g} \omega_k \mathcal{N}(y_t; \nu_{k,t}\tau_k^{-1}) \cdot 1_{\{z_t=k\}}.
$$
Therefore, we have the full conditional probability that the observation \( y_t \) has been generated by the \( k \)-th component is

\[
\Pr(z_t = k \mid y_t, \varphi) = \frac{\pi(z_t = k \mid y) dN(y_t; \nu_{k,t}, \tau_k^{-1})}{\sum_{k=1}^{g} \pi(z_t = k \mid y) dN(y_t; \nu_{k,t}, \tau_k^{-1})} = \frac{1}{C_1} \omega_k \sqrt{\tau_k} \exp \left[-\frac{\tau_k}{2}(y_t - \nu_{k,t})^2\right], \hspace{1em} (k = 1, \ldots, g),
\]

(3.17)

where \( C_1 = \sum_{u=1}^{g} \omega_u \sqrt{\tau_u} \exp \left[-\frac{\tau_u}{2}(y_t - \nu_{u,t})^2\right] \) denotes the normalizing constant, \( \nu_{k,t} \) is defined as in the equation (3.4).

### 3.6.2 Move II: updating \( \omega \)

Let \( \varphi_{-\omega} \) denote the remaining parameters except \( \omega \). Then the full conditional posterior distribution \( \pi(\omega \mid \varphi_{-\omega}, y, z) \) can be obtained by applying a Gibbs-type move as below:

\[
\pi(\omega \mid \varphi_{-\omega}, y, z) \propto \pi(\omega, \mu, \tau, \phi, p, z, y)
\]

\[
\propto \mathcal{L}(\omega, \mu, \tau, \phi, p, z) \pi(\omega) \pi(\omega) \pi(\phi \mid p) \pi(p) \pi(\tau \mid \lambda) \pi(\lambda)
\]

\[
\propto \pi(z \mid \omega) \pi(\omega)
\]

\[
\propto \prod_{k=1}^{g} \omega_k^{n_k} \prod_{k=1}^{g} \omega_k^{\delta_k-1}, \hspace{1em} \text{where} \hspace{1em} n_k = \sum_{t: z_t = k} 1
\]

\[
\propto \prod_{k=1}^{g} \omega_k^{\delta_k+n_k-1}.
\]

Hence, the posterior distribution is
\[ \omega \mid \varphi_{-\omega}, y, z \sim \mathcal{D}(\omega; \delta_1 + n_1, \ldots, \delta_g + n_g), \]

(3.18)

where \( n_k = \sum_{t: z_t = k} 1, \ (k = 1, \ldots, g) \).

### 3.6.3 Move III: updating \( \mu \)

Let \( \varphi_{-\mu} \) denote the remaining parameters except \( \mu \). Then the full conditional posterior distribution, \( \pi(\mu \mid \varphi_{-\mu}, y, z) \), can be obtained by applying a Gibbs-type move as below:

\[
\pi(\mu \mid \varphi_{-\mu}, y, z) \propto \mathcal{L}(y, z \mid \varphi) \pi(\mu)
\]

\[= \prod_{k=1}^{g} \prod_{t: z_t = k} \mathcal{N}(y_t; \nu_{k,t}, \tau_k^{-1}) \prod_{k=1}^{g} \mathcal{N}(\mu_k; \zeta, \kappa^{-1}),\]

i.e. we have the full conditional posterior for the \( k \)-th component as below:

\[
\pi(\mu_k \mid \varphi_{-\mu_k}, y, z) \propto \prod_{t: z_t = k} \mathcal{N}(y_t; \nu_{k,t}, \tau_k^{-1}) \cdot \mathcal{N}(\mu_k; \zeta, \kappa^{-1})
\]

\[\propto \exp\left[ -\frac{\tau_k}{2} \sum_{t: z_t = k} \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t-i} - \mu_k) \right\}^2 \right] \]

\[(*) \]

\[\times \exp\left[ -\frac{\kappa}{2} (\mu_k - \zeta)^2 \right].\]
Now we simplify the term (*) as below:

\[
\sum_{t: z_t = k} \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t-i} - \mu_k) \right\}^2
\]

\[
= \sum_{t: z_t = k} \left\{ \left( y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i} \right) - \left( 1 - \sum_{i=1}^{p_k} \phi_{k,i} \right) \mu_k \right\}^2
\]

\[
:= e_{k,t}, \text{ say} \\
:= b_k, \text{ say}
\]

\[
= \sum_{t: z_t = k} (e_{k,t} - b_k \mu_k)^2 = \sum_{t: z_t = k} (e_{k,t} - \bar{e}_k + \bar{e}_k - b_k \mu_k)^2,
\]

where \( \bar{e}_k = \frac{1}{n_k} \sum_{t: z_t = k} e_{k,t} \)

\[
= \sum_{t: z_t = k} \left\{ (e_{k,t} - \bar{e}_k)^2 + 2 (e_{k,t} - \bar{e}_k) (\bar{e}_k - b_k \mu_k) + (\bar{e}_k - b_k \mu_k)^2 \right\}
\]

\[
= \sum_{t: z_t = k} (e_{k,t} - \bar{e}_k)^2 + n_k (\bar{e}_k - b_k \mu_k)^2,
\]

since \( \sum_{t: z_t = k} (e_{k,t} - \bar{e}_k) = 0 \) and \( n_k = \sum_{t: z_t = k} 1. \)
Substituting the simplified form of (*) in the equation we get:

\[
\pi(\mu_k \mid \varphi_{-\mu_k}, y, z) \propto \exp \left[ -\frac{\tau_k}{2} \sum_{t:z_t=k} (e_{k,t} - \bar{e}_k)^2 + n_k (\bar{e}_k - b_k \mu_k)^2 \right] \cdot \exp \left[ -\frac{\kappa}{2} (\mu_k - \zeta)^2 \right]
\]

\[
\propto \exp \left[ -\frac{n_k \tau_k}{2} (\bar{e}_k - b_k \mu_k)^2 - \frac{\kappa}{2} (\mu_k - \zeta)^2 \right]
\]

\[
\propto \exp \left[ -\frac{1}{2} \left( \tau_k n_k b_k^2 + \kappa \right) \mu_k^2 + \left( \tau_k n_k \bar{e}_k b_k + \kappa \zeta \right) \mu_k \right]
\]

\[
:= A, \text{ say}
\]

\[
\propto \exp \left[ -A \mu_k^2 + 2B \mu_k \right] = \exp \left[ -A \left\{ \mu_k^2 - 2B \frac{B}{A} \mu_k \right\} \right]
\]

\[
= \exp \left[ -A \left\{ \mu_k - B \frac{B}{A} \right\}^2 - \frac{B^2}{A^2} \right]
\]

\[
\propto \exp \left[ -A \left( \mu_k - B \frac{B}{A} \right)^2 \right]
\]

\[
\propto \exp \left[ -\frac{\tau_k n_k b_k^2 + \kappa}{2} \left( \mu_k - \frac{\tau_k n_k \bar{e}_k b_k + \kappa \zeta}{\tau_k n_k b_k^2 + \kappa} \right)^2 \right].
\]

Hence, the full conditional (component-wise) posterior distribution of \( \mu \) is:

\[
\mu_k \mid \varphi_{-\mu_k}, y, z \sim N \left( \mu_k; \frac{\tau_k n_k \bar{e}_k b_k + \kappa \zeta}{\tau_k n_k b_k^2 + \kappa}, \frac{1}{\tau_k n_k b_k^2 + \kappa} \right), \quad (k = 1, \ldots, g),
\]

(3.19)

where

\[
\bar{e}_k = \frac{1}{n_k} \sum_{t:z_t=k} e_{k,t}, \text{ where}
\]
e_{k,t} = y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i}, \text{ and}

n_k = \sum_{t: z_t = k} 1.

b_k = 1 - \sum_{i=1}^{p_k} \phi_{k,i}.

3.6.4 Move IV: updating $\phi$

Here we update $\phi_k$, \((k = 1, \ldots, g)\) by a Metropolis-Hasting (M-H) mechanism. A candidate $\phi_{k_1}^*$ is generated by a normal density and centered in the current state of the chain $\phi_{k_1}$ so that we can define the proposal density as below:

$$q(\phi_{k_1}, \phi_{k_1}^*) = \mathcal{N}(l, u) \left( \phi_{k_1}; \phi_{k_1}, \tau_q^{-1} \right) 1_{\mathcal{MAr}}(\text{Stable}),$$

\((k = 1, \ldots, g; i = 0, 1, \ldots, p_k), \quad (3.20)\)

where $1_{\mathcal{MAr}}(\text{Stable})$ is the indicator function which takes unity if the model is stable otherwise zero (see Boshnakov (2011)), and $\mathcal{N}(l, u)(\cdot)1_{\mathcal{MAr}}(\text{Stable})$ represents that the normal distribution $\mathcal{N}(\cdot)$ is truncated at the boundary \((l, u)\) of the support of the indicator function $1_{\mathcal{MAr}}(\text{Stable})$. The precision $\tau_q$ is tuned so that a satisfactory acceptance rate (20% or more, see Sampietro (2006)) can be reached.

Using the distribution defined above by (3.20), we generate a vector $\phi_k^* = (\phi_{k_1}^*, \phi_{k_1}^*, \ldots, \phi_{k_p}^*)$ and the acceptance probability for this new $\phi_k^*$ is calculated as below:

$$\mathcal{A}(\phi_k, \phi_k^*) = \min \left( 1, \Omega^\phi \right),$$

where

$$\Omega^\phi = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}.$$
In this case, the likelihood ratio is simply given by (see equation (3.10)):

\[
\text{likelihood ratio} = \frac{\prod_{t : z_t = k} \mathcal{N}(y_t; \nu^*_k; \tau^{-1}_k)}{\prod_{t : z_t = k} \mathcal{N}(y_t; \nu_k; \tau^{-1}_k)} = \frac{\sqrt{\frac{\tau_k}{2\pi}} \exp\left[-\frac{\tau_k}{2} \sum_{t : z_t = k} (y_t - \nu^*_k)^2\right]}{\sqrt{\frac{\tau_k}{2\pi}} \exp\left[-\frac{\tau_k}{2} \sum_{t : z_t = k} (y_t - \nu_k)^2\right]}
\]
\[= \exp\left[-\frac{\tau_k}{2} \sum_{t : z_t = k} \left\{ (y_t - \nu^*_k)^2 - (y_t - \nu_k)^2 \right\}\right], \tag{3.21}\]

where \(\nu^*_k = \mu_k + \phi^*_k(y_{t-1} - \mu_k) + \cdots + \phi^*_{kp_k}(y_{t-p_k} - \mu_k)\) as defined in the model (3.4).

Since all the parameters do not change except the AR coefficients, so from the equation (3.5) the prior ratio is the ratio of the two normal densities:

\[
\text{prior ratio} = \frac{\prod_{i = 1}^{p_k} \mathcal{N}(\phi^*_k; \mu, \tau^{-1})}{\prod_{i = 1}^{p_k} \mathcal{N}(\phi_k; \mu, \tau^{-1})} = \exp\left[-\frac{\tau}{2} \sum_{i = 1}^{p_k} \left\{ (\phi^*_k - \mu)^2 - (\phi_k - \mu)^2 \right\}\right], \tag{3.22}\]

and, eventually from the equation (3.20) the proposal ratio is:

\[
\text{proposal ratio} = \frac{\prod_{i = 1}^{p_k} \mathcal{N}(u; \phi^*_k; \tau^{-1})}{\prod_{i = 1}^{p_k} \mathcal{N}(l; \phi^*_k; \tau^{-1})} = \frac{\prod_{i = 1}^{p_k} \Phi_{\mathcal{N}}(u; \phi^*_k, \tau^{-1}) - \Phi_{\mathcal{N}}(l; \phi^*_k, \tau^{-1})}{\prod_{i = 1}^{p_k} \Phi_{\mathcal{N}}(u; \phi_k, \tau^{-1}) - \Phi_{\mathcal{N}}(l; \phi_k, \tau^{-1})} = \prod_{i = 1}^{p_k} \frac{\Phi_{\mathcal{N}}(u; \phi^*_k, \tau^{-1}) - \Phi_{\mathcal{N}}(l; \phi^*_k, \tau^{-1})}{\Phi_{\mathcal{N}}(u; \phi_k, \tau^{-1}) - \Phi_{\mathcal{N}}(l; \phi_k, \tau^{-1})}, \tag{3.23}\]

where \(\Phi_{\mathcal{N}}\) is the normal cumulative distribution function.

By using (3.21), (3.22) and (3.23) we have
\[\Omega_\phi = \exp \left[ -\frac{\tau_k}{2} \sum_{t: z_t = k} \left\{ \left( y_t - \nu^*_k,t \right)^2 - \left( y_t - \nu_{k,t} \right)^2 \right\} \right] \times \exp \left[ -\frac{\tau_\phi}{2} \sum_{i=1}^{p_k} \left\{ \left( \phi^*_i - \mu^\phi \right)^2 - \left( \phi_{ki} - \mu^\phi \right)^2 \right\} \right] \times \prod_{i=1}^{p_k} \Phi_N \left( u; \phi^*_{k,i}, \tau_q^{-1} \right) - \Phi_N \left( l; \phi^*_{k,i}, \tau_q^{-1} \right) \Phi_N \left( u; \phi^*_{k,i}, \tau_q^{-1} \right) - \Phi_N \left( l; \phi^*_{k,i}, \tau_q^{-1} \right), \]

(3.24)

where \( \nu^*_{k,t} = \mu_k + \phi^*_1(y_{t-1} - \mu_k) + \cdots + \phi^*_{kp_k}(y_{t-p_k} - \mu_k) \) and \( \Phi_N \) is the normal cumulative distribution function.

### 3.6.5 Move V: updating \( \lambda \) and \( \tau \)

Let \( \varphi_{-\lambda} \) denote the remaining parameters except \( \lambda \) and so \( \varphi_{-\tau} \) denotes the remaining parameters except \( \tau \). Firstly, the full conditional posterior distribution \( \pi(\lambda | \varphi_{-\lambda}, y, z) \) can be obtained by applying a Gibbs-type move as below:

\[
\pi(\lambda | \varphi_{-\lambda}, y, z) \propto \pi(\omega, \mu, \tau, \phi, p, z, y)
\]

\[
\propto \mathcal{L}(\omega; \mu, \tau, \phi, p, z) \pi(z | \omega) \pi(\omega) \pi(\mu) \pi(\phi | p) \pi(p) \pi(\tau | \lambda) \pi(\lambda)
\]

\[
\propto \pi(\tau | \lambda) \pi(\lambda)
\]

\[
\propto \prod_{k=1}^{g} \mathcal{G}(\tau_k; c, \lambda) \cdot \mathcal{G}(\lambda; a, b)
\]

\[
= \left( \prod_{k=1}^{g} \frac{\lambda^c}{\Gamma(c)} \tau_k^{c-1} \exp \left\{ -\lambda \tau_k \right\} \right) \cdot \left( \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp \left\{ -b \lambda \right\} \right)
\]

\[
\propto \lambda^{(a+gc)-1} \exp \left\{ -\left( b + \sum_{k=1}^{g} \tau_k \right) \lambda \right\}.
\]

Hence, the full conditional posterior distribution for \( \lambda \) is:
\begin{equation}
\lambda \mid \varphi_{-\lambda}, y, z \sim \mathcal{G}(\lambda; a + gc, b + \sum_{k=1}^{g} \tau_{k}),
\end{equation}

(3.25)

where \(a, b, c\) are fixed constant set in the priors of \(\lambda\) and \(\tau\) (see equation (3.13)).

Secondly, the full conditional posterior distribution of \(\tau\) can be obtained as below:

\[
\pi(\tau \mid \varphi_{-\tau}, y, z) \propto \mathcal{L}(\varphi \mid y) \pi(\tau \mid \lambda)
\]

\[
\propto \prod_{k=1}^{g} \prod_{t: z_{t}=k} \mathcal{N}(y_{t}; \nu_{k,t}, \tau_{k}^{-1}) \cdot \prod_{k=1}^{g} \mathcal{G}(\tau_{k}; c, \lambda).
\]

So, the full conditional posterior distribution of \(\tau\) of the \(k\)-th component:

\[
\pi(\tau_{k} \mid \varphi_{-\tau}, y, z) \propto \prod_{t: z_{t}=k} \mathcal{N}(y_{t}; \nu_{k,t}, \tau_{k}^{-1}) \cdot \mathcal{G}(\tau_{k}; c, \lambda)
\]

\[
\propto \tau_{k}^{n_{k}/2} \exp\left\{-\frac{\tau_{k}}{2} \sum_{t: z_{t}=k} (y_{t} - \nu_{k,t})^{2}\right\} \cdot \tau_{k}^{c-1} \exp\{-\lambda \tau_{k}\}
\]

\[
\propto \tau_{k}^{(c+n_{k})/2} \exp\left[-\left\{\lambda + \frac{1}{2} \sum_{t: z_{t}=k} (y_{t} - \nu_{k,t})^{2}\right\} \tau_{k}\right].
\]

Thus, the full conditional posterior distribution of \(\tau\) is:

\[
\tau_{k} \mid \varphi_{-\tau}, y, z \sim \mathcal{G}(\tau_{k}; c + \frac{n_{k}}{2}, \lambda + \frac{1}{2} \sum_{t: z_{t}=k} (y_{t} - \nu_{k,t})^{2}), \quad (k = 1, \ldots, g),
\]

(3.26)

\(\nu_{k,t}\) is defined as in the equation (3.4).

### 3.7 MCMC moves for model determination

To get full Bayesian flavour we must get the posterior distribution of the model index \(\Lambda\) which can be written and factorized as:
\[ \pi(\Lambda \mid y) = \pi(p \mid g, y) \times \pi(g \mid y). \] (3.27)

So, it is enough to update AR order \( p \) and component number \( g \) separately to update the model index \( \Lambda \). We develop two additional MCMC moves to complete our Bayesian analysis as follows:

3.7.1 Move VI: updating of \( p \)

To change (increase or decrease) the AR order of the \( k \)-th component of the \( \mathcal{M}_{AR} \) model, we apply the RJMCMC (Reversible Jump Markov chain Monte Carlo) technique introduced by Green (1995). While designing the appropriate moves, we ensure that the dimensions can be balanced properly, the moves can be simulated conveniently, and the acceptance ratio can be computed economically.

At first we need to choose a component, say \( k^* \), randomly chosen in \( \{1, \ldots, g\} \). Then we can write the proposal for the new AR order of the \( k^* \)-th component, \( p_{k^*}^* \), as:

\[
p_{k^*}^* = \begin{cases} 
p_{k^*} - 1, & \text{with probability } d(p_{k^*}), \\
p_{k^*} + 1, & \text{with probability } b(p_{k^*}), \end{cases} \quad (3.28)
\]

where \( b(p_k) = 1 - d(p_k) \), for \( k = 1, \ldots, g \), \( d(0) = 0 \) and \( b(p_{\text{max}}) = 0 \). In other words, we can say that the AR order \( p_{k^*} \) of randomly selected component \( k^* \) increases by one with probability \( b(p_{k^*}) \) and decrease by one with probability \( d(p_{k^*}) \). Therefore, we call \( b(x) \) birth function while \( d(x) \) is call death function.

Definition of \( b(x) \):

Now we propose two forms for the probability function \( b(x) \) which are given below:


(i) 
\[ b(x) = \begin{cases} 
\frac{e}{c+x}, & \text{if } x \in [0, p_{\text{max}}), \ c \in \mathbb{R}, \\
0, & \text{if } x = p_{\text{max}}, 
\end{cases} \tag{3.29} \]

(ii) 
\[ b(x) = \begin{cases} 
\min\{1, \frac{p(x+1)}{p(x)}\}, & \text{if } x \in [0, p_{\text{max}}), \\
0, & \text{if } x = p_{\text{max}}, 
\end{cases} \tag{3.30} \]

where \( p(x) \) denotes the Poisson density function which has the following form:
\[ p(x) = \frac{e^{-u}u^x}{x!}, \ u \in \mathbb{R}. \]

And thus we can define \( d(x) \) by the transformation \( d(x) = 1 - b(x) \). Obviously, \( d(k) \) is monotonically increasing function and \( b(k) \) is monotonically decreasing function and both \( b(x) \) and \( d(x) \) satisfy the required properties \( d(0) = 0 \) and \( b(p_{\text{max}}) = 0 \) (see Figures 3.2 and 3.3).

![Figure 3.2: Characteristics of the probability function \( d(x) \) and \( b(x) \) where \( b(x) \) is defined by (3.29) and \( d(x) = 1 - b(x) \). If we consider \( x > 0 \) and fixed, then \( d \) increases as \( c \) decreases. On the other hand, \( b \) decreases as \( c \) decreases for any fixed \( x > 0 \).](image)

Let \( \phi_k^* \) be the proposal vector of the AR coefficients. When \( p_{k^*}^* = p_k^* - 1 \), the autoregressive coefficients \( \phi_k^* = (\phi_{k^*,1}, \ldots, \phi_{k^*,p_{k^*}^*}) \) is updated by simply discarding the last AR coefficient in \( \phi_{k^*} = (\phi_{k^*,1}, \ldots, \phi_{k^*,p_{k^*}}) \). On the other hand, when \( p_{k^*}^* = p_k^* + 1 \), the corresponding AR coefficient vector \( \phi_k^* = (\phi_{k^*,1}, \ldots, \phi_{k^*,p_{k^*}}, \phi_{k^*,p_{k^*}^*}) \) is updated by adding new parameter \( \phi_{k^*,p_{k^*}^*} \) in \( \phi_{k^*} = (\phi_{k^*,1}, \ldots, \phi_{k^*,p_{k^*}}) \) with
\[ \phi_{k^*,p_{k^*}^*} \overset{\text{ind}}{\sim} \mathcal{N} \left( \phi_{k^*,p_{k^*}^*}; \mu^\phi, \sigma^\phi_{-1} \right) \mathbf{1}_{\text{MAR}}(\text{Stable}). \]
Figure 3.3: Characteristics of the probability function \( d(x) \) and \( b(x) \) where \( b(x) \) is defined by (3.30) with \( u = 2 \) and \( d(x) = 1 - b(x) \).

To update the AR order of the model by using RJMCMC method, we have to calculate the acceptance probability,

\[
A_{p^\pm} = \min \{1, \Omega_{p^\pm} \},
\]

and \( \Omega_{p^\pm} \) is described as in section 2.7.1:

\[
\Omega_{p^\pm} = \text{likelihood ratio} \times \text{prio ratio} \times \text{proposal ratio} \times \text{jacobian}.
\]

The likelihood ratio is the same as of equation (3.21):

\[
\text{likelihood ratio} = \frac{\prod_{t: z_t = k^*} \mathcal{N}(y_t; \nu_{k^*, t}^*, \tau_{k^*}^{-1})}{\prod_{t: z_t = k^*} \mathcal{N}(y_t; \nu_{k^*, t}, \tau_{k^*}^{-1})}
\]

\[
= \exp \left[ -\frac{\tau_k}{2} \sum_{t: z_t = k^*} \left\{ (y_t - \nu_{k^*, t}^*)^2 - (y_t - \nu_{k^*, t})^2 \right\} \right]
\]

but the conditional mean \( \nu_{k^*, t}^* \) is

\[
\nu_{k^*, t}^* = \mu_{k^*} + \phi_{k^*, 1}^* (y_{t-1} - \mu_{k^*}) + \cdots + \phi_{k^*, p_{k^*}}^* (y_{t-p_{k^*}} - \mu_{k^*})
\]

as the AR order of the \( k^* (\in \{1, \ldots, g\}) \)-th component of the model is updated with this move.
The prior ratio is given by:

\[
\frac{\pi(p_k^* \pi(\phi_k^*)}{\pi(p_k)\pi(\phi_k)} = \frac{p_k^*}{p_k} \prod_{i=1}^{k^*} \frac{\pi(\phi_{k^*-i}^*)}{\pi(\phi_{k-i})} = \begin{cases} \frac{1}{\pi(\phi_{k^*} p_k^*)} & \text{if } p_k^* = p_k^* - 1, \\ \pi(\phi_{k^*} p_k^*) & \text{if } p_k^* = p_k^* + 1, \end{cases}
\]

because of the uniform prior on the AR order of the model and only the \(k^*\)-th component is updated.

To calculate the proposal ratio we have to take into account the probabilities \(b(\cdot)\) and \(d(\cdot)\) and the proposal density, which is chosen equal to the prior on \(\phi\).

\[
\text{propostion ratio} = \begin{cases} \frac{b(p_k^*) \pi(\phi_{k^*} p_k^*)}{d(p_k^*)} & \text{if } p_k^* = p_k^* - 1, \\ \frac{d(p_k^*)}{b(p_k^*) \pi(\phi_{k^*} p_k^*)} & \text{if } p_k^* = p_k^* + 1. \end{cases}
\]

Finally, the jacobian is equal to 1 because the invertible function \(g(\cdot)\) is considered as the identity function.

Therefore, multiplying likelihood, prior and proposal ratios, we derive ratio in the acceptance probability for the both the cases:

\[
\Omega_{p^*} = \begin{cases} L \times \frac{b(p_k^*) \pi(\phi_{k^*} p_k^*)}{d(p_k^*)} & \text{if } p_k^* = p_k^* - 1, \\ L \times \frac{d(p_k^*)}{b(p_k^*) \pi(\phi_{k^*} p_k^*)} & \text{if } p_k^* = p_k^* + 1. \end{cases}
\]

(3.31)

where

\[
L = \exp \left[ -\frac{\tau_k}{2} \sum_{t:z_t=k} \left\{ (y_t - \nu_{k^*,t})^2 - (y_t - \nu_{k^*,t})^2 \right\} \right],
\]

with \(\nu_{k^*,t} = \mu_{k^*} + \phi_{k^*-1}(y_{t-1} - \mu_{k^*}) + \cdots + \phi_{p_k^*}(y_{t-p_k^*} - \mu_{k^*}).\)

Therefore, \(\pi(p \mid g, y)\) is simply estimated by the proportions of every possible value for \(p\) in the sample obtained by the previous complete MCMC algorithm.
3.7.2 Move VII: updating of $g$

This is the final move to get the complete Bayesian analysis of the $M_{\text{AR}}$ model. According to Bayes’ theorem, the marginal posterior distribution of $g$ is:

$$\pi(g \mid y) \propto f(y \mid g) \pi(g),$$  \hspace{1cm} (3.32)

where $\pi(g)$ is the prior on $g$ and $f(y \mid g)$ is marginal likelihood given by

$$f(y \mid g) = \sum_p \int f(y \mid \varphi, p, g) \pi(\varphi, p \mid g) d\varphi.$$  \hspace{1cm} (3.33)

The idea of the method, developed by Chib and Jeliazkov (2001), arises from the basic marginal likelihood identity. For notational convenience we can write the marginal likelihood by omitting the model index $g$:

$$f(y) = \frac{f(y \mid \varphi^*, p^*) \pi(\varphi^*, p^*)}{\pi(\varphi^*, p^* \mid y)} = \frac{f(y \mid \varphi^*, p^*) \pi(\varphi^* \mid p^*) \pi(p^*)}{\pi(\varphi^* \mid p^*, y) \pi(p^* \mid y)}$$  \hspace{1cm} (3.34)

for any fixed point $(\varphi^*, p^*)$ which have the maximum density.

Note that the only quantity in (3.34) that is not already available is $\pi(\varphi^* \mid p^*, y)$. Chib and Jeliazkov (2001) proposed an efficient method to produce the estimate $\hat{\pi}(\varphi^* \mid p^*, y)$ using the output from a MCMC simulation with fixed $p^*$.

First of all, $\hat{\pi}(\varphi^* \mid p^*, y)$ is factorized as

$$\hat{\pi}(\varphi^* \mid p^*, y) = \hat{\pi}(\phi^* \mid p^*, y) \times \hat{\pi}(\mu^* \mid \phi^*, p^*, y) \times \hat{\pi}(\tau^* \mid \mu^*, \phi^*, p^*, y) \times \hat{\pi}(\omega^* \mid \tau^*, \mu^*, \phi^*, p^*, y).$$  \hspace{1cm} (3.35)

Suppose we have a sample $\{\varphi^{(i)}, z^{(i)}\}, (i = 1, \ldots, N)$, from the MCMC algorithm for a given $p^*$ (i.e. a sample from $\pi(\varphi \mid p^*)$). Then the terms of (3.35) are estimated by the following steps:

**Estimation of $\hat{\pi}(\phi^* \mid p^*, y)$**

Let $\mathcal{U}_{k-1} = (p, \phi_1, \ldots, \phi_{k-1})$ and $\mathcal{U}^{k+1} = (\phi_{k+1}, \ldots, \phi_g, \mu, \tau, \omega)$ be the two partitions of the parameters for any $k \in \{1, \ldots, g\}$.

Let $\{\mathcal{U}^{k+1, (i)}, z^{(i)}\}, (i = 1, \ldots, N)$, are drawn from a reduced MCMC algorithm...
CHAPTER 3. MIXTURE OF GAUSSIAN AR COMPONENTS

with distribution of interest \( \pi(\tilde{U}^{k+1}, z \mid \tilde{U}_k^*, y) \).

Also draw \( \tilde{\phi}_k^{(i)} \), for \( i = 1, \ldots, N \), from the proposal density

\[
q_p(\phi_k^*, \tilde{\phi}_k^{(i)}) = \prod_{j=1}^{p_k} q(\phi_{kj}, \tilde{\phi}_{kj}^{(i)}),
\]

where \( q(\cdot, \cdot) \) is defined by the Equation (3.20) in Section 3.6.4.

We set

\[
\pi(\phi^* \mid \tau_{(i)}, \phi_1^*, \ldots, \phi_{k-1}^*) = \frac{1}{N-B} \sum_{i=B+1}^{N} \hat{A}(\phi_k^{(i)}, \phi_k^*) q_p(\phi_k^{(i)}, \phi_k^*) \prod_{k=1}^{g} \hat{A}(\phi_k^*, \phi_k^*),
\]

where \( \hat{A}(\cdot, \cdot) = \min(1, \tilde{\Omega}(\cdot, \cdot)) \) with \( \tilde{\Omega}(\cdot, \cdot) \) defined in Equation (3.24).

Set \( \tilde{U}^{k+1, (i)} = \tilde{U}^{k+1, (i)} \) and \( z^{(i)} = z^{(i)} \), for \( i = 1, \ldots, N \). Repeat this step for \( k = 1, \ldots, g \) and finally set

\[
\hat{\pi}(\phi^* \mid p^*, y) = \prod_{k=1}^{g} \hat{\pi}(\phi_k^* \mid p^*, \phi_1^*, \ldots, \phi_{k-1}^*). \tag{3.36}
\]

Estimation of \( \hat{\pi}(\mu^* \mid p^*, y) \)

Let we have a sample \((\tau^{(i)}, z^{(i)})\) from the last iteration of the previous step. i.e. they are marginally from \( \pi(\tau, \omega, z \mid \phi^*, p^*, y) \):

\[
\hat{\pi}^{(i)}(\mu^* \mid \phi^*, p^*, y) = \frac{1}{N-B} \sum_{i=B+1}^{N} \prod_{k=1}^{g} \pi(\mu_k^* \mid \phi^*, \tau^{(i)}, z^{(i)}, p^*, y), \tag{3.37}
\]

where \( \pi(\mu_k^* \mid \phi^*, \tau^{(i)}, z^{(i)}, p^*, y) \) is given by the same distribution as Equation (3.19) in Section 3.6.3.
Chapter 3. Mixture of Gaussian AR Components

81

Estimation of $\hat{\pi}(\tau^* | \phi^*, \mu^*, p^*, y)$

Sample $\{\tau^{(j)}, \omega^{(j)}, z^{(j)}\}$, for $j = 1, \ldots, N$, from a reduced MCMC algorithm with distribution of interest $\pi(\tau, \omega, z | \phi^*, \mu^*, p^*, y)$ and set

$$\hat{\pi}(\tau^* | \phi^*, \mu^*, p^*, y) = \frac{1}{N - B} \sum_{j=B+1}^{N} \prod_{k=1}^{g} \pi(\tau_k^* | \phi^*, \mu^*, z^{(j)}, p^*, y),$$  \hspace{1cm} (3.38)

where $\pi(\tau_k^* | \phi^*, \mu^*, z^{(u)}, p^*, y)$ is given by the Equation (3.26) and we rewrite this with new settings as below:

$$\tau_k^* \sim \mathcal{N}(\tau_k^* | c + \frac{\mu^{(j)}}{2}, \lambda + \frac{1}{2} \sum_{t: z^{(j)} = k} \left( y_t - \nu_{k,t}^* \right)^2), \hspace{1cm} (k = 1, \ldots, g),$$

where $n_k^{(j)} = \sum_{t: z^{(j)} = k} 1$ and $\nu_{k,t}^* = \mu_k + \phi_{k,1}(y_{t-1} - \mu_k) + \cdots + \phi_{k,p_k}(y_{t-p_k} - \mu_k)$.

Estimation of $\hat{\pi}(\omega^* | \phi^*, \mu^*, \tau^*, p^*, y)$

Let $\{\omega^{(v)}, z^{(v)}\}$, $(v = 1, \ldots, N)$, are drawn from a reduced MCMC algorithm with distribution of interest $\pi(\omega, z | \phi^*, \mu^*, \tau^*, p^*, y)$ and set

$$\hat{\pi}(\omega^* | \phi^*, \mu^*, \tau^*, p^*, y) = \frac{1}{V} \sum_{v=1}^{N} \pi(\omega^* | z^{(v)}, p^*, y),$$  \hspace{1cm} (3.39)

where $\pi(\omega^* | z^{(v)}, p^*, y)$ is given by the Equation (3.18) and we re write this here with these new settings for convention:

$$\omega^* \sim \mathcal{D}(\omega | \delta_1 + n_1^{(v)}, \ldots, \delta_k + n_k^{(v)}), \hspace{1cm} where \hspace{1cm} n_k^{(v)} = \sum_{t: z^{(v)} = k} 1.$$

Substituting all these above estimated values obtained by, (3.36), (3.38), (3.37) and (3.39) into the Equation (3.35), we can get an estimated value of the marginal likelihood, $f(y | g)$. Finally, the posterior probability of the component number of the $\mathcal{M}_{AR}$ model can be estimated from the Equation (3.32).
3.8 Examples

In this section, we examine our MCMC (Markov Chain Monte Carlo) moves that we obtained in Section 3.6 for estimating $M_{AR}$ parameters. We have written R programming code for this simulation experiment. Then we also use some related packages that are available in R-cran. In particular, we would like to mention the package ‘coda’ developed by Plummer et al. (2006) for displaying the summary output of the model parameters not only in this chapter but also in the rest of this thesis. For this, simulated and real data are considered.

In Bayesian analysis, we have immense opportunity to present the estimated results graphically as well as numerically. It is because of considering all parameters as stochastic variables. Therefore, we need to estimate their densities as well. The Highest Posterior Density (HPD) intervals for the parameters in an MCMC sample have been calculated as well. We report some important results for each of the example here and the rest can be found in the appendix.

3.8.1 Experiment with simulated data sets

Here, we consider two 2-component models and a 3-component model for checking MCMC moves for estimating $M_{AR}$ model parameters. For our experiment, we generate time-series data $\{y_1, \ldots, y_T\}$ from the models. We set the startup values, $y_0 = y_{-1} = \cdots = 0$. We also discard a sufficiently large number of observations to remove transient effects in generating time-series observations.

Experiment 1: Two-component $M_{AR}$ model

We consider the following two $M_{AR}(2; 1, 1)$ models:

\[
\begin{align*}
ny_t & | G_{t-1} \sim 0.5 \mathcal{N}(y_t; \nu_{1,t}, 1^2) + 0.5 \mathcal{N}(y_t; \nu_{2,t}, 2^2), \\
\nu_{1,t} & = -0.5 \ y_{t-1}, \\
\nu_{2,t} & = 1.1 \ y_{t-1},
\end{align*}
\]

So, we have:

$$(\omega_1, \omega_2) = (0.5, 0.5), (\mu_1, \mu_2) = (0.0, 0.0),$$

$$(\sigma_1, \sigma_2) = (1, 2), (\phi_{11}, \phi_{21}) = (-0.5, 1.1).$$

Model (I) was considered by Wong and Li (2000) and Wong et al. (2009), named as Model (A) and analysed by EM algorithm for estimating parameters. Obviously, our approach is Bayesian. We call the simulated series from model (I) as $y_{MoI}$.
Our MCMC procedure is performed as follows. We generated \( N = 50000 \) runs of three Markov chains with built-in initial values of the hyperparameters (discussed in Section 3.5.1) to determine the best model index. Although different initial values generated automatically for each chain, we get similar results in all chains and we discarded initial 10000 runs as a burn-in. By using the MCMC moves discussed in Section 3.7, we found that \( \mathcal{M}_{AR}(2; 1, 1) \) has the highest probability. i.e. it is the best fit for the series \( y_{MoI} \).

For parameter estimation we also generate three MCMC chains with \( N = 100000 \) runs and discarded 50000 runs as burn-in. We used several convergence diagnostics to ensure the convergence of the MCMC chains. For details of these diagnostics see Section 2.4. One of those diagnostics was proposed by Raftery and Lewis (1992) based on a criterion of accuracy of estimation of the quantile \( q \). Table 3.1 gives the Raftery and Lewis convergence diagnostic information about convergence to the stationary distribution and estimates the run-lengths needed to accurately estimate quantiles of functions of the parameters. Dependence factors (\( I \)) less than 5.0 indicate convergence passed in all cases (see Raftery and Lewis (1992)). Table 3.3 shows the Heidelberger and Welch’s convergence diagnostic results (only the first chain’s results are reported) which is based on a criterion of relative accuracy for the estimate of the mean. It implements a convergence diagnostic, and removes up to half the chain in order to ensure that the means are estimated from a chain that has converged. All these indicate that the chains have converged in all cases.

Table 3.2 shows that the acceptance rate of Metropolis-Hastings moves of the AR coefficients. Since the rate is in between 20% and 50%, so the result is satisfactory for the related move.

Figure 3.5 and 3.6 show the convergence diagrams of the posterior sample for each parameter along with their marginal densities without label-switched and with label-switched respectively. Observe that the posterior converges perfectly and does not shows artificial multi-modality in without label-switched case while the convergence is disappear for the AR coefficients in the with label-switched case. This suggests that the label-switching is not required for the \( y_{MoI} \) series. Moreover, our improved Bayesian technique produces smooth and almost symmetric marginal posterior densities of the parameter \( \varphi = (\omega, \mu, \sigma, \phi) \).

Table 3.4 shows the parameter estimation results for the posterior sample (only chain 1 is reported here since the rest produces almost the same results). Column 3-6 in this table show the mean, standard deviation, median and Highest Posterior Density (HPD) interval of the posterior distribution of each parameter respectively. Note that these estimates are close to the true value of the parameter. Computational time required to finish the MCMC algorithm for 100000 runs in each of three chains is shown at the bottom of the table.
CHAPTER 3. MIXTURE OF GAUSSIAN AR COMPONENTS

Table 3.1: Raftery and Lewis’s diagnostic of parameters sampled from the MCMC moves when fitting the $\mathcal{M}_{\text{AR}}(2; 1, 1)$ model to the simulated series $\text{yMoI}$ with $N = 100000$ and $B = 1000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Chain 1</th>
<th>Chain 2</th>
<th>Chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{11}$</td>
<td>9 3746 3.46</td>
<td>2 3746 1.06</td>
<td>2 3746 1.06</td>
</tr>
<tr>
<td>$\omega_{21}$</td>
<td>12 3746 3.75</td>
<td>8 3746 2.55</td>
<td>8 3746 2.51</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>3 3746 1.17</td>
<td>3 3746 1.14</td>
<td>3 3746 1.14</td>
</tr>
<tr>
<td>$\gamma_{21}$</td>
<td>3 3746 1.11</td>
<td>3 3746 1.14</td>
<td>3 3746 1.09</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>4 3746 1.28</td>
<td>4 3746 1.26</td>
<td>4 3746 1.23</td>
</tr>
<tr>
<td>$\sigma_{21}$</td>
<td>6 3746 1.28</td>
<td>6 3746 2.10</td>
<td>6 3746 2.15</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>12 3746 3.85</td>
<td>14 3746 3.8</td>
<td>12 3746 3.82</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>12 3746 3.85</td>
<td>10 3746 2.9</td>
<td>12 3746 3.90</td>
</tr>
</tbody>
</table>

We also generate three MCMC with new prior and posterior setup for the $\phi$ parameter as proposed by Sampietro (2006) keeping all the setup as before. We did convergence diagnostic tests as well (not reported here). Figures 3.7 and 3.8 show the convergence diagrams and marginal posterior densities of each of the parameter without label-switched and with label-switched respective. Label-switching is not necessary in this case as well. Observe that the convergence diagrams and marginal posterior densities of $\phi$ in not better than that of $\phi$ in our case.

Moreover, Table 3.5 shows parameter estimation results for this new settings. Though it requires less computation time for the same MCMC runs, the estimated results are not very close to the true values as ours. Finally, we use the Monte Carlo estimation techniques for both ours and Sampietro’s methods and similar results has been found here as well (see Table 3.6). These indicate that our improved Bayesian technique for the $\mathcal{M}_{\text{AR}}$ model is more effective.
Table 3.2: Acceptance Rate for Metropolis-Hastings chains of AR coefficients when fitting the $\mathcal{M}_{AR}(2; 1, 1)$ model to the $y_{MoI}$ simulated series.

<table>
<thead>
<tr>
<th></th>
<th>N=50000</th>
<th></th>
<th>N=100000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\phi_{11}$</td>
<td>$\phi_{21}$</td>
<td>$\phi_{11}$</td>
</tr>
<tr>
<td>chain1</td>
<td>0.3513470</td>
<td>0.3696874</td>
<td>0.3675937</td>
</tr>
<tr>
<td>chain2</td>
<td>0.3514670</td>
<td>0.3658073</td>
<td>0.3674037</td>
</tr>
<tr>
<td>chain3</td>
<td>0.3532471</td>
<td>0.3713874</td>
<td>0.3643936</td>
</tr>
</tbody>
</table>

Table 3.3: Heidelberger and Welch’s convergence diagnostic of sample values of the parameters when fitting the $\mathcal{M}_{AR}(2; 1, 1)$ model to the simulated series $y_{MoI}$.

<table>
<thead>
<tr>
<th></th>
<th>Stationarity test</th>
<th>Start iteration</th>
<th>p-value</th>
<th>Halfwidth test</th>
<th>Halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{11}$</td>
<td>passed</td>
<td>1</td>
<td>0.779</td>
<td>passed</td>
<td>0.000673</td>
</tr>
<tr>
<td>$\omega_{21}$</td>
<td>passed</td>
<td>1</td>
<td>0.729</td>
<td>passed</td>
<td>0.000673</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>passed</td>
<td>1</td>
<td>0.753</td>
<td>passed</td>
<td>0.00157</td>
</tr>
<tr>
<td>$\gamma_{21}$</td>
<td>passed</td>
<td>1</td>
<td>0.705</td>
<td>passed</td>
<td>0.00251</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>passed</td>
<td>1</td>
<td>0.6562</td>
<td>passed</td>
<td>0.00362</td>
</tr>
<tr>
<td>$\sigma_{21}$</td>
<td>passed</td>
<td>5001</td>
<td>0.0507</td>
<td>passed</td>
<td>0.00456</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>passed</td>
<td>1</td>
<td>0.562</td>
<td>passed</td>
<td>0.00730</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>passed</td>
<td>1</td>
<td>0.779</td>
<td>passed</td>
<td>0.00879</td>
</tr>
</tbody>
</table>

† only the first chain is reported here.
Table 3.4: Estimation results of parameters when fitting the $\mathcal{M}_{AR}(2; 1, 1)$ model to the $y_{MoI}$ series with $N = 100000$ and burn-in $= 50000$.

<table>
<thead>
<tr>
<th>True value</th>
<th>Posterior statistics</th>
<th>Mean</th>
<th>s.d.</th>
<th>Median</th>
<th>HPD†</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$ 0.5</td>
<td>0.5943 (0.000197)†</td>
<td>0.04404</td>
<td>0.5944</td>
<td>[0.5094032, 0.6823211]</td>
<td></td>
</tr>
<tr>
<td>$\omega_2$ 0.5</td>
<td>0.4057 (0.000197)†</td>
<td>0.04404</td>
<td>0.4056</td>
<td>[0.3176789, 0.4905968]</td>
<td></td>
</tr>
<tr>
<td>$\mu_1$ 0.0</td>
<td>-0.06075 (0.0005143)</td>
<td>0.1029</td>
<td>-0.06135</td>
<td>[-0.2614122, -0.2614122]</td>
<td></td>
</tr>
<tr>
<td>$\mu_2$ 0.0</td>
<td>0.58443 (0.0779762)</td>
<td>15.5954</td>
<td>0.73104</td>
<td>[-34.4951866, 36.695812]</td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$ 1</td>
<td>1.692 (0.001301)</td>
<td>0.2601</td>
<td>1.662</td>
<td>[1.237865, 2.219494]</td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$ 2</td>
<td>2.340 (0.001629)</td>
<td>0.3258</td>
<td>2.282</td>
<td>[1.826032, 2.967795]</td>
<td></td>
</tr>
<tr>
<td>$\phi_{11}$ -0.5</td>
<td>-0.4245 (0.0003115)</td>
<td>0.06966</td>
<td>-0.4302</td>
<td>[-0.5510161, -0.2808187]</td>
<td></td>
</tr>
<tr>
<td>$\phi_{21}$ 1.1</td>
<td>1.0302 (0.0002697)</td>
<td>0.06030</td>
<td>1.0172</td>
<td>[0.9237175, 1.1711671]</td>
<td></td>
</tr>
</tbody>
</table>

Computational time: 51.54263 mins

† Highest Posterior Density interval,
‡ Standard Error.
Table 3.5: Estimation results of parameters when fitting the $\mathcal{M}_{AR}(2; 1,1)$ model with Sampietro’s setup to the $y_{MoI}$ series with $N = 100000$ and burn-in = 50000.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Posterior statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.5</td>
<td>0.5895</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0001964)</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.5</td>
<td>0.4105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0001964)</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.0</td>
<td>-0.06003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0004632)</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.0</td>
<td>-7.30651</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0814908)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>1.690</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001164)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>2</td>
<td>2.368</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001482)</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.5</td>
<td>-0.4267</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0003065)</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>1.1</td>
<td>0.9781</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0001202)</td>
</tr>
</tbody>
</table>

Computational time: 40.03449 mins

† Highest Posterior Density interval
Table 3.6: Estimation results of parameters when fitting the $\mathcal{M}_{AR}(2; 1, 1)$ model to the $y_{MoI}$ series by using Monte Carlo technique.

<table>
<thead>
<tr>
<th></th>
<th>True value</th>
<th>Our Method</th>
<th>Sampietro’s Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean s.d. Median</td>
<td>Mean s.d. Median</td>
<td>Mean s.d. Median</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.5</td>
<td>0.5638 0.0371 0.5603</td>
<td>0.5633 0.0365 0.5597</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.53, 0.61]$^\dagger$</td>
<td>[0.53, 0.61]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.5</td>
<td>0.4362 0.0371 0.4397</td>
<td>0.4367 0.0365 0.4403</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.39, 0.47]</td>
<td>[0.39, 0.47]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.0</td>
<td>0.0047 0.1533 0.0053</td>
<td>0.0036 0.1582 0.0040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.13, 0.17]</td>
<td>[-0.18, 0.16]</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.0</td>
<td>-0.01480 0.1990 -0.0122</td>
<td>-0.0130 0.1957 -0.0104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.26, 0.18]</td>
<td>[-0.29, 0.23]</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>1.7013 0.2842 1.667</td>
<td>1.7363 0.3039 1.6969</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1.30, 2.13]</td>
<td>[1.34, 2.22]</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>2</td>
<td>2.1410 0.3464 2.1354</td>
<td>2.1811 0.3565 2.1772</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1.88, 2.49]</td>
<td>[1.83, 2.45]</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.5</td>
<td>-0.1310 0.4216 -0.2405</td>
<td>-0.1437 0.4230 -0.2387</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.43, 0.53]</td>
<td>[-0.47, 0.42]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>1.1</td>
<td>0.7432 0.4367 0.8607</td>
<td>0.7120 0.4230 0.8297</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.14, 1.03]</td>
<td>[0.17, 0.97]</td>
</tr>
</tbody>
</table>

Computational time: 2.61272 hours

$^\dagger$ Highest Posterior Density interval
CHAPTER 3. MIXTURE OF GAUSSIAN AR COMPONENTS

Experiment 2: Three-component \( \mathcal{M}_{AR} \) model

We also illustrate our MCMC techniques with another simulated series generated from the following \( \mathcal{M}_{AR}(3; 2, 1, 3) \) model:

\[
\begin{align*}
   y_t \mid \mathcal{G}_{t-1} & \sim 0.40 \mathcal{N}(y_t; \nu_{1,t}, 1) + 0.30 \mathcal{N}(y_t; \nu_{2,t}, 1) + 0.30 \mathcal{N}(y_t; \nu_{3,t}, 5^2), \\
   \nu_{1,t} & = \frac{0.9}{y_{t-1}} - 0.6y_{t-2}; \\
   \nu_{2,t} & = -0.5 \ y_{t-1}; \\
   \nu_{3,t} & = -5.0 + 1.5 \ y_{t-1} - 0.74y_{t-2} + 0.12y_{t-3}
\end{align*}
\]

(II)

We will call this simulated series by \( y_{MoII} \) hereafter.

We run three-MCMC chains with \( N = 100000 \) and discarded initial 50000 runs as burn-in. Note that the initial values \( \varphi^{(0)} \) are generated automatically based on the formulae discussed in Section 3.5.1.

Figures 3.10 and 3.11 show the convergence diagrams and marginal densities of the posterior sample (from chain 1) for each parameter for the series \( y_{MoII} \) without label-switched and with label-switched respectively. We obtained similar plots for the other two chains which are not shown here to save space. Observe that our Bayesian technique captures the symmetry and asymmetry of the posterior distributions of the parameter \( \varphi \) in the case of without label-switched.

Table 3.7 shows the parameter estimation results for the posterior sample (only chain 3 is reported here since the rest produces almost the same results) for the simulated series \( y_{MoII} \) without label-switched. Column 2, 3, 5 and 6 in this table show the mean and median, standard deviation and lower and upper bound Highest Posterior Density (HPD) interval of the posterior distribution of each parameter respectively. Note that these estimates are close to the true value of the parameter. Computational time required to finish the MCMC algorithm for the specified runs and chains is shown at the bottom of the table. Similar Table is obtained in the case of with label-switched which is not shown here to save space. Moreover, the parameter estimation results are not very close in this case.

Likewise the \( y_{MoI} \) series, we implemented Sampietro’s algorithm and obtained less accurate parameter estimation results than ours (not reported here). In this circumstances, we can say that our improved MCMC algorithm for the \( \mathcal{M}_{AR} \) model fits well again with the simulated series \( y_{MoII} \) without label-switching problem.
3.8.2 Experiment with real datasets

Example 1: IBM stock prices

We consider the daily IBM common stock closing price from May 17, 1961 to November 2, 1962 (total 369 observations), taken from Box and Jenkins (1976). Among many researchers who had analysed this time series, we found Wong and Li (2000) who analysed it to fit a $\mathcal{M}_{AR}(3; 1, 1, 0)$ model for its first-differenced series by EM estimation method (see Figures 3.12). Note that our goal is to check the MCMC techniques that we have developed in this chapter for $\mathcal{M}_{AR}$ models. We call this first-differenced series by $y_{IBM}$ hereafter.

Table 3.8 shows that the estimated results is very close to the result obtained by Wong and Li (2000). In addition, it gives more informations including the HPD about the parameters. Furthermore, we have convergence diagrams and estimated marginal densities for each of the parameter (see Figure 3.13).

Figures 3.13 and 3.14 show the convergence diagrams and marginal densities of the posterior sample (from chain 2) for each parameter for the series $y_{IBM}$. We obtained similar plots for the other two chains which are not shown here to save space. Observe that our Bayesian technique captures the symmetry and asymmetry of the posterior distributions of the parameter $\varphi$.

Table 3.8 shows the parameter estimation results for the posterior sample. Column 3-6 in this table show the mean, standard deviation, median and Highest Posterior Density (HPD) interval of the posterior distribution of each parameter respectively. Note that these estimates are close to the true value of the parameter. Computational time required to finish the MCMC algorithm is shown at the bottom of the table.

In conclusion, we can say that our Bayesian approach for the $\mathcal{M}_{AR}$ model fit reasonably well with $y_{IBM}$ data compare with results obtained in Wong and Li (2000) by using EM algorithm.

Example 2: The 3-year Treasury Constant Maturity Rate

This is an daily interest rate series, published by the Federal Reserve Board, from January 2, 1996 and December 30, 2005 (2502 observations). The original data series, say $y_t$, is non-stationary while the first differences of the logarithm of the series, $r_t = \log y_t - \log y_{t-1}$, seems to be stationary (see Figure 3.15). We call this series as $y_{DGS3}$ throughout the thesis.

We run three MCMC chains for the data $y_{DGS3}$ with the same simulation number and discarded initial half of it as burn-in. We found $\mathcal{M}_{AR}(3; 0, 6, 7)$ be the best fit model for the data $y_{DGS3}$. Due to the large number of parameters...
involved in this model, longer computational time was required (2.574774 hours reported at the end of computation).

Figure 3.17 shows the convergence diagrams and marginal densities of the posterior sample for $\omega$, $\mu$, $\gamma$ and $\sigma$ for the series $y_{DGS}$ (only one chain reported here). We found the label-switching problem in this case (see Figure 3.17) and successfully solved as well by imposing the mixing weight constraint $\omega_1 < \omega_2 < \omega_3$. Observe that our Bayesian technique captures the symmetry (in most cases here) and asymmetry of the posterior distributions of the parameter $\omega$, $\mu$, $\gamma$ and $\sigma$. Since we consider a very higher AR order model for the $y_{DGS3}$ data we avoid presenting the corresponding Figures and Tables for $\phi$ here.
Figure 3.4: Graphical analysis of the simulated series $y_{MoI}$. 

