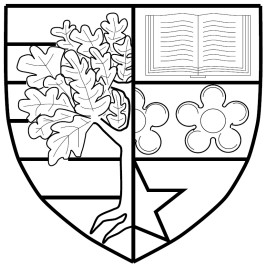
Small Population Bias and Sampling Effects in Stochastic Mortality Modelling

Liang Chen

Submitted for the degree of

Doctor of Philosophy

Heriot-Watt University



Department of Actuarial Mathematics & Statistics,

School of Mathematical and Computer Sciences.

July 2017.

Abstract

Pension schemes are facing more difficulties on matching their underlying liabilities with assets, mainly due to faster mortality improvements for their underlying populations, better environments and medical treatments and historically low interest rates. Given most of the pension schemes are relatively much smaller than the national population, modelling and forecasting the small populations' longevity risk become urgent tasks for both the industrial practitioners and academic researchers. This thesis starts with a systematic analysis on the influence of population size on the uncertainties of mortality estimates and forecasts with a stochastic mortality model, based on a parametric bootstrap methodology with England and Wales males as our benchmark population. The population size has significant effect on the uncertainty of mortality estimates and forecasts. The volatilities of small populations are over-estimated by the maximum likelihood estimators. A Bayesian model is developed to improve the estimation of the volatilities and the predictions of mortality rates for the small populations by employing the information of larger population with informative prior distributions. The new model is validated with the simulated small death scenarios. The Bayesian methodologies generate smoothed estimations for the mortality rates. Moreover, a methodology is introduced to use the information of large population for obtaining unbiased volatilities estimations given the underlying prior settings. At last, an empirical study is carried out based on the Scotland mortality dataset.

KEYWORDS — Small population, age effect, period effect, cohort effect, bootstrap, parameter uncertainty, systematic parameter difference, likelihood ratio test, power of test, Bayesian inference, MCMC, informative prior, longevity risk, sampling variation, ARIMA

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Prof. Andrew Cairns and Dr. Torsten Kleinow for their continuous support of my Ph.D study and relative research, for their patience, encouragement, enlightenment and professionalism, for their immense knowledge, meticulous and rigorous scholarship. One million thanks also go to Dr. George Streftaris for examining my Ph.D progress every year.

My sincere thanks also goes to my industry supervisor Mr. Steven Baxter from Hymans Robertson for his generous support and precious suggestions in this research project.

I would like to acknowledge the Actuarial Research Centre, the Faculty of Actuaries Endowment Fund and the Institute and Faculty of Actuaries for funding this research and providing such a great opportunity to work on this cutting edge topic on actuarial science together with so many smart and talented brains, for financially supporting and providing all those chances where I have presented our research findings all around the world. It has always been an honour to be a CERA of the Institute and Faculty of Actuaries where I pursue my actuarial researching career. I thank Mr. Kevin Mclver from the Institute and Faculty of Actuaries for his continuous coordinating work for the faculties funded Ph.D students.

I appreciate all the academics in our Department of Actuarial Mathematics and Statistics, whom I have worked with and learned from since I started my Ph.D

research. I also appreciate all the administrative staffs in the School of Mathematics and Computer Science for improving the life for the Ph.D students. I thank Heriot Watt University for offering the facilities and services for my research.

Many thanks go to my office mates. It was a pleasure to share an office with these mathematicians and statisticians, Jeff Pollock, Ruairi Donnelly, Efthalia Tzitzili, Max Lau. Particular thanks to Jeff for being my best mate and shining me a light about using the Hamiltonian Monte Carlo, which saves lots of my time.

One million thanks must also go to the University College Cork where I finished my undergraduate degree and University of Kent where I pursued my master degree, for the knowledges they delivered to me and the unforgettable campus life.

Last but not least, tremendous thanks to my family members. I could not thank more to my parents for their supporting and understanding my decisions. Thanks to my daughter, your birth gives me great motivation for keeping on my research. Thanks to my wife for her respect and taking good care of our daughter so that I can concentrate on my research.

Contents

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Thesis Outline 5

2 Mortality Model and Notations **8**

2.1 Introduction to the Parametric Mortality Models 8

2.2 Introduction to the Stochastic Mortality Models 10

2.2.1 Family of Lee-Carter Models 10

2.2.2 Age-Period-Cohort Model 13

2.2.3 P-Spline Method 14

2.2.4 Family of the Cairns-Blake-Dowd (CBD) Models 14

2.3 Introduction to Two-stage and Bayesian Approaches 19

2.3.1 Two-Stage Approach 19

2.3.2 Bayesian Approach and Updating Method 20

2.4 Stochastic Model, Notations for This Study 22

2.4.1 Measures of the Mortality 22

2.4.2 The Model 23

3 The Impact of Small Population on Parameter Uncertainty **27**

3.1 Parameter Estimation 27
3.2 Distribution of MLE in Finite Samples 29
3.2.1 MLE 30
3.3 Mortality Projections 38
3.3.1 Projecting Period Effects 40
3.3.2 Projecting the Cohort Effect 48
3.3.3 Projected Mortality Rates 51
3.4 Summary 56

4 Likelihood Ratio Test for Systematic Parameter Difference 58

4.1 Review of Likelihood Ratio Test 59
4.2 Finite Sample Distribution of LRT 60
4.3 Power of the Likelihood Ratio Test 63
4.4 Impact of Parameter Misspecification on Mortality Rates and Annuities 68
4.5 Likelihood Ratio Test for the Cohort Effect 73
4.6 Empirical Examples 77
4.6.1 Females vs. Males in England and Wales 77
4.6.2 Male Mortality in Scotland vs. England and Wales 78
4.7 Summary 79

5 Application of Bayesian Statistics 81

5.1 Review for the Previous Chapters 81
5.2 Review of Bayesian Statistics 84
5.3 Bayesian Statistics and Markov Chain Monte Carlo 87
5.3.1 Hamiltonian Monte Carlo 87

5.4 Posterior Distribution for $\boldsymbol{\theta}$ 92

5.4.1 Prior Distributions for $\boldsymbol{\kappa}$, $\boldsymbol{\mu}$ and $\boldsymbol{V_\epsilon}$ 93

5.4.2 Prior Distributions for $\boldsymbol{\gamma}$, α_γ and σ_γ 101

5.4.3 Joint Posterior Density 103

5.4.4 Parameter Initialization and Model Fitting 106

5.4.5 Projection 108

5.5 Results of Fitting Benchmark Population with Bayesian Approach . . 110

5.5.1 Convergence 110

5.5.2 Compare with the Two-Stage Approach 111

5.5.3 Sensitivity Test 136

5.5.4 Conclusion 140

5.6 Summary 141

6 Fitting Small Population with Bayesian Approach 142

6.1 Convergence 144

6.2 Compare with the Two-Stage Approach 147

6.2.1 Period Effect $\boldsymbol{\kappa}$ and the Hyper-Parameters $\boldsymbol{\mu}$, $\boldsymbol{V_\epsilon}$ 147

6.2.2 Cohort effect $\boldsymbol{\gamma}$ and the Hyper-Parameters σ_γ, α_γ 189

6.2.3 Fitted Death Rate 204

6.2.4 Summary 219

6.2.5 Projection 222

6.3 Summary 258

7 Empirical Case Study 262

7.1 Introduction 262

7.2 Data 263
7.3 Modelling 263
7.4 Results 266
7.5 Summary 285

8 Summary **287**

Appendix A Notations Introduced in Each Chapter **289**

A.1 Chapter 2 290
A.2 Chapter 3 292
A.3 Chapter 4 295
A.4 Chapter 5 297
A.5 Chapter 6 303
A.6 Chapter 7 316

Appendix B For Chapter 3 **321**

B.1 Applying Identifiability Constraints for Parameter Estimation 322

Appendix C For Chapter 5 **324**

C.1 Metropolis-Hastings Algorithm and Gibbs Sampler 325
C.1.1 Review of the Metropolis-Hastings Algorithm 325
C.1.2 Case Study 326
C.2 Hamiltonian Monte Carlo 357
C.2.1 Energy Function 357
C.2.2 Potential Energy 357
C.3 The Impact of Identifiability Constraints on Posterior Distribution . . 357

C.4 England and Wales MCMC 358

Appendix D For Chapter 6 **360**

Appendix E For Chapter 7 **373**

List of Tables

3.1 The exposure for males in England and Wales (EW) in year 2011 at selected ages. 29

3.2 The finite-sample SD of random walk drift with different population size. 45

4.1 The table contains the size of shift required for 50% power when each parameter is shifted separately, with respect to population $w = 1, 0.1, 0.01$ 68

4.2 The impact of shifting each parameter separately on the price of a twenty five-year temporary annuity for an individual aged at 65. . . . 71

4.3 The impact of shifting each parameter separately on the price of a ten-year deferred twenty five-year temporary annuity for an individual aged at 55. 71

5.1 Mean and standard deviation of the posterior distribution for the random walk drift and co-variance matrix and the respected MLEs for the England and Wales data. 114

5.2 Characteristic statistics of the posterior distribution of the hyper-parameters of the cohort effect as well as the respected MLEs. 121

5.3 Empirical death rates for generation aged 50 in year 1972, including the crude rates, mean, median, 90% credibility intervals of MCMC and the MLE. Note that the upper and lower bound of the sixth column is the 95% and 5% quantile of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ respectively. . . . 124

6.1 Characteristic statistics of the posterior distribution for the random walk volatility. 152

6.2 Characteristic statistics of the distribution of the mean of the posterior distribution for the random walk drift and volatility driven by the sampling variation, based on Method 3. 155

6.3 The mean and the standard deviation of the prior distribution for the random walik volatility. 157

6.4 Characteristic statistics of: the posterior distribution for the first random walk drift of England and Wales data and selected death scenario. 164

6.5 Characteristic statistics of the posterior distribution for the third random walk volatility based on Method 4. 179

6.6 Characteristic statistics of the posteriro distribution for the volatility of the AR(1) process for the cohort effect based on Method 4 and 6. . 193

6.7 Characteristic statistics of the distribution driven by the sampling variation of the posterior mean for the coefficient and the volatility of the AR(1) process for the cohort effect based on Method 3. 196

6.8 Characteristic statistics of the coefficient of the AR(1) process for the cohort effect based on Method 4 and 6. 197

6.9 The characteristic statistics of the price of the 25-year term annuity for an individual aged exactly 65 at the beginning of year 2012 with respect to different levels of interest rates. 248

6.10 The characteristic statistics of the price of the 25-year term annuity for an individual aged exactly 65 at the beginning of year 2012 with respect to different levels of interest rates. 250

6.11 The characteristic statistics of the price of the 25-year term annuity deferred by 10 years for an individual aged exactly 55 at the beginning of year 2012 with respect to different levels of interest rates. 252

6.12 The characteristic statistics of the price of the 25-year term annuity deferred by 10 years for an individual aged exactly 55 at the beginning of year 2012 with respect to different levels of interest rates. 254

6.13 The longevity risk (in percentage) of the temporary annuity. 255

6.14 The longevity risk (in percentage) of the deferred annuity. 255

7.1 The mean and the standard deviation of the prior distribution of the random walk volatility for England and Wales and Scotland during year 1971-2011, aged 50-89 last birthday. 284

7.2 The mean and the standard deviation of the posterior distribution of the random walk drift for England and Wales and Scotland during year 1971-2011, aged 50-89 last birthday. 285

7.3 The mean and the standard deviation of the posterior distribution of the random walk volatilityfor England and Wales and Scotland during year 1971-2011, aged 50-89 last birthday. 286

List of Figures

3.1 Empirical correlation matrix for different population size 33

3.2 Finite-sample MLEs for the first period effect with respect to different population sizes 34

3.3 Finite-sample MLEs for the second period effect with respect to different population sizes 35

3.4 Finite-sample MLEs for the third period effect with respect to different population sizes 36

3.5 Finite-sample MLEs for the cohort effect with respect to different population sizes 37

3.6 The distribution of the k^{th} order difference of MLEs 39

3.7 The distribution of random walk drift with different population size . 44

3.8 Effect of removing short cohorts 49

3.9 Distributions of hyper-parameters for the cohort effects 50

3.10 Projections of latent parameters with different population size 50

3.11 Mortality projection with different population size. 53

3.12 SD of projected mortality rates at year 2030 with different population size. 54

3.13 The projected mortality rates for the four experiments. 56

4.1 Testing full parameters: the distribution of likelihood ratio test statistics and p-values for different population sizes. 62

4.2 Power of LR test for different population sizes. 66

4.3 The impact of shifting each parameter separately on the estimated death rate of England and Wales. 69

4.4 Testing cohort effect, the distribution of LR test statistics and p-values with different population sizes 76

4.5 The estimates of cohort effect, for England and Wales males (solid line) and females (dashed line), age 50 to 89 last birthday, over year 1961 to 2011 78

4.6 The estimates of cohort effect, for the males of England and Wales (solid line) and Scotland (dotted line), age 50 to 89 last birthday, over year 1961 to 2011. The dashed lines are the CI for the cohort effect of Scotland. The upper bound is 95% quantile of the distribution and the lower bound is 5% quantile. 79

5.1 Trace plots for the selected parameters, including the period and cohort effect with selected year (1961, 2011) and year of birth (1872, 1901, 1961) respectively, the drift of multi-variate random walk, the diagonal of co-variance matrix, coefficient and standard deviation of AR(1) model. The benchmark data fitted is the England and Wales data, males during year 1961 to 2011 aged 50 to 89 last birthday. . . 112

5.2 CDF: The posterior distribution of the first random walk drift (upper, dotted CDF) and the respected volatility (lower, dotted CDF) for the England and Wales data. The vertical dotted-dashed lines are the respected MLEs 113

5.3 Credibility intervals for the posterior distribution of the first period effect for England and Wales data. 115

5.4 CDF: The posterior distribution of the second random walk drift (lower, dotted CDF) and the respected volatility (upper, dotted CDF) for the England and Wales data. The vertical dotted-dashed lines are the respected MLEs 116

5.5 Credibility intervals for the posterior distribution of the second period effect for England and Wales data. 117

5.6 CDF: The posterior distribution of the third random walk drift (lower, dotted CDF) and the respected volatility (upper, dotted CDF) for the England and Wales data. The vertical dotted-dashed lines are the respected MLEs 118

5.7 Credibility intervals for the posterior distribution of the third period effect for England and Wales data. 119

5.8 Credibility intervals for the posterior distribution of the cohort effect for England and Wales data. 120

5.9 CDF of Bayesian estimates for the coefficient (left plot) and the volatility (right plot) of the cohort effect. 122

5.10 Compare the deaths rates between: the fitted deaths rates given the empirical estimates with maximum likelihood method; the 90% confidence interval of fitted deaths rates given samples from posterior distribution (or the MCMC); crude death rates with Poisson error bar, with respected to year range in the dataset for selected ages (50, 60, 80 89). The error bar is set to 1.96 times of the standard error ($\sqrt{\frac{m(\hat{\theta}^{\mathrm{EW}},t,x)}{E^{\mathrm{EW}}(t,x)}}$). 126

5.11 The heat plot of the ratio $r(t,x)$: historical fitted $m(t,x)$ given the MLE over the posterior mean, for all the observed ages and years. In plot (b), the red dots represents the ratio less than one but greater than 0.995; the green and black dots are the ratios on $[1, 1.005]$ and greater than 1.005 respectively. Ratios less than 0.995 are coloured white. 127

5.12 The ratio of historical fitted $m(t,x)$ of MLE over the posterior mean with respect to year, age and cohort. 128

5.13 The standardized residuals of the MCMC mean (left) and ML method (right). The black dots represents a positive residuals and the grey ones are for negative residuals. 129

5.14 Compare the residuals of MCMC mean (black) and ML estimates (red) with respect to age (a), year (b), cohort (c). Figure (d) shows the correlation between the MCMC and ML residuals. 130

5.15 Comparison between the credibility intervals (dashed lines) and the confidence intervals (solid lines) for projecting the period and cohort effects for the males in England and Wales, given the Bayesian and MLE approach respectively. The upper bound of the interval is 95% quantile of the projected samples and the lower bound is 5% quantile. Both the period and cohort effects are projected fifty years forward. . 132

5.16 Comparison between the credibility intervals (dashed lines) and the confidence intervals (solid lines) for projecting the cohort effect for the males in England and Wales, given the Bayesian and MLE approach respectively. 133

5.17 The predictive intervals of log-scaled projected death rates given MCMC (dashed) and ML method (solid) for selected ages (50,60,80,89) with respect to year. 135

5.18 Cohort mortality for cohorts initialy aged 55 (upper) and 65 (lower) at year 2000. 137

5.19 CDF of sensitivity test: The posterior distribution of random walk drift (first row) and co-variance matrix (second and third row) for the males in England and Wales during year 1961-2011, aged 50-89 last birthday, given the non-informative Jeffrey's prior (dashed curve) and the informative Inverse Wishart prior (dotted curve) for the co-variance matrix. The vertical line is the corresponding point estimates generated by the MLE of the latent parameters. 139

5.20 CDF of sensitivity test: The posterior distribution of coefficient (left) and variance (right) of the cohort effect for the males in England and Wales during year 1961-2011, aged 50-89 last birthday, given the non-informative Jeffrey's prior (dashed curve) and the informative Inverse Wishart prior (dotted curve) for the co-variance matrix. The vertical line is the corresponding point estimates generated by the MLE of the latent parameters. 139

6.1 Trace plots for the selected parameters given Jeffrey's prior for the joint density of random drift and co-variance matrix. 145

6.2 Trace plots for the selected parameters given informative prior for the random walk volatility. 146

6.3 The posterior distribution of the first random walk volatility, based on Method 4 and 6. 148

6.4 The influence of the sampling variation on the posterior distribution of the first random walk volatility, based on Method 5. 150

6.5 Credibility intervals of the posterior distribution for the first period effect based on Method 4 and 6. 151

6.6 The scatter plot: the posterior mean vs. skewness (lower); the posterior mean vs. standard deviation (upper) for the first random walk volatility 156

6.7 The posterior distribution for the first random walk drift based on Method 4 and 6. 158

6.8 Influence of the sampling variation on the posterior distribution for the first random walk drift based on Method 5. 161

6.9 The scatter plot: the posterior mean vs. skewness (lower); the posterior mean vs. standard deviation (upper) for the first random walk drift. 163

6.10 Scatter plot of the posterior median of the first random walk volatility vs. the posterior standard deviation of the drift. 165

6.11 Influence of the sampling variation on the Posterior distribution of the first period effect at selected years based on Method 5. 165

6.12 Posterior distribution of the second random walk volatility based on Method 4 and 6. 166

6.13 Credibility intervals of the posterior distribution for the second period effect based on Method 4 and 6. 167

6.14 Influence of the sampling variation on the posterior distribution of the second random walk volatility based on Method 5. 169

6.15 The scatter plot: the posterior mean vs. skewness (lower); the posterior mean vs. standard deviation (upper) of the second random walk volatility. 171

6.16 Posterior distribution of the second random walk drift based on Method 4 and 6. 172

6.17 Influence of the sampling variation on the posterior distribution of the second random walk drift, based on Method 5. 174

6.18 The scatter plot: the posterior mean vs. skewness (lower); the posterior mean vs. standard deviation (upper) for the second random walk drift. 175

6.19 Scatter plot of the posterior median of the second random walk volatility vs. the posterior standard deviation of the drift. 176

6.20 Influence of the sampling variation on the posterior distribution for the second period effect at selected years based on Method 5. 177

6.21 Posterior distribution of the third random walk volatility based on Method 4 and 6. 178

6.22 Credibility intervals of the posterior distribution for the third period effect based on Method 4 and 6. 180

6.23 Influence of the sampling variation on the posterior distribution for the third random walk volatility based on Method 5. 182

6.24 The scatter plot: the posterior mean vs. skewness (lower); the posterior mean vs. standard deviation (upper) for the third random walk volatility. 184

6.25 Posterior distribution of the third random walk drift based on Method 4 and 6. 185

6.26 Influence of the sampling variation on the posterior distribution of the third random walk drift based on Method 5. 186

6.27 The scatter plot: the posterior mean vs. skewness (lower); the posterior mean vs. standard deviation (upper) for the third random walk drift. 187

6.28 Scatter plot of the posterior median of the third volatility vs. the posterior SD of the drift. 188

6.29 Influence of the sampling variation on the posterior distribution of the third period effect at selected years based on Method 5. 189

6.30 Posterior distribution of the volatility of the AR(1) process for the cohort effect based on Method 6. 190

6.31 Posterior distribution of the volatility of the AR(1) process for the cohort effect based on Method 4. 191

6.32 Influence of the sampling variation on the posterior distribution of the volatility of the AR(1) proces for the cohort effect based on Method 5.194

6.33 Scatter plot of: posterior mean vs. variance; posterior mean vs. skewness, of the volatility of the AR(1) process for the cohort effect. . . . 195

6.34 Posterior distribution of the coefficient of the AR(1) process for the cohort effect based on Method 4 and 6. 198

6.35 Influence of the sampling variation on the posterior distribution of the coefficient of the AR(1) process for the cohort effect based on Method 5. 200

6.36 Scatter plot of: posterior mean vs. variance; posterior mean vs. skewness, of the coefficient of the AR(1) process for the cohort effect. . . . 201

6.37 Influence of the sampling variation on the posterior distribution of the stationary variance of the AR(1) process for the cohort effect based on Method 5. 202

6.38 Credibility intervals of the posterior distribution for the cohort effect based on Method 4 and 6. 203

6.39 Influence of the sampling variation on the posterior distribution of the cohort effect at selected years of birth based on Method 5. 204

6.40 Fitted death rates at age 65 based on Method 4 and 6. 206

6.41 Fitted death rates based on Method 4 (upper) and the influence of sampling variation on the central fitted rates based on Method 5 at age 55. 208

6.42 Influence of the sampling variation on the upper and lower bound of the MCMC fitted rates ate age 55 based on Method 5. 210

6.43 Fitted death rates based on Method 4 (upper) and the influence of sampling variation on the central fitted rates based on Method 5 at age 65. 211

6.44 Influence of the sampling variation on the upper and lower bound of the MCMC fitted rates ate age 65 based on Method 5. 213

6.45 Fitted death rates based on Method 4 (upper) and the influence of sampling variation on the central fitted rates based on Method 5 at age 75. 214

6.46 Influence of the sampling variation on the upper and lower bound of the MCMC fitted rates ate age 75 based on Method 5. 215

6.47 Fitted death rates based on Method 4 (upper) and the influence of sampling variation on the central fitted rates based on Method 5 at age 85. 216

6.48 Influence of the sampling variation on the upper and lower bound of the MCMC fitted rates ate age 85 based on Method 5. 217

6.49 Posterior distribution of the projected first period effect. 224

6.50 Posterior distribution of the projected second period effect. 225

6.51 Posterior distribution of the projected third period effect. 226

6.52 Posterior distribution of the cohort effect. 228

6.53 Influence of the sampling variation on the central projection for the cohort effect based on Method 5. 229

6.54 Upper: Credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 65, 75 and 85; Lower: The difference between and upper and lower bound at year 2061 for the age range of 50-89. 231

6.55 90% credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 55 based on Method 4. 233

6.56 90% credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 65 based on Method 4. 234

6.57 90% credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 75 based on Method 4. 236

6.58 90% credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 85 based on Method 4. 238

6.59 90% credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 89 based on Method 4. 239

6.60 Influence of the sampling variation on the MCMC central projection at age 55, 65, 75, 85 and 89 based on Method 5. 240

6.61 Influence of the sampling variation on the upper bound of the MCMC projection at age 55, 65, 75, 85 and 89 based on Method 5. 242

6.62 Influence of the sampling variation on the lower bound of the MCMC projection at age 55, 65, 75, 85 and 89 based on Method 5. 242

6.63 The mean of the logit-scaled mortality improvement with respect to age. 244

6.64 The standard deviation of 1000 death scenarios' central logit-scaled mortality improvement rate with respect to age. 245

6.65 The log-scaled standard deviation of the distribution of log central prediction according to 1000 death scenarios as a function of age x at year 2061. 246

6.66 Survival Index of individual aged 65 exact at the beginning of year 2012. 247

6.67 Survival Index of individual aged 55 exact at the beginning of year 2012. 251

6.68 Influence of the sampling variation on the MCMC temporary and deferred annuity prices at 1% interest rate, based on Method 5. . . . 256

6.69 Influence of the sampling variation on the MCMC longevity risk based on Method 5. 258

7.1 Posterior distribution of the first random walk volatility (upper) and drift (lower) of Scotland given different prior settings for the covariance matrix. 267

7.2 Posterior distribution of the coefficient and volatility of the AR(1) process for the cohort effect of Scotland given different prior settings for the co-variance matrix. 268

7.3 Comparing the first random walk volatility of Scotland with England and Wales based on Method 5. 269

7.4 Comparing the second random walk volatility of Scotland with England and Wales based on Method 5. 270

7.5 Comparing the third random walk volatility of Scotland with England and Wales based on Method 5. 272

7.6 Comparing the first random walk drift of Scotland with England and Wales based on Method 5. 274

7.7 Comparing the second random walk drift of Scotland with England and Wales based on Method 5. 276

7.8 Comparing the third random walk drift of Scotland with England and Wales based on Method 5. 277

7.9 Log-scaled $m(t, x)$ at age 65 for England and Wales (black) and Scotland (red). 278

7.10 Log-scaled $m(t, x)$ at age 75 for England and Wales (black) and Scotland (red). 279

7.11 Log-scaled $m(t, x)$ at age 85 for England and Wales (black) and Scotland (red). 281

7.12 Log-scaled survival index at age 65 for England and Wales (black) and Scotland (red). 282

C.1 The trace plots for Experiment 1. 335

C.2 The trace plots for Experiment 2. 336

C.3 The trace plots for Experiment 3. 338

C.4 The trace plots for Experiment 4. 339

C.5 The trace plots for Experiment 5. 340

C.6 The trace plots for Experiment 6. 342

C.7 The trace plots for Experiment 7. 343

C.8 The trace plots for Experiment 8. 344

C.9 The trace plots for Experiment 9. 345

C.10 [The trace plots for Experiment 10. 346

C.11 The trace plots for Experiment 11. 347

C.12 The trace plots for Experiment 12. 348

C.13 The trace plots for Experiment 13. 349

C.14 The trace plots for Experiment 14. 350

C.15 The trace plots for Experiment 15. 351

C.16 The trace plots for Experiment 16. 352

C.17 The trace plots for Experiment 17. 353

C.18 The trace plots for Experiment 18. 354

C.19 The trace plots for Experiment 19. 355

C.20 The trace plots of the latent parameters with selected ages and years and the hyper-parameters with HMC. 356

C.21 Posterior distributions for the covariance of the period effects for the England and Wales. 359

D.1 Influence of the sampling variation on the posterior distribution of the first period effect at all the years based on Method 5. 360

D.2 Influence of the sampling variation on the posterior distribution of the second period effect at all the years based on Method 5. 361

D.3 Influence of the sampling variation on the posterior distribution of the third period effect at all the years based on Method 5. 362

D.4 Influence of the sampling variation on the posterior distribution of the cohort effect at all the years based on Method 5. 363

D.5 Fitted MCMC mortality rates at age 55, 75 and 85 based on Method 4.364

D.6 Influence of the sampling variation on the temporary and deferred annuity prices at rate 4% based on Method 5. 365

D.7 Influence of the sampling variation on the temporary and deferred annuity prices at rate 2% based on Method 5. 366

D.8 Influence of the sampling variation on the temporary and deferred annuity prices at rate 0.5% based on Method 5. 367

D.9 Influence of the sampling variation on the temporary and deferred annuity prices at rate 0% based on Method 5. 368

D.10 Influence of the sampling variation on the longevity risk at rate 2% based on Method 5. 369

D.11 Influence of the sampling variation on the longevity risk at rate 4% based on Method 5. 370

D.12 Influence of the sampling variation on the longevity risk at rate 0.5% based on Method 5. 371

D.13 [Influence of the sampling variation on the longevity risk at rate 0% based on Method 5. 372

E.1 Scotland B-EW: Trace plot for selected parameters, including the period and cohort effects with selected year (1971, 2011) and year of birth (1911,1931,1961) respectively, the drift of multi-variate random walk, the diagonal of co-variance matrix, coefficient and standard deviation of AR(1) model. 374

E.2 Scotland B-UK: Trace plot for selected parameters, including the period and cohort effects with selected year (1971, 2011) and year of birth (1911,1931,1961) respectively, the drift of multi-variate random walk, the diagonal of co-variance matrix, coefficient and standard deviation of AR(1) model. 375

E.3 Scotland A-UK: Trace plot for selected parameters, including the period and cohort effects with selected year (1971, 2011) and year of birth (1911,1931,1961) respectively, the drift of multi-variate random walk, the diagonal of co-variance matrix, coefficient and standard deviation of AR(1) model. 376

E.4 Posterior distribution of the second random walk drift and volatility for the Scotland, given varied prior settings for the volatility. 377

E.5 Posterior distribution of the third random walk drift and volatility for the Scotland, given varied prior settings for the volatility. 378

E.6 Posterior distribution of the covariance of the three period effects for the Scotland, given varied prior settings for the volatility. 379

E.7 Posterior distribution of the covariance of the three period effects for the Scotland vs. England and Wales, given varied prior settings for the volatility. 380

Chapter 1

Introduction

1.1 Background and Motivation

Stochastic mortality models are widely used as risk management tools in the insurance and pensions industry with the main application being the generation of plausible scenarios for future mortality rates. Many stochastic mortality models have been introduced in the last few decades. When new models have been developed the objective was mostly to improve the goodness of fit of the model to mortality data observed in relatively large populations: the Lee-Carter model and its refinements (e.g. Lee & Carter (1992); Renshaw & Haberman (2003); and Booth et al. (2006)) have been developed to provide a good fit to the mortality rates observed in the United States, England and Wales and the population of UK male assured lives; the generalized linear model with smoothing techniques (e.g. Currie et al. (2004) and Currie (2006)) has been developed for fitting and forecasting the mortality data of UK insurance and pensions; while the Cairns-Blake-Dowd (Cairns et al. (2006a)) model (CBD) and its generalizations (e.g. Cairns et al. (2009) and Plat (2009)) was introduced for modelling the England and Wales males population at higher ages.

In contrast, actuaries will often face the problem of modelling the mortality experience of much smaller populations, for example, the members of a pension scheme. Empirical research has found that mortality rates of smaller populations exhibit significantly more variability compared to the observed rates in larger populations. Furthermore, models that fit large countries well, might not be appropriate for smaller populations, for example, Booth et al. (2006) showed that the Lee-Carter model provides a rather poor fit to the mortality experience of smaller populations. A related issue is that empirical data from smaller populations might only be available for a relatively short period, which makes mortality projections rather uncertain. As a result, a number of recent papers have aimed to develop models specifically for smaller populations: for example, the Saint Model of Jarner & Kryger (2011).

A common assumption for many of the proposed models is that the observed numbers of deaths are realisations of random variables with a Poisson distribution given the underlying mortality rates. The estimation of parameters of any such model is therefore based on samples from a Poisson distribution, and, as always in statistics, parameter uncertainty is related to the sample size. Furthermore, many results about the distribution of estimators and corresponding confidence intervals rely on the Maximum Likelihood estimates and large sample sizes.

The increased uncertainty about estimated parameters for small populations results in high levels of uncertainty about projected mortality rates. As a consequence future realised mortality rates will not only diverge from projected rates due to future sampling variation caused by the Poisson distribution, but might also diverge from projections since the projections themselves are uncertain. The information involved in the mortality data of the smaller population might not be sufficient enough to support the estimation and additional information is required, e.g. through the Bayesian inference. In the previous literature, the Bayesian approach and the simulation techniques are adopted to deal with the parameter uncertainty, e.g. Klugman

(1989), Yashin et al. (2000), Dellaportas et al. (2001), Girosi & King (2003), to name just a few.

In the actuarial literature, simulation techniques have been proposed for dealing with uncertain parameters and projected mortality rates. For example, Liu & Braun (2011) investigated mortality uncertainty by applying a block bootstrap method on the Lee-Carter model, and Brouhns et al. (2005) proposed Poisson bootstrap methods for mortality forecasting. Cairns et al. (2006a) studied the parameter uncertainty of the two factor CBD model by adopting a Bayesian approach. Czado et al. (2005) and Pedroza (2006) carried out the first Bayesian analyses using Markov Chain Monte Carlo (MCMC) of the Lee-Carter model, with further work by Kogure et al. (2009), Kogure & Kurachi (2010). Reichmuth & Sarferaz (2008) applied MCMC to a version of the Renshaw & Haberman (2003) model. Cairns et al. (2011b) applied MCMC to a two-population Age-Period-Cohort model by combining the Poisson likelihood for the deaths counts with time series likelihood functions for the latent random period and cohort effects.

However, to the best of our knowledge, bootstrap methods have not been applied in a systematic way to investigate the impact of the size of a population on parameter and projection uncertainty. We firstly apply Poisson parametric bootstrap methods to investigate how the variation of parameter estimates and projections is affected by the size of a population. We vary the size of the population by assigning weights to a chosen benchmark population, e.g. England and Wales males. In simulation studies we find that the size of the population has a significant effect on the variation of parameter estimates and projections.

Although we apply a weight to the benchmark population (i.e. scale it down), we ensure that the mortality rates of the constructed small populations are equal to the fitted mortality rates of the benchmark population. In such a situation, uncertainty in projected mortality rates will be reduced if information from the benchmark population parameter estimates can be used for fitting smaller populations. This raises

the question of how we can test for systematic differences between the parameters driving mortality rates in a small population and a given null hypothesis about those parameters, where the null hypothesis might have been obtained from a model fitted to a much larger population. The full null hypothesis is specified in Chapter 4 on page 60. If no significant differences can be found then it seems reasonable to use elements of the large population model fit to assist in generation of scenarios for the small population. We therefore investigate the properties of a likelihood ratio (LR) test for all or some of the estimated parameters, and, in particular, consider the distribution of the test statistic based on the bootstrap simulations. This allows us to investigate the power of the LR test and the effect of varying population sizes on the rejection rates. We find that the population size has a strong effect on the probability of a type II error. This is particularly relevant for pension schemes since the acceptance of an incorrect null hypothesis might lead to inaccurate mortality assumptions. To investigate the financial consequences of the resulting misspecified model, we consider annuity prices based on different assumptions about the underlying parameters of our model. We apply the LR test in an empirical study. The null hypothesis for that study is the estimated cohort effect for males in England and Wales. With this null hypothesis we then carry out hypothesis tests using, first, mortality data for females in England and Wales and, second, males in Scotland to check if their cohort effects are significantly different from the estimated cohort effect for males in England and Wales. We find for both populations that the estimated cohort effect is significantly different from that in the null hypothesis.

We then focus on developing a Bayesian model for estimating and forecasting the mortality data of the smaller populations that combines the fitting and projecting stages into one and most importantly uses relatively strong informative prior distributions which employ our prior knowledge on the shape of the stochastic processes (known as latent parameters in Bayesian context) of the stochastic model and the information of a referencing larger population to improve the estimation for the respected "hyper-parameters." To our best knowledge, bootstrap methods

have not been applied in a systematic way to investigate the impact of the sampling variation, especially of the smaller population, on the posterior distribution of the parameters and further how it will affect the predictive posterior distribution for the mortality data. We apply the Poisson parametric bootstrap methods to investigate how the posterior estimations of the parameters and hence the posterior predictive distribution for the mortality forecasts are affected by the Poisson sampling variation of the deaths counts when the population is small. We find that by employing the information of the referencing population for smaller population modelling we greatly smooth the latent-parameter estimations and better estimate the respected volatilities, especially for the volatilities of the period effects that dominate the forecasting uncertainty. We also observe that the parameter estimates and the mortality forecasts are shifted by the sampling variation and sometimes such shift can be very large.

1.2 Thesis Outline

The structure of the thesis is organised as follows.

In Chapter 2, we start with literature reviews on various stochastic mortality models, the widely used two-stage fitting and forecasting approach as well as the previous study on the application of the Bayesian approach for fitting and forecasting the mortality data. We then introduce the stochastic model, assumptions, definitions and notations adopted in our research. The specific mortality model that we consider is a second generation CBD model with added cohort effect.