(a) Original series

(b) Histogram

(c) ACF of the original series
CHAPTER 3. MIXTURE OF GAUSSIAN AR COMPONENTS

Figure 3.5: MCMC output of sampled values of mixing proportions, shift, scale and AR coefficients when fitting the $\mathcal{M}_{AR}(2; 1, 1)$ model to the simulated series $y_{MoI}$, obtained by setting our priors, using the priors given in Section 3.5 and the MCMC moves described in Section 3.6, showing no label-switching problem at all.
Figure 3.6: MCMC output of sampled values of mixing proportions, shift, scale and AR coefficients when fitting the $M_{AR}(2; 1, 1)$ model to the simulated series $y_{MoI}$, obtained by setting our priors, using the priors given in Section 3.5 and the MCMC moves described in Section 3.6, imposing the identifiability constraint $\omega_1 < \omega_2$ which is **not required** in this case. On the other hand, convergence of the AR coefficients have been lost by imposing the constraint.
Figure 3.7: Graphical output of sampled values of mixing proportions, shift, scale and AR coefficients when fitting the $\mathcal{M}_{AR}(2; 1, 1)$ model to the simulated series $y_{MoI}$, obtained by setting a different prior only for the AR coefficients suggested by Sampietro (2006), obtained by without label-switched.
Figure 3.8: Graphical output of sampled values of mixing proportions, shift, scale and AR coefficients when fitting the \( \mathcal{M}_{\text{AR}}(2; 1, 1) \) model to the simulated series \( y_{\text{MoI}} \), obtained by setting our priors, using the priors given in Section 3.5 and the MCMC moves described in Section 3.6, obtained by setting a different prior only for the AR coefficients suggested by Sampietro (2006), imposing the identifiability constraint \( \omega_1 < \omega_2 \) which is not required in this case. On the other hand, convergence of the AR coefficients have been lost by imposing the constraint.
CHAPTER 3. MIXTURE OF GAUSSIAN AR COMPONENTS

Figure 3.9: Graphical analysis of the simulated series $\text{yMoII}$.
Table 3.7: Estimation results of parameters when fitting the model $\mathcal{M}_{\text{AR}}(3; 2, 1, 3)$ to the series $y_{\text{MoII}}$ with $N = 100000$ and burn-in = 10000.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>HPD†</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.5542</td>
<td>0.04778</td>
<td>0.0002137</td>
<td>0.4637623</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.2535</td>
<td>0.03667</td>
<td>0.0001640</td>
<td>0.1818711</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.1924</td>
<td>0.03883</td>
<td>0.0001737</td>
<td>0.1175513</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.5263</td>
<td>0.412</td>
<td>0.001842</td>
<td>-1.336940</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.9344</td>
<td>1.008</td>
<td>0.004507</td>
<td>-2.925702</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-5.5479</td>
<td>1.634</td>
<td>0.007306</td>
<td>-8.811863</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>3.868</td>
<td>0.6038</td>
<td>0.002700</td>
<td>2.743558</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>6.887</td>
<td>1.2799</td>
<td>0.005724</td>
<td>4.403334</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>7.893</td>
<td>1.6306</td>
<td>0.007292</td>
<td>5.284299</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>0.9115</td>
<td>0.04553</td>
<td>0.0002036</td>
<td>0.8211350</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>-0.6332</td>
<td>0.04529</td>
<td>0.0002025</td>
<td>-0.7190058</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>-0.3109</td>
<td>0.09505</td>
<td>0.0004251</td>
<td>-0.4936217</td>
</tr>
<tr>
<td>$\phi_{31}$</td>
<td>1.3155</td>
<td>0.16465</td>
<td>0.0007363</td>
<td>0.9641369</td>
</tr>
<tr>
<td>$\phi_{32}$</td>
<td>-0.7033</td>
<td>0.15749</td>
<td>0.0007043</td>
<td>-0.9811393</td>
</tr>
<tr>
<td>$\phi_{33}$</td>
<td>0.1004</td>
<td>0.15621</td>
<td>0.0006986</td>
<td>-0.1981949</td>
</tr>
</tbody>
</table>

Computational time: 34.67351 mins

† Highest Posterior Density
‡ Median
CHAPTER 3. MIXTURE OF GAUSSIAN AR COMPONENTS

Figure 3.10: Graphical output of sampled values of mixing proportions, shift, scale, and mean, AR coefficients respectively when fitting the $\mathcal{M}_{AR}(3; 2, 1, 3)$ model to the simulated series $y_{MoII}$. 
Figure 3.11: MCMC output of the posterior sampled values of mixing proportions, shift, scale, and mean, AR coefficients respectively when fitting the $M_{AR}(3; 2, 1, 3)$ model to the simulated series $y_{MoII}$, imposing the identifiability constraint $\omega_1 < \omega_2 < \omega_3$ to solve label-switching problem.
Figure 3.12: Graphical output of *yIBM* series, showing that the original series is not stationary while the first-differenced is stationary.
Table 3.8: Estimation results of parameters when fitting the $M_{AR}(3; 2,1,3)$ model to the series $y_{IBM}$ with $N = 100000$ and burn-in = 10000.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior statistics</th>
<th>HPD$^\dagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.5411</td>
<td>0.04354</td>
</tr>
<tr>
<td></td>
<td>(0.5705)$^\dagger$</td>
<td></td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.3163</td>
<td>0.03392</td>
</tr>
<tr>
<td></td>
<td>(0.3394)</td>
<td></td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.1426</td>
<td>0.03338</td>
</tr>
<tr>
<td></td>
<td>(0.1641)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.3059</td>
<td>0.5834</td>
</tr>
<tr>
<td></td>
<td>(0.3214)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-1.5774</td>
<td>1.6993</td>
</tr>
<tr>
<td></td>
<td>(-1.6337)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.5735</td>
<td>2.9116</td>
</tr>
<tr>
<td></td>
<td>(-0.5202)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>4.845</td>
<td>0.4554</td>
</tr>
<tr>
<td></td>
<td>(5.089)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>6.510</td>
<td>0.7254</td>
</tr>
<tr>
<td></td>
<td>(6.980)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>10.421</td>
<td>1.8033</td>
</tr>
<tr>
<td></td>
<td>(11.517)</td>
<td></td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.3253</td>
<td>0.1859</td>
</tr>
<tr>
<td></td>
<td>(-0.2177)</td>
<td></td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.7706</td>
<td>0.1511</td>
</tr>
<tr>
<td></td>
<td>(0.8998)</td>
<td></td>
</tr>
</tbody>
</table>

Computational time: 55.52408 mins

$^\dagger$ Highest Posterior Density
$^\ddagger$ Median
Figure 3.13: MCMC output of sampled values of parameters when fitting the $\mathcal{M}_{\text{AR}}(3; 1, 1, 0)$ model to the series $y_{\text{IBM}}$, showing the presence of the so called label-switching problem for mixture models.
Figure 3.14: MCMC output of sampled values of parameters when fitting the $M_{AR}(3; 1, 1, 0)$ model to the series $y_{IBM}$, the label-switching problem by imposing the identifiability constraint $\omega_1 < \omega_2 < \omega_3$. 

(a) trace and density of $\omega$  
(b) trace and density of $\gamma$  
(c) trace and density of $\sigma$  
(d) trace and density of $\phi$
Figure 3.15: Graphs of The 3-year Treasury Constant Maturity Rate series which is named as $y_{DGS3}$ series.
Figure 3.16: MCMC output of sampled values of mixing proportions, shift, mean and scale parameter when fitting the model $\mathcal{M}_{AR}(3; 0, 6, 7)$ to the $\text{yDGS3}$ series, using the priors and posteriors given in Sections 3.5 and 3.6 respectively, showing the label-switching problem in a small range.
CHAPTER 3. MIXTURE OF GAUSSIAN AR COMPONENTS

Figure 3.17: MCMC output of sampled values of mixing proportions, shift, mean and scale parameter when fitting the model $M_{AR}(3; 0, 6, 7)$ to the $y_{DGS3}$ series, using the priors and posteriors given in Sections 3.5 and 3.6 respectively, showing that the identifiability constraint $\omega_1 < \omega_2 < \omega_3$ has been successfully imposed.
Chapter 4

Mixture of Gaussian ARMA components

4.1 Introduction

In this chapter, we discuss finite mixtures of autoregressive moving average (ARMA) models which can be considered as extension of $\mathcal{M}_{\text{AR}}$ models. We denote this class of nonlinear time series model by $\mathcal{M}_{\text{ARMA}}$. Likewise the previous chapter, we develop a complete Bayesian analysis for $\mathcal{M}_{\text{ARMA}}$ models with different prior settings for the AR coefficients. In addition, we have to consider new parameters, moving average (MA) coefficients and MA orders along with respective priors for them. Therefore, one can analyse the $\mathcal{M}_{\text{AR}}$ model with this new prior settings of AR coefficients provided the MA order is fixed to zero and can compare the results obtained with different setup in the previous chapter.

In the literature, we found very few research paper about mixtures of ARMA models (e.g. see Xiong and Yeung (2002)). But none of them considered Bayesian tools for estimating the model parameters. So, we are delighted to develop a Bayesian treatment for $\mathcal{M}_{\text{ARMA}}$ model in this chapter. Eventually, the analysis of the $\mathcal{M}_{\text{ARMA}}$ model will provide the more general structures including the $\mathcal{M}_{\text{AR}}$ model. Moreover, we could analyse a mixture of MA components by setting the AR order to zero.
4.2 Background

An autoregressive moving average (ARMA) process $y_t$ of order $p$ and $q$, is usually denoted by ARMA($p, q$) and is defined as

$$y_t = \mu + \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu) + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \tau^{-1}), \quad (4.1)$$

where $\tau$ is the precision of the error term; $\phi_i$s are the coefficients of the AR($p$) process, $\theta_j$s are the coefficients of the MA($q$) process.

Let $g(L) = \sum_{i=1}^{p} \phi_i L^i$, and $h(L) = \sum_{j=1}^{q} \theta_j L^j$, where $L$ is the lag operator. Since all moving average MA processes with finite coefficients are stationary, the stationarity of ARMA process depends only on the autoregressive AR part of the model. To ensure stationarity and invertibility, it is very common practice to impose the following two constraints in the ARMA model:

(CS): All roots of $1 - g(L) = 0$ must lie outside the unit circle.

(CI): All roots of $1 + h(L) = 0$ must lie outside the unit circle.

Now, if we denote:

$$\nu_t = \mu + \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu) + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad (4.2)$$

the conditional distribution of $y_t$ can be written as

$$y_t \mid \mathcal{G}_{t-1} \sim \mathcal{N}(y_t; \nu_t, \tau^{-1}), \quad (4.3)$$

where “$\mid \mathcal{G}_{t-1}$” is used to denote conditioning on the past observations and on all other variables.
CHAPTER 4. MIXTURE OF GAUSSIAN ARMA COMPONENTS

4.3 The $\mathcal{M}_{\text{ARMA}}$ model

Let $y = \{y_t\}_{t=1}^T$ be a time series. Then the $g$-components mixture of ARMA model (see Xiong and Yeung (2002)) is defined as follows:

\[
\begin{align*}
\mathbb{P}(y_t \mid G_{t-1}) = & \sum_{k=1}^g \omega_k \mathcal{N}(y_t | \nu_{k,t}^{-1}, \nu_{k,t}), \\
\nu_{k,t} = & \mu_k + \sum_{j=1}^{p_k} \phi_{kj} (y_{t-j} - \mu_k) + \sum_{j=1}^{q_k} \theta_{kj} \varepsilon_{t-j}, \quad (k = 1, \ldots, g), \\
\omega_k > 0, \quad (k = 1, \ldots, g); \quad \sum_{k=1}^g \omega_k = 1
\end{align*}
\]

where where " $G_{t-1}$" is used to denote conditioning on the past observations and on all other variables; $\mu_k$ denotes the mean of the $k$th component; $\omega_k$ denotes the mixing weight of the $k$th component; $\mathcal{N}(x; a, b^{-1})$ denotes the univariate normal density function of $x$ with mean $a$ and precision $b$. For details, see Section 2.11.5.

Clearly, the $\mathcal{M}_{\text{ARMA}}$ model given by (4.4) is a generalization of the model $\mathcal{M}_{\text{AR}}$ model. Indeed, if we set $q_k = 0$, for $k = 1, \ldots, g$ the model (4.4) reduces to the model of $\mathcal{M}_{\text{AR}}$ model with Gaussian error term.

Now we establish notation for all parameters as:

\[
\varphi = (\omega, \mu, \tau, \phi, \theta),
\]

where $\omega = (\omega_1, \ldots, \omega_g)$, denotes the mixing weight vector of the model; $\mu = (\mu_1, \ldots, \mu_g)$, denotes the mean vector; $\tau = (\tau_1, \ldots, \tau_g)$, denotes the precision vector of the error term of the model; $\phi_k = (\phi_{k,1}, \ldots, \phi_{k,p_k})$, for $(k = 1, \ldots, g)$, denotes the AR coefficients of $k$-th component, and $\theta_k = (\theta_{k,1}, \ldots, \theta_{k,q_k})$, for $(k = 1, \ldots, g)$, denotes the MA coefficients of the $k$-th component.

We also denote $p = (p_1, \ldots, p_g)$, the vector of orders of the AR part of the model, and $q = (q_1, \ldots, q_g)$, as the vector of orders of the MA part of the model.

Likewise the $\mathcal{M}_{\text{AR}}$ and $\mathcal{M}'_{\text{AR}}$ model, the conditional expectation of $y_t$ given the past information is

\[
E(y_t \mid G_{t-1}) = \sum_{k=1}^g \omega_k \nu_{k,t}, \quad (4.5)
\]

which depends on the past values of the time series and the shape of the conditional distributions may change from unimodal to multi-modal. Therefore, the conditional expectation of $y_t$ given by (4.5) may not be the best predictor of the future values.
4.3.1 Hypotheses of the model

Let $G_k(L) = \sum_{j=1}^{p_k} \phi_{kj} L^j$ and $H_k(L) = \sum_{j=1}^{q_k} \theta_{kj} L^j$ for $k = 1, \ldots, g$, where $L$ is the lag operator. We assume that the following hypotheses are true for our $\mathcal{M}_{\text{ARMA}}^t$ model (4.4):

To ensure stationarity and invertibility, it is very common practice to impose the following two constraints in the $\mathcal{M}_{\text{ARMA}}$ model (Nakatsuma (2000)):

$(C_S)$: all roots of 
\[1 - G_k(L) = 0, \quad (k = 1, \ldots, g)\]
are outside the unit circle.

$(C_I)$: all roots of 
\[1 + H_k(L) = 0, \quad (k = 1, \ldots, g)\]
are outside the unit circle.

The stationarity condition is the same as for the $\mathcal{M}_{\text{AR}}$ model and does not involve MA parameters.

4.4 Missing data formulation and the likelihood function

For the given (observed) time series $y = \{y_t\}_{t=1}^T$ and the unobserved series $\varepsilon = \{\varepsilon_t\}_{t=1}^T$, where $\varepsilon_t \sim \mathcal{N}(0, \tau^{-1})$, $(*) = 1, \ldots, g$, the likelihood function of the $\mathcal{M}_{\text{ARMA}}$ model takes a complicated form. But it can be simplified by introducing the usual missing data formulation for mixture setups (see e.g. Diebolt and Robert (1994)), where a set of allocation variables $z = \{z_t\}_{t=1}^T$ are defined such that

\[
\Pr(z_t = k \mid g, \omega_1, \ldots, \omega_g) = \omega_k, \quad (k = 1, \ldots, g). \tag{4.6}
\]

Therefore, our observed and unobserved series $y$ and $\varepsilon$ are being completed with the help of a missing data set $z$. This means that conditional on $z$, $y$ and $\varepsilon$ are labelled with specific components of the mixture. For example, if $z_t = k$, then $y_t$
and $\varepsilon_t$ are assumed to arise from the $k$th component of the mixture. i.e.

$$y_t \mid g_{t-1}, z_t \sim \begin{cases} 
N(y_t; \nu_{1,t}, \tau_1^{-1}) & \text{if } z_t = 1, \\
\vdots & \\
N(y_t; \nu_{k,t}, \tau_k^{-1}) & \text{if } z_t = k, \\
\vdots & \\
N(y_t; \nu_{g,t}, \tau_g^{-1}) & \text{if } z_t = g.
\end{cases} \quad (4.7)$$

As a consequence, the full conditional likelihood function separates into $g$ parts where each one concerning the data assigned to each of the $g$ mixture components,

$$\mathcal{L}(\varphi \mid y, z) = \prod_{k=1}^{g} \prod_{t: z_t = k} \mathcal{N}(y_t; \nu_{k,t}, \tau_k^{-1})$$

$$= \prod_{k=1}^{g} \prod_{t: z_t = k} \sqrt{\frac{\tau_k}{2\pi}} \exp \left[ -\frac{\tau_k}{2} (y_t - \hat{\nu}_{k,t})^2 \right], \quad (4.8)$$

where

$$\hat{\nu}_{k,t} = \mu_k + \sum_{j=1}^{p_k} \phi_{kj} (y_{t-j} - \mu_k) + \sum_{j=1}^{q_k} \theta_{kj} \hat{\varepsilon}_{t-j}, \quad (4.9)$$

and the innovation term $\hat{\varepsilon}_t$ is defined as:

$$\hat{\varepsilon}_t = \begin{cases} 
0, & (t \leq 0), \\
y_t - \hat{\nu}_{k,t}, & (t = 1, \ldots, T).
\end{cases} \quad (4.10)$$

### 4.5 Priors setup

For fixed component number $g$, we will have a fixed ordered $\mathcal{M}_{ARMA}$ model. We set all the prior distributions, $\pi(\varphi \mid g)$, of the model parameters as same as that of the $\mathcal{M}_{AR}$ model except the parameter $\phi$ and $\theta$. We state these priors again here for convenience (for details see Section 3.5):
4.5.1 Priors of $\omega$, $\mu$ and $\tau$

$$\omega | g \sim D(\omega; \delta_1, \ldots, \delta_g)$$  \hspace{1cm} (4.11)

$$\mu_k \overset{\text{iid}}{\sim} \mathcal{N}(\mu_k; \zeta, \kappa^{-1}), \quad (k = 1, \ldots, g)$$  \hspace{1cm} (4.12)

$$\tau_k | \lambda \overset{\text{iid}}{\sim} \mathcal{G}(\tau_k; c, \lambda), \quad (k = 1, \ldots, g)$$

$$\lambda \overset{\text{iid}}{\sim} \mathcal{G}(\lambda; a, b).$$  \hspace{1cm} (4.13)

4.5.2 Priors for $\phi$ and $\theta$

Let $\Xi_p$ and $\Xi_q$, where $p = \max\{p_1, \ldots, p_g\}$ and $q = \max\{q_1, \ldots, q_g\}$, for the regions in which these conditions hold, so that the stationarity and invertibility region is $\Xi_p \times \Xi_q$. The form of the region $\Xi_i$ is very complex for $i \geq 3$, while $\Xi_1$ denotes the interval $(-1, 1)$ and $\Xi_2$ denotes a triangle.

We specify prior distributions for $\phi = (\phi_1, \ldots, \phi_g)$ and that for $\theta = (\theta_1, \ldots, \theta_g)$ based on the method proposed by Jones (1987). The method consists of a one-to-one correspondence between the AR parameters (or the MA parameters) and its partial autocorrelations. It comprises generating the partial autocorrelations as independent beta variates, with appropriate parameter values, and transforming to get the parameters of the process. This equivalents to taking them from a uniform distribution on $\Xi_p \times \Xi_q$.

Following Jones (1987) and Barndorff-Nielsen and Schou (1973), we choose the priors for $\phi = (\phi_1, \ldots, \phi_g)$ and $\theta = (\theta_1, \ldots, \theta_g)$:

$$\phi_k \sim \text{Uniform on } \Xi_p \iff \rho_{ki}^\phi \overset{\text{iid}}{\sim} \mathcal{B}_{(-1,+1)} \left( \rho_{ki}^\phi; \left\lfloor \frac{i + 1}{2} \right\rfloor, \left\lceil \frac{i}{2} \right\rceil + 1 \right),$$

$$(k = 1, \ldots, g, i = 1, \ldots, q_k)$$  \hspace{1cm} (4.14)

$$\theta_k \sim \text{Uniform on } \Xi_q \iff \rho_{ki}^\theta \overset{\text{iid}}{\sim} \mathcal{B}_{(-1,+1)} \left( \rho_{ki}^\theta; \left\lfloor \frac{i + 1}{2} \right\rfloor, \left\lceil \frac{i}{2} \right\rceil + 1 \right),$$

$$(k = 1, \ldots, g, i = 1, \ldots, q_k)$$  \hspace{1cm} (4.15)

where $\rho_{ki}^\phi$ be the partial autocorrelation coefficients at lag $i$ for the $k$th component; $\mathcal{B}_{(-1,+1)}(x; a, b)$ represents a generalized beta distribution defined on $(-1, +1)$.
and where \([a]\) means the integer part of \(a\).

### 4.5.3 Priors for \(g\), \(p\) and \(v\)

To complete Bayesian analysis of our parameters, we stress that \(g\), \(p\) and \(v\) are stochastic also. The priors for these can be set as:

\[
g \sim U(g; 1, g_{\max}) \tag{4.16}
\]

\[
p_k | g \overset{iid}{\sim} U(p_k; 0, p_{\max}), \quad (k = 1, \ldots, g) \tag{4.17}
\]

\[
q_k | g \overset{iid}{\sim} U(q_k; 0, q_{\max}), \quad (k = 1, \ldots, g) \tag{4.18}
\]

where \(U(x | a, b)\) denotes the discrete uniform distribution for \(x = a, a + 1, \ldots, b\) and \(g_{\max}, p_{\max}, q_{\max}\) and \(v_{\max}\) is fixed. For details of the distributions mentioned above see Section 2.11.

### 4.5.4 Formulae for hyperparameters

In the prior settings, the \(M_{ARMA}\) model has the hyperparameters as that of the \(M_{AR}\) model. We re-written the formulae here (for details, see Section 3.5.1):

\[
\delta_1 = \cdots = \delta_g = 1,
\]

\[
R_y = \max(y) - \min(y), \quad \zeta = \min(y) + \frac{R_y}{2}, \quad \kappa = \frac{1}{R_y},
\]

\[
c = 2, \quad a = 0.2, \quad b = \frac{100}{c R_y^2}. \tag{4.19}
\]

### 4.6 MCMC moves for parameter estimation

Let \(\Lambda = (g, p, q)\) be the model index. We apply MCMC component-wise techniques to calculate the full conditional posterior distribution \(\pi(\varphi | y, \Lambda)\) in which each of the parameter will update separately. Moreover, we need to update the latent variable \(z\) as well. There, we have a set of \textbf{NINE} moves in total which are discussed below:
4.6.1 Move I: updating $\omega$

This is a Gibbs-type move and let $\varphi_{-\omega}$ denotes the remaining parameters except $\omega$. Then the full conditional posterior distribution $\pi(\omega | \varphi_{-\omega}, y, z)$ can be obtained in similar fashion as in Section 3.6.2:

$$\omega | \varphi_{-\omega}, y, z \sim D(\omega; \delta_1 + n_1, \ldots, \delta_g + n_g),$$

(4.20)

where $n_k = \sum_{t:z_t=k} 1$.

4.6.2 Move II: updating $\mu$

Let $\varphi_{-\mu}$ denotes the remaining parameters except $\mu$. Then the full conditional posterior distribution $\pi(\mu | \varphi_{-\mu}, y, z)$ can be obtained by applying a Gibbs-type
move as below:

\[ \pi(\mu \mid \varphi-\mu, y, z) \propto \mathcal{L}(\varphi, z)\pi(\mu) \]

\[ = \prod_{k=1}^{g} \prod_{t: z_t = k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, \tau_k^{-1}) \prod_{k=1}^{g} \mathcal{N}(\mu_k; \zeta, \kappa^{-1}) \]

Therefore, the full conditional component-wise posterior distribution for the \( k \)-th component is:

\[ \pi(\mu_k \mid \varphi-\mu_k, y, z) \propto \prod_{t: z_t = k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, \tau_k^{-1}) \cdot \mathcal{N}(\mu_k; \zeta, \kappa^{-1}) \]

\[ \propto \exp\left[ -\frac{\tau_k}{2} \sum_{t: z_t = k} \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t-i} - \mu_k) + \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \right\}^2 \right] \]

\[ \cdot \exp\left[ -\frac{\kappa}{2} (\mu_k - \zeta)^2 \right] \]

Now the term \((**\)) can be simplified in a similar fashion as in the case of \(\mathcal{M}_{An} \) model (see Section 3.6.3):

\[ \sum_{t: z_t = k} \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t-i} - \mu_k) + \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \right\}^2 \]

\[ = \sum_{t: z_t = k} \left\{ \left( y_{t} - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i} + \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \right) - \left( 1 - \sum_{i=1}^{p_k} \phi_{k,i} \right) \mu_k \right\}^2 \]

\[ := e_{k,t}, \text{ say} \]

\[ := b_k, \text{ say} \]

\[ = \sum_{t: z_t = k} (e_{k,t} - \bar{e}_k)^2 + n_k (\bar{e}_k - b_k \mu_k)^2, \]

since \( \sum_{t: z_t = k} (e_{k,t} - \bar{e}_k) = 0 \), and \( n_k = \sum_{t: z_t = k} 1 \).
Substituting the simplified form of (**) in the equation back and after few steps of simplification (see, Section 3.6.3), we have the full conditional posterior distribution of $\mu$ as below:

$$
\mu_k \mid \phi_{-\mu_k}, y, z \sim N\left(\mu_k; \frac{\tau_k n_k \bar{e}_k b_k + \kappa \zeta}{\tau_k n_k b_k^2 + \kappa}, \frac{1}{\tau_k n_k b_k^2 + \kappa}\right), \quad (k = 1, \ldots, g)
$$

(4.21)

where

- $\bar{e}_k = \frac{1}{n_k} \sum_{t: z_t = k} e_{k,t}$
- $e_{k,t} = y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i} - \sum_{j=1}^{q_k} \theta_{k,j} \hat{e}_{t-j}$
- $n_k = \sum_{t: z_t = k} 1$
- $b_k = 1 - \sum_{i=1}^{p_k} \phi_{k,i}$

4.6.3 Move III: updating $\phi$

Here we update $\rho_{k,i}^\phi$, $k = 1, \ldots, g$ by a Metropolis-Hasting mechanism which is equivalent to update $\phi$. The algorithm for this move are as follows:

- A candidate $\rho_{k,i}^\phi^* = (\rho_{k,i}^\phi)^*$ is generated by a normal density truncated in $(-1, 1)$ and centred in the current state of the chain $\rho_{k,i}^\phi$ so that we can define the proposal density as below:

$$
q_{\phi}(\rho_{k,i}^\phi, \rho_{k,i}^\phi^*) = N_{(-1,+1)} \left(\rho_{k,i}^\phi; \rho_{k,i}^\phi^*, \tau_{q_{\phi}^{-1}}\right), \quad (k = 1, \ldots, g; i = 1, \ldots, p_k)
$$

(4.22)

where the precision $\tau_{q_{\phi}^{-1}}$ is chosen in order to obtain a satisfactory acceptance rate (20% or more, see Sampietro (2006)).

Using the distribution defined above by (4.22), we generate a vector $\rho_{k}^\phi^* = (\rho_{k1}^\phi^*, \ldots, \rho_{kp_k}^\phi^*)$ for the partial autocorrelations which will be then used to derive
the corresponding parameter vector \( \phi_k^* = (\phi_{k1}, \ldots, \phi_{kp_k}) \) through the transformation proposed by Barndorff-Nielsen and Schou (1973).

Recall that the ratio in the acceptance probability for a M-H algorithm can be defined as
\[
A(\phi_k, \phi_k^*) = \min(1, \Omega),
\]
(4.23)
where
\[
\Omega = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}.
\]

First of all, from the Equation (4.8), the likelihood ratio is simply given by:
\[
\text{likelihood ratio} = \frac{\prod_{t:z_t=k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, \tau^{-1}_k) \prod_{t:z_t=k} \mathcal{N}(y_t; \nu_{k,t}, \tau^{-1}_k)}{\prod_{t:z_t=k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, \tau^{-1}_k)}
\]
\[
= \exp\left[-\frac{\tau_k}{2} \sum_{t:z_t=k} \left\{ (y_t - \hat{\nu}_{k,t})^2 - (y_t - \nu_{k,t})^2 \right\}\right],
\]
(4.24)
where \( \hat{\nu}_{k,t} = \mu_k + \sum_{j=1}^{p_k} \phi_{kj}^* (y_{t-j} - \mu_k) - \sum_{j=1}^{q_k} \theta_{kj} \epsilon_{t-j} \).

Since all the parameters but the \( \phi \) do not change, the prior ratio in terms of the partial autocorrelations \( \rho^\phi \) reduces to the ratio as below (see the Equation (4.14)):
\[
\text{prior ratio} = \frac{\prod_{i=1}^{p_k} \mathcal{B}_{(-1,1)}(\rho_{ki}; \left[ \frac{i}{2} \right], \left[ \frac{i+1}{2} \right] + 1)}{\prod_{i=1}^{p_k} \mathcal{B}_{(-1,1)}(\rho_{ki}; \left[ \frac{i}{2} \right], \left[ \frac{i+1}{2} \right] + 1)}
\]
\[
= \prod_{i=1}^{p_k} \frac{\Gamma\left(\frac{i+1}{2}\right) \Gamma\left(\frac{i+1}{2} + 1\right)}{\Gamma\left(\frac{i}{2} + 1\right) \Gamma\left(\frac{i}{2} + 1 + 1\right)} \cdot \frac{\rho_{ki}^* - (-1)^{\left[\frac{i}{2}\right]} \rho_{ki}}{(1 - (-1)^{\left[\frac{i}{2}\right]} \rho_{ki})^{\left[\frac{i}{2}\right]} + 1}
\]
\[
= \prod_{i=2}^{p_k} \frac{(\rho_{ki}^* + 1)^{\left[\frac{i}{2}\right]} - (1 - \rho_{ki})^{\left[\frac{i}{2}\right]} - 1}{(\rho_{ki}^* + 1)^{\left[\frac{i}{2}\right]} - (1 - \rho_{ki})^{\left[\frac{i}{2}\right]} - 1}
\]
(4.25)
and, eventually from the Equation (4.22) the proposal ratio is:

$$\text{proposal ratio} = \frac{\prod_{i=1}^{p_k} \mathcal{N}(-1,+1) \left( \rho_{ki}^{\phi}; \rho_{ki}^{\phi*}, \tau_{q^\phi}^{-1} \right)}{\prod_{i=1}^{p_k} \mathcal{N}(-1,+1) \left( \rho_{ki}^{\phi*}; \rho_{ki}^{\phi}, \tau_{q^\phi}^{-1} \right)}$$

By using (4.24), (4.25) and (4.26) we have

$$\Omega_{\phi} = \exp \left[ -\frac{\tau_k}{2} \sum_{t:z_t=k} \left\{ (y_t - \nu_{k,t}^*)^2 - (y_t - \nu_{k,t})^2 \right\} \right]$$

$$\times \prod_{i=2}^{p_k} \left( \rho_{ki}^{\phi*} + 1 \right)^{-1} (1 - \rho_{ki}^{\phi*}) \left[ \frac{1}{2} \right]$$

$$\times \prod_{i=1}^{p_k} \Phi_N \left( +1; \rho_{ki}^{\phi}, \tau_{q^\phi}^{-1} \right) - \Phi_N \left( -1; \rho_{ki}^{\phi}, \tau_{q^\phi}^{-1} \right),$$

where $\Phi_N$ is the normal cumulative distribution function.

By using (4.24), (4.25) and (4.26) we have

$$\Omega_{\phi} = \exp \left[ -\frac{\tau_k}{2} \sum_{t:z_t=k} \left\{ (y_t - \nu_{k,t}^*)^2 - (y_t - \nu_{k,t})^2 \right\} \right]$$

$$\times \prod_{i=2}^{p_k} \left( \rho_{ki}^{\phi*} + 1 \right)^{-1} (1 - \rho_{ki}^{\phi*}) \left[ \frac{1}{2} \right]$$

$$\times \prod_{i=1}^{p_k} \Phi_N \left( +1; \rho_{ki}^{\phi}, \tau_{q^\phi}^{-1} \right) - \Phi_N \left( -1; \rho_{ki}^{\phi}, \tau_{q^\phi}^{-1} \right),$$

where $\Phi_N$ is the normal cumulative distribution function.

By using (4.24), (4.25) and (4.26) we have

$$\Omega_{\phi} = \exp \left[ -\frac{\tau_k}{2} \sum_{t:z_t=k} \left\{ (y_t - \nu_{k,t}^*)^2 - (y_t - \nu_{k,t})^2 \right\} \right]$$

$$\times \prod_{i=2}^{p_k} \left( \rho_{ki}^{\phi*} + 1 \right)^{-1} (1 - \rho_{ki}^{\phi*}) \left[ \frac{1}{2} \right]$$

$$\times \prod_{i=1}^{p_k} \Phi_N \left( +1; \rho_{ki}^{\phi}, \tau_{q^\phi}^{-1} \right) - \Phi_N \left( -1; \rho_{ki}^{\phi}, \tau_{q^\phi}^{-1} \right),$$

(4.27)

### 4.6.4 Move IV: updating $\theta$

Here we update $\rho_{ki}^{\phi}$, $k = 1, \ldots, g$ by a Metropolis-Hasting mechanism. This is exactly the same technique as discussed in the previous move, Section 4.6.3 and
we just need to replace $\phi$ by $\theta$ in all notations. This saves us from further complex calculation and time as well.

A candidate $\rho_{ki}^\theta$ is generated by a normal density truncated in $(-1, +1)$ and centred in the current state of the chain $\rho_{ki}^\theta$ so that we can define the proposal density as below:

$$q_\theta(\rho_{ki}^\theta, \rho_{ki}^\theta) = \mathcal{N}(-1, +1) \left( \rho_{ki}^\theta; \rho_{ki}^\theta, \tau^{-1}_{q_\theta} \right), \quad (k = 1, \ldots, g; i = 1, \ldots, p_k),$$

where the precision $\tau^{-1}_{q_\theta}$ is chosen in order to obtain a satisfactory acceptance rate (20% or more, see Sampietro (2006)). Using the distribution defined above by (4.28), we generate a vector $\rho_k^\theta = (\rho_{k1}^\theta, \ldots, \rho_{kp_k}^\theta)$ for the partial autocorrelations which will be then used to derive the corresponding parameter vector $\theta_k^* = (\theta_{k1}^*, \ldots, \theta_{kp_k}^*)$ through the transformation proposed by Barndorff-Nielsen and Schou (1973).

Therefore, we have the acceptance probability for this move:

$$A(\theta_k, \theta_k^*) = \min(1, \Omega_\theta), \quad (4.29)$$

where

$$\Omega_\theta = \exp \left[ -\frac{\tau_k}{2} \sum_{t: z_t = k} \left\{ (y_t - \hat{\nu}_{k,t}^*)^2 - (y_t - \hat{\nu}_{k,t})^2 \right\} \right] \times \prod_{i=2}^{p_k} \left( \rho_{ki}^\theta + 1 \right)^{\frac{i+1}{2} - 1} \left( 1 - \rho_{ki}^\theta \right)^{\frac{i}{2}} 
\times \prod_{i=1}^{p_k} \Phi_N \left( +1; \rho_{ki}^\theta, \tau^{-1}_{q_\theta} \right) - \Phi_N \left( -1; \rho_{ki}^\theta, \tau^{-1}_{q_\theta} \right) \times \Phi_N \left( +1; \rho_{ki}^\theta, \tau^{-1}_{q_\theta} \right) - \Phi_N \left( -1; \rho_{ki}^\theta, \tau^{-1}_{q_\theta} \right),$$

(4.30)

where $\hat{\nu}_{k,t}^* = \mu_k + \sum_{j=1}^{p_k} \phi_{kj} (y_{t-j} - \mu_k) - \sum_{j=1}^{q_k} \theta_{kj}^* z_{t-j}$ and $\Phi_N$ is the normal cumulative distribution function.
CHAPTER 4. MIXTURE OF GAUSSIAN ARMA COMPONENTS

4.6.5 Move V: updating $\lambda$ and $\tau$

Let $\varphi_{-\lambda}$ denote the remaining parameters except $\lambda$ and $\varphi_{-\tau}$ denotes the remaining parameters except $\tau$. The full conditional posterior distribution of $\lambda$ and $\tau$ can be obtained by applying Gibbs-typed move in the similar way as described for the $\mathcal{M}_{\text{AR}}$ model in Section 3.6.5. We state the results here only.

Firstly, the full conditional posterior distribution of $\lambda$ is:

$$
\lambda \mid \varphi_{-\lambda}, y, z \sim \mathcal{G}(\lambda; a + gc, b + \sum_{k=1}^{g} \tau_k), \quad (k = 1, \ldots, g).
$$

(4.31)

Secondly, the full conditional posterior distribution of $\tau$ is:

$$
\tau_k \mid \varphi_{-\tau}, y, z \sim \mathcal{G}(\tau_k; c + \frac{n_k}{2}, \lambda + \frac{1}{2} \sum_{t: z_t = k} (y_t - \hat{\nu}_{k,t})^2), \quad (k = 1, \ldots, g)
$$

(4.32)

where $\nu_{k,t}$ is defined as in Equation (4.4). For details of these derivations, see Section 3.6.5.

4.6.6 Move VI: updating $z$

This is a Gibbs-type move. From (4.6) and (4.7), the full conditional posterior probability of the location variables $z_t = k$ (i.e. the observations $y_t$ has been generated by the $k$-th mixture component) is,

$$
\Pr(z_t = k \mid y_t, \varphi) = \frac{1}{C_3} \omega_k \sqrt{\tau_k} \exp \left[ -\frac{\tau_k}{2} (y_t - \hat{\nu}_{k,t})^2 \right], \quad (k = 1, \ldots, g),
$$

(4.33)

where $C_3 = \sum_{u=1}^{g} \omega_u \exp \left[ -\frac{\tau_u}{2} (y_t - \nu_{u,t})^2 \right]$, is the normalizing constant; $\nu_{*,t}$ is defined as in the Equation (4.9).
4.7 MCMC moves for model determination

To get complete Bayesian flavour we must get the posterior distribution of the model index \( \Lambda = (g, p, q) \) which can be written and factorized as:

\[
\pi(\Lambda | y) = \pi(p | g, q, y) \times \pi(q | g, p, y) \times \pi(g | y)
\]  

(4.34)

So, it is enough to update AR order \( p \), MA order \( q \) and component number \( g \) separately to update the model index \( \Lambda \). We develop two additional MCMC moves to complete our Bayesian analysis as follows.

4.7.1 Move VII: updating of \( p \)

In a similar fashion, we can update the parameter \( p \) as discussed for the \( M_{AR} \) model in Section 3.7.1.

At first we need to choose a component, say \( k^{\dagger} \), randomly chosen in \( \{1, \ldots, g\} \). Then we can write the proposal for the new AR order of the \( k^{\dagger} \)-th component, \( p_{k^{\dagger}} \), as:

\[
p_{k^{\dagger}} = \begin{cases} 
p_{k^{\dagger}} - 1, & \text{with probability } d(p_{k^{\dagger}}), \\
p_{k^{\dagger}} + 1, & \text{with probability } b(p_{k^{\dagger}}), 
\end{cases}
\]  

(4.35)

where \( b(p_k) = 1 - d(p_k) \), for \( k = 1, \ldots, g \), \( d(0) = 0 \) and \( b(p_{\text{max}}) = 0 \).

Let \( \rho^\phi_{k^{\dagger}} \) be the proposal vector of the partial autocorrelations. When \( p_{k^{\dagger}} = p_{k^{\dagger}} - 1 \), the autoregressive coefficients \( \rho^\phi_{k^{\dagger}} = (\rho^\phi_{k^{\dagger}1}, \ldots, \rho^\phi_{k^{\dagger}p_{k^{\dagger}}}) \) is updated by simply discarding the last partial autocorrelation in \( \rho^\phi_{k^{\dagger}} = (\rho^\phi_{k^{\dagger}1}, \ldots, \rho^\phi_{k^{\dagger}p_{k^{\dagger}}}) \). On the other hand, when \( p_{k^{\dagger}} = p_{k^{\dagger}} + 1 \), the corresponding autocorrelations \( \rho^\phi_{k^{\dagger}} = (\rho^\phi_{k^{\dagger}1}, \ldots, \rho^\phi_{k^{\dagger}p_{k^{\dagger}}}, \rho^\phi_{k^{\dagger}p_{k^{\dagger}}}) \) is updated by adding new parameter \( \rho^\phi_{k^{\dagger}p_{k^{\dagger}}+1} \) in \( \rho^\phi_{k^{\dagger}} = (\rho^\phi_{k^{\dagger}1}, \ldots, \rho^\phi_{k^{\dagger}p_{k^{\dagger}}+1}) \) with

\[
\rho^\phi_{k^{\dagger}p_{k^{\dagger}}+1} \sim \mathcal{B}_{(-1, +1)}(\rho^\phi_{k^{\dagger}p_{k^{\dagger}}}, \left[\frac{p_{k^{\dagger}}}{2}\right], \left\lceil\frac{p_{k^{\dagger}}}{2}\right\rceil + 1).
\]

Then update the corresponding AR coefficients parameters by using the reparametrization discussed in Section 4.5.

To update the AR order of the model by using RJMCMC method, we have the
acceptance probability

\[ A_{p^\perp} = \min\{1, \Omega_{p^\perp}\} \]

and \( \Omega_{p^\perp} \) can be obtained as (for details, see Section 3.7.1):

\[
\Omega_{p^\perp} = \left\{
\begin{array}{ll}
L^\dagger \frac{b(p_{k^\perp}^\dagger)}{a(p_{k^\perp})} & \text{if } p_{k^\perp}^\dagger = p_{k^\perp} - 1, \\
L^\dagger \frac{d(p_{k^\perp}^\dagger)}{b(p_{k^\perp})} & \text{if } p_{k^\perp}^\dagger = p_{k^\perp} + 1,
\end{array}
\right.
\]

(4.36)

where

\[ L^\dagger = \exp\left[-\frac{\tau_k}{2} \sum_{t:z_t=k} \left\{ (y_t - \hat{\nu}_{k^\perp,t}^\dagger)^2 - (y_t - \hat{\nu}_{k^\perp,t}^\dagger - 1)^2 \right\}\right], \]

along with

\[ \hat{\nu}_{k^\perp,t}^\dagger = \mu_{k^\perp} + \sum_{j=1}^{q_{k^\perp}^\dagger} \phi_{k^\perp,j}^\dagger (y_t - \mu_{k^\perp}) - \sum_{j=1}^{q_{k^\perp}^\dagger} \theta_{k^\perp,j}^\dagger \hat{\epsilon}_{k^\perp,t-j}, \]

and \( \hat{\nu}_{k^\perp,t}^\dagger \) is as defined in Equation (4.4).

Note that the Equation (4.36) has quite simple form and \( \pi(p \mid g, y) \) is simply estimated by the proportions of every possible value for \( p \) in the sample obtained by the previous complete MCMC algorithm.

### 4.7.2 Move VIII: updating of \( q \)

This is almost the same as discussed in Section 4.7.1. For convention, we present that discussion very shortly here.

Likewise the previous move, at first we need to choose a component randomly, say \( k^\perp \), in \( \{1, \ldots, g\} \). Then we can write the proposal for the new MA order of the \( k^\perp \)-th component, \( q_{k^\perp}^\dagger \), as:

\[
q_{k^\perp}^\dagger = \left\{
\begin{array}{ll}
q_{k^\perp} - 1, & \text{with probability } d(q_{k^\perp}), \\
q_{k^\perp} + 1, & \text{with probability } b(q_{k^\perp}),
\end{array}
\right.
\]

(4.37)

where \( b(q_k) = 1 - d(q_k) \), for \( k = 1, \ldots, g \), \( d(0) = 0 \) and \( b(q_{\max}) = 0 \).
Let $\rho^\theta_{k^t}$ be the proposal vector of the partial autocorrelations. When $q_{k^t}^\dagger = q_{k^t} - 1$, the autoregressive coefficients $\rho^\theta_{k^t} = (\rho^\theta_{k^t1}, \ldots, \rho^\theta_{k^t,q_{k^t}})$ is updated by simply discarding the last partial autocorrelation in $\rho^\theta_{k^t} = (\rho^\theta_{k^t1}, \ldots, \rho^\theta_{k^t,q_{k^t}})$. On the other hand, when $q_{k^t}^\dagger = q_{k^t} + 1$, the corresponding autocorrelations $\rho^\theta_{k^t} = (\rho^\theta_{k^t1}, \ldots, \rho^\theta_{k^t,q_{k^t}}, \rho^\theta_{k^t1,q_{k^t}})$ is updated by adding new parameter $\rho^\theta_{k^t1,q_{k^t}}$ in $\rho^\theta_{k^t} = (\rho^\theta_{k^t1}, \ldots, \rho^\theta_{k^t,q_{k^t}})$ with $\rho^\theta_{k^t1,q_{k^t}} \sim B(-1, +1)$.

Then update the corresponding MA coefficients parameters by using the reparametrization discussed in Section 4.5.

To update the MA order of the model by using RJMCMC method, we have the acceptance probability

$$A_{q^*} = \min\{1, \Omega_{q^*}\},$$

and $\Omega_{q^*}$ can be obtained as in Section 3.7.1:

$$\Omega_{q^*} = \begin{cases} 
\exp \left[ -\frac{\tau}{2} \sum_{t:z_t = k} \left\{ (y_t - \tilde{\nu}_{k^t,t}^\dagger)^2 - (y_t - \hat{\nu}_{k^t,t})^2 \right\} \right] b(q_{k^t}^\dagger) & \text{if } q_{k^t}^\dagger = q_{k^t} - 1, \\
\exp \left[ -\frac{\tau}{2} \sum_{t:z_t = k} \left\{ (y_t - \tilde{\nu}_{k^t,t}^\dagger)^2 - (y_t - \hat{\nu}_{k^t,t})^2 \right\} \right] \frac{d(q_{k^t}^\dagger)}{b(q_{k^t}^\dagger)} & \text{if } q_{k^t}^\dagger = q_{k^t} + 1,
\end{cases}$$

(4.38)

where $\tilde{\nu}_{k^t,t}^\dagger = \mu_{k^t} + \sum_{j=1}^{p_{k^t}} \phi_{k^t,j} (y_{t-j} - \mu_{k^t}) - \sum_{j=1}^{q_{k^t}^\dagger} \theta_{k^t,j}^\dagger \tilde{e}_{k^t,t-j}$ and $\nu_{k^t,t}$ is as defined in Equation (4.4).

Similarly, the posterior density $\pi(q \mid g, y)$ is simply estimated by the proportions of every possible value for $q$ in the sample obtained by the previous complete MCMC algorithm.
Finally, we are at the move by which we can complete our Bayesian analysis. According to Bayes’ theorem, the marginal posterior distribution of $g$ is:

$$
\pi(g \mid y) \propto f(y \mid g)\pi(g),
$$

(4.39)

where $\pi(g)$ is the prior on $g$ and $f(y \mid g)$ is marginal likelihood given by

$$
f(y \mid g) = \sum_p \sum_q \int f(y \mid \varphi, p, q, g)\pi(\varphi, p, q \mid g) d\varphi.
$$

(4.40)

Now we are going to apply the method proposed by Chib and Jeliazkov (2001). First, applying the basic marginal likelihood identity and suppress the model index $g$ for notational convenience, we have

$$
f(y) = \frac{f(y \mid \varphi^*, p^*, q^*)\pi(\varphi^*, p^*, q^*)}{\pi(\varphi^*, p^*, q^* \mid y)}
$$

$$
= \frac{f(y \mid \varphi^*, p^*, q^*)\pi(\varphi^* \mid p^*, q^*)\pi(p^*)\pi(q^*)}{\pi(\varphi^* \mid p^*, q^*, y)\pi(p^* \mid y)\pi(q^* \mid y)}
$$

(4.41)

for any fixed point $(\varphi^*, p^*, q^*)$ which have the maximum density.

Note that the only quantity in (4.41) that is not already available is $\pi(\varphi^* \mid p^*, q^*, y)$. To produce the estimate $\hat{\pi}(\varphi^* \mid p^*, y)$ using the output from a MCMC simulation with fixed $p^*$ and $q^*$, we apply an efficient method proposed by Chib and Jeliazkov (2001).

First of all, $\hat{\pi}(\varphi^* \mid p^*, q^*, y)$ is factorized as

$$
\hat{\pi}(\varphi^* \mid p^*, q^*, y) = \hat{\pi}(\rho^{\phi^*} \mid p^*, y) \times \hat{\pi}(\rho^{\theta^*} \mid q^*, y) \times \hat{\pi}(\mu^* \mid \phi^*, \theta^*, p^*, q^*, y)
$$

$$
\times \hat{\pi}(\tau^* \mid \mu^*, \phi^*, \theta^*, p^*, q^*, y) \times \hat{\pi}(\omega^* \mid \tau^*, \mu^*, \phi^*, p^*, q^*, y).
$$

(4.42)

Suppose we have a sample $\{\varphi^{(i)}, z^{(i)}\}, (i = 1, \ldots, N)$, from the MCMC algorithm for a given $p^*$ and $q^*$ (i.e. a sample from $\pi(\varphi \mid p^*, q^*)$).

Then the terms of (4.42) are estimated by the following steps:
CHAPTER 4. MIXTURE OF GAUSSIAN ARMA COMPONENTS

Estimation of $\hat{\pi}(\rho^{\phi^*} \mid p^*, y)$

Let $\mathcal{U}_k^{\phi} = (p, \rho_k^\phi, \ldots, \rho_{k-1}^\phi)$ and $\mathcal{U}_{k+1}^{\phi} = (\rho_{k+1}^\phi, \ldots, \rho_{g}^\phi, q, \rho^\phi, \mu, \tau, \omega)$ be the two partitions of the parameters for any $k \in \{1, \ldots, g\}$.

Suppose $\{\mathcal{U}_{k+1}^{\phi}, z^{(i)}\}$, $(i = 1, \ldots, N)$, are the sample drawn from a reduced MCMC algorithm with distribution of interest $\pi(\mathcal{U}_{k+1}^{\phi}, z \mid \mathcal{U}_k^{\phi^*}, y)$.

We also draw $\hat{\rho}_k^{\phi^*}$, $(i = 1, \ldots, N)$, from the proposal density $q_p(\rho_k^{\phi^*}, \hat{\rho}_k^{\phi}) = \prod_{j=1}^{p_k} q(\rho_{kj}^{\phi^*}, \hat{\rho}_{kj}^{\phi})$, where $q(\cdot, \cdot)$ is defined by the Equation (4.22) in Section 4.6.3. We set

$$
\hat{\pi}(\rho_k^{\phi^*} \mid p^*, \rho_1^{\phi^*}, \ldots, \rho_{k-1}^{\phi^*}) = \frac{1}{N-B} \sum_{i=1}^{N} A(\rho_k^{\phi(i)}, \hat{\rho}_k^{\phi}) q_p(\rho_k^{\phi(i)}, \hat{\rho}_k^{\phi})
$$

where $A(\cdot, \cdot) = \min(1, \Omega)$ with $\Omega$ defined in Equation (4.27) and $B$ is the burn-in period.

Set $\mathcal{U}_{k+1}^{\phi(i)} = \tilde{\mathcal{U}}_{k+1}^{\phi(i)}$ and $z^{(i)} = \tilde{z}^{(i)}$, $(i = 1, \ldots, N)$. Repeat this step for $k = 1, \ldots, g$ and finally set

$$
\hat{\pi}(\rho^{\phi^*} \mid p^*, y) = \prod_{k=1}^{g} \hat{\pi}(\phi_k^* \mid p^*, \rho_1^{\phi^*}, \ldots, \rho_{k-1}^{\phi^*}). \tag{4.44}
$$

Estimation of $\pi(\rho^{\theta^*} \mid q^*, y)$

Let $\mathcal{U}_k^{\theta} = (p, \rho^\theta, q, \rho_1^\theta, \ldots, \rho_{k-1}^\theta)$ and $\mathcal{U}_{k+1}^{\theta} = (\rho_{k+1}^\theta, \ldots, \rho_{g}^\theta, \mu, \tau, \omega)$ be the two partitions of the parameters for any $k \in \{1, \ldots, g\}$.

Assume $\{\mathcal{U}_{k+1}^{\theta(i)}, z^{(i)}\}$, $(i = 1, \ldots, N)$, are the sample drawn from a reduced MCMC algorithm with distribution of interest $\pi(\mathcal{U}_{k+1}^{\theta^*}, z \mid \mathcal{U}_k^{\theta^*}, y)$.

We also draw $\hat{\rho}_k^{\theta^*}$, $(i = 1, \ldots, N)$, from the proposal density $q_p(\rho_k^{\theta^*}, \hat{\rho}_k^{\theta}) = \prod_{j=1}^{p_k} q(\rho_{kj}^{\theta^*}, \hat{\rho}_{kj}^{\theta})$, where $q(\cdot, \cdot)$ is defined by the Equation (4.28) in Section 4.6.4.

$$
\hat{\pi}(\theta^{\phi^*} \mid p^*, \rho_1^{\phi^*}, \ldots, \rho_{k-1}^{\phi^*}) = \frac{1}{N-B} \sum_{i=1}^{N} A(\hat{\theta}_k^{\phi(i)} \mid \rho_k^{\phi}, \hat{\theta}_k^{\phi}) q_p(\hat{\theta}_k^{\phi(i)} \mid \rho_k^{\phi}, \hat{\theta}_k^{\phi})
$$

where $A(\cdot, \cdot) = \min(1, \Omega)$ with $\Omega$ defined in Equation (4.27) and $B$ is the burn-in period.
We set
\[
\pi(\rho^* | \mathbf{p}, \mathbf{q}, \rho_1^*, \ldots, \rho_{k-1}^*) = \frac{1}{N-B} \sum_{i=B+1}^{N} \mathcal{A}(\rho^*_k, \rho_k^*) q_p(\rho^*_k, \rho_k^*) \frac{1}{N-B} \sum_{i=B+1}^{N} \mathcal{A}(\rho^*_k, \rho^{(i)}_k), \tag{4.45}
\]
where \(\mathcal{A}(*, *) = \min(1, \Omega)\) with \(\Omega\) defined in Equation (4.30).

Set \(\tilde{\theta}^{(i)}_{k+1} = \tilde{\theta}^{(i)}_{k+1}\) and \(z^{(i)} = \tilde{z}^{(i)}\), for \(i = 1, \ldots, N\). Repeat this step for \(k = 1, \ldots, g\) and finally set
\[
\pi(\mu^* | \phi^*, \theta^*, \mathbf{p}, \mathbf{q}, y) = \prod_{k=1}^{g} \pi(\mu^*_k | \phi^*, \theta^*, \mathbf{p}, \mathbf{q}, \rho_1^*, \ldots, \rho_{k-1}^*) \tag{4.46}
\]

**Estimation of \(\pi(\mu^* | \phi^*, \theta^*, \mathbf{p}, \mathbf{q}, y)\)**

Let the sample \((\tau^{(i)}, z^{(i)}), (i = 1, \ldots, N)\), are drawn from the last iteration of the previous step (i.e. marginally from \(\pi(\tau, \omega, z | \phi^*, \mathbf{p}, y)\)). The have the following estimation:
\[
\hat{\pi}(\mu^* | \phi^*, \theta^*, \mathbf{p}, \mathbf{q}, y) = \frac{1}{N-B} \sum_{i=B+1}^{N} \prod_{k=1}^{g} \pi(\mu^*_k | \phi^*, \theta^*, \tau^{(i)}, z^{(i)}), \mathbf{p}, \mathbf{q}, y) \tag{4.47}
\]
where \(\pi(\mu^*_k | \phi^*, \theta^*, \tau^{(i)}, z^{(i)}, \mathbf{p}, \mathbf{q}, y)\) is has the same distribution as (4.21) in Section 4.6.2.

**Estimation of \(\pi(\tau^* | \phi^*, \theta^*, \mu^*, \mathbf{p}, \mathbf{q}, y)\)**

Suppose we have the sample \(\{\tau^{(u)}, \omega^{(u)}, z^{(u)}\}, (u = 1, \ldots, N)\), from a reduced MCMC algorithm with distribution of interest \(\pi(\tau, \omega, z | \phi^*, \mu^*, \mathbf{p}, y)\). Then the estimator is:
\[
\hat{\pi}(\tau^* | \phi^*, \theta^*, \mu^*, \mathbf{p}, \mathbf{q}, y) = \frac{1}{N-B} \sum_{u=B+1}^{N} \prod_{k=1}^{g} \pi(\tau^*_k | \phi^*, \theta^*, \mu^*, z^{(u)}, \mathbf{p}, \mathbf{q}, y) \tag{4.48}
\]
where \( \pi \left( \tau_k^* \mid \phi^*, \theta^*, \mu^*, z^{(u)}, p^*, q^*, y \right) \) is given by the Equation (4.32).

**Estimation of \( \widehat{\pi}(\omega^* \mid \phi^*, \theta^*, \mu^*, \tau^*, p^*, q^*, y) \)**

Sample \( \{ \omega^{(v)}, z^{(v)} \}, (v = 1, \ldots, N) \), from a reduced MCMC algorithm with distribution of interest \( \pi(\omega, z \mid \phi^*, \theta^*, \mu^*, \tau^*, p^*, q^*, y) \) and set

\[
\widehat{\pi}(\omega^* \mid \phi^*, \theta^*, \mu^*, \tau^*, p^*, q^*, y) = \frac{1}{N - B} \sum_{v=B+1}^{N} \pi(\omega^* \mid z^{(v)}, p^*, q^*, y)
\]

where \( \pi(\omega^* \mid z^{(v)}, p^*, q^*, y) \) is given by the Equation (4.20).

### 4.8 Examples

We illustrate the MCMC techniques for the \( \mathcal{M}_{\text{ARMA}} \) model with the same simulated series and real data that has been considered for \( \mathcal{M}_{\text{AR}} \) as well in the previous chapter. For details of these series and data see Section 3.8.

We run three MCMC chains for the all series each with \( N = 100000 \) simulations and discarded initial 20000 as burn-in with MCMC moves discussed in Section 4.6 and in Section 4.7. Initial values was generated automatically while we run our algorithm based on the formulae discussed in Section 4.5.4.

#### 4.8.1 Experiment with simulated data sets

We consider both the simulated series \( y_{\text{MoI}} \) and \( y_{\text{MoII}} \) which are best fitted with \( \mathcal{M}_{\text{ARMA}}(2; 1, 1; 0, 0) \) and \( \mathcal{M}_{\text{ARMA}}(3; 2, 1, 3; 1, 0, 1) \) respectively.

Table 4.1 shows that the acceptance rate of Metropolis-Hastings moves of the AR coefficients. Since the rate is more that 20%, so the result is satisfactory for the related move.

Table 4.2 shows the parameter estimation results for the posterior sample for the series \( y_{\text{MoI}} \) fitted with the mixture model \( \mathcal{M}_{\text{ARMA}}(2; 1, 1; 0, 0) \). Our MCMC method captures the convergency of the trace plot of the posterior sample and asmetry and symmetry of the marginal posterior densities as well (not reported here, see Appendix).
Table 4.1: Acceptance Rate for Metropolis-Hastings chains of AR coefficients when fitting the $\mathcal{M}_{\text{ARMA}}(2; 1, 1; 0, 0)$ model to the simulated series $y_{\text{MoI}}$.

<table>
<thead>
<tr>
<th></th>
<th>N=50000</th>
<th>N=100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>chain1</td>
<td>$\phi_{11}$</td>
<td>$\phi_{21}$</td>
</tr>
<tr>
<td></td>
<td>0.3513470</td>
<td>0.3696874</td>
</tr>
<tr>
<td>chain2</td>
<td>0.3514670</td>
<td>0.3658073</td>
</tr>
<tr>
<td>chain3</td>
<td>0.3532471</td>
<td>0.3713874</td>
</tr>
</tbody>
</table>

Table 4.3 gives the Raftery and Lewis convergence diagnostic information about convergence to the stationary distribution and estimates the run-lengths needed to accurately estimate quantiles of functions of the parameters. Dependence factors ($I$) less than 5.0 indicates convergence passed in all cases (see Raftery and Lewis (1992)).

Convergence diagnostics and parameter estimation results of the above fitted model has been reported in Table 4.3. Table 4.4 shows the parameter estimation results for the posterior sample for the series $y_{\text{MoII}}$ fitted with the model $\mathcal{M}_{\text{ARMA}}(3; 2, 1, 3; 1, 0, 1)$. Columns 3-6 in Table 4.4 show the mean, standard deviation, median and Highest Posterior Density (HPD) interval of the posterior distribution of each parameter respectively.

Convergence diagrams and marginal posterior densities of the parameters are shown in Figure 4.2 and Figure 4.3. A decreasing weight constraint was imposed to solve the label switching problem.

These conclude that our Bayesian approach for the $\mathcal{M}_{\text{ARMA}}$ model also fits well with both the series $y_{\text{MoI}}$ and $y_{\text{MoII}}$.

### 4.8.2 Experiment with real datasets

We consider the data $y_{\text{IBM}}$ and $y_{\text{DGS3}}$ in this section. For details of these data see Section 3.8. Our Bayesian technique suggested $\mathcal{M}_{\text{ARMA}}(3; 1, 1, 0; 1, 1, 1)$ and $\mathcal{M}_{\text{ARMA}}(3; 0, 2, 2; 1, 1, 1)$ as their best fit models respectively.

Figure 4.4 shows the convergence diagrams and marginal densities of the posterior sample for each parameter when fitting $\mathcal{M}_{\text{ARMA}}(3; 1, 1, 0; 1, 1, 1)$ model to the data $y_{\text{IBM}}$. We obtained similar plots for the other two chains which are not shown here to save space. Observe that our Bayesian technique captures the symmetry and asymmetry of the marginal posterior densities of the parameter.
Table 4.2: Estimation results of MCMC samples of the parameters when fitting the $M_{\text{ARMA}}(2; 1, 1; 0, 0)$ model to the simulated series $y_{\text{MoI}}$, using $N = 10000$ and $B = 10000$.

-----------------------------------------------

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-series SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>0.5937</td>
<td>0.04357</td>
<td>7.955e-05</td>
<td>0.0002355</td>
</tr>
<tr>
<td>$w_2$</td>
<td>0.4063</td>
<td>0.04357</td>
<td>7.955e-05</td>
<td>0.0002355</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.1539</td>
<td>2.158</td>
<td>0.00394</td>
<td>0.004857</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-5.7805</td>
<td>16.483</td>
<td>0.03009</td>
<td>0.129362</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.08659</td>
<td>0.1494</td>
<td>0.0002728</td>
<td>0.0004791</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.11610</td>
<td>0.2080</td>
<td>0.0003798</td>
<td>0.0007581</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.3321</td>
<td>0.2450</td>
<td>0.0002728</td>
<td>0.004135</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.8651</td>
<td>0.3099</td>
<td>0.0005658</td>
<td>0.018379</td>
</tr>
</tbody>
</table>

-----------------------------------------------

2. Quantiles for each variable:

<table>
<thead>
<tr>
<th>Variable</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>0.5120</td>
<td>0.5628</td>
<td>0.5928</td>
<td>0.6233</td>
<td>0.6811</td>
</tr>
<tr>
<td>$w_2$</td>
<td>0.3189</td>
<td>0.3767</td>
<td>0.4072</td>
<td>0.4372</td>
<td>0.4880</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.3361</td>
<td>-0.1397</td>
<td>-0.06452</td>
<td>0.009336</td>
<td>0.1714</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-44.7153</td>
<td>-12.5200</td>
<td>-3.10853</td>
<td>0.322732</td>
<td>28.6716</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.3819</td>
<td>-0.1856</td>
<td>-0.08654</td>
<td>0.01234</td>
<td>0.2067</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.5702</td>
<td>-0.2414</td>
<td>-0.09029</td>
<td>0.01309</td>
<td>0.2645</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.57520</td>
<td>-0.4657</td>
<td>-0.4070</td>
<td>-0.3010</td>
<td>0.1793</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>-0.01665</td>
<td>0.9463</td>
<td>0.9807</td>
<td>0.9926</td>
<td>0.9993</td>
</tr>
</tbody>
</table>

-----------------------------------------------
Table 4.3: Convergence diagnostics of MCMC sampled values of $\omega$, $\mu$ and $\sigma$ when fitting the $\mathcal{M}_{\text{ARMA}}(3; 2, 1, 3; 1, 0, 1)$ model to the $y_{\text{MoII}}$ series, using $N = 50000$ and burn-in = 10000.

<table>
<thead>
<tr>
<th></th>
<th>Stationarity test</th>
<th>Start iteration</th>
<th>p-value</th>
<th>Halfwidth test</th>
<th>Mean</th>
<th>Halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>passed</td>
<td>1</td>
<td>0.140</td>
<td>passed</td>
<td>0.524</td>
<td>0.00413</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>passed</td>
<td>1</td>
<td>0.118</td>
<td>passed</td>
<td>0.281</td>
<td>0.00296</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>passed</td>
<td>1</td>
<td>0.329</td>
<td>passed</td>
<td>0.195</td>
<td>0.00210</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>passed</td>
<td>1</td>
<td>0.405</td>
<td>passed</td>
<td>-1.045</td>
<td>0.0396</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>passed</td>
<td>1</td>
<td>0.531</td>
<td>passed</td>
<td>-0.717</td>
<td>0.0513</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>passed</td>
<td>1</td>
<td>0.651</td>
<td>passed</td>
<td>-15.341</td>
<td>0.8097</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>passed</td>
<td>1</td>
<td>0.591</td>
<td>passed</td>
<td>3.91</td>
<td>0.0260</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>passed</td>
<td>1</td>
<td>0.922</td>
<td>passed</td>
<td>6.07</td>
<td>0.0341</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>passed</td>
<td>1</td>
<td>0.427</td>
<td>passed</td>
<td>8.79</td>
<td>8.79</td>
</tr>
</tbody>
</table>

The Geweke diagnostic:

<table>
<thead>
<tr>
<th></th>
<th>Z-score</th>
<th></th>
<th>Z-score</th>
<th></th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>-0.3093970</td>
<td>$\mu_1$</td>
<td>0.5098861</td>
<td>$\sigma_1$</td>
<td>0.2744771</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.8876903</td>
<td>$\mu_2$</td>
<td>0.8754738</td>
<td>$\sigma_2$</td>
<td>-0.1070980</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>-0.4410602</td>
<td>$\mu_3$</td>
<td>-0.4812748</td>
<td>$\sigma_3$</td>
<td>0.2868377</td>
</tr>
</tbody>
</table>
Table 4.4: Estimation results of parameters of the fitted model $\mathcal{M}_{\text{ARMA}}(3; 2, 1, 3; 1, 0, 1)$ to the simulated series $\mathbf{y}_{\text{MoII}}$ when $N = 50000$ and $B = 10000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.5622</td>
<td>0.06024</td>
<td>0.000269</td>
<td>0.5628</td>
<td>[0.44451422, 0.6806210]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.2915</td>
<td>0.04958</td>
<td>0.0002217</td>
<td>0.2901</td>
<td>[0.19590761, 0.3883456]</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.1463</td>
<td>0.03909</td>
<td>0.0001748</td>
<td>0.1434</td>
<td>[0.07382923, 0.2236892]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-1.1080</td>
<td>0.6718</td>
<td>0.003005</td>
<td>-1.0656</td>
<td>[-0.000790, 0.000749]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-0.7397</td>
<td>5.0766</td>
<td>0.022703</td>
<td>-0.4977</td>
<td>[-0.002532, 0.002026]</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-15.6930</td>
<td>20.3983</td>
<td>0.091224</td>
<td>-11.3587</td>
<td>[-42.707513, 1.5832250]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.7967</td>
<td>0.4632</td>
<td>0.002071</td>
<td>-0.7798</td>
<td>[-1.720022, 0.08437412]</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.7169</td>
<td>1.0312</td>
<td>0.004612</td>
<td>-0.6670</td>
<td>[-2.773888, 1.29088836]</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-4.5624</td>
<td>2.4020</td>
<td>0.010742</td>
<td>-4.5597</td>
<td>[-9.105997, 0.32139525]</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>4.075</td>
<td>0.683</td>
<td>0.003055</td>
<td>4.033</td>
<td>[2.788796, 5.410800]</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>6.177</td>
<td>1.899</td>
<td>0.008494</td>
<td>5.841</td>
<td>[3.127152, 9.982757]</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>10.162</td>
<td>3.433</td>
<td>0.015355</td>
<td>9.586</td>
<td>[5.426699, 15.635495]</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>0.87992</td>
<td>0.1420</td>
<td>0.0006349</td>
<td>0.8963</td>
<td>[0.6369107, 1.10219562]</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>-0.61282</td>
<td>0.1234</td>
<td>0.0005518</td>
<td>-0.6221</td>
<td>[-0.8398059, -0.37694165]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>-0.27596</td>
<td>0.2658</td>
<td>0.0011889</td>
<td>-0.3527</td>
<td>[-0.6281380, 0.33625167]</td>
</tr>
<tr>
<td>$\phi_{31}$</td>
<td>1.05476</td>
<td>0.4119</td>
<td>0.0018421</td>
<td>0.9653</td>
<td>[0.4279093, 1.91812074]</td>
</tr>
<tr>
<td>$\phi_{32}$</td>
<td>-0.46016</td>
<td>0.4701</td>
<td>0.0021025</td>
<td>-0.3629</td>
<td>[-1.5565401, 0.04001462]</td>
</tr>
<tr>
<td>$\phi_{33}$</td>
<td>0.05869</td>
<td>0.2706</td>
<td>0.0012102</td>
<td>0.04123</td>
<td>[-0.4210252, 0.68959747]</td>
</tr>
<tr>
<td>$\theta_{11}$</td>
<td>0.1687</td>
<td>0.2361</td>
<td>0.001056</td>
<td>0.1848</td>
<td>[-0.3039040, -0.3039040]</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.2322</td>
<td>0.2612</td>
<td>0.001168</td>
<td>0.1993</td>
<td>[-0.2036651, 0.7987750]</td>
</tr>
</tbody>
</table>
Table 4.5 shows the parameter estimation results for the posterior sample for the yIBM which includes the mean, standard deviation, median and Highest Posterior Density (HPD) interval of the posterior distribution of each parameter respectively. Note that these estimates are close to the true value of the parameter.

On the other hand, Figure 4.5 shows the convergence diagrams and marginal densities of the posterior sample for $\omega, \gamma$ and $\sigma$ for the series yDGS3. We obtained similar plots for the other parameters $\phi$ and $\theta$ but they are not shown here to save space. Observe that our Bayesian technique captures the symmetry (in most cases) and asymmetry of the posterior distributions of the parameter.
Figure 4.3: MCMC output of AR and MA coefficients when fitting the $M_{ARMA}(3; 2, 1, 3; 1, 0, 1)$ model to the $y_{MoII}$ series.

In Table 4.6 the parameter estimation results for the posterior sample are shown. It includes the mean, standard deviation, median and Highest Posterior Density (HPD) interval of the posterior distribution of each parameter respectively. These concludes that the Bayesian method for the $M_{ARMA}$ model fits better with the data $y_{DGS3}$. 
Table 4.5: Estimation results of parameters of the fitted model \( \mathcal{M}_{\text{ARMA}}(3; 1, 1, 0; 1, 1, 1) \) to the \( y_{\text{IBM}} \) data when \( N = 50000 \) and \( B = 10000 \).

<table>
<thead>
<tr>
<th>Posterior statistics</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_1 )</td>
<td>0.5648</td>
<td>0.08408</td>
<td>0.002658</td>
<td>0.5512</td>
<td>[0.43253139, 0.7481310]</td>
</tr>
<tr>
<td>( \omega_2 )</td>
<td>0.3218</td>
<td>0.08002</td>
<td>0.002529</td>
<td>0.3836</td>
<td>[0.16743826, 0.4559032]</td>
</tr>
<tr>
<td>( \omega_3 )</td>
<td>0.1134</td>
<td>0.04468</td>
<td>0.001412</td>
<td>0.1037</td>
<td>[0.05039192, 0.2034097]</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>0.2064</td>
<td>0.8291</td>
<td>0.003708</td>
<td>0.2762</td>
<td>[-1.540659, 1.767814]</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-0.6079</td>
<td>1.6570</td>
<td>0.007410</td>
<td>-0.6898</td>
<td>[-3.854221, 2.788555]</td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>-1.9251</td>
<td>3.1078</td>
<td>0.013899</td>
<td>-1.9562</td>
<td>[-7.978585, 4.516651]</td>
</tr>
<tr>
<td>( \sigma_1 )</td>
<td>5.390</td>
<td>0.5805</td>
<td>0.002596</td>
<td>5.357</td>
<td>[4.305660, 6.547276]</td>
</tr>
<tr>
<td>( \sigma_2 )</td>
<td>6.485</td>
<td>1.3311</td>
<td>0.005953</td>
<td>6.229</td>
<td>[4.370619, 9.234221]</td>
</tr>
<tr>
<td>( \sigma_3 )</td>
<td>11.134</td>
<td>2.5706</td>
<td>0.011496</td>
<td>10.943</td>
<td>[6.203935, 16.183060]</td>
</tr>
<tr>
<td>( \phi_{11} )</td>
<td>-0.02123</td>
<td>0.1897</td>
<td>0.0008484</td>
<td>0.00000</td>
<td>[-0.4709458, 0.3821788]</td>
</tr>
<tr>
<td>( \phi_{21} )</td>
<td>0.23072</td>
<td>0.3910</td>
<td>0.0017486</td>
<td>0.03084</td>
<td>[-0.3153974, 0.9997872]</td>
</tr>
<tr>
<td>( \theta_{11} )</td>
<td>-0.02667</td>
<td>0.2509</td>
<td>0.001122</td>
<td>-0.06623</td>
<td>[-0.4120708, 0.5197699]</td>
</tr>
<tr>
<td>( \theta_{21} )</td>
<td>0.12154</td>
<td>0.3651</td>
<td>0.001633</td>
<td>0.06693</td>
<td>[-0.4693954, 0.7916791]</td>
</tr>
<tr>
<td>( \theta_{31} )</td>
<td>-0.09567</td>
<td>0.4489</td>
<td>0.002008</td>
<td>-0.15599</td>
<td>[-0.8768510, 0.8342363]</td>
</tr>
</tbody>
</table>

\( \mathcal{M}_{\text{ARMA}} \) denotes a mixture of Gaussian ARMA components.
Table 4.6: Estimation results of parameters when fitting the $M_{ARMA}(3; 0, 2, 2; 1, 1, 1)$ model to the $y_{DGS3}$ data with $N = 30000$ and $B = 10000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.7421</td>
<td>0.02470</td>
<td>1.426e-04</td>
<td>0.7434</td>
<td>[0.69451300, 0.7893212]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.0021975</td>
<td>0.0038594</td>
<td>2.228e-05</td>
<td>0.1419</td>
<td>[0.11021720, 0.1785349]</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.1145</td>
<td>0.01356</td>
<td>7.831e-05</td>
<td>0.1139</td>
<td>[0.08846234, 0.1409121]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.0004820</td>
<td>0.0003328</td>
<td>1.921e-06</td>
<td>-0.0004812</td>
<td>[-0.001119151, 0.0001774533]</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.0021975</td>
<td>0.0038594</td>
<td>2.228e-05</td>
<td>0.0017921</td>
<td>[-0.004298617, 0.0107140492]</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.0007406</td>
<td>0.0041728</td>
<td>2.409e-05</td>
<td>-0.0004583</td>
<td>[-0.009827193, 0.0071739102]</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.01117</td>
<td>0.0004602</td>
<td>2.657e-06</td>
<td>0.01117</td>
<td>[0.01031515, 0.01203795]</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.02744</td>
<td>0.0035280</td>
<td>2.037e-05</td>
<td>0.02754</td>
<td>[0.02101529, 0.03409178]</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.03012</td>
<td>0.0034595</td>
<td>1.997e-05</td>
<td>0.03026</td>
<td>[0.02285155, 0.03642802]</td>
</tr>
</tbody>
</table>
Figure 4.4: MCMC output of parameter when fitting the $\mathcal{M}_{\text{ARMA}}(3; 2, 1, 3; 1, 1, 0; 1, 1, 1)$ model to the $y_{\text{IBM}}$ data, using the priors given in Section 4.5 and the MCMC moves discussed in Section 4.6.
CHAPTER 4. MIXTURE OF GAUSSIAN ARMA COMPONENTS

Figure 4.5: MCMC output of parameters while fitting the $\mathcal{M}_{\text{ARMA}(3; 0, 2, 2; 1, 1, 1)}$ model to the $y_{\text{DGS3}}$ data.
Chapter 5

Mixture of Gaussian ARMA-GARCH components

5.1 Introduction

The name of the model, “Mixture of Gaussian ARMA-GARCH components”, suggests that it consists of a mixture of finite ARMA components with generalized autoregressive conditional heteroscedasticity; that is, the conditional mean of the process variable follows a mixture ARMA process, whereas the conditional variance of the process variable follows a mixture GARCH process. We denote this class of model by the notation $\mathcal{M}_{\text{ARMA-GARCH}}$. It can be considered as a generalised form of the $\mathcal{M}_{\text{AR-ARCH}}$ model.

It has been observed by researchers that financial time series exhibits some common patterns like volatility clustering, heavy-tailed distributions, large kurtosis and extreme observations. There is a huge amount of theoretical and empirical research on analysis of financial time series now-a-days. The ARCH (autoregressive conditional heteroskedastic) by Engle (1982) and its generalization, the GARCH model by Bollerslev (1986), are considered as a benchmark for modeling the time-varying volatility in financial time series. However, normal GARCH model is not consistent with high kurtosis, heavy tails and extreme events. On the other hand, these features can be well-captured by the mixture GARCH model of Ausín and Galeano (2007), the $\mathcal{M}_{\text{AR-ARCH}}$ model of Wong and Li (2001). Recently, Tang et al. (2003) proposed the mixture ARMA-GARCH model which can be considered as an extension of the mixture of AR-GARCH model developed by Xu (1998). It has been shown that the $\mathcal{M}_{\text{ARMA-GARCH}}$ model yields better prediction results than both the $\mathcal{M}_{\text{AR}}$ and the $\mathcal{M}_{\text{AR-ARCH}}$ models by experimenting some simulations. But inference on these models has been carried out using GEM (generalized expectation maximization) algorithm. Therefore, we have taken the opportunity by analysing the $\mathcal{M}_{\text{ARMA-GARCH}}$ model from the full Bayesian perspective in this
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS

5.2 Background

Given a time series \( y = \{y_t\}_{t=1}^T \). Then the normal ARMA-GARCH model of order \((p, q; r, s)\) can be defined as

\[
\begin{align*}
  y_t \mid \mathcal{G}_{t-1} &\sim \mathcal{N} (y_t; \nu_t, \tau_t^{-1}) \\
  \nu_t &= \mu + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \\
  \tau_t^{-1} &= \alpha_0 + \sum_{j=1}^{r} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2
\end{align*}
\]

where \( \mathcal{G}_t \) denotes all information set up to time \( t \); \( \phi_j \) are the coefficients of the AR\((p)\) process, \( \theta_j \) are the coefficients of the MA\((q)\) process; \( \alpha_j \)s and \( \beta_j \)s are the coefficients of the GARCH\((r, s)\) process; \( \mu \in \mathbb{R} \) and \( \nu > 0 \) are shift coefficients of the model.

That is, the residual term \( \varepsilon_t \) assumed gaussian white noise with variance follows a GARCH model as defined above.

Let \( G(L) = \sum_{j=1}^{p} \phi_j L^j \), \( H(L) = \sum_{j=1}^{q} \theta_j L^j \), \( A(L) = \sum_{j=1}^{r} \alpha_j L^j \) and \( B(L) = \sum_{j=1}^{s} \beta_j L^j \), where \( L \) is the lag operator.

To ensure stationarity and invertibility, it is very common practice to impose the following two constraints in the ARMA model:

\((C_S)\): all roots of \( 1 - G(L) = 0 \) must lie outside the unit circle.

\((C_I)\): all roots of \( 1 + H(L) = 0 \) must lie outside the unit circle.

To ensure the conditional variance \( \sigma_t^2 \) is always positive, we often impose the following two constraints:

\((C_{\alpha})\): \( \alpha_i > 0 \) for \( i = 0, 1, \ldots, r \)

\((C_{\beta})\): \( \beta_i > 0 \) for \( i = 1, \ldots, s \)

And finally, for the finiteness of the unconditional variance of \( \varepsilon_t \), we need to impose the constraint.

\((C_{\varepsilon})\): \( A(1) + B(1) < 1 \) (i.e. \( \sum \alpha_j + \sum \beta_j < 1 \)).
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS

5.3 The $\mathcal{M}_{\text{ARMA-GARCH}}$ model

Let $y = \{y_t\}_{t=1}^T$ be an observed time series. Then the $g$-component $\mathcal{M}_{\text{ARMA-GARCH}}$ model (see Tang et al. (2003)) can be defined as below:

$$
\begin{align*}
    y_t \mid \mathcal{G}_{t-1} &\sim \sum_{k=1}^{g} \omega_k \mathcal{N} \left( y_t; \nu_{k,t}, \tau_{k,t}^{-1} \right), \quad (t = 1, \ldots, T), \\
    \nu_{k,t} &= \mu_k + \sum_{j=1}^{p_k} \phi_{kj} (y_{t-j} - \mu_k) + \sum_{j=1}^{q_k} \theta_{kj} \varepsilon_{t-j}, \quad (k = 1, \ldots, g), \\
    \tau_{k,t}^{-1} &= \alpha_{k0} + \sum_{j=1}^{r_k} \alpha_{kj} \varepsilon_{t-j}^2 + \sum_{j=1}^{s_k} \beta_{kj} \tau_{k,t-j}^{-1}, \quad (k = 1, \ldots, g), \\
    \omega_k > 0, \quad (k = 1, \ldots, g); \quad \sum_{k=1}^{g} \omega_k = 1,
\end{align*}
$$

(5.2)

where “$\mid \mathcal{G}_{t-1}$” is used to denote conditioning on the past observations and on all other variables; $\mathcal{N}(x; a, b^{-1})$ denotes the univariate normal distribution with mean $a$ and precision $b$ (for details, see Section 2.11); $\omega_k$ the $k$-th mixing weight of the mixture; $\tau_{k,t}^{-1}$ is the conditional precision of the $k$-th component; $\phi_{kj}$s are the AR coefficients of the $k$-th component; $\theta_{kj}$s are the MA coefficients of the $k$-th component; $\alpha_{kj}$s and $\beta_{kj}$s are the GARCH coefficients of the $k$-th component. We denote this model by the notation $\mathcal{M}_{\text{ARMA-GARCH}}(g; p_1, \ldots, p_g; q_1, \ldots, q_g; r_1, \ldots, r_g; s_1, \ldots, s_g)$.

The $\mathcal{M}_{\text{ARMA-GARCH}}$ model given by (5.2) can be treated as a generalization of the mixture of $\mathcal{M}_{\text{AR-ARCH}}$ process (see Wong and Li (2001)). Indeed, if we set all $q_k = 0$ and $s_k = 0$, for $k = 1, \ldots, g$ the model (5.2) reduces to a mixture of AR-ARCH model with Gaussian error term. In fact, it can also be considered as a generalized version of the $\mathcal{M}_{\text{AR}}$ and $\mathcal{M}_{\text{ARMA}}$ models as well.

For simplicity in notation, we can assume $p = \max\{p_1, \ldots, p_g\}$, $q = \max\{q_1, \ldots, q_g\}$, $r = \max\{r_1, \ldots, r_g\}$, and $s = \max\{s_1, \ldots, s_g\}$ with the settings $\phi_{kj} = 0$ if $j > p_k$, $\theta_{kj} = 0$ if $j > q_k$, $\alpha_{kj} = 0$ if $j > r_k$, and $\beta_{kj} = 0$ if $j > s_k$ for $k = 1, \ldots, g$.

Likewise all other mixture models discussed in this thesis, the conditional expectation of $y_t$ given the past information is the same. i.e.

$$
E(y_t \mid \mathcal{G}_{t-1}) = \sum_{k=1}^{g} \omega_k \nu_{k,t}.
$$

(5.3)
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS

For notational convenience of our model, we set \( \omega = (\omega_1, \ldots, \omega_g) \) denotes the mixing weight vector of the model; \( \mu = (\mu_1, \ldots, \mu_g) \), denotes the component’s mean vector of the model; \( \phi_k = (\phi_{k,1}, \ldots, \phi_{k,p_k}) \), \( k = 1, \ldots, g \), denotes the AR coefficients vector of \( k \)-th component; \( \theta_k = (\theta_{k,1}, \ldots, \theta_{k,q_k}) \), \( k = 1, \ldots, g \) denotes the MA coefficients vector of \( k \)-th component; \( \gamma = (\gamma_1, \ldots, \gamma_g) \), \( \alpha_k = (\alpha_{k,1}, \ldots, \alpha_{k,q_k}) \) and \( \beta_k = (\beta_{k,1}, \ldots, \beta_{k,q_k}) \) for \( k = 1, \ldots, g \), denotes the GARCH coefficients. Finally, we can set \( \varphi = (\omega, \mu, \phi, \theta, \alpha, \beta) \), be the parameter vector of model (5.2).

We also denote \( p = (p_1, \ldots, p_g) \), as the vector of orders of the AR part of the model, \( q = (q_1, \ldots, q_g) \), as the vector of orders of the MA part of the model, and \( r = (r_1, \ldots, r_g) \), \( s = (s_1, \ldots, s_g) \) as the vector of order of the GARCH part of the model. Therefore, we set \( \Lambda = (g, p, q, r, s) \) as the model index.

5.3.1 Hypotheses for the model

Let \( G_k(L) = \sum_{j=1}^{p_k} \phi_{kj} L^j \), \( H_k(L) = \sum_{j=1}^{q_k} \theta_{kj} L^j \), \( A_k(L) = \sum_{j=1}^{r_k} \alpha_{kj} L^j \) and \( B_k(L) = \sum_{j=1}^{s_k} \beta_{kj} L^j \), for \( k = 1, \ldots, g \), where \( L \) is the lag operator. We assume that the following hypotheses are true for our \( \mathcal{M}_{\text{ARMA-GARCH}} \) model (5.2):

\((C_S)\): All roots of each of the characteristic equations
\[ 1 - G_k(L) = 0, \quad (k = 1, \ldots, g) \]
must lie outside the unit circle.

\((C_T)\): All roots of each of the characteristic equations,
\[ 1 + H_k(L) = 0, \quad (k = 1, \ldots, g) \]
must lie outside the unit circle.

\((C_\alpha)\): \( \alpha_{ki} > 0, \quad (k = 1, \ldots, g, i = 0, 1, \ldots, r_k) \)

\((C_\beta)\): \( \beta_{ki} > 0, \quad (k = 1, \ldots, g, i = 1, \ldots, s_k) \)

\((C_{\varepsilon})\): \( \sum_{j}^{r_k} \alpha_{kj} + \sum_{j}^{s_k} \beta_{kj} < 1, \quad (k = 1, \ldots, g) \).

The reasons for implementing the above hypotheses are very common in the literature. Note that \((C_S)\) ensures the stationarity and \((C_T)\) assures invertibility
in the ARMA part of the model; \((C_\alpha)\) and \((C_\beta)\) ensure that the conditional variance \(\tau_{k,t}^{-1}\) is always positive which is required for the model; and finally, \((C_{\varepsilon})\) is for the finiteness of the unconditional variance of \(\varepsilon_t\).

### 5.4 Missing data formulation and the likelihood function

For the given (observed) time series \(y = \{y_t\}_T^{t=1}\) and the residual (unobserved) series \(\varepsilon = \{\varepsilon_t\}_T^{t=1}\), where \(\varepsilon_t \iid \mathcal{N}(0, \tau_{k,t}^{-1})\), the likelihood function of the \(\mathcal{M}_{\text{ARMA-GARCH}}\) takes a very complicated form. But it can be simplified by introducing the usual missing data formulation for mixture setups (see e.g. Diebolt and Robert (1994)), where a set of allocation variables \(z = \{z_t\}_T^{t=1}\) are defined such that

\[
z_t = \begin{cases} 
1, & \text{with probability } \omega_1, \\
\vdots & \\
k, & \text{with probability } \omega_k, \\
\vdots & \\
g, & \text{with probability } \omega_g,
\end{cases} \quad (t = 1, \ldots, T). \tag{5.4}
\]

Therefore, our observed and unobserved series \(y\) and \(\varepsilon\) are being completed with the help of a missing data set \(z\). So, conditional on \(z\), both the series \(y\) and \(\varepsilon\) are labelled with a specific component of the mixture. For example, if \(z_t = k\), then \(y_t\) and \(\varepsilon_t\) are assumed to arise from the \(k\)th component of the mixture. i.e.

\[
y_t \mid G_{t-1}, z_t \sim \begin{cases} 
\mathcal{N}(y_t; \nu_{1,t}, \tau_{1,t}^{-1}) & \text{if } z_t = 1, \\
\vdots & \\
\mathcal{N}(y_t; \nu_{k,t}, \tau_{k,t}^{-1}) & \text{if } z_t = k, \quad (t = 1, \ldots, T) \\
\vdots & \\
\mathcal{N}(y_t; \nu_{g,t}, \tau_{g,t}^{-1}) & \text{if } z_t = g.
\end{cases} \tag{5.5}
\]

We divide the parameters into two groups such as \(\varphi^\dagger = (\omega, \mu, \phi, \theta)\) and \(\varphi^{\dagger\dagger} = (\alpha, \beta)\) for convention. The update of the parameters in the first group \(\varphi^\dagger\) are based on the original \(\mathcal{M}_{\text{ARMA-GARCH}}\) model which is given by the equation (5.2). But we assume that the conditional variances \(\{\tau_{k,t}\}_{t=1}^{n_k}\) are fixed and known. So, we can update the corresponding posterior density \(\pi(\varphi^\dagger \mid y, \varepsilon, z)\) by using the full
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS

The conditional likelihood of the model (5.2),

$$\mathcal{L}(\phi^1 \mid y, z) = \prod_{k=1}^g \prod_{t: z_t = k} \mathcal{N}(y_t; \nu_{k,t}, \tau^{-1}_{k,t})$$

is given by

$$= \prod_{k=1}^g \prod_{t: z_t = k} \sqrt{\frac{\tau_{k,t}}{2\pi}} \exp \left[ -\frac{\tau_{k,t}}{2} (y_t - \hat{\nu}_{k,t})^2 \right]$$

(5.6)

where

$$\hat{\nu}_{k,t} = \mu_k + \sum_{j=1}^{p_k} \phi_{kj} (y_{t-j} - \mu_k) + \sum_{j=1}^{q_k} \theta_{kj} \hat{\varepsilon}_{t-j},$$

(5.7)

with

$$\hat{\varepsilon}_t = \begin{cases} 0, & (t \leq 0), \\ y_t - \hat{\nu}_{k,t}, & (t = 1, \ldots, T) \end{cases}.$$  

(5.8)

That is, the likelihood function separates into $g$ parts where each one concerning the data assigned to each of the $g$ mixture components. On the other hand, the update of the parameters in the second group $\phi^1$ are based on an approximated mixture of GARCH model. Now, we recall the $k$-th GARCH component model from our main model defined by (5.2):

$$\tau^{-1}_{k,t} = \alpha_{k0} + \sum_{i=1}^{r_k} \alpha_{ki} \varepsilon^2_{t-i} + \sum_{j=1}^{s_k} \beta_{kj} \tau^{-1}_{k,t-j},$$

(5.9)

As shown by Bollerslev (1986), an equivalent representatin of the GARCH($r_k, s_k$) process (5.9) is given by

$$\begin{cases} \varepsilon^2_t = \alpha_{k0} + \sum_{i=1}^{r_k} \alpha_{ki} \varepsilon^2_{t-i} + \sum_{j=1}^{s_k} \beta_{kj} \varepsilon^2_{t-j} - \sum_{j=1}^{s_k} \beta_{kj} \tilde{u}_{k,t-j} + \tilde{u}_{k,t}, \text{ and} \\ \tilde{u}_{k,t} = \varepsilon^2_t - \tau^{-1}_{k,t} = \left( \frac{\varepsilon^2_t}{\tau^{-1}_{k,t}} - 1 \right) \tau^{-1}_{k,t}, \quad \text{where } \eta_t \text{iid } \mathcal{N}(0, 1). \end{cases}$$

(5.10)

Clearly, by definition $\tilde{u}_{k,t} \sim (\chi^2(1) - 1) \tau^{-1}_{k,t}$, conditionally on $\mathcal{G}_{t-1}$ and $z_t$, and so the conditional mean is $E(\tilde{u}_{k,t} \mid \mathcal{G}_{t-1}) = 0$ and the conditional variance is $\text{Var}(\tilde{u}_{k,t} \mid \mathcal{G}_{t-1}) = 2\tau^{-2}_{k,t}$. Similarly to Nakatsuma (2000), we replace $\tilde{u}_{k,t}$ with $u_{k,t} \sim \mathcal{N}(0, 2\tau^{-2}_{k,t})$, a normal variable with the same mean and variance. Then the
equation (5.10) becomes:

$$
\varepsilon_t^2 = \alpha_{k0} + \sum_{i=1}^{l_k}(\alpha_{ki} + \beta_{ki})\varepsilon_{t-i}^2 + u_{k,t} - \sum_{j=1}^{s_k}\beta_{kj}u_{k,t-j}, \quad u_{k,t} \sim \mathcal{N}(0, 2\tau_{k,t}^{-2}), \quad (5.11)
$$

where $l_k = \max\{r_k, s_k\}$, $\alpha_{k,i} > 0$ for $i > r_k$, and $\beta_{k,j} > 0$ for $j > s_k$.

Hence, the GARCH($r_k, s_k$) process (5.9) is approximated by an ARMA process in $\varepsilon_t^2$ of order $(l_k, s_k)$.

The full likelihood function of the approximated GARCH model (5.11) is

$$
\mathcal{L}(\varphi^\dagger | \Lambda, \varphi, z, \varepsilon_t^2) = \prod_{k=1}^{g} \prod_{t:z_t=k} \mathcal{N}(\varepsilon_t^2; \nu_{k,t}^2, 2\tau_{k,t}^{-2}) \mathbf{1}_{\{\varepsilon_t^2 > 0\}} \quad (5.12)
$$

where

$$
\nu_{k,t}^2 = \alpha_{k0} + \sum_{j=1}^{l_k}(\alpha_{kj} + \beta_{kj})\varepsilon_{t-j}^2 - \sum_{j=1}^{s_k}\beta_{kj}\hat{u}_{t-j}, \quad (5.13)
$$

with

$$
\hat{u}_t = \begin{cases} 
0, & \text{if } t \leq 0, \\
\varepsilon_t^2 - \alpha_{k0} - \sum_{i=1}^{l_k}(\alpha_{ki} + \beta_{ki})\varepsilon_{t-i}^2 + \sum_{j=1}^{s_k}\beta_{kj}\hat{u}_{t-j}, & \text{if } t = 1, \ldots, T.
\end{cases} \quad (5.14)
$$

### 5.5 Priors setup

For fixed $g$, we will have a specific $\mathcal{M}_{\text{ARMA-GARCH}}$. The prior distribution for the $\mathcal{M}_{\text{ARMA}}$ part of the model given by (5.2) will be the same as we specified for the $\mathcal{M}_{\text{ARMA}}$ model given by (4.4). We state these priors again here for convenience (for details, see Section 4.5). In addition, we discuss the priors of other new parameters.

#### 5.5.1 Priors of $\omega$ and $\mu$

$$
\omega | g \sim \mathcal{D}(\omega; \delta_1, \ldots, \delta_g) \quad (5.15)
$$

$$
\mu_k \overset{\text{iid}}{\sim} \mathcal{N}(\mu_k; \zeta, \kappa^{-1}), \quad (k = 1, \ldots, g) \quad (5.16)
$$
where $\mathcal{D}(\cdot; \delta_1, \ldots, \delta_g)$ and $\mathcal{N}(\cdot; \zeta, \kappa^{-1})$ denote the Dirichlet distribution and the normal distribution respectively. For details of these distributions, see Section 2.11.

### 5.5.2 Prior for $\phi$ and $\theta$

\[
\phi_k \sim \text{Uniform on } \Xi_p \iff \rho_{ki}^{\phi} \sim \mathcal{B}(\xi_{-1}, +1)\left(\rho_{ki}^{\phi}; \left\lfloor \frac{i+1}{2} \right\rfloor, \left\lceil \frac{i}{2} \right\rceil + 1\right),
\]
\[
(k = 1, \ldots, g, i = 1, \ldots, q_k) \quad (5.17)
\]

\[
\theta_k \sim \text{Uniform on } \Xi_q \iff \rho_{ki}^{\theta} \sim \mathcal{B}(\xi_{-1}, +1)\left(\rho_{ki}^{\theta}; \left\lfloor \frac{i+1}{2} \right\rfloor, \left\lceil \frac{i}{2} \right\rceil + 1\right),
\]
\[
(k = 1, \ldots, g, i = 1, \ldots, q_k) \quad (5.18)
\]

where $\rho_{ki}^{\phi}$ be the partial autocorrelation coefficients at lag $i$ for the $k$th component; $\mathcal{B}(\xi_{-1}, +1)(x; a, b)$ represents a generalized beta distribution defined on $(-1, +1)$, where $[a]$ means the integer part of $a$; $\Xi_p \times \Xi_q$ denotes the stationarity and invertibility region.

The above prior settings suggest that there is one-to-one relation between the partial autocorrelations and the AR coefficients (and MA) and the transformation is done by the method proposed by Barndorff-Nielsen and Schou (1973).

### 5.5.3 Priors for $\alpha$ and $\beta$

We set priors of the ARCH and GARCH coefficients based on approximated ARMA($l_k, s_k$) model (5.11):

\[
\alpha_k \sim (\text{Uniform on } \Xi_r) 1_{(c_{\alpha})} \iff \rho_{ki}^{\alpha} \sim \mathcal{B}(\xi_{-1}, +1)\left(\rho_{ki}^{\alpha}; \left\lfloor \frac{i+1}{2} \right\rfloor, \left\lceil \frac{i}{2} \right\rceil + 1\right),
\]
\[
(k = 1, \ldots, g, i = 1, \ldots, r_k) \quad (5.19)
\]
\[ \beta_k \sim (\text{Uniform on } \Xi_s) 1(C_\beta) \iff \rho_{ki}^\beta \sim B(-1, +1) \left( \rho_{ki}^\beta; \left[ \frac{i + 1}{2} \right], \left[ \frac{i}{2} \right] + 1 \right), \]

\[(k = 1, \ldots, g, i = 1, \ldots, s_k) \] (5.20)

where \( \rho_{ki}^\beta \) be the partial autocorrelation coefficients at lag \( i \) for the \( k \)th component; \( B(-1, +1)(x; a, b) \) represents a generalized beta distribution defined on \((-1, +1)\), where \([a]\) means the integer part of \( a \); \( \Xi_r \times \Xi_s \) denotes the stationarity and invertibility region.

The above prior settings suggest that there is one-to-one relation between the partial autocorrelations and the ARCH coefficients (and GARCH) and the transformation is done by the method proposed by Barndorff-Nielsen and Schou (1973).

### 5.5.4 Priors for \( \Lambda \)

However, by setting the model index \( \Lambda = (p, q, r, s) \) as random quantities we will get a complete Bayesian structure for our model. As a consequences, we set the priors for these:

\[ g \sim U(g; 1, g_{\text{max}}) \] (5.21)

\[ p_k \mid g \sim U(p_k; 0, p_{\text{max}}), \quad (k = 1, \ldots, g) \] (5.22)

\[ q_k \mid g \sim U(q_k; 0, q_{\text{max}}), \quad (k = 1, \ldots, g) \] (5.23)

\[ r_k \mid g \sim U(r_k; 0, r_{\text{max}}), \quad (k = 1, \ldots, g) \] (5.24)

\[ s_k \mid g \sim U(s_k; 0, s_{\text{max}}), \quad (k = 1, \ldots, g) \] (5.25)

where \( U(x \mid a, b) \) denotes the discrete uniform distribution for \( x = a, a + 1, \ldots, b \) and \( g_{\text{max}}, p_{\text{max}}, q_{\text{max}}, r_{\text{max}}, s_{\text{max}} \) are fixed.

For details of the above distributions mentioned above, see Section 2.11.
5.5.5 Formulae for hyperparameters

In the prior settings, the parameters $\delta, \zeta, \kappa, c, a$ and $b$ are fixed and they are defined by the same formulae as discussed in Section 3.5.1:

\[
\delta_1 = \cdots = \delta_g = 1 \\
\mathcal{R}_y = \max(y) - \min(y) \\
\zeta = \min(y) + \frac{\mathcal{R}_y}{2}, \quad \kappa = \frac{1}{\mathcal{R}_y} \\
c = 2, \quad a = 0.2, \quad b = \frac{100}{c \mathcal{R}_y^2}.
\] (5.26)

5.6 MCMC moves for parameter estimation

First we consider $g$, the number of component of the model $\mathcal{M}_{\text{ARMA-GARCH}}$, is fixed. We apply MCMC component-wise techniques to calculate the full conditional posterior distrionbution $\pi(\varphi \mid y, \Lambda)$ in which each of the parameter will update separately. Moreover, we need to update the latent variable $z$ as well.

Now, we can look forward to update our model parameters individually with help of MCMC techniques. The moves are discussed below.

5.6.1 Move I: updating $z$

This is a Gibbs-type move. From (5.4) and (5.4), the full conditional posterior probability of the location variables $z_t = k$ (i.e. the observations $y_t$ has been generated by the $k$-th mixture component) is,

\[
\Pr(z_t = k \mid y_t, \varphi) = \frac{\omega_k \sqrt{\tau_{k,t}} \exp \left[ - \frac{\tau_{k,t}}{2} (y_t - \hat{\nu}_{k,t})^2 \right]}{\sum_{u=1}^{g} \omega_u \sqrt{\tau_{u,t}} \exp \left[ - \frac{\tau_{u,t}}{2} (y_t - \hat{\nu}_{u,t})^2 \right]}, \quad (k = 1, \ldots, g),
\] (5.27)

where $\hat{\nu}_{k,t}$ is defined as in the equation (5.7). For details of the above derivation see Section 4.6.6.
Fig. 5.1: Directed Acyclic Graph of the $\mathcal{M}_{\text{ARMA-GARCH}}$ model. Squared nodes refer to constant parameters, circled nodes refer to stochastic components and the bold squared node indicates observed data of the model. The arrows indicate the conditional independence structure of the model.
5.6.2 Move II: updating \( \omega \)

This is a Gibbs-type move and let \( \varphi - \omega \) denotes the remaining parameters except \( \omega \). Then the full conditional posterior distribution \( \pi(\omega \mid \varphi - \omega, y, z) \) can be obtained in similar fashion as in Section 3.6.2:

\[
\omega \mid \varphi - \omega, y, z \sim \mathcal{D}(\omega; \delta_1 + n_1, \ldots, \delta_g + n_g),
\]

(5.28)

where \( n_k = \sum_{t:z_t = k} 1 \).

5.6.3 Move III: updating \( \mu \)

This is a Gibbs-type move and the full conditional distribution \( \pi(\mu \mid \varphi^\dagger - \mu, y, z) \), where \( \varphi^\dagger - \mu \) denote the remaining parameter excepts \( \mu \), is can be obtained as below:

\[
\pi(\mu \mid \varphi^\dagger - \mu, y, z) \propto \mathcal{L}(\varphi^\dagger - \mu \mid y, z) \pi(\mu)
= \prod_{k=1}^{g} \prod_{t:z_t = k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, \tau_{k,t}^{-1}) \prod_{k=1}^{g} \mathcal{N}(\mu_k; \zeta, \kappa^{-1})
\]

Therefore, we have the full conditional posterior distribution of \( \mu_k \):

\[
\pi(\mu_k \mid \varphi^\dagger_{-\mu_k}, y, z) \propto \prod_{t:z_t = k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, \tau_{k,t}^{-1}) \cdot \mathcal{N}(\mu_k; \zeta, \kappa^{-1})
\]

\( \propto \exp \left[ -\frac{1}{2} \sum_{t:z_t = k} \tau_{k,t} \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i}(y_{t-i} - \mu_k) + \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \right\}^2 \right] \)

(****)

\( \times \exp \left[ -\frac{\kappa}{2} (\mu_k - \zeta)^2 \right] \).

Now the term (****) can be simplified in a similar fashion as in the case of \( M_{AR} \)
model (see Section 3.6.3):

\[
\sum_{t \mid z_t = k} \tau_{k,t} \left\{ y_t - \mu_k - \phi_{k,t} (y_{t-i} - \mu_k) + \theta_{k,j} \hat{\epsilon}_{t-j} \right\}^2
\]

\[
= \sum_{t \mid z_t = k} \tau_{k,t} \left\{ \left( y_t - \phi_{k,i} y_{t-i} - \theta_{k,j} \hat{\epsilon}_{t-j} \right) - \left( 1 - \phi_{k,i} \right) \mu_k \right\}^2
\]

\[
= \sum_{t \mid z_t = k} \tau_{k,t} \left( e_{k,t} - b_k \mu_k \right)^2 = \sum_{t \mid z_t = k} \tau_{k,t} \left( \tilde{e}_{k,t} + \bar{e}_k - b_k \mu_k \right)^2,
\]

where \( \bar{e}_k = \frac{1}{n_k} \sum_{t \mid z_t = k} e_{k,t} \) provided \( n_k = \sum_{t \mid z_t = k} 1 \)

\[
= \sum_{t \mid z_t = k} \tau_{k,t} \left\{ (e_{k,t} - \bar{e}_k)^2 + 2(e_{k,t} - \bar{e}_k)(\tilde{e}_k - b_k \mu_k) + (\tilde{e}_k - b_k \mu_k)^2 \right\}.
\]

Substituting the simplified form of (****) in the equation back and after few steps of simplification (see Section 3.6.3), we have the full conditional posterior distribution of \( \mu_k \) as follows:

\[
\pi(\mu_k \mid \varphi; \mu_k, y, z) \propto \exp \left[ - \frac{1}{2} \left( \bar{e}_k - b_k \mu_k \right)^2 \sum_{t \mid z_t = k} \tau_{k,t} \right] \times \exp \left[ - \frac{\kappa}{2} (\mu_k - \zeta)^2 \right] \propto \exp \left[ - \frac{1}{2} \left( b_k^2 d_k^1 + \kappa \right) \mu_k^2 + \left( \bar{e}_k b_k d_k^1 + \kappa \zeta + b_k c_k^1 \right) \mu_k \right]
\]
\[ \propto \exp \left[ -A\mu_k^2 + 2B\mu_k \right] = \exp \left[ -A\left\{ \mu_k^2 - 2\frac{B}{A}\mu_k \right\} \right] \]

\[ = \exp \left[ -A\left( \mu_k - \frac{B}{A} \right)^2 \right] = \exp \left[ -A\left( \mu_k - \frac{B}{A} \right)^2 \right] \]

\[ \propto \exp \left[ -\frac{b_k^2d_k^1 + \kappa}{2} \left( \mu_k - \frac{\bar{e}_k b_k d_k^1 + \kappa \zeta + b_k c_k^1}{b_k^2 d_k^1 + \kappa} \right)^2 \right]. \]

Hence, the full conditional posterior distribution of \( \mu \) is:

\[
\begin{align*}
\mu_k \mid \varphi_{-\mu_k}, y, z, \xi & \sim \mathcal{N}(\mu_k; \frac{\bar{e}_k b_k d_k^1 + \kappa \zeta + b_k c_k^1}{b_k^2 d_k^1 + \kappa}, \frac{1}{b_k^2 d_k^1 + \kappa}), \\
&(k = 1, \ldots, g), \quad (5.29)
\end{align*}
\]

where

- \( \bar{e}_k = \frac{1}{n_k} \sum_{t: z_t = k} e_{k,t} \), where
  
  (i) \( e_{k,t} = y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i} - \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \), and

  (ii) \( n_k = \sum_{t: z_t = k} 1 \);

- \( b_k = 1 - \sum_{i=1}^{p_k} \phi_{k,i} \); \( c_k^1 = \sum_{t: z_t = k} \tau_{k,t}(e_{k,t} - \bar{e}_k) \); \( d_k^1 = \sum_{t: z_t = k} \tau_{k,t} \).

where \( \tau_{k,t} \) is fixed and defined in the Equation (5.2).