In Chapter 3, we simulate finite samples of death scenarios with the Poisson parametric bootstrap method by ensuring that the true mortality rates of the constructed population are equal to the fitted mortality rates of the benchmark population. The two-stage approach is employed to fit the model and forecast the mortality rates of the constructed populations. That is, we estimate the latent parameters for each

simulated death scenario with the maximum likelihood method and the MLEs are then used in the candidate projecting models for calculating the point estimates of the respected parameters in the projecting models. The impact of the size of a population on parameter and projection uncertainty is studied according to the distribution of the finite-sample estimates and forecasts.

In Chapter 4, we start with the review of the generalized version of the LR test followed by a test with respect to population size for systematic differences between all of the parameters driving mortality rates (latent parameters) of constructed populations and a given null hypothesis about these parameters, where the null hypothesis has been obtained from the selected stochastic model fitted to the benchmark population. The properties of the distribution of the test statistic with respect to the population size are studied according to the distribution of the finite-sample test statistics based on the bootstrap simulated death scenarios of different sized constructed populations. We then investigate how the population size affects the power of LRT. We consider four alternative models (i.e. misspecified models), under which we evaluate the power of the LRT with a parametric bootstrap procedure similar to the one for gaining the distribution of test statistics, by shifting or scaling one of the latent parameters estimated from the benchmark population with various levels of degrees. The finical implication of the resulting misspecified model is studied by calculating the annuity prices of a temporary and deferred annuity based on alternative models where the shift or scale is determined when it results in 50% power with respect to population size. We end this chapter by applying the LR test in an empirical study

In Chapter 5, we start with the review of the Bayesian statistics and an introduction of the dynamic Hamiltonian Monte Carlo updating algorithm. We then develop a Bayesian model for fitting and forecasting the mortality data for small populations. It combines the fitting and forecasting stages into one and allows us to use our prior knowledge and the information of the benchmark population. In par-

ticular, we introduce the candidate time series likelihood for the latent parameters; the prior distribution for the hyper-parameters, especially an informative prior distribution for the volatilities of the period effects; the structure of our joint posterior density; the method of simulating the predictive posterior distribution for mortality forecasting. At the end of this chapter, we compare the performance of the two-stage and the Bayesian approach fitting and forecasting the benchmark population mortality data.

In Chapter 6, we fit and forecast the mortality data of the constructed population which is as small as one percent of the benchmark population with our Bayesian model. In particular, the information for the volatilities of the benchmark population's period effects are employed to provide an improved estimation for the small population, compared with the estimation based on the two-stage approach. We compare the Bayesian estimations and the forecasts for the small population's mortality data with those for the benchmark population to investigate the impact of the population size on the influence of the time series likelihood and prior distributions on the joint posterior distribution. With the Poisson parametric bootstrap method, we investigate how the sampling variation affects the estimations and the forecasts given the Bayesian approach. We calculate the annuities price to study the financial implication of our Bayesian approach and sampling variation on the small population mortality forecasts according to the Poisson parametric bootstrap method.

In Chapter 7, we carry out an empirical study by fitting and forecasting the mortality data of the Scotland males given the Bayesian model with the information of the England and Wales males. The Poisson parametric bootstrap method is used for investigating if both populations have the same level of volatility.

In Chapter 8, we make a summary of the contribution of the research and plan for the future research.

Chapter 2

Mortality Model and Notations

2.1 Introduction to the Parametric Mortality Models

Gompertz (1825) introduced the linear relationship between the log-scaled force of mortality and the age, known as the "law of human mortality". More specifically, the parametric model can be written as:

$$\log \mu(x) = \log(\alpha) + \log(\beta)x,$$

where $\mu(x)$ is the force of mortality at age x, α is a given referencing mortality rate and β measures the speed of senescent. Makeham (1860) extended the law by adding an age independent component λ. During the following decades many parametric mortality models are developed, see Tabeau et al. (2002) and Forfar (2004) for a comprehensive list of parametric models. In particular, Heligman & Pollard (1980) developed the well known Heligman-Pollard 8-component model:

$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + D\exp\left[-E\left\{\log\frac{x}{F}\right\}^2\right] + GH^x,$$

where q_x is the mortality rate at age x, the components A to H are the eight parameters. Rogers & Planck (1983) firstly introduced the nine-parameter multi-exponential mortality model, more specifically

$$m_x = a_0 + a_1 \exp(-\alpha_1 x) + a_2 \exp\{-\alpha_2(x-\mu) - \exp[-\lambda(x-\mu)]\} + a_3 \exp(-\alpha_3 x),$$

According to the original paper, m_x is the central mortality rate at age x; a_0 measures the level of the mortality rate; the a_i for $i = 1, 2, 3$ are the scale parameters; α_i for $i = 1, 2, 3$, μ and λ define the mortality characteristic.

There are many other complicated parametric models, for instance: the eight-parameter Carriere (1992) model with parameters representing the age-related causes of deaths, eight-parameter Gage & Mode (1993) model, five-parameter Siler (1983) model, ten-parameter Mode & Jacobson (1984) model, Hannerz (1999) five-parameter model, etc.

The advantages for using parametric modes are: only small number of parameters required in the model (e.g. three to ten parameters); the parameters of some models could be interpreted easily (e.g. eight-parameter Carriere (1992) model); easy to be generalized (e.g. the nine-parameter Rogers & Planck (1983) model includes the components for the accident introduced by Coale & McNeil (1972) and the Gompertz-law formed senescent factors); fit certain populations well (e.g. the study of Mode & Busby (1982); Hartmann (1987); Kostaki (1988); Rogers & Gard (1990) show the Heligman-Pollard model fits a range of populations well).

On the other hand, many parametric models are hard to fit, e.g. the Heligman-Pollard model. Empirical studies (e.g. Hartmann (1987), Pollard et al. (1987), Keyfitz (1991), McNown & Rogers (1992), McNown et al. (1995)) show that both the Heligman-Pollard and the multi-exponential models are not useful for forecasting, mainly due to the high correlations between the parameters that compromise the advantage of easily interpreted parameters (Booth & Tickle (2008)).

2.2 Introduction to the Stochastic Mortality Models

Unlike the parametric mortality models that pre-specify functions of age for the mortality rate, the stochastic mortality models (non-parametric) are time series models that try to estimate the patterns of the effects that directly drive the mortality rate (latent parameters), for example the age, period and cohort effects. Booth & Tickle (2008) make a comprehensive review of a range of mortality models' performance on fitting and forecasting the mortality rates.

Most of the stochastic models that have been developed during the last few decades could be in general grouped into the following families: Cairns-Blake-Dowd (CBD) model (Cairns et al. (2006a)) and its extensions; Lee-Carter model (Lee & Carter (1992)) and its variations; models smoothing both the the age and period dimensions incorporated with the P-spline technique (Currie (2006)).

2.2.1 Family of Lee-Carter Models

Lee & Carter (1992) proposed the following model (labelled by Cairns et al. (2009) as M1) for fitting the US mortality dataset:

$$\log m(t,x) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)},$$

whereas: $m(t,x)$ is the death rate at year t, age x; $\beta_x^{(1)}$ and $\beta_x^{(2)}$ are the functions for the age effect; $\kappa_t^{(2)}$ is the function for the period effect. The Lee-Carter model allows roughness in the mortality rates between ages and has only one random effect, i.e. the $\kappa_t^{(2)}$. In the original paper, the authors estimated the latent parameters with the singular value decomposition methodology (SVD) and forecast the random period

effect with a random walk model with a drift, more specifically

$$\kappa_t^{(2)} = \mu + \kappa_{t-1}^{(2)} + \epsilon_t,$$

where μ and ϵ_t are the drift and the error term respectively.

Note that the latent parameters are defined as the parameters that directly drive the mortality rate, for example the $\kappa_t^{(2)}$ in the Lee-Carter model. Parameters that determine the stochastic process of the latent parameters are defined as the hyper-parameters, for example the μ in the above random walk model for the period effect $\kappa_t^{(2)}$.

There is an identifiability problem with this model as one could have the same fitted value with different parameter estimations. The authors proposed the following constraints (not unique) on the $\beta_x^{(2)}$ and $\kappa_t^{(2)}$ such that $\sum \beta_x^{(2)} = 1$ and $\sum \kappa_t^{(2)} = 0$ and therefore $\beta_x^{(1)}$ is the mean of the $\log m(t, x)$ over the underlying period.

Variants are introduced to the original Lee-Carter method by the following study. For example, Lee & Miller (2001) introduce modifications include forecasting from the observed rates and Booth et al. (2002) modify the Lee-Carter method with a conditional maximum likelihood procedure. An extended version of the original Lee-Carter model is introduced in Booth et al. (2002, 2001), more specifically;

$$\log m(t, x) = \sum_{i=1}^{n} \beta_x^{(i)} \kappa_t^{(i)},$$

where in the original paper $\kappa_t^{(1)} = 1$. Hyndman & Ullah (2007) extend the Lee-Carter method with a functional data paradigm incorporated with non-parametric smoothing component to reduce the data randomness and a robust principal components to deal with the outlying years.

Renshaw & Haberman (2003) extend the Lee-Carter model to allow for the age-specific enhancement and Renshaw & Haberman (2006) extend the Lee-Carter

model with another source of randomness, a cohort effect (labelled by Cairns et al. (2009) as M2). More specifically,

$$\log m(t,x) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_{t-x}^{(3)},$$

whereas $\gamma_{t-x}^{(3)}$ is the function for the cohort effect at the year of birth $t-x$. In the original paper, the authors proposed a two-stage (see further discussion in Section 2.3.1) fitting strategy in which $\beta_x^{(1)}$ is firstly fixed to the average of the $\log m(t,x)$ over the time and the rest of the parameters are then estimated according to the following set of constraints (not unique):

$$\sum_x \beta_x^{(2)} = 1;\ \sum_x \beta_x^{(3)} = 1;\ \kappa_{t_1}^{(2)} = 0.$$

For the Model M2, Cairns et al. (2009) adopt the mean of the $\log m(t,x)$ as an initial value of he $\beta_x^{(1)}$ and calculate the maximum likelihood estimates (MLE) for all the parameters with $\beta_x^{(1)}$ included in the updating algorithm. The constraints they set are as follows:

$$\sum_t \kappa_t^{(2)} = 0;\ \sum_x \beta_x^{(2)} = 1;\ \sum_c \gamma_c^{(3)} = 0;\ \sum_x \beta_x^{(3)} = 1.$$

They claim that there remains some sort of identifiability problem since the parameters are observed to converge to their MLEs very slowly.

The advantages of the Lee-Carter method include: relatively simple and easy to fit compared with those complicated parametric models; the forecasting random walk model is appropriate and allows to produce stochastic forecasts with prediction intervals, etc.

It is worth noticing that Wilmoth (1993) firstly develop a maximum likelihood method for fitting the Lee-Carter model by assuming the deaths follows a Poisson distribution, which literately becomes a common methodology for fitting the

stochastic mortality models to the death data in the following studies.

The shortcomings of the Lee-Carter model include: the model assumes that the mortality improvement rate at each age level is unchanged over the period while there is evidence suggests this assumption may not be realistic (Booth & Tickle (2008)); the allowance of the roughness on the age dimension could cause smoothness problem on the forecasting and empirical study (Girosi & King (2003)) shows that the forecasts becomes far less smoothness over time; Cairns et al. (2011a, 2009) note that the forecasts uncertainty is proportional to the average improvement rate $\beta_x^{(2)}$, which could results in a too low uncertainty for high ages given the empirical improvement rates have always been low at these ages (Plat (2009)).

2.2.2 Age-Period-Cohort Model

One problem of including the cohort factor is the linear relationship between the age, period and cohort factors, i.e. identifiability problem (Tabeau (2001)). Willekens & Baydar (1984) develop the original Age-Period-Cohort model, which is fitted to the so-called "double-classified data" collection in the original paper to avoid the inter-dependence problem. The model (labelled by Cairns et al. (2009) as M3) is shown as follows:

$$\log m(t,x) = \beta_x^{(1)} + \kappa_t^{(2)} + \gamma_{t-x}^{(3)}.$$

Currie (2006) fit and forecast the mortality rates with the APC model incorporated with a P-spline smoothness. Cairns et al. (2009) analysed the discrete version of the Model M3 without imposing any smoothness. This model has an identifiability problem and the constraints proposed by Cairns et al. (2009) are that with the $\kappa_t^{(2)}$ and $\gamma_c^{(3)}$ restricted by the following two constraints:

$$\sum_t \kappa_t^{(2)} = 0; \ \sum_c \gamma_c^{(3)} = 0,$$

they adjusted the latent-parameter estimates as follows:

$$\begin{aligned}
\tilde{\kappa}_t^{(2)} &= \kappa_t^{(2)} - \delta(t - \bar{t}) \\
\tilde{\gamma}_c^{(3)} &= \gamma_c^{(3)} + \delta\Big((t - \bar{t}) - (x - \bar{x})\Big) \\
\tilde{\beta}_x^{(1)} &= \beta_x^{(1)} + \delta(x - \bar{x}).
\end{aligned}$$

See more discussion and the derivation of the parameter δ in Cairns et al. (2009).

2.2.3 P-Spline Method

Currie et al. (2004) demonstrate smoothing the estimation and forecasting with the method of the P-splines. In the original paper, they proposed a penalized generalized linear model with Poisson error as follows (labelled by Cairns et al. (2009) as M4):

$$\log m(t, x) = \sum_{i,j} \theta_{ij} B_{ij}^{ay}(x, t)$$

with θ_{ij} smoothed in the age and cohort directions. The $B_{ij}^{ay}(x, t)$ is the two-dimensional B-spline regression matrix that can be written as the Kronecker product of the two one-dimensional B-spline regression matrices $B_j^a(t)$ and $B_i^y(t)$ for the age x and period t respectively, more specifically $B_{ij}^{ay}(x, t) = B_i^y(t) \otimes B_j^a(t)$, see Currie et al. (2004), Cairns et al. (2009) for the details on constructing the B-splines.

2.2.4 Family of the Cairns-Blake-Dowd (CBD) Models

Cairns et al. (2006a) propose the original two-factor Cairns-Blake-Dowd (CBD) model (labelled by Cairns et al. (2009) as M5) for fitting the mortality rate $q(t, x)$

of the England and Wales males, especially at the higher ages. More specifically:

$$\text{logit}q(t,x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}),$$

where $\bar{x}$ is the mean of the age range of the mortality data.

The features of the CBD model include: a logit transformation is used to guarantee the forecasts are strictly with the range of $[0, 1]$; the model fits the mortality rate $q(t, x)$ instead of the death rate $m(t, x)$ as $m(t, x)$ is not necessarily less than one and when the age is very high or low, both rates could be very higher and therefore $q(t, x)$ could be much differed from the $m(t, x)$; the model has a Gompertz typed parametric age effect $(x - \bar{x})$ with the randomness from the two time series process ($\kappa_t^{(1)}$ and $\kappa_t^{(2)}$); the period effects construct a non-trivial correlation structure (Plat (2009)); the model assumes that there is smoothness in the mortality rates between the ages; the model is relatively simple; no identifiability problem with the original CBD model. In other words, the CBD model and its extensions can be viewed as an extended Gompertz model families with uncertainties from the time series process over time.

In the original paper, the authors projected $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ with a two-dimensional random walk model with a drift as follows;

$$\begin{pmatrix} \kappa_t^{(1)} \\ \kappa_t^{(2)} \end{pmatrix} = \begin{pmatrix} \kappa_{t-1}^{(1)} \\ \kappa_{t-1}^{(2)} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + L \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix},$$

where $(\mu_1, \mu_2)^T$ is the drift vector, $(\epsilon_1, \epsilon_2)^T$ is the vector of two-dimensional multivariate normal error independent of t and L is the Cholesky decomposition of the co-variance matrix $V = LL^T$. The authors projected the survival index by including and excluding the parameter uncertainty for the drift and the co-variance matrix and concluded that the CBD model enables us to simulate the survival index efficiently under either the real-world or some sort of risk-adjusted measures and therefore

allows us to analyse the longevity risk of the longevity bonds.

Efforts have been paid to explain the cohort effect since it has been identified by many countries, e.g. England and Wales. Willets (2004) in 2004 systematically and statistically discussed the cohort effect of the U.K. population and the cohort-related causes of deaths. Cairns et al. (2009) introduce the following model to capture the cohort effect (labelled by Cairns et al. (2009) as M6):

$$\text{logit } q(t,x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}.$$

This model has an identifiability problem such that one can add a linear term with respect to the cohort $t - x$ to the cohort effect and make corresponding adjustment to the two period effects without changing the fitted value. In the original paper, in order to avoid an arbitrary choice of the linear term the authors proposed the following constraints (not unique):

$$\sum_c \gamma_c^{(3)} = 0; \ \sum_c c\gamma_c^{(3)} = 0$$

such that the least square estimates of the two coefficients of the linear term are zero.

Cairns et al. (2009) proposed the third generalization of the CBD model as follows (labelled as M8):

$$\text{logit}q(t,x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x}^{(3)}(x_c - x),$$

where x_c is some constant parameter to be estimated. The authors restricted the estimates for the cohort effect such that $\sum_c \gamma_c^{(3)} = 0$.

Plat (2009) proposes a CBD typed 4-factor model with a non-parametric factor

$\beta_x^{(0)}$ to allow a fit to younger ages. More specifically,

$$\log m(t,x) = \beta_x^{(0)} + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x} - x) + \kappa_t^{(3)}(\bar{x} - x)^+ + \gamma_c^{(4)},$$

where $(\bar{x} - x)^+ = \max(\bar{x} - x, 0)$. The author adds the factor $\kappa_t^{(3)}$ to capture the dynamics of mortality rates at younger ages. The four stochastic processes are modelled with suitable mean-reverting ARIMA process. In the original paper, the author follows the approach in Cairns et al. (2009) for Model M6 and adopts the following constraints:

$$\sum_c \gamma_c^{(4)} = 0; \ \sum_c c\gamma_c^{(4)} = 0; \ \sum_t \kappa_t^{(3)} = 0$$

to deal with the identifiability problem.

Model M7

Cairns et al. (2009) introduce the second generalization of the CBD model (labelled by Cairns et al. (2009) as M7) as follows:

$$\text{logit } q(t,x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x}^{(4)},$$

where $\hat{\sigma}_x^2$ is the mean of the $(x - \bar{x})^2$.

Model M7 extends the CBD model with a cohort effect $\gamma_{t-x}^{(4)}$, another Gomperts typed parametric quadratic function for the age effect $((x - \bar{x})^2 - \hat{\sigma}_x^2)$ and the third stochastic process for the period effect $\kappa_t^{(3)}$. The adoption of the quadratic term tries to capture the wave pattern of the mortality rates over the age dimension observed in the U.S. data (Cairns et al. (2009)). The influence of the $\kappa_t^{(3)}$ on the mortality rate improvement increases when the age approaches very old or young ages and therefore relatively slightly better fit for those ages.

Model M7 also has an identifiability problem. One example provided by Cairns et al. (2009) is that one can replace the cohort effect $\gamma_{t-x}^{(4)}$ with $\tilde{\gamma}_{t-x}^{(4)}$ by adding a quadratic term with respect to the year of birth $t-x$ such that:

$$\tilde{\gamma}_{t-x}^{(4)} = \gamma_{t-x}^{(4)} + \phi_1 + \phi_2(t-x-\bar{x}) + \phi_3(t-x-\bar{x})^2,$$

and make corresponding adjustment to the other three period effects without changing the fitted value of logit $q(t,x)$. In order to avoid an arbitrary choice for the ϕ_1, ϕ_2, ϕ_3, Cairns et al. (2009) propose the following constraints (not unique) on the $\gamma_{t-x}^{(4)}$ such that:

$$\sum_c \gamma_c^{(4)} = 0; \ \sum_c c\gamma_c^{(4)} = 0; \ \sum_c c^2\gamma_c^{(4)} = 0,$$

where $c = t-x$. This set of constraints ensure that the least square estimates for ϕ_1, ϕ_2, ϕ_3 are zero. The consequence of this set of constraints is that the estimated cohort effect is restricted to be fluctuating around zero without any linear and quadratic curvature trend. As we have mentioned, there are other candidate choices of constraints. For example, one could fix the first three cohorts to be zero or even set constraints on the period effects. One could also set the following constraints:

$$\sum_t \sum_x \gamma_c^{(4)} = 0; \ \sum_t \sum_x c\gamma_c^{(4)} = 0; \ \sum_t \sum_x c^2\gamma_c^{(4)} = 0.$$

Cairns et al. (2009) systematically compare the performance of the stochastic models M1-M8 fitting the data from the England and Wales and the United States according to various criteria. The authors claim that no single model bests the others under all the criteria and each model has its own strength. However, they do conclude that the Model M7 with a cohort effect and a quadratic term for the age fits both the England and Wales and the United States well. The Model M7 stands out with the best robustness than the rest of the models, which implies that we could, at least to some extent, trust on the result of the projection generated by M7. As a complementary, Cairns et al. (2011a) systematically compared the performance of

mortality forecasting of the models M1-M8 given the mortality data of the England and Wales and the United States. The authors emphasised that providing a good fit to the historical data is not the only criteria for the model selection. One should also consider if the model could generate plausible forecasts with comparable forecasts uncertainty with the historical uncertainty that are biologically reasonable and of good robustness to the historical data fitted. The authors claimed that the Model M5 and M7 provides robust and biologically reasonable forecasts for both populations.

Therefore, we select Model M7 for our study in this thesis based on the features of M7 and to reflect the work of Cairns et al. (2009) namely that we want to use a model that fits the benchmark population (the males from England and Wales) in our study well. We use the same constraints with Cairns et al. (2009) so that: it is consistent with the idea that cohort effect picks up the residuals of the model and other effects not captured by the period effects rather than the dominant drivers of the mortality rates; it corresponds to the prior distribution we are using for the cohort effect such that it follows a zero mean-reverting auto-regression process around zero.

2.3 Introduction to Two-stage and Bayesian Approaches

2.3.1 Two-Stage Approach

We have briefly introduced the two-stage approach in the introduction chapter. It means that the fitting and forecasting procedures are carried out separately. The latent parameters (see Page 11) that directly drive the mortality models (e.g. age, period and cohort effects) are estimated from the mortality data while the hyper-parameters (see Page 11) of the projecting model for the respected latent parameters

(e.g. random walk drift, variance for the period effect) are then calculated based on the estimates of the latent parameters using the Box-Jenkins time series method. The two-stage approach is widely used in many empirical studies. Let us mention the works by Lee & Carter (1992), Lee & Miller (2001), Booth et al. (2002, 2001), Renshaw & Haberman (2006), Cairns et al. (2006a), Cairns et al. (2009), Cairns et al. (2011a), to name just a few.

As we have discussed in the previous chapter, the estimation for the hyper-parameters could be compromised if the noise to the respected latent parameters is large when the population is not sufficiently large.

2.3.2 Bayesian Approach and Updating Method

In the previous literature, the Bayesian approach and the simulation techniques are adopted to deal with the parameter uncertainty, e.g. Klugman (1989) (1989), Yashin et al. (2000) (2000), Dellaportas et al. (2001) (2001), Girosi & King (2003) (2003), to name just a few.

Possible incoherence may arise from the two-stage approach. Czado et al. (2005) model the French males mortality data with the Lee-Carter model and the two stages are combined into one with a Bayesian approach. The parameter estimates are obtained with Markov Chain Monte Carlo (MCMC) simulation. In the original paper, the authors use an autoregressive prior distribution with a linear mean for the period effect $\kappa_t^{(2)}$. They conclude that the Bayesian credibility intervals for the forecasts always include both Poisson Goodman and Lee-Carter's projections and the Bayesian point projections for the cohort 1961 are always higher than those generated by the Lee-Carter model and two-stage approach.

The prediction randomness given the two-stage approach (e.g. Lee & Carter (1992)) does not include the estimation error of the parameters in the model (Pedroza (2006)). There could be missing data in the mortality data. Therefore in order

to handle the missing data and incorporate the parameter uncertainty for the mortality forecasting, Pedroza (2006) proposed a Bayesian forecasting model based on a Lee-Carter model based state-space model [1] firstly introduced in Pedroza (2002) for predicting the U.S. males mortality data, incorporate with the Kalman filter updating algorithm. In the original paper, the author simulates the period effect $\kappa_t^{(2)}$ by following the approach introduced by Harrison & West (1999). The examples of the following work are by Kogure et al. (2009), Kogure & Kurachi (2010). Reichmuth & Sarferaz (2008) applied MCMC to a version of the Renshaw & Haberman (2003) model.

Cairns et al. (2011b) firstly applied MCMC to a two-population Age-Period-Cohort model by combining the Poisson likelihood for the death counts with the time series likelihood functions for the latent random period and cohort effects. In the original paper, the authors considered the scenario such that one population is significantly larger than the other. They adopted the methodology that the latent parameters of the larger population is estimated by a standard single-population model and the differences of the parameters between the smaller and the larger population are then modelled with a Bayesian approach incorporated with MCMC. By doing this, the authors employed the information of the larger population for estimating the parameters of the smaller population. Their research enlightens our study on to what extent we could "borrow" the information of a referencing large population to improve our fitting and forecasting.

In the actuarial literature, the Metropolis-Hastings (M-H) algorithm and its special version Gibbs sampler are widely adopted as the updating tool of MCMC for simulating the posterior distribution when it is not possible to directly draw samples from a complicated posterior distribution, e.g. Czado et al. (2005), Pedroza (2006), Cairns et al. (2011b), etc.

However, in this research a dynamical sampling method, known as Hamiltonian

[1] See Frühwirth-Schnatter (1994) and Carter & Kohn (1994) for simulating the state vector.

Monte Carlo (HMC) is used as a replacement of M-H method due to its limitations, e.g. random walk behaviour, which results in extremely slow convergence. Alder & Wainwright (1959) firstly derives the dynamical aspects and decades later Andersen (1980) introduce a stochastic element to replace sampling from the canonical distribution. We derive how the HMC algorithm is adopted for the simulation given our Bayesian model in Section 5.3.1, Chapter 5. Full detailed derivation, justification and discussion for the principle of HMC can be found in Neal (1993) and the principle of Hamiltonian dynamic system can be found in De Almeida (1990).

2.4 Stochastic Model, Notations for This Study

2.4.1 Measures of the Mortality

Let $m_c(t,x)$ be the crude death rate for age x in calendar year t, where $t = t_1, \ldots, t_{n_y}$ and $x = x_1, \ldots, x_{n_a}$ respectively. More specifically $m_c(t,x)$ can be written as

$$m_c(t,x) = \frac{\text{Number of deaths in calendar year t at age x}}{\text{Average exposure in calendar year t at age x}}.$$

We also define $q(t,x)$ to be the corresponding mortality rate, which is the probability that a person aged exactly x will die during year t.

The last measure is the force of mortality, denoted as $\mu(t,x)$, which represents the death rate for a group of individuals aged exactly x during year t to $t+dt$ given $dt \to 0$.

The relationship between $m(t,x)$, $q(t,x)$ and $\mu(t,x)$ can be written as:

$$m(t,x) = \mu(t,x), \tag{2.1}$$

$$m(t,x) = -\log(1-q(t,x)). \tag{2.2}$$

Equation (2.1) and (2.2) hold given the following assumptions:

- $\mu(t, x)$ remains constant over each t and x.
- The size of population remains constant at all ages over time.

Equation (2.2) is useful in our study since our stochastic model is formulated in terms of $q(t, x)$.

2.4.2 The Model

We denote by $D(t, x)$ the number of deaths during calendar year $t = t_1, \ldots, t_{n_y}$ at age $x = x_1, \ldots, x_{n_a}$ and by $E(t, x)$ the corresponding central exposure to risk.

We will fit the following Poisson model to the observed death data, see Cairns et al. (2009):

$$D(t,x)|\theta_1 \sim \text{Pois}(m(\theta_1,t,x)E(t,x)) \tag{2.3}$$

$$m(\theta_1,t,x) = -\log(1-q(\theta_1,t,x)) \tag{2.4}$$

$$\text{logit } q(\theta_1,t,x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_c^{(4)} \tag{2.5}$$

where we have the following interpretations:

- $\kappa_t^{(i)}$ is a period effect in year $t = t_1, \ldots, t_{n_y}$ for each $i = 1, 2, 3$,
- $\kappa = \{\kappa^{(1)}, \kappa^{(2)}, \kappa^{(3)}\}$, where $\kappa^{(i)} = \{\kappa_t^{(i)}\}_{t=t_1,\ldots t_{n_y}}$ for $i = 1, 2, 3$,
- $\gamma_c^{(4)}$ is the cohort effect for the cohort born in year $c = t - x$,
- $\gamma^{(4)} = \{\gamma_c^{(4)}\}_{c=t_1-x_{n_a},\ldots,t_{n_y}-x_1}$,
- $\bar{x}$ is the mean of the age range we use for our analysis,

- $\hat{\sigma}_x^2$ is the mean of $(x - \bar{x})^2$, and
- $\theta_{1,t,x} = (\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_c^{(4)})^T$

For convenience purpose, we let $\theta_1 = \theta_{1,t,x}$ and these two notations are used interchangeably. The notation $m(\theta_{1,t,x}, t, x)$ and $q(\theta_{1,t,x}, t, x)$ emphasis their dependence on θ_1 in year t aged x.

Recall Cairns et al. (2006a) firstly introduced the CBD model for an improved fit for the mortality rates $q(t, x)$ at a higher ages. The crude death rates $m_c(t, x)$ could be any non-negative number given advanced ages while the mortality rates are strictly non-negative numbers that are no greater than one. We therefore follow the same idea with Cairns et al. (2006a) and model the $q(t, x)$ instead. The logit function is applied to ensure the fitted rates are positive and no greater than one.

The vectorized version of the notations for data and parameters are as follows:

- $\boldsymbol{E} = \{E(t, x)\}_{t=t_1,\ldots,t_{n_y};x=x_1,\ldots,x_{n_a}}$, the $(t_{n_y} - t_1 + 1) \times (x_{n_a} - x_1 + 1)$ dimensioned matrix of exposure. More specifically

$$\boldsymbol{E} = \begin{bmatrix} E(t_1, x_1) & \ldots & E(t_1, x_{n_a}) \\ \vdots & \ddots & \vdots \\ E(t_{n_y}, x_1) & \ldots & E(t_{n_y}, x_{n_a}) \end{bmatrix}$$

- $\boldsymbol{D} = \{D(t, x)\} t = t_1, \ldots, t_{n_y}; x = x_1, \ldots, x_{n_a}$, the corresponding matrix of death counts. Similarly $\boldsymbol{D}$ can be written as

$$\boldsymbol{D} = \begin{bmatrix} D(t_1, x_1) & \ldots & D(t_1, x_{n_a}) \\ \vdots & \ddots & \vdots \\ D(t_{n_y}, x_1) & \ldots & D(t_{n_y}, x_{n_a}) \end{bmatrix}$$

- $\boldsymbol{t} = (t_1, \ldots, t_{n_y})$.

- $\boldsymbol{x} = (x_1, \ldots, x_{n_a})$.
- $\boldsymbol{\kappa}^{(i)} = (\kappa_{t_1}^{(i)}, \ldots, \kappa_{t_{n_y}}^{(i)})$ for $i = 1, 2, 3$, is the vector of period parameters. The operator T is the transpose of a matrix.
- $\boldsymbol{\kappa} = (\boldsymbol{\kappa}^{(1)}, \boldsymbol{\kappa}^{(2)}, \boldsymbol{\kappa}^{(3)})$.
- $\boldsymbol{\gamma}^{(4)} = (\gamma_{t_1 - x_{n_a}}^{(4)}, \ldots, \gamma_{t_{n_y} - x_1}^{(4)})$ is the vector of cohort parameters.
- $\boldsymbol{\theta}_1 = (\boldsymbol{\kappa}^{(1)}, \boldsymbol{\kappa}^{(2)}, \boldsymbol{\kappa}^{(3)}, \boldsymbol{\gamma}^{(4)})$, the vector of all latent parameters.
- $\boldsymbol{q}(\boldsymbol{\theta}_1) = \{q(\theta_1, t, x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, the matrix of mortality rates.
- $\boldsymbol{\theta}_{11} = \boldsymbol{\kappa}_{t=t_1} = (\kappa_{t_1}^{(1)}, \kappa_{t_1}^{(2)}, \kappa_{t_1}^{(3)})$.
- $\boldsymbol{\theta}_{12} = (\boldsymbol{\kappa}_{t_2}, \ldots, \boldsymbol{\kappa}_{t_{n_y}})$, where $\boldsymbol{\kappa}_t = (\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)})$ for $t = t_2, \ldots, t_{n_y}$.
- $\boldsymbol{\theta}_{13} = \gamma_{t_1 - x_{n_a}}^{(4)}$, the first cohort year in the data.
- $\boldsymbol{\theta}_{14} = (\gamma_{t_1 - x_{n_a}+1}^{(4)}, \ldots, \gamma_{t_{n_y} - x_1}^{(4)})$.
- $\boldsymbol{\theta}_2 =$ the vector of all the hyper-parameters.
- $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, the vector of complete parameters consists of sub-vectors for the latent parameters and sub-vectors for the hyper-parameters.

The reason for including the cohort effect is that it is a well established feature in some populations such as England and Wales, see Cairns et al. (2011b). We do not claim that this model is necessarily the best model for the datasets to be considered. However we select the model based on a particular set of model selection criterion studied in Cairns et al. (2009), including the BIC, robustness, residuals etc. The choice of "M7" here thus reflects the work of Cairns et al. (2009) namely that we want to use a model that fits the males from England and Wales well.

It is well known that the parameters in model (2.5) are not identifiable without imposing constraints on their values. It means that we can add a quadratic term with respect to $\boldsymbol{t}$ and $\boldsymbol{x}$ to the cohort effects $\boldsymbol{\gamma}$ and with corresponding adjustments

to the period effects $\boldsymbol{\kappa}$ without changing the value of $\boldsymbol{q}(\boldsymbol{\theta}_1)$. More specifically we can add $\phi_1 + \phi_2(t-x-\bar{x}) + \phi_3(t-x-\bar{x})^2$ to $\gamma^{(4)}_{t-x}$ and adjustments to $\kappa^{(i)}_t$ for $i = 1, 2, 3$ such that the fitted value of $q(\theta_1, t, x)$ remain unchanged, where ϕ_1, ϕ_2 and ϕ_3 are arbitrary values.

Nielsen & Nielsen (2014) discuss the impact of identifiability problems within stochastic mortality models on parameter estimation, hypothesis testing and forecasting. Currie (2016) discusses modelling with M7 by writing the model as a generalized linear model with a non-full rank design matrix. We follow Cairns et al. (2009) and apply the following constraints on $\boldsymbol{\theta}_1$:

$$\sum_{c\in C} \gamma^{(4)}_c = 0, \qquad \sum_{c\in C} c\gamma^{(4)}_c = 0, \qquad \sum_{c\in C} c^2\gamma^{(4)}_c = 0 \tag{2.6}$$

where $C = (t_1 - x_{n_a}, \ldots, t_{n_y} - x_1)$ is the set of all years of birth in a given dataset. (2.6) ensures that the least square estimator for ϕ_i is zero, that is $\hat{\phi}_i = 0$ for $i = 1, 2, 3$ if we fit the quadratic term $\phi_1 + \phi_2(t-x-\bar{x}) + \phi_3(t-x-\bar{x})^2$ to $\gamma^{(4)}_{t-x}$. More specifically, $\sum_{c\in C} \gamma^{(4)}_c = 0$ ensures that the estimates of $\gamma^{(4)}_c$ will fluctuate around zero. $\sum_{c\in C} c\gamma^{(4)}_c = 0$ and $\sum_{c\in C} c^2\gamma^{(4)}_c = 0$ remove both the linear and quadratic trend from the estimator. See the procedure of applying (2.6) in Appendix B.1.