### 5.6.4 Move IV-A: updating \( \phi \)

Here we update \( \rho_k^{\phi} \), \( k = 1, \ldots, g \) by a Metropolis-Hasting mechanism which is equivalent to update \( \phi \). The algorithm for this move are as follows:
A candidate \( \rho_{ki}^* \) is generated by a normal density truncated in \((-1, +1)\) and centred in the current state of the chain \( \rho_{ki}^\phi \) so that we can define the proposal density as below:

\[
q(\rho_{ki}^\phi, \rho_{ki}^*) = N(-1, +1) \left( \rho_{ki}^*; \rho_{ki}^\phi, \tau_{q^\phi}^{-1} \right), \quad (k = 1, \ldots, g; i = 1, \ldots, p_k),
\]

(5.30)

where the precision \( \tau_{q^\phi}^{-1} \) is chosen in order to obtain a satisfactory acceptance rate (20% or more, see Sampietro (2006)).

Using the distribution defined above by (5.30), we generate a vector \( \rho_k^* = (\rho_{k1}^*, \ldots, \rho_{kp_k}^*) \) for the partial autocorrelations which will be then used to derive the corresponding parameter vector \( \phi_k^* = (\phi_{k1}^*, \ldots, \phi_{kp_k}^*) \) through the transformation proposed by Barndorff-Nielsen and Schou (1973).

Now, the ratio in the acceptance probability for a M-H algorithm can be defined as

\[
A(\phi_k^*, \phi_k) = \min(1, \Omega),
\]

where

\[
\Omega = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}.
\]

In this case, the likelihood ratio is simply given by (see Equation (5.6)):

\[
\text{likelihood ratio} = \frac{\prod_{t:z_t=k} N(y_t; \hat{\nu}_{k,t}^*, \tau_{k,t}^{-1})}{\prod_{t:z_t=k} N(y_t; \hat{\nu}_{k,t}^*, \tau_{k,t}^{-1})}
\]

\[
= \exp \left[ -\frac{1}{2} \sum_{t:z_t=k} \tau_{k,t} \left\{ (y_t - \hat{\nu}_{k,t})^2 - (y_t - \hat{\nu}_{k,t})^2 \right\} \right], \quad (5.31)
\]

where \( \hat{\nu}_{k,t}^* = \mu_k + \sum_{j=1}^{p_k} \phi_{kj}^* (y_{t-j} - \mu_k) - \sum_{j=1}^{q_k} \theta_{kj} \hat{\varepsilon}_{t-j} \) and \( \hat{\varepsilon}_t \) is defined by the equation (5.8).

Since all the parameters but the \( \phi \) do not change, the prior ratio in terms of the partial autocorrelations \( \rho^\phi \) reduces to the ratio as below (see the Equation...
(5.17)):

$$\text{prior ratio} = \frac{\prod_{i=1}^{pk} B(\mu_{t}^i; [\frac{i}{2}, \frac{i+1}{2}])}{\prod_{i=1}^{pk} B(\mu_{t}; [\frac{i}{2}, \frac{i+1}{2}])}$$

$$= \prod_{i=2}^{pk} (1 - \rho_{ki}^\phi) \frac{1}{(1 - \rho_{ki}^\phi)}$$  \hspace{1cm} (5.32)

and, eventually from the equation (5.30) the proposal ratio is:

$$\text{proposal ratio} = \frac{\prod_{i=1}^{pk} \mathcal{N}(\mu_{t}; \rho_{ki}^\phi, \tau_{q\phi}^\phi)}{\prod_{i=1}^{pk} \mathcal{N}(\mu_{t+1}; \rho_{ki}^\phi, \tau_{q\phi}^\phi)}$$

$$= \prod_{i=1}^{pk} \frac{\Phi_{N}(\mu_{t+1}; \rho_{ki}^\phi, \tau_{q\phi}^\phi) - \Phi_{N}(\mu_{t}; \rho_{ki}^\phi, \tau_{q\phi}^\phi)}{\Phi_{N}(\mu_{t+1}; \rho_{ki}^\phi, \tau_{q\phi}^\phi) - \Phi_{N}(\mu_{t}; \rho_{ki}^\phi, \tau_{q\phi}^\phi)}$$  \hspace{1cm} (5.33)

where $\Phi_{N}$ is the normal cumulative distribution function.

By using (5.31), (5.32) and (5.33) we have

$$\Omega_{\phi} = \exp \left[ -\frac{1}{2} \sum_{t; zt=k} \tau_{k,t} \left\{ (y_t - \hat{v}_{k,t})^2 - (y_t - \hat{v}_{k,t})^2 \right\} \right]$$

$$\times \prod_{i=2}^{pk} \frac{(\rho_{ki}^\phi + 1)[\frac{i+1}{2} - 1](1 - \rho_{ki}^\phi)[\frac{i}{2}]}{(\rho_{ki}^\phi + 1)[\frac{i+1}{2} - 1](1 - \rho_{ki}^\phi)[\frac{i}{2}]}$$

$$\times \prod_{i=1}^{pk} \frac{\Phi_{N}(\mu_{t+1}; \rho_{ki}^\phi, \tau_{q\phi}^\phi) - \Phi_{N}(\mu_{t}; \rho_{ki}^\phi, \tau_{q\phi}^\phi)}{\Phi_{N}(\mu_{t+1}; \rho_{ki}^\phi, \tau_{q\phi}^\phi) - \Phi_{N}(\mu_{t}; \rho_{ki}^\phi, \tau_{q\phi}^\phi)}.$$  \hspace{1cm} (5.34)
5.6.5 Move IV-B: updating $\theta$

Here we update $\rho_k^\theta$, $k = 1, \ldots, g$ by a Metropolis-Hasting mechanism in a similar way as described in the previous section. Therefore, we have a very brief discussion in this case as follows.

A candidate $\rho_{k,i}^*$ is generated by a normal density truncated in $(-1, +1)$ and centred in the current state of the chain $\rho_k^\theta$ so that we can define the proposal density as below:

$$q_\theta(\rho_{k,i}^\theta, \rho_{k,i}^*) = \mathcal{N}(-1,+1) \left( \rho_{k,i}^\theta, \rho_{k,i}^*, \tau_{q_k}^{-1} \right), \quad (k = 1, \ldots, g, i = 1, \ldots, q_k),$$

(5.35)

Using the distribution defined above by (5.43), we generate a vector $\rho_k^* = (\rho_{k,1}^*, \ldots, \rho_{k,p_k}^*)$ for the partial autocorrelations which will be then used to derive the corresponding parameter vector $\theta_k^* = (\theta_{k,1}^*, \ldots, \theta_{k,p_k}^*)$ through the transformation as discussed in Barndorff-Nielsen and Schou (1973).

Therefore, the ratio in the acceptance probability for a M-H algorithm can be defined as $A_\theta \min(1, \Omega_\theta)$, where

$$\Omega_\theta = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}$$

$$= \exp \left[ -\frac{1}{2} \sum_{t:z_t=k} \tau_{k,t} \left\{ \left( y_t - \hat{\nu}_{k,t}^* \right)^2 - \left( y_t - \nu_{k,t} \right)^2 \right\} \right]$$

$$\times \prod_{i=2}^{p_k} \left( \rho_{k,i}^* + 1 \right)^{[\frac{i+1}{2}]-1} \left( 1 - \rho_{k,i}^* \right)^{[\frac{i}{2}]}$$

$$\times \prod_{i=1}^{q_k} \Phi_N \left( +1 \mid \rho_{k,i}^* \tau_{q_k}^{-1} \right) - \Phi_N \left( -1 \mid \rho_{k,i}^* \tau_{q_k}^{-1} \right)$$

$$\times \prod_{i=1}^{q_k} \Phi_N \left( +1 \mid \rho_{k,i}^* \tau_{q_k}^{-1} \right) - \Phi_N \left( -1 \mid \rho_{k,i}^* \tau_{q_k}^{-1} \right),$$

(5.36)

where $\hat{\nu}_{k,t} = \mu_k + \sum_{j=1}^{p_k} \phi_k (y_{t-j} - \mu_k) + \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j}$, $\hat{\epsilon}_t$ is defined by the Equation (5.8) and $\Phi_N$ denotes the cumulative distribution function.
5.6.6 Move V-A: updating $\alpha$

Here we update $\rho_k^\alpha, \ k = 1, \ldots, g$ by a Metropolis-Hasting mechanism in a similar way. First, a candidate $\rho_k^{\alpha*}$ is generated by a normal density truncated in $(-1, +1)$ and centred in the current state of the chain $\rho_k^\alpha$ so that we can define the proposal density as below:

$$q_\alpha(\rho_k^\alpha, \rho_k^{\alpha*}) = \mathcal{N}_{(-1,+1)} \left( \rho_k^\alpha; \rho_k^{\alpha*}, \tau_k^{-2} \right), \ (k = 1, \ldots, g, i = 1, \ldots, r_k).$$ 

(5.37)

Using the distribution defined above by (5.37), we generate a vector $\rho_k^{\alpha*} = (\rho_k^{\alpha*1}, \ldots, \rho_k^{\alpha* p_k})$ for the partial autocorrelations which will be then used to derive the corresponding parameter vector $\alpha_k^* = (\alpha_k^{*1}, \ldots, \alpha_k^{* p_k})$ through the transformation as proposed in Barndorff-Nielsen and Schou (1973).

So, the ratio in the acceptance probability for a M-H algorithm can be defined as

$$A_\alpha(\alpha_k^*, \alpha_k) = \min(1, \Omega_{\alpha}),$$

(5.38)

where

$$\Omega_{\alpha} = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}.$$ 

In this case, the likelihood ratio depends on the likelihood (5.12) and is given by:

$$\text{likelihood ratio} = \frac{\prod_{t:z_t = k} \mathcal{N} \left( \varepsilon_t^2; \hat{\nu}_{k,t}^\varepsilon, 2\tau_{k,t}^{-2} \right)}{\prod_{t:z_t = k} \mathcal{N} \left( \varepsilon_t^2; \nu_{k,t}^\varepsilon, 2\tau_{k,t}^{-2} \right)}$$

$$= \exp \left[ -\frac{1}{4} \sum_{t:z_t = k} \tau_{k,t}^2 \left( \varepsilon_t^2 - \nu_{k,t}^{\varepsilon*} \right)^2 - \left( \varepsilon_t^2 - \nu_{k,t}^{\varepsilon*} \right)^2 \right],$$

(5.39)

where

$$\nu_{k,t}^{\varepsilon*} = \alpha_k^{*0} + \sum_{j=1}^{l_k} \left( \alpha_k^{*j} + \beta_k^{*j} \right) \varepsilon_{t-j}^2 - \sum_{j=1}^{s_k} \beta_k^{*j} \hat{u}_{t-j}^*, \ \text{where} \ l_k = \max \{ r_k, s_k \}$$

and $\hat{u}_{t}^*$ is defined as in the Equation (5.57).
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS

Since all the parameters but the $\alpha$ do not change, the prior ratio in terms of the partial autocorrelations $\rho^\alpha$ reduces to the ratio as below:

$$
\text{prior ratio} = \frac{\prod_{i=1}^{p_k} \mathcal{B}_{[-1,+1]}(\rho^\alpha_{k,i}; \left[\frac{i+1}{2}\right], \left[\frac{i}{2}\right] + 1)}{\prod_{i=1}^{p_k} \mathcal{B}_{[-1,+1]}(\rho^\alpha_{k,i}; \left[\frac{i+1}{2}\right], \left[\frac{i}{2}\right] + 1)}
$$

$$
= \prod_{i=2}^{p_k} (\rho^\alpha_{k,i} + 1)^{\left[\frac{i+1}{2}\right]}(1 - \rho^\alpha_{k,i})^{\left[\frac{i}{2}\right]}
$$

and, finally from (5.37) we have:

$$
\text{proposal ratio} = \frac{\prod_{i=1}^{p_k} \mathcal{N}_{(-1,+1)}(\rho^\alpha_{k,i}; \rho^\ast_{k,i}, \tau_{q^\ast})}{\prod_{i=1}^{p_k} \mathcal{N}_{(-1,+1)}(\rho^\alpha_{k,i}; \rho^\ast_{k,i}, \tau_{q^\ast})}
$$

$$
= \prod_{i=1}^{p_k} \frac{\Phi_N(\rho^\ast_{k,i}; \tau_{q^\ast}) - \Phi_N(-1; \rho^\ast_{k,i}, \tau_{q^\ast})}{\Phi_N(-1; \rho^\ast_{k,i}, \tau_{q^\ast}) - \Phi_N(-1; \rho^\ast_{k,i}, \tau_{q^\ast})}
$$

By using (5.39), (5.40) and (5.41) we have

$$
\Omega \alpha = \exp \left[ -\frac{1}{4} \sum_{\ell : z_{\ell} = k} \tau_{k,\ell}^2 \left( \varepsilon^2_{\ell} - \nu^*_{k,\ell} \right)^2 - \left( \varepsilon^2_{\ell} - \nu^*_{k,\ell} \right)^2 \right]
$$

$$
\times \prod_{i=2}^{p_k} (\rho^\alpha_{k,i} + 1)^{\left[\frac{i+1}{2}\right]}(1 - \rho^\alpha_{k,i})^{\left[\frac{i}{2}\right]}
$$

$$
\times \prod_{i=1}^{p_k} \frac{\Phi_N(\rho^\ast_{k,i}; \tau_{q^\ast}) - \Phi_N(-1; \rho^\ast_{k,i}, \tau_{q^\ast})}{\Phi_N(-1; \rho^\ast_{k,i}, \tau_{q^\ast}) - \Phi_N(-1; \rho^\ast_{k,i}, \tau_{q^\ast})}
$$

(5.42)
5.6.7 Move V-B: updating $\beta$

Here we update $\rho^\beta_k$, $k = 1, \ldots, g$ by a Metropolis-Hasting mechanism in a similar way. A candidate $\rho^\beta_{k,i}^{\dagger}$ is generated by a normal density truncated in $(-1, +1)$ and centred in the current state of the chain $\rho^\beta_k$ so that we can define the proposal density as below:

$$q_{\beta}(\rho^\beta_{k,i}; \rho^\beta_{k,i}^{\dagger}) = \mathcal{N}(-1, +1) \left( \rho^\beta_{k,i}; \rho^\beta_{k,i}^{\dagger}, \tau^{-1} \right), \quad (i = 1, \ldots, s_k, k = 1, \ldots, g)$$

(5.43)

Using the distribution defined above by (5.43), we generate a vector $\rho^\beta_{k}^{\dagger} = (\rho^\beta_{k,1}^{\dagger}, \ldots, \rho^\beta_{k,p_k}^{\dagger})$ for the partial autocorrelations which will be then used to derive the corresponding parameter vector $\beta^\dagger_k = (\beta^\dagger_{k,1}, \ldots, \beta^\dagger_{k,p_k})$ through the transformation.

Therefore, the ratio in the acceptance probability for a M-H algorithm can be defined as

$$A_{\beta}(\beta^*_{k}, \beta_{k}) = \min(1, \Omega_{\beta}),$$

(5.44)

where

$$\Omega_{\beta} = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}$$

$$= \exp \left[ -\frac{1}{4} \sum_{t:i=k} \tau_{k,t}^2 \left\{ (\epsilon_t^2 - \nu^{\dagger}_{k,t})^2 - (\epsilon_t^2 - \nu_{k,t})^2 \right\} \right]$$

$$\times \prod_{i=2}^{p_k} \left( \rho^\beta_{k,i}^{\dagger} + 1 \right) \left[ \frac{\nu_{k,i}^{\dagger}}{\nu_{k,i}} \right]^{-1} \left( 1 - \rho^\beta_{k,i}^{\dagger} \right) \left[ \frac{\nu_{k,i}^{\dagger}}{\nu_{k,i}} \right]$$

$$\times \prod_{i=1}^{p_k} \frac{\Phi_N(+1; \rho^\beta_{k,i}, \tau_{q^\beta}^{-1}) - \Phi_N(-1; \rho^\beta_{k,i}, \tau_{q^\beta}^{-1})}{\Phi_N(+1; \rho^\beta_{k,i}^{\dagger}, \tau_{q^\beta}^{-1}) - \Phi_N(-1; \rho^\beta_{k,i}^{\dagger}, \tau_{q^\beta}^{-1})},$$

(5.45)

where

$$\nu^{\dagger}_{k,t} = \alpha_{k0} + \sum_{j=1}^{l_k} (\alpha_{kj} + \beta^\dagger_{kj}) \epsilon^2_{t-j} - \sum_{j=1}^{s_k} \beta^\dagger_{kj} \tilde{u}^2_{t-j}, \text{ where } l_k = \max\{r_k, s_k\}$$

(5.46)
and $\hat{u}_t^{\dagger}$ is defined as in the Equation (5.57).

## 5.7 MCMC moves for model determination

Model determination for the $\mathcal{M}^{t}_{AR}$ model is the same as that for the $\mathcal{M}_{AR}$ model. We recall all techniques here in short. For details, see Section 3.7. To get full Bayesian flavour we must get the posterior distribution of the model index $\Lambda = (g, p, v)$ which can be written and factorized as:

$$
\pi(\Lambda | y) = \pi(p | g, q, y) \times \pi(q | g, p, y) \times \pi(v | g, y) \times \pi(g | y) \quad (5.47)
$$

So, it is enough to update $p, q, v$ and and $g$ separately to update the model index $\Lambda$. We develop two additional MCMC moves to complete our Bayesian analysis as follows.

### 5.7.1 Move VI-A: updating of $p$

In a similar fashion, we can update the parameter $p$ as discussed for the $\mathcal{M}_{ARMA}$ model in Section 4.7.1.

At first we need to choose a component, say $k^{\dagger}$, randomly chosen in $\{1, \ldots, g\}$. Then we can write the proposal for the new AR order of the $k^{\dagger}$-th component, $p_{k^{\dagger}}$, as:

$$
p_{k^{\dagger}} = \begin{cases} 
p_{k}^{\dagger} - 1, & \text{with probability } d(p_{k}^{\dagger}), \\
p_{k}^{\dagger} + 1, & \text{with probability } b(p_{k}^{\dagger}),
\end{cases} \quad (5.48)
$$

where $b(p_{k}) = 1 - d(p_{k})$, for $k = 1, \ldots, g$, $d(0) = 0$ and $b(p_{\text{max}}) = 0$. Let $\rho_{k}^{\phi}$ be the proposal vector of the partial autocorrelations. When $p_{k}^{\dagger} = p_{k}^{\dagger} - 1$, the autoregressive coefficients $\rho_{k}^{\phi} = (\rho_{k1}^{\phi}, \ldots, \rho_{k,p_{k}^{\dagger}}^{\phi})$ is updated by simply discarding the last partial autocorrelation in $\rho_{k}^{\phi} = (\rho_{k1}^{\phi}, \ldots, \rho_{k,p_{k}^{\dagger}}^{\phi})$. On the other hand, when $p_{k}^{\dagger} = p_{k}^{\dagger} + 1$, the corresponding autocorrelations $\rho_{k}^{\phi} = (\rho_{k1}^{\phi}, \ldots, \rho_{k,p_{k}^{\dagger}}^{\phi}, \rho_{k,p_{k}^{\dagger}+1}^{\phi})$ is updated by adding new parameter $\rho_{k}^{\phi}_{k1,p_{k}^{\dagger}}$ in $\rho_{k}^{\phi} = (\rho_{k1}^{\phi}, \ldots, \rho_{k,p_{k}^{\dagger}}^{\phi})$ with

$$
\rho_{k1,p_{k}^{\dagger}}^{\phi} \overset{\text{ind}}{\sim} \mathcal{B}(-1, +1) \left( \rho_{k1,p_{k}^{\dagger}}^{\phi}; \left[\frac{p_{k}^{\dagger} + 1}{2}\right], \left[\frac{p_{k}^{\dagger}}{2}\right] + 1 \right).
$$
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS

Then update the corresponding AR coefficients parameters by using the reparametrization discussed in Section 4.5.

To update the AR order of the model by using RJMCMC method, we have the acceptance probability

\[ A_{p^\uparrow} = \min\{1, \Omega_{p^\uparrow}\} \]

and \( \Omega_{p^\uparrow} \) can be obtained as (for details see section 3.7.1):

\[
\Omega_{p^\uparrow} = \begin{cases} 
L^1 \frac{b(p^\uparrow_{k^\uparrow})}{d(p^\uparrow_{k^\uparrow})} & \text{if } p^\uparrow_{k^\uparrow} = p_{k^\uparrow} - 1, \\
L^1 \frac{d(p^\uparrow_{k^\uparrow})}{b(p^\uparrow_{k^\uparrow})} & \text{if } p^\uparrow_{k^\uparrow} = p_{k^\uparrow} + 1,
\end{cases}
\]

(5.49)

where

\[ L^1 = \exp \left[ -\frac{\tau_{k^\uparrow,t}}{2} \sum_{t:z_t=k} \left\{ (y_t - \hat{\nu}_{k^\uparrow,t}^{\uparrow})^2 - (y_t - \hat{\nu}_{k^\uparrow,t}^{\uparrow})^2 \right\} \right], \]

where \( \hat{\nu}_{k^\uparrow,t}^{\uparrow} = \mu_{k^\uparrow} + \sum_{j=1}^{q_{k^\uparrow}} \hat{\phi}_{k^\uparrow,j} (y_{t-j} - \mu_{k^\uparrow}) - \sum_{j=1}^{q_{k^\uparrow}} \theta_{k^\uparrow,j} \hat{\epsilon}_{k^\uparrow,t-j} \), and \( \hat{\nu}_{k^\uparrow,t}^{\uparrow} \) is as defined in Equation (5.7).

Note that the equation (5.49) has quite simple form and \( \pi(p \mid g, y) \) is simply estimated by the proportions of every possible value for \( p \) in the sample obtained by the previous complete MCMC algorithm.

5.7.2 Move VI-B: updating of q

This is almost the same as discussed in Section 4.7.1. For convention, we present that discussion very shortly here.

Likewise the previous move, at first we need to choose a component randomly, say \( k^\uparrow \), in \{1, \ldots, g\}. Then we can write the proposal for the new MA order of the \( k^\uparrow \)-th component, \( q^\uparrow_{k^\uparrow} \), as:

\[
q^\uparrow_{k^\uparrow} = \begin{cases} 
q_{k^\uparrow} - 1, & \text{with probability } d(q_{k^\uparrow}), \\
q_{k^\uparrow} + 1, & \text{with probability } b(q_{k^\uparrow}),
\end{cases}
\]

where \( b(q_k) = 1 - d(q_k) \), for \( k = 1, \ldots, g \), \( d(0) = 0 \) and \( b(q_{\text{max}}) = 0 \).
Let $\rho_{k^1}^\theta$ be the proposal vector of the partial autocorrelations. When $q^\dagger_{k^1} = q_{k^1} - 1$, the autoregressive coefficients $\rho_{k^1}^\theta = (\rho_{k^11}^\theta, \ldots, \rho_{k^1 q_{k^1}^\dagger}^\theta)$ is updated by simply discarding the last partial autocorrelation in $\rho_{k^1}^\theta = (\rho_{k^11}^\theta, \ldots, \rho_{k^1 q_{k^1}^\dagger}^\theta)$. On the other hand, when $q^\dagger_{k^1} = q_{k^1} + 1$, the corresponding autocorrelations $\rho_{k^1}^\theta = (\rho_{k^11}, \ldots, \rho_{k^1 q_{k^1}^\dagger}, \rho_{k^1 q_{k^1}^\dagger + 1}^\theta)$ is updated by adding new parameter $\rho_{k^1 q_{k^1}^\dagger}^\theta$ in $\rho_{k^1}^\theta = (\rho_{k^11}, \ldots, \rho_{k^1 q_{k^1}^\dagger})$ with

$$\rho_{k^1 q_{k^1}^\dagger}^\theta \sim B(-1, 1)(\rho_{k^1 q_{k^1}^\dagger}^\theta; \frac{q^\dagger_{k^1} + 1}{2}, \frac{q^\dagger_{k^1}}{2} + 1).$$

Then update the corresponding MA coefficients parameters by using the reparametrization discussed in Section 4.5.

To update the MA order of the model by using RJMCMC method, we have the acceptance probability

$$A_{q^\dagger} = \min\{1, \Omega_{q^\dagger}\},$$

and $\Omega_{q^\dagger}$ can be obtained as in Section 6.6.1:

$$\Omega_{q^\dagger} = \begin{cases} L^* \frac{d(q^\cdot)}{d(q^\cdot)} & \text{if } q^\cdot = q^\cdot - 1, \\
 L^* \frac{d(q^\cdot)}{d(q^\cdot)} & \text{if } q^\cdot = q^\cdot + 1, 
\end{cases}$$

(5.50)

where

$$L^* = \exp \left[ -\frac{\tau_{k^\cdot t}}{2} \sum_{t_{\cdot j} = k} \left\{ (y_t - \hat{\nu}^*_{k^\cdot t})^2 - (y_t - \nu_{k^\cdot t})^2 \right\} \right],$$

where $\hat{\nu}^*_{k^\cdot t} = \mu_{k^\cdot} \sum_{j=1}^{\nu_{k^\cdot}} \phi_{k^\cdot j} (y_{t-j} - \mu_{k^\cdot}) - \sum_{j=1}^{\nu_{k^\cdot}} \theta_{k^\cdot j} \hat{\epsilon}_{k^\cdot, t-j}$ and $\nu_{k^\cdot, t}$ is as defined in Equation (7.8).

Similarly, the posterior density $\pi(q \mid g, y)$ is simply estimated by the proportions of every possible value for $q$ in the sample obtained by the previous complete MCMC algorithm.

### 5.7.3 Move VII-A: updating of $r$

This is a similar move like the previous one but it is based on the different likelihood function defined by the Equation (5.12). First, select a component, say
\( k^* \), randomly chosen in \( \{ 1, \ldots, g \} \). Then we can write the proposal for the new ARCH order of the \( k^* \)-th component, \( r_{k^*}^* \), as:

\[
\begin{align*}
    r_{k^*}^* &= \begin{cases} 
        r_{k^*} - 1, & \text{with probability } d(r_{k^*}), \\
        r_{k^*} + 1, & \text{with probability } b(r_{k^*}),
    \end{cases} \\
    \text{provided that } b(r_{k^*}) &= 1 - d(r_{k^*}), \quad (k = 1, \ldots, g), \quad d(0) = 0 \text{ and } b(r_{\text{max}}) = 0.
\end{align*}
\]

Let \( \rho_k^\alpha \) be the proposal vector of the partial autocorrelations. When \( r_{k^*}^* = r_{k^*} \), the autoregressive coefficients \( \rho_k^\alpha = (\rho_k^\alpha, \rho_{k^*}^\alpha) \) is updated by simply discarding the last partial autocorrelation in \( \rho_k^\alpha \). On the other hand, when \( r_{k^*}^* = r_{k^*} + 1 \), the corresponding autocorrelations \( \rho_k^\alpha = (\rho_{k^*}^\alpha, \rho_{k^*}^\alpha, \rho_{k^*}^\alpha, \rho_{k^*}^\alpha) \) is updated by adding new parameter \( \rho_{k^*}^\alpha \) in \( \rho_k^\alpha \ = (\rho_k^\alpha, \rho_{k^*}^\alpha) \) with

\[
\rho_{k^*}^\alpha \overset{\text{ind}}{\sim} B(-1, +1) \left( \rho_{k^*}^\alpha, \left[ \frac{r_{k^*}^* + 1}{2} \right], \left[ \frac{r_{k^*}^*}{2} \right] + 1 \right).
\]

Then the corresponding ARCH coefficients \( \alpha_{k^*} \) is generated by using the reparametrization discussed in Section 4.5.

Now the acceptance probability is given by,

\[
\Omega_{r^+} = \min \{ 1, \Omega_{r^+} \},
\]

where

\[
\Omega_{r^+} = \text{likelihood ratio} \times \text{prio ratio} \times \text{proposal ratio} \times \text{jacobian}.
\]

From (5.12), we have:

\[
\text{likelihood ratio} = \frac{\prod_{t: z_t = k^*} \mathcal{N} \left( \varepsilon_t^2; \nu^\varepsilon_{k^*}, 2\tau_{k^*}^{-2} \right)}{\prod_{t: z_t = k^*} \mathcal{N} \left( \varepsilon_t^2; \nu^\varepsilon_{k^*}, 2\tau_{k^*}^{-2} \right)} = \exp \left[ -\frac{1}{4} \sum_{t: z_t = k^*} \tau_{k^*}^2 \left( \left( \varepsilon_t^2 - \nu^\varepsilon_{k^*} \right)^2 - \left( \varepsilon_t^2 - \nu^\varepsilon_{k^*} \right)^2 \right) \right].
\]
where
\[ \nu_{*,k,t}^* = \alpha_{k,0}^* + \sum_{j=1}^{s_{k,*}} (\alpha_{k,*j}^* + \beta_{k,*j}^*) \hat{\epsilon}_{t-j}^2 - \sum_{j=1}^{s_{k,*}} \beta_{k,*j}^* \hat{u}_{t-j}^*, \quad l_{k,*}^* = \max\{r_{k,*}, s_{k,*}\} \]

provided
\[ \hat{u}_t^* = \begin{cases} 0, & (t \leq 0), \\ \hat{\epsilon}_t^2 - \nu_{*,k,t}^*, & (t = 1, \ldots, T), \end{cases} \quad (5.52) \]

and \( \hat{\epsilon}_t \) is as defined in Equation (5.8).

The prior ratio is given by:
\[ \text{prior ratio} = \prod_{i=1}^{r_{k,*}} \frac{\pi(\alpha_{k,*i})}{\pi(\alpha_{k,*})} = \prod_{i=1}^{r_{k,*}} \frac{1}{\pi(\alpha_{k,*})} = \begin{cases} \frac{1}{\pi(\alpha_{k,*})} & \text{if } r_{k,*}^* = r_{k,*} - 1, \\ \pi(\alpha_{k,*}) & \text{if } r_{k,*}^* = r_{k,*} + 1, \end{cases} \]

because of the uniform prior on the ARCH order of the model and only the \( k^* \)-th component is updated.

To calculate the proposal ratio we have to take into account the probabilities \( b(\cdot) \) and \( d(\cdot) \) and the proposal density, which is chosen equal to the prior on \( \alpha \).

\[ \text{proposal ratio} = \begin{cases} \frac{b(r_{k,*}^*)}{d(r_{k,*})} \pi(\alpha_{k,*r_{k,*}^*}) & \text{if } r_{k,*}^* = r_{k,*} - 1, \\ \frac{d(r_{k,*}^*)}{b(r_{k,*})} \pi(\alpha_{k,*r_{k,*}^*}) & \text{if } r_{k,*}^* = r_{k,*} + 1. \end{cases} \]

Finally, the jacobian is equal to 1 because the invertible function \( g(\cdot) \) is considered as the identity function.

Therefore, multiplying likelihood, prior and proposal ratios, we derive ratio in the acceptance probability for both cases:
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS 164

\[ \Omega_{\tau^*} = \begin{cases} 
L^* \frac{b(r_{k^s*})}{d(r_{k^s*})} & \text{if } r_{k^s*} = r_{k^s} - 1, \\
L^* \frac{d(r_{k^s*})}{b(r_{k^s*})} & \text{if } r_{k^s*} = r_{k^s} + 1.
\end{cases} \] (5.53)

where

\[ L^* = \exp \left[ - \frac{1}{2} \sum_{t,2t=k^s} \left( \varepsilon_t^2 - \nu_{k^s*,t}^\rho \right)^2 \right] \]

along with \( \nu_{k^s*,t}^\rho = \alpha_{k^s*0} + \sum_{j=1}^{l^*_{k^s*}} (\alpha_{k^s*j} + \beta_{k^s*j}) \varepsilon_{t-j}^2 - \sum_{j=1}^{s_{k^s*}} \beta_{k^s*j} \hat{\varepsilon}_{t-j}, \quad l^*_{k^s*} = \max \{ r_{k^s*}, s_{k^s*} \} \)

and \( \nu_{k^s*,t}^\rho \) is defined as in Equation (5.56).

Therefore, \( \pi(\mathbf{r} \mid g, y) \) is simply estimated by the proportions of every possible value for \( \mathbf{r} \) in the sample obtained by the previous complete MCMC algorithm.

5.7.4 Move VII-B: updating of \( s \)

This is a similar move like the previous one as it is also based on the same likelihood function defined by the Equation (5.12). We give a very short description for this move here.

First, select a component, say \( k^s* \), randomly chosen in \( \{1, \ldots, g\} \). Then we can write the proposal for the new ARCH order of the \( k^s*\)-th component, \( s_{k^s*} \), as:

\[ s_{k^s*} = \begin{cases} 
-1, & \text{with probability } d(s_{k^s*}), \\
1, & \text{with probability } b(s_{k^s*}),
\end{cases} \] (5.54)

provided that \( b(s_k) = 1 - d(r_k), \quad (k = 1, \ldots, g), \quad d(0) = 0 \) and \( b(s_{\max}) = 0. \)

Let \( \rho_{k^s,*}^{\beta} \) be the proposal vector of the partial autocorrelations. When \( s_{k^s*} = r_{k^s*} - 1, \) the autoregressive coefficients \( \rho_{k^s,*}^{\beta} = (\rho_{k^s*1}^\beta, \ldots, \rho_{k^s*s_{k^s*}}^\beta) \) is updated by simply discarding the last partial autocorrelation in \( \rho_{k^s*}^{\beta} = (\rho_{k^s*1}^\beta, \ldots, \rho_{k^s*s_{k^s*}}^\beta). \)

On the other hand, when \( s_{k^s*} = s_{k^s*} + 1, \) the corresponding autocorrelations \( \rho_{k^s,*}^{\beta} = (\rho_{k^s*1}^\beta, \ldots, \rho_{k^s*s_{k^s*}}^\beta, \rho_{k^s,s_{k^s*}+1}^\beta) \) is updated by adding a new parameter \( \rho_{k^s,s_{k^s*}+1}^\beta \) in \( \rho_{k^s,*}^{\beta} = (\rho_{k^s*1}^\beta, \ldots, \rho_{k^s*s_{k^s*}}^\beta) \) with

\[ \rho_{k^s,s_{k^s*}+1}^\beta \sim \mathcal{B}(-1,+1) \left( \rho_{k^s,s_{k^s*}+1}^\beta; \left[ \frac{s_{k^s*} + 1}{2} \right], \left[ \frac{s_{k^s*}}{2} + 1 \right] \right). \]
Then the corresponding GARCH coefficients $\beta^*_{k\star}$ is obtained by transforming $\rho_{k\star}^\beta$ through the reparametrization technique discussed in Section 4.5 and the acceptance probability is given by,

$$A_{s\star} = \min\{1, \Omega_{s\star}\},$$

and

$$\Omega_{s\star} = \text{likelihood ratio} \times \text{prio ratio} \times \text{proposal ratio} \times \text{jacobian}$$

$$= \left\{ \begin{array}{ll}
L^* \frac{b(s_{k\star}^* \epsilon_{k\star}, t)}{d(s_{k\star}^* \epsilon_{k\star})} & \text{if } s_{k\star}^* = s_k - 1, \\
L^* \frac{d(s_{k\star}^* \epsilon_{k\star})}{b(s_{k\star}^* \epsilon_{k\star}, t)} & \text{if } s_{k\star}^* = s_k + 1,
\end{array} \right.$$

(5.55)

where

$$\exp \left[ -\frac{\tau_{k\star, t}^2}{4} \sum_{t: t = k\star} \left\{ \left( \epsilon_t^2 - \nu_{k\star, t}^\delta \right)^2 - \left( \epsilon_t^2 - \nu_{k\star, t}^\delta \right)^2 \right\} \right]$$

along with

$$\nu_{k\star, t}^\delta = \alpha_{k\star} + \sum_{j=1}^{l_{k\star}} (\alpha_{k\star j} + \beta_{k\star j}) \nu_{k\star j}^2 - \sum_{j=1}^{l_{k\star}} \beta_{k\star j} \hat{u}_{t-j}^* - l_{k\star} = \max\{r_{k\star}, s_{k\star}^*\}$$

(5.56)

provided

$$\hat{u}_t^* = \left\{ \begin{array}{ll}
0, & (t \leq 0), \\
\epsilon_t^2 - \nu_{k\star, t}^\delta, & (t = 1, \ldots, T),
\end{array} \right.$$ 

(5.57)

and $\hat{e}_t$ is as defined in Equation (5.8).

Therefore, $\pi(s \mid g, y)$ is simply estimated by the proportions of every possible value for $s$ in the sample obtained by the previous complete MCMC algorithm.

### 5.7.5 Move VIII: updating of $g$

Finally, we are at the move by which we can complete our Bayesian analysis. According to Bayes’ theorem, the marginal posterior distribution of $g$ is:

$$\pi(g \mid y) \propto f(y \mid g)\pi(g),$$

(5.58)
where \( \pi(g) \) is the prior on \( g \) and \( f(y \mid g) \) is marginal likelihood given by

\[
f(y \mid g) = \sum_p \sum_q \int f(y \mid \varphi, p, q, v) \pi(\varphi, p, q, v \mid g) d\varphi. \tag{5.59}\]

Now we are going to apply the method proposed by Chib and Jeliazkov (2001). First, applying the basic marginal likelihood identity and suppress the model index \( g \) for notational convenience, we have

\[
f(y) = \frac{f(y \mid \varphi^*, p^*, q^*, v^*) \pi(\varphi^*, p^*, q^*, v^*)}{\pi(\varphi^*, p^*, q^*, v^*)}
= \frac{f(y \mid \varphi^*, p^*, q^*, v^*) \pi(\varphi^* \mid p^*, q^*, v^*) \pi(p^*) \pi(q^*) \pi(v^*)}{\pi(\varphi^* \mid p^*, q^*, v^*) \pi(p^*) \pi(q^*) \pi(v^*)} \tag{5.60}\]

for any fixed point \((\varphi^*, p^*, q^*, v^*)\) which have the maximum density.

We apply the same methodology and the discussion would be the same as it was the case for \( \mathcal{M}_{\text{ARMA}} \) model. We just need consider another condition parameter \( v \) here and the estimation will be followed by the corresponding equations for this model. For details, see Section 4.7.3.

### 5.8 Examples

We will illustrate our MCMC techniques here with a new simulated series generated from the following \( \mathcal{M}_{\text{ARMA-GARCH}}(2; 1,1; 1,0; 1,1; 1,1) \) model:

\[
\begin{aligned}
y_t & \mid \mathcal{G}_{t-1} \sim 0.6 \mathcal{N}(y_t; \nu_{1,t}, \tau^{-1}_{1,t}) + 0.4 \mathcal{N}(y_t; \nu_{2,t}, \tau^{-1}_{2,t}), \\

\nu_{1,t} & = -0.4 y_{t-1} + 0.2 \varepsilon_{t-1}; \quad \nu_{2,t} = 0.1 y_{t-1} \\

\tau^{-1}_{1,t} & = 0.5 + 0.6 \varepsilon^2_{t-1} + 0.3 \tau^{-1}_{1,t-1}; \quad \tau^{-1}_{2,t} = 0.5 + 0.7 \varepsilon^2_{t-1} + 0.2 \tau^{-1}_{1,t-1}.
\end{aligned} \tag{III}
\]

We will call this simulated series by \( \text{yMoIII} \) hereafter.

We run three-MCMC chains with \( N = 100000 \) and discarded initial 50000 runs as burn-in. Note that the initial values \( \varphi^{(0)} \) are generated automatically based on the formulae discussed in Section 3.5.1.

We illustrate our MCMC method in Bayesian analysis of the \( \mathcal{M}_{\text{ARMA-GARCH}} \) fitted with one simulated series \( \text{yMoIII} \) and with real data \( \text{yDGS3} \) (for details, see Section 3.8).
We run the Bayesian algorithm discussed in Section 5.6 and in Section 5.7 for both series with different Simulation numbers. For yMoIII, we run three MCMC chains with \( N = 30000 \) while we run three MCMC chains with \( N = 20000 \) and \( N = 50000 \) for the yDGS data. We discarded initial 10000 simulations as burn-in.

Figure 5.3 displays the convergence diagrams and the marginal posterior densities for the sampled values of each of the parameter of the best fitted \( M_{\text{ARMA-GARCH}}(2; 1,1; 1,0; 1,1; 1,1) \) model for the yMoIII series. Since the GARCH parameters are restricted to be non-negative we got exponential distribution for these parameters. This also shows that our Bayesian procedure captures symmetry and asymmetry of the marginal posterior distributions very well. The trace plot of parameter also indicates strong convergency of the posterior sample values.

Table 5.1 represents parameter estimation results for each parameter of the model \( M_{\text{ARMA-GARCH}}(2; 1,1; 1,0; 1,1; 1,1) \) fitted with the series yMoIII. The table contains mean, median and HPD values of each parameter. All these output shows that our MCMC approach fit well with the series yMoIII.

On the other hand, our Bayesian approach captures the symmetry and asymmetry of the marginal posterior distributions of each parameter for the model \( M_{\text{ARMA-GARCH}}(3; 0,2,2; 0,2,2; 1,1,1; 1,1,1) \) fitted to the series yDGS3 (see Figure 5.4, Figure 5.5, Figure 5.6 and Figure 5.7). Figure 5.4 shows the comparison between the MCMC output of \( \omega \) and \( \mu \) when fitting \( M_{\text{ARMA-GARCH}} \) to the series yDGS3 with different \( N = 20000 \) and \( N = 50000 \) respectively. Clearly, there is significant difference in the outputs with different simulation numbers. The rest Figures shows the same comparison for the rest parameters.

Table 5.2 and Table 8.2 show parameter estimation results for the MCMC samples of each parameter for the yDGS3 series for \( n = 20000 \) and \( N = 50000 \) respectively. Observe that there is a very small difference in the estimation of mean, median and HPD of each parameter with the increment of \( N \).

Due to the existence of many parameters involved in these \( M_{\text{ARMA-GARCH}} \) models, the algorithm is quite bigger and so the computational time was very long in each cases (e.g. one computational time is recorded at bottom of the Table 8.2). But our Bayesian algorithm produces good results in all aspect.
Table 5.1: Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}(2; 1, 1; 1, 0; 1, 1; 1, 1)$ model to the $y_{\text{MoIII}}$ series with $N = 50000$ and $B = 20000$.

<table>
<thead>
<tr>
<th>Posterior statistics</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.5795</td>
<td>0.05634</td>
<td>0.00023</td>
<td>0.5708</td>
<td>[0.5000306, 0.6828810]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.4205</td>
<td>0.05634</td>
<td>0.00023</td>
<td>0.4292</td>
<td>[0.3171190, 0.4999694]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.0003072</td>
<td>0.1119</td>
<td>0.000289</td>
<td>-0.01349</td>
<td>[-0.1479089, 0.2078197]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.0873596</td>
<td>0.8067</td>
<td>0.002083</td>
<td>0.04388</td>
<td>[-0.3011251, 0.5360288]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.01689</td>
<td>0.09854</td>
<td>0.0004023</td>
<td>-0.02197</td>
<td>[-0.1840407, 0.1871625]</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.03465</td>
<td>0.14857</td>
<td>0.0006065</td>
<td>0.04948</td>
<td>[-0.2429879, 0.2960132]</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.36281</td>
<td>0.3039</td>
<td>0.00805</td>
<td>-0.4388</td>
<td>[-0.8856398, 0.2622847]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.10632</td>
<td>0.4023</td>
<td>0.01264</td>
<td>0.1296</td>
<td>[-0.8354060, 0.7064437]</td>
</tr>
<tr>
<td>$\theta_{11}$</td>
<td>0.05819</td>
<td>0.1991</td>
<td>0.001408</td>
<td>0.0000</td>
<td>[-0.1668725, 0.7198981]</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>[0.0000, 0.0000]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.5436</td>
<td>0.2763</td>
<td>0.0007133</td>
<td>0.5593</td>
<td>[7.875578e-02, 0.9997872]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.4630</td>
<td>0.2858</td>
<td>0.0007379</td>
<td>0.4452</td>
<td>[7.629032e-05, 0.9365904]</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.6035</td>
<td>0.2748</td>
<td>0.0007096</td>
<td>0.6434</td>
<td>[1.172085e-01, 0.9999995]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.5150</td>
<td>0.2876</td>
<td>0.0007426</td>
<td>0.5237</td>
<td>[3.993412e-02, 0.9868172]</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.3374</td>
<td>0.2616</td>
<td>0.001170</td>
<td>0.2804</td>
<td>[1.528451e-06, 0.8477229]</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.2472</td>
<td>0.2717</td>
<td>0.001215</td>
<td>0.1827</td>
<td>[9.540563e-06, 0.8651784]</td>
</tr>
</tbody>
</table>

Computational time: 8.389423 hours
Table 5.2: Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 1, 1, 1; 1, 1, 1)$ model to the yDGS3 data, using $N = 20000$ and $B = 10000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.3643</td>
<td>0.02077</td>
<td>8.478e-05</td>
<td>0.3599</td>
<td>[0.3347748, 0.4058357]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.3328</td>
<td>0.01286</td>
<td>5.250e-05</td>
<td>0.3331</td>
<td>[0.3064483, 0.3591364]</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.3029</td>
<td>0.01990</td>
<td>8.123e-05</td>
<td>0.3070</td>
<td>[0.2637127, 0.3322760]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.001364</td>
<td>0.1958</td>
<td>0.001385</td>
<td>-0.007373</td>
<td>[-0.3248866, 0.3089162]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-0.001715</td>
<td>0.1425</td>
<td>0.001008</td>
<td>-0.006797</td>
<td>[-0.3467075, 0.3222914]</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.002578</td>
<td>0.1534</td>
<td>0.001085</td>
<td>-0.005494</td>
<td>[-0.3664909, 0.3348755]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.22657</td>
<td>0.6048</td>
<td>0.004276</td>
<td>0.00000</td>
<td>[-0.9970538, 1.6190313]</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.10269</td>
<td>0.4212</td>
<td>0.002978</td>
<td>0.00000</td>
<td>[-0.9980060, 0.6812085]</td>
</tr>
<tr>
<td>$\phi_{31}$</td>
<td>0.26180</td>
<td>0.6365</td>
<td>0.004501</td>
<td>0.00000</td>
<td>[-0.9709549, 1.6653710]</td>
</tr>
<tr>
<td>$\phi_{32}$</td>
<td>1.05476</td>
<td>0.4119</td>
<td>0.0018421</td>
<td>0.9653</td>
<td>[-0.9965988, 0.6570900]</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>0.1903</td>
<td>0.5988</td>
<td>0.004234</td>
<td>0.00000</td>
<td>[-0.9510087, 1.5930183]</td>
</tr>
<tr>
<td>$\theta_{22}$</td>
<td>-0.1297</td>
<td>0.4226</td>
<td>0.002988</td>
<td>0.00000</td>
<td>[-0.9998950, 0.6016461]</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.1939</td>
<td>0.6083</td>
<td>0.004302</td>
<td>0.00000</td>
<td>[-0.9408175, 1.6017361]</td>
</tr>
<tr>
<td>$\theta_{32}$</td>
<td>-0.1253</td>
<td>0.4389</td>
<td>0.003104</td>
<td>0.00000</td>
<td>[-0.9970318, 0.6396732]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.7205</td>
<td>0.2516</td>
<td>0.001779</td>
<td>0.7978</td>
<td>[0.001662, 0.9999594]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.4076</td>
<td>0.2882</td>
<td>0.002038</td>
<td>0.3622</td>
<td>[1.271054e-03, 0.9257736]</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.7565</td>
<td>0.2351</td>
<td>0.001662</td>
<td>0.00000</td>
<td>[2.496914e-01*, 0.9999594]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.1297</td>
<td>0.4226</td>
<td>0.002988</td>
<td>0.00000</td>
<td>[1.271005e-05, 0.9144708]</td>
</tr>
<tr>
<td>$\alpha_{31}$</td>
<td>0.1939</td>
<td>0.6083</td>
<td>0.004302</td>
<td>0.00000</td>
<td>[2.902335e-01, 0.9999818]</td>
</tr>
<tr>
<td>$\alpha_{32}$</td>
<td>-0.1253</td>
<td>0.4389</td>
<td>0.003104</td>
<td>0.00000</td>
<td>[2.082577e-05, 0.9173782]</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.7205</td>
<td>0.2357</td>
<td>0.001667</td>
<td>0.8570</td>
<td>[0.2332162, 0.9999876]</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.7910</td>
<td>0.2237</td>
<td>0.001582</td>
<td>0.8735</td>
<td>[0.2763551, 0.9999253]</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.8160</td>
<td>0.2028</td>
<td>0.001434</td>
<td>0.8906</td>
<td>[0.3612366, 0.9999367]</td>
</tr>
</tbody>
</table>

* $e - \#$ should read as $10^{-\#}$
Table 5.3: Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 1, 1, 1; 1, 1, 1)$ model to the $y_{\text{DGS3}}$ data with $N = 50000$ and $B = 10000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.3632</td>
<td>0.01963</td>
<td>8.780e-05</td>
<td>0.3591</td>
<td>[0.3345154, 0.4008474]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.3332</td>
<td>0.01250</td>
<td>5.590e-05</td>
<td>0.3332</td>
<td>[0.3078511, 0.3594812]</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.3036</td>
<td>0.01937</td>
<td>8.663e-05</td>
<td>0.3077</td>
<td>[0.2655839, 0.3317639]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.003050</td>
<td>0.1359</td>
<td>0.0006793</td>
<td>-0.008156</td>
<td>[-0.3198857, 0.3317496]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-0.004146</td>
<td>0.1449</td>
<td>0.0007243</td>
<td>-0.007569</td>
<td>[-0.3423053, 0.3318999]</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.003904</td>
<td>0.1512</td>
<td>0.0007558</td>
<td>-0.006507</td>
<td>[-0.3726280, 0.3295131]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.22567</td>
<td>0.6233</td>
<td>0.003117</td>
<td>0.00000</td>
<td>[-0.8903918, 1.7296938]</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.08511</td>
<td>0.4336</td>
<td>0.002168</td>
<td>0.00000</td>
<td>[-0.9999920, 0.6816749]</td>
</tr>
<tr>
<td>$\phi_{31}$</td>
<td>0.22389</td>
<td>0.6421</td>
<td>0.003211</td>
<td>0.00000</td>
<td>[-0.9234100, 1.7661058]</td>
</tr>
<tr>
<td>$\phi_{32}$</td>
<td>-0.08601</td>
<td>0.4436</td>
<td>0.002218</td>
<td>0.9653</td>
<td>[-0.9996972, 0.6831918]</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>0.1722</td>
<td>0.5929</td>
<td>0.002964</td>
<td>0.00000</td>
<td>[-0.9624248, 1.5776749]</td>
</tr>
<tr>
<td>$\theta_{22}$</td>
<td>-0.1088</td>
<td>0.4139</td>
<td>0.002069</td>
<td>0.00000</td>
<td>[-0.9999058, 0.6103684]</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.1780</td>
<td>0.5976</td>
<td>0.002988</td>
<td>0.00000</td>
<td>[-0.9656674, 1.5635401]</td>
</tr>
<tr>
<td>$\theta_{32}$</td>
<td>-0.1113</td>
<td>0.4271</td>
<td>0.002135</td>
<td>0.00000</td>
<td>[-0.9999306, 0.6123990]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.7297</td>
<td>0.2513</td>
<td>0.001256</td>
<td>0.8106</td>
<td>[1.970079e-01, 0.9999998]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.4078</td>
<td>0.2888</td>
<td>0.001444</td>
<td>0.3652</td>
<td>[1.476110e-04, 0.9170495]</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.7489</td>
<td>0.2408</td>
<td>0.001204</td>
<td>0.8301</td>
<td>[2.330945e-01, 0.9999998]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.4195</td>
<td>0.2919</td>
<td>0.001460</td>
<td>0.3800</td>
<td>[1.472996e-04, 0.9246623]</td>
</tr>
<tr>
<td>$\alpha_{31}$</td>
<td>0.7713</td>
<td>0.2316</td>
<td>0.001158</td>
<td>0.3800</td>
<td>[2.574370e-01, 0.9999998]</td>
</tr>
<tr>
<td>$\alpha_{32}$</td>
<td>0.4246</td>
<td>0.2897</td>
<td>0.001449</td>
<td>0.3920</td>
<td>[1.090606e-05, 0.9216717]</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.7715</td>
<td>0.2349</td>
<td>0.001175</td>
<td>0.8574</td>
<td>[0.2428083, 0.9999989]</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.7918</td>
<td>0.2218</td>
<td>0.001109</td>
<td>0.8738</td>
<td>[0.2923498, 0.9999989]</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.8096</td>
<td>0.2106</td>
<td>0.001053</td>
<td>0.8880</td>
<td>[0.3318563, 0.9999889]</td>
</tr>
</tbody>
</table>

Computational Time: 1.66 days
Figure 5.2: Graphical analysis of the simulated series yMoIII.
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS

Figure 5.3: MCMC output of sampled values of the parameter when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}(2; 1, 1; 1, 0; 1, 1; 1, 1)$ model to the yMoIII series, using the methods discussed in Section 5.5 and in Section 5.6, imposing the identifiability constraint $\omega_1 < \omega_2$ to solve the label-switching problem.
Figure 5.4: Comparison between the MCMC output of sampled values of mixing weights and means when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 1, 1, 1; 1, 1, 1)$ model to the $y_{\text{DGS3}}$ data with different simulations numbers, in each case burn-in is 10000.
Figure 5.5: Comparison between the MCMC output of sampled values of AR coefficients when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 1, 1, 1; 1, 1, 1)$ model to the \texttt{yDGS3} data with different simulations number, in each case burn-in is 10000.
CHAPTER 5. MIXTURE OF GAUSSIAN ARMA-GARCH COMPONENTS

Figure 5.6: Comparison between the MCMC sampled values of MA coefficients when fitting the \( \mathcal{M}_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 1, 1, 1; 1, 1, 1) \) model to the \( y_{\text{DGS3}} \) data with different simulations number, in each case burn-in is 10000.
Figure 5.7: Comparison between the MCMC outputsampled values of GARCH coefficients when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 1, 1, 1; 1, 1, 1)$ model to the $\textbf{yDGS3}$ data with different simulations number, in each case burn-in is 10000.
Chapter 6

Mixture of Student-\(t\) AR components

6.1 Introduction

In this chapter, we extend the \(M_{AR}\) model by replacing the Gaussian error term with the Student-\(t\) distribution and we denote this class of models by \(M'_{AR}\). Very recently, Wong et al. (2009) introduced this class of mixture model and analysed the model by classical approach. Our treatment will be a complete Bayesian analysis in this case for both the parameter estimation and model determination. This could be a benchmark model from Bayesian point of view.