In Chapters 3-4, identifiability constraints are defined as part of our model system to ensure all parameters are identifiable and provide a coherent framework for the consideration of confidence intervals and for hypothesis testing. One can freely adopt any reasonable set of constraints to the model and the study would be focusing on the results given the selected constraints. For the proposed Bayesian model in Chapters 5-7, the identifiability constraints are replaced by an auto-regressive zero mean-reverting time series prior distribution for the cohort effect.

Chapter 3

The Impact of Small Population on Parameter Uncertainty

In the previous chapter, we reviewed the empirical studies on a variety of methods for fitting and forecasting mortality data and introduced the stochastic model and notations we are using for the rest of the thesis. In this chapter, we will investigate the impact of a small population size on the uncertainty of parameter estimates and mortality projections.

3.1 Parameter Estimation

To estimate the parameters in model (2.5) on page 23 we apply maximum likelihood estimation. The log-likelihood function for our model (2.3) on page 23 is:

$$l_1(\boldsymbol{\theta}_1|\boldsymbol{D},\boldsymbol{E}) = \sum_{t,x} D(t,x)\log[E(t,x)m(\theta_1,t,x)] - E(t,x)m(\theta_1,t,x) - \log[D(t,x)!] \tag{3.1}$$

where $m(\theta_1,t,x)$ is given by (2.4) and (2.5). It is worth noticing that both the fitted mortality rates and the log-likelihood function $l(\boldsymbol{\theta}_1|\boldsymbol{D},\boldsymbol{E})$ are invariant to the choice

of the identifiability constraints mentioned in Chapter 2.

As mentioned earlier, in this chapter we are concerned with the consequences of small exposures, or population sizes, on the distribution of the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}_1$ of $\boldsymbol{\theta}_1$. To study the distribution of the MLE $\hat{\boldsymbol{\theta}}_1$ we will simulate death data $\boldsymbol{D} = \{D(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$ from the model in (2.3)–(2.5) using different exposure sizes and a given parameter vector $\boldsymbol{\theta}_{1,0} = \{\theta_{1,0}(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, where $\theta_{1,0}(t,x) = (\kappa_{t,0}^{(1)}, \kappa_{t,0}^{(2)}, \kappa_{t,0}^{(3)}, \gamma_{t-x,0}^{(4)})$. Note that for simplicity purposes, $\theta_{1,0}(t,x)$ and $\theta_{1,0}$ are used interchangeably for the given latent parameters at year t and age x.

To ensure that our results are relevant for typical values of $\boldsymbol{\theta}_1$ we first fit our model to death and exposure data observed in England and Wales during the years 1961 to 2011 for males aged 50 to 89. Note that we do not claim that this is the only choice of dataset. Any large population plus any model that is known to fit it well can be used for this study and we expect similar conclusions could be obtained given other referencing populations. The reason for choosing this particular dataset is that we have familiarity with the England and Wales data and the selected model fits a similar dataset well, see Cairns et al. (2009). We then fix $\boldsymbol{\theta}_{1,0}$ to be equal to the estimated parameter vector $\hat{\boldsymbol{\theta}}_1^{\text{EW}}$ for these data. Note that this is only an example for the true parameter vector $\boldsymbol{\theta}_{1,0}$ and our analysis can be applied to other choice of $\boldsymbol{\theta}_{1,0}$. Mortality data for England and Wales are obtained from the Human Mortality Database.[1] Note that we do not exclude short cohorts (cohort year with observations less than six) from the estimation since we are interested in how maximum likelihood (ML) method fits the short cohorts and the impact of small population sizes on the estimates.

The different exposure sizes used to simulate data in the remainder of this study will be relative to the exposure $\boldsymbol{E}_0 = \{E_0(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$ for a benchmark popula-

[1]Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at *www.mortality.org* or *www.humanmortality.de* (data downloaded on 16 February, 2014).

tion. For reasons of practical relevance and consistency with our choice of $\boldsymbol{\theta}_{1,0}$ the benchmark population is the male population in England and Wales unless stated otherwise.

3.2 Distribution of MLE in Finite Samples

For any given parameter vector $\boldsymbol{\theta}_{1,0}$ and benchmark exposure $\boldsymbol{E}_0$ we define the small-sample exposure as

$$\boldsymbol{E}^w = w\boldsymbol{E}_0$$

for a constant $w \leq 1$. More specifically, we have $\boldsymbol{E}^w = \{E^w(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, where $E^w(t,x) = wE_0(t,x)$. Table 3.1 shows the exposures for males in England and Wales for the year 2011 at selected ages. The total exposure for males in England and Wales in 2011 across all ages from 50 to 89 is $9,049,613$. The weights we consider in this study are $w = 1, 0.1, 0.01, 0.001$. The smallest population will therefore have an exposure of 43 at age 89 and 382 at age 50.

Age x		50	60	70	80	89
Exposure	EW	$381,797$	$307,825$	$213,455$	$134,966$	$42,640$

Table 3.1: The exposure for males in England and Wales (EW) in year 2011 at selected ages.

We then simulate N_1 scenarios for the death counts $\boldsymbol{D}^w = \{D^w(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$ using the model in (2.3)–(2.5) with $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,0}$. Through our simulation we obtain N_1 independent scenarios $\boldsymbol{D}_j^w = \{D_j^w(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$ for the death counts with

$$D_j^w(t,x) \sim \text{Pois}\big(m(\theta_{1,0},t,x)wE_0(t,x)\big) \qquad \text{for all } j = 1,\ldots,N_1. \tag{3.2}$$

A more general approach would be to consider a weights matrix $\boldsymbol{W} = \{w(t,x)\}$ for $t = t_1,\ldots,t_{n_y}$ and $x = x_1,\ldots,x_{n_a}$ allowing for weights to depend on age and calendar year. This would be particularly relevant when our proposed methodology

is applied to investigate the mortality of members of a pension scheme with a very different age structure than the age structure of the overall population in England and Wales. However, for clarity of presentation, we only consider a constant weight applied to all ages and calendar years.

3.2.1 MLE

Denote by

$$\hat{\boldsymbol{\theta}}^{w}_{1,j} = (\hat{\kappa}^{(1),w}_{t_1,j}, \ldots, \hat{\kappa}^{(1),w}_{t_{ny},j}, \hat{\kappa}^{(2),w}_{t_1,j}, \ldots, \hat{\kappa}^{(2),w}_{t_{ny},j}, \hat{\kappa}^{(3),w}_{t_1,j}, \ldots, \hat{\kappa}^{(3),w}_{t_{ny},j}, \hat{\gamma}^{(4),w}_{t_1-x_{na},j}, \ldots, \hat{\gamma}^{(4),w}_{t_{ny}-x_1,j})$$

the estimated parameter vector of MLEs for each simulated scenario j and each population size w. To obtain $\hat{\boldsymbol{\theta}}^{w}_{1,j}$ we maximise the log-likelihood function $l_1(\boldsymbol{\theta}_1|\boldsymbol{D}^w_j, \boldsymbol{E}^w)$ in Equation (3.1) subject to the constraints in (2.6) on page 26, that is,

$$\hat{\boldsymbol{\theta}}^{w}_{1,j} := \arg\max_{\boldsymbol{\theta}_1} l(\boldsymbol{\theta}_1|\boldsymbol{D}^w_j, \boldsymbol{E}^w). \tag{3.3}$$

See Appendix B.1 for the details of the procedure for applying the identifiability constrains in (2.6).

Asymptotic Distribution of the MLEs

Classical sampling theory[2] tells us that

$$\sqrt{w}\big(\hat{\boldsymbol{\theta}}^{w}_{1,j} - \boldsymbol{\theta}_{1,0}\big) \xrightarrow{\text{Dist}} N(\mathbf{0}, \boldsymbol{H}) \;\text{ as } w \to \infty$$

for some positive semi definite matrix $\boldsymbol{H}$. The derivation of $\boldsymbol{H}$ is shown as follows.

[2]The details of consistency, asymptotic normality and the efficiency of ML estimators are discussed in Kendall et al. (1987).

Rewrite Equation (3.1) for each w:

$$l_1(\boldsymbol{\theta}_1|\boldsymbol{D}_j^w, \boldsymbol{E}^w) = \sum_{t,x} f(\theta_1, t, x) - g(\theta_1, t, x) + h(t, x) \tag{3.4}$$

where

$$f(\theta_1, t, x) = D_j^w(t, x) \log[m(\theta_1, t, x)] \tag{3.5}$$

$$g(\theta_1, t, x) = E^w(t, x) m(\theta_1, t, x) \tag{3.6}$$

and $h(t, x)$ is independent of $\boldsymbol{\theta}_1$. Recall that the form of the parameter vector $\boldsymbol{\theta}_1$ with $4n_y + n_a - 1$ dimensions is

$$\boldsymbol{\theta}_1 = (\kappa_{t_1}^{(1)}, \ldots, \kappa_{t_{ny}}^{(1)}, \kappa_{t_1}^{(2)}, \ldots, \kappa_{t_{ny}}^{(2)}, \kappa_{t_1}^{(3)}, \ldots, \kappa_{t_{ny}}^{(3)}, \gamma_{t_1 - x_{n_a}}^{(4)}, \ldots, \gamma_{t_{ny} - x_1}^{(4)})^T.$$

The second derivative of $l_1(\boldsymbol{\theta}_1|\boldsymbol{D}_j^w, \boldsymbol{E}^w)$ with respect to $\boldsymbol{\theta}_1$ is

$$\frac{\partial^2 l_1}{\partial \boldsymbol{\theta}_1^2} = \sum_{t,x} \frac{\partial^2 f}{\partial \boldsymbol{\theta}_1^2} - \frac{\partial^2 g}{\partial \boldsymbol{\theta}_1^2}$$

It is worth to note that for every pair of (t, x), $m(\theta_1, t, x)$ is a single value. Thus the second derivative of f and g with respect to $\boldsymbol{\theta}_1$ is a Hessian matrix with $4n_y + n_a - 1$ rows and columns:

$$\frac{\partial^2 f}{\partial \boldsymbol{\theta}_1^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial \kappa_{t_1}^{(1)\ 2}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_1}^{(1)} \partial \kappa_{t_{ny}}^{(2)}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_1}^{(1)} \partial \kappa_{t_{ny}}^{(3)}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_1}^{(1)} \partial \gamma_{t_{ny} - x_1}^{(4)}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial \kappa_{t_{ny}}^{(2)} \partial \kappa_{t_1}^{(1)}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_{ny}}^{(2)\ 2}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_{ny}}^{(2)} \partial \kappa_{t_{ny}}^{(3)}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_{ny}}^{(2)} \partial \gamma_{t_{ny} - x_1}^{(4)}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial \kappa_{t_{ny}}^{(3)} \partial \kappa_{t_1}^{(1)}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_{ny}}^{(3)} \partial \kappa_{t_{ny}}^{(2)}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_{ny}}^{(3)\ 2}} & \cdots & \frac{\partial^2 f}{\partial \kappa_{t_{ny}}^{(3)} \partial \gamma_{t_{ny} - x_1}^{(4)}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial \gamma_{t_{ny} - x_1}^{(4)} \partial \kappa_{t_1}^{(1)}} & \cdots & \frac{\partial^2 f}{\partial \gamma_{t_{ny} - x_1}^{(4)} \partial \kappa_{t_{ny}}^{(2)}} & \cdots & \frac{\partial^2 f}{\partial \gamma_{t_{ny} - x_1}^{(4)} \partial \kappa_{t_{ny}}^{(3)}} & \cdots & \frac{\partial^2 f}{\partial \gamma_{t_{ny} - x_1}^{(4)\ 2}} \end{bmatrix}$$

and similarly for $\frac{\partial^2 g}{\partial \boldsymbol{\theta}_1^2}$. Thus the form of the element at row p and column q is

$$\frac{\partial^2 f}{\partial \kappa_p^{(i)} \partial \kappa_q^{(j)}} = -\frac{D_j^w(t,x)}{m(\theta_1,t,x)^2} \frac{\partial m(\theta_1,t,x)}{\partial \kappa_p^{(i)}} \frac{\partial m(\theta_1,t,x)}{\partial \kappa_p^{(j)}} + \frac{D_j^w(t,x)}{m(\theta_1,t,x)} \frac{\partial^2 m(\theta_1,t,x)}{\partial \kappa_p^{(i)} \partial \kappa_q^{(j)}},$$

where $i, j \in (1,2,3,4)$ and are not necessarily the same. Note that $\kappa_c^{(4)} = \gamma_c^{(4)}$ represents cohort effect for convenience for $c = t_1 - x_{n_a}, \ldots, t_{n_y} - x_1$. Same derivation can be done for function g and we have

$$\frac{\partial^2 g}{\partial \kappa_p^{(i)} \partial \kappa_q^{(j)}} = E^w(t,x) \frac{\partial^2 m(\theta_1,t,x)}{\partial \kappa_p^{(i)} \partial \kappa_q^{(j)}}.$$

Thus for each pair of (t,x), the expected value of the element of the second derivative of l_1 is:

$$E\left[\frac{\partial^2 (f-g)}{\partial \kappa_p^{(i)} \partial \kappa_q^{(j)}}\right] = -w \frac{E_0(t,x)}{m(\theta_1,t,x)} \frac{\partial m(\theta_1,t,x)}{\partial \kappa_p^{(i)}} \frac{\partial m(\theta_1,t,x)}{\partial \kappa_q^{(j)}}.$$

Thus we have the fisher information matrix given $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,0}$:

$$I(\boldsymbol{\theta}_{1,0}) = w \sum_{t,x} \frac{E_0(t,x)}{m(\theta_{1,0},t,x)} \left[\frac{\partial m(\theta_1,t,x)}{\partial \theta_1} \left(\frac{\partial m(\theta_1,t,x)}{\partial \theta_1}\right)^T\right]\Bigg|_{\theta_1=\theta_{1,0}}.$$

The asymptotic distribution of $(\hat{\boldsymbol{\theta}}_{1,j}^w - \boldsymbol{\theta}_{1,0})$ is a multi-variate normal distribution with mean vector $\mathbf{0}$ and co-variance matrix $I^{-1}(\boldsymbol{\theta}_{1,0})$ for $w \to \infty$. More specifically

$$\sqrt{w}\big(\hat{\boldsymbol{\theta}}_{1,j}^w - \boldsymbol{\theta}_{1,0}\big) \xrightarrow{\text{Dist}} N(\mathbf{0}, wI^{-1}(\boldsymbol{\theta}_{1,0})), \quad \text{as } w \to \infty$$

and hence $\boldsymbol{H} = wI^{-1}(\boldsymbol{\theta}_{1,0})$.

Therefore, we would expect that, even in a finite sample, the co-variance of the distribution of $\hat{\boldsymbol{\theta}}_{1,j}^w$ is approximately $w^{-1}\boldsymbol{H}$ and the correlations between different components of $\hat{\boldsymbol{\theta}}_{1,j}^w$ are approximately independent of the relative population size w. Using the simulated sample $\hat{\boldsymbol{\theta}}_{1,1}^w, \ldots, \hat{\boldsymbol{\theta}}_{1,N_1}^w$ for $\hat{\boldsymbol{\theta}}_1^w$, where

$$\hat{\boldsymbol{\theta}}_1^w = (\hat{\kappa}_{t_1}^{(1),w}, \ldots, \hat{\kappa}_{t_{n_y}}^{(1),w}, \hat{\kappa}_{t_1}^{(2),w}, \ldots, \hat{\kappa}_{t_{n_y}}^{(2),w}, \hat{\kappa}_{t_1}^{(3),w}, \ldots, \hat{\kappa}_{t_{n_y}}^{(3),w}, \hat{\gamma}_{t_1-x_{n_a}}^{(4),w}, \ldots, \hat{\gamma}_{t_{n_y}-x_1}^{(4),w})^T,$$

we can investigate the finite-sample co-variance and correlation matrices of $\hat{\boldsymbol{\theta}}_1^w$. In Figure 3.1 we plot a graphical representation of the correlation matrices of $\hat{\boldsymbol{\theta}}_1^w$ that we obtain for two values of w.

(a) Correlation between estimates, w=1

(b) Correlation between estimates, w=0.001

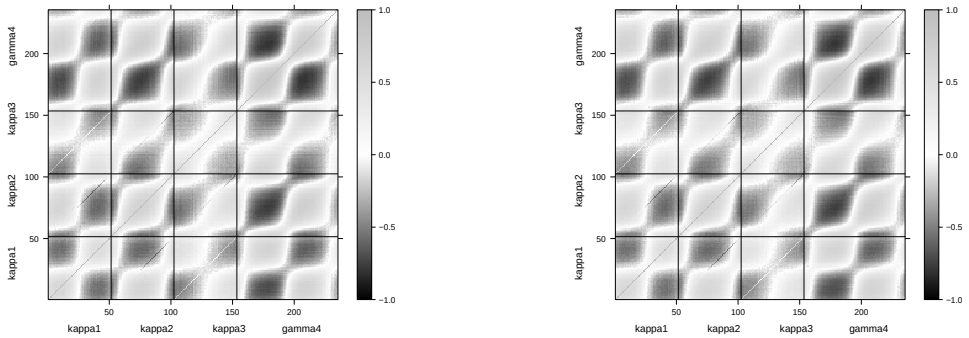


Figure 3.1: The empirical correlation matrix of the simulated parameter estimates $\hat{\boldsymbol{\theta}}_1^w$ for different values of the population size $w = 1$ and $w = 0.001$. The grid lines at 51.5, 102.5 and 153.5 are used to visually separate the parameters $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$, $\gamma_c^{(4)}$ from each other in both dimensions. For instance, the bottom left rectangle contains the correlations for $\hat{\kappa}_t^{(1),w}$ for the 51 years from 1960 to 2011.

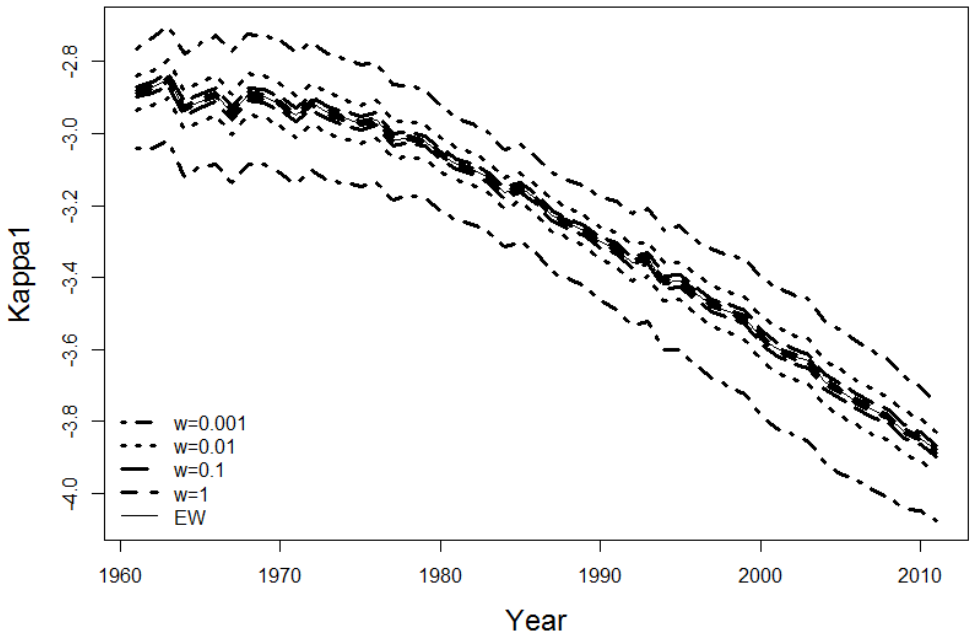
We conclude from Figure 3.1 that there are no significant differences between the empirical correlation matrices obtained from different population size, as predicted. However, we would not expect individual components of $\hat{\boldsymbol{\theta}}_1^w$ are independent from each other due to the identifiability constraints in Equation (2.6).

Distribution of the Finite-Sample MLEs $\hat{\boldsymbol{\theta}}_1^w$

To investigate the finite-sample distribution of the MLEs $\hat{\boldsymbol{\theta}}_1^w$ obtained from the bootstrap simulations further, we plot the empirical mean together with 90% confidence intervals for each of the components of $\hat{\boldsymbol{\theta}}_1^w$ in Figure 3.2-3.5.

We find for all population sizes considered that the empirical means of the estimates in the simulated death scenarios fluctuate around the true parameter value $\boldsymbol{\theta}_{1,0}$ (solid line), which indicates that the MLE is approximately unbiased for all considered population sizes. However, the standard deviation of the estimator depends

(a) Estimates of $\kappa_t^{(1)}$



(b) Log standard deviation of $\kappa_t^{(1)}$

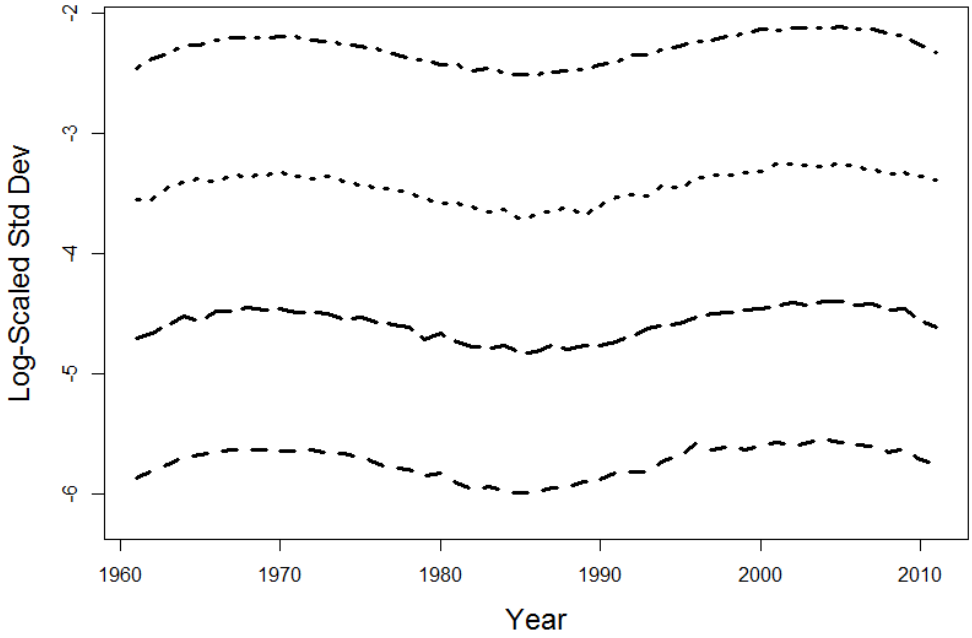
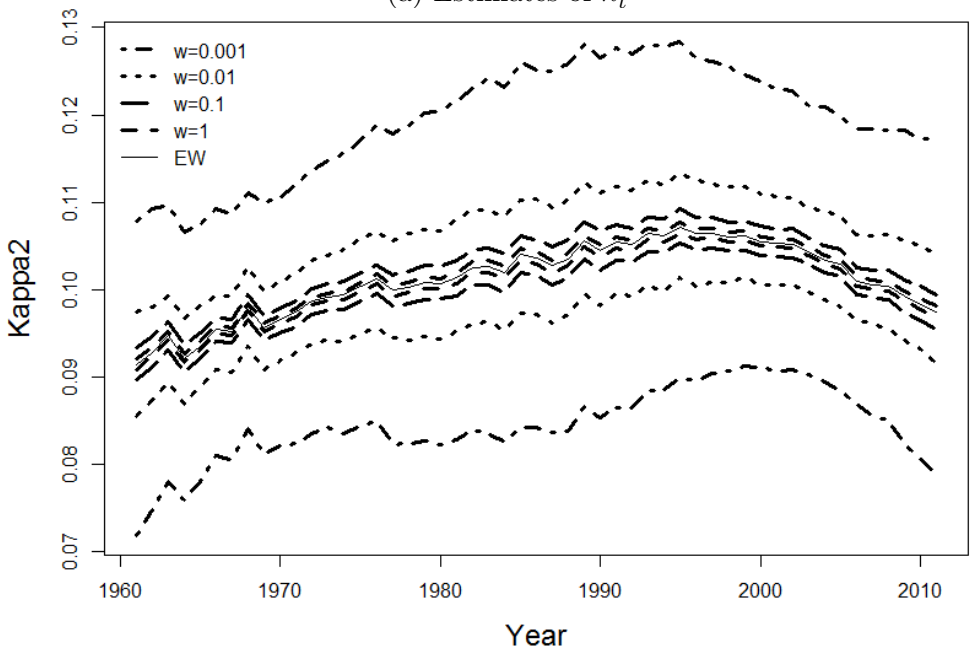


Figure 3.2: The distribution of MLEs: The mean and confidence interval (upper) and the log-scaled standard deviation (lower) of the MLEs of $\kappa_t^{(1)}$, with respect to year t, of populations with $w = 1$ (dashed line), $w = 0.1$ (long dashed line), 0.01 (dotted line), 0.001 (dot dashed line), together with the parameter estimates for the England and Wales population (solid line). Note: The upper bound of the CI in the left column is the 95% quantile of the distribution and the lower bound is the 5% quantile.

strongly on the size of the population, increasing significantly as the exposures get smaller as can be seen from the width of the confidence intervals.

(a) Estimates of $\kappa_t^{(2)}$



(b) Log standard deviation of $\kappa_t^{(2)}$

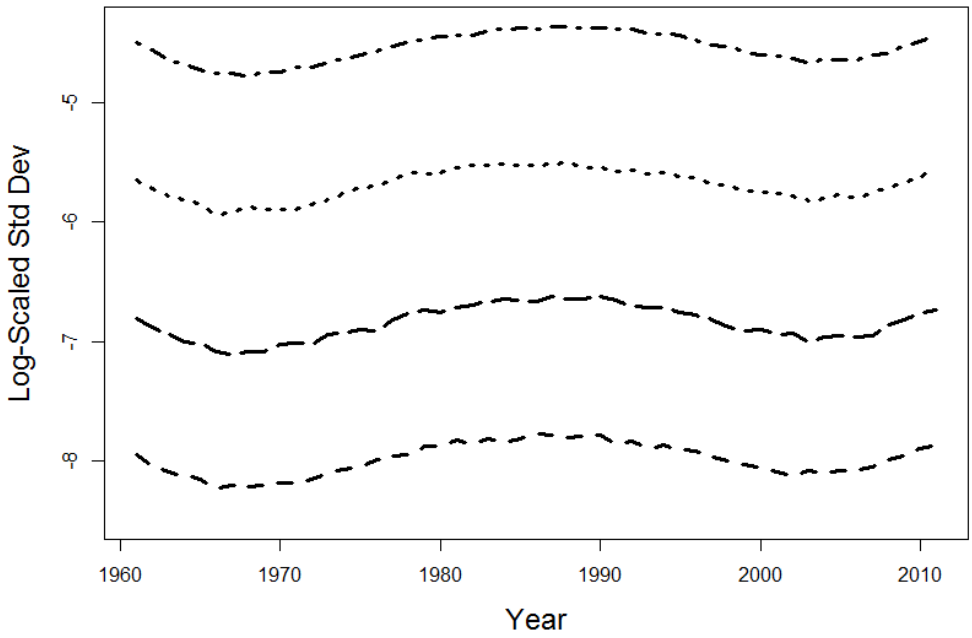
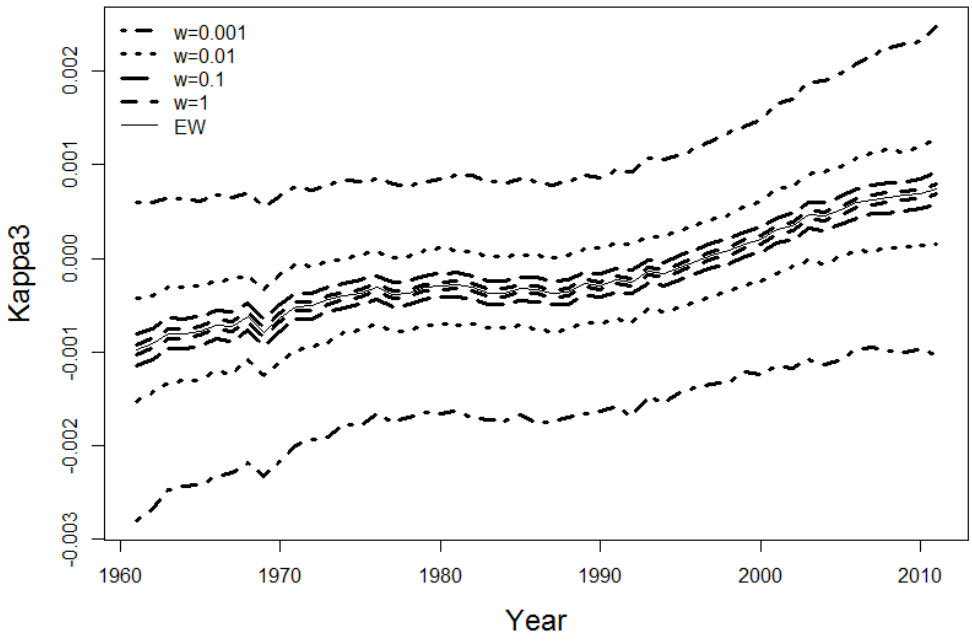


Figure 3.3: The distribution of MLEs: The mean and confidence interval (upper) and the log-scaled standard deviation (lower) of the MLEs of $\kappa_t^{(2)}$, with respect to year t, of populations with $w = 1$ (dashed line), $w = 0.1$ (long dashed line), 0.01 (dotted line), 0.001 (dot dashed line), together with the parameter estimates for the England and Wales population (solid line). Note: The upper bound of the CI in the left column is the 95% quantile of the distribution and the lower bound is the 5% quantile.

The relative levels of the parallel lines in the graphs on the right hand side of Figure 3.2-3.5 show that the level of fluctuation increases approximately by a

(a) Estimates of $\kappa_t^{(3)}$



(b) Log standard deviation of $\kappa_t^{(3)}$

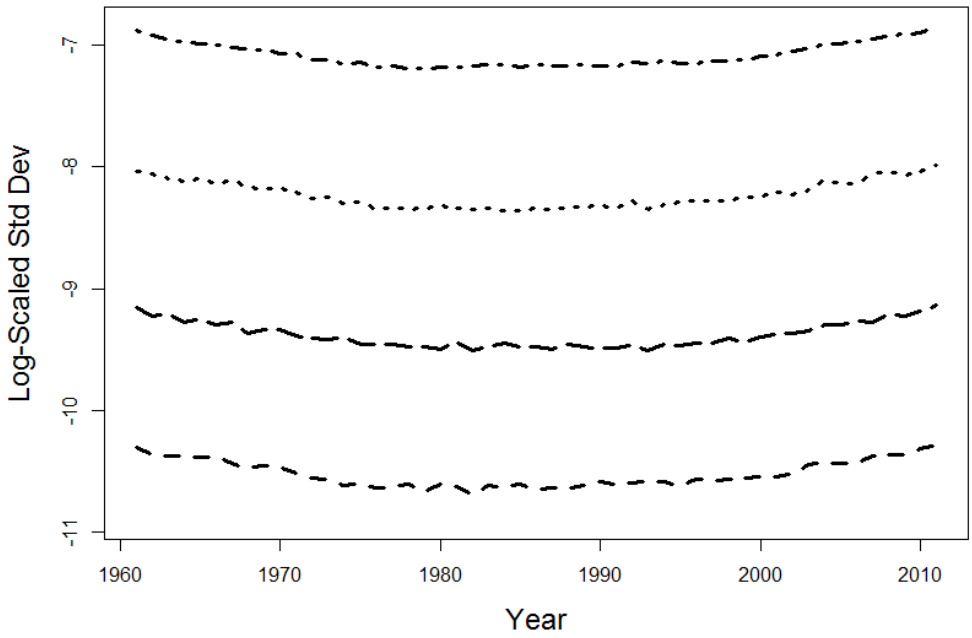
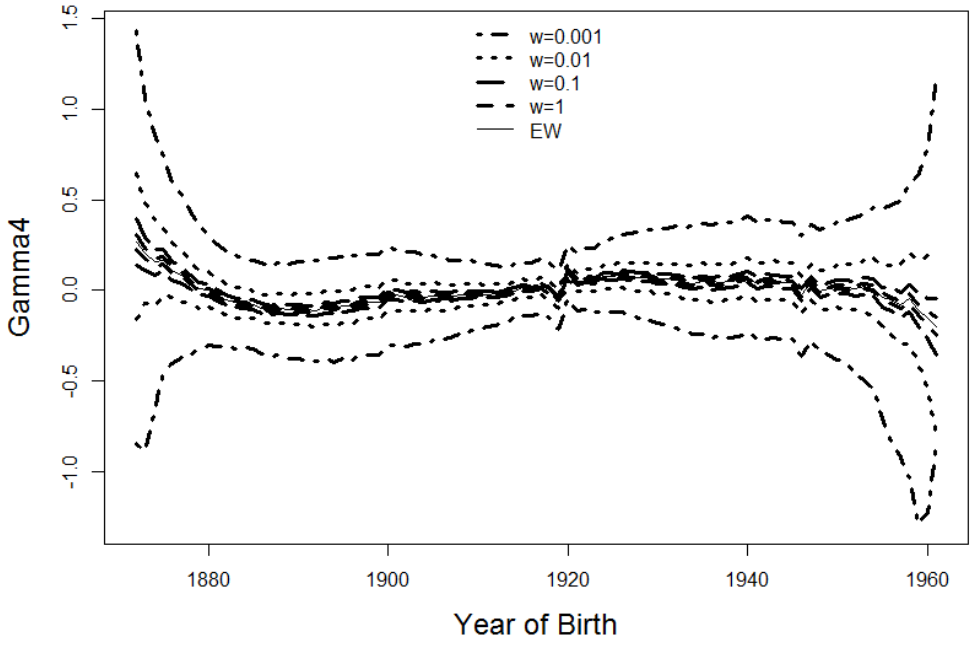


Figure 3.4: The distribution of MLEs: The mean and confidence interval (upper) and the log-scaled standard deviation (lower) of the MLEs of $\kappa_t^{(3)}$ with respect to year t, of populations with $w = 1$ (dashed line), $w = 0.1$ (long dashed line), 0.01 (dotted line), 0.001 (dot dashed line), together with the parameter estimates for the England and Wales population (solid line). Note: The upper bound of the CI in the left column is the 95% quantile of the distribution and the lower bound is the 5% quantile.

factor $\sqrt{n}$ if the population size is reduced by a factor $1/n$, which is consistent with the asymptotic co-variance matrix being proportional to $1/w$. It also suggests

(a) Estimates of $\gamma_c^{(4)}$



(b) Log standard deviation of $\gamma_c^{(4)}$

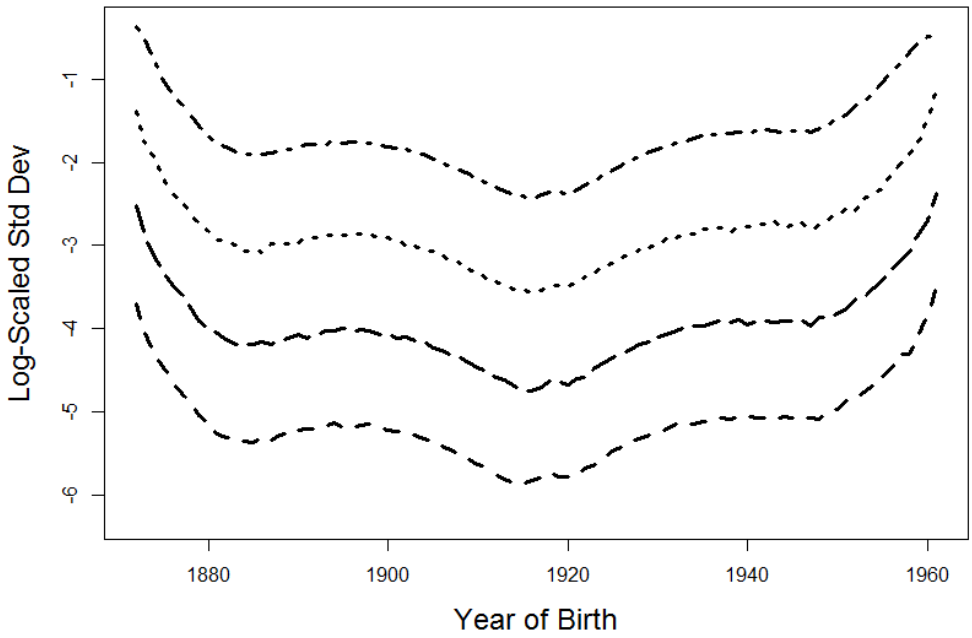


Figure 3.5: The distribution of MLEs: The mean and confidence interval (upper) and the log-scaled standard deviation (lower) of the MLEs of $\gamma_c^{(4)}$ with respect to year of birth c, of populations with $w = 1$ (dashed line), $w = 0.1$ (long dashed line), 0.01 (dotted line), 0.001 (dot dashed line), together with the parameter estimates for the England and Wales population (solid line). Note: The upper bound of the CI in the left column is the 95% quantile of the distribution and the lower bound is the 5% quantile.

that the variance is generally stable for all the period effects over years, which is not the case for the cohort effect with a wave shaped pattern. We notice that

the standard deviation of $\hat{\gamma}_c^{(4),w}$ widens out considerably at both ends, reflecting the reducing number of observations that we have for the younger and older cohorts. It is worth recalling that the finite-sample distribution of the MLE $\hat{\theta}^w$ varies if different sets of constraints are defined in the model system: that is, for a given w, the shapes of the various confidence intervals might be different if other identifiability constraints are used. The impact of the identifiability constraints in our study can be removed by calculating the following quantities of the point estimates: $\Delta^3\hat{\kappa}_t^{(1),w}$, $\Delta^2\hat{\kappa}_t^{(2),w}$, $\Delta\hat{\kappa}_t^{(3),w}$ and $\Delta^3\hat{\gamma}_c^{(4),w}$ for $w = 1, 0.1, 0.01, 0.001$, where Δ^k represents the k^{th} order difference[3]. The finite-sample distribution of these quantities and the corresponding standard deviation are shown in Figure 3.6, where unsurprisingly the right column implies that our conclusion regarding the proportional relationship between the variance and the population size holds.

3.3 Mortality Projections

While fitting the model in (2.3)–(2.5) to observed mortality data only requires the estimation of the period effects $\boldsymbol{\kappa}_t = (\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)})^T$ and the cohort effect $\gamma_c^{(4)}$, projecting mortality rates into the future requires a model for values of κ_t for $t > t_{n_y}$ where t_{n_y} is the last year for which mortality data are available. Similarly, future values of the cohort effect $\gamma^{(4)}$ are also required.

The most common approach to obtain future values of κ and $\gamma^{(4)}$ is to consider these parameter vectors as observed trajectories of stochastic processes and fit a parametric time series model to each trajectory. In the following we will fit a three-dimensional random walk to $\boldsymbol{\kappa}_t$ and a stationary AR(1) model to $\gamma_c^{(4)}$, as in Cairns et al. (2009). We will then discuss the estimation of the parameters of those models based on the values of $\boldsymbol{\theta}_{1,0}$ and $\hat{\boldsymbol{\theta}}_{1,j}^{w}$ for different values of w. This will allow us to investigate the impact of the relative population size w on the estimators for the

[3]See Appendix B.1 for how the effect of the constraints are removed.

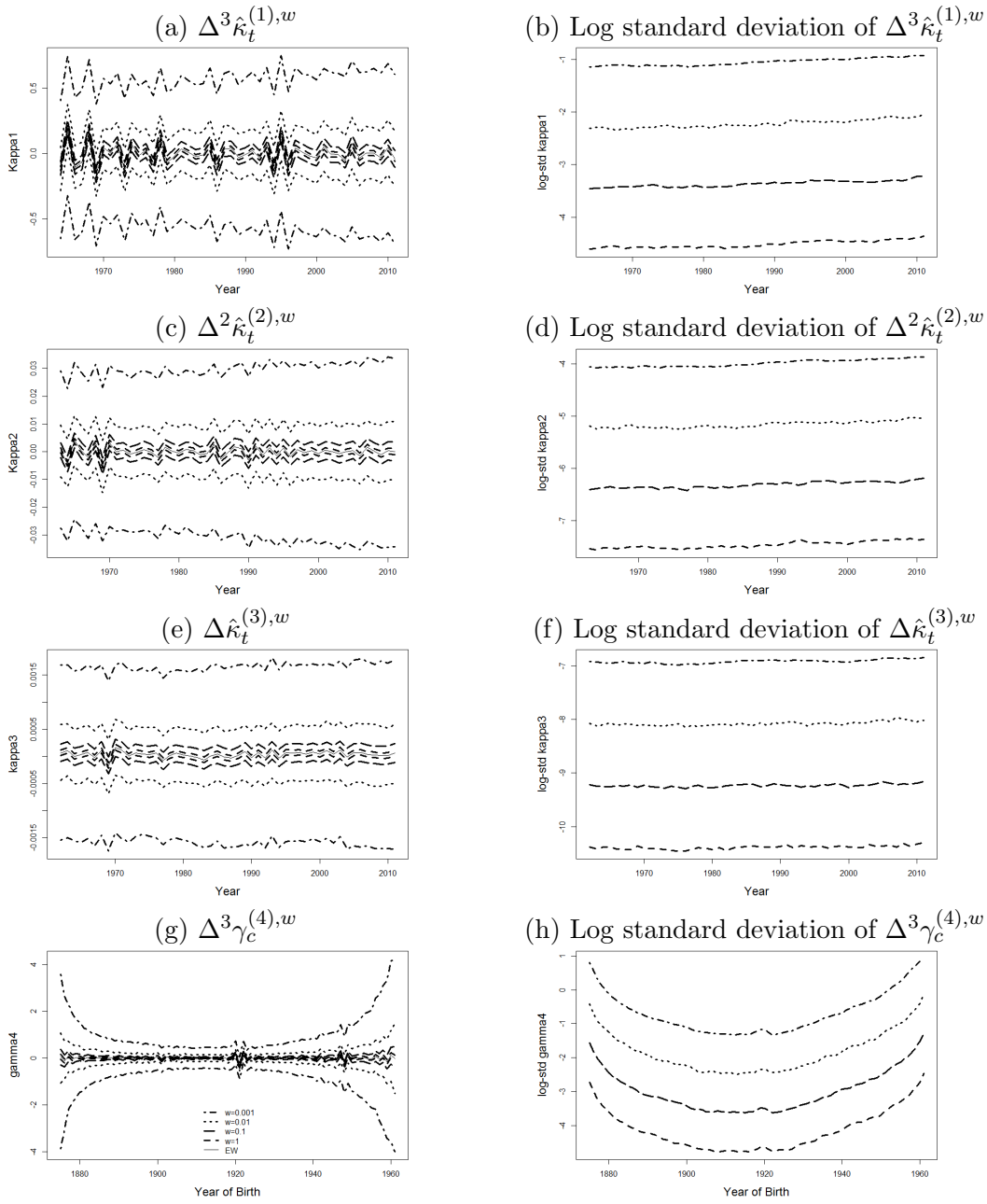


Figure 3.6: The distribution of k^{th} order difference of MLEs: The mean and confidence interval (left column) and the log-scaled standard deviation (right column) of $\Delta^3\hat{\kappa}_t^{(1),w}$, $\Delta^2\hat{\kappa}_t^{(2),w}$, $\Delta\hat{\kappa}_t^{(3),w}$ and $\Delta^3\hat{\gamma}_c^{(4),w}$, with respect to year t and year of birth c respectively, of populations with $w = 1$ (dashed line), $w = 0.1$ (long dashed line), 0.01 (dotted line), 0.001 (dot dashed line), together with the parameter estimates for the England and Wales population (solid line). Note: The upper bound of the CI in the left column is the 95% quantile of the distribution and the lower bound is the 5% quantile.

parameters of the κ and $\gamma^{(4)}$ process.

For the estimation of those parameters and the projections of the period effects

and the cohort effect we will consider two approaches. Firstly, we will use a frequent approach to obtain point estimates of the process parameters ignoring any uncertainty about those estimates. In our further analysis we will follow a Bayesian approach to incorporate parameter uncertainty into our mortality projections.

3.3.1 Projecting Period Effects

As mentioned above, we model the period effects $\boldsymbol{\kappa}_t$ as a three-dimensional normal random walk.

$$\Delta\boldsymbol{\kappa}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \tag{3.7}$$

where $\Delta\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1}$ and the $\boldsymbol{\epsilon}_t$ are independent identically distributed $(i.i.d)$ random vectors with a multi-variate normal distribution with mean vector $\mathbf{0}$ and co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}} = \{V_\epsilon(i,k)\}$ for $i,k = 1,2,3$. More specifically $\boldsymbol{\epsilon}_t \sim MVN(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\epsilon}})$. The parameter vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^T$ is the 3×1 drift vector of the random walk.

Point Estimators

Having generated N_1 scenarios for the number of deaths according to (3.2) and having estimated the parameter vector $\hat{\boldsymbol{\theta}}^w_{1,j}$ in each scenario j as in (3.3), we can now apply the random walk model to the period effects in $\boldsymbol{\theta}_{1,0}$ and $\hat{\boldsymbol{\theta}}^w_{1,j}$ for every generated scenario j.

Denote as $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)^T$ the point estimator for $\boldsymbol{\mu}$ and $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}} = \{\hat{V}_\epsilon(i,k)\}$ the point estimator for the co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, where $i,k = 1,2,3$.

We then apply the usual (i.e. maximum likelihood) point estimators $\hat{\boldsymbol{\mu}}^w_j$ and $\hat{\boldsymbol{V}}^{\,w}_{\boldsymbol{\epsilon},j}$ for each simulated scenario of $\boldsymbol{D}^w_j$. For $t = t_1, \ldots, t_{n_y}$, the estimators for the

three components of $\hat{\boldsymbol{\mu}}_j^w$ (scenario j) are

$$\hat{\mu}_{i,j}^w = \frac{1}{n_y - 1} \sum_{t=t_2}^{t_{ny}} \left(\hat{\kappa}_{t,j}^{(i),w} - \hat{\kappa}_{t-1,j}^{(i),w} \right); \qquad i = 1, 2, 3 \tag{3.8}$$

and the entries of the estimated 3×3 co-variance matrix $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon},j}^{\ w}$ are

$$\hat{V}_{\epsilon,j}^w(i,k) = \frac{1}{n_y - 1} \sum_{t=t_2}^{t_{n_y}} \Big[\left(\Delta\hat{\kappa}_{t,j}^{(i),w} - \hat{\mu}_{i,j}^w \right) \left(\Delta\hat{\kappa}_{t,j}^{(k),w} - \hat{\mu}_{k,j}^w \right) \Big]; \qquad i,k = 1,2,3. \tag{3.9}$$

The corresponding estimators for $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the true trajectory $\boldsymbol{\theta}_{1,0}$ are defined similarly.

Bayesian Estimation – Parameter Uncertainty

As mentioned earlier we model uncertainty about the parameters $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ by applying a Bayesian approach to estimation. We denote by p the density of the prior joint distribution of the two parameters.

The choice of the prior distribution for $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ follows the non-informative prior described in Gelman et al. (2014). We start with a short review of a non-informative prior distribution for a multi-variate normal model given unknown mean and variance. For a multi-variate normal distribution:

$$\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3.10}$$

where $\boldsymbol{y}$ is a vector with length d and $\boldsymbol{\Sigma}$ is a $d \times d$ co-variance matrix. Given unknown drift, $\boldsymbol{\mu}$ and co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, a common choice of a non-informative prior distribution for $(\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}})$ is the Jeffreys prior density,

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\frac{d+1}{2})} \tag{3.11}$$

where $|.|$ is the determinant of a matrix. Note that $|\boldsymbol{\Sigma}|$ always exists as a co-variance matrix that is positive definite and symmetric. Then the corresponding posterior distribution for $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma}|\boldsymbol{y} \sim \text{Inverse Wishart}(n-1, \boldsymbol{S}), \tag{3.12}$$

where n is the number of samples of $\boldsymbol{y}$ and

$$\boldsymbol{S} = \sum_{i=1}^{n} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T. \tag{3.13}$$

The conditional posterior for $\boldsymbol{\mu}$ is

$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{y} \sim MVN(\bar{\boldsymbol{y}}, \frac{\boldsymbol{\Sigma}}{n}). \tag{3.14}$$

See further details of derivation and comments on Page 81 in Gelman et al. (2014).

Assuming that we have no prior knowledge about the true values of $\boldsymbol{\mu}$ and $\boldsymbol{V_\epsilon}$, we use the Jeffreys prior density

$$p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon}) \propto |\boldsymbol{V_\epsilon}|^{-\frac{4}{2}},$$

where $|\boldsymbol{V_\epsilon}|$ is the determinant of $\boldsymbol{V_\epsilon}$ (see for example, page 81, Gelman et al. (2014)).

Denote by

$$\Delta\hat{\boldsymbol{\kappa}}_j^w = \begin{bmatrix} \Delta\hat{\boldsymbol{\kappa}}_j^{(1),w} \\ \Delta\hat{\boldsymbol{\kappa}}_j^{(2),w} \\ \Delta\hat{\boldsymbol{\kappa}}_j^{(3),w} \end{bmatrix}$$

where $(\Delta\hat{\boldsymbol{\kappa}}_j^{(i),w})^T = (\Delta\hat{\kappa}_{t_2,j}^{(i),w}, \ldots, \Delta\hat{\kappa}_{t_{ny},j}^{(i),w})$ for $i = 1, 2, 3$. Using this prior distribution in each scenario j, the posterior distribution is given by the Inverse Wishart

distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ and a multivariate normal distribution for $\boldsymbol{\mu}$, that is,

$$\left(\tilde{\boldsymbol{V}}^{w}_{\boldsymbol{\epsilon},j}\right)^{-1}|\Delta\hat{\boldsymbol{\kappa}}^{w}_{j} \sim \text{Wishart}(n_y-2,(n_y-1)^{-1}(\hat{\boldsymbol{V}}^{w}_{\boldsymbol{\epsilon},j})^{-1}) \tag{3.15}$$

$$\tilde{\boldsymbol{\mu}}^{w}_{j}|\tilde{\boldsymbol{V}}^{w}_{\boldsymbol{\epsilon},j},\Delta\hat{\boldsymbol{\kappa}}^{w}_{j} \sim N(\hat{\boldsymbol{\mu}}^{w}_{j},(n_y-1)^{-1}\tilde{\boldsymbol{V}}^{w}_{\boldsymbol{\epsilon},j}) \tag{3.16}$$

where $\hat{\boldsymbol{\mu}}^{w}_{j}$ and $\hat{\boldsymbol{V}}^{w}_{\boldsymbol{\epsilon},j}$ are the estimates obtained from $\hat{\boldsymbol{\theta}}^{w}_{1,j}$ as defined in (3.8) and (3.9) (See for example, Box & Tiao (2011) or Gelman et al. (2014) for the derivation).

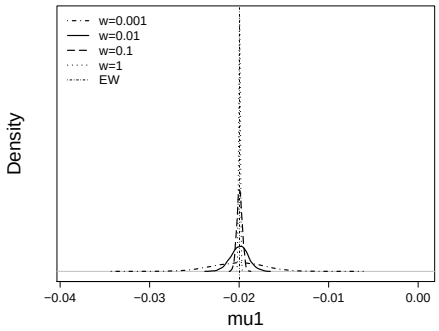
Empirical Comparison

For our empirical study we simulate $N_1 = 1000$ scenarios for different values of w and plot the empirical density of the point estimator $\hat{\boldsymbol{\mu}}^{w}$ in (3.8) based in the sample $\hat{\boldsymbol{\mu}}^{w}_{1},\ldots,\hat{\boldsymbol{\mu}}^{w}_{N_1}$ on the left hand side of Figure 3.7, where $\hat{\boldsymbol{\mu}}^{w}_{j} = (\hat{\mu}^{w}_{1,j},\hat{\mu}^{w}_{2,j},\hat{\mu}^{w}_{3,j})^{T}$ for $j = 1,\ldots,N_1$ defined in (3.8) and $\hat{\boldsymbol{\mu}}^{w} = (\hat{\mu}^{w}_{1},\hat{\mu}^{w}_{2},\hat{\mu}^{w}_{3})^{T}$. To incorporate parameter uncertainty we draw a further sample of size $M = 100$ from the posterior distribution of $\tilde{\boldsymbol{\mu}}^{w}_{j}$ in (3.16) in each scenario $j = 1,\ldots,N_1$. The empirical density of $\tilde{\boldsymbol{\mu}}^{w} = (\tilde{\mu}^{w}_{1},\tilde{\mu}^{w}_{2},\tilde{\mu}^{w}_{3})^{T}$ from these $N_1 \times M$ realisations is shown on the right hand side of Figure 3.7.

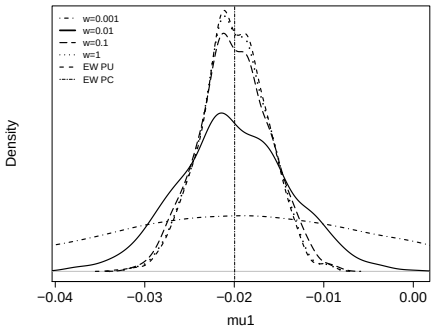
By comparing the densities in the two columns of that figure we observe that the additional parameter uncertainty increases the variance of the empirical distributions of the drift estimators. This can be explained by investigating the source of uncertainty to the drift. The variation to the point estimator $\hat{\mu}^{w}_{i}$ with no allowance for parameter uncertainty comes from the Poisson noise in the number of deaths from the bootstrap simulations, while the variance of the Bayesian estimator $\tilde{\mu}^{w}_{i}$ with allowance for extra parameter uncertainty also includes the uncertainty (Equation (3.16)) from the posterior distribution given the Poisson noise.

We also find in Figure 3.7 that the size of a population affects the uncertainty about the drift vector $\boldsymbol{\mu}$. The variance of the empirical finite sample distribution of

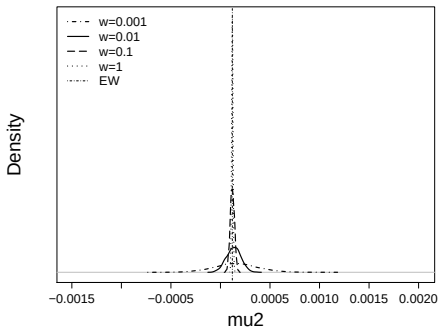
(a) Density of $\hat{\mu}_1^w$ with No Allowance to Parameter Uncertainty



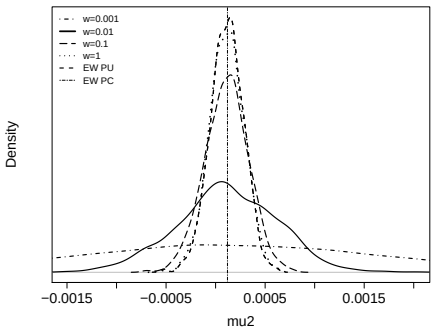
(b) Density of $\tilde{\mu}_1^w$ with Allowance to Parameter Uncertainty



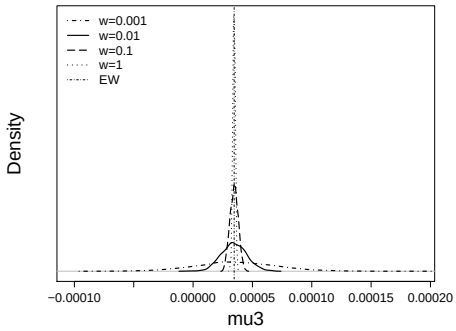
(c) Density of $\hat{\mu}_2^w$ with No Allowance to Parameter Uncertainty



(d) Density of $\tilde{\mu}_2^w$ with Allowance to Parameter Uncertainty



(e) Density of $\hat{\mu}_3^w$ with No Allowance to Parameter Uncertainty



(f) Density of $\tilde{\mu}_3^w$ with Allowance to Parameter Uncertainty

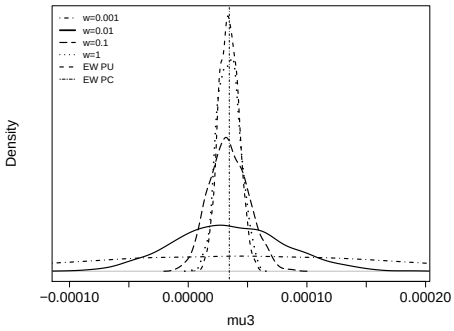


Figure 3.7: The impact of population size on the distribution of the random walk drift, from population of $w = 1$ (dotted line), $w = 0.1$ (long-dashed line), $w = 0.01$ (solid line), $w = 0.001$ (dot dashed line) and England and Wales (vertical line). The left column is the density of drift without allowance to the parameter uncertainty; the right column is the density of drift with allowance to the parameter uncertainty. Note the right column also includes the density for $\tilde{\mu}_3^{\mathrm{EW}}$ (dashed line) for the England and Wales dataset. PU means parameter uncertain and PC represents parameter certain.

both estimators, $\hat{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}$ decreases significantly when the population size increases, although the difference between $w = 1$ and $w = 0.01$ is rather small as is obvious for the Bayesian estimator $\tilde{\boldsymbol{\mu}}$.

However, for smaller values of w we find that the population size has a much more pronounced effect on the variance. For example, the range of likely values of $\tilde{\boldsymbol{\mu}}^{0.001}$ is significantly wider than the range of values of $\tilde{\boldsymbol{\mu}}^{0.1}$ and $\tilde{\boldsymbol{\mu}}^{1}$ reflecting the uncertainty about $\boldsymbol{\mu}$ that we have already observed in Figure 3.2a. The same argument applies to the point estimators $\hat{\boldsymbol{\mu}}$.

To investigate parameter uncertainty further we calculate the standard deviations for the distributions of $\hat{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}$ in Figure 3.7. Those standard deviations are shown in Table 3.2. We observe that the standard deviation of the point estimator

		$i = 1$	$i = 2$	$i = 3$
Point estimator $\hat{\mu}_i^w$	w=1	0.0000966	0.0000071	0.00000113
	w=0.1	0.0003050	0.0000217	0.00000343
	w=0.01	0.0009777	0.0000727	0.00001068
	w=0.001	0.0028787	0.0002206	0.00003387
Bayesian estimator $\tilde{\mu}_i^w$	w=1	0.00369	0.000173	0.00000936
	w=0.1	0.00396	0.000222	0.0000162
	w=0.01	0.00620	0.000505	0.0000458
	w=0.001	0.01689	0.001566	0.0001478

Table 3.2: The finite sample standard deviation of $\hat{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}$.

$\boldsymbol{\mu}$ is increased approximately by a factor $\sqrt{10}$ if the population size is reduced by a factor 10. The situation becomes more complicated when for the Bayesian estimator $\tilde{\boldsymbol{\mu}}$ since the variance of the posterior distribution affects the finite sample variance of the estimator. There is no obvious proportional relationship between population size and variation, which suggests that the size of the population is not the only determinant of the variance of $\tilde{\boldsymbol{\mu}}$.

To investigate the impact of the relative population size w and the inclusion of parameter uncertainty on the empirical distribution of the estimated co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ of the random walk in (3.7) we compare the empirical means of $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$

and the Bayesian estimator matrix $\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}} = \{\tilde{V}_{\boldsymbol{\epsilon}}(i,k)\}$, where $i,k = 1,2,3$, obtained for different values of w, denoted by $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ w}$ and $\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}$ respectively. The means of the estimated co-variance matrix $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ w}$ are:

$$E[\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ 1}] = \begin{bmatrix} 6.82\times10^{-4} & 2.12\times10^{-5} & 5.42\times10^{-7} \\ 2.12\times10^{-5} & 1.41\times10^{-6} & 2.99\times10^{-8} \\ 5.42\times10^{-7} & 2.99\times10^{-8} & 4.30\times10^{-9} \end{bmatrix}$$

$$E[\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ 0.01}] = \begin{bmatrix} 18.7\times10^{-4} & -1.52\times10^{-5} & 4.61\times10^{-6} \\ -1.52\times10^{-5} & 1.25\times10^{-5} & -1.89\times10^{-7} \\ 4.61\times10^{-6} & -1.89\times10^{-7} & 0.99\times10^{-7} \end{bmatrix}$$

and the mean values of the Bayesian estimator $\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}$ are

$$E[\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{1}] = \begin{bmatrix} 7.58\times10^{-4} & 2.37\times10^{-5} & 6.06\times10^{-7} \\ 2.37\times10^{-5} & 1.58\times10^{-6} & 3.35\times10^{-8} \\ 6.06\times10^{-7} & 3.35\times10^{-8} & 4.79\times10^{-9} \end{bmatrix}$$

$$E[\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{0.01}] = \begin{bmatrix} 20.90\times10^{-4} & -1.74\times10^{-5} & 5.11\times10^{-6} \\ -1.74\times10^{-5} & 1.39\times10^{-5} & -2.10\times10^{-7} \\ 5.11\times10^{-6} & -2.10\times10^{-7} & 1.11\times10^{-7} \end{bmatrix}.$$

The corresponding estimated co-variance matrices, $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ EW}$, for England and Wales based on the single sample paths of $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}^{(4)}$ and the mean of Bayesian estimator

are

$$\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{EW} = \begin{bmatrix} 6.70 \times 10^{-4} & 2.16 \times 10^{-5} & 4.94 \times 10^{-7} \\ 2.16 \times 10^{-5} & 1.31 \times 10^{-6} & 3.18 \times 10^{-8} \\ 4.94 \times 10^{-7} & 3.18 \times 10^{-8} & 3.30 \times 10^{-9} \end{bmatrix}$$

$$E[\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{EW}] = \begin{bmatrix} 5.49 \times 10^{-4} & 1.80 \times 10^{-5} & 1.05 \times 10^{-7} \\ 1.80 \times 10^{-5} & 1.07 \times 10^{-6} & 2.19 \times 10^{-8} \\ 1.05 \times 10^{-7} & 2.19 \times 10^{-8} & 3.06 \times 10^{-9} \end{bmatrix}.$$

Comparing the mean values of $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$ and $\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$ with the estimates obtained from the England and Wales data we find significant differences in the estimated co-variance. In particular, for smaller populations (e.g. w=0.01) sampling variation pushes up significantly estimates of the co-variance matrix. In addition, sampling variation also widens the distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ around these mean values for smaller values of w. On the other hand, for a given value of w, the inclusion of full Bayesian parameter uncertainty moving from $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$ to $\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$ has rather less of an impact.

Finally, the projected parameters based on the Bayesian estimates $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$ are shown in Figure 3.10. As we expected, the prediction intervals reflecting the uncertainty about future values of the period effects are very wide for small populations. The plots also suggest that the means of the co-variances are right biased compared to the estimate for England and Wales. The variance of projection for all the populations are much higher than the estimates, due to the additional normal randomness added in the forecasting model by simulating the sample paths for $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}^{(4)}$. However, the left column shows that there is no obvious proportional relationship between the population size and projection variance. By investigating the mean co-variance matrices, we find that the increase of $E[\boldsymbol{V}_{\boldsymbol{\epsilon}}^{w}(3,3)]$ from $w = 0.01$ to $w = 1$ is of the highest among the three period effects, which suggests that the standard deviation of projection for $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\kappa}^{(2)}$ is not as sensitive as $\boldsymbol{\kappa}^{(3)}$ to the

change of population size.

3.3.2 Projecting the Cohort Effect

As mentioned earlier we fit an $AR(1)$ model to the cohort effect. We will not investigate how additional parameter uncertainty influences mortality projections, but will only use point estimates for the parameters in the $AR(1)$ model. To be precise, our model is given by

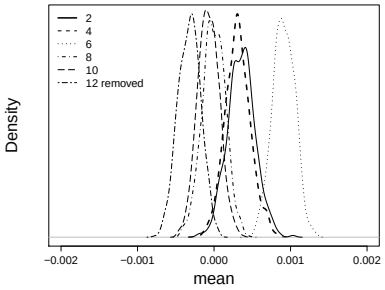
$$\gamma_{c+1}^{(4)} = \alpha_0 + \alpha_\gamma \left(\gamma_c^{(4)} - \alpha_0\right) + \epsilon_{c+1} \tag{3.17}$$

Figure 3.5a shows that the variance of the estimated cohort effect is very large for the very early and very late years of birth, in particular, for $w = 0.01$ and 0.001. This is a consequence of the very few observations available for those cohorts. We therefore remove the cohorts with six or less observations. Cohorts are removed equally from the beginning and the end.

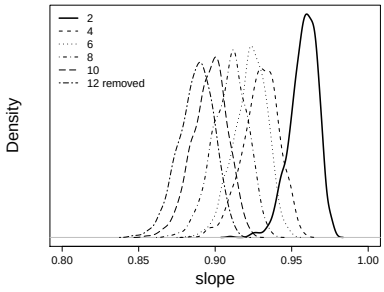
However, removing short cohorts could significantly influence the estimated values of the parameters. To investigate the effect of removing short cohorts in more detail we plot the empirical densities for the parameters in (3.17) based on the estimated parameters in each simulated scenario j for $w = 1$ (Figure 3.8). We find that the distribution of $\hat{\alpha}_0$ is not significantly affected by removing cohorts which is also the case for the estimated variance of $\epsilon_c$when more than 4 cohorts are removed. Further we notice that the variance of the estimators for all three parameters stays approximately unchanged regardless of how many cohorts are removed.

After having removed cohorts with six or less observations from the data, we fit the $AR(1)$ model in (3.17) to the rest of the cohort effects. The resulting density of the parameter estimates of the model are shown in Figure 3.9. All of the parameter estimates and the standard deviation of error terms appear to be biased relative to

(a) The Density of $\hat{\alpha}_0^w$



(b) The Density of $\hat{\alpha}_\gamma^w$



(c) The Density of the Standard Deviation of ϵ_c^w

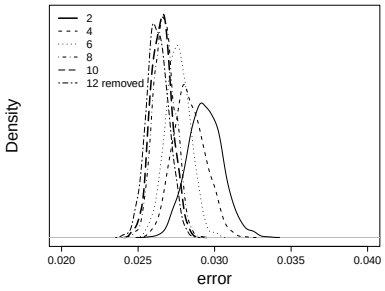


Figure 3.8: The effect of removing the cohort effects of short cohorts on the distribution of the parameter estimators of the $AR(1)$ model for $w = 1$. We investigate the distribution of parameter estimates when the first and last $1, 2, 3, 4, 5, 6$ cohorts are removed

the estimate for England and Wales, regardless of the size of population. However, we find that reducing the population size will greatly increase the mean bias as well as the uncertainty.

We now forecast the cohort effect from $\hat{\gamma}^{(4),w}_{t_{n_y}-x_1-6} = \hat{\gamma}^{(4),w}_{1955}$ instead of $\hat{\gamma}^{(4),w}_{1961}$ and the result is shown in Figure 3.10d. The variation in the projected cohort effects for the years 1956 to 1961 now comes from the Poisson and Normal randomness, which is not as great as variation at the two tails of the estimates observed in Figure 3.5a where no cohorts have been removed. Within the sample, the confidence intervals are narrower for cohorts with greater number of observed years (ranging from 7 to 40) and greater numbers of deaths since variance is reduced by having more number of observations.

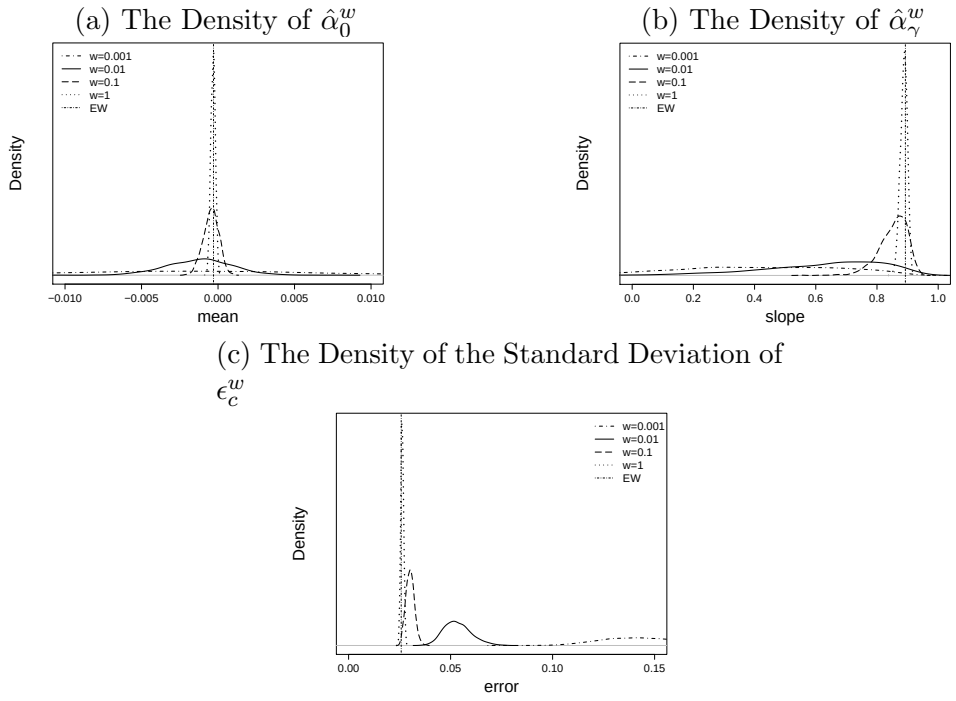


Figure 3.9: The comparison of the distribution of the parameter estimates of the $AR(1)$ model between the constructed populations $w = 1$ (dotted line), $w = 0.1$ (long dashed), $w = 0.01$ (solid), $w = 0.001$ (dot dashed) and England and Wales (vertical line)

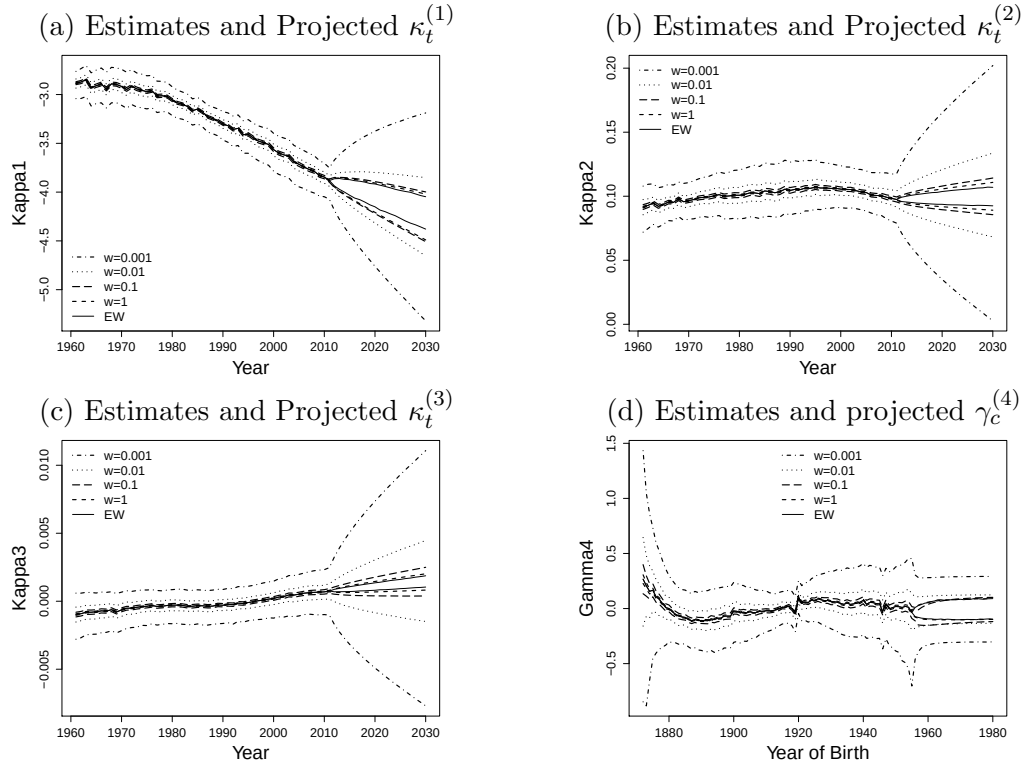


Figure 3.10: The comparison of twenty-year forward projection of $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}^{(4)}$, of weight $w = 1, 0.1, 0.01, 0.001$ with England and Wales. Note: We forecast the cohort effect from the last sixth cohort instead of the very last one due to the cohort removal. The upper bound of the CI is the 95% quantile of the distribution and lower bound is the 5% quantile. Parameter uncertainty is allowed in the projection.