There are several strong points in support of choosing the Student-\(t\) distribution (see Wong et al. (2009)). First of all, it provides heavier tails than normal distribution and would be able to accommodate some aberrant returns occasionally observed in financial markets. Secondly, the precision of Student-\(t\) distribution depends on its degrees of freedom which will interfere the estimation of the component precision. Thirdly, the degrees of freedom is treated as parameter in the \(M'_{AR}\) model which allows to adjust the degree of thickness of the conditional distributions according to the actually observed data. Finally, as degrees of freedom in a Student-\(t\) distribution approaches infinity, the distribution approaches normal. Consequently, the \(M_{AR}\) model is a limiting case of the \(M'_{AR}\) model.

Since only difference between \(M_{AR}\) and \(M'_{AR}\) models is the error term, we have similarity in the posterior derivation of some of the parameters. Therefore, we avoid the full derivations in such cases.
6.2 The $\mathcal{M}^t_{AR}$ model

Given $\{y_t\}$ be a (observed) time series. Then, the $g$-components $\mathcal{M}^t_{AR}$ model can be defined in the following form (see different form shown in Wong et al. (2009)):

$$ y_t \mid \psi_{t-1} \sim \sum_{k=1}^{g} \omega_k S_t(y_t; \nu_{k,t}, \tau_k^{-1}, v_k), \quad (t = 1, \ldots, T), $$

$$ \nu_{k,t} = \mu_k + \sum_{j=1}^{p_k} \phi_{kj}(y_{t-j} - \mu_k), \quad (k = 1, \ldots, g), \quad (6.1) $$

$$ \omega_k > 0, \quad (k = 1, \ldots, g); \quad \sum_{k=1}^{g} \omega_k = 1, $$

where $S_t(x; a, b^{-1}, c)$ denotes the univariate standardised Student-$t$ density function with mean $a$, precision $b$ and degrees of freedom $c$ (for more details, see Section 2.11.7); $v_k$ denotes the degrees of freedom of the $k$th component and the rest notations are the same as described for the $\mathcal{M}_{AR}$ model in Section 3.3.

We denote this model by $\mathcal{M}^t_{AR}(g; p_1, \ldots, p_g; v_1, \ldots, v_g)$. For simplicity, we can assume $p = \max\{p_1, \ldots, p_g\}$ by setting $\phi_{kj} = 0$, if $j > p_k$ for $k = 1, \ldots, g$.

Likewise the $\mathcal{M}_{AR}$ model we set notations for all parameters as $\varphi = (\omega, \mu, \tau, \phi)$. We also denote $p = (p_1, \ldots, p_g)$ and $v = (v_1, \ldots, v_g)$ as the vector of AR order and the vector of degrees of freedom of the model respectively.

The conditional expectation and conditional variance of $y_t$ given the past information are the same as in the $\mathcal{M}_{AR}$ model (also see Wong (1998)):

$$ E(y_t \mid \mathcal{G}_{t-1}) = \sum_{k=1}^{g} \omega_k \nu_{k,t}, \quad (6.2) $$

and,

$$ \text{Var}(y_t \mid \mathcal{G}_{t-1}) = \sum_{k=1}^{g} \omega_k \tau_k^{-1} + \sum_{k=1}^{g} \omega_k \nu_{k,t}^2 - \left( \sum_{k=1}^{g} \omega_k \nu_{k,t} \right)^2. \quad (6.3) $$

6.2.1 Hypothesis for the $\mathcal{M}^t_{AR}$ model

By definitions, the hypothesis for $\mathcal{M}^t_{AR}$ models is exactly the same as that for $\mathcal{M}_{AR}$ models discussed in Section 3.3.2. For reader, we state the less restrictive necessary and sufficient condition for first and second order stationarity of the $\mathcal{M}^t_{AR}$ model which is as follows:
(\(C^*_{\text{S}}\)) : the largest eigenvalue of \(\sum_{k=1}^{g} \omega_k A_k \otimes A_k\) is smaller than 1,

where

\[
A_k = \begin{pmatrix}
\phi_{k1} & \phi_{k2} & \cdots & \phi_{kp-1} & \phi_{kp} \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 \\
\end{pmatrix}, \quad (k = 1, \ldots, g).
\]

Then the \(\mathcal{M}_A\) model is said to be Stable if the constraint \((C^*_{\text{S}})\) is satisfied. For more details see Theorem 3 in Boshnakov (2011).

### 6.3 Missing data formulation and the likelihood function

Clearly, the likelihood function of the \(\mathcal{M}_A\) model (6.1) takes more complicated form than that of the \(\mathcal{M}_A\) model (3.4). The so-called direct application (see Titterington et al. (1985)) of finite mixture models refers to the situation in which we assume that each of the observed variables \(y_t\)'s belongs to only one of the \(g\) regimes. From this point of view, \(S_t(y_t; \nu_k, \tau_k^{-1}, v_k)\) represents the probability density of \(y_t\) given that the observation comes from the \(k\)-th category where the probability that each observation comes from the \(k\)-th component is \(\omega_k\).

Since \(y = \{y_t\}_{t=1}^{T}\) is not observed directly we consider a latent variable \(z = \{z_t\}_{t=1}^{T}\) which is regarded as “component label” so that:

\[
z_t = \begin{cases}
1 & \text{with probability } \omega_1 \\
\vdots & \vdots \\
k & \text{with probability } \omega_k, \quad (t = 1, \ldots, T). \\
\vdots & \vdots \\
g & \text{with probability } \omega_g
\end{cases}
\]  

(6.4)

With this approach, our observed series \(\{y_t\}_{t=1}^{T}\) becomes completed with a missing data set \(\{z_t\}_{t=1}^{T}\), indicating the specific component of the mixture from which every observation is assumed to arise (see Diebolt and Robert (1994)).

In the light of the above characterization of the Student-\(t\) distribution, it is convenient to view the observed data augmented by the latent variable \(z_t\) as still being incomplete. Therefore, we introduce a new missing variable \(\xi = \{\xi_t\}_{t=1}^{T}\)
which corresponds to the mixing random variables in the Student-\( t \) distribution and control the magnitude of the variance in the normal density function (see McLachlan and Peel (2000)).

Conditionally on \( \xi_t \) which is conditional on \( z_t \), the observations are drawn from their respective individual subpopulation, i.e.

\[
y_t \mid \xi_t, z_t = k; \mathcal{G}_{t-1} \overset{\text{ind}}{\sim} \mathcal{N} \left( y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1} \right),
\]

where

\[
\xi_t \mid z_t = k \overset{\text{ind}}{\sim} \mathcal{G} \left( \xi_t; \frac{\nu_k}{2}, \frac{v_k - 2}{2} \right), \quad v_k > 2.
\]

Therefore, we have

\[
y_t \mid \mathcal{G}_{t-1}, z_t \sim \begin{cases} 
\mathcal{N} \left( y_t; \nu_{1,t}, (\tau_1 \xi_t)^{-1} \right) \mathcal{G} \left( \xi_t; \frac{\nu_1}{2}, \frac{v_1 - 2}{2} \right) & \text{if } z_t = 1, \\
\vdots & \\
\mathcal{N} \left( y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1} \right) \mathcal{G} \left( \xi_t; \frac{\nu_k}{2}, \frac{v_k - 2}{2} \right) & \text{if } z_t = k, \\
\vdots & \\
\mathcal{N} \left( y_t; \nu_{g,t}, (\tau_g \xi_t)^{-1} \right) \mathcal{G} \left( \xi_t; \frac{\nu_g}{2}, \frac{v_g - 2}{2} \right) & \text{if } z_t = g,
\end{cases}
\]

provided \( v_k > 2, \quad (k = 1, \ldots, g) \). That is, conditional on the information about the source of \( y_t \), only the corresponding component is relevant.

Consequently, the full conditional likelihood separates into \( g \) parts where each one concerning the data assigned to each of the \( g \) mixture components:

\[
\mathcal{L}(\varphi \mid y, z, \xi) = \prod_{k=1}^{g} \prod_{t: z_t = k} \mathcal{N} \left( y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1} \right) \mathcal{G} \left( \xi_t; \frac{\nu_k}{2}, \frac{v_k - 2}{2} \right), \quad v > 2.
\]

6.4 Priors setup

6.4.1 Priors for \( \omega, \mu, \) and \( \tau \)

For fixed \( g \), we will have a specific \( M_{AR}^t \) model. We set the prior distributions, \( \pi(\varphi \mid g) \), of the model parameters as same as that of the \( M_{AR} \) model. We state
these priors briefly again here for convention (for details please see Section 3.5):
\[\omega \mid g \sim D(\omega; \delta_1, \ldots, \delta_g) \quad (6.9)\]
\[\mu_k \overset{\text{iid}}{\sim} \mathcal{N}(\mu_k; \zeta, \kappa^{-1}), \quad (k = 1, \ldots, g) \quad (6.10)\]
\[\tau_k \mid \lambda \overset{\text{iid}}{\sim} \mathcal{G}(\tau_k; c, \lambda), \quad (k = 1, \ldots, g) \quad (6.11)\]
\[\phi_{ki} \overset{\text{iid}}{\sim} \mathcal{N}\left(\phi_{ki}; \mu_\phi, \tau^{-1}_\phi\right) \mathbf{1}_{\mathcal{M}_{AR}}(\text{Stable}) \quad (k = 1, \ldots, g; \ i = 1, \ldots, p_k) \quad (6.12)\]

### 6.4.2 Priors for \(g, p\) and \(v\)

For complete bayesian analysis of our parameters, we stress that \(g, p\) and \(v\) are stochastic also. The priors for these can be set as:
\[g \sim \mathcal{U}(g; 1, g_{\max}) \quad (6.13)\]
\[p_k \mid g \overset{\text{iid}}{\sim} \mathcal{U}(p_k; 0, p_{\max}), \quad (k = 1, \ldots, g) \quad (6.14)\]
\[v_k \mid g \overset{\text{iid}}{\sim} \mathcal{U}(v_k; 3, v_{\max}), \quad (k = 1, \ldots, g) \quad (6.15)\]

where \(\mathcal{U}(x \mid a, b)\) denotes the discrete uniform distribution for \(x = a, a + 1, \ldots, b\) and \(g_{\max}, p_{\max}\) and \(v_{\max}\) is fixed. For details of the above distributions mentioned above see Section 2.11.

### 6.4.3 Formulae for hyperparameters

In the prior settings, both the \(\mathcal{M}_{AR}\) and \(\mathcal{M}_{AR}^t\) models have some common hyper-parameters. We re-written the formulae here (for details, see Section 3.5.1):
\[\delta = 1\]
\[\mathcal{R}_y = \max(y) - \min(y)\]
\[\zeta = \min(y) + \frac{\mathcal{R}_y}{2}, \quad \kappa = \frac{1}{\mathcal{R}_y}\]
\[c = 2, \quad a = 0.2, \quad b = \frac{100 \ a}{c \mathcal{R}_y^2} \quad (6.16)\]
6.5 MCMC moves for parameter estimation

We let $\Lambda = (g, p, v)$ be the model index and $\varphi = (\omega, \mu, \tau, \phi)$ be the vector of parameters. We apply MCMC component-wise techniques to calculate the full conditional posterior distribution $\pi(\varphi | y, \Lambda)$ in which each of the parameter will update separately. Moreover, we need to update the latent variables $z$ and $\xi$ as well. There, we have a set of NINE moves in total which is discussed below.

6.5.1 Move I: updating $\omega$

Let $\varphi_{-\omega}$ denote the remaining parameters except $\omega$. Then the full conditional posterior distribution $\pi(\omega | \varphi_{-\omega}, y, z)$ can be obtained by applying a Gibbs-type move in similar fashion as in Section 3.6.2:

$$\omega | \varphi_{-\omega}, y, z \sim D(\omega | \delta_1 + n_1, \ldots, \delta_k + n_k),$$

(6.17)
CHAPTER 6. MIXTURE OF STUDENT-T AR COMPONENTS

where \( n_k = \sum_{t: z_t = k} 1 \).

6.5.2 Move II: updating \( \mu \)

Let \( \varphi_{-\mu} \) denote the remaining parameters except \( \mu \). Then the full conditional posterior distribution \( \pi(\mu \mid \varphi_{-\mu}, y, z) \) can be obtained by applying a Gibbs-type move as below:

\[
\pi(\mu \mid \varphi_{-\mu}, y, z, \xi) \propto L(\varphi \mid y, z, \xi) \pi(\mu)
\]

\[
\propto \prod_{k=1}^{g} \prod_{t: z_t = k} N\left(y_t; \mu_{k,t}, (\tau_k \xi_t)^{-1}\right) \mathcal{G}\left(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}\right)
\]

\[
\times \prod_{k=1}^{g} N\left(\mu_k; \zeta, \kappa^{-1}\right).
\]

Therefore, we have the full conditional posterior for the \( k \)-th component as below:

\[
\pi(\mu_k \mid \varphi_{-\mu_k}, y, z, \xi) \propto \prod_{t: z_t = k} N\left(y_t; \mu_{k,t}, (\tau_k \xi_t)^{-1}\right) \mathcal{G}\left(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}\right) \times N\left(\mu_k; \zeta, \kappa^{-1}\right)
\]

\[
\propto \exp\left[ -\frac{\tau_k}{2} \sum_{t: z_t = k} \xi_t \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t-i} - \mu_k) \right\}^2 \right]
\]

\[
\cdot \exp\left[ -\frac{\kappa}{2} (\mu_k - \zeta)^2 \right].
\]

Now the term \((#)\) is similar as in the case of \( \mathcal{M}_{AR} \) model. The details of simplification of this term are described in Section 3.6.3. We state the simplified result
as below:

\[
\sum_{t:z_t=k} \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_t - \mu_k) \right\}^2 \xi_t
\]

\[
= \sum_{t:z_t=k} \left\{ \left( y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i} \right) - \left( 1 - \sum_{i=1}^{p_k} \phi_{k,i} \right) \mu_k \right\}^2 \xi_t
\]

\[
= \sum_{t:z_t=k} \xi_t (e_{k,t} - b_k \mu_k)^2 = \sum_{t:z_t=k} \xi_t (e_{k,t} - \bar{e}_k + \bar{e}_k - b_k \mu_k)^2,
\]

where \( \bar{e}_k = \frac{1}{n_k} \sum_{t:z_t=k} e_{k,t} \) provided \( n_k = \sum_{t:z_t=k} 1 \)

\[
= \sum_{t:z_t=k} \xi_t \left\{ (e_{k,t} - \bar{e}_k)^2 + 2 (e_{k,t} - \bar{e}_k) (\bar{e}_k - b_k \mu_k) + (\bar{e}_k - b_k \mu_k)^2 \right\}.
\]
Substituting the simplified results into the equation (\#) we have,

\[
\pi(\mu_k \mid \varphi_{-\mu_k}, y, z,\xi) \propto \exp \left[ -\tau_k (\bar{e}_k - b_k\mu_k) \sum_{t: z_t = k} \xi_t (e_{k,t} - \bar{e}_k) - \frac{\tau_k}{2} (\bar{e}_k - b_k\mu_k)^2 \sum_{t: z_t = k} \xi_t \right]
\]

\[
\times \exp \left[ -\frac{\kappa}{2} (\mu_k - \zeta)^2 \right]
\]

\[
\propto \exp \left[ -\frac{1}{2} \left( \tau_k b_k^2 d_k + \kappa \right) \mu_k^2 + \left( \tau_k \bar{e}_k b_k d_k + \kappa \zeta + \tau_k b_k c_k \right) \mu_k \right]
\]

\[
: = A, \text{ say}
\]

\[
\propto \exp \left[ -A \mu_k^2 + 2B \mu_k \right] = \exp \left[ -A \left( \mu_k - \frac{B}{A} \right)^2 \right]
\]

\[
= \exp \left[ -A \left( \mu_k - \frac{B}{A} \right)^2 \right] = \exp \left[ -A \left( \mu_k - \frac{B}{A} \right)^2 \right]
\]

\[
\times \exp \left[ -\frac{\tau_k b_k^2 d_k + \kappa}{2} \left( \mu_k - \frac{\tau_k \bar{e}_k b_k d_k + \kappa \zeta + \tau_k b_k c_k}{\tau_k b_k^2 d_k + \kappa} \right)^2 \right].
\]

Hence, the full conditional posterior distribution of \( \mu \) is:

\[
\mu_k \mid \varphi_{-\mu_k}, y, z, \xi \sim \mathcal{N}(\mu_k; \frac{\tau_k b_k (\bar{e}_k d_k + c_k) + \kappa \zeta}{\tau_k b_k^2 d_k + \kappa}, \frac{1}{\tau_k b_k^2 d_k + \kappa}),
\]

\[(k = 1, \dotsc, g), \quad (6.18)\]

where

- \( \bar{e}_k = \frac{1}{n_k} \sum_{t: z_t = k} e_{k,t} \) with \( e_{k,t} = y_t - \sum_{i=1}^{p_k} \phi_{k,i}y_{t-i} \), and \( n_k = \sum_{t: z_t = k} 1 \);

- \( b_k = 1 - \sum_{i=1}^{p_k} \phi_{k,i}; \quad c_k = \sum_{t: z_t = k} \xi_t (e_{k,t} - \bar{e}_k); \quad d_k = \sum_{t: z_t = k} \xi_t. \)
6.5.3 Move III: updating $\phi$

Here we update $\phi_k$, $(k = 1, \ldots, g)$ by a Metropolis-Hasting (M-H) mechanism in a similar fashion as described in the case of the $\mathcal{M}_{AR}$ model. We discuss the algorithm briefly as below (for details, see Section 3.6.4):

A candidate $\phi^*_k$ is generated by a normal density and centred in the current state of the chain $\phi_k$ so that we can define the proposal density as below:

$$q(\phi_{ki}, \phi^*_{ki}) = \mathcal{N}_{(l,u)}(\phi^*_{ki}; \phi_{ki}, \tau_q^{-1}) \mathbf{1}_{\mathcal{M}_{AR}} \text{(Stable)}, \quad (k = 1, \ldots, g; i = 0, 1, \ldots, p_k).$$

(6.19)

where $\mathbf{1}_{\mathcal{M}_{AR}} \text{(Stable)}$ is the indicator function which takes unity if the model is stable otherwise zero (i.e. it satisfies the constraint $(C_S^\phi)$ discussed in Section 6.2.1), and $\mathcal{N}_{(l,u)}(\cdot) \mathbf{1}_{\mathcal{M}_{AR}} \text{(Stable)}$ represents that the normal distribution $\mathcal{N}(\cdot)$ is truncated at the boundary $(l, u)$ of the support of the indicator function $\mathbf{1}_{\mathcal{M}_{AR}} \text{(Stable)}$.

The precision $\tau_q$ is tuned so that a satisfactory acceptance rate can be reached.

Using the distribution defined above by (6.19), we generate a vector $\phi^*_k = (\phi^*_k, \phi^*_{k1}, \ldots, \phi^*_{kp_k})$ and the acceptance probability for this new $\phi^*_k$ is calculated as below:

$$A(\phi_k, \phi^*_k) = \min(1, \Omega),$$

where $\Omega = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}$.

First, from (6.8), the likelihood ratio is simply given by:

$$\text{likelihood ratio} = \frac{\prod_{t: z_t = k} \mathcal{N}(y_t; \nu^*_{k,t}, (\tau_k \xi_t)^{-1}) G\left(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}\right)}{\prod_{t: z_t = k} \mathcal{N}(y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1}) G\left(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}\right)}$$

$$= \exp\left[-\frac{\tau_k}{2} \sum_{t: z_t = k} \xi_t \left\{ (y_t - \nu^*_{k,t})^2 - (y_t - \nu_{k,t})^2 \right\} \right],$$

(6.20)

where $\nu^*_{k,t} = \mu_k + \phi^*_{k,1} (y_{t-1} - \mu_k) + \cdots + \phi^*_{k,p_k} (y_{t-p_k} - \mu_k)$.

Since all the parameters do not change except the AR coefficients, so from (6.12) the prior ratio is the ratio of the two normal densities:

$$\text{prior ratio} = \exp\left[-\frac{\tau}{2} \sum_{i=1}^{p_k} \{(\phi^*_{ki} - \mu_\phi)^2 - (\phi_{ki} - \mu_\phi)^2\}\right],$$

(6.21)
and, eventually from the Equation (6.19) the proposal ratio is:

\[
\text{proposal ratio} = \prod_{i=1}^{p_k} \frac{\Phi_N(u; \phi_{k,i}, \tau_q^{-1}) - \Phi_N(l; \phi_{k,i}, \tau_q^{-1})}{\Phi_N(u; \phi_{k,i}^*, \tau_q^{-1}) - \Phi_N(l; \phi_{k,i}^*, \tau_q^{-1})},
\]

(6.22)

where \( \Phi_N \) is the normal cumulative distribution function.

By using (6.20), (6.21) and (6.22) we have,

\[
\Omega = \exp \left[ -\frac{\tau_k}{2} \sum_{t: z_t = k} \xi_t \left\{ (y_t - \nu_{k,t}^*)^2 - (y_t - \nu_{k,t})^2 \right\} \right]
\]

\[
\times \exp \left[ -\frac{\tau_c}{2} \sum_{i=1}^{p_k} \left\{ (\phi_{ki}^* - \mu_c)^2 - (\phi_{ki} - \mu_c)^2 \right\} \right]
\]

\[
\times \prod_{i=1}^{p_k} \frac{\Phi_N(u; \phi_{k,i}^*, \tau_q^{-1}) - \Phi_N(l; \phi_{k,i}^*, \tau_q^{-1})}{\Phi_N(u; \phi_{k,i}, \tau_q^{-1}) - \Phi_N(l; \phi_{k,i}, \tau_q^{-1})},
\]

(6.23)

### 6.5.4 Move IV: updating \( \lambda \) and \( \tau \)

Let \( \varphi_{-\lambda} \) denote the remaining parameters except \( \lambda \) and \( \varphi_{-\tau} \) denotes the remaining parameters except \( \tau \). Firstly, the full conditional posterior distribution \( \pi(\lambda | \varphi_{-\lambda}, y, z, \xi) \) can be obtained by applying a Gibbs-type move as below:

\[
\lambda | \varphi_{-\lambda}, y, z \sim \mathcal{G}(\lambda; a + gc, b + \sum_{k=1}^{g} \tau_k), \quad (k = 1, \ldots, g)
\]

(6.24)

For details of the derivation, see Section 3.6.5.

Secondly, the full conditional posterior distribution of \( \tau \) can be obtained as below:

\[
\pi(\tau | \varphi_{-\tau}, y, z, \xi) \propto \mathcal{L}(\varphi | y) \pi(\tau | \lambda)
\]

\[
\propto \prod_{k=1}^{g} \prod_{t: z_t = k} \mathcal{N}(y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1}) \mathcal{G}(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}) \prod_{k=1}^{g} \mathcal{G}(\tau_k; c, \lambda)
\]
Consequently, we have the full conditional posterior density for the \( k \)-th component:

\[
\pi(\tau_k \mid \varphi_{-\tau_k}, y, z, \xi) \propto \prod_{t: z_t = k} \mathcal{N}(y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1}) \mathcal{G}\left(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}\right) \cdot \mathcal{G}(\tau_k; c, \lambda)
\]

\[
\propto (\tau_k \xi_t)^{n_k/2} \exp\left[-\frac{\tau_k}{2} \sum_{t: z_t = k} \xi_t (y_t - \nu_{k,t})^2\right] \cdot \tau_k^{c-1} \exp\left[-\lambda \tau_k\right]
\]

\[
\propto (\tau_k)^{c+n_k/2-1} \exp\left[-\left(\lambda + \frac{1}{2} \sum_{t: z_t = k} \xi_t (y_t - \nu_{k,t})^2\right) \tau_k\right]
\]

Hence, the full conditional posterior distribution of \( \boldsymbol{\tau} \) is:

\[
\tau_k \mid \varphi_{-\tau_k}, y, z, \xi \sim \mathcal{G}(\tau_k; c + \frac{n_k}{2}, \lambda + \frac{1}{2} \sum_{t: z_t = k} \xi_t (y_t - \nu_{k,t})^2), \quad (k = 1, \ldots, g).
\]

(6.25)

### 6.5.5 Move V: updating \( z_t \)

This is also a Gibbs-type move by which we have the full conditional posterior distribution of \( z \):

\[
\pi(z \mid \varphi, y) \propto \mathcal{L}(\varphi \mid y, z) \pi(z \mid \omega)
\]

\[
\propto \prod_{t=1}^{n} \sum_{k=1}^{g} \omega_k \mathcal{N}(y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1}) \mathcal{G}\left(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}\right) \cdot 1_{\{z_t = k\}}
\]

\[
\propto \prod_{t=1}^{n} \sum_{k=1}^{g} \omega_k \mathcal{N}(y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1}) \cdot 1_{\{z_t = k\}}
\]

Therefore, we have the full conditional probability that the observation \( y_t \) has been generated by the \( k \)-th mixture component is:
\[
\Pr(z_t = k \mid y_t, \varphi) = \frac{1}{C_2} \omega_k \sqrt{\tau_k} \exp \left[ -\frac{\tau_k \xi_t}{2} (y_t - \nu_{k,t})^2 \right], \\
(t = 1, \ldots, T; k = 1, \ldots, g-1), \quad (6.26)
\]

where \( C_2 = \sum_{u=1}^{g} \omega_u \sqrt{\tau_u} \exp \left[ -\frac{\tau_u \xi_t}{2} (y_t - \nu_{u,t})^2 \right] \), is the normalizing constant; \( \nu_{\#t} \) is defined as in the equation (6.1). Obviously, the probability of having been generated by the last component is,

\[
\Pr(z_t = g \mid y_t, \varphi) = 1 - \sum_{k=1}^{g-1} \Pr(z_t = k \mid y_t, \varphi).
\]

**6.5.6 Move VI: updating \( \xi \)**

To get a full conditional posterior distribution of \( \xi \), we apply the Gibbs-typed move as below:

\[
\pi(\xi \mid \varphi, y, z) \propto \mathcal{L}(\varphi \mid y, z) \pi(z \mid \omega) \\
= \prod_{t=1}^{n} \sum_{k=1}^{g} \omega_k \mathcal{N}(y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1}) \mathcal{G}(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}) \cdot 1_{\{z_t = k\}}.
\]

Therefore, the full conditional distribution of \( \xi_t \mid z_t = k \) is:

\[
\pi(\xi_t \mid \varphi, y, z_t = k) \propto \omega_k \mathcal{N}(y_t; \nu_{k,t}, (\tau_k \xi_t)^{-1}) \mathcal{G}(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}) \\
\propto \omega_k (\tau_k \xi_t)^{1/2} \exp \left[ -\frac{\tau_k \xi_t}{2} (y_t - \nu_{k,t})^2 \right] \\
\cdot \xi_t^{\nu_k/2-1} \exp \left[ -\left(\frac{v_k - 2}{2}\right) \xi_t \right] \\
\propto \xi_t^{\nu_k/2+\frac{1}{2}-1} \exp \left[ -\left\{ \frac{\tau_k}{2} (y_t - \nu_{k,t})^2 + \frac{v_k - 2}{2} \right\} \xi_t \right].
\]

Hence, we have the full conditional posterior distribution of \( \xi \):
\[ \xi_t | \varphi, y, z_t = k \sim G \left( \xi_t ; \frac{v_k}{2} + \frac{1}{2}, \frac{\tau_k}{2} (y_t - \nu_{k,t})^2 + \frac{v_k - 2}{2} \right), \]

\[(t = 1, \ldots, T; k = 1, \ldots, g), \quad (6.27)\]

where \( \nu_{k,t} \) is as defined in equation (6.1).

### 6.6 MCMC moves for model determination

Model determination for the \( \mathcal{M}_{\text{AR}}^t \) model is the same as that for the \( \mathcal{M}_{\text{AR}} \) model. We recall all techniques here in short. For details, see Section 3.7. To get full Bayesian flavour we must get the posterior distribution of the model index \( \Lambda = (g, p, \nu) \) which can be written and factorized as:

\[ \pi(\Lambda | y) = \pi(p | g, y) \times \pi(\nu | g, y) \times \pi(g | y) \quad (6.28) \]

So, it is enough to update AR order \( p \) and component number \( g \) separately to update the model index \( \Lambda \). We develop two additional MCMC moves to complete our Bayesian analysis as follows:

#### 6.6.1 Move VII: updating of \( p \)

This is almost the same as discussed in Section 3.7.1. For the reader, we present that discussion very shortly here.

To update the AR order of the \( k \)-th component of the \( \mathcal{M}_{\text{AR}}^t \) model, we apply the RJMCMC (Reversible Jump Markov chain Monte Carlo) technique introduced by Green (1995). While designing the appropriate moves, we ensure that the dimensions can be balanced properly, the moves can be simulated conveniently, and the acceptance ratio can be computed economically.

At first we need to choose a component, say \( k^* \), randomly chosen in \( \{1, \ldots, g\} \). Then we can write the proposal for the new AR order of the \( k^* \)-th component, \( p^*_{k^*} \), as:

\[ p^*_{k^*} = \begin{cases} 
    p_{k^*} - 1, & \text{with probability } d(p_{k^*}), \\
    p_{k^*} + 1, & \text{with probability } b(p_{k^*}),
\end{cases} \quad (6.29) \]
where \( b(p_k) = 1 - d(p_k) \), for \( k = 1, \ldots, g \). \( d(0) = 0 \) and \( b(p_{\max}) = 0 \). The function \( b(x) \) has been defined in section 3.7.1.

Let \( \phi_{k^*}^* \) be the proposal vector of the AR coefficients. When \( p_{k^*}^* = p_{k^*} - 1 \), the autoregressive coefficients \( \phi_{k^*}^* = (\phi_{k^*1}, \ldots, \phi_{k^*p_{k^*}^*}) \) is updated by simply discarding the last AR coefficient in \( \phi_{k^*}^* = (\phi_{k^*1}, \ldots, \phi_{k^*p_{k^*}^*}) \). On the other hand, when \( p_{k^*}^* = p_{k^*} + 1 \), the corresponding arcoefficients \( \phi_{k^*}^* = (\phi_{k^*1}, \ldots, \phi_{k^*p_{k^*}^*}, \phi_{k^*p_{k^*}^*+1}^*) \) is updated by adding new parameter \( \phi_{k^*p_{k^*}^*+1}^* \) in \( \phi_{k^*}^* = (\phi_{k^*1}, \ldots, \phi_{k^*p_{k^*}^*}) \) with

\[
\phi_{k^*p_{k^*}^*+1}^* \sim \mathcal{N}(\phi_{k^*p_{k^*}^*+1}^*; \mu, \tau^{-1}),
\]

To update the AR order of the model by using RJMCMC method, we have to calculate the acceptance probability,

\[
A_p = \min\{1, \Omega_p\},
\]

and \( \Omega_p \) can be obtained as in Section 3.7.1:

\[
\Omega_p = \begin{cases} 
L^* \frac{b(p_{k^*}^*)}{d(p_{k^*}^*)} & \text{if } p_{k^*}^* = p_{k^*} - 1, \\
L^* \frac{d(p_{k^*}^*)}{b(p_{k^*}^*)} & \text{if } p_{k^*}^* = p_{k^*} + 1,
\end{cases}
\]

where

\[
L^* = \exp \left[ -\frac{\tau_k}{2} \sum_{t=1}^{\xi_t} \left\{ (y_t - \nu_{k^*t}^*)^2 - (y_t - \nu_{k't}^*)^2 \right\} \right]
\]

with \( \nu_{k't}^* = \mu_{k^*} + \phi_{k't1}^* (y_{t-1} - \mu_{k^*}) + \cdots + \phi_{k'tp_{k't}^*}^* (y_{t-p_{k't}^* - 1} - \mu_{k^*}) \).

Note that the Equation (6.30) has quite simple form and \( \pi(p \mid g, y) \) is simply estimated by the proportions of every possible value for \( p \) in the sample obtained by the previous complete MCMC algorithm.

### 6.6.2 Move VIII: updating of \( v \)

We will apply the same technique to update the parameter \( v \) as discussed in Section 3.7.1.

At first we need to choose a component, say \( k^* \), randomly chosen in \( \{1, \ldots, g\} \).

Then the corresponding degrees of freedom \( v_{k^*} \) increases or decreases by one with the following rule:
CHAPTER 6. MIXTURE OF STUDENT-T AR COMPONENTS

\[ v^{*}_{k*} = \begin{cases} 
  v_{k*} - 1, & \text{with probability } d(v_{k*}), \\
  v_{k*} + 1, & \text{with probability } b(v_{k*}),
\end{cases} \tag{6.31} \]

where \( b(x) = 1 - d(x) \), for \( k = 1, \ldots, g \), \( d(3) = 0 \) and \( b(v_{\text{max}}) = 0 \) because \( v_k \in [3, v_{\text{max}}] \). The function \( b(x) \) is defined in Section 3.7.1.

Let \( \xi^{*}_t \) be the proposal vector conditional to \( z = k^{*} \). When \( v^{*}_{k*} = v_{k*} - 1 \), then the corresponding missing variable \( \xi^{*}_t \mid z_t = k^{*} \sim (\xi_t; \frac{v_{k*}-1}{2}, \frac{v_{k*}-3}{2}) \) and when \( v^{*}_{k*} = v_{k*} + 1 \), then \( \xi^{*}_t \mid z_t = k^{*} \sim (\xi_t; \frac{v_{k*}+1}{2}, \frac{v_{k*}-1}{2}) \). To update the degrees of freedom parameter, \( v \), of the model by using RJMCMC method, we have to calculate the acceptance probability,

\[ A_{v^{*}} = \min\{1, \Omega_{v^{*}}\}, \]

and \( \Omega_{v^{*}} \) is described as in Section 2.7.1:

\[ \Omega_{v^{*}} = \text{likelihood ratio} \times \text{prio ratio} \times \text{proposal ratio} \times \text{jacobian}. \]
First, from (6.8) the likelihood ratio becomes:

\[
\text{likelihood ratio} = \frac{\prod_{t: z_t = k^*} \mathcal{N}(y_t; \nu_{k^*, t}, (\tau_{k^*} \xi_t)^{-1}) \mathcal{G}(\xi_t^*; \frac{v_{k^*}^*}{2}, \frac{v_{k^*}^* - 2}{2})}{\prod_{t: z_t = k^*} \mathcal{N}(y_t; \nu_{k^*, t}, (\tau_{k^*} \xi_t)^{-1}) \mathcal{G}(\xi_t; \frac{v_{k^*}}{2}, \frac{v_{k^*} - 2}{2})}
\]

\[
= \left(\frac{\sqrt{\tau_{k^*} \xi_t}}{2\pi}\right)^{\frac{n_{k^*}}{2}} \sum_{t: z_t = k^*} \exp \left[ -\frac{1}{2} \sum_{t: z_t = k^*} \left\{ (y_t - \nu_{k^*, t})^2 - (y_t - \nu_{k^*, t})^2 \right\} \right]
\]

\[
\times \frac{\left(\frac{v_{k^*} - 2}{2}\right)^\frac{n_{k^*}}{2}}{\Gamma\left(\frac{v_{k^*}}{2}\right)} \left(\frac{\tau_{k^*} \xi_t}{2}\right)^{\frac{n_{k^*} - v_{k^*}^*}{2}} \sum_{t: z_t = k^*} \exp \left[ -\frac{1}{2} \sum_{t: z_t = k^*} \left\{ (y_t - \nu_{k^*, t})^2 - (y_t - \nu_{k^*, t})^2 \right\} \right]
\]

\[
= \left(\frac{\xi_t^*}{\xi_t}\right)^{n_{k^*}} \times \left(\frac{\Gamma(\nu_{k^*}/2)}{\Gamma(v_{k^*}/2)}\right)^{n_{k^*}} \frac{(v_{k^*} - 2)^{n_{k^*} - v_{k^*}^*/2}}{(v_{k^*} - 2)^{n_{k^*} - v_{k^*}^*/2}} \times \frac{\xi_t^{(v_{k^*}/2) - (v_{k^*}/2) - 1)n_{k^*}}{\xi_t^{(v_{k^*}/2) - 1)n_{k^*}}}
\]

\[
\times \frac{\exp\left[-\frac{1}{2} \sum_{t: z_t = k^*} \left( (y_t - \nu_{k^*, t})^2 - (y_t - \nu_{k^*, t})^2 \right) \right]}{\exp\left[-\frac{1}{2} \sum_{t: z_t = k^*} \left( (y_t - \nu_{k^*, t})^2 - (y_t - \nu_{k^*, t})^2 \right) \right]}, \text{ since } \sum_{t: z_t = k^*} 1 = n_{k^*}
\]

\[
= C \exp\left[-\frac{n_{k^*}}{2} \left\{ (v_{k^*} - 2)\xi_t^* - (v_{k^*} - 2)\xi_t \right\} \right],
\]

where 

\[
C = \left(\frac{\xi_t^*}{\xi_t}\right)^{n_{k^*}} \times \left(\frac{\Gamma(\nu_{k^*}/2)}{\Gamma(v_{k^*}/2)}\right)^{n_{k^*}} \frac{(v_{k^*} - 2)^{n_{k^*} - v_{k^*}^*/2}}{(v_{k^*} - 2)^{n_{k^*} - v_{k^*}^*/2}} \times \frac{\xi_t^{(v_{k^*}/2) - (v_{k^*}/2) - 1)n_{k^*}}{\xi_t^{(v_{k^*}/2) - 1)n_{k^*}}}.
\]

The prior ratio is given by:

\[
\text{prior ratio} = \frac{\pi(v_{k^*}^*)\pi(\xi_t^* | z_t = k^*)}{\pi(v_{k^*})\pi(\xi_t | z_t = k^*)} = \frac{\pi(\xi_t^* | z_t = k^*)}{\pi(\xi_t | z_t = k^*)}
\]

because of the uniform prior on the AR order of the model and only the $k^*$-th component is updated.
CHAPTER 6. MIXTURE OF STUDENT-T AR COMPONENTS

To calculate the proposal ratio we have to take into account the probabilities \( b(\cdot) \) and \( d(\cdot) \) and the proposal density, which is chosen equal to the prior on \( \xi \).

\[
\text{proposal ratio} = \begin{cases} 
\frac{b(v_{k^*}) \pi(\xi_{t|z=t=k^*})}{d(v_{k^*})} & \text{if } v_{k^*} = v_{k^*} - 1, \\
\frac{d(v_{k^*}^*)}{b(v_{k^*}) \pi(\xi_{t|z=t=k^*})} & \text{if } v_{k^*} = v_{k^*} + 1.
\end{cases}
\]

Finally, the Jacobian is equal to 1 because the invertible function \( g(\cdot) \) is considered as the identity function.

Therefore, multiplying likelihood, prior and proposal ratios, we derive ratio in the acceptance probability for the both the cases:

\[
\Omega_{p^*} = \begin{cases} 
D \exp \left[ -\frac{n_{k^*}}{2} \left\{ (v_{k^*}^* - 2)\xi_{t}^* - (v_{k^*}^* - 2)\xi_{t} \right\} \right] \frac{b(v_{k^*})}{d(v_{k^*})} & \text{if } v_{k^*} = v_{k^*} - 1, \\
E \exp \left[ -\frac{n_{k^*}}{2} \left\{ (v_{k^*}^* - 2)\xi_{t}^* - (v_{k^*}^* - 2)\xi_{t} \right\} \right] \frac{d(v_{k^*}^*)}{b(v_{k^*})} & \text{if } v_{k^*} = v_{k^*} + 1,
\end{cases}
\]

where \( D = C \prod_{t:z=t=k^*} G(\xi_t; \frac{v_{k^*}}{2}, \frac{v_{k^*}}{2}) \) and \( E = C/ \prod_{t:z=t=k^*} G(\xi_t; \frac{v_{k^*}}{2}, \frac{v_{k^*}}{2}) \).

Therefore, \( \pi(p \mid g, y) \) is simply estimated by the proportions of every possible value for \( p \) in the sample obtained by the previous complete MCMC algorithm.

6.6.3 Move IX: updating of \( g \)

This move will complete our Bayesian analysis for the \( \mathcal{M}_{\text{AR}} \) model. According to Bayes’ theorem, the marginal posterior distribution of \( g \) is:

\[
\pi(g \mid y) \propto f(y \mid g)\pi(g),
\]

where \( \pi(g) \) is the prior on \( g \) and \( f(y \mid g) \) is marginal likelihood given by

\[
f(y \mid g) = \sum_p \int f(y \mid \varphi, p, v, g)\pi(\varphi, p, v \mid g)d\varphi.
\]

We apply the method which is described by Chib and Jeliazkov (2001) and it is based on the basic marginal likelihood identity. We suppress the model index \( g \)
CHAPTER 6. MIXTURE OF STUDENT-T AR COMPONENTS

for notational convenience:

\[ f(y) = \frac{f(y | \varphi^*, p^*, v^*) \pi(\varphi^*, p^*, v^*)}{\pi(\varphi^*, p^*, v | y)} \]

\[ = \frac{f(y | \varphi^*, p^*) \pi(\varphi^* | p^*, v^*) \pi(p^*) \pi(v^*)}{\pi(\varphi^* | p^*, v^*) \pi(p^* | y) \pi(v^* | y)} \]  \hspace{1cm} (6.35)

for any fixed point \((\varphi^*, p^*, v^*)\) which have the maximum density.

Note that the only quantity in (6.35) that is not already available is \(\pi(\varphi^* | p^*, v^*, y)\). Chib and Jeliazkov (2001) proposed an efficient method to produce the estimate \(\widehat{\pi}(\varphi^* | p^*, v^*, y)\) using the output from a MCMC simulation with fixed \(p^*\).

First of all, \(\widehat{\pi}(\varphi^* | p^*, v^*, y)\) is factorized as

\[
\widehat{\pi}(\varphi^* | p^*, v^*, y) = \widehat{\pi}(\phi^* | p^*, y) \times \widehat{\pi}(\mu^* | \phi^*, p^*, v^*, y) \\
\times \widehat{\pi}(\tau^* | \mu^*, \phi^*, p^*, v^*, y) \times \widehat{\pi}(\omega^* | \tau^*, \mu^*, \phi^*, p^*, v^*, y).
\]  \hspace{1cm} (6.36)

Suppose we have a sample \(\{\varphi^{(i)}, z^{(i)}, \xi^{(i)}\}\), for \(i = 1, \ldots, N_1\), from the MCMC algorithm for a given \(p^*\) and \(v^*\) (i.e. a sample from \(\pi(\varphi | p^*, v^*)\)). Then the terms of (6.36) are estimated by the following steps:

**Estimation of \(\widehat{\pi}(\phi^* | p^*, v^*, y)\):**

Let \(\mathcal{U}_{k-1} = (p, \phi_1, \ldots, \phi_{k-1})\) and \(\mathcal{U}^{k+1} = (\phi_{k+1}, \ldots, \phi_g, v, \mu, \tau, \omega)\) be the two partitions of the parameters for any \(k \in \{1, \ldots, g\}\).

Let \(\{\tilde{\mathcal{U}}^{k+1, (i)}, \tilde{z}^{(i)}\}\), for \(i = 1, \ldots, N\), are drawn from a reduced MCMC algorithm with distribution of interest \(\pi(\tilde{\mathcal{U}}^{k+1}, z | \tilde{\mathcal{U}}^*_k, y)\).

Also draw \(\tilde{\phi}^{(i)}_k\), for \(i = 1, \ldots, N\), from the proposal density \(q_p(\phi^*_k, \tilde{\phi}^{(i)}_k) = \prod_{j=1}^{p_k} q(\phi^{*_{kj}}, \tilde{\phi}^{(i)}_{kj})\), where \(q(\cdot, \cdot)\) is defined by the Equation (6.19) in Section 6.5.3.
We set
\[
\pi(\phi^*_k \mid p^*, \phi_1^*, \ldots, \phi_{k-1}^*) = \frac{1}{N-B} \sum_{i=B+1}^{N} A(\phi_k^{(i)}, \phi_k^*) q_p(\phi_k^{(i)}, \phi_k^*)
\]
where \(A(\cdot, \cdot) = \min(1, \bar{\Omega}(\cdot, \cdot))\) with \(\bar{\Omega}(\cdot, \cdot)\) defined in Equation (6.23).

Now we set \(\tilde{\phi}_{k+1}^*, (i) \) = \(\tilde{\phi}_{k+1}^*, (i) \) and \(z^*(i) = \tilde{z}^*(i)\), for \(i = 1, \ldots, N\). Repeat this step for \(k = 1, \ldots, g\) and finally set
\[
\pi(\mu^* \mid \phi^*, p^*, y) = \prod_{k=1}^{g} \pi(\phi_k^* \mid p^*, \phi_1^*, \ldots, \phi_{k-1}^*).
\]
(6.37)

**Estimation of \(\pi(\mu^* \mid \phi^*, p^*, y)\):**

Let \((\tau^{(i)}, z^{(i)}, \xi^{(i)})\) are drawn from the last iteration of the previous step (thus they are marginally from \(\pi(\tau, \omega, z, \xi \mid \phi^*, p^*, y)\)) and consequently,
\[
\pi(\mu_k^* \mid \phi^*, p^*, y) = \frac{1}{N-B} \sum_{i=B+1}^{N} \prod_{k=1}^{g} \pi(\mu_k^* \mid \phi^*, \tau^{(i)}, z^{(i)}, \xi^{(i)}, p^*, y),
\]
(6.38)

where \(\pi(\mu_k^* \mid \phi^*, \tau^{(i)}, z^{(i)}, \xi^{(i)}, p^*, y)\) is given by the same distribution as (6.18).

**Estimation of \(\pi(\tau^* \mid \phi^*, \mu^*, p^*, y)\):**

Let \(\{\tau^{(u)}, \omega^{(u)}, z^{(u)}, \xi^{(u)}\}\), for \(u = 1, \ldots, N\), are drawn from a reduced MCMC algorithm with distribution of interest \(\pi(\tau, \omega, z, \xi \mid \phi^*, \mu^*, p^*, y)\). Then we have,
\[ \hat{\pi}(\tau^* | \phi^*, \mu^*, p^*, v^*, y) \]

\[ = \frac{1}{N - B} \sum_{u=B+1}^{N} \prod_{k=1}^{g} \pi(\tau_k^* | \phi^*, \mu^*, z^{(u)}, \xi^{(u)}, p^*, v^*, y), \]

(6.39)

where \( \pi(\tau^*_k | \phi^*, \mu^*, z^{(u)}, \xi^{(u)}, p^*, v^*, y) \) is given by the Equation (6.25).

Estimation of \( \hat{\pi}(\omega^* | \phi^*, \mu^*, \tau^*, p^*, v^*, y) \):

Let \( \{\omega^{(s)}, z^{(s)}, \xi^{(s)}\} \), for \( s = 1, \ldots, N \), are drawn from a reduced MCMC algorithm with distribution of interest \( \pi(\omega, z | \phi^*, \mu^*, \tau^*, p^*, y) \). So, we have:

\[ \hat{\pi}(\omega^* | \phi^*, \mu^*, \tau^*, p^*, v^*, y) = \frac{1}{N - B} \sum_{s=B+1}^{N} \pi(\omega^* | z^{(s)}, \xi^{(s)} p^*, v^*, y), \]

(6.40)

where \( \pi(\omega^* | z^{(s)}, p^*, y) \) is given by the Equation (6.17) in Section 6.5.1.

Using (6.37), (6.38), (6.39), and (6.40) into (6.35) and then finally, we can estimate the posterior probability, \( \pi(g | y) \) from (6.33).

6.7 Examples

To illustrate our Bayesian algorithm for the \( \mathcal{M}_{AR} \) model, we consider one simulated series and one the real data such \( y_{MoI} \) and \( y_{DGS3} \) respectively. Details of these can be found in Section 3.8. For the simulated series we generated \( N = 100000 \) runs of the three Markov chain and for the data we generated \( N = 50000 \) runs of the three Markov chains. We discarded initial 10000 from each chain as burn-in.

Several diagnostic tests were performed to assess the convergence of the Markov chains. Convergence diagnostics of sampled values of parameters for the data \( y_{DGS3} \) are only reported here in Table 6.2. We found the chains have considered in all cases that is not reported here for saving space.

Table 6.1 represents parameter estimation results of sampled values for the model \( \mathcal{M}_{AR}^t(2; 1, 1; 4, 8) \) fitted the simulated series \( y_{MoI} \). Figure 6.2 shows convergence
Table 6.1: Estimation results of sampled values of parameters when fitting the $\mathcal{M}_{AR}^t(2; 1, 1; 4, 8)$ model to the $y_{MoI}$ series with $N = 100000$ and burn-in=50000.

<table>
<thead>
<tr>
<th>True value</th>
<th>Posterior statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>(0.0001001)</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>(0.0001001)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>(0.0003389)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>(0.0005227)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.001163)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(0.001312)</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>(0.001551)</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>(0.001584)</td>
</tr>
</tbody>
</table>

\(*\) Highest Posterior Density of the parameter of the MCMC sample

Diagrams and marginal posterior densities of each parameters of the best fitted model $\mathcal{M}_{AR}^t(2; 1, 1; 4, 8)$ for the simulated series $y_{MoI}$. Observe that our Bayesian algorithm fit well for $\mathcal{M}_{AR}^t$ model.

Figure 6.3 shows convergence diagrams and marginal posterior densities of parameters $\omega, \mu, \sigma$ for the fitted model $\mathcal{M}_{AR}^t(3; 0, 4, 4)$ for the $y_{DGS3}$ data. Figure 6.4 shows MCMC output of sampled values of $\phi$ for the data $y_{DGS3}$.

Finally, Table 6.3 shows parameter estimation results of MCMC sampled values of parameters while fitting $\mathcal{M}_{AR}^t$ to the data $y_{DGS3}$. These suggests that our Method is performed well for the data $y_{DGS3}$ also.
Table 6.2: Convergence diagnostics of sampled values of parameters when fitting the $\mathcal{M}^{t}_{\text{AR}}(3; 0, 4, 4; 5, 5, 5)$ model to the $\text{yDGS3}$ data with $N = 50000$ and burn-in = 10000.