3.3.3 Projected Mortality Rates

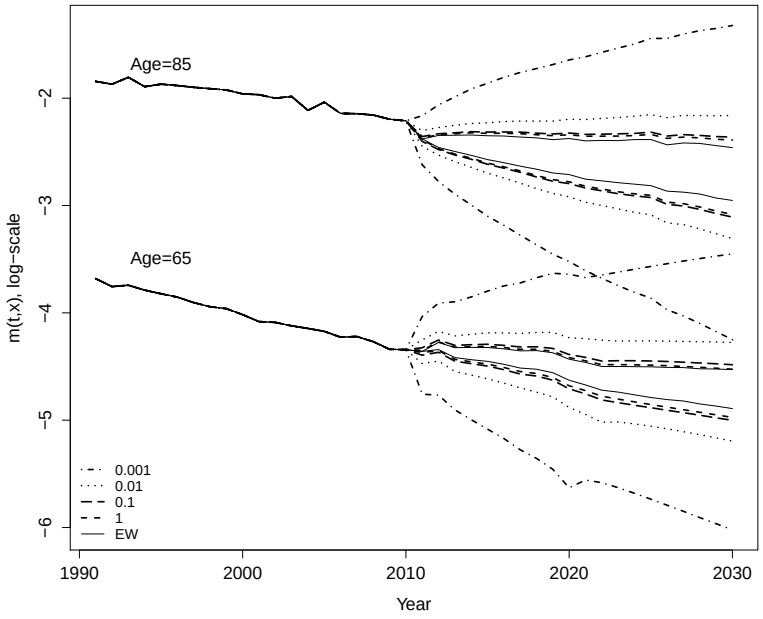
Based on the projected period and cohort effects we can now turn to the projection of mortality rates using our model in (2.3)–(2.5). Figure 3.11 shows that the twenty-year forward projections of mortality rates at age 65 and 85. We compare the predicted rates with and without the allowance for parameter uncertainty for all the constructed populations with the projections based on the England and Wales data. Unsurprisingly, the uncertainty about future mortality rates increases as the forecast horizon increases. The other two factors which significantly influence the projection uncertainty are age and population size.

Reducing the population size results in greater uncertainty about mortality forecasts for both ages. For example, the uncertainty is much greater for the smaller populations ($w = 0.01, 0.001$) at both ages 65 and 85. This means that there is considerable uncertainty about future mortality scenarios for a relatively small pension scheme with significant implications for the risk management of such a scheme.

Comparing parts (a) and (b) of Figure 3.11 we find that the inclusion of parameter uncertainty for the drift parameter $\boldsymbol{\mu}$ adds further uncertainty about the projected mortality rates. This reflects the additional randomness from not having a sufficiently long period of observed rates. We notice that the difference of variance between including and excluding parameter uncertainty increases as time increases. Thus parameter uncertainty becomes much less important when only relatively short forecast horizons are considered. Similar results can be found in Figure 6 of Cairns et al. (2006a) which shows the log scaled variance of both, with and without parameter uncertainty, for the survival index. Our findings are also in line with results obtained by Kleinow & Richards (2016) who have found that the uncertainty about the drift of the period effect in a Lee-Carter model has little impact on the uncertainty of short term projections while it significantly affects the uncertainty of long-term projections. This supports our conclusion that the differences in the

variances are tiny when the projection horizon t is very small, and become more significant for long term projection. We notice that for age 65 the intervals are not smooth in some years due to the cohort effect.

(a) Projected mortality with allowance for Parameter Uncertainty



(b) Projected mortality with no allowance for Parameter Uncertainty

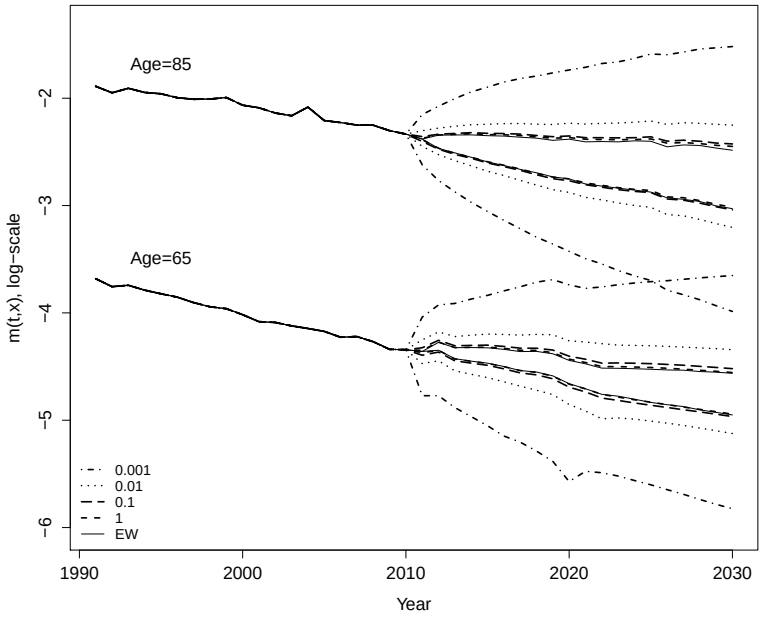
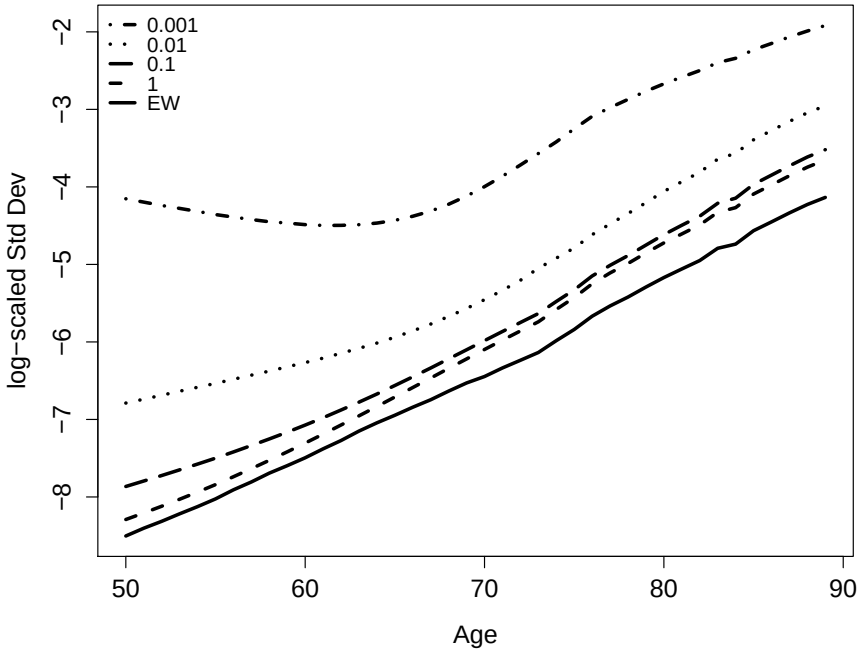


Figure 3.11: The log-scaled 90% prediction intervals of twenty-year forward mortality rate projections with (upper plot, a) and without (lower plot, b) allowance for parameter uncertainty at ages 65 and 85, for population size $w = 1$ (dashed line), 0.1 (long dashed line), 0.01 (dotted), 0.001 (dot dashed line) and England and Wales (solid line). Note that the solid line at the left end is the estimated mortality rate of the England and Wales population, with length of 20 years. The upper bound of the prediction interval is the 95% quantile of the distribution and lower bound is the 5% quantile.

(a) Log-scaled standard deviation of projected mortality with allowance for Parameter Uncertainty



(b) Log-scaled standard deviation of projected mortality with no allowance for Parameter Uncertainty

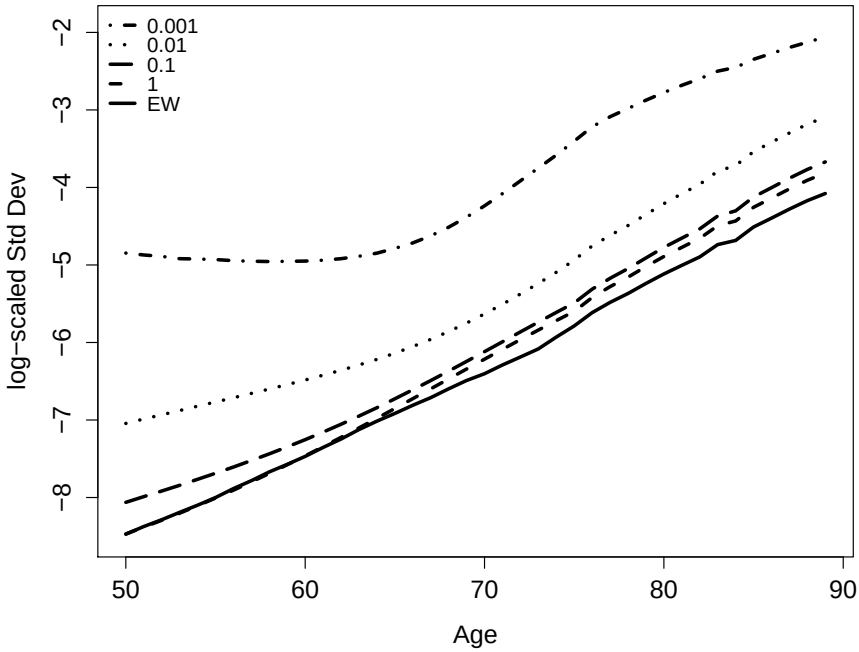


Figure 3.12: The log-scaled standard deviation of projected mortality rates with (upper plot) and without (lower plot) allowance for parameter uncertainty in year 2030 with respect to age for population size $w = 1$ (dashed line), 0.1 (long dashed line), 0.01 (dotted), 0.001 (dot dashed line) and England and Wales (solid line).

We also notice that age seems to affect the amount of uncertainty around the central projection differently in small and large populations. To illustrate this further we consider the standard deviation of projected mortality rates as a function of age for a fixed projection horizon. Figure 3.12 shows the log-scaled standard deviation of the projected mortality rates in the calendar year 2030 with respect to age. We find in this figure that the variance is an increasing function of age if the population size is rather large. In contrast, we find for the smallest population ($w = 0.001$) that the variance only starts to increase from about age 70 while it is constant or slightly decreasing for younger ages. As we found in Figure 3.11, at age 65 (and also at age 85), the three largest populations have prediction intervals which are of similar width. However, Figure 3.12 shows that the much wider prediction intervals for the two smaller populations seem to be less affected by age with the relative increase in the standard deviation from age 65 to 85 being smaller than for the large populations.

We are also interested in how much of the forecast variation is due to the impact of sampling variation and parameter uncertainty about the co-variance matrix, $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ and the drift, $\boldsymbol{\mu}$.To investigate this, we consider four experiments outlined below. Note that we still projected the cohort effect, given the point estimates for population w with the method introduced in Section 3.3.2 and we sample from the empirical distribution (generated by different death scenarios) of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$ without considering the Bayesian posterior.

1. Project mortality rates for each constructed population, while fixing the parameters $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ of the random walk to the estimates obtained from the England and Wales data.

2. Project mortality rates for each constructed population, while fixing only the drift $\boldsymbol{\mu}$ to the corresponding EW estimates and sample realisations of $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$ from its empirical distribution.

3. Project mortality rates for each constructed population, while fixing only the variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the corresponding EW estimates and sample the drift parameter from the empirical distribution of $\hat{\boldsymbol{\mu}}$.

4. Project mortality rates when both $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ and $\boldsymbol{\mu}$ are samples from the empirical distribution of $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}$ and $\hat{\boldsymbol{\mu}}$.

The results are shown in Figure 3.13. We find that fixing parameters has a significant effect on mortality forecasting when populations are very small (w=0.001) in Figure 3.13a. We can see that the widths of prediction intervals for our experiments 1 and 3 are much narrower than for experiments 2 and 4, and the difference of variance is greater for long term projections. The major difference between these two scenarios is that we fix the co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to its estimate obtained from England and Wales data in experiments 1 and 3. Thus we conclude that a major source of uncertainty for our mortality forecasts comes from the bias in the estimated covariance matrix for small populations.

(a) Projected Mortality for $w = 0.001$ (b) Projected Mortality for $w = 0.1$

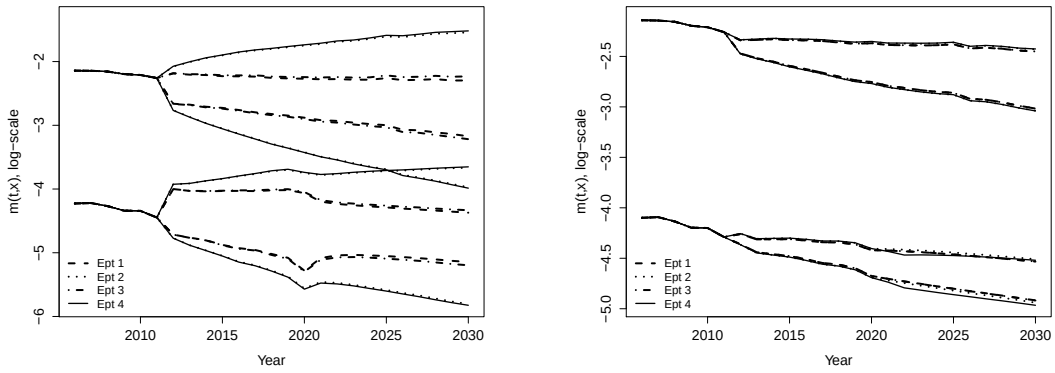


Figure 3.13: The projected mortality rates at age 65 and 85 for population sizes $w = 0.001$ (left), $w = 0.1$ (right) for the four experiments outlined in Table 3.3.3. The upper bound of the prediction interval is the 95% quantile of the distribution and the lower bound is the 5% quantile. Note that the solid line at the left end is the true mortality rate of the England and Wales population, up to year 2011.

3.4 Summary

In this Chapter, we studied the finite sample distribution of the MLEs for the parameters of the underlying stochastic mortality model. We found that the size of a population has a significant effect on the uncertainty of the estimated parameters

and the projected mortality rates. In particular, we found that estimating and projecting the parameters separately over-estimate the co-variance matrix of the random walk fitted to the period effects when the size of population is small.

As a consequence, prediction intervals are rather wide for small populations even when the parameter uncertainty is ignored. To summarise, forecasts levels of uncertainty in future mortality are biased upwards for two reasons. First, and the most obvious, the Poisson noise in the data leads to biased estimates of the random walk covariance matrix to a significant extent (Figure 3.13). Second, when we include a Bayesian analysis of parameter uncertainty, uncertainty about the drift of the random walk resulting from observations over a relatively small number of years is increased by the small population bias in the covariance matrix, $\boldsymbol{V}_{\boldsymbol{\epsilon}}$. This has its greatest impact on long term projections, and less impact in the short term.

Chapter 4

Likelihood Ratio Test for Systematic Parameter Difference

We have seen that the size of a population has a substantial impact on the level of uncertainty about the parameters of the model in (2.3)–(2.5) in Chapter 2 when this model is fitted to the population's mortality data. This raises the question whether the estimated period and cohort effects in $\theta_1 = (\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_c^{(4)})$ for a small population are significantly different from those in a given, typically much larger, reference population. To address this question we apply a likelihood ratio test to test for significant deviations of estimated parameters from a given null hypothesis using the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{1,j}^{w}$ defined in (3.3) for simulated mortality data $\boldsymbol{D}_j^w$ as in (3.2), Chapter 3. We are particularly interested in the finite sample distribution of the test statistic as compared to its asymptotic distribution. As in Section 3.2 of Chapter 3 we will use simulated deaths scenarios to investigate the finite sample distribution and the power of the likelihood ratio test (LR test) applied to mortality data. We will start with a short review of the LR test.

4.1 Review of Likelihood Ratio Test

The LR test used in this study follows the generalized form of the LR test as defined in Kendall et al. (1987). For a random variable $\boldsymbol{X}$ with a distribution that depends on a parameter vector $\boldsymbol{\varphi}$, the likelihood function is defined as usual:

$$L(\boldsymbol{x}|\boldsymbol{\varphi}) := \prod_{i=1}^{n} f_i(\boldsymbol{x}_i|\boldsymbol{\varphi}), \tag{4.1}$$

where $f_i(.|\boldsymbol{\varphi})$ is the probability density function of $\boldsymbol{X}_i$ given the parameter vector $\boldsymbol{\varphi}$. We assume that $\boldsymbol{\varphi} := (\boldsymbol{\varphi}_r, \boldsymbol{\varphi}_s)$ is a vector of $r + s$ parameters. The null hypothesis and alternative for the LR test concern only the parameters in $\boldsymbol{\varphi}_r$, that is,

$$H_0 : \boldsymbol{\varphi}_r = \boldsymbol{\varphi}_{r0};\ H_1 : \boldsymbol{\varphi}_r \neq \boldsymbol{\varphi}_{r0}. \tag{4.2}$$

In order to calculate the test statistic, we first find the MLE of $(\hat{\boldsymbol{\varphi}}_r, \hat{\boldsymbol{\varphi}}_s)$, which leads to the unconditional maximum of the likelihood function

$$\hat{\boldsymbol{\varphi}} := (\hat{\boldsymbol{\varphi}}_r, \hat{\boldsymbol{\varphi}}_s) := \arg\max_{(\boldsymbol{\varphi}_r, \boldsymbol{\varphi}_s)} L(\boldsymbol{x} \mid \boldsymbol{\varphi}_r, \boldsymbol{\varphi}_s). \tag{4.3}$$

We then find the MLE of $\boldsymbol{\varphi}_s$ assuming that the null hypothesis is fulfilled, that is,

$$\tilde{\boldsymbol{\varphi}}_s := \arg\max_{\boldsymbol{\varphi}_s} L(\boldsymbol{x} \mid \boldsymbol{\varphi}_{r0}, \boldsymbol{\varphi}_s). \tag{4.4}$$

In general $\tilde{\boldsymbol{\varphi}}_s \equiv \tilde{\boldsymbol{\varphi}}_s(\boldsymbol{\varphi}_{r0}) \neq \hat{\boldsymbol{\varphi}}_s$. We use the notation $\tilde{\boldsymbol{\varphi}}_s(\boldsymbol{\varphi}_{r0})$ to emphasise that $\tilde{\boldsymbol{\varphi}}_s$ is conditional on the value of $\boldsymbol{\varphi}_{r0}$.

We now define the test statistic in the usual way:

$$\Gamma := -2\log\frac{L(\boldsymbol{x} \mid \boldsymbol{\varphi}_{r0}, \tilde{\boldsymbol{\varphi}}_s)}{L(\boldsymbol{x} \mid \hat{\boldsymbol{\varphi}}_r, \hat{\boldsymbol{\varphi}}_s)} \tag{4.5}$$

Wilks (1938) proved that when H_0 holds, Γ asymptotically follows a central χ^2

distribution with r degrees of freedom. From the central limit theorem, it follows that the χ^2_r distribution can be approximated by a normal distribution with mean r, given r is sufficiently large.[1] Thus we expect that the distribution of Γ should approximately be symmetric around r.

Before we start testing our null hypothesis, it is worth considering the testability of the hypothesis.[2] In our approach the constraints in Equation (2.6) in Chapter 2 are part of the model and therefore the effective number of parameters that are identifiable is the total number of parameters reduced by the number of constraints. In this thesis, we formulate the constraints in terms of the cohort effect $\boldsymbol{\gamma}$ since we will in particular consider the case $\boldsymbol{\varphi}_r = \boldsymbol{\gamma}$ in our empirical study. If the test is about one of the period effects we could reformulate the constraints in terms of that period effect (strictly, therefore, a different model). In that way, the constraints are always fulfilled under H_0. In short, the constraints should be chosen such that the null hypothesis fulfils the constraints. In other words, we are testing the null hypothesis that the mortality experience is generated by mortality rates that follow model M7 with the constraints in Equation (2.6) and $\boldsymbol{\varphi}_r = \boldsymbol{\varphi}_{r_0}$.

In the remainder of this section we will consider a null hypothesis about the entire parameter vector $\boldsymbol{\varphi}$ setting $s = 0$. In Section 4.5 we will then consider a null hypothesis about the cohort effect $\boldsymbol{\gamma}$ only, that is $s > 0$.

4.2 Finite Sample Distribution of LRT

As mentioned above, we now consider a test for systematic parameter differences involving all period effects and the cohort effect, that is, $s = 0$ and $\boldsymbol{\varphi} = \boldsymbol{\varphi}_r = \boldsymbol{\theta}_1 = (\boldsymbol{\kappa}^{(1)}, \boldsymbol{\kappa}^{(2)}, \boldsymbol{\kappa}^{(3)}, \boldsymbol{\gamma}^{(4)})$. The null hypothesis and alternative are given in (4.2), and the

[1] See Kendall et al. (1987) for more details about the likelihood ratio test and the asymptotic distribution of the LRT statistic

[2] See Searle (1971) for more details about testable hypotheses.

LR test statistic is defined in (4.5) which simplifies to

$$\Gamma = -2\log\frac{L(\boldsymbol{x} \mid \boldsymbol{\varphi}_{r0})}{L(\boldsymbol{x} \mid \hat{\boldsymbol{\varphi}}_r)} \tag{4.6}$$

since $s = 0$.

As in Section 3.2 of Chapter 3, we choose the male population in England and Wales as our base case and set $\boldsymbol{\theta}_{1,0} = \hat{\boldsymbol{\theta}}_1^{\mathrm{EW}}$. Therefore, the relative null hypothesis and alternative for the LRT can be written as:

$$H_0 : \boldsymbol{\varphi}_r = \boldsymbol{\theta}_{1,0};\ H_1 : \boldsymbol{\varphi}_r \neq \boldsymbol{\theta}_{1,0}.$$

To investigate the finite sample properties of the LR test in small populations we apply a parametric bootstrap procedure in which we simulate N_1 mortality scenarios, estimate the parameter vector $\boldsymbol{\theta}_1$ as in Section 3.2 and apply the LR test in each scenario. More specifically we use the following steps to find a bootstrap approximation of the finite sample distribution of Γ: For different values of w and for each scenario $j = 1, \ldots, N_1$ we

1. simulate $\boldsymbol{D}_j^w$ as in (3.2) on page 29,
2. find the estimate $\hat{\boldsymbol{\theta}}_{1,j}^w$ as in (3.3),
3. calculate the realisation of the LR test statistic Γ_j^w as in (4.6) with $\boldsymbol{\varphi}_{r0} = \boldsymbol{\theta}_{1,0}$ and $\hat{\boldsymbol{\varphi}}_r = \hat{\boldsymbol{\theta}}_{1,j}^w$, and
4. calculate the p-value P_j^w based on the asymptotic χ^2-distribution as $P_j^w = \mathbf{P}[X > \Gamma_j^w]$ where X is has χ^2-distribution with α degrees of freedom.

The degrees of freedom of the χ^2-distribution in step 4 should be the effective number of parameters denoted by α, which is the total number of parameters r less

the number of constraints, that is

$$\alpha = 3n_y + n_c - 3$$

where n_y is the number of years, and $n_c = n_y + n_a - 1$ is the total number of cohorts in a given dataset without removing short cohorts. In our empirical study, $n_y = 51$, $n_a = 40$, $n_c = 51 + 40 - 1 = 90$, hence $\alpha = 240$. After applying the parametric bootstrap method we can study the distribution of the test statistic.

For any population size w we find the empirical distribution function of Γ^w based on the sample $\Gamma_1^w, \ldots, \Gamma_{N_1}^w$. We expect that the distribution of Γ^w should be approximately symmetric around 240. Furthermore, if the asymptotic χ^2 approximation is accurate, the p-values $P_1^w, \ldots, P_{N_1}^w$ should be independent and uniformly distributed on $[0, 1]$. The cumulative distribution of the test statistic Γ^w and the p-values P^w for all considered population sizes w are shown in Figure 4.1 for $N_1 = 1000$. Figure 4.1a shows that the empirical distribution of Γ^w is indeed centred around $\alpha = 240$. We also observe in Figure 4.1b that the cumulative distribution function of the p-values resembles the distribution function of the uniform distribution on $[0, 1]$. Both results indicate that the χ^2 approximation for the distribution of Γ^w under the null hypothesis is very good for all values of w considered.

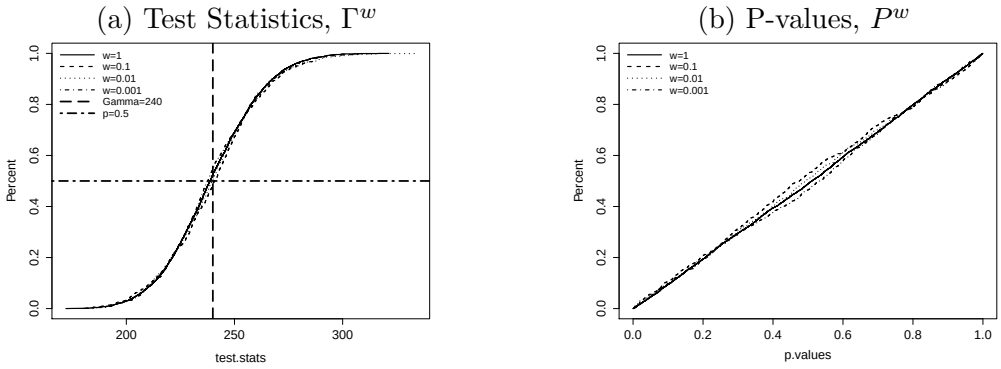


Figure 4.1: Likelihood ratio test: (a) empirical CDF's of test statistics for sample size $N_1 = 1000$. (b) empirical CDF's of asymptotic p-values. Results shown for populations of $w = 1$ (solid line), $w = 0.1$ (dashed line), $w = 0.01$ (dotted line) and $w = 0.001$ (dot dashed line). The mean of the asymptotic χ^2_{240}- distribution is also shown as the vertical dashed line in plot (a).

4.3 Power of the Likelihood Ratio Test

In the last section, we carried out the likelihood ratio test for the parameter difference and found that the χ^2 approximation does not fail to capture the feature of the test statistic Γ^w when H_0 holds. We will now investigate how the population size affects the power of LR test.

We start with a short review of the power of the hypothesis test. In general, the power of a binary hypothesis is the probability of correctly accepting the alternative hypothesis when it is true[3]. We follow the definition in Kendall et al. (1987). It is well known that there are two types of errors (Type I and II) we may made in a statistical hypothesis test:

Type I error: Reject H_0 when it is true;

Type II error: Do not reject H_0 when H_1 is true.

The probability of Type II error is denoted as β and $1-\beta$ is defined as the power of the test of H_0 against H_1, that is:

$$\text{Power of test} = \text{Prob (Reject } H_0 \mid H_1 \text{ is True).}$$

To evaluate the power of the LRT with a parametric bootstrap procedure similar to the one used in the previous section we need to generate scenarios under the alternative. So far we have considered a very general alternative $\boldsymbol{\varphi}_r \neq \boldsymbol{\varphi}_{r0}$. We will now need to specify this alternative further. To this end we define four alternative models and investigate the power assuming that the "true" data generating model is one of those alternatives. The four models we consider for the alternative shift or scale one of the period effects or the cohort effect estimated from the England and Wales data.

[3]See, for example, Ellis (2010) for more details on statistical power.

More specifically, the alternatives we consider are:

- $\boldsymbol{\theta}_1^{(1)}(\lambda) = (\boldsymbol{\kappa}_0^{(1)} + \lambda, \boldsymbol{\kappa}_0^{(2)}, \boldsymbol{\kappa}_0^{(3)}, \boldsymbol{\gamma}_0^{(4)})$
- $\boldsymbol{\theta}_1^{(2)}(\lambda) = (\boldsymbol{\kappa}_0^{(1)}, \boldsymbol{\kappa}_0^{(2)} + \lambda, \boldsymbol{\kappa}_0^{(3)}, \boldsymbol{\gamma}_0^{(4)})$
- $\boldsymbol{\theta}_1^{(3)}(\lambda) = (\boldsymbol{\kappa}_0^{(1)}, \boldsymbol{\kappa}_0^{(2)}, \boldsymbol{\kappa}_0^{(3)} + \lambda, \boldsymbol{\gamma}_0^{(4)})$
- $\boldsymbol{\theta}_1^{(4)}(\lambda) = (\boldsymbol{\kappa}_0^{(1)}, \boldsymbol{\kappa}_0^{(2)}, \boldsymbol{\kappa}_0^{(3)}, \lambda\boldsymbol{\gamma}_0^{(4)})$

Further we define $\theta_1^{(i)}(\lambda) = \theta_1^{(i)}(\lambda, t, x)$ at time t age x as:

$$\theta_1^{(i)}(\lambda) = (\kappa_{t,0}^{(1)} + \lambda\mathbb{1}_{i=1}, \kappa_{t,0}^{(2)} + \lambda\mathbb{1}_{i=2}, \kappa_{t,0}^{(3)} + \lambda\mathbb{1}_{i=3}, (\lambda\mathbb{1}_{i=4} + 1)\gamma_{c,0}^{(4)})$$

where

$$\mathbb{1}_{i=k} = \begin{cases} 1 & i = k \\ 0 & \text{otherwise,} \end{cases} \quad \text{for } k = 1, 2, 3, 4.$$

Note that for $i = 1, 2, 3$, $\kappa_{t,0}^{(i)}$ and $\gamma_{c,0}^{(4)}$ are the t^{th} and c^{th} elements of $\boldsymbol{\kappa}_0^{(i)}$ and $\boldsymbol{\gamma}_0^{(4)}$ respectively.

We then evaluate the power of the LR test against each of those alternatives with different values of λ. Note that we scaled the cohort effect by λ units instead of shifting it since shifting the cohort effect would result in the same fitted mortality rates as shifting $\boldsymbol{\kappa}_0^{(1)}$ in $\boldsymbol{\theta}_1^{(1)}(\lambda)$. We note that a more general alternative could be considered by allowing for combinations of the above. However, we wish to focus on the impact of misspecifying individual parameters and the power of the test to detect those misspecification.

We can now proceed as in the previous section with simulating death counts and then apply the LR test for different alternatives and different values of λ. We define matrix $\boldsymbol{D}_j^{w,(i)} = \{D_j^{w,(i)}(t, x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$ to be the simulated deaths in scenario $j = 1, \ldots, N_1$ for the population of size $w\boldsymbol{E}_0$ using the parameter $\boldsymbol{\theta}_1^{(i)}(\lambda)$ in our

model, that is,

$$D_j^{w,(i)}(t,x) \sim \text{Pois}(m(\theta_1^{(i)}(\lambda))wE_0(t,x)) \tag{4.7}$$

for $i = 1, 2, 3, 4$, where m is defined in (2.4) and (2.5). Note that death counts also depend on λ.

Using the simulated death counts $\boldsymbol{D}_j^{w,(i)}$ we obtain the MLE $\hat{\boldsymbol{\theta}}_{1,j}^{w,(i)}$ as in (3.3). We the use the asymptotic χ^2-distribution to test the null hypothesis that the parameters of our model are equal to the parameters obtained from the England and Wales population. The p-values $P_j^{w,(i)} = P_j^{w,(i)}(\lambda)$ are then calculated as in step 4 in the previous section, and the null hypothesis is rejected in any scenario j for which $P_j^{w,(i)} < 0.05$, that is, the significance level of the test is 0.05.

The power of the LRT for any fixed alternative i, relative population size w and fixed λ is the proportion of the simulated p-values which are less than 0.05, that is, we count the number of scenarios for which the null hypothesis is rejected. More specifically, we define the random variables

$$\begin{aligned} R_j^{w,(i)}(\lambda) &= \begin{cases} 1 & \text{if } P_j^{w,(i)}(\lambda) < 0.05 \ (H_0 \text{ rejected}) \\ 0 & \text{otherwise} \end{cases} \\ R^{w,(i)}(\lambda) &= \frac{1}{N_1} \sum_{j=1}^{N_1} R_j^{w,(i)}(\lambda) \end{aligned} \tag{4.8}$$

so that $R^{w,(i)}(\lambda)$ is the proportion of scenarios in which the null hypothesis is rejected among N_1 simulated scenarios. We call $R^{w,(i)}(\lambda)$ the empirical rejection rate. Since we are considering independent scenarios, $N_1 R^{w,(i)}(\lambda)$ has a Binomial distribution,

$$N_1 R^{w,(i)}(\lambda) \sim \text{Bin}\Big(N_1, p^{w,(i)}(\lambda)\Big) \tag{4.9}$$

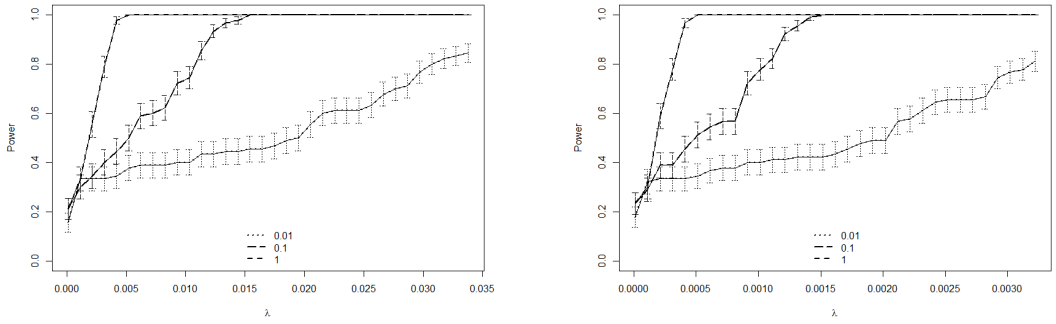
where $p^{w,(i)}(\lambda)$ is the (unknown) power of the LR test if alternative $\boldsymbol{\theta}_1^{(i)}(\lambda)$ with parameter λ is the true parameter set for the simulated death counts. Therefore, the empirical rejection rate $R^{w,(i)}(\lambda)$ is an unbiased estimator for the power $p^{w,(i)}(\lambda)$

and the estimated standard deviation of $R^{w,(i)}(\lambda)$ can easily be found from (4.9) in the usual way.

Then we investigate the sensitivity of the power with respect to the size of λ and the size of the population w. For each of the four cases, $\boldsymbol{\theta}_1^{(1)}(\lambda), \ldots, \boldsymbol{\theta}_1^{(4)}(\lambda)$, we consider a set of values for λ that are regularly spaced.

Figure 4.2 shows the obtained estimates $R_j^{w,(i)}(\lambda)$ for the power as a function of λ for different relative population sizes w. Note that for each alternative $\boldsymbol{\theta}_1^{(i)}(\lambda)$ and any fixed λ we have simulated $N_1 = 100$ scenarios, which is less than in the previous section. The reason is that we need to simulate those scenarios for each combination of i (alternative) and λ, which makes the total number of simulated scenarios very large.

(a) The Power of LR test when $\kappa^{(1)}$ shifted (b) The Power of LR test when $\kappa^{(2)}$ shifted



(c) The Power of LR test when $\kappa^{(3)}$ shifted (d) The Power of LR test when $\gamma^{(4)}$ scaled

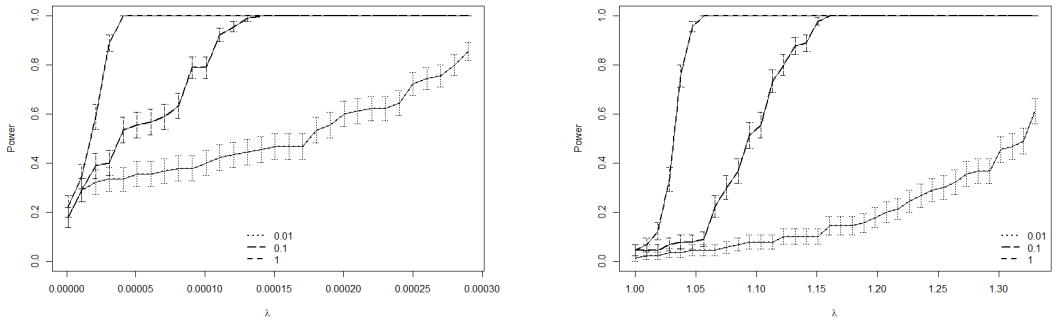


Figure 4.2: The empirical rejection rates $R^{w,(i)}(\lambda)$ under the LRT together with error bars for relative population sizes $w = 1$ (dashed line), $w = 0.1$ (long dashed line) and $w = 0.01$ (dotted line). The width of the error bars is one standard error based on (4.9).

Unsurprisingly, the power of the LRT is increasing in λ for any $\boldsymbol{\theta}_1^{(i)}(\lambda)$ and relative population size w; the more we shift/scale the null hypothesis, the easier it

is for the test to detect any shift/scaling. For the three period effects, decreasing the population size will greatly reduce the capability of LR test to detect the same amount of shift to a single parameter. We can also compare these plots with the earlier Figure **??** (page **??**) which includes distributions of parameter estimates resulting from sampling variation. By way of example, for $w = 0.01$ the width of the confidence interval in Figure **??** (e) for $\kappa_t^{(3)}$ is about 0.005. This is much larger than the shifts that are considered in the power plot in Figure 4.2. The reason why the latter values are so much lower is because we apply a systematic adjustment to all of the $\kappa_t^{(3)}$, in contrast to random adjustments (due to sampling variation) in the former.

4.4 Impact of Parameter Misspecification on Mortality Rates and Annuities

We now investigate how significant the impact of shifting and scaling parameters is on the fitted mortality rates and corresponding annuity prices. We consider again the four alternatives in the previous section. For each of those and for each relative population size w we determine the value of λ that results in a power of 50% of the LR test, that is, there is a 50% probability that the LR test will detect the wrong model and reject the null hypothesis. Those values are denoted by $\lambda_{0.5}^{w,(i)}$ and shown in Table 4.1.

Parameter shifted	$w = 1$	$w = 0.1$	$w = 0.01$
$\lambda_{0.5}^{w,(1)}$	0.003	0.006	0.02
$\lambda_{0.5}^{w,(2)}$	0.0003	0.0006	0.002
$\lambda_{0.5}^{w,(3)}$	0.0000025	0.000005	0.00018
$\lambda_{0.5}^{w,(4)}$	1.03	1.09	1.32

Table 4.1: The table contains the size of shift required for 50% power when each parameter is shifted separately, with respect to population $w = 1, 0.1, 0.01$

We then calculate fitted mortality rates using the model in (2.4) and (2.5) (see page 23) with the following parameter constellations:

- $\boldsymbol{\theta}_1^{w,(1)}(\lambda_{0.5}^{w,(1)}) = (\boldsymbol{\kappa}_0^{(1)} + \lambda_{0.5}^{w,(1)}, \boldsymbol{\kappa}_0^{(2)}, \boldsymbol{\kappa}_0^{(3)}, \boldsymbol{\gamma}_0^{(4)})$
- $\boldsymbol{\theta}_1^{w,(2)}(\lambda_{0.5}^{w,(2)}) = (\boldsymbol{\kappa}_0^{(1)}, \boldsymbol{\kappa}_0^{(2)} + \lambda_{0.5}^{w,(2)}, \boldsymbol{\kappa}_0^{(3)}, \boldsymbol{\gamma}_0^{(4)})$
- $\boldsymbol{\theta}_1^{w,(3)}(\lambda_{0.5}^{w,(3)}) = (\boldsymbol{\kappa}_0^{(1)}, \boldsymbol{\kappa}_0^{(2)}, \boldsymbol{\kappa}_0^{(3)} + \lambda_{0.5}^{w,(3)}, \boldsymbol{\gamma}_0^{(4)})$
- $\boldsymbol{\theta}_1^{w,(4)}(\lambda_{0.5}^{w,(4)}) = (\boldsymbol{\kappa}_0^{(1)}, \boldsymbol{\kappa}_0^{(2)}, \boldsymbol{\kappa}_0^{(3)}, \lambda_{0.5}^{w,(4)}\boldsymbol{\gamma}_0^{(4)})$

and similarly we define $\theta_1^{w,(i)}(\lambda_{0.5}^{w,(i)})(t,x)$ at time t age x to be:

$$\theta_1^{w,(i)}(\lambda_{0.5}^{w,(i)}) = (\kappa_{t,0}^{(1)} + \lambda_{0.5}^{w,(i)}\mathbb{1}_{i=1}, \kappa_{t,0}^{(2)} + \lambda_{0.5}^{w,(i)}\mathbb{1}_{i=2}, \kappa_{t,0}^{(3)} + \lambda_{0.5}^{w,(i)}\mathbb{1}_{i=3}, (\lambda_{0.5}^{w,(i)}\mathbb{1}_{i=4} + 1)\gamma_{c,0}^{(4)})$$

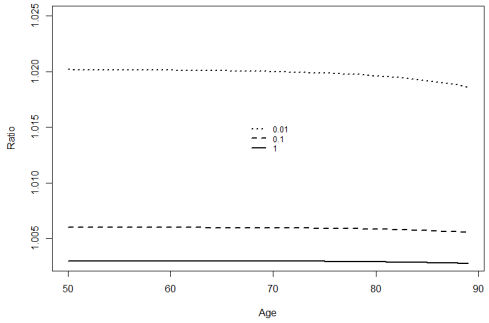
for $i = 1, 2, 3, 4$.

To quantify the change in fitted mortality rates we calculate the following ratio

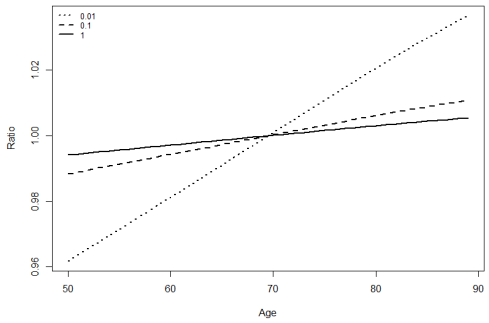
$$\rho_{t,x}^{w,(i)} = \frac{m(\theta_1^{w,(i)}(\lambda_{0.5}^{w,(i)}), t, x)}{m(\theta_{1,0}, t, x)}$$

for each $i = 1, 2, 3, 4$ and different values of w. We expect that shifting $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$ will result in a parallel shift upwards, tilting rates in an anti clockwise direction and add some concavity to the rates respectively. This can indeed be seen in Figure 4.3 where we plot the ratio $\rho_{t,x}^{w,(i)}$ for the year $t = 2011$. Figure 4.3d suggests that scaling $\gamma_c^{(4)}$ tilts and introduces more fluctuation to the ratio. For all the four parameters, reducing the relative population size w increased the relative change $\rho_{t,x}^{w,(i)}$ since $\lambda_{0.5}^{w,(i)}$ increases. This confirms the intuitive idea that even misspecified parameters which produce significant changes in the fitted mortality rates are hard to detect with an LR test when the exposures are small.

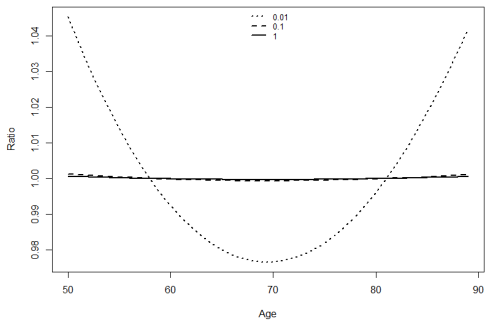
(a) The impact of $\kappa_t^{(1)}$ shifted at year 2011



(b) The impact of $\kappa_t^{(2)}$ shifted at year 2011



(c) The impact of $\kappa_t^{(3)}$ shifted at year 2011



(d) The impact of $\gamma_c^{(4)}$ scaled at year 2011

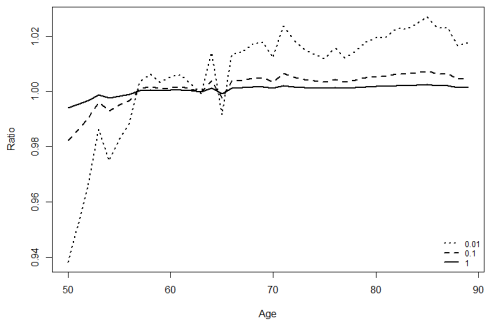


Figure 4.3: The impact of shifting each parameter separately on the estimated death rate of England and Wales. The shift is determined when it results in 50% power for each population $w = 1$ (solid line), $w = 0.1$ (dashed line) and $w = 0.01$ (dotted line).

From a financial point of view the effect on fitted mortality rates is only relevant in so far as annuity prices are affected. We will therefore consider the following annuities and discuss the effect of the four alternatives specified above on their values:

- A temporary annuity of £1 per annum payable annually in arrears to a life now aged 65 exactly, starting at the beginning of year 2012 with term of 25 years. Its expected present value is calculated as:

$$a_{65:\overline{25}|} = \sum_{j=1}^{25} S(T+j, 65)v^j$$

- An annuity of £1 per annum payable annually in arrears to a life now aged 55 exactly, deferred for 10 years, starting at the beginning of year 2012 with term of 25 years. Its expected present value is:

$$_{10|}a_{55:\overline{25}|} = \sum_{j=11}^{35} S(T+j, 55)v^j$$

where v is the discount factor, $S(T+t, x)$ is the survival index for the probability of an individual aged x exactly at the start of year T, that will survive for the next t years. We assume the interest rate of $i = 2\%$. The reason for investigating the deferred annuity is that Figure 3.5a (see page 37) suggests that the estimates of cohort effect at $c = 1946$ is approximately zero and the effect of scaling cohort estimates may not be obvious on the annuity price $a_{65:\overline{25}|}$ but more obvious for $_{10|}a_{55:\overline{25}|}$.

We project the period and cohort effects in $\boldsymbol{\theta}_1^{w,(i)}(\lambda_{0.5}^{w,(i)})$ $(i = 1, 2, 3, 4)$ and $\hat{\boldsymbol{\theta}}_1^{\text{EW}}$ forward for 35 years as in Section 3.3 of Chapter 3 where we use the point estimates defined in (3.8) and (3.9) (see page 41) for the parameters of the random walk for the shifted period effects, that is, we do not consider uncertainty about the drift and variance matrix of the random walk. Annuity prices are calculated for each

sample path and we then calculate the average annuity price for each w with the i^{th} parameter shifted or scaled. The results are shown in Table 4.2 and Table 4.3.

Parameter shifted	England and Wales	$w = 1$	$w = 0.1$	$w = 0.01$
$\kappa^{(1)}$	14.76756	14.75694	14.74631	14.69650
$\kappa^{(2)}$	14.76756	14.76181	14.75605	14.72905
$\kappa^{(3)}$	14.76756	14.76811	14.76866	14.80665
$\gamma^{(4)}$	14.76756	14.76291	14.75362	14.71787

Table 4.2: The impact of shifting each parameter separately on the price of a twenty five-year temporary annuity for an individual aged at 65. The shift is determined when it results in 50% power for each population $w = 1, 0.1, 0.01$, which are shown in Table 4.1. We assume an interest rate of 2%.

Parameter shifted	England and Wales	$w = 1$	$w = 0.1$	$w = 0.01$
$\kappa^{(1)}$	12.04525	12.03591	12.02656	11.98271
$\kappa^{(2)}$	12.04525	12.04338	12.04150	12.03251
$\kappa^{(3)}$	12.04525	12.04563	12.04602	12.07256
$\gamma^{(4)}$	12.04525	12.04779	12.05288	12.07226

Table 4.3: The impact of shifting each parameter separately on the price of a ten-year deferred twenty five-year temporary annuity for an individual aged at 55. The shift is determined when it results in 50% power for each population $w = 1, 0.1, 0.01$, which are shown in Table 4.1. We assume an interest rate of 2%.

The effects of shifting the period effects and scaling the cohort effect are somewhat varied. As might be expected, the impact on prices is most obvious for $w = 0.01$. The impact on both types of annuity is straightforward to see. For $\boldsymbol{\kappa}^{(1)}$: the shift pushes up mortality rates at all ages and lowers prices. For $\boldsymbol{\kappa}^{(2)}$ there is more impact on the age-65 annuity than the age-55 deferred annuity as the shift lowers mortality at younger ages and raises it at higher ages. For $\boldsymbol{\kappa}^{(3)}$, also, the impact is different at different ages. Finally, for $\boldsymbol{\gamma}^{(4)}$, the impact of scaling simply depends on the sign and magnitude of the value of $\boldsymbol{\gamma}^{(4)}$ for the cohort being priced.

Generally shifting or scaling the parameter estimates has no obvious effect on the annuity price even for $w = 0.01$. Thus the financial consequence of not rejecting a false null hypothesis are relatively small for the considered alternatives. In other words, the fact that we have accepted H_0 means that $\boldsymbol{\varphi}_r$ for $w = 0.01$, while not

identical, must be very close to $\hat{\boldsymbol{\theta}}_1^{\mathrm{EW}}$, and that, therefore, any error in pricing will also be very small.

4.5 Likelihood Ratio Test for the Cohort Effect

The general form of the LR test as reviewed in Section 4.1 allows us to test a null hypothesis about parts of the parameter vector $\boldsymbol{\theta}_1$ (restricted by the specified identifiability constraints as part of the model) rather than the entire $\boldsymbol{\theta}_1 = (\boldsymbol{\kappa}^{(1)}, \boldsymbol{\kappa}^{(2)}, \boldsymbol{\kappa}^{(3)}, \boldsymbol{\gamma}^{(4)})$. Testing parts of $\boldsymbol{\theta}_1$ is particularly relevant if mortality rates in a rather small population are modelled using estimated period or cohort effects from a larger population. Setting one or more of the components of $\boldsymbol{\theta}_1$ equal to the function of corresponding parameters estimated from the large population reduces the dimension of the parameter vector which needs to be estimated from the small population where parameter uncertainty is rather strong as we have seen in Section 3.2. The example we have in mind is a pension fund that uses national mortality data to improve its mortality models, or when the mortality experience in a small country is modelled based on the combined experience of other similar countries.

In the reminder of this section we will use the LR test to test a null hypothesis about the cohort effect $\boldsymbol{\gamma}^{(4)}$. In our general setting of Section 4.1 this means that

$$\boldsymbol{\varphi}_r = \boldsymbol{\gamma}^{(4)} \text{ and } \boldsymbol{\varphi}_s = (\boldsymbol{\kappa}^{(1)}, \boldsymbol{\kappa}^{(2)}, \boldsymbol{\kappa}^{(3)}).$$

Our null hypothesis is then that $\boldsymbol{\gamma}^{(4)} = \boldsymbol{\gamma}_0^{(4)}$ where $\boldsymbol{\gamma}_0^{(4)}$ is a given vector of cohort effects, for which we later use an estimated cohort effect from a different population. We can now write the hypotheses as in (4.2) and proceed as in Section 4.2 to find the distribution of the LR test statistic in (4.5) for a finite sample of death counts from a small population.

For practical relevance we base our simulation study on the female and male populations in England and Wales. We choose $\boldsymbol{\gamma}_0^{(4)} = \hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$, which is the estimated cohort effect from the mortality data for males in England and Wales. It is worth noting that, as $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$ already satisfies the identifiability constraints, the null

hypothesis $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0^{(4)}$ has no testability problems under the given identifiability constraints defined in the model system. To investigate finite sample properties of Γ we will need to specify a full parameter vector $\boldsymbol{\theta}_1$ to simulate scenarios for the death counts. Having fixed the cohort effect $\boldsymbol{\gamma}_0^{(4)}$ we choose the period effects to be the estimated period effects from data for the female population in England and Wales assuming that the cohort effect for those data is actually $\boldsymbol{\gamma}_0^{(4)}$. As we are mainly interested in small populations we will consider deaths count scenarios for populations which have exposures equal to $w\boldsymbol{E}_0$ where $\boldsymbol{E}_0$ is here the exposure for the female population in England and Wales.

More specifically, we first find the MLE $\tilde{\boldsymbol{\varphi}}_s = \arg\max_{\boldsymbol{\varphi}_s} L(\boldsymbol{x} \mid \boldsymbol{\varphi}_{r0} = \boldsymbol{\gamma}_0^{(4)}, \boldsymbol{\varphi}_s)$ of the period effect $\boldsymbol{\varphi}_s = (\boldsymbol{\kappa}^{(1)}, \boldsymbol{\kappa}^{(2)}, \boldsymbol{\kappa}^{(3)})$ from the data for females assuming that the cohort effect is indeed $\boldsymbol{\gamma}_0^{(4)}$ (which is the estimated cohort effect for males), see (4.4). Note that no constraints are applied for finding $\tilde{\boldsymbol{\varphi}}_s$ since the cohort is fixed and therefore there is no identifiability problem. We then generate N_1 realizations of the value of the test statistic Γ^w for different values of the relative population size w using the following algorithm:

1. Simulate death counts $\boldsymbol{D}_j^w$ as in (3.2) using the parameter vector

$$\tilde{\boldsymbol{\varphi}} = (\tilde{\boldsymbol{\varphi}}_s, \boldsymbol{\varphi}_{r0}) = (\tilde{\boldsymbol{\kappa}}^{(1)}, \tilde{\boldsymbol{\kappa}}^{(2)}, \tilde{\boldsymbol{\kappa}}^{(3)}, \boldsymbol{\gamma}_0^{(4)}),$$

   where $\tilde{\boldsymbol{\varphi}}_s = (\tilde{\boldsymbol{\kappa}}^{(1)}, \tilde{\boldsymbol{\kappa}}^{(2)}, \tilde{\boldsymbol{\kappa}}^{(3)})$ and $\tilde{\boldsymbol{\kappa}}^{(i)} = (\tilde{\boldsymbol{\kappa}}_{t_1}^{(i)}, \ldots, \tilde{\boldsymbol{\kappa}}_{t_{ny}}^{(i)})$ for $i = 1, 2, 3$, to obtain scenario $\boldsymbol{D}_j^w$ for different values of the relative population size w. The period effects $\tilde{\boldsymbol{\kappa}}^{(i)}$ for $i = 1, 2, 3$ are estimated from data for females with the cohort effect fixed to $\boldsymbol{\gamma}_0^{(4)}$. The exposure is $w\boldsymbol{E}_0$ where $\boldsymbol{E}_0$ is the exposure for the female population in England and Wales.

2. Find the MLE $\tilde{\boldsymbol{\varphi}}_{s,j}^w$ of period effects $\boldsymbol{\kappa}$ in scenario j for relative population size w assuming that the null hypothesis holds, as in (4.4).

3. Find the unrestricted MLE $\hat{\boldsymbol{\varphi}}_j^w$ in scenario j for relative population size w, as

in (4.3).

4. Calculate the value of the test statistic Γ_j^w in (4.5) in each scenario j.

5. Calculate the p-values P_j^w based on the asymptotic χ^2-distribution with α degrees of freedom, where α is the number of parameters (cohorts) r minus the number of constraints as in Section 4.2. For our data set we obtain $\alpha = 87$.

The reason for having $\alpha = 87$ number of effective parameters is that the likelihood for the alternative hypothesis, i.e. the unconditional likelihood without fixing parameters to any values, will not be changed by either the choice of identifiability constraints or whether or not to apply the constraints. Thus the test statistic for the LR test will not be changed without applying the constraints either. However, in this model (M7), three of the cohorts can be written as a linear combination of the rest of the cohorts. Such feature of the model itself indicates that the number of effective parameters should be the total number of parameters minus three.

The simulated distribution functions of the LR test statistic Γ^w and the p-values P_j^w are shown in Figure 4.4. The results suggest that changing the size of the population has no significant impact on the distribution of Γ^w and that the p-values are roughly uniformly distributed for all w, which is an indication that the χ^2-approximation works well for our data set as we have also found in Section 4.2 where the full parameter vector was tested.

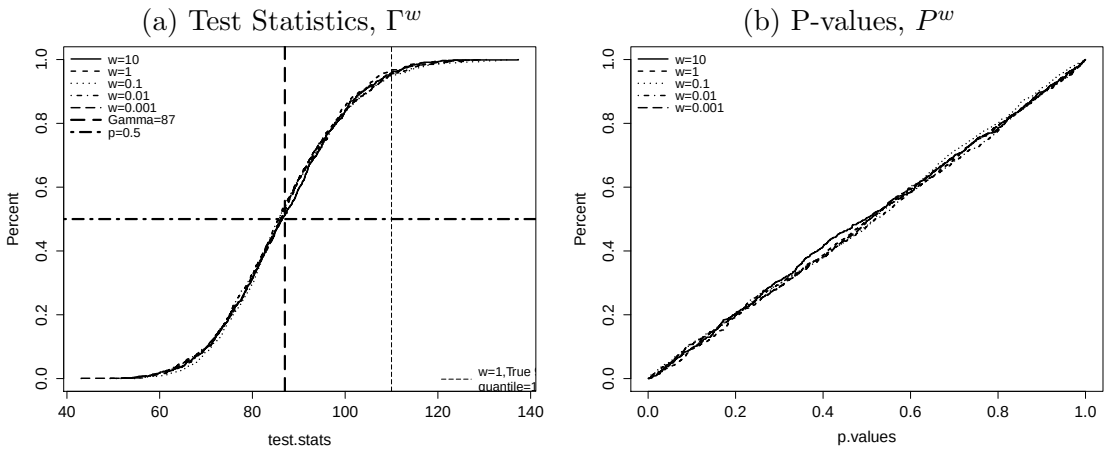


Figure 4.4: The results of likelihood ratio test, with distributions of test statistics (a) and p-values (b), for the population of $w = 10$ (solid line), $w = 1$ (dashed line), $w = 0.1$ (dotted line), $w = 0.01$ (dashed dotted line) and $w = 0.001$. The left vertical dashed line is the mean of normal approximation for the χ^2_{87}, at $x = 87$. The right dashed line at $x = 110$ is the true 95% quantile of population $w = 1$.

4.6 Empirical Examples

We apply the LR test for the cohort effect in two empirical studies.

4.6.1 Females vs. Males in England and Wales

The population for which we wish to test the cohort effect first is the female population in England and Wales that we already considered in our simulation study. Our null hypothesis is therefore that the true cohort effect for the female population in England and Wales is equal to the estimated cohort effect for males in England and Wales. Note that this is different from testing the hypothesis that the male and female population share the same (true) cohort effect since we ignore the uncertainty about the estimated cohort effect for males.

To illustrate the difference between the two cohort effects we plot in Figure 4.5 the estimated cohort effects for females and males. There are fairly strong similarities between the two curves after about 1910, but there are also significant qualitative differences before 1900. To check empirically, that these differences are not simply the result of the identifiability constraints, one can plot $\hat{\boldsymbol{\gamma}}^{(4),M} - \boldsymbol{\gamma}^{(4),F}$. If this looks quadratic then the differences could, simply, be due to the identifiability constraints. But for these data, a plot of $\hat{\boldsymbol{\gamma}}^{(4),M} - \boldsymbol{\gamma}^{(4),F}$ would clearly not be quadratic (exhibiting more of a cubic shape).

This difference can be confirmed more formally using the LR test with the null hypothesis that the females have the same cohort effect as the previously estimated males cohort effect. The test statistic Γ is approximately 6311, which is an extremely high value for a χ^2-distribution with 87 degrees of freedom and is also very high compared to the values of Γ observed in our simulation study, see Figure 4.4. The p-value is therefore very close to zero, and we reject the null hypothesis that the cohort effect fro the mortality of the female population is the same as the previously

Figure 4.5: The estimates of cohort effect, for England and Wales males (solid line) and females (dashed line), age 50 to 89 last birthday, over year 1961 to 2011

estimated cohort effect for the male population.

4.6.2 Male Mortality in Scotland vs. England and Wales

A second, and more intriguing, empirical example concerns the cohort effects estimated from mortality data for the male population in England and Wales versus the male population in Scotland. Figure 4.6 compares the independently-estimated cohort effects with a confidence interval added around the Scottish estimates. Compared to Figure 4.5, the two curves here look much more similar, with the pattern of $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}} - \hat{\boldsymbol{\gamma}}^{(4),\text{SL}}$ again not like a quadratic function with respect to cohort year c. On the other hand, we find that most of the cohort effects for males in England and Wales lie outside of the confidence interval calculated for Scottish males. This suggests that although the two populations have similar pattern for the cohort estimates, the difference might still be significant.

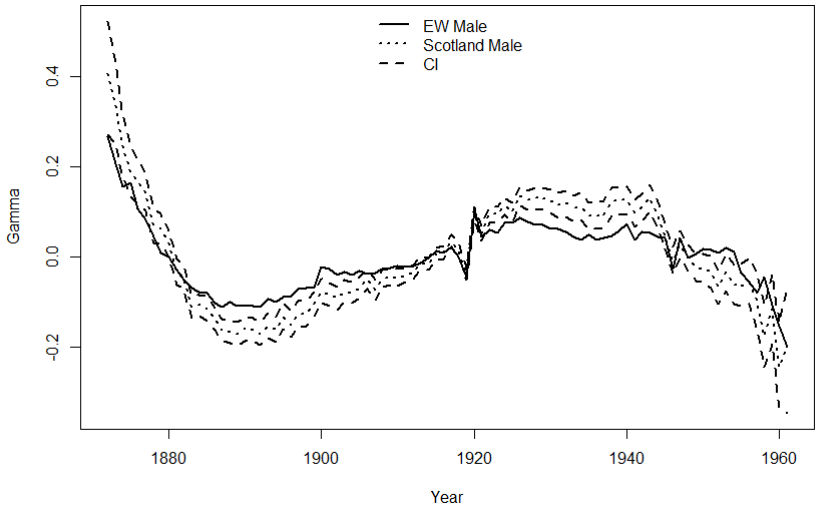


Figure 4.6: The estimates of cohort effect, for the males of England and Wales (solid line) and Scotland (dotted line), age 50 to 89 last birthday, over year 1961 to 2011. The dashed lines are the CI for the cohort effect of Scotland. The upper bound is 95% quantile of the distribution and the lower bound is 5% quantile.

For the LR test we again choose $\boldsymbol{\gamma}_0^{(4)} = \hat{\boldsymbol{\gamma}}^{EW}$ and then test the hypothesis that the true cohort effect for Scottish males is equal to $\boldsymbol{\gamma}_0^{(4)}$. The 99% quantile of a χ^2-distribution with 87 degrees of freedom is approximately 121. For the test statistic we find $\Gamma = 193.37$ and we therefore reject the null hypothesis and conclude that the cohort effect in Scotland is significantly different from the estimated cohort effect for England and Wales. This indicates that there might be factors in the Scottish male population that result in significant differences throughout time. However, we might speculate that there is a common cohort effect, that is, for some reason, magnified in Scotland. Investigating this in detail is beyond the scope of this paper, but we speculate that a magnified effect might be the result of socio-economic differences between the two populations: for example, cohort effects might be greater in lower socio-economic groups.

4.7 Summary

In this chapter, we investigated how a likelihood ratio test performs when applied to the mortality experience of a small population. To summarise, the finite sample

distribution of the test statistic is very close to the asymptotically correct χ^2 distribution; the power of the test depends strongly on the population size; the ability of the test to detect deviations from the null hypothesis being significantly reduced when the size of the underlying populations is small. A brief investigation of annuity prices has shown that the misspecification of parameters has a limited financial impact.

Chapter 5

Application of Bayesian Statistics

5.1 Review for the Previous Chapters

In Chapters 3 and 4, we investigated the finite sample distribution of the maximum likelihood estimators for the parameters of a stochastic mortality model. We found that the size of a population has a significant effect on the uncertainty about the estimated parameters and mortality projections. In particular, we found that there exists a bias in the estimated covariance matrix of the random walk fitted to the period effects when the size of the underlying population is small. As a consequence, prediction intervals are rather wide for small populations even when parameter uncertainty is ignored.

To investigate if parameters estimated from larger populations can be used to generate scenarios for smaller populations we investigated how a likelihood ratio test performs when applied to the mortality experience of a small population. We found that the finite sample distribution of the test statistic is very close to the asymptotically correct χ^2-distribution and, therefore, the observed rejection rates are close to the chosen significance level. However, we also found that the power of the test depends strongly on the population size with the ability of the test to

detect deviations from the null hypothesis being significantly reduced when the size of the underlying populations is small.

A brief investigation of annuity prices has shown that the misspecification of parameters has a limited financial impact. Considering shifts in the parameter values which the LR test would detect with a 50% chance we have seen that the impact of a small population size is significant for deferred annuities. To have a complete picture of possible further financial consequences, a more detailed study is required, which is beyond the scope of this study.

In our empirical analysis we then applied the LR test, and found that neither of the mortality rates of the female population in England and Wales and the male population in Scotland should be modelled with a cohort effect estimated from the male population in England and Wales.

In Chapters 3 and 4, we used the traditional two-stage fitting approach whereby the period and cohort effects are estimated using the Poisson maximum likelihood method in the first stage and a time series model is fitted to these effects in the second stage. We have found that sampling variation in the small population datasets has significant impact, which can then obscure the true signal in those effects, and giving rise to misleading forecasts. Bayesian approaches that combine the two stages into one, e.g., Pedroza (2006), Cairns et al. (2011b) and Czado et al. (2005)) can be used to provide a way to address this problem. However, as use of the two-stage approach is widespread (perhaps because of its relative simplicity) we have, here, attempted the first systematic analysis of the impact of population size on parameter estimates and forecasts using the two-stage approach. In this way, users of the two-stage approach will be better informed about its limitations as well as understanding how the likelihood ratio test might be used to exploit data from larger populations.

The limitations of the two-stage approach give us the motivation to combine these two stages into one, more specifically, adding the time series prior for the latent

parameters to the Poisson likelihood, and investigate the joint posterior distribution of all the parameters conditional on the observed dataset on a Bayesian inference. For large population when Poisson likelihood dominates the posterior distribution, the impact of the time series prior is little. However, it will compete with the Poisson likelihood given a much smaller population such that the parameter estimation looks more like it is generated from the proposed time series model[1], that is we balance the huge noise of the latent parameter estimation due to the large sampling variation with the smoothness involved in the ARIMA models.

The Bayesian approach allows us to include short cohorts with few observations. For instance, the first step of the two-stage approach that maximises only the Poisson likelihood will greatly drive the estimates for the cohorts with only one observation far away (either increase or decrease) from the estimates for the rest of the cohorts and hence significantly increases the variance of the projecting model. In contrast, the low level information contained by the one observation will be balanced by the ARIMA likelihood with a wider posterior distribution.

Most importantly, the Bayesian approach allows us to employ the knowledge of a related larger population (e.g. England and Wales) by choosing a more informative prior distribution for the parameter we are interested in, which may provide a more accurate estimation. This is meaningful for the managers of a small pension scheme if they believe, referring to evidence, that their population shares common characteristics with a large reference population.

In this and the following chapters, we combine the two stages into one by adding time series prior distributions for the latent parameters (i.e. period and cohort effects) to obtain a joint posterior distribution of the parameters we use and study the impact of the time series prior on parameter estimation and mortality projection.

24

[1]See Cairns et al. (2011b)

5.2 Review of Bayesian Statistics

See details of Bayesian statistics in Gelman et al. (2014).

Denote as $p(\cdot|\cdot)$ a conditional density with its argument determined by the following context and similarly as $p(\cdot)$ the marginal density. Given a data vector $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ with n data points, that can be modelled by a parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m) = (\theta_1, \ldots, \theta_K)$ with length $K \geq m$, such that both $\boldsymbol{y}_i$ and $\boldsymbol{\theta}_j$ can either be a single scalar or a vector.

Given the observed value of $\boldsymbol{y}$, the Bayes' rule tells us that the posterior distribution of $\boldsymbol{\theta}$ conditional on $\boldsymbol{y}$ is

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta}, \boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})}, \tag{5.1}$$

where

$$p(\boldsymbol{y}) = \int p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta} \tag{5.2}$$

for continuous $\boldsymbol{\theta}$ and $p(\boldsymbol{y}) = \sum_{\boldsymbol{\theta}} p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$ for discrete $\boldsymbol{\theta}$ is independent on $\boldsymbol{\theta}$. Thus (5.1) can be written as

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{5.3}$$

$p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood of $\boldsymbol{y}$ conditional on $\boldsymbol{\theta}$ and $p(\boldsymbol{\theta})$ is the prior distribution for $\boldsymbol{\theta}$.

In general, we usually have a $\boldsymbol{\theta}$ with length $m > 1$ and $\boldsymbol{\theta}_i$ a vector. Denote as:

- $\boldsymbol{\theta}_{-i} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_m)$;

- $p(\boldsymbol{\theta}_{-i})$ the joint density of $\boldsymbol{\theta}_{-i}$;

- $\boldsymbol{\theta}_{i,-k} = (\theta_{i,1}, \ldots, \theta_{i,k-1}, \theta_{i,k+1}, \ldots, \theta_{i,K_i})$.

With the chain rule we can re-write $p(\boldsymbol{\theta})$ as

$$\begin{aligned} p(\boldsymbol{\theta}) &= p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i})p(\boldsymbol{\theta}_{-i}) \\ &= p(\theta_{i,k}|\boldsymbol{\theta}_{i,-k},\boldsymbol{\theta}_{-i})p(\boldsymbol{\theta}_{i,-k}|\boldsymbol{\theta}_{-i})p(\boldsymbol{\theta}_{-i}) \\ &= p(\theta_{i,k}|\boldsymbol{\theta}_{i,-k})p(\boldsymbol{\theta}_{i,-k})p(\boldsymbol{\theta}_{-i}), \end{aligned} \tag{5.4}$$

by assuming that each $\theta_{i,k} \in \boldsymbol{\theta}_i = (\theta_{i,1}, \ldots, \theta_{i,K_i})$ only directly depends on $\boldsymbol{\theta}_{i,-k}$, that is $\theta_{i,k}$ is conditionally independent given $\boldsymbol{\theta}_{-i}$. Thus by repeatedly applying the chain rule to $p(\boldsymbol{\theta}_{i,-k})$ and $p(\boldsymbol{\theta}_{-i})$ in Equation (5.4),

$$p(\boldsymbol{\theta}) = p(\theta_{1,1}|\boldsymbol{\theta}_{1,-1})\cdots p(\theta_{m,K_m}) \tag{5.5}$$

and eventually we have

$$\int p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} \propto \underbrace{\int_{\theta_{1,1}} \cdots \int_{\theta_{m,K_m}}}_{K \text{ in total}} p(\boldsymbol{y}|\boldsymbol{\theta})p(\theta_{1,1}|\boldsymbol{\theta}_{1,-1}) \cdots p(\theta_{m,K_m})d\theta_{m,K_m} \cdots d\theta_{1,1}. \tag{5.6}$$

Thus the inference about $\boldsymbol{\theta}$ is based on the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ and (5.6) shows that obtaining the joint posterior distribution for $\boldsymbol{\theta}$ conditional on the observed data $\boldsymbol{y}$ requires integrating the joint posterior density $p(\boldsymbol{\theta}|\boldsymbol{y})$ with respect to each $\theta_{i,k}$ for $i = 1, \ldots, m$ and $k = 1, \ldots, K_i$. This process can be complicated if the number of parameters concerned is large.