<table>
<thead>
<tr>
<th>Stationarity test</th>
<th>Start iteration</th>
<th>p-value</th>
<th>Halfwidth test</th>
<th>Mean</th>
<th>Halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>passed</td>
<td>0.140</td>
<td>passed</td>
<td>0.524</td>
<td>0.00413</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>passed</td>
<td>0.118</td>
<td>passed</td>
<td>0.281</td>
<td>0.00296</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>passed</td>
<td>0.329</td>
<td>passed</td>
<td>0.195</td>
<td>0.00210</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>passed</td>
<td>0.3750</td>
<td>passed</td>
<td>-2.62e-05</td>
<td>1.34e-05</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>passed</td>
<td>0.0657</td>
<td>passed</td>
<td>-3.19e-04</td>
<td>5.70e-05</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>passed</td>
<td>0.2419</td>
<td>passed</td>
<td>-3.66e-04</td>
<td>6.00e-05</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>passed</td>
<td>0.277</td>
<td>passed</td>
<td>0.0106</td>
<td>3.51e-05</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>passed</td>
<td>0.308</td>
<td>passed</td>
<td>0.0161</td>
<td>8.77e-05</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>passed</td>
<td>0.553</td>
<td>passed</td>
<td>0.0253</td>
<td>1.34e-04</td>
</tr>
</tbody>
</table>

The Geweke diagnostic:

<table>
<thead>
<tr>
<th>Z-score</th>
<th>Z-score</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>-0.09951167</td>
<td>$\mu_1$</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.33799317</td>
<td>$\mu_2$</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>-0.20718795</td>
<td>$\mu_3$</td>
</tr>
</tbody>
</table>
Table 6.3: Estimation results of MCMC sampled values of parameters while fitting the $M_{ar}^t(3; 0, 4, 4)$ model to the $y_{DGS3}$ data with $N = 50000$ and $B = 10000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.5258</td>
<td>0.04000</td>
<td>0.0001789</td>
<td>0.5252</td>
<td>[0.449958, 0.602300]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.2805</td>
<td>0.02940</td>
<td>0.0001315</td>
<td>0.2800</td>
<td>[0.224189, 0.337226]</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.1938</td>
<td>0.02519</td>
<td>0.0001127</td>
<td>0.1938</td>
<td>[0.145232, 0.242860]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>$-2.497 \times 10^{-5}$</td>
<td>0.0003908</td>
<td>$1.748 \times 10^{-6}$</td>
<td>$-2.679 \times 10^{-5}$</td>
<td>[-0.000790, 0.000749]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>$-3.163 \times 10^{-4}$</td>
<td>0.0011589</td>
<td>$5.183 \times 10^{-6}$</td>
<td>$-3.120 \times 10^{-4}$</td>
<td>[-0.002532, 0.002026]</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>$-3.477 \times 10^{-4}$</td>
<td>0.0013812</td>
<td>$6.177 \times 10^{-6}$</td>
<td>$-3.536 \times 10^{-4}$</td>
<td>[-0.003032, 0.002421]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>$-2.497 \times 10^{-5}$</td>
<td>0.0003908</td>
<td>$1.748 \times 10^{-6}$</td>
<td>$-2.672 \times 10^{-4}$</td>
<td>[-0.000790, 0.000749]</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>$-2.708 \times 10^{-4}$</td>
<td>0.0009476</td>
<td>$4.238 \times 10^{-6}$</td>
<td>-0.00002</td>
<td>[-0.002083, 0.001597]</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>$-3.911 \times 10^{-4}$</td>
<td>0.0015497</td>
<td>$6.930 \times 10^{-6}$</td>
<td>$-4.164 \times 10^{-4}$</td>
<td>[-0.003455, 0.002624]</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.01061</td>
<td>0.0005686</td>
<td>$2.543 \times 10^{-6}$</td>
<td>0.01060</td>
<td>[0.009532, 0.011665]</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.01627</td>
<td>0.0016929</td>
<td>$7.571 \times 10^{-6}$</td>
<td>0.01607</td>
<td>[0.013402, 0.019080]</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.02521</td>
<td>0.0023510</td>
<td>$1.051 \times 10^{-5}$</td>
<td>0.02519</td>
<td>[0.021017, 0.029936]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.121368</td>
<td>0.08324</td>
<td>0.0003723</td>
<td>0.1108964</td>
<td>[-0.161095, 0.302492]</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.375314</td>
<td>0.08645</td>
<td>0.0003866</td>
<td>-0.3877293</td>
<td>[-0.494292, -0.287995]</td>
</tr>
<tr>
<td>$\phi_{23}$</td>
<td>-0.018596</td>
<td>0.06191</td>
<td>0.0002769</td>
<td>-0.0144039</td>
<td>[-0.131950, 0.082196]</td>
</tr>
<tr>
<td>$\phi_{24}$</td>
<td>0.416073</td>
<td>0.11257</td>
<td>0.0005034</td>
<td>0.4290372</td>
<td>[0.305450, 0.526759]</td>
</tr>
<tr>
<td>$\phi_{31}$</td>
<td>0.052273</td>
<td>0.12210</td>
<td>0.0005461</td>
<td>0.0680850</td>
<td>[-0.253695, 0.257609]</td>
</tr>
<tr>
<td>$\phi_{32}$</td>
<td>0.277151</td>
<td>0.11331</td>
<td>0.0005067</td>
<td>0.2848711</td>
<td>[0.117789, 0.459679]</td>
</tr>
<tr>
<td>$\phi_{33}$</td>
<td>-0.001978</td>
<td>0.10115</td>
<td>0.0004523</td>
<td>-0.0004673</td>
<td>[-0.192026, 0.189198]</td>
</tr>
<tr>
<td>$\phi_{34}$</td>
<td>-0.510443</td>
<td>0.13648</td>
<td>0.0006103</td>
<td>-0.5193292</td>
<td>[-0.721276, -0.344216]</td>
</tr>
</tbody>
</table>
CHAPTER 6. MIXTURE OF STUDENT-T AR COMPONENTS

201

Figure 6.2: MCMC output of sampled values of the parameters while fitting the $\mathcal{M}_{\text{AR}}(2; 1, 1; 4, 8)$ model to the simulated series $y_{\text{MoI}}$ with $N = 100000$ and burn-in=50000.
CHAPTER 6. MIXTURE OF STUDENT-T AR COMPONENTS

Figure 6.3: MCMC output of sampled values of parameters while fitting the $\mathcal{M}_\text{AR}(3; 0, 4, 4)$ model to the yDGS3 data, using $N = 100000$ and $B = 50000$. 

(a) trace and density of $\omega$
(b) trace and density of $\gamma$
(c) trace and density of $\mu$
(d) trace and density of $\sigma$
CHAPTER 6. MIXTURE OF STUDENT-T AR COMPONENTS 203

Figure 6.4: MCMC output of sampled values of the parameter $\phi$ while fitting the $\mathcal{M}_{AR}(3; 0, 4; 4, 5, 5)$ model to the $y_{DGS3}$ data with $N = 50000$ and burn-in $= 10000$.
Chapter 7

Mixture of Student-\(t\) ARMA components

7.1 Introduction

We discussed the exact Bayesian analysis of the class of Student-\(t\) mixture Autoregressive models \(\mathcal{M}_{\text{AR}}^t\) in chapter 6 for nonlinear time series analysis. We recall that \(\mathcal{M}_{\text{AR}}^t\) is able to capture serial correlations; time-varying means and volatilities; and the shape of the conditional distributions can be time varied from short-tailed to long-tailed, or from unimodal to multimodal. In this consequence, we are motivated to develop a complete Bayesian analysis for the mixture of autoregressive moving average components model with Student-\(t\) error. We denote this class of model by \(\mathcal{M}_{\text{ARMA}}^t\) and name this new model, “Student-\(t\) typed Mixture of Autoregressive Moving Average Model”. Therefore, the analysis of this new class \(\mathcal{M}_{\text{ARMA}}^t\) model will provide the more general structures including the \(\mathcal{M}_{\text{AR}}^t\) model.

The degrees of freedom is treated as parameter in the \(\mathcal{M}_{\text{ARMA}}^t\) model which allows to adjust the degree of thickness of the conditional distributions according to the actually observed data. Finally, as degrees of freedom in a student \(t\) distribution approaches infinity, the distribution approaches normal. Consequently, the \(\mathcal{M}_{\text{ARMA}}\) model is a limiting case of the \(\mathcal{M}_{\text{ARMA}}^t\) model.
CHAPTER 7. MIXTURE OF STUDENT-T ARMA COMPONENTS

7.2 The $M_{ARMA}^t$ model

Given \( \{y_t\}, \quad (t = 1, \ldots, T) \) be a (observed) time series. Then a \( g \)-components $M_{ARMA}^t$ model can be defined as follows:

\[
\begin{align*}
\begin{cases}
y_t \mid G_{t-1} & \sim \sum_{k=1}^{g} \omega_k \mathcal{S}_t(y_t; \nu_{k,t}, \tau_k^{-1}, v_k), \quad (t = 1, \ldots, T), \\
\nu_{k,t} &= \mu_k + \sum_{i=1}^{p_k} \phi_{ki}(y_{t-i} - \mu_k) + \sum_{j=1}^{q_k} \theta_{kj}\epsilon_{t-j}, \quad (k = 1, \ldots, g), \quad (7.1) \\
\omega_k > 0, \quad (k = 1, \ldots, g); \quad \sum_{k=1}^{g} \omega_k = 1,
\end{cases}
\end{align*}
\]

where “ \( \mid G_{t-1} \)” is used to denote conditioning on the past observations and on all other variables; \( \mathcal{S}_t(x|a, b^{-1}, c) \) denotes the univariate standardised Student-t distribution with mean \( a \), precision \( b \) and degrees of freedom \( c \)(for more details, see Section 2.11.7); \( \nu_{k,t} \) denotes the conditional mean of the \( k \)-component; \( \omega_k \) denotes the mixing weight of the \( k \)-th component.

For the simplicity for model, we can assume \( p = \max\{p_1, \ldots, p_g\} \) and \( q = \max\{q_1, \ldots, q_g\} \) with the settings \( \phi_{kj} = 0 \) if \( j > p_k \) and \( \theta_{kj} = 0 \) if \( j > q_k \), for \( k = 1, \ldots, g \) respectively.

Clearly, the conditional expectation of \( y_t \) given the past information is

\[
E(y_t \mid G_{t-1}) = \sum_{k=1}^{g} \omega_k \nu_{k,t}. \quad (7.2)
\]

This is exactly the same as it is in the case of the $M_{AR}^t$ model. Since the conditional means of the components depend on the past values of the time series and the shape of the conditional distributions may change from unimodal to multi-modal, the conditional expectation of \( y_t \) given by (7.2) may not be the best predictor of the future values.

7.3 Missing data formulation and the likelihood function

The setup of the likelihood function of the $M_{ARMA}^t$ model is the same as that of the likelihood function of the $M_{AR}^t$ model. We repeat the discussion here briefly
CHAPTER 7. MIXTURE OF STUDENT-T ARMA COMPONENTS

for convenience. For details, see Section 6.3.

Since \( y = \{y_t\}_{t=1}^T \) is not observed directly we consider a latent variable \( z = \{z_t\}_{t=1}^T \) which is regarded as “component label” so that:

\[
z_t = \begin{cases} 
1 & \text{with probability } \omega_1 \\
\vdots & \vdots \\
k & \text{with probability } \omega_k \\
\vdots & \vdots \\
g & \text{with probability } \omega_g \\
\end{cases}, \quad (t = 1, \ldots, T). \tag{7.3}
\]

With this approach, the observed series \( \{y_t\}_{t=1}^T \) and the unobserved (residual) series \( \{\varepsilon_t\}_{t=1}^T \) becomes completed with a missing data set \( \{z_t\}_{t=1}^T \), indicating the specific component of the mixture from which every observation is assumed to arise (see Diebolt and Robert (1994)).

In the light of the above characterization of the Student-t distribution, it is convenient to view the observed data augmented by the latent variable \( z_t \) as still being incomplete. Therefore, likewise the \( \mathcal{M}_{AN} \) model we introduce a new missing variable \( \xi_t = \{\xi_t\}_{t=1}^T \) which corresponds to the mixing random variables in the Student-t distribution and control the magnitude of the variance in the normal density function (see McLachlan and Peel (2000)).

Conditionally on \( \xi_t \) which is conditional on \( z_t \), the observations are drawn from their respective individual subpopulation, i.e.

\[
y_t \mid \xi_t, z_t = k, \mathcal{G}_{t-1} \xrightarrow{\text{ind}} \mathcal{N} (y_t; \nu_{k,t}, (\tau_{k}\xi_t)^{-1}) , \tag{7.4}
\]

and

\[
\xi_t \mid z_t = k \xrightarrow{\text{ind}} \mathcal{G} \left( \xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2} \right), \quad v_k > 2. \tag{7.5}
\]

i.e.

\[
y_t \mid \mathcal{G}_{t-1}, z_t \sim \begin{cases} 
\mathcal{N} (y_t; \nu_{1,t}, (\tau_1\xi_t)^{-1}) \mathcal{G} \left( \xi_t; \frac{v_1}{2}, \frac{v_1 - 2}{2} \right) & \text{if } z_t = 1, \\
\vdots & \vdots \\
\mathcal{N} (y_t; \nu_{k,t}, (\tau_k\xi_t)^{-1}) \mathcal{G} \left( \xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2} \right) & \text{if } z_t = k, \\
\vdots & \vdots \\
\mathcal{N} (y_t; \nu_{g,t}, (\tau_g\xi_t)^{-1}) \mathcal{G} \left( \xi_t; \frac{v_g}{2}, \frac{v_g - 2}{2} \right) & \text{if } z_t = g,
\end{cases} \tag{7.6}
\]

provided \( v_k > 2, \quad (k = 1, \ldots, g) \). That is, conditional on the information about the source of \( y_t \), only the corresponding component is relevant.

As a consequence, the full conditional likelihood function separates into \( g \) parts
where each one concerning the data assigned to each of the $g$ mixture components,

$$
\mathcal{L}(\mathbf{\phi} \mid y, z, \xi) = \prod_{k=1}^{g} \prod_{t:z_t=k} N(y_t; \hat{\nu}_{k,t}, (\tau_k \xi_t)^{-1}) \mathcal{G}(\xi_t; \frac{v_k}{2}, \frac{v_k-2}{2}), \quad v > 2. \quad (7.7)
$$

where

$$
\hat{\nu}_{k,t} = \mu_k + \sum_{j=1}^{p_k} \phi_{kj}(y_{t-j} - \mu_k) + \sum_{j=1}^{q_k} \theta_{kj} \hat{\epsilon}_{t-j}, \quad (7.8)
$$

and the residual term $\hat{\epsilon}_t$ is defined as:

$$
\hat{\epsilon}_t = \begin{cases} 
0, & (t \leq 0), \\
y_t - \mu_k - \sum_{j=1}^{p_k} \phi_{kj}(y_{t-j} - \mu_k) - \sum_{j=1}^{q_k} \theta_{kj} \hat{\epsilon}_{t-j}, & (t = 1, \ldots, T), 
\end{cases} \quad (7.9)
$$

### 7.3.1 Hypotheses of the model

Let $G_k(L) = \sum_{j=1}^{p_k} \phi_{kj} L^j$ and $H_k(L) = \sum_{j=1}^{q_k} \theta_{kj} L^j$ for $k = 1, \ldots, g$, where $L$ is the lag operator. We assume that the following hypotheses are true for our $M_t\text{ARMA}$ model (7.1).

To ensure stationarity and invertibility, it is very common practice to impose the following two constraints in the ARMA model (see Nakatsuma (2000)):

(C$_S$): all roots of $1 - G_k(L) = 0, \quad (k = 1, \ldots, g)$ are outside the unit circle.

(C$_I$): all roots of $1 + H_k(L) = 0, \quad (k = 1, \ldots, g)$ are outside the unit circle.

### 7.4 Priors setup

For fixed $\Lambda = (g, p, q, v)$, we will have a specific $M_{\text{ARMA}}$ model. We set all the prior distributions, $\pi(\mathbf{\phi} \mid \Lambda)$, of the model parameters as same as that of the $M_{\text{AR}}$ model except the parameter $\theta$. We state these priors again here for convenience (for details see Section 4.5):
7.4.1 Priors for $\omega$, $\mu$, and $\tau$

\[
\omega \mid g \sim D(\delta_1, \ldots, \delta_g) \tag{7.10}
\]

\[
\mu_k \overset{\text{iid}}{\sim} N(\mu_k; \zeta, \kappa^{-1}), \quad (k = 1, \ldots, g) \tag{7.11}
\]

\[
\tau_k \mid \lambda \overset{\text{iid}}{\sim} G(\tau_k; c, \lambda), \quad (k = 1, \ldots, g)
\]

\[
\lambda \overset{\text{iid}}{\sim} G(\lambda; a, b) \tag{7.12}
\]

7.4.2 Priors for $\phi$ and $\theta$

We specify prior distributions for $\phi = (\phi_1, \ldots, \phi_g)$ and that for $\theta = (\theta_1, \ldots, \theta_g)$ as described for the $\mathcal{M}_{\text{ARMA}}$ model in section 4.5.2. Following Jones (1987) and Barndorff-Nielsen and Schou (1973), we choose the priors for $\phi = (\phi_1, \ldots, \phi_g)$ and $\theta = (\theta_1, \ldots, \theta_g)$:

\[
\phi_k \sim \text{Uniform on } \Xi_p \iff \rho_{ki}^{\phi} \overset{\text{iid}}{\sim} \mathcal{B}(-1, +1) \left( \rho_{ki}^{\phi}; \left\lfloor \frac{i}{2} \right\rfloor + 1 \right), \quad (k = 1, \ldots, g, i = 1, \ldots, q_k) \tag{7.13}
\]

\[
\theta_k \sim \text{Uniform on } \Xi_q \iff \rho_{ki}^{\theta} \overset{\text{iid}}{\sim} \mathcal{B}(-1, +1) \left( \rho_{ki}^{\theta}; \left\lfloor \frac{i}{2} \right\rfloor + 1 \right), \quad (k = 1, \ldots, g, i = 1, \ldots, q_k) \tag{7.14}
\]

where $\rho_{ki}$ be the partial autocorrelation coefficients at lag $i$ for the $k$th component; $\mathcal{B}(-1, +1) (x; a, b)$ represents a generalized beta distribution defined on $(-1, +1)$ and where $[a]$ means the integer part of $a$. 
7.4.3 Priors for $g$, $p$ and $v$

For complete Bayesian analysis of our parameters, we stress that $g, p, q$ and $v$ are stochastic also. The priors for these can be set as:

\[ g \sim U(g; 1, g_{\text{max}}) \] (7.15)

\[ p_k \mid g \sim \text{iid } U(p_k; 0, p_{\text{max}}) , \quad (k = 1, \ldots, g) \] (7.16)

\[ q_k \mid g \sim \text{iid } U(q_k; 0, q_{\text{max}}) , \quad (k = 1, \ldots, g) \] (7.17)

\[ v_k \mid g \sim \text{iid } U(v_k; 3, v_{\text{max}}) , \quad (k = 1, \ldots, g) \] (7.18)

where $U(x \mid a, b)$ denotes the discrete uniform distribution for $x = a, a + 1, \ldots, b$ and $g_{\text{max}}, p_{\text{max}}, q_{\text{max}}$ and $v_{\text{max}}$ is fixed. For details of the above distributions mentioned above see Section 2.11.

7.4.4 Formulae for hyperparameters

In the prior settings, the $M_{\text{ARMA}}$ model has the hyperparameters as that of the $M_{\text{AR}}$ model. We re-written the formulae here (for details, see Section 3.5.1):

\[ \delta = 1 \]

\[ R_y = \max(y) - \min(y) \]

\[ \zeta = \min(y) + \frac{R_y}{2}, \quad \kappa = \frac{1}{R_y} \]

\[ c = 2, \quad a = 0.2, \quad b = \frac{100 a}{c R_y^2}. \] (7.19)

7.5 MCMC moves for parameter estimation

We let $\Lambda = (g, p, q, v)$ be the model index and $\varphi = (\omega, \mu, \tau, \phi)$ be the vector of parameters. We apply MCMC component-wise techniques to calculate the full conditional posterior distribution $\pi(\varphi \mid y, \Lambda)$ in which each of the parameter will update separately. Moreover, we need to update the latent variable $z$ as well. There, we have a set of five moves in total which is discussed below:
7.5.1 Move I: updating \( z \)

This is a Gibbs-type move. From (7.3) and (7.6), the full conditional posterior probability of the location variables \( z_t = k \) (i.e. the observations \( y_t \) has been generated by the \( k \)-th mixture component) is, This is also a Gibbs-type move by which we have the full conditional posterior distribution of \( z \):

\[
\pi(z \mid \varphi_{-\tau}, y) \propto \mathcal{L}(\varphi \mid y, z) \pi(z \mid \omega)
\]

\[
\propto \prod_{t=1}^{n} \sum_{k=1}^{g} \omega_k \mathcal{N}(y_t; \hat{\nu}_{k,t}, (\tau_k \xi_t)^{-1}) \mathcal{G}(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}) \cdot 1_{\{z_t = k\}}
\]

\[
\propto \prod_{t=1}^{n} \sum_{k=1}^{g} \omega_k \mathcal{N}(y_t; \hat{\nu}_{k,t}, (\tau_k \xi_t)^{-1}) \cdot 1_{\{z_t = k\}}
\]

Therefore, we have the full conditional probability that the observation \( y_t \) has been generated by the \( k \)-th mixture component is:
\[
\Pr(z_t = k \mid y_t, \varphi) = \frac{1}{C_4} \omega_k \exp \left[ -\frac{\tau_k \xi_t}{2} (y_t - \hat{\nu}_{k,t})^2 \right],
\]
\[(t = 1, \ldots, T; k = 1, \ldots, g), \quad (7.20)\]

where \( C_4 = \sum_{u=1}^{g} \omega_u \exp \left[ -\frac{\tau_u \xi_t}{2} (y_t - \hat{\nu}_{u,t})^2 \right] \), is the normalizing constant and \( \hat{\nu}_{*,t} \) is defined as in the Equation (7.8).

### 7.5.2 Move II: updating \( \omega \)

This is a Gibbs-type move where the full conditional distribution of \( \omega \) is obtained in similar fashion as in Section 3.6.2:

\[
\omega \mid \varphi_{-\omega}, y, z \sim D(\omega; \delta_1 + n_1, \ldots, \delta_k + n_k),
\]
\[(7.21)\]

where \( n_k = \sum_{t: z_t = k} 1 \) and \( \varphi_{-\omega} \) denotes all the parameters except \( \omega \).

### 7.5.3 Move III: updating \( \mu \)

This is a Gibbs-type move where the full conditional distribution of \( \mu \) can be obtained as below:

\[
\pi(\mu \mid \varphi_{-\mu}, y, z) \propto \mathcal{L}(\varphi, z, \xi \mid y) \pi(\mu)
\]
\[
\propto \prod_{k=1}^{g} \prod_{t: z_t = k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, (\tau_k \xi_t)^{-1}) \mathcal{G} \left( \xi_k; \frac{v_k}{2}, \frac{v_k - 2}{2} \right)
\]
\[
\times \prod_{k=1}^{g} \mathcal{N}(\mu_k; \zeta, \kappa^{-1})
\]
Therefore, the full conditional posterior density for the $k$-th component is:

$$
\pi(\mu_k \mid \varphi_{-\mu_k}, y, z) \propto \prod_{t: z_t = k} \mathcal{N}(y_t; \hat{v}_{k,t}, (\tau_k \xi_t)^{-1}) \cdot \mathcal{N}(\mu_k; \zeta, \kappa^{-1})
$$

$$
\propto \exp \left[ -\frac{\tau_k}{2} \sum_{t: z_t = k} \xi_t \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t-i} - \mu_k) - \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \right\}^2 \right]
$$

Now the term $(##)$ is the same as in the case of $\mathcal{M}_{\text{ARMA}}$ model. The details of simplification of this term are described in Section 6.5.2. We state the simplified result as below:

$$
\sum_{t: z_t = k} \xi_t \left\{ y_t - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t-i} - \mu_k) - \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \right\}^2
$$

$$
= \sum_{t: z_t = k} \xi_t \left\{ y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i} - \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \right\} - \left( 1 - \sum_{i=1}^{p_k} \phi_{k,i} \right) \mu_k \right\}^2
$$

$$
= \sum_{t: z_t = k} \xi_t \left( e_{k,t} - b_k \mu_k \right)^2
$$

where $\bar{e}_k = \frac{1}{n_k} \sum_{t: z_t = k} e_{k,t}$ provided $n_k = \sum_{t: z_t = k} 1$

$$
\sum_{t: z_t = k} \xi_t \left\{ (e_{k,t} - \bar{e}_k)^2 + 2 (e_{k,t} - \bar{e}_k) (\bar{e}_k - b_k \mu_k) + (\bar{e}_k - b_k \mu_k)^2 \right\}
$$

Substituting this into the equation $(##)$ and simplify exactly as described in Section 6.5.2 we have the full conditional posterior distribution of $\mu$: 
\[ \mu_k \mid \varphi_{-\mu_k}, y, z, \xi \sim \mathcal{N}(\mu_k; \frac{\tau_k b_k (\bar{e}_k d_k + c_k) + \kappa \zeta}{\tau_k b_k^2 d_k + \kappa}, \frac{1}{\tau_k b_k^2 d_k + \kappa}), \] 

\text{(k = 1, \ldots, g), (7.22)}

where

- \( \bar{e}_k = \frac{1}{n_k} \sum_{t: z_t = k} e_{k,t} \), where \( e_{k,t} = y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i} - \sum_{j=1}^{q_k} \theta_{k,j} \hat{\varepsilon}_{t-j} \) and \( n_k = \sum_{t: z_t = k} 1 \);

- \( b_k = 1 - \sum_{i=1}^{p_k} \phi_{k,i}; \ c_k = \sum_{t: z_t = k} \xi_t (e_{k,t} - \bar{e}_k); \ d_k = \sum_{t: z_t = k} \xi_t \).

### 7.5.4 Move IV-A: updating \( \phi \)

Here we update \( \rho_k^\phi, \ k = 1, \ldots, g \) by a Metropolis-Hasting (M-H) mechanism. A candidate \( \rho_k^\phi^* \) is generated by a normal density truncated in \((-1, +1)\) and centred in the current state of the chain \( \rho_k^\phi \) so that we can define the proposal density as below:

\[ q_\phi(\rho_k^\phi, \rho_k^\phi^*) = \mathcal{N}_{(-1,+1)}(\rho_k^\phi^* \mid \rho_k^\phi, \tau_{q_\phi}^{-1}), \quad (i = 1, \ldots, p_k; \ k = 1, \ldots, g) \]  

\text{(7.23)}

Using the distribution defined above by (7.23), we generate a vector \( \rho_k^\phi^* = (\rho_{k,1}^\phi^*, \ldots, \rho_{k,p_k}^\phi^*) \) for the partial autocorrelations which will be then used to derive the corresponding parameter vector \( \phi_k^* = (\phi_{k,1}^*, \ldots, \phi_{k,p_k}^*) \) through the transformation.

So, the ratio in the acceptance probability for a M-H algorithm can be defined as

\[ A(\phi_k, \phi_k^*) = \min(1, \Omega_\phi), \]  

\text{(7.24)}

where

\[ \Omega_\phi = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio} \]
Firstly, from (7.7) the likelihood ratio is defined as:

\[
\text{likelihood ratio} = \frac{\prod_{t:z_t=k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, (\tau_k \xi_t)^{-1}) G\left(\frac{v_k}{2}, \frac{v_k - 2}{2}\right)}{\prod_{t:z_t=k} \mathcal{N}(y_t; \hat{\nu}_{k,t}, (\tau_k \xi_t)^{-1}) G\left(\frac{v_k}{2}, \frac{v_k - 2}{2}\right)}
\]

\[
= \exp\left[-\frac{\tau_k}{2} \sum_{t:z_t=k} \xi_t \left\{ \left(y_t - \hat{\nu}_{k,t}^*\right)^2 - \left(y_t - \hat{\nu}_{k,t}\right)^2 \right\}\right] \quad (7.25)
\]

where \(\hat{\nu}_{k,t}^* = \mu_k + \sum_{i=1}^{p_k} \phi_{k,i}^* (y_{t-i} - \mu_k) - \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j}\) as defined in equation (7.8).

Since all the parameters but the \(\phi\) do not change, so from (7.13) the prior ratio in terms of the partial autocorrelations \(\rho^\phi\) reduces to the ratio as below:

\[
\text{prior ratio} = \frac{\prod_{i=1}^{p_k} \mathcal{B}_{(-1,+1)}\left(\rho_{k,i}^*, \left\lfloor \frac{i}{2} \right\rfloor + 1\right)}{\prod_{i=1}^{p_k} \mathcal{B}_{(1,-1)}\left(\rho_{k,i}; \left\lfloor \frac{i}{2} \right\rfloor + 1\right)}
\]

\[
= \prod_{i=2}^{p_k} \frac{(\rho_{k,i}^* + 1)^{\lfloor \frac{i}{2} \rfloor - 1} (1 - \rho_{k,i}^*)^{\left\lfloor \frac{i}{2} \right\rfloor}}{(\rho_{k,i} + 1)^{\lfloor \frac{i}{2} \rfloor - 1} (1 - \rho_{k,i})^{\left\lfloor \frac{i}{2} \right\rfloor}} \quad (7.26)
\]

and, finally from (7.23) we have

\[
\text{proposal ratio} = \frac{\prod_{i=1}^{p_k} \mathcal{N}_{(-1,+1)}\left(\rho_{k,i}; \rho_{k,i}^*, \tau_{q^\phi}^{-1}\right)}{\prod_{i=1}^{p_k} \mathcal{N}_{(-1,+1)}\left(\rho_{k,i}; \rho_{k,i}, \tau_{q^\phi}^{-1}\right)}
\]

\[
= \prod_{i=1}^{p_k} \frac{\Phi_X\left( +1; \rho_{k,i}^*, \tau_{q^\phi}^{-1}\right) - \Phi_X\left( -1; \rho_{k,i}^*, \tau_{q^\phi}^{-1}\right)}{\Phi_X\left( +1; \rho_{k,i}, \tau_{q^\phi}^{-1}\right) - \Phi_X\left( -1; \rho_{k,i}, \tau_{q^\phi}^{-1}\right)} \quad (7.27)
\]

By using (7.25), (7.26) and (7.27) we have
\[ \Omega_\phi = \exp \left[ -\frac{\tau_k}{2} \sum_{t: z_t = k} \xi_t \left\{ (y_t - \hat{v}_{k,t}^*)^2 - (y_t - \hat{v}_{k,t})^2 \right\} \right] \]

\[ \times \prod_{i=2}^{p_k} \left( \rho^\phi_{k,i} + 1 \right)^\left( \frac{q_k^\phi - 1}{2} \right) \left( 1 - \rho^\phi_{k,i} \right)^\left( \frac{p_k - 1}{2} \right) \]

\[ \times \prod_{i=1}^{p_k} \Phi_N \left( +1; \rho^\phi_{k,i}, \tau_{q^\phi}^{-1} \right) - \Phi_N \left( -1; \rho^\phi_{k,i}, \tau_{q^\phi}^{-1} \right) \]

\[ \times \prod_{i=1}^{q_k^\phi} \Phi_N \left( +1; \rho^\phi_{i,k}, \tau_{q^\phi}^{-1} \right) - \Phi_N \left( -1; \rho^\phi_{i,k}, \tau_{q^\phi}^{-1} \right) \]

(7.28)

where \( \hat{v}_{k,t}^* = \mu_k + \sum_{i=1}^{p_k} \phi^*_k (y_{t-i} - \mu_k) - \sum_{j=1}^{q_k} \theta_{k,j} \hat{\epsilon}_{t-j} \) as defined in Equation (7.8) and \( \Phi_N \) denotes the cumulative normal distribution function.

### 7.5.5 Move IV-B: updating \( \theta \)

Here we update \( \rho^\phi_{k}, \ k = 1, \ldots, g \) by a Metropolis-Hasting mechanism in a similar way. A candidate \( \rho^\phi_{k,i}^* \) is generated by a normal density truncated in \((-1, +1)\) and centred in the current state of the chain \( \rho^\phi_{k,i} \) so that we can define the proposal density as below:

\[
q_\theta(\rho^\phi_{k,i}, \rho^\phi_{k,i}^*) = \mathcal{N}_{(-1, +1)} \left( \rho^\phi_{k,i}^*; \rho^\phi_{k,i}, \tau_{q^\phi}^{-1} \right) \quad (i = 1, \ldots, p_k; k = 1, \ldots, g)
\]

(7.29)

Using the distribution defined above by (7.29), we generate a vector \( \rho^\phi_{k}^* = (\rho^\phi_{k,1}^*, \ldots, \rho^\phi_{k,p_k}^*) \) for the partial autocorrelations which will be then used to derive the corresponding parameter vector \( \theta^\phi_{k}^* = (\theta^\phi_{k,1}, \ldots, \theta^\phi_{k,p_k}) \) through the transformation.

The ratio in the acceptance probability for a M-H algorithm can be defined as

\[ A(\theta_k, \theta^\phi_{k}) = \min(1, \Omega_\theta), \]

where
\[ \Omega_\theta = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio} \]
\[ = \exp \left[ -\frac{\tau_k}{2} \sum_{t=1}^{z_t=k} \xi_t \left( (y_t - \hat{\nu}_{k,t}^1)^2 - (y_t - \hat{\nu}_{k,t})^2 \right) \right] \]
\[ \times \prod_{i=2}^{p_k} \left( \rho_{k,i}^\theta + 1 \right)^{\left[ \frac{\pi + 1}{2} \right]} - 1 \left( 1 - \rho_{k,i}^\theta \right)^{\left[ \frac{\pi}{2} \right]} \]
\[ \times \prod_{i=1}^{p_k} \frac{\Phi_N\left( +1; \rho_{k,i}^\theta \tau_{q_i}^{-1} \right) - \Phi_N\left( -1; \rho_{k,i}^\theta \tau_{q_i}^{-1} \right)}{\Phi_N\left( +1; \rho_{k,i}^\theta \tau_{q_i}^{-1} \right) - \Phi_N\left( -1; \rho_{k,i}^\theta \tau_{q_i}^{-1} \right)} \]
\[ \times \prod_{i=1}^{q_k} G\left( \xi_t; \nu_{k,i}, \left( \tau_{k,i} \right)^{-1} \right) \]
\[ (7.30) \]

where \( \hat{\nu}_{k,t}^1 = \mu_k + \sum_{i=1}^{p_k} \phi_{k,i} (y_{t-i} - \mu_k) - \sum_{j=1}^{q_k} \theta_{k,j}^1 \xi_{t-j} \) as defined in Equation (7.8).

### 7.5.6 Move V: updating \( \lambda \) and \( \tau \)

Let \( \varphi_{-\lambda} \) denote the remaining parameters except \( \lambda \) and \( \varphi_{-\tau} \) denotes the remaining parameters except \( \tau \). The full conditional posterior distribution of \( \lambda \) and \( \tau \) can be obtained by applying Gibbs-typed move in the similar way as described for the \( \mathcal{M}_{\text{AR}} \) model in Section 3.6.5. We state the results here only.

Firstly, the full conditional posterior distribution of \( \lambda \) is:

\[ \lambda \mid \varphi_{-\lambda}, y, z \sim \mathcal{G}\left( \lambda; a + gc, b + \sum_{k=1}^{g} \tau_k \right), \quad (k = 1, \ldots, g) \quad (7.31) \]

Secondly, the full conditional posterior distribution of \( \tau \) can be obtained as below:

\[ \pi(\tau \mid \varphi_{-\tau}, y, z, \xi) \propto \mathcal{L}(\varphi \mid y) \pi(\tau \mid \lambda) \]
\[ \propto \prod_{k=1}^{g} \prod_{t,z_t=k} N\left( y_t; \hat{\nu}_{k,t}, \left( \tau_k \xi_t \right)^{-1} \right) \mathcal{G}\left( \xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2} \right) \prod_{k=1}^{g} \mathcal{G}\left( \tau_k; c, \lambda \right) \]
Consequently, we have the full conditional posterior density for the \( k \)-th component:

\[
\pi(\tau_k \mid \varphi_{-\tau_k}, y, z, \xi) \propto \prod_{t : z_t = k} N(y_t; \hat{\nu}_{k,t}, (\tau_k \xi_t)^{-1}) G(\xi_t; \frac{\nu_k}{2}, \frac{\nu_k - 2}{2}) \cdot G(\tau_k; c, \lambda)
\]

\[
\propto (\tau_k \xi_t)^{nk/2} \exp \left[ -\frac{\tau_k}{2} \sum_{t : z_t = k} \xi_t (y_t - \hat{\nu}_{k,t})^2 \right] \cdot \tau_k^{c-1} \exp \left[ -\lambda \tau_k \right]
\]

\[
\propto (\tau_k)^{c+nk/2-1} \exp \left[ -\left( \lambda + \frac{1}{2} \sum_{t : z_t = k} \xi_t (y_t - \hat{\nu}_{k,t})^2 \right) \tau_k \right]
\]

Hence, the full conditional posterior distribution of \( \tau \) is:

\[
\tau_k \mid \varphi_{-\tau_k}, y, z, \xi \sim G(\tau_k; c + \frac{nk}{2}, \lambda + \frac{1}{2} \sum_{t : z_t = k} \xi_t (y_t - \hat{\nu}_{k,t})^2), \quad (k = 1, \ldots, g),
\]

\[(7.32)\]

where \( \hat{\nu}_{k,t} \) is as defined in Equation (7.8).

### 7.5.7 Move VI: updating \( z_t \)

This is also a similar Gibbs-typed move as described in Section 6.5.5. To avoid repetition we only state the full conditional posterior probability \( z_t = k \mid y, \varphi \) as below:

\[
\Pr(z_t = k \mid y, \varphi) = \frac{1}{C_4} \omega_k \sqrt{\tau_k} \exp \left[ -\frac{\tau_k \xi_t}{2} (y_t - \hat{\nu}_{k,t})^2 \right], \quad (t = 1, \ldots, T; k = 1, \ldots, g),
\]

\[(7.33)\]

where \( C_4 = \sum_{u=1}^g \omega_u \sqrt{\tau_u} \exp \left[ -\frac{\tau_u \xi_t}{2} (y_t - \hat{\nu}_{u,t})^2 \right] \) is the normalizing constant and \( \hat{\nu}_{u,t} \) is defined as in the equation (7.8).
7.5.8 Move VII: updating $\xi$

This is also a Gibbs-typed move as discussed in Section 6.5.6. Here we only state the full conditional posterior distribution of $\xi$ which is:

$$
\xi_t \mid \varphi_{-\xi_t}, y, z_t = k \sim \mathcal{G}\left(\frac{v_k + 1}{2}, \frac{\tau_k}{2} (y_t - \hat{\nu}_{k,t})^2 + \frac{v_k - 2}{2}\right),
$$

\hspace{1cm} (t = 1, \ldots, T; k = 1, \ldots, g),

(7.34)

where $\hat{\nu}_{k,t}$ is defined as in the Equation (7.8).

7.6 MCMC moves for model determination

We determine the model index in Bayesian context by applying the techniques described by Sampietro (2006). Model determination for the $\mathcal{M}_{\text{ARMA}}^t$ model is the same as that for the $\mathcal{M}_{\text{ARMA}}$ model. We recall all techniques here in short. For details, see Section 3.7. To get full Bayesian flavour we must get the posterior distribution of the model index $\Lambda = (g, p, q, v)$ which can be written and factorized as:

$$
\pi(\Lambda \mid y) = \pi(p \mid g, y) \times \pi(q \mid g, y) \times \pi(v \mid g, y) \times \pi(g \mid y)
$$

(7.35)

So, it is enough to update $p$, $q$, $v$, and $g$ separately to update the model index $\Lambda$. We develop two additional MCMC moves to complete our Bayesian analysis as follows.

7.6.1 Move VIII-A: updating of $p$

In a similar fashion, we can update the parameter $p$ as discussed for the $\mathcal{M}_{\text{ARMA}}$ model in Section 4.7.1.

At first we need to choose a component, say $k^\dagger$, randomly chosen in $\{1, \ldots, g\}$. Then we can write the proposal for the new AR order of the $k^\dagger$-th component, $p_{k^\dagger}$, as:

$$
p_{k^\dagger} = \begin{cases} 
p_{k^\dagger} - 1, & \text{with probability } d(p_{k^\dagger}), \\
p_{k^\dagger} + 1, & \text{with probability } b(p_{k^\dagger}),
\end{cases}
$$

(7.36)

where $b(p_k) = 1 - d(p_k)$, for $k = 1, \ldots, g$, $d(0) = 0$ and $b(p_{\text{max}}) = 0$. 


Let $\rho^\phi_{k_1}$ be the proposal vector of the partial autocorrelations. When $p^*_{k_1} = p_{k_1} - 1$, the autoregressive coefficients $\rho^\phi_{k_1} = (\rho^\phi_{k_1 1}, \ldots, \rho^\phi_{k_1 p^*_{k_1}})$ is updated by simply discarding the last partial autocorrelation in $\rho^\phi_{k_1} = (\rho^\phi_{k_1 1}, \ldots, \rho^\phi_{k_1 p_{k_1}})$. On the other hand, when $p^*_{k_1} = p_{k_1} + 1$, the corresponding autocorrelations $\rho^\phi_{k_1} = (\rho^\phi_{k_1 1}, \ldots, \rho^\phi_{k_1 p_{k_1}})$ is updated by adding new parameter $\rho^\phi_{k_1 p^*_{k_1}}$ in $\rho^\phi_{k_1} = (\rho^\phi_{k_1 1}, \ldots, \rho^\phi_{k_1 p_{k_1}})$ with

$$
\rho^\phi_{k_1 p^*_{k_1}} \sim B(-1, +1) \left( \rho^\phi_{k_1 1}, \ldots, \rho^\phi_{k_1 p_{k_1}} \right).
$$

Then update the corresponding AR coefficients parameters by using the reparametrization discussed in Section 4.5.

To update the AR order of the model by using RJMCMC method, we have the acceptance probability

$$
A_{p^\pm} = \min \{ 1, \Omega_{p^\pm} \},
$$

and $\Omega_{p^\pm}$ can be obtained as (for details see section ??):

$$
\Omega_{p^\pm} = \begin{cases} 
\exp \left[-\frac{\tau}{2} \sum_{t: z_t = k} \left\{ (y_t - \hat{\nu}_{k_1, t})^2 - (y_t - \hat{\nu}_{k_1, t})^2 \right\} \right] \frac{b(p^\dagger_{k_1})}{d(p^\dagger_{k_1})} & \text{if } p^\dagger_{k_1} = p_{k_1} - 1, \\
\exp \left[-\frac{\tau}{2} \sum_{t: z_t = k} \left\{ (y_t - \hat{\nu}_{k_1, t})^2 - (y_t - \hat{\nu}_{k_1, t})^2 \right\} \right] \frac{d(p^\dagger_{k_1})}{b(p^\dagger_{k_1})} & \text{if } p^\dagger_{k_1} = p_{k_1} + 1,
\end{cases}
$$

(7.37)

where $\hat{\nu}_{k_1, t} = \mu_{k_1} + \sum_{j=1}^{p_{k_1}} \phi^\dagger_{k_1, j} (y_{t-j} - \mu_{k_1}) - \sum_{j=1}^{q_{k_1}} \theta^\dagger_{k_1, j} \hat{\epsilon}_{k_1, t-j}$, and $\nu_{k_1, t}$ is as defined in Equation (7.8).

Note that the Equation (7.37) has simple form and $\pi(p \mid g, y)$ is simply estimated by the proportions of every possible value for $p$ in the sample obtained by the previous complete MCMC algorithm.

### 7.6.2 Move VIII-B: updating of q

This is almost the same as discussed in Section 8.6.1. For convention, we present that discussion very shortly here.

Likewise the previous move, at first we need to choose a component randomly,
say $k^\dagger$, in $\{1, \ldots, g\}$. Then we can write the proposal for the new MA order of the $k^\dagger$-th component, $q_{k^\dagger}$, as:

$$q_{k^\dagger} = \begin{cases} 
q_{k^\dagger} - 1, & \text{with probability } d(q_{k^\dagger}), \\
q_{k^\dagger} + 1, & \text{with probability } b(q_{k^\dagger}), 
\end{cases}$$

(7.38)

where $b(q_k) = 1 - d(q_k)$, for $k = 1, \ldots, g$, $d(0) = 0$ and $b(q_{\max}) = 0$.

Let $\rho_{k^\dagger}^\theta$ be the proposal vector of the partial autocorrelations. When $q_{k^\dagger} = q_{k^\dagger} - 1$, the autoregressive coefficients $\rho_{k^\dagger}^\theta = (\rho_{k^\dagger 1}^\theta, \ldots, \rho_{k^\dagger q_{k^\dagger}^\theta}^\theta)$ is updated by simply discarding the last partial autocorrelation in $\rho_{k^\dagger}^\theta = (\rho_{k^\dagger 1}^\theta, \ldots, \rho_{k^\dagger q_{k^\dagger}^\theta}^\theta)$. On the other hand, when $q_{k^\dagger} = q_{k^\dagger} + 1$, the corresponding autocorrelations $\rho_{k^\dagger}^\theta = (\rho_{k^\dagger 1}^\theta, \ldots, \rho_{k^\dagger 1}^\theta, \rho_{k^\dagger q_{k^\dagger}^\theta}^\theta)$ is updated by adding new parameter $\rho_{k^\dagger q_{k^\dagger}^\theta}^\theta$ in $\rho_{k^\dagger}^\theta = (\rho_{k^\dagger 1}^\theta, \ldots, \rho_{k^\dagger q_{k^\dagger}^\theta}^\theta)$ with

$$\rho_{k^\dagger q_{k^\dagger}^\theta}^\theta \sim \mathcal{B}_{-1,+1}(\rho_{k^\dagger q_{k^\dagger}^\theta}^\theta, [q_{k^\dagger}^\theta + 1, 2], [q_{k^\dagger}^\theta + 1/2, 1]).$$

Then update the corresponding MA coefficients parameters by using the reparametrization discussed in Section 4.5.

Then update all MA parameters by using the reparametrization discussed in Section 4.5.

To update the MA order of the model by using RJMCMC method, we have the acceptance probability

$$A_{q^\dagger} = \min\{1, \Omega_{q^\dagger}\},$$

and $\Omega_{q^\dagger}$ can be obtained as in Section 6.6.1:

$$\Omega_{q^\dagger} = \begin{cases} 
\exp\left[-\frac{7}{2} \sum_{t: z_t = k} \left\{(y_t - \hat{\nu}_{k^\dagger t, t})^2 - (y_t - \hat{\nu}_{k^\dagger, t})^2\right\}\right] \frac{b(q_{k^\dagger})}{d(q_{k^\dagger})} & \text{if } q_{k^\dagger} = q_{k^\dagger} - 1, \\
\exp\left[-\frac{7}{2} \sum_{t: z_t = k} \left\{(y_t - \hat{\nu}_{k^\dagger t, t})^2 - (y_t - \hat{\nu}_{k^\dagger, t})^2\right\}\right] \frac{d(q_{k^\dagger})}{b(q_{k^\dagger})} & \text{if } q_{k^\dagger} = q_{k^\dagger} + 1, 
\end{cases}$$

(7.39)

where $\hat{\nu}_{k^\dagger, t} = \mu_{k^\dagger} + \sum_{j=1}^{p_{k^\dagger}} \phi_{k^\dagger j} (y_{t-j} - \mu_{k^\dagger}) - \sum_{j=1}^{q_{k^\dagger}} \theta_{k^\dagger j} \hat{\epsilon}_{k^\dagger, t-j}$ and $\nu_{k^\dagger, t}$ is as defined in Equation (7.8).
Similarly, the posterior density \( \pi(q \mid g, y) \) is simply estimated by the proportions of every possible value for \( q \) in the sample obtained by the previous complete MCMC algorithm.

### 7.6.3 Move VIII-C: updating of \( v \)

We will apply the same technique to update the parameter \( v \) as discussed in Section 3.7.1 and we got a similar result as mentioned in Section 6.6.2 for the \( \mathcal{M}_{AR} \) model. So, we will discuss this move briefly here.

At first we need to choose a component, say \( k^* \), randomly chosen in \( \{1, \ldots, g\} \). Then the corresponding degrees of freedom \( v_{k^*} \) increases or decreases by one with the following rule:

\[
v_{k^*} = \begin{cases} v_{k^*} - 1, & \text{with probability } d(v_{k^*}), \\ v_{k^*} + 1, & \text{with probability } b(v_{k^*}), \end{cases}
\]

where \( b(x) = 1 - d(x) \), for \( k = 1, \ldots, g, \ d(3) = 0 \) and \( b(v_{\max}) = 0 \) because \( v_k \in [3, v_{\max}] \). The function \( b(x) \) is defined in Section 3.7.1.

Let \( \xi_t^* \) be the proposal vector conditional to \( z = k^* \). When \( v_{k^*} = v_{k^*} - 1 \), then the corresponding missing variable \( \xi_t^* \mid z_t = k^* \sim \left( \xi_t; \frac{v_{k^*} - 1}{2}, \frac{v_{k^*} - 3}{2} \right) \) and when \( v_{k^*} = v_{k^*} + 1 \), then \( \xi_t^* \mid z_t = k^* \sim \left( \xi_t; \frac{v_{k^*} + 1}{2}, \frac{v_{k^*} - 1}{2} \right) \).

So, the acceptance probability is

\[
A_{v^\pm} = \min\{1, \Omega_{v^\pm}\},
\]

and \( \Omega_{v^\pm} \) is described as in Section 2.7.1:

\[
\Omega_{v^\pm} = \text{likelihood ratio} \times \text{prio ratio} \times \text{proposal ratio} \times \text{jacobian}.
\]

The likelihood ratio is the same as of Equation (3.21):

\[
\text{likelihood ratio} = \frac{\prod_{t: z_t = k^*} \mathcal{N}(y_t; \hat{v}_{k^*, t}, (\tau_{k^*} \xi_t^*)^{-1}) g\left(\xi_t; \frac{v_{k^*}}{2}, \frac{v_{k^*} - 2}{2}\right)}{\prod_{t: z_t = k^*} \mathcal{N}(y_t; \hat{v}_{k^*, t}, (\tau_{k^*} \xi_t)^{-1}) g\left(\xi_t; \frac{v_{k^*}}{2}, \frac{v_{k^*} - 2}{2}\right)}
\]

\[
= C \exp\left[-\frac{n_{k^*}}{2} \left\{ (v_{k^*}^* - 2) \xi_t^* - (v_{k^*}^* - 2) \xi_t \right\}\right],
\]
where $$C = \left( \frac{\xi}{\xi_t} \right)^{\frac{n_k \cdot}{2}} \times \left( \frac{\Gamma(\frac{n_k \cdot}{2})}{\Gamma(\frac{2}{2})} \right)^{\frac{n_k \cdot \cdot - 2}{2}} \times \frac{\Xi_t^2}{\xi_t} \times \frac{\Gamma(\frac{2}{2})}{\Gamma(\frac{n_k \cdot}{2})} \times \frac{\Xi_t^2}{\xi_t} \times \frac{\Gamma(\frac{2}{2})}{\Gamma(\frac{n_k \cdot}{2} - 1)}_{n_k \cdot} .$$

The prior ratio is given by:

$$\text{prior ratio} = \frac{\pi(v^*_k \cdot) \cdot \pi(\xi^*_t | z_t = k^*)}{\pi(v_{k^*} \cdot) \cdot \pi(\xi_t | z_t = k^*)} = \frac{\pi(\xi^*_t | z_t = k^*)}{\pi(\xi_t | z_t = k^*)}$$

because of the uniform prior on the AR order of the model and only the $$k^*$$-th component is updated.

To calculate the proposal ratio we have to take into account the probabilities $$b(\cdot)$$ and $$d(\cdot)$$ and the proposal density, which is chosen equal to the prior on $$\xi$$.

$$\text{proposal ratio} = \begin{cases} \frac{b(v^*_k \cdot) \cdot \pi(\xi_t | z_t = k^*)}{d(v^*_k \cdot)} & \text{if } v^*_k \cdot = v_{k^*} - 1, \\ \frac{d(v^*_k \cdot)}{b(v_{k^*} \cdot) \cdot \pi(\xi_t | z_t = k^*)} & \text{if } v^*_k \cdot = v_{k^*} + 1. \end{cases}$$

Finally, the jacobian is equal to 1 because the invertible function $$g(\cdot)$$ is considered as the identity function.

Therefore, multiplying likelihood, prior and proposal ratios, we derive ratio in the acceptance probability for the both the cases:

$$\Omega_{p^*} = \begin{cases} D \exp \left[ - \frac{n_k \cdot}{2} \left\{ (v^*_k - 2) \xi^*_t - (v^*_k - 2)\xi_t \right\} \right] \frac{b(v^*_k \cdot)}{d(v^*_k \cdot)} & \text{if } v^*_k \cdot = v_{k^*} - 1, \\ E \exp \left[ - \frac{n_k \cdot}{2} \left\{ (v^*_k - 2) \xi^*_t - (v^*_k - 2)\xi_t \right\} \right] \frac{d(v^*_k \cdot)}{b(v_{k^*} \cdot)} & \text{if } v^*_k \cdot = v_{k^*} + 1, \end{cases}$$

where $$D = C \prod_{t: z_t = k^*} \mathcal{G}(\xi_t; \frac{v^*_k - 2}{2}, \frac{v^*_k - 2}{2})$$ and $$E = C / \prod_{t: z_t = k^*} \mathcal{G}(\xi_t; \frac{v_{k^*} - 2}{2}, \frac{v_{k^*} - 2}{2})$$.

Therefore, $$\pi(p^* | g, y)$$ is simply estimated by the proportions of every possible value for $$p$$ in the sample obtained by the previous complete MCMC algorithm.

### 7.6.4 Move IX: updating of $$g$$

Finally, we are at the move by which we can complete our Bayesian analysis. According to Bayes' theorem, the marginal posterior distribution of $$g$$ is:

$$\pi(g | y) \propto f(y | g) \pi(g), \quad (7.41)$$
where \( \pi(g) \) is the prior on \( g \) and \( f(y \mid g) \) is marginal likelihood given by

\[
f(y \mid g) = \sum_p \sum_q \int f(y \mid \varphi, p, q, v) \pi(\varphi, p, q, v \mid g) d\varphi
\]

(7.42)

Now we are going to apply the method proposed by Chib and Jeliazkov (2001). First, applying the basic marginal likelihood identity and suppress the model index \( g \) for notational convenience, we have

\[
f(y) = \frac{f(y \mid \varphi^*, p^*, q^*, v^*) \pi(\varphi^*, p^*, q^*, v^*)}{\pi(\varphi^* \mid p^*, q^*, v^*, y) \pi(p^* \mid y) \pi(q^* \mid y) \pi(v^* \mid y)}
\]

(7.43)

for any fixed point \( (\varphi^*, p^*, q^*, v^*) \) which have the maximum density.

We apply the same methodology and the discussion would be the same as it was the case for \( \mathcal{M}_{\text{ARMA}} \) model. We just need consider another condition parameter \( v \) here and estimations will be followed by the corresponding equations for this model. For details, see Section 4.7.3.

### 7.7 Examples

First we implement our Bayesian procedure of the \( \mathcal{M}_{\text{ARMA}} \) model, discussed in Section 7.5 for the simulated series \( y_{\text{MoI}} \). Secondly, we illustrate it with the data \( y_{\text{DGS3}} \). The details of these series are available in Section 3.8.

For each series, we generated \( N = 50000 \) runs of three Markov chains with different initial values that are generated automatically and discarded the initial 10000 runs as burn-in. We also assessed the convergence of the Markov chains by compiled several diagnostic tests such as Geweke diagnostic test (not reported here, see Appendix) and the chains have converged in all cases.

Table 7.1 contains parameter estimation results of sampled values for the simulated series \( y_{\text{MoI}} \). It shows mean, median and HPD intervals for each of the parameter by which we can argue the better fittings of the series \( y_{\text{MoI}} \) with \( \mathcal{M}_{\text{ARMA}}(2; 1, 1; 1, 0; 4, 8) \) model than that with the \( \mathcal{M}_{\text{AR}}(2; 1, 1; 4, 8) \) model.

Figure 7.2 shows the convergence diagrams and marginal densities of the posterior sampled values of \( \varphi \) for the simulated series \( y_{\text{MoI}} \). Observe that the asymmetry of the marginal distributions are well captured by our MCMC technique.

On the other hand, we present graphical analysis of the MCMC output of each
Table 7.1: Estimation results of sampled values of parameters while fitting the $\mathcal{M}_{\text{AR}}(2; 1, 1; 1, 0; 4, 8)$ model to the simulated series $y_{\text{MoI}}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Posterior statistics</th>
<th>Mean</th>
<th>s.d.</th>
<th>Median</th>
<th>HPD†</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.5</td>
<td>Mean</td>
<td>0.551</td>
<td>0.03259</td>
<td>0.5475</td>
<td>[0.5000000, 0.6093632]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.</td>
<td>(0.0001031)†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.5</td>
<td>Mean</td>
<td>0.449</td>
<td>0.03259</td>
<td>0.4525</td>
<td>[0.3906368, 0.5000000]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.</td>
<td>(0.0001031)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.0</td>
<td>Mean</td>
<td>-0.14431</td>
<td>0.1934</td>
<td>-0.13950</td>
<td>[-0.5403449, 0.2223246]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.</td>
<td>(0.0006116)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.0</td>
<td>Mean</td>
<td>-0.07919</td>
<td>0.1751</td>
<td>-0.04798</td>
<td>[-0.4709479, 0.2487256]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.</td>
<td>(0.005538)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1</td>
<td>Mean</td>
<td>1.435</td>
<td>0.3979</td>
<td>1.318</td>
<td>[1.280780, 1.904066]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.</td>
<td>(0.001258)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>2</td>
<td>Mean</td>
<td>2.345</td>
<td>0.5083</td>
<td>2.404</td>
<td>[1.752201, 2.691374]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.</td>
<td>(0.001607)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.5</td>
<td>Mean</td>
<td>-0.3503</td>
<td>0.4686</td>
<td>-0.5097</td>
<td>[-0.5607706, 0.9999902]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.</td>
<td>(0.001482)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>1.1</td>
<td>Mean</td>
<td>0.8211</td>
<td>0.4704</td>
<td>0.9898</td>
<td>[-0.5238992, 0.9999902]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.</td>
<td>(0.001488)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Highest Posterior Density interval

The parameter without and with label-switched situations. Figure 7.3 shows the convergence diagrams and marginal posterior densities of $\omega$, $\mu$, and $\sigma$ while Figure 7.4 shows that of $\phi$ and finally, Figure 7.5 shows that of $\theta$ for the $y_{\text{DGS3}}$ series without label-switched and with label-switched cases. It is clear from the diagrams that the label switching problem happened at the beginning of each chain and solved easily by imposing the identifiability constraint $\omega_1 < \omega_2 < \omega_3$. Table of parameter estimation results are obtained but not reported here to save space.

At the end, we can argue that our Bayesian algorithm fits not only with the simulated series but also with the real data.
Figure 7.2: MCMC output of sampled values of the parameters while fitting the $\mathcal{M}_\text{ARMA}(2; 1, 1; 1, 0; 4, 8)$ model to the simulated series $y_{\text{MoI}}$. 
CHAPTER 7. MIXTURE OF STUDENT-T ARMA COMPONENTS

226

Figure 7.3: MCMC output of sampled values of mixing proportions, means and scale while fitting the $\mathcal{M}_{\text{ARMA}}(3; 0, 2, 2; 1, 1, 1; 5, 8, 10)$ model to the yDGS3 data without label-switching and when label-switched side by side, using simulation $N = 50000$ and burn-n = 10000; graphs show some improvement after we dealt with the label-switching.
CHAPTER 7. MIXTURE OF STUDENT-T ARMA COMPONENTS

Figure 7.4: MCMC output of sampled values of AR coefficients while fitting the $\mathcal{M}_{\text{ARMA}}^{t}(3; 0, 2, 2; 1, 1, 1; 5, 8, 10)$ model to the $yDGS3$ data without label-switching and when label-switched, using simulation $N = 50000$ and burn-n = 10000; Clearly graphs show a negative impact of label-switching problem for $\phi$. 

(a) trace of $\phi$ without label-switched
(b) density of $\phi$ without label-switched
(c) trace and density of $\phi$ with label-switched
(d) density of $\phi$ with label-switched
Figure 7.5: MCMC output of sampled values of MA coefficients while fitting the $M_{\text{ARMA}}(3; 0, 2, 2; 1, 1, 1; 5, 8, 10)$ model to the $y_{\text{DGS3}}$ data without label-switching and when label-switched, using simulation $N = 50000$ and burn-n = 10000; Clearly graphs show the impact of label-switching problem.
Chapter 8

Mixture of Student- $t$ ARMA-GARCH components

8.1 Introduction

In this chapter, we consider a model named, “Mixture of Student- $t$ ARMA-GARCH components”, where the conditional mean of the process variable follows a mixture ARMA process, and the conditional variance of the that follows a mixture GARCH process with Student- $t$ errors. We denote this class of model by the notation $M^t_{\text{ARMA-GARCH}}$. Clearly, this class of model is different from the class of $M_{\text{ARMA-GARCH}}$ by its error term. There are several strong points in support of choosing the Student- $t$ distribution (see Wong et al. (2009)). First of all, it provides heavier tails than normal distribution and would be able to accommodate some aberrant returns occasionally observed in financial markets. Secondly, the precision of Student- $t$ distribution depends on its degrees of freedom which will interfere the estimation of the component precision. Finally, as degrees of freedom in a Student- $t$ distribution approaches infinity, the distribution approaches normal. Consequently, the $M_{\text{ARMA-GARCH}}$ model is a limiting case of the $M^t_{\text{ARMA-GARCH}}$ model.

Since only difference between $M_{\text{ARMA-GARCH}}$ and $M^t_{\text{ARMA-GARCH}}$ models is the error term, we have similarity in the posterior derivation of some of the parameters. Therefore, we avoid the full derivations in such cases.
8.2 The $\mathcal{M}_t^{\text{ARMA-GARCH}}$ model

Let $y = \{y_t\}_{t=1}^T$ be an observed time series. Then we define $g$-component $\mathcal{M}_{\text{ARMA-GARCH}}$ model as below:

\[
\begin{aligned}
y_t \mid \mathcal{G}_{t-1} &\sim \sum_{k=1}^g \omega_k \mathcal{S}_t (y_t; \nu_{k,t}, \tau_{k,t}^{-1}, v_k), \quad (t = 1, \ldots, T), \\
v_{k,t} &= \mu_k + \sum_{j=1}^{p_k} \phi_{kj}(y_{t-j} - \mu_k) + \sum_{j=1}^{q_k} \theta_{kj}\varepsilon_{t-j}, \quad (k = 1, \ldots, g), \\
\tau_{k,t}^{-1} &= \alpha_k + \sum_{j=1}^{r_k} \alpha_{kj}\varepsilon_{t-j}^2 + \sum_{j=1}^{s_k} \beta_{kj}\tau_{k,t-j}^{-1}, \quad (k = 1, \ldots, g), \\
\omega_k &> 0, \quad (k = 1, \ldots, g); \quad \sum_{k=1}^g \omega_k = 1,
\end{aligned}
\]  

(8.1)

where “$\mid \mathcal{G}_{t-1}$” is used to denote conditioning on the past observations and on all other variables; $\mathcal{S}_t(x; a, b^{-1}, c)$ denotes the univariate standardised student-$t$ distribution with mean $a$, precision $b$ and degrees of freedom $c$ (for more details, see Section 2.11.7); $\nu_{k,t}$ denotes the conditional mean of the $k$-component; $\omega_k$ the $k$-th mixing weight of the mixture; $\tau_{k,t}^{-1}$ is the conditional precision of the $k$-th component; $\phi_{kj}$’s are the AR coefficients of the $k$-th component; $\theta_{kj}$’s are the MA coefficients of the $k$-th component; $\alpha_{kj}$ and $\beta_{kj}$’s are the GARCH coefficients of the $k$-th component. We denote this model by the notation $\mathcal{M}_t^{\text{ARMA-GARCH}} (g; p_1, \ldots, p_g, q_1, \ldots, q_g; r_1, \ldots, r_g, s_1, \ldots, s_g; v_1, \ldots, v_g)$.

The $\mathcal{M}_t^{\text{ARMA-GARCH}}$ model given by (8.1) can be treated as a generalized version of the models so far discussed in this thesis. i.e. $\mathcal{M}_t^{\text{AR}}$, $\mathcal{M}_t^{\text{AR}}$, $\mathcal{M}_t^{\text{ARMA}}$, $\mathcal{M}_t^{\text{ARMA}}$ and $\mathcal{M}_t^{\text{ARMA-GARCH}}$ models.

For the simplicity for model, we can assume $p = \max\{p_1, \ldots, p_g\}$, $q = \max\{q_1, \ldots, q_g\}$, $r = \max\{r_1, \ldots, r_g\}$, and $s = \max\{s_1, \ldots, s_g\}$ with the settings $\phi_{kj} = 0$ if $j > p_k$, $\theta_{kj} = 0$ if $j > q_k$, $\alpha_{kj} = 0$ if $j > r_k$, and $\beta_{kj} = 0$ if $j > s_k$ for $k = 1, \ldots, g$. In addition, we set also $v = \max\{v_1, \ldots, v_g\}$.

The conditional expectation of $y_t$ given the past information is the same. i.e.

\[
E(y_t \mid \mathcal{G}_{t-1}) = \sum_{k=1}^g \omega_k \nu_{k,t}.
\]

(8.2)

For notational convention of our model, we set $\omega = (\omega_1, \ldots, \omega_g)$ denotes the
mixing weight vector of the model; \( \mu = (\mu_1, \ldots, \mu_g) \), denotes the component’s mean vector of the model; \( \phi_k = (\phi_{k,1}, \ldots, \phi_{k,p_k}) \), \( k = 1, \ldots, g \), denotes the AR coefficients vector of \( k \)-th component; \( \theta_k = (\theta_{k,1}, \ldots, \theta_{k,q_k}) \), \( k = 1, \ldots, g \) denotes the MA coefficients vector of \( k \)-th component; \( \alpha_k = (\alpha_{k,1}, \ldots, \alpha_{k,q_k}) \) and \( \beta_k = (\beta_{k,1}, \ldots, \beta_{k,q_k}) \) for \( k = 1, \ldots, g \), denotes the GARCH coefficients. Finally, we can set \( \phi = (\omega, \mu, \phi, \theta, \alpha, \beta) \), be the parameter vector of model (5.2).

We also denote \( p = (p_1, \ldots, p_g) \), as the vector of orders of the AR part of the model, \( q = (q_1, \ldots, q_g) \), as the vector of orders of the MA part of the model, and \( r = (r_1, \ldots, r_g) \), \( s = (s_1, \ldots, s_g) \) as the vector of order of the GARCH part of the model and \( \nu = (v_1, \ldots, v_g) \). Therefore, we set \( \Lambda = (g, p, q, r, s) \) as the model index.

### 8.2.1 Hypotheses for the model

Let \( G_k(L) = \sum_{j=1}^{p_k} \phi_{kj} L^j \), \( H_k(L) = \sum_{j=1}^{q_k} \theta_{kj} L^j \), \( A_k(L) = \sum_{j=1}^{r_k} \alpha_{kj} L^j \) and \( B_k(L) = \sum_{j=1}^{s_k} \beta_{kj} L^j \), for \( k = 1, \ldots, g \), where \( L \) is the lag operator. We assume that the following hypotheses are true for our \( \mathcal{M}_{\text{ARMA-GARCH}} \) model (5.2):

\((C_S)\): All roots of each of the characteristic equations

\[
1 - G_k(L) = 0, \quad (k = 1, \ldots, g)
\]

must lie outside the unit circle.

\((C_T)\): All roots of each of the characteristic equations,

\[
1 + H_k(L) = 0, \quad (k = 1, \ldots, g)
\]

must lie outside the unit circle.

\((C_\alpha)\): \( \alpha_{ki} > 0, \quad (k = 1, \ldots, g, i = 0, 1, \ldots, r_k) \)

\((C_\beta)\): \( \beta_{ki} > 0, \quad (k = 1, \ldots, g, i = 1, \ldots, s_k) \)

\((C_{\varepsilon})\): \( \sum_j^{r_k} \alpha_{kj} + \sum_j^{s_k} \beta_{kj} < 1, \quad (k = 1, \ldots, g) \).

The reasons for implementing the above hypotheses are very common in the literature. Note that \((C_S)\) ensures the stationarity and \((C_T)\) assures invertibility in the ARMA part of the model; \((C_\alpha)\) and \((C_\beta)\) ensure that the conditional
8.3 Missing data formulation and the likelihood function

It is evident that the likelihood function of the $M_{t \text{ARMA-GARCH}}$ takes a very complicated form. But it can be simplified by introducing the usual missing data formulation for mixture setups, where a set of allocation variables $z = \{z_t\}_{t=1}^T$ are defined such that

$$ z_t = \begin{cases} 1 & \text{with probability } \omega_1 \\ \vdots & \vdots \\ k & \text{with probability } \omega_k \\ \vdots & \vdots \\ g & \text{with probability } \omega_g \end{cases}, \quad (t = 1, \ldots, T) \quad (8.3) $$

With this approach, the observed series $y = \{y_t\}_{t=1}^T$ and the unobserved (residual) series $\varepsilon = \{\varepsilon_t\}_{t=1}^T$ becomes completed with a missing data set $z = \{z_t\}_{t=1}^T$, indicating the specific component of the mixture from which every observation is assumed to arise (see Diebolt and Robert (1994)).

In the light of the above characterization of the Student-$t$ distribution, it is convenient to view the observed data augmented by the latent variable $z_t$ as still being incomplete. Therefore, likewise the $M_{t \text{AR}}$ and $M_{t \text{ARMA}}$ models we introduce a new missing variable $\xi = \{\xi_t\}_{t=1}^T$ which corresponds to the mixing random variables in the Student-$t$ distribution and control the magnitude of the variance in the normal density function (see McLachlan and Peel (2000)).

Consequently, conditionally on $\xi_t$ which is conditional on $z_t$, the observations are drawn from their respective individual subpopulation, i.e.

$$ y_t \mid \mathcal{G}_{t-1}, z_t \sim \begin{cases} \mathcal{N} \left( y_t; \nu_{1,t}, (\tau_{1,t}\xi_t)^{-1} \right) \mathcal{G} \left( \xi_t; \frac{\nu_1}{2}, \frac{\nu_1-2}{2} \right) & \text{if } z_t = 1, \\ \vdots & \vdots \\ \mathcal{N} \left( y_t; \nu_{k,t}, (\tau_{k,t}\xi_t)^{-1} \right) \mathcal{G} \left( \xi_t; \frac{\nu_k}{2}, \frac{\nu_k-2}{2} \right) & \text{if } z_t = k, \\ \vdots & \vdots \\ \mathcal{N} \left( y_t; \nu_{g,t}, (\tau_{g,t}\xi_t)^{-1} \right) \mathcal{G} \left( \xi_t; \frac{\nu_g}{2}, \frac{\nu_g-2}{2} \right) & \text{if } z_t = g, \end{cases} \quad (8.4) $$

provided $v_k > 2, \ (k = 1, \ldots, g)$.

We divide the parameters into two groups such as $\varphi^\dagger = (\omega, \mu, \phi, \theta)$ and $\varphi^\ddagger = (\alpha, \beta)$ for convention. The update of the parameters in the first group $\varphi^\dagger$ are...
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

based on the original \( \mathcal{M}_{\text{ARMA-GARCH}} \) model which is given by the Equation (5.2). But we assume that the conditional variances \( \{\tau_{k,t}\}_{i=1}^{n_k} \) are fixed and known. So, we can update the corresponding posterior density \( \pi(\phi^{\dagger} \mid g, y, z) \) by using the full conditional likelihood of the model (8.1):

\[
\mathcal{L}(\phi^{\dagger} \mid \Lambda, y, z, \xi) = \prod_{k=1}^{g} \prod_{t: z_{t}=k} \mathcal{N}(y_{t}; \nu_{k,t}, (\tau_{k,t}\xi_{t})^{-1}) \mathcal{G}(\xi_{t}; v_{k}/2, (v_{k} - 2)/2), \quad v > 2. \tag{8.5}
\]

where

\[
\nu_{k,t} = \mu_{k} + \sum_{j=1}^{p_k} \phi_{kj}(y_{t-j} - \mu_{k}) + \sum_{j=1}^{q_k} \theta_{kj}\hat{\epsilon}_{t-j}, \tag{8.6}
\]

and the residual term \( \hat{\epsilon}_{t} \) is defined as:

\[
\hat{\epsilon}_{t} = \begin{cases} 
0, & (t \leq 0), \\
y_{t} - \nu_{k,t}, & (t = 1, \ldots, T),
\end{cases} \tag{8.7}
\]

Note that the likelihood function mentioned above (8.5) separates into \( g \) parts where each one concerning the data assigned to each of the \( g \) mixture components.

On the other hand, the update of the parameters in the second group \( \phi^{\dagger\dagger} \) are based on an approximated mixture of GARCH model. Now, we recall the \( k \)th GARCH component model from our main model defined by (8.1) is:

\[
\tau_{k,t}^{-1} = \alpha_{k0} + \sum_{j=1}^{r_k} \alpha_{kj}\hat{\epsilon}_{t-j}^2 + \sum_{j=1}^{s_k} \beta_{kj}\tau_{k,t-j}^{-1}. \tag{8.8}
\]

As shown in Section 5.4, the above GARCH(\( r_{k}, s_{k} \)) process can be interpreted as an ARMA process in \( \hat{\epsilon}_{t}^2 \) of order (\( l_{k}, s_{k} \)) as follows

\[
\hat{\epsilon}_{t}^2 = \alpha_{k0} + \sum_{i=1}^{l_k} (\alpha_{ki} + \beta_{ki})\hat{\epsilon}_{t-i}^2 + u_{k,t} - \sum_{j=1}^{s_k} \beta_{kj}u_{k,t-j}, \quad u_{k,t} \sim \mathcal{N}(0, 2\tau_{k,t}^{-2}), \tag{8.9}
\]

where \( l_{k} = \max\{r_{k}, s_{k}\}, \alpha_{ki} > 0 \) for \( i > r_{k} \), and \( \beta_{kj} > 0 \) for \( j > s_{k} \).

Hence, the full likelihood function of the approximated GARCH model (8.9) is

\[
\mathcal{L}(\phi^{\dagger\dagger} \mid \Lambda, \phi, z, \hat{\epsilon}_{t}^2) = \prod_{k=1}^{g} \prod_{t: z_{t}=k} \mathcal{N}(\hat{\epsilon}_{t}^2; \nu_{k,t}^2, 2\tau_{k,t}^{-2}) \tag{8.10}
\]
where
\[ \nu_{k,t}^\varepsilon = \alpha_{k0} + \sum_{j=1}^{l_k}(\alpha_{kj} + \beta_{kj})\bar{\varepsilon}_{t-j}^2 - \sum_{j=1}^{s_k}\beta_{kj}\hat{\mu}_{t-j}, \tag{8.11} \]

with
\[ \hat{\mu}_t = \begin{cases} 0, & (t \leq 0), \\ \nu_{k,t}^\varepsilon, & (t = 1, \ldots, T), \end{cases} \tag{8.12} \]

8.4 Priors setup

For fixed \( g \), we will have a specific \( M_{\text{ARMA-GARCH}} \). The prior distributions of \((\omega, \mu, \phi, \theta, \alpha, \beta, p, q, r, s)\) are the same as that of the \( M_{\text{ARMA-GARCH}} \) model discussed in section 5.5. We state these priors again here for convenience. In addition, we define priors for \( \nu \).

8.4.1 Priors of \( \omega, \mu \)

\[ \omega \mid g \sim D(\omega; \delta_1, \ldots, \delta_g) \tag{8.13} \]

\[ \mu_k \mid \text{id} \sim N(\mu_k; \zeta, \kappa^{-1}), \quad (k = 1, \ldots, g) \tag{8.14} \]

where \( D(\cdot; \delta_1, \ldots, \delta_g) \) and \( N(\cdot; \zeta, \kappa^{-1}) \) denote the Dirichlet distribution and the normal distribution respectively. For details of these distributions see section 2.11.

8.4.2 Priors for \( \phi, \theta, \alpha, \text{and} \beta \)

The priors of the AR and MA coefficients as set based on the original \( M_{\text{ARMA-GARCH}} \) given by (8.1) model as below:

\[ \phi_k \sim \text{Uniform on } \Xi_{\rho} \iff \rho_{ki}^\phi \mid \text{id} \sim B(-1, +1) \left( \rho_{ki}^\phi, \frac{i + 1}{2} \right), \quad (k = 1, \ldots, g, i = 1, \ldots, q_k) \tag{8.15} \]
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

\[
\theta_k \sim \text{Uniform on } \Xi_q \iff \rho_{ki}^\theta \sim \mathcal{B}_{(-1,+1)}\left(\rho_{ki}^\theta; \left[\frac{i+1}{2}\right], \left[\frac{i}{2}\right] + 1\right),
\]

\[(k = 1, \ldots, g, i = 1, \ldots, q_k) \quad (8.16)\]

We set priors of the ARCH and GARCH coefficients based on approximated ARMA(\(l_k, s_k\)) model (8.9):

\[
\alpha_k \sim \text{Uniform on } \Xi_r \iff \rho_{ki}^\alpha \sim \mathcal{B}_{(-1,+1)}\left(\rho_{ki}^\alpha; \left[\frac{i+1}{2}\right], \left[\frac{i}{2}\right] + 1\right),
\]

\[(k = 1, \ldots, g, i = 1, \ldots, r_k) \quad (8.17)\]

\[
\beta_k \sim \text{Uniform on } \Xi_s \iff \rho_{ki}^\beta \sim \mathcal{B}_{(-1,+1)}\left(\rho_{ki}^\beta; \left[\frac{i+1}{2}\right], \left[\frac{i}{2}\right] + 1\right),
\]

\[(k = 1, \ldots, g, i = 1, \ldots, s_k) \quad (8.18)\]

where \(\rho_{ki}^\phi\) be the partial autocorrelation coefficients at lag \(i\) for the \(k\)th component; \(\mathcal{B}_{(-1,+1)}(x; a, b)\) represents a generalized beta distribution defined on \((-1, +1)\), where \([a]\) means the integer part of \(a\); \(\Xi_p \times \Xi_q\) denotes the stationarity and invertibility region of \(\phi\) and \(\theta\) while \(\Xi_r \times \Xi_s\) denotes the stationarity and invertibility region of \(\alpha\) and \(\beta\).

The above prior settings suggest that there is one-to-one relation between the partial autocorrelations and the \(#,(\# = \phi, \theta, \alpha, \beta)\) and the transformation is done by the method proposed by Barndorff-Nielsen and Schou (1973).

### 8.4.3 Priors for \(\Lambda\)

However, by setting the model index \(\Lambda = (p, q, r, s, v)\) as random quantities we will get a complete Bayesian structure for our model. As a consequences, we set
the priors for these (see Sections 5.5 and 7.4):

\[ g \sim \mathcal{U}(g; 1, g_{\text{max}}) \quad (8.19) \]

\[ p_k \mid g \sim \mathcal{U}(p_k; 0, p_{\text{max}}), \quad (k = 1, \ldots, g) \quad (8.20) \]

\[ q_k \mid g \sim \mathcal{U}(q_k; 0, q_{\text{max}}), \quad (k = 1, \ldots, g) \quad (8.21) \]

\[ r_k \mid g \sim \mathcal{U}(r_k; 0, r_{\text{max}}), \quad (k = 1, \ldots, g) \quad (8.22) \]

\[ s_k \mid g \sim \mathcal{U}(s_k; 0, s_{\text{max}}), \quad (k = 1, \ldots, g) \quad (8.23) \]

\[ v_k \mid g \sim \mathcal{U}(v_k; 3, v_{\text{max}}), \quad (k = 1, \ldots, g) \quad (8.24) \]

where \( \mathcal{U}(x \mid a, b) \) denotes the discrete uniform distribution for \( x = a, a+1, \ldots, b \) and \( g_{\text{max}}, p_{\text{max}}, q_{\text{max}}, r_{\text{max}}, s_{\text{max}} \) and \( v_{\text{max}} \) are fixed.

For details of the above distributions mentioned above see Section 2.11.

### 8.4.4 Formulae for hyperparameters

In the prior settings, the parameters \( \delta, \zeta, \kappa, c, a \) and \( b \) are fixed and they are defined by the same formulae as discussed in Section 3.5.1:

\[ \delta_1 = \cdots = \delta_g = 1 \]
\[ R_y = \max(y) - \min(y) \]
\[ \zeta = \min(y) + \frac{R_y}{2}, \quad \kappa = \frac{1}{R_y} \]
\[ c = 2, \quad a = 0.2, \quad b = \frac{100}{c \cdot R_y^2} \quad (8.25) \]

### 8.5 MCMC moves for parameter estimation

First we consider the number of component of the model \( \mathcal{M}_{\text{ARMA-GARCH}}^t, g \), is fixed. We apply MCMC component-wise techniques to calculate the full conditional posterior distribution \( \pi(\varphi \mid y, \Lambda) \) in which each of the parameter will update separately. Moreover, we need to update the latent variables \( z \) and \( \xi \) as well. The moves are discussed below:
Fig. 8.1: Directed Acyclic Graph of the $\mathcal{M}_{\text{ARMA-GARCH}}^t$ model. Squared nodes refer to constant parameters, circled nodes refer to stochastic components and the bold squared node indicates observed data of the model. The arrows indicate the conditional independence structure of the model.
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

8.5.1 Move I: updating $z$

As described in Section 7.5.7, this is also a Gibbs-typed move. From (8.3) and (8.4), the full conditional posterior probability that the observations $y_t$ has been generated by the $k$-th mixture component is,

$$\Pr(z_t = k \mid y_t, \varphi) = \frac{1}{C_6} \omega_k \sqrt{\tau_{k,t}} \exp \left[ -\frac{\tau_{k,t}\xi_t}{2} (y_t - \hat{\nu}_{k,t})^2 \right],$$

where $C_6 = \sum_{l=1}^{g} \omega_l \sqrt{\tau_{l,t}} \exp \left[ -\frac{\tau_{l,t}\xi_t}{2} (y_t - \nu_{l,t})^2 \right]$ denotes the normalizing constant; $\hat{\nu}_{*,t}$ is defined as in the Equation (8.6). For details see Section 4.6.6.

8.5.2 Move II: updating $\omega$

This is a Gibbs-type move similar as in Section 3.6.2. Then the full conditional posterior distribution $\pi(\omega \mid \varphi_{-\omega}, y, z)$ is:

$$\omega \mid \varphi_{-\omega}, y, z \sim \mathcal{D}(\omega; \delta_1 + n_1, \ldots, \delta_g + n_g),$$

where $n_k = \sum_{t: z_t = k} 1$ and $\varphi_{-\omega}$ denotes the remaining parameters except $\omega$.

8.5.3 Move III: updating $\mu$

This is a Gibbs-type move and the full conditional distribution $\pi(\mu \mid \varphi_{-\mu}, y, z)$, where $\varphi_{-\mu}$ denote the remaining parameter excepts $\mu$, is can be obtained as
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

below:

\[ \pi(\mu | \varphi_{-\mu}, y, z, \xi) \propto \mathcal{L}(\varphi_{-\mu} | y, z) \pi(\mu) \]

\[ \propto \prod_{k=1}^{g} \prod_{t:z_t=k} \mathcal{N}(y_{t i} | \bar{v}_{k,i}, (\tau_{k,t} \xi_t)^{-1}) \mathcal{G}(\xi_t; \nu_k/2, (v_k - 2)/2) \]

\[ \times \prod_{k=1}^{g} \mathcal{N}(\mu_k; \zeta, \kappa^{-1}) \]

Accordingly, the full conditional posterior density for the \( k \)-th component is:

\[ \pi(\mu_k | \varphi_{-\mu_k}, y, z, \xi) \propto \prod_{t:z_t=k} \mathcal{N}(y_{t i} | \bar{v}_{k,i}, (\tau_{k,t} \xi_t)^{-1}) \cdot \mathcal{N}(\mu_k; \zeta, \kappa^{-1}) \]

\[ \propto \exp\left[ -\frac{1}{2} \sum_{t:z_t=k} \tau_{k,t} \xi_t \left\{ y_{t i} - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t i - i} - \mu_k) + \sum_{j=1}^{q_k} \theta_{k,j} \hat{\varepsilon}_{t i - j} \right\}^2 \right] \]

\[ \cdot \exp\left[ -\frac{\kappa}{2} (\mu_k - \zeta)^2 \right] \]

Now the term \((****)\) can be simplified in a similar fashion as in the case of \( \mathcal{M}_t^{\text{AR}} \) model (see Section 6.5.2):

\[ \sum_{t:z_t=k} \tau_{k,t} \xi_t \left\{ y_{t i} - \mu_k - \sum_{i=1}^{p_k} \phi_{k,i} (y_{t i - i} - \mu_k) + \sum_{j=1}^{q_k} \theta_{k,j} \hat{\varepsilon}_{t i - j} \right\}^2 \]

\[ = \sum_{t:z_t=k} \tau_{k,t} \xi_t \left( y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t - i} - \sum_{j=1}^{q_k} \theta_{k,j} \hat{\varepsilon}_{t - j} - (1 - \sum_{i=1}^{p_k} \phi_{k,i}) \mu_k \right)^2 \]

\[ = \sum_{t:z_t=k} \tau_{k,t} \xi_t (e_{k,t} - b_k \mu_k)^2 = \sum_{t:z_t=k} \tau_{k,t} \xi_t (e_{k,t} - \bar{e}_k + \bar{e}_k - b_k \mu_k)^2, \]

where \( \bar{e}_k = \frac{1}{n_k} \sum_{t:z_t=k} e_{k,t} \) provided \( n_k = \sum_{t:z_t=k} 1 \)
\[ = \sum_{t: z_t \equiv k} \tau_{k,t} \xi_t \left\{ (e_{k,t} - \bar{e}_k)^2 + 2(e_{k,t} - \bar{e}_k)(\bar{e}_k - b_k \mu_k) + (\bar{e}_k - b_k \mu_k)^2 \right\} \]

Substituting the simplified results into the equation (\#) we have,

\[
\pi(\mu_k | \varphi_{-\mu_k}, y, z, \xi) \propto \exp\left[ -\left( \bar{e}_k - b_k \mu_k \right)^2 \right]
\]

\[
\propto \exp\left[ -\frac{1}{2} \left( \left( e_{k,t} - \bar{e}_k b_k d^t_k + \kappa \mu_k + b_k e^t_k \right) \mu_k \right) \right]
\]

\[
\propto \exp\left[ -A \mu_k^2 + 2B \mu_k \right] = \exp\left[ -A \left( \mu_k^2 - 2B \mu_k \right) \right]
\]

\[
= \exp\left[ -A \left( \mu_k - \frac{B}{A} \right)^2 \right] = \exp\left[ -A \left( \mu_k - \frac{B}{A} \right)^2 \right]
\]

\[
\propto \exp\left[ -\frac{b_k d^t_k + \kappa}{2} \left( \mu_k - \frac{\bar{e}_k b_k d^t_k + \kappa \zeta + b_k e^t_k}{b_k d^t_k + \kappa} \right)^2 \right].
\]

Hence, the full conditional posterior distribution of \( \mu \) is:

\[
\mu_k | \varphi_{-\mu_k}, y, z, \xi \sim N(\mu_k; \frac{\bar{e}_k b_k d^t_k + \kappa \zeta + b_k e^t_k}{b_k d^t_k + \kappa}, \frac{1}{b_k d^t_k + \kappa}),
\]

\[(k = 1, \ldots, g), \quad (8.28)\]

where

\[ \bar{e}_k = \frac{1}{n_k} \sum_{t: z_t \equiv k} e_{k,t}, \]
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

\[ e_{k,t} = y_t - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i} - \sum_{j=1}^{q_k} \theta_{k,j} \hat{e}_{t-j}, \text{and} \]

\[ n_k = \sum_{t:z_t=k} 1; \]

\[ b_k = 1 - \sum_{i=1}^{p_k} \phi_{k,i}; \]

\[ c^+_k = \sum_{t:z_t=k} \tau_{k,t} \xi_t (e_{k,t} - \bar{e}_k); \]

\[ d^+_k = \sum_{t:z_t=k} \tau_{k,t} \xi_t. \]

where \( \tau_{k,t} \) is fixed and known defined in the Equation (8.1).

### 8.5.4 Move IV-A: updating \( \phi \)

This is a similar move by applying Metropolis-Hasting mechanism discussed in Section 5.6.4. Based on the prior set up for \( \phi \) defined in (8.15), the update of this is equivalent to the update \( \phi \). The algorithm is discussed here briefly as follows:

A candidate \( \phi^*_k \) is generated by a normal density truncated in \((-1, +1)\) and centred in the current state of the chain \( \phi_k \) so that we can define the proposal density as below:

\[
g_{\phi}(\phi_k; \phi^*_k) = \mathcal{N}_{(-1,+1)}(\phi_k^*_k; \tau^{-1}_q), \quad (k = 1, \ldots, g; i = 1, \ldots, p_k), \quad (8.29)
\]

where the precision \( \tau^{-1}_q \) is chosen in order to obtain a satisfactory acceptance rate.

Using the distribution defined above by (8.29), we generate a vector \( \phi^*_k = (\rho^*_{k1}, \ldots, \rho^*_{kq_k}) \) for the partial autocorrelations which will be then used to derive the corresponding parameter vector \( \phi^*_k = (\phi^*_{k1}, \ldots, \phi^*_{kp_k}) \) through the transformation proposed by Barndorff-Nielsen and Schou (1973).

Now, the ratio in the acceptance probability for a M-H algorithm can be defined as

\[
A_{\phi}(\phi^*_k; \phi_k) = \min(1, \Omega_{\phi}),
\]

where

\[
\Omega_{\phi} = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}
\]
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

In this case, the likelihood ratio is simply given by (see Equation (5.6)):

\[
\text{likelihood ratio} = \frac{\prod_{t:z_t=k} \mathcal{N}(y_t; \hat{\nu}_{k,t}^*, (\tau_{k,t}\xi_t)^{-1}) \mathcal{G}(\xi_t; v_k/2, (v_k - 2)/2)}{\prod_{t:z_t=k} \mathcal{N}(y_t; \hat{\nu}_{k,t}^*, (\tau_{k,t}\xi_t)^{-1}) \mathcal{G}(\xi_t; v_k/2, (v_k - 2)/2)}
\]

\[= \exp \left[ -\frac{1}{2} \sum_{t:z_t=k} \tau_{k,t} \xi_t \left\{ (y_t - \hat{\nu}_{k,t}^*)^2 - (y_t - \hat{\nu}_{k,t})^2 \right\} \right], \quad (8.30)\]

where \(\hat{\nu}_{k,t}^* = \mu_k + \sum_{j=1}^{p_k} \phi_{kj}^* (y_{t-j} - \mu_k) - \sum_{j=1}^{q_k} \theta_{kj} \check{\epsilon}_{t-j} \) and \(\check{\epsilon}_t\) is defined as in the Equation (8.7).

From (8.15) we have (for details, see Section 5.6.4):

\[
\text{prior ratio} = \prod_{i=2}^{p_k} \left( \rho_{k_i}^{\phi^*} + 1 \right) \left[ \frac{1 + i}{2} \right]^{-1} (1 - \rho_{k_i}^{\phi^*})^{\left[ \frac{i}{2} \right]}
\]

\[= \prod_{i=2}^{p_k} \frac{\Phi_N(1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1}) - \Phi_N(-1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1})}{\Phi_N(-1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1}) - \Phi_N(1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1})}, \quad (8.31)\]

and, eventually from the equation (8.29) the proposal ratio is:

\[
\text{proposal ratio} = \prod_{i=1}^{p_k} \frac{\Phi_N(1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1}) - \Phi_N(-1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1})}{\Phi_N(-1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1}) - \Phi_N(1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1})}, \quad (8.32)\]

where \(\Phi_N\) is the normal cumulative distribution function (for details, see Section 5.6.5).

By using (8.30), (8.31) and (8.32) we have

\[
\Omega_{\phi} = \exp \left[ -\frac{1}{2} \sum_{t:z_t=k} \tau_{k,t} \xi_t \left\{ (y_t - \hat{\nu}_{k,t}^*)^2 - (y_t - \hat{\nu}_{k,t})^2 \right\} \right]
\]

\[\times \prod_{i=2}^{p_k} \left( \rho_{k_i}^{\phi^*} + 1 \right) \left[ \frac{1 + i}{2} \right]^{-1} (1 - \rho_{k_i}^{\phi^*})^{\left[ \frac{i}{2} \right]}
\]

\[\times \prod_{i=1}^{p_k} \frac{\Phi_N(1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1}) - \Phi_N(-1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1})}{\Phi_N(-1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1}) - \Phi_N(1; \rho_{k_i}^{\phi}, \tau_{q_i}^{-1})}.
\]

(8.33)
8.5.5 Move IV-B: updating $\theta$

This is exactly the same typed Metropolis-Hasting move discussed in the previous section. We present a short discussion below with appropriate notations.

A candidate $\rho_{k,i}^*$ is generated by a normal density truncated in $(-1, +1)$ and centred in the current state of the chain $\rho_{k,i}^0$ so that we can define the proposal density as below:

$$q_\theta(\rho_{k,i}^0, \rho_{k,i}^*) = N_{(-1, +1)}(\rho_{k,i}^0 | \rho_{k,i}^0, \tau_q^{-1}) ; (k = 1, \ldots, g; i = 1, \ldots, q_k)$$

(8.34)

Using the distribution defined above by (8.39), we generate a vector $\rho_{k}^* = (\rho_{k,1}^*, \ldots, \rho_{k,p_k}^*)$ for the partial autocorrelations which will be then used to derive the corresponding parameter vector $\theta_{k}^* = (\theta_{k,1}^*, \ldots, \theta_{k,p_k}^*)$ through the transformation as discussed in Barndorff-Nielsen and Schou (1973).

Therefore, the ratio in the acceptance probability for a M-H algorithm can be defined as $A_\theta \min(1, \Omega_\theta)$, where

$$\Omega_\theta = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}$$

$$= \exp\left[- \frac{1}{2} \sum_{t: z_t = k} \tau_{k,t} \xi_t \left\{ (y_t - \hat{\nu}_{k,t}^*)^2 - (y_t - \nu_{k,t})^2 \right\} \right]$$

$$\times \prod_{i=2}^{p_k} \left( \frac{1}{\rho_{k,i}^0 + 1} \right)^{-1/2} \left( 1 - \rho_{k,i}^0 \right)^{1/2}$$

$$\times \prod_{i=1}^{p_k} \Phi_N\left( +1; \rho_{k,i}^0, \tau_q^{-1} \right) - \Phi_N\left( -1; \rho_{k,i}^0, \tau_q^{-1} \right)$$

(8.35)

where $\hat{\nu}_{k,t}^* = \mu_k + \sum_{j=1}^{p_k} \phi_{kj} (y_{t-j} - \mu_k) + \sum_{j=1}^{q_k} \theta_{k,i}^* \hat{\epsilon}_{t-j} ; \hat{\epsilon}_t$ is defined by the equation (8.7) and $\Phi_N$ denotes the cumulative distribution function.
8.5.6 Move V-A: updating $\alpha$

The update of $\rho_{k,i}^{\alpha}$, $(k = 1, \ldots, g)$ by a Metropolis-Hasting mechanism in a similar way as discussed for the $\mathcal{M}_{\text{ARMA-GARCH}}$ model. For details of this move please see Section 5.6.6.

If a candidate $\rho_{k,i}^{\alpha^*}$ is generated by a normal density truncated in $(-1, +1)$ and centred in the current state of the chain $\rho_{k,i}^{\alpha}$ so that we can define the proposal density as below:

$$q_{\alpha}(\rho_{k,i}^{\alpha}, \rho_{k,i}^{\alpha^*}) = \mathcal{N}(-1, +1) \left( \rho_{k,i}^{\alpha^*}, \tau_{q^*}^{-1} \right), \quad (k = 1, \ldots, g, i = 1, \ldots, r_k)$$

(8.36)

Using the distribution defined above by (5.37), we generate a vector $\rho_{k}^{\alpha^*} = (\rho_{k,1}^{\alpha^*}, \ldots, \rho_{k,p_k}^{\alpha^*})$ for the partial autocorrelations which will be then used to derive the corresponding parameter vector $\alpha_k^{*} = (\alpha_{k,1}^{*}, \ldots, \alpha_{k,p_k}^{*})$ through the transformation as proposed in Barndorff-Nielsen and Schou (1973).

Then the acceptance probability for a M-H algorithm can be defined as

$$A_{\alpha}(\alpha_k^*, \alpha_k) = \min(1, \Omega_{\alpha}),$$

(8.37)

where

$$\Omega_{\alpha} = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}$$

$$= \exp \left[ -\frac{1}{4} \sum_{t=1-i}^i \left\{ \left( z_t^2 - \nu_{k,t}^{\alpha^*} \right)^2 - \left( z_t^2 - \nu_{k,t}^{\alpha} \right)^2 \right\} \right]$$

$$\times \prod_{i=2}^{p_k} (\rho_{k,i}^{\alpha^*} + 1)^{\left[ \frac{i-1}{2} \right]} (1 - \rho_{k,i}^{\alpha^*})^{\left[ \frac{i}{2} \right]}$$

$$\times \prod_{i=1}^{p_k} \Phi_N \left( -1; \rho_{k,i}^{\alpha^*}, \tau_{q^*}^{-1} \right) - \Phi_N \left( +1; \rho_{k,i}^{\alpha^*}, \tau_{q^*}^{-1} \right)$$

$$\times \prod_{i=1}^{p_k} \Phi_N \left( +1; \rho_{k,i}^{\alpha}, \tau_{q}^{-1} \right) - \Phi_N \left( -1; \rho_{k,i}^{\alpha}, \tau_{q}^{-1} \right),$$

(8.38)
\[ \hat{\nu}_{k,t}^2 = \alpha_{k0}^* + \sum_{j=1}^{l_k} (\alpha_{kj}^* + \beta_{kj}) \hat{\varepsilon}_{t-j}^2 - \sum_{j=1}^{s_k} \beta_{kj} \hat{u}_{t-j}^*, \text{ where } l_k = \max\{r_k, s_k\} \]

and \( \hat{u}_t^* \) is defined as in the Equation (8.12).

### 8.5.7 Move V-B: updating \( \beta \)

This is also a similar Metropolis-Hasting move discussed in the previous section. For details, see Section 5.6.7.

To update \( \rho_k^\beta, \ k = 1, \ldots, g \), we first select a candidate \( \rho_{k,i}^{\beta^1} \) that is generated by a normal density truncated in \((-1, +1)\) and centred in the current state of the chain \( \rho_{k,i}^\beta \), so that we can define the proposal density as below:

\[
q_\beta(\rho_{k,i}^{\beta^1}; \rho_{k,i}^\beta) = \mathcal{N}_{(-1,+1)} \left( \rho_{k,i}^{\beta^1}; \rho_{k,i}^\beta, \tau_{q_{\beta}}^{-1} \right), \quad (i = 1, \ldots, s_k, k = 1, \ldots, g).
\]  

Using the distribution defined above by (8.39), we generate a vector \( \rho_{k}^{\beta^1} = (\rho_{k,1}^{\beta^1}, \ldots, \rho_{k,p_k}^{\beta^1}) \) for the partial autocorrelations which will be then used to derive the corresponding parameter vector \( \beta_{k}^{\dagger} = (\beta_{k,1}^{\dagger}, \ldots, \beta_{k,p_k}^{\dagger}) \) through the transformation.

Therefore, the ratio in the acceptance probability for a M-H algorithm can be defined as

\[
A_\beta = \min(1, \Omega_\beta),
\]

where
\[ \Omega_\beta = \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio} \]

\[ = \exp \left[ -\frac{\tau_{k,t}^2}{4} \sum_{t:z_t=k} \left\{ \left( \varepsilon_t^2 - \hat{\nu}_{k,t}^\dagger \right)^2 - \left( \hat{\epsilon}_{t}^2 - \hat{\nu}_{k,t}^\dagger \right)^2 \right\} \right] \]

\[ \times \prod_{i=2}^{p_k} (\rho_{k,i}^\alpha + 1)^{\left[ \frac{i+1}{2} \right] - 1} \left( 1 - \rho_{k,i}^\alpha \right)^{\left[ \frac{i}{2} \right]} \]

\[ \times \prod_{i=1}^{p_k} \Phi_N \left( +1; \rho_{k,i}^\alpha \tau_{q-1}^{-1} \right) - \Phi_N \left( -1; \rho_{k,i}^\alpha \tau_{q-1}^{-1} \right) \]

\[ \times \prod_{i=1}^{p_k} \Phi_N \left( +1; \rho_{k,i}^\alpha \tau_{q-1}^{-1} \right) - \Phi_N \left( -1; \rho_{k,i}^\alpha \tau_{q-1}^{-1} \right), \]

(8.41)

where

\[ \hat{\nu}_{k,t}^\dagger = \alpha_{k0} + \sum_{j=1}^{l_k} (\alpha_{kj} + \beta_{kj}^\dagger) \hat{\epsilon}_{t-j}^2 - \sum_{j=1}^{s_k} \beta_{kj}^\dagger \hat{u}_{t-j}^\dagger, \text{ where } l_k = \max\{r_k, s_k\} \]  

(8.42)

and \( \hat{u}_t^\dagger \) is defined as in the Equation (8.12).

### 8.5.8 Move VI: updating \( \xi \)

To get a full conditional posterior distribution of \( \xi \), we apply the Gibbs-typed move as below:

\[ \pi(\xi | \phi, y, z) \propto \mathcal{L}(\phi | y, z) \pi(z | \omega) \]

\[ = \prod_{t=1}^{n} \sum_{k=1}^{q} \omega_k \mathcal{N} \left( y_t; \hat{\nu}_{k,t}, (\tau_{k,t}\xi_t)^{-1} \right) \mathcal{G} \left( \xi_t, \psi_k, \frac{\psi_k - 2}{2} \right) \cdot 1_{\{z_t=k\}}. \]
Therefore, the full conditional distribution of \( \xi_t \mid z_t = k \) is:

\[
\pi(\xi_t \mid \varphi, y, z_t = k) \propto \omega_k \mathcal{N}(y_t; \hat{\nu}_{k,t}, (\tau_{k,t}\xi_t)^{-1}) \mathcal{G}\left(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}\right)
\]

\[
\propto \omega_k (\tau_{k,t}\xi_t)^{1/2} \exp\left[-\frac{\tau_{k,t}\xi_t}{2} (y_t - \hat{\nu}_{k,t})^2\right]
\]

\[
\cdot \xi_t^{\frac{v_k}{2} - 1} \exp\left[-\left(\frac{v_k}{2}\right)\xi_t\right]
\]

\[
\propto \xi_t^{\frac{v_k}{2} + \frac{1}{2} - 1} \exp\left[-\left\{\frac{\tau_{k,t}}{2} (y_t - \hat{\nu}_{k,t})^2 + \frac{v_k - 2}{2}\right\}\xi_t\right].
\]

Hence, we have the full conditional posterior distribution of \( \xi_t \):

\[
\xi_t \mid \varphi, y, z_t = k \sim \mathcal{G}\left(\xi_t; \frac{v_k}{2} + \frac{1}{2}, \frac{\tau_{k,t}}{2} (y_t - \hat{\nu}_{k,t})^2 + \frac{v_k - 2}{2}\right),
\]

\[(t = 1, \ldots, T; k = 1, \ldots, g), \quad (8.43)\]

where \( \hat{\nu}_{k,t} \) is defined as in the Equation (8.6).

### 8.6 MCMC moves for model determination

Model determination for the \( \mathcal{M}_t^{AR} \) model is the same as that for the \( \mathcal{M}_t^{AR} \) model. We recall all techniques here in short. For details see Section 3.7. To get full Bayesian flavour we must get the posterior distribution of the model index \( \Lambda = (g, p, q, r, s, v) \) which can be written and factorized as:

\[
\pi(\Lambda \mid y) = \pi(p \mid \Lambda_{-p}, g, y) \times \pi(q \mid \Lambda_{-q}, g, y) \times \pi(\alpha \mid \Lambda_{-r}, g, y)
\]

\[
\times \pi(\beta \mid \Lambda_{-s}, g, y) \times \pi(v \mid \Lambda_{-v}, g \mid y)
\]

\[(8.44)\]

where \( \Lambda_{-\#} \) denotes all the model index parameters except \( \# = p/q/r/s \)

Evidently, We need to develop FIVE additional MCMC moves to make our Bayesian analysis complete.
8.6.1 Move VII-A: updating of $p$

In a similar fashion, we can update the parameter $p$ as discussed for the $\mathcal{M}_{\text{ARMA-GARCH}}$ model in Section 4.7.1.

At first we need to choose a component, say $k^\dagger$, randomly chosen in $\{1, \ldots, g\}$. Then we can write the proposal for the new AR order of the $k^\dagger$-th component, $p_{k^\dagger}$, as:

\[
p_{k^\dagger} = \begin{cases} 
p_{k^\dagger} - 1, & \text{with probability } d(p_{k^\dagger}), \\
p_{k^\dagger} + 1, & \text{with probability } b(p_{k^\dagger}), 
\end{cases}
\] (8.45)

where $b(p) = 1 - d(p)$, for $k = 1, \ldots, g$, $d(0) = 0$ and $b(p_{\text{max}}) = 0$.

Let $\rho_{k^\dagger}^{\phi}$ be the proposal vector of the partial autocorrelations. When $p_{k^\dagger} = p_{k^\dagger} - 1$, the autoregressive coefficients $\rho_{k^\dagger}^{\phi} = (\rho_{k^\dagger 1}^{\phi}, \ldots, \rho_{k^\dagger p_{k^\dagger}}^{\phi})$ is updated by simply discarding the last partial autocorrelation in $\rho_{k^\dagger}^{\phi} = (\rho_{k^\dagger 1}^{\phi}, \ldots, \rho_{k^\dagger p_{k^\dagger}}^{\phi})$. On the other hand, when $p_{k^\dagger} = p_{k^\dagger} + 1$, the corresponding autocorrelations $\rho_{k^\dagger}^{\phi} = (\rho_{k^\dagger 1}^{\phi}, \ldots, \rho_{k^\dagger p_{k^\dagger}}^{\phi}, \rho_{k^\dagger p_{k^\dagger} + 1}^{\phi})$ is updated by adding new parameter $\rho_{k^\dagger p_{k^\dagger} + 1}^{\phi}$ in $\rho_{k^\dagger}^{\phi} = (\rho_{k^\dagger 1}^{\phi}, \ldots, \rho_{k^\dagger p_{k^\dagger}}^{\phi})$ with

\[
\rho_{k^\dagger p_{k^\dagger} + 1}^{\phi} \overset{\text{ind}}{\sim} \mathcal{B}_{(-1,+1)} \left( \rho_{k^\dagger p_{k^\dagger} + 1}^{\phi}; \left[ \frac{p_{k^\dagger} + 1}{2} \right], \left[ \frac{p_{k^\dagger} + 1}{2} + 1 \right] \right).
\]

Then update the corresponding AR coefficients parameters by using the reparametrization discussed in Section 4.5.

To update the AR order of the model by using RJMCMC method, we have the acceptance probability

\[
\mathcal{A}_{p^\dagger} = \min \{ 1, \Omega_{p^\dagger} \},
\]

and $\Omega_{p^\dagger}$ can be obtained as (for details, see Section 3.7.1):

\[
\Omega_{p^\dagger} = \begin{cases} 
L_{\dagger} \frac{d(p_{k^\dagger})}{d(p_{k^\dagger - 1})} & \text{if } p_{k^\dagger} = p_{k^\dagger} - 1, \\
L_{\dagger} \frac{b(p_{k^\dagger})}{b(p_{k^\dagger - 1})} & \text{if } p_{k^\dagger} = p_{k^\dagger} + 1,
\end{cases}
\] (8.46)
where
\[
L^† = \exp \left\{ -\frac{\tau_{k^†, t} \xi_t}{2} \sum_{t: z_t = k} \left\{ (y_t - \hat{\nu}_{k^†, t})^2 - (y_t - \hat{\nu}_{k^†, t})^2 \right\} \right\}
\]
with \( \hat{\nu}_{k^†, t} = \mu_{k^†} + \sum_{j=1}^{q_{k^†}} \phi_{k^† j}^\tau (y_{t-j} - \mu_{k^†}) - \sum_{j=1}^{q_{k^†}} \theta_{k^† j} \xi_{k^†, t-j} \), and \( \nu_{k^†, t} \) is as defined in Equation (7.8).

Note that the Equation (8.46) has simple form and \( \pi(p | g, y) \) is simply estimated by the proportions of every possible value for \( p \) in the sample obtained by the previous complete MCMC algorithm.

### 8.6.2 Move VII-B: updating of q

This is almost the same as discussed in Section 8.6.1. For convention, we present that discussion very shortly here.

Likewise the previous move, at first we need to choose a component randomly, say \( k^‡ \), in \( \{1, \ldots, g\} \). Then we can write the proposal for the new MA order of the \( k^‡ \)-th component, \( q^‡_{k^‡} \), as:

\[
q^‡_{k^‡} = \begin{cases} 
q_{k^‡} - 1, & \text{with probability } d(q_{k^‡}), \\
q_{k^‡} + 1, & \text{with probability } b(q_{k^‡}), 
\end{cases}
\]  

(8.47)

where \( b(q_k) = 1 - d(q_k) \), for \( k = 1, \ldots, g \), \( d(0) = 0 \) and \( b(q_{\text{max}}) = 0 \).