To implement the Bayesian approach, we combine the two stage approach into one by combining the Poisson likelihood for the death counts $\boldsymbol{D}$ conditional on the parameter vector $\boldsymbol{\theta}$ with the ARIMA likelihood functions (or densities) for the latent parameters for period and cohort effects. More specifically our model[2] is:

[2]Note that this is not the only choice for the model structure. More dispersion would be allowed by using, for example, a Poisson-LogNormal(LN)-CBD typed model, where we can assume the death rate $m(t, x)$ is a log-normal random variable. However, in this thesis we will be focusing on our underlying model settings.

$$D(t,x)|\theta_1 \sim \text{Pois}(m(\theta_1,t,x)E(t,x)) \tag{5.7}$$

$$m(\theta_1,t,x) = -\log(1-q(\theta_1,t,x)) \tag{5.8}$$

$$\text{logit } q(\theta_1,x,t) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_c^{(4)} \tag{5.9}$$

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \text{ for } t \geq t_2 \tag{5.10}$$

$$\gamma_c^{(4)} = \alpha_\gamma \gamma_{c-1}^{(4)} + \epsilon_c, \text{ for } c > t_1 - x_{n_a}, \tag{5.11}$$

that is the period effects $\boldsymbol{\kappa}_t$ are modelled by a multi-variate random walk model with drift $\boldsymbol{\mu} = (\mu_1,\mu_2,\mu_3)^T$ and an $i.i.d$ three dimensional multi-variate normal error term, $\boldsymbol{\epsilon}_t \sim MVN(\mathbf{0},\boldsymbol{V}_{\boldsymbol{\epsilon}})$. $\boldsymbol{V}_{\boldsymbol{\epsilon}} = \{V_\epsilon(i,k)\}$ for $i,k = 1,2,3$ is the co-variance matrix for $\boldsymbol{\epsilon}_t$, which is positive definite and independent of t. An AR(1) model with zero mean reverting is fitted to the cohort effect $\gamma_c^{(4)}$ and ϵ_c is $i.i.d$ normal with mean 0 and variance σ_γ^2. See further discussion in Section 5.4.2. The Equations (5.10)-(5.11) introduce $\boldsymbol{\theta}_2 = (\boldsymbol{\mu},\boldsymbol{V}_{\boldsymbol{\epsilon}},\alpha_\gamma,\sigma_\gamma^2)$ as the hyper-parameters.

The posterior density for $\boldsymbol{\theta}$ conditional on the observed death counts and exposures, $p(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{E}) = p(\boldsymbol{\kappa},\boldsymbol{\gamma},\boldsymbol{\mu},\boldsymbol{V}_{\boldsymbol{\epsilon}},\alpha_\gamma,\sigma_\gamma|\boldsymbol{D},\boldsymbol{E})$, can be expressed as:

$$p(\boldsymbol{\kappa},\boldsymbol{\gamma},\boldsymbol{\mu},\boldsymbol{V}_{\boldsymbol{\epsilon}},\alpha_\gamma,\sigma_\gamma|\boldsymbol{D},\boldsymbol{E}) \propto p(\boldsymbol{D},\boldsymbol{E}|\boldsymbol{\kappa},\boldsymbol{\gamma},\boldsymbol{\mu},\boldsymbol{V}_{\boldsymbol{\epsilon}},\alpha_\gamma,\sigma_\gamma)p(\boldsymbol{\kappa},\boldsymbol{\gamma},\boldsymbol{\mu},\boldsymbol{V}_{\boldsymbol{\epsilon}},\alpha_\gamma,\sigma_\gamma) \tag{5.12}$$

$$= p(\boldsymbol{D},\boldsymbol{E}|\boldsymbol{\kappa},\boldsymbol{\gamma})p(\boldsymbol{\kappa},\boldsymbol{\gamma}|\boldsymbol{\mu},\boldsymbol{V}_{\boldsymbol{\epsilon}},\alpha_\gamma,\sigma_\gamma)p(\boldsymbol{\mu},\boldsymbol{V}_{\boldsymbol{\epsilon}},\alpha_\gamma,\sigma_\gamma). \tag{5.13}$$

The last equation implies referring to (5.7) that $(\boldsymbol{D},\boldsymbol{E})$ is conditionally independent of all the hyper-parameters.

5.3 Bayesian Statistics and Markov Chain Monte Carlo

Obtaining the joint posterior distribution of $\boldsymbol{\theta} = (\boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{V_\epsilon}, \alpha_\gamma, \sigma_\gamma)$ given the observed data $(\boldsymbol{D}, \boldsymbol{E})$ requires integrating Equation (5.13) with respect to all the $\boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{V_\epsilon}, \boldsymbol{\mu}, \alpha_\gamma$ and σ_γ. Integration with respect to such high dimensioned parameter vector is technically complicated, in particular the total number of parameters in our model is 257 which, on the other hand, makes drawing samples directly from the joint posterior distribution almost impossible.

Instead of directly drawing samples from such high-dimensioned joint distribution, Markov Chain Monte Carlo (MCMC) method is employed to simulate samples that can be termed as if they are drawn from the true posterior distribution, with density $p(\boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{V_\epsilon}, \alpha_\gamma, \sigma_\gamma | \boldsymbol{D}, \boldsymbol{E})$. The aim of MCMC, based on Gelman et al. (2014), is to simulate a random walk path in the space of $\boldsymbol{\theta}$ that eventually converges to our target distribution $p(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E})$. Gibbs sampler and Metropolis-Hastings (M-H) algorithm have been widely used to draw samples from $p(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E})$ in previous studies, e.g. Cairns et al. (2011b), Czado et al. (2005), Pedroza (2006). Cairns et al. (2011b) employed the Bayesian and MCMC framework for a two-population modelling problem, where one population dominates the other. In Pedroza (2006), Kalman filter was adopted to estimate and project the state-space model under the Bayesian and MCMC framework.

5.3.1 Hamiltonian Monte Carlo

In this study, we adopted the dynamical sampling, in particular the Hamiltonian Monte Carlo (HMC) method as our updating algorithm rather than the MH algo-

rithm.[3] Dynamical sampling is based on a physical analogy.[4] One big advantage of HMC over the M-H algorithm is that it avoids the random walk behaviour involved in the M-H method.

In this section we only briefly introduce the algorithm of HMC. Since our study focuses on modelling the longevity risk and projecting the mortality rates for small population under the Bayesian inference with the aid of MCMC updating methodology, justifying the principle of HMC method itself is beyond the scope of our study. Detailed derivation, justification and discussion for the principle of HMC can be found in Neal (1993) and the principle of Hamilton dynamic system can be found in De Almeida (1990).

Canonical Function

Suppose we wish to sample from the joint posterior distribution with density $f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{y})$ for an arbitrary vector of random variables $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ with length n conditional on the observed data $\boldsymbol{y}$.

Define as $E_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ and $f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ any energy function[5] and posterior density function for $\boldsymbol{X}$ respectively. $E_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ is assumed to be differentiable with respect to x_i for $i = 1, \ldots, n$. We can transfer the energy function to density with the canonical function, such as

$$f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z_E} \exp(-E_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})) \tag{5.14}$$

$$\propto \exp(-E_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})) \tag{5.15}$$

where Z_E is some constant that ensures the density function integrates to one. In

[3] Appendix C.1 outlines the MH algorithm and illustrates why it does not work so well for this problem.

[4] "The gradient of the potential energy for a physical system with respect to its configuration coordinates defines the 'force' that acts to change this configuration, via its effect on the momentum." See Neal (1993).

[5] See Appendix C.2.1 in C.2

this study, we are concerned to draw samples from the joint posterior distribution of $\boldsymbol{\theta}$ conditional on $(\boldsymbol{D}, \boldsymbol{E})$, that is we fix $\boldsymbol{x}|\boldsymbol{y} = \boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E}$.

Hamiltonian Dynamics and Total Energy Function

We then define $U_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ as the potential energy function[6] of $\boldsymbol{X}|\boldsymbol{y}$ and is assumed to be differentiable with respect to x_i for $i = 1, \ldots, n$. (5.15) shows that the link between the density and the potential energy is

$$f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) \propto \exp(-U_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{y})). \tag{5.16}$$

We then introduce another arbitrary vector of random variables, denoted as $\boldsymbol{P} = (P_1, \ldots, P_n)$. In physics, P_i is called a momentum variable. Note that the length of $\boldsymbol{P}$ is identical with $\boldsymbol{X}$ and each P_i is referred to X_i for $i = 1, \ldots, n$. Define as $K_{\boldsymbol{P}}(\boldsymbol{p})$ the kinetic energy function and the density of $\boldsymbol{P}$ can be written as

$$f_{\boldsymbol{P}}(\boldsymbol{p}) \propto \exp(-K_{\boldsymbol{P}}(\boldsymbol{p})). \tag{5.17}$$

The choice of the density for $\boldsymbol{P}$ is irrelevant to the sampling for $\boldsymbol{X}$ (Equation (5.19)) and therefore $K_{\boldsymbol{P}}(\boldsymbol{p})$ can be any function given Equation (5.17) holds.

In physics, the combination of $\boldsymbol{X}|\boldsymbol{y}$ and $\boldsymbol{P}$ is known as a phase space and the total energy function for points on this space can be written as the sum of the potential and kinetic energy. More specifically,

$$H_{\boldsymbol{X}|\boldsymbol{y},\boldsymbol{P}}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{p}) = U_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) + K_{\boldsymbol{P}}(\boldsymbol{p}). \tag{5.18}$$

Define as $f_{\boldsymbol{X}|\boldsymbol{y},\boldsymbol{P}}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{p})$ the joint density of $\boldsymbol{X}|\boldsymbol{y}$ and $\boldsymbol{P}$. Based on (5.15), we

[6]See Appendix C.2

can transform it to:

$$\begin{aligned} f_{\boldsymbol{X}|\boldsymbol{y},\boldsymbol{P}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{p}) &\propto \exp(-H_{\boldsymbol{X}|\boldsymbol{y},\boldsymbol{P}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{p})) \\ &= \exp(-U_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}) - K_{\boldsymbol{P}}(\boldsymbol{p})) \\ &= \exp(-U_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) \exp(-K_{\boldsymbol{P}}(\boldsymbol{p})) \\ &\propto f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) f_{\boldsymbol{P}}(\boldsymbol{p}). \end{aligned} \tag{5.19}$$

(5.19) shows that $\boldsymbol{X}|\boldsymbol{y}$ is independent of $\boldsymbol{P}$ and the posterior distribution from which we wish to draw sample for $\boldsymbol{X}$ is the marginal distribution of $\boldsymbol{X}|\boldsymbol{y}$ with respect to the phase space. It means that if we can find a methodology to sample $(\boldsymbol{X}|\boldsymbol{y},\boldsymbol{P})$ from $f_{\boldsymbol{X}|\boldsymbol{y},\boldsymbol{P}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{p})$, the value of the sample for $\boldsymbol{X}$ is exactly drawn from the posterior distribution $f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ if we ignore the value for $\boldsymbol{P}$. (5.19) implies that the choice of the distribution of $\boldsymbol{P}$ will not affect the sampling for $\boldsymbol{X}$. Thus the distribution for $\boldsymbol{P}$ can be determined arbitrarily and literately P_i is assumed to be $i.i.d$ normal distribution with zero mean and unit variance, more specifically $P_i \sim N(0,1)$ for $i = 1, \ldots, n$.

We then define the Hamiltonian dynamics,

$$\frac{\partial\ x_i|\boldsymbol{y}}{\partial\tau} = \frac{\partial H}{\partial p_i} = \frac{\partial K_{\boldsymbol{P}}(\boldsymbol{p})}{\partial p_i} \tag{5.20}$$

$$\frac{\partial p_i}{\partial\tau} = -\frac{\partial H}{\partial\ x_i|\boldsymbol{y}} = -\frac{U_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})}{\partial x_i}. \tag{5.21}$$

In physics, τ is the time parameter and it is artificial in statistic inference. In this study, τ represents the state of the iteration in the MCMC algorithm.

The Leapfrog Discretization

The Hamiltonian dynamics is defined by (5.20) and (5.21) is hard to follow in practice and we need to discretize the state change from state τ to $\tau + 1$ with some non-zero steps. It is known that discretizing differentiation inevitably introduces errors. Literately a leapfrog discretization method is used such that the Liouville's theorem[7] still holds.

Define as $p_i(\tau)$ and $x_i(\tau)$ the current value of p_i and x_i respectively at iteration τ. Also define as L the number of equal-length small steps between τ to $\tau + 1$ with step size δ such that $\delta = \frac{1}{L}$. The algorithm of leapfrog method are as follows:

1. Calculate a half step with step size $\frac{\delta}{2}$ forward for $p_i(\tau)$, that is

$$p_i(\tau + \frac{\delta}{2}) = p_i(\tau) - \frac{\delta}{2} \frac{\partial U_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})}{\partial x_i}\Big|_{\boldsymbol{x}=\boldsymbol{x}(\tau)} \tag{5.22}$$

2. Take a full step δ for $x_i(\tau)$:

$$x_i(\tau + \delta) = x_i(\tau) + \delta \frac{\partial K_{\boldsymbol{P}}(\boldsymbol{p})}{\partial p_i}\Big|_{\boldsymbol{p}=\boldsymbol{p}(\tau+\frac{\delta}{2})} \tag{5.23}$$

3. Take another half step for $p_i(\tau + \frac{\delta}{2})$:

$$p_i(\tau + \delta) = p_i(\tau + \frac{\delta}{2}) - \frac{\delta}{2} \frac{\partial U_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})}{\partial x_i}\Big|_{\boldsymbol{x}=\boldsymbol{x}(\tau+\delta)}. \tag{5.24}$$

The choice of L needs to be sufficiently large so that the value of δ is small enough to give an acceptable error.[8]

[7] Liouville's theorem:

$$\sum_i \Big[\frac{\partial}{\partial x_i}\Big(\frac{dx_i}{d\tau}\Big) + \frac{\partial}{\partial p_i}\Big(\frac{dp_i}{d\tau}\Big)\Big] = \sum_i \Big[\frac{\partial H}{\partial x_i \partial p_i} - \frac{\partial H}{\partial p_i \partial x_i}\Big] = 0$$

See more details on Liouville's theorem and its application in Neal (1993).

[8] See Neal (1993) for further discussion on how the size of δ will affect the magnitude of the error.

Updating Algorithm

Recall we fix $\boldsymbol{x}|\boldsymbol{y} = \boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E}$ where $\boldsymbol{\theta}$ is the parameter vector including all the latent and hyper parameters. Given the discretization method introduced in (5.22)-(5.24) the updating algorithm is as follows:

- Given the current value for $\boldsymbol{\theta}$ and $\boldsymbol{p}$ at iteration τ, denoted as $(\hat{\boldsymbol{\theta}}(\tau), \hat{\boldsymbol{p}}(\tau))$, sample a new momentum variable through the canonical function $\tilde{\boldsymbol{p}}(\tau) \sim f_{\boldsymbol{P}}(\hat{\boldsymbol{p}}(\tau))$.

- From state τ to $\tau + 1$, perform step (5.22)-(5.24) by L times with step size δ. Denote as $(\boldsymbol{\theta}^*, \boldsymbol{p}^*)$ the ending value.

- Calculate the Metropolis acceptance probability:

$$\alpha = \min\Big\{1, \exp\Big(- U_{\boldsymbol{\theta}}(\boldsymbol{\theta}^*) + U_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) - K_{\boldsymbol{P}}(\boldsymbol{p}^*) + K_{\boldsymbol{P}}(\hat{\boldsymbol{p}})\Big)\Big\} \tag{5.25}$$

- Draw a random number $u \sim U(0, 1)$ where $U(\cdot)$ represents a uniform distribution. If $u \leq \alpha$, accept the new state value $(\boldsymbol{\theta}^*, \boldsymbol{p}^*)$ as the current value for iteration $\tau + 1$, else $(\hat{\boldsymbol{\theta}}(\tau), \hat{\boldsymbol{p}}(\tau))$ is kept for state $\tau + 1$.

5.4 Posterior Distribution for $\boldsymbol{\theta}$

In this section, we discuss the components of the posterior distribution for $\boldsymbol{\theta}$. Recall that (5.13) is the posterior density of $\boldsymbol{\theta}$ conditional on the observed death counts

and exposures, which can be written further as:

$$\begin{aligned} p(\boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{V_\epsilon}, \alpha_\gamma, \sigma_\gamma | \boldsymbol{D}, \boldsymbol{E}) &\propto p(\boldsymbol{D}, \boldsymbol{E} | \boldsymbol{\kappa}, \boldsymbol{\gamma}) p(\boldsymbol{\kappa}, \boldsymbol{\gamma} | \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_2). \\ &= p(\boldsymbol{D}, \boldsymbol{E} | \boldsymbol{\kappa}, \boldsymbol{\gamma}) p(\boldsymbol{\kappa} | \boldsymbol{\theta}_2, \boldsymbol{\gamma}) p(\boldsymbol{\gamma} | \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_2) \qquad (5.26) \\ &= p(\boldsymbol{D}, \boldsymbol{E} | \boldsymbol{\kappa}, \boldsymbol{\gamma}) p(\boldsymbol{\kappa} | \boldsymbol{\mu}, \boldsymbol{V_\epsilon}) p(\boldsymbol{\gamma} | \alpha_\gamma, \sigma_\gamma) p(\boldsymbol{\theta}_2) \end{aligned}$$

(5.27)

where $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}, \boldsymbol{V_\epsilon}, \alpha_\gamma, \sigma_\gamma)$ is the vector of all the hyper-parameters. The last equation holds refer to (5.10) and (5.11) such that $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$ are conditionally independent of $(\alpha_\gamma, \sigma_\gamma)$ and $(\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ respectively.

5.4.1 Prior Distributions for $\boldsymbol{\kappa}$, $\boldsymbol{\mu}$ and $\boldsymbol{V_\epsilon}$

We start with the prior distributions for the period effects and their corresponding hyper-parameters. Recall in (5.10) the period effects are modelled by a multi-variate random walk for $t \geq t_2$, more specifically

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \text{for } t \geq t_2$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^T$ and $\boldsymbol{\epsilon}_t \sim MVN(\boldsymbol{0}, \boldsymbol{V_\epsilon})$ an $i.i.d$ three dimensional multi-variate normal error term independent of t. Let $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{t_2}, \ldots, \boldsymbol{\Delta}_{t_{n_y}})$ where $\boldsymbol{\Delta}_t = \boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1}$ for $t \geq t_2$ is $i.i.d$ and independent of t refer to our random walk model. The distribution of $\boldsymbol{\Delta}_t | \boldsymbol{\mu}, \boldsymbol{V_\epsilon}$ is

$$\boldsymbol{\Delta}_t | \boldsymbol{\mu}, \boldsymbol{V_\epsilon} \sim MVN(\boldsymbol{\mu}, \boldsymbol{V_\epsilon}). \qquad (5.28)$$

By assuming the prior for $\boldsymbol{\kappa}_{t_1} \propto 1$, the prior density of $\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}$ can be written as

$$p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}) \propto |\boldsymbol{V}_{\boldsymbol{\epsilon}}|^{-\frac{n_y-1}{2}} \exp\Big(-\frac{1}{2}\sum_{t=t_2}^{t_{n_y}} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})^T \boldsymbol{V}_{\boldsymbol{\epsilon}}^{-1} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})\Big). \quad (5.29)$$

Note that $p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}})$ is a well-defined distribution given $n_y \geq 4$. We are not claiming that a flat uniform prior is the only choice for $\boldsymbol{\kappa}_{t_1}$. For example it is a common practice to set identifiability constraints by fixing it to a constant, e.g. $\boldsymbol{\kappa}_{t_1} = 0$, otherwise there is a convergence problem by employing the Metropolis-Hastings algorithm. However there is no such problem with the Hamiltonian Monte Carlo[9]. Further it makes more sense for the industrial practice to investigate the joint posterior distribution of all the parameters rather than partially fixing certain parameters. It is hard to interpret the rationale for fixing parameters and the choice of parameters that should be fixed.

We then need to determine the prior distribution for $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ by balancing (a) reflecting as much information as possible contained in the data, that is "letting the data speak for themselves" and (b) combining our prior knowledge about the underlying data. Attention should also be paid to the MCMC convergence if a noninformative prior is employed for small population.

Noninformative Prior Density for $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$

The biggest advantage of noninformative prior density is its convenience and an improper prior distribution is desirable if no prior knowledge is available for the parameters and the prior distribution is required to play a minimum role in the posterior distribution, that is to let the data speak for themselves.

[9]This is mainly because: the Hamiltonian Monte Carlo (HMC) does not have the random walk behaviour as the Metropolis-Hastings algorithm does. It obtains the gradient of the log joint posterior density. Assuming a flat uniform prior effectively contributes a constant to the joint log-density. Further, our time series prior distributions provide sufficient information for our model to determine the posterior distribution, especially when the population size is small. Given the observations for the death counts, our likelihood is not absolutely flat.

The facts that (a) we have no prior knowledge about the true parameters for the parameters of benchmark population and (b) the Poisson likelihood dominates the posterior distribution when the population is large and therefore the influence of the prior distribution is little imply that we may not need to work especially hard to find an informative prior. Literately the choice for the joint prior density $p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ is the Jeffreys prior distribution given no prior knowledge is available. More specifically, for the benchmark population:

$$p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon}) \propto |\boldsymbol{V_\epsilon}|^{-2}. \tag{5.30}$$

Combining with the prior density $p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ in (5.29) we have the joint prior density for $\boldsymbol{\kappa}$, $\boldsymbol{\mu}$ and $\boldsymbol{V_\epsilon}$:

$$p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V_\epsilon})p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon}) \propto |\boldsymbol{V_\epsilon}|^{-\frac{n_y+4-1}{2}} \exp\Big(-\frac{1}{2}\sum_{t=t_2}^{t_{n_y}} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})^T \boldsymbol{V_\epsilon}^{-1} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})\Big). \tag{5.31}$$

However, it is worth noticing that the importance of the prior distribution increases as the population size decreases. As for the small population when the prior distribution has a strong impact on the posterior distribution, an uninformative prior distribution may cause problem of convergence for MCMC, given an under-identified model or non-identified parameters. Our model is under-identified given the observed deaths $\boldsymbol{D}$ and exposure $\boldsymbol{E}$ because the likelihood function is nearly equal for a range of values of $\boldsymbol{\theta}$ without applying the identifiability constraints. The reason for not applying the constraints is discussed in Section 5.4.2. Theoretically there is no single point to which the posterior distribution can converge given an under-identified model.[10] In practice, the MCMC estimate for the period effects converges to a straight line for the simulated death scenarios where $w = 0.01$, given the Jeffrey's prior for $p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ and this is possible for all w. Our solution to deal this problem is to employ the external information implied in the benchmark population

[10]See Gelman et al. (2014)

to create a more informative prior distribution.

Informative Prior Density for $\boldsymbol{V_\epsilon}$

There are several advantages to apply an informative prior with information in a large population. It allows us to adjust the value of the parameters of the prior distribution so that the prior is strong enough (lower variance) to guarantee a convergent MCMC process.

We will have a more accurate estimation for $\boldsymbol{\theta}$ by adopting information from the benchmark population. That is, if the managers of a small pension scheme believe, referring to reliable evidences, that their underlying exposure is a subset of a much larger population with similar characteristics or if the unknown true parameters are a function (e.g. linearly related) of the estimates of the large population (e.g. the parameter estimates of males in England and Wales, $\hat{\boldsymbol{\theta}}^{\text{EW}}$), then including such information in the prior distribution (e.g. assuming a particular statistic of the prior is a function of $\hat{\boldsymbol{\theta}}^{\text{EW}}$) will probably provide better estimation. This study is a special case where the true parameters of the small population death scenarios are set equal to the point estimates of the benchmark population. Although it is unrealistic to be aware of the exact true parameters in real problems (otherwise we are done!), we are aiming to show a methodology for employing external information and how much improvement on the parameter estimation there will be compared with the two-stage approach.

Further, choosing an informative prior allows us to investigate how sensitive our posterior distribution is with respect to the selection of prior distribution. We are interested in how the posterior distribution differs when the information of large population is adopted in different forms.

Recall that we simulate death scenarios from (3.2) (see page 29):

$$D_j^w(t,x) \sim \text{Pois}\big(m(\theta_{1,0},t,x)wE_0(t,x)\big) \qquad \text{for all } j = 1,\ldots,N_1,$$

where we fix the give parameter vector $\theta_{1,0} = \hat{\theta}_1^{\text{EW}}$ to ensure that the true parameters of $D_j^w(t,x)$ are the latent parameter estimates of the benchmark population (males in England and Wales). We employ the Bayesian approach for one percent of the benchmark population, that is $w = 0.01$. Similar studies can be carried for other values of w. The prior for the co-variance matrix $\boldsymbol{V_\epsilon}$ for $w = 0.01$ is chosen to be an Inverse-Wishart distribution, more specifically

$$\boldsymbol{V_\epsilon} \propto \text{Inv- Wishart}(\nu, \boldsymbol{\Sigma}),$$

where the degrees of freedom ν and the 3×3 scale matrix $\boldsymbol{\Sigma}$ are assumed to be specific. The density of $\boldsymbol{V_\epsilon}$ can then be written as

$$p(\boldsymbol{V_\epsilon}) \propto |\boldsymbol{V_\epsilon}|^{-\frac{\nu+3+1}{2}} \exp\big(-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}\boldsymbol{V_\epsilon}^{-1})\big)$$

with mean

$$\text{E}[\boldsymbol{V_\epsilon}] = \frac{\boldsymbol{\Sigma}}{\nu-3-1} = \frac{\boldsymbol{\Sigma}}{\nu-4}, \text{for } \nu > 4,$$

and mode

$$\arg\max_{\boldsymbol{V_\epsilon}} p(\boldsymbol{V_\epsilon}) = \frac{\boldsymbol{\Sigma}}{\nu+3+1} = \frac{\boldsymbol{\Sigma}}{\nu+4}.$$

The variance of $V_\epsilon(i,k)$ for $i,k = 1,2,3$ is

$$\text{Var}[V_\epsilon(i,k)] = \frac{(\nu-3+1)V_\epsilon^2(i,k) + (\nu-3-1)V_\epsilon(i,i)V_\epsilon(k,k)}{(\nu-3)(\nu-3-1)^2(\nu-3-3)},$$

which can be simplified to

$$\frac{2V_\epsilon^2(i,i)}{(\nu-3-1)^2(\nu-3-3)} \text{ for } i = k.$$

We employ the information from the benchmark population by choosing the ν and $\boldsymbol{\Sigma}$ such that:

MCMC-Mean (MME): the mean of the prior distribution for $\boldsymbol{V_\epsilon}$ is set equal to the true parameters for the simulated death scenario, that is the estimated covariance matrix for England and Wales data, $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ which is calculated based on Equation (3.9) (see page 41). More specifically,

$$\text{E}[\boldsymbol{V_\epsilon}] = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}; \tag{5.32}$$

MCMC-Mode (MMO): the mode of the prior distribution for $\boldsymbol{V_\epsilon}$ is set equal to the true parameters for the simulated death scenario. More specifically,

$$\arg\max_{\boldsymbol{V_\epsilon}} p(\boldsymbol{V_\epsilon}) = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}. \tag{5.33}$$

It is worth to notice that the mean of the prior distribution for $\boldsymbol{V_\epsilon}$, given MMO is $\frac{\nu+4}{\nu-4}$ times higher than MME[11]. We run the MCMC given the MMO as a sensitivity test so that the impact of changing the prior distribution on the posterior distribution can be investigated.

This approach is essentially similar with an empirical Bayes approach where we determine the prior distribution for $\boldsymbol{V_\epsilon}$ based on the information from the data set of the larger population. In our case, death scenarios are simulated to be sub-populations, whose true parameters are the estimates of the benchmark population. We therefore demonstrate a methodology of how this kind of information could be used to form a prior distribution for the practitioners whose underlying populations

[11]The derivation is as follows: referring to 5.33 the scale matrix for MMO can be written as

$$\boldsymbol{\Sigma} = (\nu + 4)\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}.$$

Then the corresponding mean can be re-written as

$$\text{E}[\boldsymbol{V_\epsilon}] = \frac{\boldsymbol{\Sigma}}{\nu - 4} = \frac{\nu + 4}{\nu - 4}\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$$

are sub-populations of and share similar characteristics with a larger population.

At last we assume a improper uniform distribution for the drift $\boldsymbol{\mu}$. The reason for the choice is that by restricting the latent parameter estimation with the time series prior, we want the Bayesian model, in particular the proposed time series models, to determine the slopes and hence the trajectories of estimations for the period effects, which on the other hand enables us to study the influence of the non-Poisson-likelihood components in the joint posterior distribution on the estimated drifts of the period effects when the population size becomes smaller. In practice, the uniform assumption causes no problem for MCMC converging. We are not claiming that this is the only choice for the prior of $\boldsymbol{\mu}$. A multi-variate normal distribution that is conditional on $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is also a good candidate for the prior distribution of $\boldsymbol{\mu}$. In this study we focus on the uniform assumption.

The joint density of $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ can now be written as

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}) &= p(\boldsymbol{\mu}|\boldsymbol{V}_{\boldsymbol{\epsilon}})p(\boldsymbol{V}_{\boldsymbol{\epsilon}}) = p(\boldsymbol{\mu})p(\boldsymbol{V}_{\boldsymbol{\epsilon}}) \qquad (5.34) \\ &\propto p(\boldsymbol{V}_{\boldsymbol{\epsilon}}) \\ &\propto |\boldsymbol{V}_{\boldsymbol{\epsilon}}|^{-\frac{\nu+3+1}{2}} \exp\big(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{V}_{\boldsymbol{\epsilon}}^{-1})\big). \qquad (5.35) \end{aligned}$$

Combining with the prior density $p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ in (5.29) we have:

$$\begin{aligned} p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V_\epsilon})p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon}) &\propto |\boldsymbol{V_\epsilon}|^{-\frac{n_y+\nu+3}{2}} \exp\Big(-\frac{1}{2}\sum_{t=t_2}^{t_{n_y}} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})^T \boldsymbol{V_\epsilon}^{-1} (\boldsymbol{\Delta}_t - \boldsymbol{\mu}) \\ &\quad -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{V_\epsilon}^{-1})\Big) \\ &= |\boldsymbol{V_\epsilon}|^{-\frac{n_y+\nu+3}{2}} \exp\Big(-\frac{1}{2}\mathrm{tr}(\boldsymbol{V_\epsilon}^{-1}S_0) - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{V_\epsilon}^{-1})\Big) \\ &= |\boldsymbol{V_\epsilon}|^{-\frac{n_y+\nu+3}{2}} \exp\Big(-\frac{1}{2}\mathrm{tr}(S_0\boldsymbol{V_\epsilon}^{-1}) - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{V_\epsilon}^{-1})\Big) \\ &= |\boldsymbol{V_\epsilon}|^{-\frac{n_y+\nu+3}{2}} \exp\Big(-\frac{1}{2}\mathrm{tr}\Big(\big(S_0 + \boldsymbol{\Sigma}\big)\boldsymbol{V_\epsilon}^{-1}\Big)\Big), \qquad (5.36) \end{aligned}$$

where

$$S_0 = \sum_{t=t_2}^{t_{n_y}} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})(\boldsymbol{\Delta}_t - \boldsymbol{\mu})^T.$$

(5.36) shows that $p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V_\epsilon})p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ is of the same family with $p(\boldsymbol{V_\epsilon})$ with degree of freedom $n_y - 1 + \nu$ and scale matrix $S_0 + \boldsymbol{\Sigma}$, and hence a proper distribution.

Summary

In summary, the period effects $\boldsymbol{\kappa}$ is assumed to follow a three-dimensional multivariate random walk with drift $\boldsymbol{\mu}$ and co-variance matrix $\boldsymbol{V_\epsilon}$. For the benchmark population, the joint prior density of $(\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ is assumed to be a non-informative Jeffrey's prior, while for the simulated death scenarios with population size equivalent to one percent of benchmark we have $\boldsymbol{V_\epsilon} \sim$ Inv- Wishart$(\nu, \boldsymbol{\Sigma})$ and $p(\boldsymbol{\mu}) \propto 1$. We choose the value of ν and $\boldsymbol{\Sigma}$ such that the mean of the prior for $\boldsymbol{V_\epsilon}$ is equivalent to the corresponding true parameter $\hat{\boldsymbol{V}_\epsilon}^{\mathrm{EW}}$. Sensitivity test is carried out by fixing the mode of the prior for $\boldsymbol{V_\epsilon}$ to the corresponding true parameters. In the next section we will introduce the prior distribution for the cohort effects and the respected hyper-parameters.

5.4.2 Prior Distributions for $\boldsymbol{\gamma}$, α_γ and σ_γ

We start with reviewing the constraints employed for the cohort effects referring to the two-stage approach. Recall that there is an identifiability problem involved in model M7 such that one add an arbitrary quadratic term with respect to cohort year to the respected cohort effect and make corresponding changes to the period effect without changing the fitted mortality rates. More specifically, one can add $\phi_1 + \phi_2(c - \bar{x}) + \phi_3(c - \bar{x})^2$ to $\gamma_c^{(4)}$ and adjust $\kappa_t^{(i)}$ for $i = 1, 2, 3$ without changing the fitted $q(\theta_1, t, x)$, where ϕ_1, ϕ_2 and ϕ_3 can be any value. Constraints on $\boldsymbol{\gamma}$ introduced by (2.6) (see page 26) have to be applied while maximising the Poisson likelihood in the two-stage approach to ensure that the least square estimator for ϕ_i is zero for $i = 1, 2, 3$.

As for the Bayesian approach, in (5.11) we fit a zero mean reverting AR(1) process to $\gamma_c^{(4)}$, more specifically:

$$\gamma_c^{(4)} = \alpha_\gamma \gamma_{c-1}^{(4)} + \epsilon_c \text{ for } c > t_1 - x_{n_a},$$

to replace the constraints (2.6). Note that this is not the only choice for dealing with the identifiability problem and fully calibrating the identifiability constraints is necessary for a fast convergent MCMC process if Metropolis-Hastings algorithm is employed.

The most important reason for not applying the constraints directly is that it will bring in much more complexity by changing the posterior distribution due to the quadratic term in M7. See discussion in Appendix C.3. We would like the model for the cohort effect to determine the posterior distribution. The zero mean-reverting AR(1) process assumed for the cohort effect is stronger than the identifiability constraints. Recall we have discussed the influence of the time series model, e.g. the AR(1) model for the cohort effect, significantly increases for smaller population and therefore a much stronger restriction such that the estimated latent

parameters, e.g. cohort effect, look more like the proposed time series models, e.g. the AR(1) zero mean-reverting model. Therefore, we expect the time series prior distributions provide sufficient information for the latent parameters while modelling the small populations. Such information and restrictions are strong enough to ensure a proper posterior density and converged MCMC by HMC, see Page 94 for the advantages of HMC. Note we are not claiming that Metropolis-Hastings can not obtain a converged MCMC given the underlying models. However, the random walk behaviour of the Metropolis-Hastings results in a very slow convergence.

Recall the model can be refined and eventually written as

$$\gamma_c^{(4)} = \alpha_\gamma \gamma_{c-1}^{(4)} + \epsilon_c, \text{ for } c > t_1 - x_{n_a}, \tag{5.37}$$

where $\epsilon_c \sim N(0, \sigma_\gamma^2)$ is an $i.i.d$ error term with its variance σ_γ^2 independent of cohort year c. The prior distribution for $\gamma_c^{(4)}$ can be written as

$$\gamma_c^{(4)} | \gamma_{c-1}^{(4)} \sim N(\alpha_\gamma \gamma_{c-1}^{(4)}, \sigma_\gamma^2), \text{ for } c > t_1 - x_{n_a} \tag{5.38}$$

$$\gamma_{t_1 - x_{n_a}}^{(4)} \sim N\Big(0, \frac{\sigma_\gamma^2}{1-\alpha_\gamma^2}\Big), \tag{5.39}$$

where (5.39) holds for a stationary AR(1) process. We introduce the hyper-parameters α_γ and σ_γ for the cohort effects.