Let \( \rho^\theta_{k^‡} \) be the proposal vector of the partial autocorrelations. When \( q^‡_{k^‡} = q_{k^‡} - 1 \), the autoregressive coefficients \( \rho^\theta_{k^‡} = (\rho^\theta_{k^‡ 1}, \ldots, \rho^\theta_{k^‡ q^‡_{k^‡} - 1}) \) is updated by simply discarding the last partial autocorrelation in \( \rho^\theta_{k^‡} = (\rho^\theta_{k^‡ 1}, \ldots, \rho^\theta_{k^‡ q_{k^‡}}) \). On the other hand, when \( q^‡_{k^‡} = q_{k^‡} + 1 \), the corresponding autocorrelations \( \rho^\theta_{k^‡} = (\rho^\theta_{k^‡ 1}, \ldots, \rho^\theta_{k^‡ q_{k^‡}}, \rho^\theta_{k^‡ q^‡_{k^‡} + 1}) \) is updated by adding new parameter \( \rho^\theta_{k^‡ q^‡_{k^‡}} \) in \( \rho^\theta_{k^‡} = (\rho^\theta_{k^‡ 1}, \ldots, \rho^\theta_{k^‡ q_{k^‡}}) \) with

\[
\rho^\theta_{k^‡ q^‡_{k^‡}} \overset{\text{ind}}{\sim} B(-1, +1) \left( \rho^\theta_{k^‡ q^‡_{k^‡} - 1}; \left[ \frac{q^‡_{k^‡} - 1}{2} \right], \left[ \frac{q^‡_{k^‡}}{2} \right] + 1 \right).
\]

Then update the corresponding MA coefficients parameters by using the reparametrization discussed in Section 4.5.

Then update all MA parameters by using the reparametrization discussed in Section 4.5.
To update the MA order of the model by using RJMCMC method, we have the acceptance probability

$$A_{q^*} = \min\{1, \Omega_{q^*}\},$$

and $$\Omega_{q^*}$$ can be obtained as in Section 6.6.1:

$$\Omega_{q^*} = \begin{cases} 
L^\frac{b(q^*_{k^*})}{a(q^*_{k^*})} & \text{if } q^*_{k^*} = q_{k^*} - 1, \\
L^\frac{d(q^*_{k^*})}{a(q^*_{k^*})} & \text{if } q^*_{k^*} = q_{k^*} + 1,
\end{cases}$$

(8.48)

where

$$L^\frac{1}{2} = \exp\left[-\frac{\tau_{k^*, t} \hat{\nu}_{k^*, t}}{2} \sum_{t=2}^{k} \left\{(y_t - \hat{\nu}_{k^*, t})^2 - (y_t - \hat{\nu}_{k^*, t})^2\right\}\right]$$

with $$\hat{\nu}_{k^*, t} = \mu_{k^*} + \sum_{j=1}^{p_{k^*}} \phi_{k^*, j} (y_{t-j} - \mu_{k^*}) - \sum_{j=1}^{q_{k^*}} \theta_{k^*, j} \hat{\epsilon}_{k^*, t-j}$$ and $$\nu_{k^*, t}$$ is as defined in Equation (7.8).

Similarly, the posterior density $$\pi(q | g, y)$$ is simply estimated by the proportions of every possible value for $$q$$ in the sample obtained by the previous complete MCMC algorithm.

### 8.6.3 Move VIII-A: updating of r

This is a exactly the same move like we derived for the $$\mathcal{M}_{\text{ARMA-GARCH}}$$ model in Section 5.7.3.

### 8.6.4 Move VIII-B: updating of s

This is a exactly the same move like we derived for the $$\mathcal{M}_{\text{ARMA-GARCH}}$$ model in Section 5.7.4.

### 8.6.5 Move VIII-C: updating of v

We will apply the same technique to update the parameter $$v$$ as discussed in Section 6.6.2 and we got a similar result as mentioned in Section 6.6.2 for the $$\mathcal{M}^v_{\text{ARMA}}$$ model. So, we will discuss this move briefly here.
At first we need to choose a component, say \( k^* \), randomly chosen in \( \{1, \ldots, g\} \). Then the corresponding degrees of freedom \( v_{k^*} \) increases or decreases by one with the following rule:

\[
v_{k^*}^* = \begin{cases} 
  v_{k^*} - 1, & \text{with probability } d(v_{k^*}), \\
  v_{k^*} + 1, & \text{with probability } b(v_{k^*}),
\end{cases}
\]

(8.49)

where \( b(x) = 1 - d(x) \), for \( k = 1, \ldots, g \). \( d(3) = 0 \) and \( b(v_{\text{max}}) = 0 \) because \( v_k \in [3, v_{\text{max}}] \). The function \( b(x) \) is defined in Section 3.7.1.

Let \( \xi^*_t \) be the proposal vector conditional to \( z = k^* \). When \( v_{k^*}^* = v_{k^*} - 1 \), then the corresponding missing variable \( \xi_t^* \mid z_t = k^* \sim (\xi_t; v_{k^*}^* - 1, v_{k^*}^* - 3) \) and when \( v_{k^*}^* = v_{k^*} + 1 \), then \( \xi_t^* \mid z_t = k^* \sim (\xi_t; v_{k^*}^* + 1, v_{k^*}^* - 1) \).

So, the acceptance probability is

\[
A_{v^*} = \min\{1, \Omega_{v^*}\},
\]

and \( \Omega_{v^*} \) is described as in Section 2.7.1:

\[
\Omega_{v^*} = \text{likelihood ratio} \times \text{prio ratio} \times \text{proposal ratio} \times \text{jacobian}.
\]

The likelihood ratio is the same as of Equation (3.21):

\[
\text{likelihood ratio} = \prod_{t: z_t = k^*} \mathcal{N}(y_t; \hat{\nu}_{k^*,t}, (\tau_{k^*,t}^2 \xi_t^*)^{-1}) \mathcal{G}(\xi_t; \frac{v_{k^*}^*}{2}, \frac{v_{k^*}^* - 2}{2}) \nonumber
\]

\[
\prod_{t: z_t = k^*} \mathcal{N}(y_t; \hat{\nu}_{k^*,t}, (\tau_{k^*,t}^2 \xi_t^*)^{-1}) \mathcal{G}(\xi_t; \frac{v_{k^*}^*}{2}, \frac{v_{k^*}^* - 2}{2}) \nonumber
\]

\[
= C_t \exp\left[-\frac{n_{k^*}}{2} \{ (v_{k^*}^* - 2)\xi_t^* - (v_{k^*}^* - 2)\xi_t \}\right],
\]

where \( C_t = \left( \frac{\xi_t}{\xi_t^*} \right)^{n_{k^*}/2} \times \left( \Gamma\left(\frac{v_{k^*}^*}{2}\right) \right)^{n_{k^*} \tau_{k^*,t}^2 / 2} \times \left( \frac{v_{k^*}^* - 2}{2} \right)^{-n_{k^*} \tau_{k^*,t}^2 / 2} \times \frac{\xi_t^{(\tau_{k^*,t}^2 - 1)n_{k^*} / 2}}{\xi_t^*^{(\tau_{k^*,t}^2 - 1)n_{k^*} / 2}}.
\]

The prior ratio is given by:

\[
\text{prior ratio} = \frac{\pi(v_{k^*}^*) \pi(\xi^*_t \mid z_t = k^*)}{\pi(v_{k^*}^*) \pi(\xi_t \mid z_t = k^*)} = \frac{\pi(\xi^*_t \mid z_t = k^*)}{\pi(\xi_t \mid z_t = k^*)}
\]

because of the uniform prior on the AR order of the model and only the \( k^* \)-th component is updated.
To calculate the proposal ratio we have to take into account the probabilities \( b(\cdot) \) and \( d(\cdot) \) and the proposal density, which is chosen equal to the prior on \( \xi \).

\[
\text{proposal ratio} = \begin{cases} 
\frac{b(v^*_k)\pi(\xi_t | z_t = k^*)}{d(v_k)} & \text{if } v^*_k = v_k - 1, \\
\frac{d(v^*_k)}{b(v_k)\pi(\xi_t | z_t = k^*)} & \text{if } v^*_k = v_k + 1.
\end{cases}
\]

Finally, the jacobian is equal to 1 because the invertible function \( g(\cdot) \) is considered as the identity function.

Therefore, multiplying likelihood, prior and proposal ratios, we derive ratio in the acceptance probability for the both the cases:

\[
\Omega_{p^\pm} = \begin{cases} 
D_t \exp \left[ - \frac{n_k}{2} \left\{ (v^*_k - 2)\xi_t^* - (v^*_k - 2)\xi_t \right\} \right] \frac{b(v^*_k)}{d(v_k)} & \text{if } v^*_k = v_k - 1, \\
E_t \exp \left[ - \frac{n_k}{2} \left\{ (v^*_k - 2)\xi_t^* - (v^*_k - 2)\xi_t \right\} \right] \frac{d(v^*_k)}{b(v_k)} & \text{if } v^*_k = v_k + 1,
\end{cases}
\]

(8.50)

where \( D_t = C_t \prod_{t:z_i = k^*} \mathcal{G}(\xi_t; \frac{v^*_k}{2}, \frac{v^*_k - 2}{2}) \) and \( E_t = C_t / \prod_{t:z_i = k^*} \mathcal{G}(\xi_t; \frac{v_k}{2}, \frac{v_k - 2}{2}) \).

Therefore, \( \pi(p | g, y) \) is simply estimated by the proportions of every possible value for \( p \) in the sample obtained by the previous complete MCMC algorithm.

### 8.6.6 Move IX: updating of \( g \)

Finally, we are at the move by which we can complete our Bayesian analysis. According to Bayes’ theorem, the marginal posterior distribution of \( g \) is:

\[
\pi(g | y) \propto f(y | g)\pi(g),
\]

(8.51)

where \( \pi(g) \) is the prior on \( g \) and \( f(y | g) \) is marginal likelihood given by

\[
f(y | g) = \sum_p \sum_q \int f(y | \varphi, p, q, v)\pi(\varphi, p, q, v | g) \, d\varphi
\]

(8.52)

Now we are going to apply the method proposed by Chib and Jeliazkov (2001). First, applying the basic marginal likelihood identity and suppress the model
index \( g \) for notational convenience, we have

\[
f(y) = \frac{f(y|\varphi^*, p^*, q^*, v^*)\pi(\varphi^*, p^*, q^*, v^*)}{\pi(\varphi^*, p^*, q^*, v^* | y)}
\]

\[
= \frac{f(y | \varphi^*, p^*, q^*, v^*)\pi(\varphi^* | p^*, q^*, v^*)\pi(p^*)\pi(q^*)\pi(v^*)}{\pi(\varphi^* | p^*, q^*, v^*, y)\pi(p^* | y)\pi(q^* | y)\pi(v^* | y)}
\]

(8.53)

for any fixed point \((\varphi^*, p^*, q^*, v^*)\) which have the maximum density.

We apply the same methodology and the discussion would be the same as it was the case for \( \mathcal{M}_{\text{ARMA}} \) model. We just need consider another condition parameter \( v \) here and Estimations will be followed by the corresponding equations for this model. For details, see Section 4.7.3.

### 8.7 Examples

We consider the simulated series \( y_{\text{MoIII}} \) as well as the real data \( y_{\text{DGS3}} \) to justify our MCMC methods in Bayesian analysis to fit them with the \( \mathcal{M}_{\text{ARMA-GARCH}} \) model. For this, we complied the Bayesian algorithm described in Section 8.5 and in Section 8.6 with three MCMC chains, each of which run \( N = 30000 \) times for \( y_{\text{MoIII}} \) and \( N = 50000 \) times for \( y_{\text{DGS3}} \) and initial 10000 runs are discarded as burn-in for each series. The initial values, \( \varphi^{(0)} \) of the parameter generated automatically inside the algorithm based on the formulae discussed in Section 8.4.4. Identifiability imposed to solve the label switching problem according after we observes the MCMC samples.

Figure 8.2 represents convergence diagrams and marginal posterior densities of the posterior sample (of one MCMC chain) of the best fitted model \( \mathcal{M}_{\text{ARMA-GARCH}}^t(2; 1, 1; 1, 0; 1, 1; 1, 1; 5, 5) \) for the \( y_{\text{MoIII}} \) series. Similar plots are obtained for the other chains which are not reported here to save space. It is clear from this figure that our Bayesian algorithm captures the asymmetry of the posterior distributions of each parameter of the model. Multi-modality is almost eliminated by imposing the identifiability constrain \( \omega_1 < \omega_2 \).

Table 8.1 shows parameter estimation results of the MCMC samples for the series \( y_{\text{MoIII}} \). It contains mean, median, standard error and HPD\(^1\) intervals of each parameter. Based on these Bayesian inference, we can argue that our Bayesian method fit well with the simulated series \( y_{\text{MoIII}} \).

On the hand, we found \( \mathcal{M}_{\text{ARMA-GARCH}}^t(3; 0, 2, 2; 0, 2, 2; 1, 1, 1; 1, 1, 1; 5, 8, 10) \) be the best fit to the series \( y_{\text{DGS3}} \). Figure 8.3 shows the comparison of convergence

---

\(^1\)Highest Posterior Density intervals
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

Table 8.1: Estimation results of parameters when fitting the $\mathcal{M}_{\text{ARMA-GARCH}}^{t}(2; 1,1; 1,0; 1,1; 1,1; 5,5)$ model to the $y_{\text{MoIII}}$ series with $N = 30000$ and $B = 1000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.6093</td>
<td>0.07654</td>
<td>0.0003423</td>
<td>0.5972</td>
<td>[0.5000022, 0.7499643]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.3907</td>
<td>0.07654</td>
<td>0.0003423</td>
<td>0.4028</td>
<td>[0.2500357, 0.4999978]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.004437</td>
<td>0.1307</td>
<td>0.0005843</td>
<td>-0.01408</td>
<td>[-0.1626482, 0.1600465]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>1.153552</td>
<td>3.7874</td>
<td>0.0169377</td>
<td>0.10428</td>
<td>[-1.5742598, 8.4114492]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.01524</td>
<td>0.08691</td>
<td>0.0003887</td>
<td>-0.01990</td>
<td>[-0.1719672, 0.1690958]</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.11726</td>
<td>0.24277</td>
<td>0.0010857</td>
<td>0.08579</td>
<td>[-0.3107913, 0.5753044]</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.3728</td>
<td>0.2349</td>
<td>0.001051</td>
<td>-0.4282</td>
<td>[-0.7271362, 0.1972014]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.2600</td>
<td>0.5149</td>
<td>0.002303</td>
<td>0.3524</td>
<td>[-0.5835619, 0.9959681]</td>
</tr>
<tr>
<td>$\theta_{11}$</td>
<td>0.05819</td>
<td>0.1991</td>
<td>0.001408</td>
<td>0.0000</td>
<td>[-0.1668725, 0.7198981]</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>[0.0000, 0.0000]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.5236</td>
<td>0.2763</td>
<td>0.0007133</td>
<td>0.5593</td>
<td>[7.875578e-02, 0.9997872]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.4330</td>
<td>0.2058</td>
<td>0.0007379</td>
<td>0.4452</td>
<td>[7.629032e-05, 0.9365904]</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.6012</td>
<td>0.2098</td>
<td>0.0007096</td>
<td>0.6434</td>
<td>[1.172085e-01, 0.9999995]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.5051</td>
<td>0.2116</td>
<td>0.0007426</td>
<td>0.5237</td>
<td>[3.993412e-02, 0.9868172]</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.3122</td>
<td>0.2016</td>
<td>0.001170</td>
<td>0.2804</td>
<td>[1.228451e-06, 0.8477229]</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.2072</td>
<td>0.2207</td>
<td>0.001215</td>
<td>0.1827</td>
<td>[6.540563e-06, 0.8651784]</td>
</tr>
</tbody>
</table>

Computational time: 9.389423 hours

diagrams and marginal posterior densities for the MCMC of $\omega$ and $\mu$ for the series $y_{\text{DGS3}}$. Similarly, Figure 8.4, Figure 8.5, and Figure 8.6 show the convergence diagrams and marginal posterior densities of $\phi$, $\theta$, $\alpha$ and $\beta$ respectively when fitting $\mathcal{M}_{\text{ARMA-GARCH}}$ and the model $\mathcal{M}_{\text{ARMA-GARCH}}^{t}$ to the series $y_{\text{DGS3}}$. Observe that our technique captures the asymmetry of the marginal distribution of the parameter while $\mathcal{M}_{\text{ARMA-GARCH}}^{t}$ model is fitted with the data.

Finally, Table 8.2 presents parameter estimation results when fitting the model $\mathcal{M}_{\text{ARMA-GARCH}}^{t}$ to the series $y_{\text{DGS3}}$ which also justifying the fitness of the prescribed model.
### CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

Table 8.2: Estimation results of parameters when fitting the $\mathcal{M}^t_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 1, 1, 1; 1, 1, 1; 5, 8, 10)$ model to the $y_{\text{DGS3}}$ data with $N = 50000$ and $B = 10000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.3468</td>
<td>0.007695</td>
<td>3.847e-05</td>
<td>0.3456</td>
<td>[0.3343212, 0.3616598]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.3333</td>
<td>0.005414</td>
<td>2.707e-05</td>
<td>0.3333</td>
<td>[0.3223942, 0.3441188]</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.3199</td>
<td>0.007619</td>
<td>3.810e-05</td>
<td>0.3211</td>
<td>[0.3055502, 0.3323756]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.003648</td>
<td>0.1377</td>
<td>0.0006883</td>
<td>-0.008108</td>
<td>[-0.3382992, 0.3120573]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-0.003409</td>
<td>0.1417</td>
<td>0.0007083</td>
<td>-0.008215</td>
<td>[-0.3336749, 0.3323010]</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.003120</td>
<td>0.1437</td>
<td>0.0007183</td>
<td>-0.007209</td>
<td>[-0.3572062, 0.3198515]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.22922</td>
<td>0.6085</td>
<td>0.003042</td>
<td>0.00000</td>
<td>[-0.8473869, 1.7234136]</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.07325</td>
<td>0.4265</td>
<td>0.002133</td>
<td>0.00000</td>
<td>[-0.9984435, 0.6844835]</td>
</tr>
<tr>
<td>$\phi_{31}$</td>
<td>0.24331</td>
<td>0.6223</td>
<td>0.003111</td>
<td>0.00000</td>
<td>[-0.8778677, 1.7292897]</td>
</tr>
<tr>
<td>$\phi_{32}$</td>
<td>-0.07990</td>
<td>0.4358</td>
<td>0.002179</td>
<td>0.00000</td>
<td>[-0.9996383, 0.6915985]</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>0.17656</td>
<td>0.5666</td>
<td>0.002833</td>
<td>0.00000</td>
<td>[-0.9510087, 1.5930183]</td>
</tr>
<tr>
<td>$\theta_{22}$</td>
<td>-0.09213</td>
<td>0.4110</td>
<td>0.002055</td>
<td>0.00000</td>
<td>[-0.9998950, 0.6016461]</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.17422</td>
<td>0.5789</td>
<td>0.002894</td>
<td>0.00000</td>
<td>[-0.9408175, 1.6017361]</td>
</tr>
<tr>
<td>$\theta_{32}$</td>
<td>-0.09728</td>
<td>0.4170</td>
<td>0.002085</td>
<td>0.00000</td>
<td>[-0.9970318, 0.6396732]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.7446</td>
<td>0.2436</td>
<td>0.001218</td>
<td>0.8262</td>
<td>[2.209428e-01, 0.9999803]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.4041</td>
<td>0.2894</td>
<td>0.001447</td>
<td>0.3592</td>
<td>[9.298713e-06, 0.9179191]</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.7490</td>
<td>0.2402</td>
<td>0.001201</td>
<td>0.8300</td>
<td>[2.320545e-01, 0.9999700]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.4018</td>
<td>0.2883</td>
<td>0.001441</td>
<td>0.3566</td>
<td>[9.298713e-06, 0.9165846]</td>
</tr>
<tr>
<td>$\alpha_{31}$</td>
<td>0.7581</td>
<td>0.2356</td>
<td>0.001178</td>
<td>0.8386</td>
<td>[2.497258e-01, 0.9999700]</td>
</tr>
<tr>
<td>$\alpha_{32}$</td>
<td>0.4011</td>
<td>0.2886</td>
<td>0.001443</td>
<td>0.3568</td>
<td>[9.298713e-06, 0.9174603]</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.7840</td>
<td>0.2287</td>
<td>0.001143</td>
<td>0.8691</td>
<td>[0.2630545, 0.9999810]</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.7883</td>
<td>0.2249</td>
<td>0.001125</td>
<td>0.8702</td>
<td>[0.2743795, 0.9999810]</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.7973</td>
<td>0.2190</td>
<td>0.001095</td>
<td>0.8781</td>
<td>[0.2961741, 0.999964]</td>
</tr>
</tbody>
</table>

† Highest Posterior Density intervals.
Figure 8.2: MCMC output of sampled values of parameters when fitting the model $M_{t}^{\text{ARMA-GARCH}}(2; 1, 1; 1, 0; 1, 1; 1, 1; 5, 5)$ to the yMoIII series.
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

Figure 8.3: MCMC output of sampled values of mixing weights and means when fitting $\mathcal{M}_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 2, 2, 2; 1, 1, 1)$ and $\mathcal{M}^t_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 2, 2, 2; 1, 1, 1; 5, 8, 10)$ models to the yDGS3 data with $N = 50000$, in each case burn-in is 10000.
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

Figure 8.4: MCMC output of sampled values of AR coefficients when fitting $\mathcal{M}_{ARMA-GARCH}(3; 0, 2, 2; 0, 2, 2; 2, 2, 2; 1, 1, 1)$ and $\mathcal{M}_{ARMA-GARCH}^t(3; 0, 2, 2; 0, 2, 2; 2, 2, 2; 1, 1, 1; 5, 8, 10)$ models to the yDGS3 data with $N = 50000$, in each case burn-in is 10000.
Figure 8.5: MCMC output of sampled values of MA coefficients when fitting $\mathcal{M}_{\text{ARMA-GARCH}}(3; 0, 2, 2; 0, 2, 2; 2, 2, 2; 1, 1, 1)$ and $\mathcal{M}_{\text{ARMA-GARCH}}'(3; 0, 2, 2; 0, 2, 2; 2, 2, 2; 1, 1, 1; 5, 8, 10)$ models to the yDGS3 data with $N = 50000$, in each case burn-in is 10000.
CHAPTER 8. MIXTURE OF STUDENT-T ARMA-GARCH COMPONENTS

Figure 8.6: MCMC output of sampled values of GARCH coefficients when fitting $M_{ARMA-GARCH}(3; 0, 2, 2; 0, 2, 2; 2, 2, 2; 1, 1, 1)$ and $M_{ARMA-GARCH}^t(3; 0, 2, 2; 2, 2, 2; 1, 1, 1; 5, 8, 10)$ models to the series $y_{DGS3}$ with $N = 50000$, in each case burn-in is 10000.
Chapter 9

Conclusion

We developed complete Bayesian techniques for six different mixture time series models involving various combinations of assumptions on the conditional mean (AR, ARMA), distribution of the error components (Gaussian, Student-t) and GARCH-type effects.

For the $M_{AR}$ model we improved the Sampietro’s algorithm by using precision instead of variance with an additional hierarchical level and changing the moves for the AR parameters so that the whole stationarity region is used.

The choice of component precision with an additional hierarchical level enhanced our Bayesian structures for the mixture models. This idea could be extended for the other parameters as well.

To the best of our knowledge, the complete MCMC approach has not been considered for these models before, except for the $M_{AR}$ model. The models $M_{ARMA}^t$ and $M_{ARMA-GARCH}^t$ are new that we proposed in this thesis.

We used component-wise technique to derive univariate posterior distributions of the parameters while multivariate posteriors could be generated. Combination of Gibbs and Metropolis-Hastings moves played a very important role in our Bayesian treatment of dealing with mixture models.

We mainly focused on developing Bayesian methods for our models and presented properties of the parameter estimates, such as posterior densities, means, HPDs. We also studied the convergence properties. Further Bayesian inferences can be carried out easily by using the MCMC samples that have been already generated for each model, e.g. prediction of the conditional volatilities with credible intervals.

The complexity of the models has made programming the methods very demanding. Also special care was needed for numerical details not obvious from the formulae. We plan to develop the R code further as R packages.
Appendix A

MCMC output

In this appendix we present some tables and figures in support of the results discussed in this thesis.

Table A.1: Acceptance Rate for Metropolis-Hastings chains of AR coefficients of the fitted $\mathcal{M}_{AR}^t(2; 1, 1; 4, 8)$ model for the simulated $y\text{MoI}$ series.

<table>
<thead>
<tr>
<th>Chain</th>
<th>$\phi_{11}$</th>
<th>$\phi_{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>chain1</td>
<td>0.2598026</td>
<td>0.2121421</td>
</tr>
<tr>
<td>chain2</td>
<td>0.2603326</td>
<td>0.2118221</td>
</tr>
<tr>
<td>chain3</td>
<td>0.2634126</td>
<td>0.2099721</td>
</tr>
</tbody>
</table>
Table A.2: The Geweke diagnostic test results of the $M_{ARMA}(2; 1, 1; 0, 0)$ model fitted with the simulated $y_{MoI}$ series.

<table>
<thead>
<tr>
<th></th>
<th>chain1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$ Z-scores</td>
<td>-1.250525</td>
<td>1.61417</td>
<td>-1.953361</td>
</tr>
<tr>
<td>$w_2$ Z-scores</td>
<td>1.250525</td>
<td>-1.61417</td>
<td>1.953361</td>
</tr>
<tr>
<td>Window From Start</td>
<td>0.100000</td>
<td>0.64873</td>
<td>0.969310</td>
</tr>
<tr>
<td>Window From Stop</td>
<td>0.500000</td>
<td>0.32530</td>
<td>0.019920</td>
</tr>
</tbody>
</table>

********** The Geweke diagnostic: *******

<table>
<thead>
<tr>
<th></th>
<th>Point est.</th>
<th>Upper C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$ PSRF</td>
<td>1.01</td>
<td>1.03</td>
</tr>
<tr>
<td>$w_2$ PSRF</td>
<td>1.01</td>
<td>1.03</td>
</tr>
<tr>
<td>Multivariate PSRF</td>
<td>1.01</td>
<td></td>
</tr>
</tbody>
</table>
Table A.3: Estimation results of parameters of the $\mathcal{M}_{AR}(3; 0, 6, 7)$ model fitted with the $yDGS3$ data with $N = 100000$ and $B = 10000$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.d.</th>
<th>s.e.</th>
<th>Median</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.7370</td>
<td>0.02377</td>
<td>$7.516 \times 10^{-5}$</td>
<td>0.7380</td>
<td>[0.69062775, 0.7828183]</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.1432</td>
<td>0.01671</td>
<td>$5.284 \times 10^{-5}$</td>
<td>0.1415</td>
<td>[0.11221517, 0.1765267]</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.1199</td>
<td>0.01261</td>
<td>$3.986 \times 10^{-5}$</td>
<td>0.1195</td>
<td>[0.09533645, 0.1445742]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.0004369</td>
<td>0.0003324</td>
<td>$1.051 \times 10^{-6}$</td>
<td>-0.00044</td>
<td>[-0.0010873, 0.00021561]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-0.0135820</td>
<td>0.0846504</td>
<td>$2.677 \times 10^{-5}$</td>
<td>-0.0214e-05</td>
<td>[-0.23140630, 0.135191997]</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.0089255</td>
<td>0.0859090</td>
<td>$2.717 \times 10^{-5}$</td>
<td>0.0598e-04</td>
<td>[-0.22506414, 0.154435717]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.00044</td>
<td>0.0003324</td>
<td>$1.051 \times 10^{-6}$</td>
<td>-0.00044</td>
<td>[-0.0010873, 0.00021561]</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.00010</td>
<td>0.0025975</td>
<td>$8.214 \times 10^{-6}$</td>
<td>-0.00002</td>
<td>[-0.0049522, 0.00536703]</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.00092</td>
<td>0.0029047</td>
<td>$9.185 \times 10^{-6}$</td>
<td>0.00078</td>
<td>[-0.0047580, 0.00670735]</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.01112</td>
<td>0.0004445</td>
<td>$1.406 \times 10^{-6}$</td>
<td>0.01111</td>
<td>[0.01025996, 0.01198882]</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.02705</td>
<td>0.0032909</td>
<td>$1.401 \times 10^{-5}$</td>
<td>0.02692</td>
<td>[0.02091435, 0.03315618]</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.02907</td>
<td>0.0031647</td>
<td>$1.001 \times 10^{-5}$</td>
<td>0.02913</td>
<td>[0.02091435, 0.03315618]</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.11283</td>
<td>0.1593</td>
<td>0.0005039</td>
<td>0.115052</td>
<td>[-0.20087196, 0.4601531]</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.15975</td>
<td>0.2213</td>
<td>0.0006998</td>
<td>-0.187102</td>
<td>[-0.57563059, 0.2551848]</td>
</tr>
<tr>
<td>$\phi_{23}$</td>
<td>0.05079</td>
<td>0.2132</td>
<td>0.0006742</td>
<td>0.072479</td>
<td>[-0.33880390, 0.4754793]</td>
</tr>
<tr>
<td>$\phi_{24}$</td>
<td>0.03329</td>
<td>0.2296</td>
<td>0.0007260</td>
<td>0.061784</td>
<td>[-0.40655264, 0.4415826]</td>
</tr>
<tr>
<td>$\phi_{25}$</td>
<td>0.10922</td>
<td>0.4875</td>
<td>0.0015416</td>
<td>0.331076</td>
<td>[-0.70526794, 0.7293053]</td>
</tr>
<tr>
<td>$\phi_{26}$</td>
<td>0.04711</td>
<td>0.1804</td>
<td>0.0005705</td>
<td>0.063780</td>
<td>[-0.28719359, 0.3743707]</td>
</tr>
<tr>
<td>$\phi_{31}$</td>
<td>0.08285</td>
<td>0.1729</td>
<td>0.0005467</td>
<td>0.08681</td>
<td>[-0.25346357, 0.4560569]</td>
</tr>
<tr>
<td>$\phi_{32}$</td>
<td>-0.01380</td>
<td>0.2426</td>
<td>0.0007673</td>
<td>0.001218</td>
<td>[-0.48954959, 0.4396667]</td>
</tr>
<tr>
<td>$\phi_{33}$</td>
<td>-0.02548</td>
<td>0.2270</td>
<td>0.0007178</td>
<td>-0.048129</td>
<td>[-0.45848040, 0.4035403]</td>
</tr>
<tr>
<td>$\phi_{34}$</td>
<td>-0.08208</td>
<td>0.2445</td>
<td>0.0007731</td>
<td>-0.098293</td>
<td>[-0.50604502, 0.4090841]</td>
</tr>
<tr>
<td>$\phi_{35}$</td>
<td>-0.10627</td>
<td>0.5114</td>
<td>0.0016172</td>
<td>-0.330393</td>
<td>[-0.77122389, 0.7447398]</td>
</tr>
<tr>
<td>$\phi_{36}$</td>
<td>0.06002</td>
<td>0.1868</td>
<td>0.0005908</td>
<td>0.064368</td>
<td>[-0.28049132, 0.4137606]</td>
</tr>
<tr>
<td>$\phi_{37}$</td>
<td>0.06133</td>
<td>0.1002</td>
<td>0.0003169</td>
<td>0.064368</td>
<td>[-0.05462484, 0.3066245]</td>
</tr>
</tbody>
</table>
Figure A.1: MCMC output of parameters while fitting the $M_{AR}(3; 0, 6, 7)$ model to the $y_{DGS3}$ data, using priors discussed in Section 3.5.
Figure A.2: MCMC output parameters while fitting the $\mathcal{M}_{\text{ARMA}}(2; 1, 1; 1, 0)$ model to the simulated $y_{\text{MoI}}$ series.
Table A.4: The Heidelberger-Welch diagnostic results of the parameter from MCMC chains of the $\mathcal{M}_{\text{ARMA}}(2;1,1;0,0)$ model fitted with the simulated $y_{\text{MoI}}$ series.

<table>
<thead>
<tr>
<th>Chain</th>
<th>epsilon</th>
<th>alpha</th>
<th>Stationarity start</th>
<th>p-value</th>
<th>iteration</th>
<th>test</th>
<th>Halfwidth</th>
<th>Mean</th>
<th>Halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.05</td>
<td>w_1 passed</td>
<td>1</td>
<td>0.147</td>
<td></td>
<td>0.596</td>
<td>0.000782</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>w_2 passed</td>
<td>1</td>
<td>0.147</td>
<td></td>
<td>0.404</td>
<td>0.000782</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.093</td>
<td>0.1</td>
<td>w_1 passed</td>
<td>1</td>
<td>0.305</td>
<td></td>
<td>0.589</td>
<td>0.000806</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>w_2 passed</td>
<td>1</td>
<td>0.305</td>
<td></td>
<td>0.411</td>
<td>0.000806</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
<td>0.05</td>
<td>w_1 passed</td>
<td>9001</td>
<td>0.201</td>
<td></td>
<td>0.597</td>
<td>0.000839</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>w_2 passed</td>
<td>9001</td>
<td>0.201</td>
<td></td>
<td>0.403</td>
<td>0.000839</td>
<td></td>
</tr>
</tbody>
</table>

********** The Heidelberger-Welch diagnostic: **********
Table A.5: The Raftery-Lewis diagnostic results of parameters while fitting the $M_{ARMA}(2; 1; 1; 0, 0)$ model to the simulated $y_{MoI}$ series, using number of chains =3, $N = 100000$ and burn-in = 10000.

| Chain 1, converge.eps = 0.001 | Quantile (q) = 0.025 | Accuracy (r) = +/- 0.005 | Probability (s) = 0.95 |
| Chain 2, converge.eps = 0.001 | Quantile (q) = 0.01 | Accuracy (r) = +/- 0.005 | Probability (s) = 0.95 |
| Chain 3, converge.eps = 0.005 | Quantile (q) = 0.25 | Accuracy (r) = +/- 0.005 | Probability (s) = 0.999 |

<table>
<thead>
<tr>
<th>Burn-in</th>
<th>Total</th>
<th>Lower bound</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>w_1</td>
<td>9</td>
<td>12990</td>
<td>3746</td>
</tr>
<tr>
<td>w_2</td>
<td>12</td>
<td>17520</td>
<td>3746</td>
</tr>
<tr>
<td>w_1</td>
<td>4</td>
<td>3348</td>
<td>1522</td>
</tr>
<tr>
<td>w_2</td>
<td>9</td>
<td>5112</td>
<td>1522</td>
</tr>
<tr>
<td>w_1</td>
<td>12</td>
<td>428448</td>
<td>81207</td>
</tr>
<tr>
<td>w_2</td>
<td>15</td>
<td>516455</td>
<td>81207</td>
</tr>
</tbody>
</table>
Table A.6: Diagnostic results of the parameter $\sigma$ of the MCMC chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted $\mathcal{M}_{AR}^t(2; 1, 1; 4, 8)$ model for the simulated $y_{MoI}$ series.

===============================================
********** The Geweke diagnostic: **********
Z-scores:

<table>
<thead>
<tr>
<th></th>
<th>chain1</th>
<th>chain2</th>
<th>chain3</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>-0.6212057</td>
<td>-0.6364944</td>
<td>-2.6027242</td>
</tr>
<tr>
<td>gamma_2</td>
<td>0.7534821</td>
<td>0.5040581</td>
<td>0.3609124</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Window From Start</td>
<td>0.1000000</td>
<td>0.8599500</td>
<td>0.5027000</td>
</tr>
<tr>
<td>Window From Stop</td>
<td>0.5000000</td>
<td>0.0143400</td>
<td>0.4602200</td>
</tr>
</tbody>
</table>

********** The Gelman-Rubin diagnostic: **********
Point est. Upper C.I.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>gamma_2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Multivariate psrf: 1

********** The Heidelberger-Welch diagnostic: **********
Chain 1, epsilon=0.1, alpha=0.05

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1 passed</td>
<td>1</td>
<td>0.516</td>
</tr>
<tr>
<td>gamma_2 passed</td>
<td>1</td>
<td>0.613</td>
</tr>
</tbody>
</table>

Halfwidth Mean Halfwidth

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1 passed</td>
<td>-0.0835</td>
<td>0.000904</td>
</tr>
<tr>
<td>gamma_2 passed</td>
<td>-0.0588</td>
<td>0.001644</td>
</tr>
</tbody>
</table>

********** The Raftery-Lewis diagnostic: **********
converge.eps = 0.001, Quantile (q) = 0.025
Accuracy (r) = +/- 0.005, Probability (s) = 0.95

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>3</td>
<td>4418</td>
<td>3746</td>
<td>1.18</td>
</tr>
<tr>
<td>gamma_2</td>
<td>6</td>
<td>8862</td>
<td>3746</td>
<td>2.37</td>
</tr>
</tbody>
</table>
Table A.7: Diagnostic results of the parameter $\gamma$ of the MCMC chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted $\mathcal{M}_{\text{ARMA}}(2; 1, 1, 0; 1, 1, 1)$ model for the yIBM data.

<table>
<thead>
<tr>
<th>Z-scores:</th>
<th>chain1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>0.6340755</td>
<td>1.2132538</td>
<td>0.6971038</td>
</tr>
<tr>
<td>gamma_2</td>
<td>-1.2447421</td>
<td>-0.4004509</td>
<td>-2.3756834</td>
</tr>
<tr>
<td>gamma_3</td>
<td>0.6402923</td>
<td>-1.9485325</td>
<td>1.9518327</td>
</tr>
</tbody>
</table>

Window From Start: 0.1000000, 0.8462500, 0.1100900
Window From Stop: 0.5000000, 0.0348500, 0.1463400

<table>
<thead>
<tr>
<th>Point est. Upper C.I.</th>
<th>gamma_1</th>
<th>gamma_2</th>
<th>gamma_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>gamma_2</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>gamma_3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Multivariate psrf: 1

<table>
<thead>
<tr>
<th>Stationarity start p-value: test iteration &quot;</th>
<th>gamma_1 passed 1 0.612</th>
<th>gamma_2 passed 1 0.608</th>
<th>gamma_3 passed 1 0.953</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halfwidth Mean Halfwidth: test</td>
<td>gamma_1 failed 0.190 0.0408</td>
<td>gamma_2 failed -0.482 0.1046</td>
<td>gamma_3 passed -2.205 0.1325</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The Raftery-Lewis diagnostic: **********</th>
<th>Chain 1, converge.eps = 0.001, Quantile (q) = 0.025</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (r) = +/- 0.005, Probability (s) = 0.95</td>
<td></td>
</tr>
<tr>
<td>Burn-in (M)</td>
<td>Total (N)</td>
</tr>
<tr>
<td>gamma_1 12</td>
<td>18564</td>
</tr>
<tr>
<td>gamma_2 16</td>
<td>20008</td>
</tr>
<tr>
<td>gamma_3 6</td>
<td>8822</td>
</tr>
</tbody>
</table>
### Table A.8: Diagnostic results of the parameter $\sigma$ of the MCMC chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted $\mathcal{M}_{\text{ARMA}}(2; 1, 1, 0; 1, 1, 1)$ model for the simulated $y_{\text{MoI}}$ series.

<table>
<thead>
<tr>
<th></th>
<th>chain1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geweke diagnostic:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z-scores:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>$-0.05474329$</td>
<td>$0.3127165$</td>
<td>$-2.1611471$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>$0.55369517$</td>
<td>$-1.2173744$</td>
<td>$1.1820576$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>$0.10455222$</td>
<td>$1.3697969$</td>
<td>$0.6747889$</td>
</tr>
<tr>
<td>Window From Start</td>
<td>$0.10000000$</td>
<td>$0.5222900$</td>
<td>$0.1855100$</td>
</tr>
<tr>
<td>Window From Stop</td>
<td>$0.50000000$</td>
<td>$0.2844000$</td>
<td>$0.3712300$</td>
</tr>
<tr>
<td><strong>Gelman-Rubin diagnostic:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potential scale reduction factors:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
</tr>
<tr>
<td><strong>Heidelberger-Welch diagnostic:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\epsilon=0.1$, $\alpha=0.05$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stationarity test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>passed</td>
<td>$1$</td>
<td>$0.697$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>passed</td>
<td>$1$</td>
<td>$0.880$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>passed</td>
<td>$1$</td>
<td>$0.936$</td>
</tr>
<tr>
<td>Halfwidth Mean test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>passed</td>
<td>$5.40$</td>
<td>$0.0316$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>passed</td>
<td>$6.47$</td>
<td>$0.0937$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>passed</td>
<td>$11.16$</td>
<td>$0.1035$</td>
</tr>
<tr>
<td><strong>Raftery-Lewis diagnostic:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{Chain 1, }$</td>
<td>$\text{converge.eps = 0.005, Quantile (q) = 0.001}$</td>
<td>$\text{Accuracy (r) = +/- 0.001, Probability (s) = 0.9}$</td>
<td></td>
</tr>
<tr>
<td>Burn-in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>$3$</td>
<td>$3301$</td>
<td>$2703$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>$3$</td>
<td>$3781$</td>
<td>$2703$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>$4$</td>
<td>$4346$</td>
<td>$2703$</td>
</tr>
</tbody>
</table>
Table A.9: Diagnostic results of the parameter $\theta$ of the mcmc chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted $\mathcal{M}_{\text{ARMA}}(2; 1, 1, 0; 1, 1, 1)$ model for the simulated $y_{\text{MoI}}$ series.

========================================================

********** The Geweke diagnostic: **********

Z-scores:

<table>
<thead>
<tr>
<th></th>
<th>chain1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{11}$</td>
<td>0.21</td>
<td>4.86</td>
<td>0.16</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>0.69</td>
<td>0.43</td>
<td>0.23</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>-0.28</td>
<td>1.53</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Window From Start 0.100000 0.0114600 0.0428800
Window From Stop 0.500000 0.2339300 0.2405800

********** The Gelman-Rubin diagnostic: **********

Potential scale reduction factors:

<table>
<thead>
<tr>
<th></th>
<th>Point est.</th>
<th>Upper C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{11}$</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>1</td>
<td>1.01</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>1</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Multivariate psrf: 1.01

********** The Heidelberger-Welch diagnostic: **********

Chain 1, epsilon=0.1, alpha=0.05

<table>
<thead>
<tr>
<th></th>
<th>Stationarity</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{11}$</td>
<td>passed</td>
<td>0.931</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>passed</td>
<td>0.937</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>passed</td>
<td>0.945</td>
</tr>
</tbody>
</table>

Halfwidth test

<table>
<thead>
<tr>
<th></th>
<th>Halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{11}$</td>
<td>-0.0264 0.0218</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>0.1301 0.0466</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>-0.0645 0.0701</td>
</tr>
</tbody>
</table>

********** The Raftery-Lewis diagnostic: **********

Chain 1, converge.eps = 0.001, Quantile (q) = 0.025

Accuracy ($r$) = +/- 0.005, Probability (s) = 0.95

Burn-in Total Lower bound Dependence

<table>
<thead>
<tr>
<th></th>
<th>(M)</th>
<th>(N)</th>
<th>(Nmin)</th>
<th>factor (I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{11}$</td>
<td>32</td>
<td>35640</td>
<td>3746</td>
<td>3.51</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>42</td>
<td>47201</td>
<td>3746</td>
<td>2.60</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>30</td>
<td>37110</td>
<td>3746</td>
<td>3.91</td>
</tr>
</tbody>
</table>

========================================================
Table A.10: Diagnostic results of the parameter $\phi$ of the MCMC chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted $\mathcal{M}_t^k(2; 1, 1; 4, 8)$ model for the simulated $y_{MoI}$ series.

<table>
<thead>
<tr>
<th></th>
<th>chain1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>phi_11</td>
<td>1.0316398</td>
<td>0.7829874</td>
<td>-0.1230307</td>
</tr>
<tr>
<td>phi_21</td>
<td>-0.7704723</td>
<td>-0.7011745</td>
<td>-0.1759021</td>
</tr>
<tr>
<td>Window From Start</td>
<td>0.1000000</td>
<td>0.3774400</td>
<td>0.3971000</td>
</tr>
<tr>
<td>Window From Stop</td>
<td>0.5000000</td>
<td>0.0819000</td>
<td>0.4421000</td>
</tr>
</tbody>
</table>

********** The Geweke diagnostic: **********

<table>
<thead>
<tr>
<th></th>
<th>chain1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point est. Upper C.I.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phi_11</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>phi_21</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Multivariate psrf: 1

********** The Gelman-Rubin diagnostic: **********

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationarity start</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phi_11 passed</td>
<td>18001</td>
<td>0.153</td>
<td></td>
</tr>
<tr>
<td>phi_21 passed</td>
<td>18001</td>
<td>0.071</td>
<td></td>
</tr>
</tbody>
</table>

Halfwidth Mean Halfwidth test

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Halfwidth Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phi_11 passed</td>
<td>-0.35</td>
<td>0.00426</td>
<td></td>
</tr>
<tr>
<td>phi_21 passed</td>
<td>0.92</td>
<td>0.00518</td>
<td></td>
</tr>
</tbody>
</table>

********** The Heidelberger-Welch diagnostic: **********

epsilon=0.1, alpha=0.05

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationarity test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phi_11 passed</td>
<td>18001</td>
<td>0.153</td>
<td></td>
</tr>
<tr>
<td>phi_21 passed</td>
<td>18001</td>
<td>0.071</td>
<td></td>
</tr>
</tbody>
</table>

Halfwidth test

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Halfwidth test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phi_11 passed</td>
<td>-0.35</td>
<td>0.00426</td>
<td></td>
</tr>
<tr>
<td>phi_21 passed</td>
<td>0.92</td>
<td>0.00518</td>
<td></td>
</tr>
</tbody>
</table>

********** The Raftery-Lewis diagnostic: **********

converge.eps = 0.001, Quantile (q) = 0.025

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (r) = +/-</td>
<td>0.005</td>
<td>Probability (s)</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Burn-in (M)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phi_11 24</td>
<td>24510</td>
<td>3746</td>
<td>4.54</td>
</tr>
<tr>
<td>phi_21 12</td>
<td>13605</td>
<td>3746</td>
<td>3.63</td>
</tr>
</tbody>
</table>
APPENDIX A. MCMC OUTPUT

Table A.11: Diagnostic results of the parameter $\omega$ of the mcmc chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted model $\mathcal{M}_{\text{ARMA}}^{t}(2; 1, 1; 1, 0; 4, 0)$ for the simulated series $y_{MoI}$

==============================================
******** The Geweke diagnostic: *********
Z-scores:
  chain1 chain 2 chain 3
$w_1$ 0.7272963 -0.9059722 0.3017725
$w_2$ -0.7272963 0.9059722 -0.3017725
Window From Start 0.1000000 0.5365500 0.3349200
Window From Stop 0.5000000 0.3208400 0.3246400

******** The Gelman-Rubin diagnostic: ********
Point est. Upper C.I.
$w_1$ 1.15 1.44
$w_2$ 1.15 1.44

Multivariate psrf: 1.14

******** The Heidelberger-Welch diagnostic: ********
Chain 1, epsilon=0.1, alpha=0.05
  Stationarity start $p$-value
  test iteration
$w_1$ passed 1 0.291
$w_2$ passed 1 0.291

  Halfwidth Mean Halfwidth
  test
$w_1$ passed 0.551 0.000354
$w_2$ passed 0.449 0.000354

******** The Raftery-Lewis diagnostic: *********
Chain 1, converge.eps = 0.001
Quantile (q) = 0.025"
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

  Burn-in Total Lower bound Dependence
  (M) (N) (Nmin) factor (I)
$w_1$ 3 4024 3746 1.07
$w_2$ 9 12420 3746 3.32

================================================
Table A.12: Diagnostic results of the parameter $\gamma$ of the MCMC chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted $\mathcal{M}_{\text{ARMA}}^t(2; 1, 1; 1, 0; 4, 8)$ model for the simulated $\text{yMoI}$ series.

---

****** The Geweke diagnostic: ******

Z-scores:

<table>
<thead>
<tr>
<th></th>
<th>chain 1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>-0.5476716</td>
<td>-1.37713211</td>
<td>0.5041446</td>
</tr>
<tr>
<td>gamma_2</td>
<td>-0.2179390</td>
<td>-0.02032288</td>
<td>0.4254284</td>
</tr>
</tbody>
</table>

Window From Start 0.1000000 0.87664000 0.90169000
Window From Stop 0.5000000 0.02989000 0.06323000

****** The Gelman-Rubin diagnostic: ******

Potential scale reduction factors:

<table>
<thead>
<tr>
<th></th>
<th>Point est.</th>
<th>Upper C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>1</td>
<td>1.01</td>
</tr>
<tr>
<td>gamma_2</td>
<td>1</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Multivariate psrf: 1

****** The Heidelberger-Welch diagnostic: ******

Chain 1, epsilon=0.1, alpha=0.05

<table>
<thead>
<tr>
<th></th>
<th>Stationarity start</th>
<th>p-value</th>
<th>iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>passed 1</td>
<td>0.553</td>
<td></td>
</tr>
<tr>
<td>gamma_2</td>
<td>passed 1</td>
<td>0.839</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Halfwidth Mean</th>
<th>Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>-0.1616</td>
<td>0.00170</td>
</tr>
<tr>
<td>gamma_2</td>
<td>-0.0814</td>
<td>0.00188</td>
</tr>
</tbody>
</table>

****** The Raftery-Lewis diagnostic: ******

Chain 1, converge.eps = 0.001

<table>
<thead>
<tr>
<th></th>
<th>Burn-in</th>
<th>Total</th>
<th>Lower bound</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(N)</td>
<td>(Nmin)</td>
<td>factor (I)</td>
</tr>
<tr>
<td>gamma_1</td>
<td>3</td>
<td>4286</td>
<td>3746</td>
<td>1.14</td>
</tr>
<tr>
<td>gamma_2</td>
<td>2</td>
<td>3965</td>
<td>3746</td>
<td>1.06</td>
</tr>
</tbody>
</table>
Table A.13: Diagnostic results of the parameter $\sigma$ of the MCMC chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted $\mathcal{M}^t_{\text{ARMA}}(2; 1, 1; 1, 0; 4, 8)$ model for the $\text{yMoI}$ series.

<table>
<thead>
<tr>
<th></th>
<th>Z-scores</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>chain1</td>
<td>chain 2</td>
<td>chain 3</td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>-0.901574</td>
<td>5.9788190</td>
<td>0.5883798</td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>1.339764</td>
<td>-0.4231879</td>
<td>-0.5881939</td>
<td></td>
</tr>
<tr>
<td>Window From Start</td>
<td>0.100000</td>
<td>0.0034600</td>
<td>0.6446800</td>
<td></td>
</tr>
<tr>
<td>Window From Stop</td>
<td>0.500000</td>
<td>0.5866300</td>
<td>0.0245100</td>
<td></td>
</tr>
</tbody>
</table>

********** The Geweke diagnostic: **********

<table>
<thead>
<tr>
<th></th>
<th>Point est. Upper C.I.</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1$</td>
<td>1.20</td>
<td>1.59</td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>1.09</td>
<td>1.20</td>
<td></td>
</tr>
</tbody>
</table>

Multivariate psrf
1.07

********** The Gelman-Rubin diagnostic: **********

<table>
<thead>
<tr>
<th></th>
<th>Stationarity start p-value</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1$</td>
<td>passed</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>passed</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Halfwidth Mean Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1$</td>
<td>passed</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>passed</td>
</tr>
</tbody>
</table>

********** The Heidelberger-Welch diagnostic: **********

<table>
<thead>
<tr>
<th></th>
<th>Burn-in (M)</th>
<th>Total (N)</th>
<th>Lower bound (Nmin)</th>
<th>Dependence factor (I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1$</td>
<td>6</td>
<td>11793</td>
<td>3746</td>
<td>3.15</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>6</td>
<td>8552</td>
<td>3746</td>
<td>2.28</td>
</tr>
</tbody>
</table>

********** The Raftery-Lewis diagnostic: **********
Table A.14: Diagnostic results of the parameter $\phi$ of the MCMC chains (chain 1 is reported here among 3 chains, each has $N = 100000$ with $B = 10000$) of the fitted $\mathcal{M}^t_{\text{ARMA}}(2; 1, 1; 1, 0; 4, 8)$ model for the $\text{yMoI}$ series.

<table>
<thead>
<tr>
<th></th>
<th>chain 1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{11}$</td>
<td>-1.0166238</td>
<td>5.53037497</td>
<td>2.9529599</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.9543418</td>
<td>-0.04239744</td>
<td>-0.2069622</td>
</tr>
<tr>
<td>Window From Start</td>
<td>0.1000000</td>
<td>0.56129000</td>
<td>0.5109000</td>
</tr>
<tr>
<td>Window From Stop</td>
<td>0.5000000</td>
<td>0.17748000</td>
<td>0.0197200</td>
</tr>
</tbody>
</table>

******* The Geweke diagnostic: *******

Z-scores:

<table>
<thead>
<tr>
<th></th>
<th>chain 1</th>
<th>chain 2</th>
<th>chain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{11}$</td>
<td>-1.0166238</td>
<td>5.53037497</td>
<td>2.9529599</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.9543418</td>
<td>-0.04239744</td>
<td>-0.2069622</td>
</tr>
</tbody>
</table>

Window From Start | 0.1000000 | 0.56129000 | 0.5109000 |
Window From Stop  | 0.5000000 | 0.17748000 | 0.0197200 |

******* The Gelman-Rubin diagnostic: *******

Point est. Upper C.I.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{11}$</td>
<td>1.10</td>
<td>1.21</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>1.16</td>
<td>1.39</td>
</tr>
</tbody>
</table>

Multivariate psrf: 1.05

******* The Heidelberger-Welch diagnostic: *******

Chain 1, epsilon=0.1, alpha=0.05

<table>
<thead>
<tr>
<th></th>
<th>test</th>
<th>iteration</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{11}$</td>
<td>passed</td>
<td>1</td>
<td>0.441</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>passed</td>
<td>1</td>
<td>0.552</td>
</tr>
</tbody>
</table>

Halfwidth Mean     | Halfwidth Mean     |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{11}$ passed</td>
<td>-0.350 0.00435</td>
</tr>
<tr>
<td>$\phi_{21}$ passed</td>
<td>0.821 0.00453</td>
</tr>
</tbody>
</table>

******* The Raftery-Lewis diagnostic: *******

Chain 1, converge.eps = 0.001
Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

<table>
<thead>
<tr>
<th></th>
<th>Burn-in</th>
<th>Total</th>
<th>Lower bound</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(N)</td>
<td>(Nmin)</td>
<td>factor (I)</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>24</td>
<td>26136</td>
<td>3746</td>
<td>4.98</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>12</td>
<td>16540</td>
<td>3746</td>
<td>4.42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
Bibliography