The prior density of cohort effects conditional on the hyper-parameters $p(\boldsymbol{\gamma}|\alpha_\gamma, \sigma_\gamma)$ can be written as:

$$\begin{aligned} p(\boldsymbol{\gamma}|\alpha_\gamma, \sigma_\gamma^2) &= p(\gamma_{t_1-x_{n_a}}^{(4)}|\alpha_\gamma, \sigma_\gamma^2) p(\gamma_{t_1-x_{n_a}+1}^{(4)}, \ldots, \gamma_{t_{n_y}-x_1}^{(4)} | \gamma_{t_1-x_{n_a}}^{(4)}, \alpha_\gamma, \sigma_\gamma^2) \\ &\propto \Big(\frac{\sigma_\gamma^2}{1-\alpha_\gamma^2}\Big)^{-\frac{1}{2}} \exp\Big(-\frac{(\gamma_{t_1-x_{n_a}}^{(4)})^2(1-\alpha_\gamma^2)}{2\sigma_\gamma^2}\Big) \\ &\quad \times (\sigma_\gamma^2)^{-\frac{n_c-1}{2}} \exp\Big(\sum_{c=t_1-x_{n_a}+1}^{t_{n_y}-x_1} \frac{(\gamma_c^{(4)} - \alpha_\gamma \gamma_{c-1}^{(4)})^2}{-2\sigma_\gamma^2}\Big), \end{aligned} \tag{5.40}$$

where $n_c = n_y + n_a - 1$ is total number of cohort years.

At last we follow the common practice to assume that α_γ has a beta-type prior on $(-1, 1)$, more specifically, $p(\alpha_\gamma) \propto (1 - \alpha_\gamma^2)^g$ for $|\alpha_\gamma| < 1$, and σ_γ^2 follows an Inv-Gamma with shape parameter a_γ and scale parameter b_γ, that is $p(\sigma_\gamma^2) \propto (\sigma_\gamma^2)^{-a_\gamma - 1} \exp\Big(- \frac{b_\gamma}{\sigma_\gamma^2}\Big)$. The values of g, a_γ and b_γ are assumed to be known.

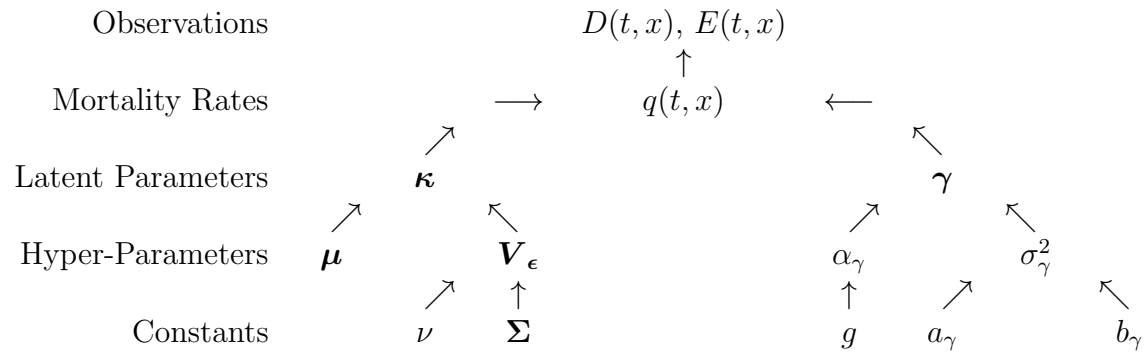
Combining with the prior density $p(\boldsymbol{\gamma}|\alpha_\gamma, \sigma_\gamma)$ in (5.40) we have:

$$p(\boldsymbol{\gamma}|\alpha_\gamma, \sigma_\gamma^2)p(\alpha_\gamma)p(\sigma_\gamma^2) \propto (\sigma_\gamma^2)^{-\frac{n_c+2a_\gamma+2}{2}} \Big(\frac{1}{1-\alpha_\gamma^2}\Big)^{-(\frac{1}{2}+g)} \exp\Big(- \frac{(\gamma_{t_1-x_{n_a}}^{(4)})^2(1-\alpha_\gamma^2)}{2\sigma_\gamma^2} - \sum_{c=t_1-x_{n_a}+1}^{t_{n_y}-x_1} \frac{(\gamma_c^{(4)} - \alpha_\gamma \gamma_{c-1}^{(4)})^2}{2\sigma_\gamma^2} - \frac{b_\gamma}{\sigma_\gamma^2}\Big), \tag{5.41}$$

which is a proper density given $\gamma_c^{(4)}$.

5.4.3 Joint Posterior Density

In this section, we start with summarising the structure of our model referring to the Bayesian approach in the following table. Note that the constants ν and $\boldsymbol{\Sigma}$ on the



fifth line is only available for the prior of co-variance matrix $\boldsymbol{V_\epsilon}$ given the simulated death scenario with population size $w = 0.01$. A non-informative prior is given to

the joint density of $\boldsymbol{\mu}$ and $\boldsymbol{V_\epsilon}$ for the benchmark population with no constant to be determined.

Recall (5.27) gives the joint posterior density of $\boldsymbol{\theta}$ conditional on the observed death counts and exposures as a product of the Poisson likelihood of data given all the parameters, density of the latent parameters given the respected hyper-parameters and the density for the hyper-parameters. More specifically,

$$p(\boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{V_\epsilon}, \alpha_\gamma, \sigma_\gamma^2 | \boldsymbol{D}, \boldsymbol{E}) \propto p(\boldsymbol{D}, \boldsymbol{E} | \boldsymbol{\kappa}, \boldsymbol{\gamma}) p(\boldsymbol{\kappa} | \boldsymbol{\mu}, \boldsymbol{V_\epsilon}) p(\boldsymbol{\gamma} | \alpha_\gamma, \sigma_\gamma^2) p(\boldsymbol{\theta}_2).$$

Note that we have not strictly proved that it is a proper distribution with uniform assumption for the $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}_{t_1}$, although we have shown that $p(\boldsymbol{\gamma}|\alpha_\gamma, \sigma_\gamma^2)p(\alpha_\gamma)p(\sigma_\gamma^2)$ is proper. However we have not identified any evidence that suggests our posterior distribution is not proper. As a matter of fact, our MCMC converges quickly using the Hamiltonian dynamic system, which implies that our prior settings for the other parameters (e.g. time series prior for the other latent parameters) are sufficiently informative without constraints on $\boldsymbol{\theta}_1$. See more discussion on convergence diagnostic for our MCMC outputs in Section 5.5.

Let $l(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E}) = \log p(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E})$ be the log-scaled joint posterior density that can be written as:

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E}) &= l_1(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E}) + l_{21}(\boldsymbol{\theta}_{11}) + l_{22}(\boldsymbol{\theta}_{12}|\boldsymbol{\theta}_{11}, \boldsymbol{\theta}_2) + l_{31}(\boldsymbol{\theta}_{13}|\boldsymbol{\theta}_2) \\ &\quad + l_{32}(\boldsymbol{\theta}_{14}|\boldsymbol{\theta}_{13}, \boldsymbol{\theta}_2) + l_4(\boldsymbol{\theta}_2) + \text{constant}, \end{aligned} \tag{5.42}$$

where

- $l_1(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E})$ = the Poisson log-likelihood function of all the parameters given the observed deaths.
- $l_{21}(\boldsymbol{\theta}_{11})$ = the un-conditional log density for $(\kappa_{t_1}^{(1)}, \kappa_{t_1}^{(2)}, \kappa_{t_1}^{(3)})^T$.

- $l_{22}(\boldsymbol{\theta}_{12}|\boldsymbol{\theta}_{11}, \boldsymbol{\theta}_2)$ = the conditional log density for $\boldsymbol{\kappa}_t$ for $t = t_2, \ldots, t_{n_y}$.

- $l_{31}(\boldsymbol{\theta}_{13}|\boldsymbol{\theta}_2)$ = conditional log density for $\gamma^{(4)}_{t_1 - x_{n_a}}$.

- $l_{32}(\boldsymbol{\theta}_{14}|\boldsymbol{\theta}_{13}, \boldsymbol{\theta}_2)$ = conditional log density for $\gamma^{(4)}_c$ for $c = t_1 - x_{n_a} + 1, \ldots, t_{n_y} - x_1$.

- $l_4(\boldsymbol{\theta}_2)$ = the log density for the hyper-parameters.

More specifically:

$$\begin{aligned} l_1(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E}) &= \sum_{t,x} \Big\{ D(t,x) \log m(\theta_1, t, x) - E(t,x) m(\theta_1, t, x) \Big\} \\ &\quad + \text{constant} \end{aligned} \tag{5.43}$$

$$\begin{aligned} l_{21}(\boldsymbol{\theta}_{11}) + l_{22}(\boldsymbol{\theta}_{12}|\boldsymbol{\theta}_{11}, \boldsymbol{\theta}_2) &= \text{constant} - \frac{1}{2} \Big\{ (n_y - 1) \log(|\boldsymbol{V}_{\boldsymbol{\epsilon}}|) \\ &\quad + \sum_{t=t_2}^{t_{ny}} (\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - \boldsymbol{\mu})^T \boldsymbol{V}_{\boldsymbol{\epsilon}}^{-1} (\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - \boldsymbol{\mu}) \Big\} \end{aligned} \tag{5.44}$$

$$\begin{aligned} l_{31}(\boldsymbol{\theta}_{13}|\boldsymbol{\theta}_2) + l_{32}(\boldsymbol{\theta}_{14}|\boldsymbol{\theta}_{13}, \boldsymbol{\theta}_2) &= -\frac{1}{2} \Big\{ \log \Big(\frac{\sigma^2_\gamma}{1 - \alpha^2_\gamma} \Big) + (n_c - 1) \log \sigma^2_\gamma \\ &\quad + \frac{(\gamma^{(4)}_{t_1 - x_{n_a}})^2 (1 - \alpha^2_\gamma)}{\sigma^2_\gamma} + \sum_{c = t_1 - x_{n_a} + 1}^{t_{ny} - x_1} \frac{(\gamma^{(4)}_c - \alpha_\gamma \gamma^{(4)}_{c-1})^2}{2\sigma^2_\gamma} \Big\} \\ &\quad + \text{constant.} \end{aligned} \tag{5.45}$$

For benchmark population

$$\begin{aligned} l_4(\boldsymbol{\theta}_2) &= -2 \log |\boldsymbol{V}_{\boldsymbol{\epsilon}}| + g \log(1 - \alpha^2_\gamma) - (a_\gamma + 1) \log \sigma^2_\gamma - \frac{b_\gamma}{\sigma^2_\gamma} \\ &\quad + \text{constant}, \end{aligned} \tag{5.46}$$

and for simulated death scenarios with population size $w = 0.01$

$$\begin{aligned} l_4(\boldsymbol{\theta}_2) &= -\frac{\nu+4}{2}\log|\boldsymbol{V}_{\boldsymbol{\epsilon}}| - \frac{1}{2}\mathrm{tr}\big(\boldsymbol{\Sigma}\boldsymbol{V}_{\boldsymbol{\epsilon}}^{-1}\big) \\ &\quad + g\log(1-\alpha_\gamma^2) - (a_\gamma+1)\log\sigma_\gamma^2 - \frac{b_\gamma}{\sigma_\gamma^2} + \text{constant}. \end{aligned} \tag{5.47}$$

5.4.4 Parameter Initialization and Model Fitting

Before carrying out the MCMC, we need to determine the starting point for both the latent and hyper parameters and the parameter value for the density of hyper-parameters.

As for the latent parameter $\boldsymbol{\theta}_1$, the initial values for benchmark population and the simulated death scenarios are the MLEs by maximising the Poisson likelihood only, more specifically, $\boldsymbol{\theta}_{1,0}$ and $\hat{\boldsymbol{\theta}}_1^w$ respectively, where $w = 0.01$. The starting values for $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ are the empirical estimates given the MLEs of the period effects. See Formulas (3.8)-(3.9), page 41. The starting point for α_γ and σ_γ^2 are estimated by fitting the model described in (5.37) to the empirical MLEs of the cohort effects.

We fix the parameter g of the prior density for α_γ to 1 for the benchmark and 3 for the simulated death scenarios. The choice of $g = 3$ for the simulated death scenarios with exposure size $w = 0.01$ is to guarantee a convergent MCMC for every death scenario. Recall that $\sigma_\gamma^2 \sim$ Inv-Gamma(a_γ, b_γ), the mean and variance of σ_γ^2 are given by $\frac{b_\gamma}{a_\gamma - 1}$ for $a_\gamma > 1$ and $\frac{b_\gamma^2}{(a_\gamma-1)^2(a_\gamma-2)}$ for $a_\gamma > 2$ respectively. We thus choose the value of a_γ and b_γ by following a similar way discussed in Czado et al. (2005). More specifically, we let a_γ be greater than but close to 2 for a significantly large variance. Then b_γ is chosen such that $b_\gamma = (a_\gamma - 1)\mathrm{Var}(\hat{\boldsymbol{\gamma}})$. In this study, we fix $a_\gamma = 2.001$ for both benchmark and simulated death scenarios. Recall that $\boldsymbol{V}_{\boldsymbol{\epsilon}} \sim$ Inv-Wishart$(\nu, \boldsymbol{\Sigma})$ for simulated death scenario $w = 0.01$, we fix $\nu = 12$ for a relatively large variance and most importantly guaranteeing that the MCMC is

convergent for every scenario. At last we let $\boldsymbol{\Sigma} = 8\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ to ensure that the mean of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is equivalent to the true parameters (MME) and $\boldsymbol{\Sigma} = 16\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ for the sensitivity test, that is the mode of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is the true parameters (MMO).

We run the MCMC for the benchmark exposure directly with the initial setting introduced above. For population size $w = 0.01$, we follow the below steps:

- Draw N_1 death scenarios from the deaths assumption

$$D^w(t,x) \sim \text{Pois}(m(\theta_{1,0},t,x)wE_0(t,x)),$$

  denoted as $D_j^w(t,x)$ for each scenario $j = 1,\ldots,N_1$.

- Run MCMC for $D_j^w(t,x)$ for $j = 1,\ldots,N_1$ with N_3 iterations. N_3 is chosen to be sufficiently large to ensure Markov chain converges to its stationary distribution.

- For each j, remove the burn-in period and draw N_2 samples from the posterior distribution.

In practice, we simulated $N_1 = 1000$ deaths scenarios. For each of the deaths scenarios we run MCMC with $N_3 = 100,000$ iterations. We remove the first $50,000$ iterations and keep every 50^{th} iteration from the second $50,000$ iteration, that is we draw $N_2 = 1000$ samples for $\boldsymbol{\theta}$ from the joint posterior distribution.

Empirical study on the samples drawn from the posterior distribution is carried out based on five methods for every parameter we are interested in. Note that more detailed descriptions for the five methods are shown on Page 143

Method 1 Study all of the possible samples as a whole. The sample size for each parameter is N_1N_2.

Method 2 For each deaths scenario, we draw one sample from the posterior distribution. We have in total N_1 samples.

Method 3 For each deaths scenario, the mean of the posterior distribution is calculated. We have in total N_1 sample means.

Method 4 Randomly draw one deaths scenario out of the N_1 and investigate its posterior distribution.

Method 5 Randomly select 100 deaths scenarios and compare their CDF's. One application of this method is that we calculate the cumulative distribution function (CDF) for each of the selected death scenario and plot all the Method4 $\times$100 CDFs in one plot.

In particular, the results of Method 4 and 5 are more attractive than the rest of the methods since we are allowed to investigate how the Poisson randomness affects the "distribution" of the posterior distribution, which drives one of our motivations to run MCMC for more than one death scenario. Understanding the impact of sampling variation on the distribution of the posterior distribution is also meaningful for practical application. It allows a small pension scheme manager, who works out the posterior distribution for his/her underlying exposure, to estimate how much confidence the he/she can have to believe that the posterior distribution is correct due to the large sampling variation.

5.4.5 Projection

In the previous sections, we have shown that the estimates, given the two-stage approach, for the parameters in the projecting model are significantly biased from the true parameters due to the large sampling variation of small population. Such bias greatly increases the variance of mortality projection. One of the advantages of Bayesian approach over the two-stage is that we can draw samples of the pro-

jected mortality rates directly from the so called posterior predictive distribution, conditional on the observed deaths counts.

We start with a brief review of the definition of the posterior predictive distribution. Similarly with Section 5.2, we denote as $\boldsymbol{y}$ the observed data depends on a K dimensional parameter vector $\boldsymbol{\theta}$. Let $\boldsymbol{y}'$ be the future unknown but observable value. Theoretically, the posterior predictive distribution $p(\tilde{\boldsymbol{y}}|\boldsymbol{y})$ is defined as:

$$p(\boldsymbol{y}'|\boldsymbol{y}) = \int p(\boldsymbol{y}', \boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} = \int p(\boldsymbol{y}'|\boldsymbol{\theta}, \boldsymbol{y}) p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} = \int p(\boldsymbol{y}'|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}.$$

The last equation holds as $\boldsymbol{y}'$ is assumed to be conditionally independent of $\boldsymbol{y}$, given $\boldsymbol{\theta}$. It implies that the uncertainty of the posterior predictive distribution consists of the future simulation risk as well as the parameter uncertainty. Obtaining the density $p(\boldsymbol{y}'|\boldsymbol{y})$ analytically by integrating $p(\boldsymbol{y}'|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})$ with respect to all the parameters can be complicated and simulation method is employed. Samples are drawn from $p(\boldsymbol{\theta}|\boldsymbol{y})$ and $p(\boldsymbol{y}'|\boldsymbol{\theta})$ to obtain the predictive distribution.

In this study, the samples drawn from the stationary distribution achieved by the Hamiltonian Monte Carlo can be treated as if they are drawn from the true posterior distribution for $\boldsymbol{\theta}$. Let $\tilde{\boldsymbol{\theta}}$ be the Bayesian estimation for $\boldsymbol{\theta}$. The mortality rates can be projected according to Equation (2.5) (see page 23).

The following steps are performed for each of the simulated death scenario. Let $\tilde{\boldsymbol{\theta}}^{(k)}$ be the k^{th} sample drawn from the posterior distribution for $\boldsymbol{\theta}$, where $k = 1, \ldots, N_2$. Denote as $n' = 1, \ldots, N'$ the number of years projected. Then for each posterior sample k,

- Generate a random sample path for the future values of $\tilde{\boldsymbol{\kappa}}^{(k)}$, by drawing a sample, denoted as $\tilde{\boldsymbol{\kappa}}^{(k)}_{t_{ny}+n'}$, from

$$\tilde{\boldsymbol{\kappa}}^{(k)}_{t_{ny}+n'} \sim MVN(\tilde{\boldsymbol{\kappa}}^{(k)}_{t_{ny}+n'-1} + \tilde{\boldsymbol{\mu}}^{(k)}, \tilde{\boldsymbol{V}}^{(k)}_{\epsilon})$$

- Generate a random sample path for the future values of $\tilde{\boldsymbol{\gamma}}^{(4,k)}$ by drawing a sample, denoted as $\tilde{\gamma}^{(4,k)}_{t_{ny}-x_1+n'}$, from

$$\tilde{\gamma}^{(4,k)}_{t_{ny}-x_1+n'} \sim N(\tilde{\alpha}^{(k)}_{\gamma}\tilde{\gamma}^{(4,k)}_{t_{ny}-x_1+n'-1}, (\tilde{\sigma}^{(k)}_{\gamma})^2)$$

- Construct the future mortality rate $q(t,x)$ with Equation (2.5), given the projected period and cohort effects.

- Repeat the above three steps to generate N_4 sample paths starting from $\tilde{\boldsymbol{\kappa}}^{(k)}_{t_{ny}}$ and $\tilde{\gamma}^{(4,k)}_{t_{ny}-x_1}$ and calculate the projected mortality rates.

5.5 Results of Fitting Benchmark Population with Bayesian Approach

In Section 5.4, we introduced the structure of our model including the time series prior for the latent parameters and the prior distribution for the hyper-parameters. In this section, the posterior distribution of $\boldsymbol{\theta}$ for the benchmark population (males in England and Wales) is investigated for a better insight to the impact of combining the ARIMA model with the Poisson likelihood for fitting large population.

5.5.1 Convergence

Recall that no identifiability constraints are explicitly applied during the updating procedure and uniform prior distributions are used for the joint posterior density. Therefore it is important to check if the MCMC obtained a stationary distribution for every parameter given the HMC algorithm. Figure 5.1 shows the trajectories of the MCMC updating for both the latent (selected years and years of birth) and hyper-parameters of the males in England and Wales during year 1961-2011 and aged 50-89 last birthday (same with the dataset fitted by two-stage approach in

Chapter 3). It suggests that at least for this study, Hamiltonian Monte Carlo is an efficient updating algorithm for the MCMC approaching the stationary posterior distribution. Recall that a non-informative prior is assumed for the joint density $p(\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}})$ while fitting the benchmark exposure given we have no prior knowledge about the drift and co-variance matrix of the big population. The convergence shown in Figure 5.1 implies that for large population the prior distributions for the other parameters are sufficiently informative based on the underlying model structure.

We select the Raftery et al. (1992) method for detecting convergence to the stationary distribution according to the study by Cowles & Carlin (1996) that comparatively review various convergence diagnostic methods. The method is chosen based on the Table 1 in Cowles & Carlin (1996) and our circumstances. The result of the diagnostic shows that for example, only $N_3 = 8220$ and 8324 iterations should be run if we are requiring the cumulative distribution of the 2.5% and 97.5% quantiles respectively be estimated to within ± 0.005 with probability 0.95 for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$. These requirements are reasonable if roughly, we want to calculate the 95% credibility intervals for a particular parameter to have its actual posterior probability between 0.94 and 0.96 (Raftery et al. (1992)). Recall in actual practice, we simulate $N_2 = 400,000$ iterations for gaining the stationary distribution for England and Wales data. The result of the diagnostic carried out for each parameter implies a fast convergence.

5.5.2 Compare with the Two-Stage Approach

In this section, the performances of fitting and forecasting the England and Wales males' mortality dataset given the Bayesian and two-stage approach are compared.

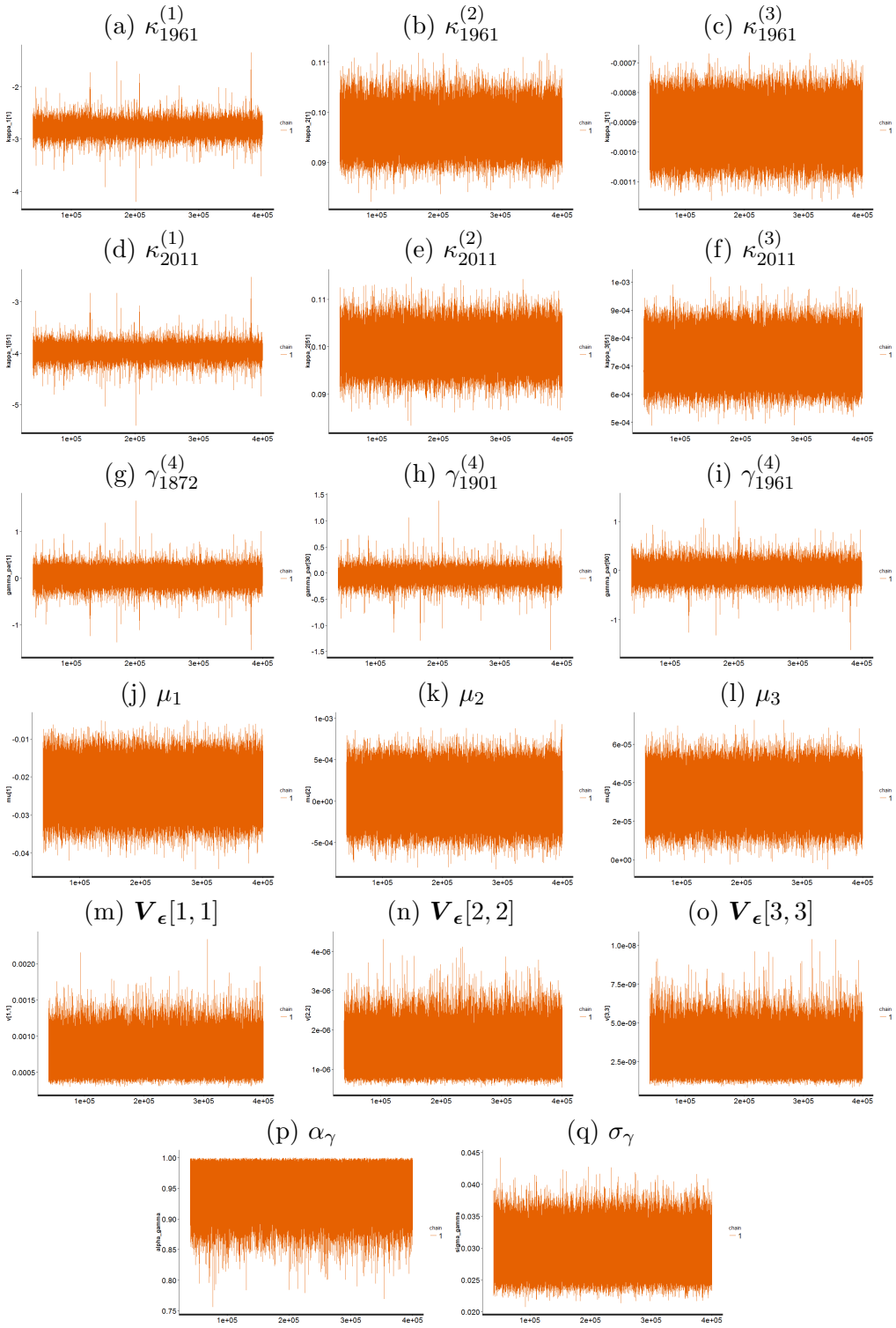


Figure 5.1: Trace plot for the selected parameters obtained by Hamiltonian Monte Carlo, including the period and cohort effect with selected year (1961, 2011) and year of birth (1872, 1901, 1961) respectively, the drift of multi-variate random walk $\boldsymbol{\mu}$, the diagonal of co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, coefficient and standard deviation of AR(1) model α_γ, σ_γ. The benchmark data fitted is the England and Wales data, males during year 1961 to 2011 aged 50 to 89 last birthday.

Bayesian Estimation for $\boldsymbol{\theta}$

Denote as $\tilde{(\cdot)}$ the posterior distribution of the parameters, e.g. $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$ represents the posterior estimation of the volatility of the first period effect $\boldsymbol{\kappa}^{(1)}$ of the England and Wales data. Note that we use the term "posterior distribution" and "posterior estimation" inter-changeably.

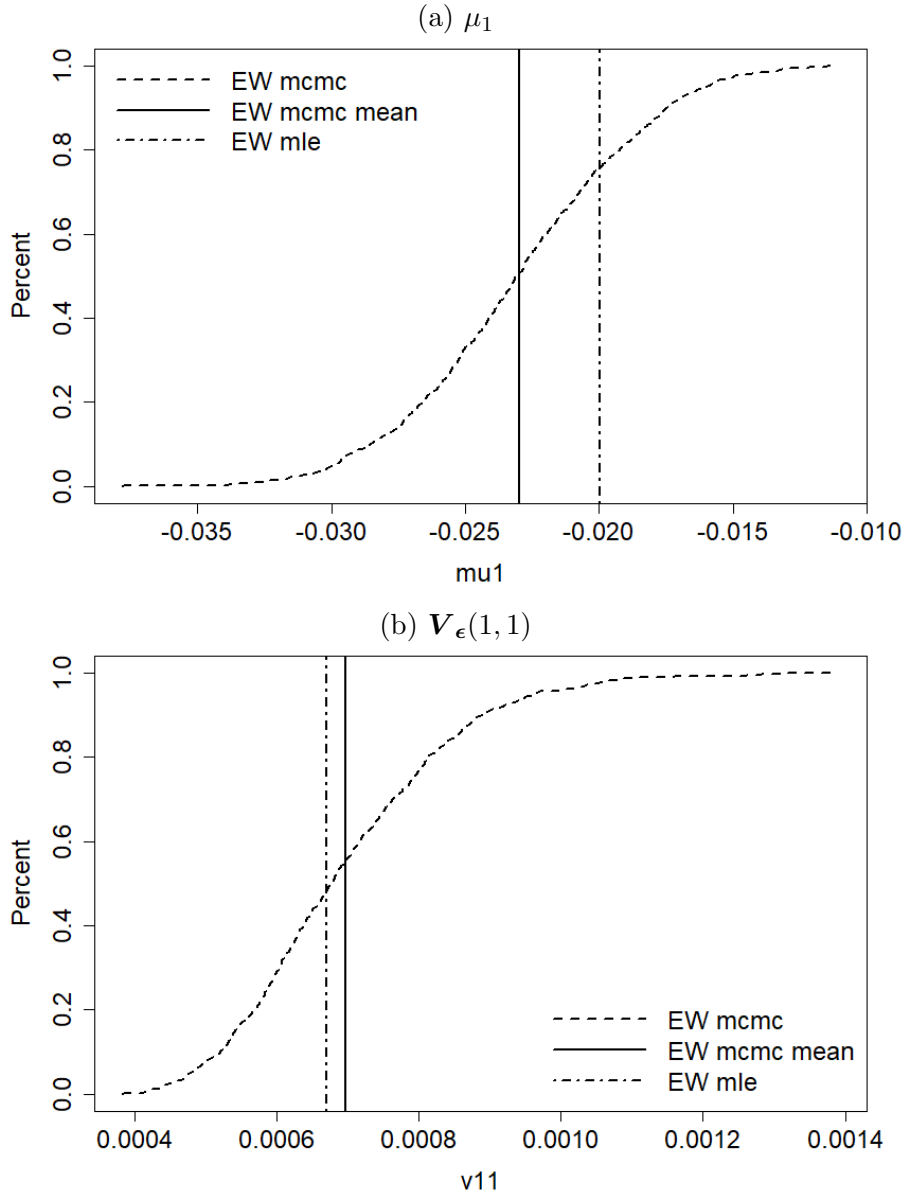


Figure 5.2: CDF: The posterior distribution of $\boldsymbol{\mu}_1$ (lower, dotted CDF) and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ (upper, dotted CDF) for the England and Wales data. The vertical dotted-dashed lines are the respected MLEs and the vertical solid line is the mean of the posterior distribution.

We demonstrate the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$ (dashed CDF) for the Eng-

land and Wales data in Figure 5.2b, which is slightly right-tailed. The respective MLE $\hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{\mathrm{EW}}(1,1)$ (vertical dotted-dashed line, derived according to Equation 3.9) lies approximately on the 50% quantile of the $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$. See Table 5.1 for the characteristic statistics. We therefore conclude that the two methods generate similar estimation for the volatility of the $\boldsymbol{\kappa}^{(1)}$ of the England and Wales data as expected when the population size is large. The Poisson likelihood dominates the joint posterior distribution and the influence of the time series models and the non-informative prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ that lets the data speak for themselves have little impact on the estimation. Therefore the posterior estimation for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ is not significantly different from the respective MLE.

	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$
MCMC Mean	6.97×10^{-4}	13.60×10^{-7}	2.80×10^{-9}
MCMC SD	1.54×10^{-4}	3.25×10^{-7}	0.77×10^{-9}
MLE	6.70×10^{-4}	13.05×10^{-7}	3.30×10^{-9}
	$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_2$	$\boldsymbol{\mu}_3$
MCMC Mean	-2.30×10^{-2}	0.67×10^{-4}	3.34×10^{-5}
MCMC SD	0.42×10^{-2}	1.90×10^{-4}	0.76×10^{-5}
MLE	-2.00×10^{-2}	1.20×10^{-4}	3.46×10^{-5}

Table 5.1: Mean and standard deviation of the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(i,i)$ and $\tilde{\boldsymbol{\mu}}_i^{\mathrm{EW}}$ and the respected MLEs for the England and Wales data, where $i = 1, 2, 3$.

Consistently, visually no obvious smoothness could be observed from the trajectories of the posterior distribution $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ of the England and Wales data (dashed lines) demonstrated in Figure 5.3.

Similar results could be observed for the estimated volatility of $\boldsymbol{\kappa}^{(2)}$ demonstrated in Figure 5.4b, which is also approximately centred around the respected MLE

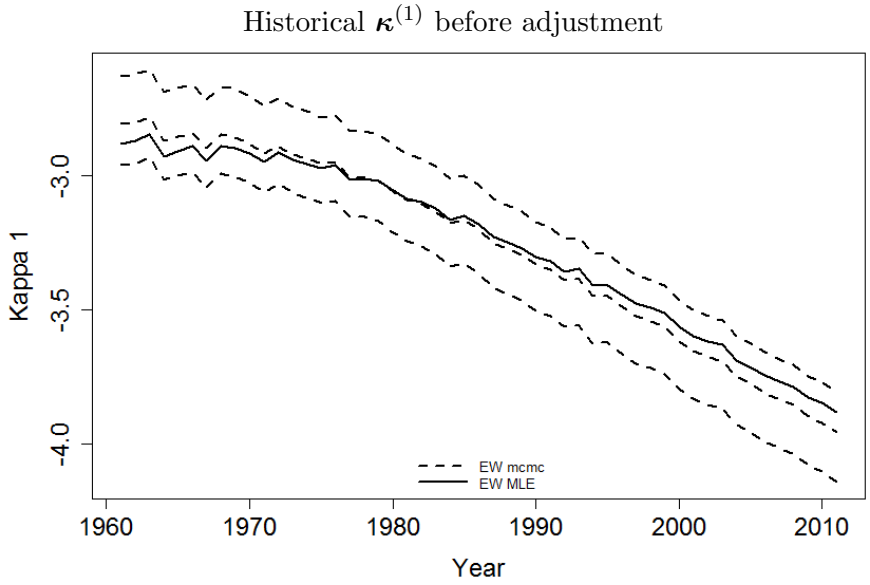


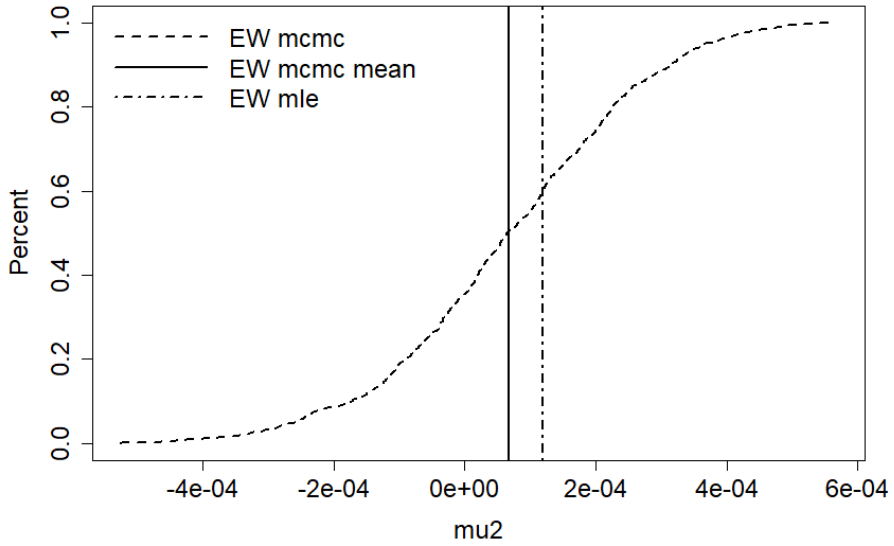
Figure 5.3: Credibility Intervals: The posterior distribution $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ (dashed lines) for the England and Wales males during year 1961-2011 aged 50-89 last birthday. The middle dashed line is the mean of $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$. The upper bound of CI is the 95% quantile of the posterior samples and the lower bound is 5% quantile. The solid line is the respect MLE $\hat{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$.

$\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(2,2)$ (vertical dotted-dashed line) and right tailed. Once again the trajectory of the posterior distribution $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ demonstrated in Figure 5.5 (dashed lines) is visually not more smoothed than that of the respected MLE $\hat{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ (solid line).

On the other hand, the MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(3,3)$ (vertical dotted-dashed line) of the volatility of $\boldsymbol{\kappa}^{(3)}$ for the England and Wales data is on the 80% quantile (roughly) of the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(3,3)$ (dashed CDF) demonstrated in Figure 5.6b, which implies, as may be expected, that the posterior estimation for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ is relatively more sensitive to the time series prior's smoothing influence than for the first and second period effects.

Consistently, the trajectory for $\tilde{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ (dashed lines) demonstrated in Figure 5.7 is slightly more smoothed compared with the respected MLE $\hat{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$. However, we could still conclude that the estimations for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ generated by the two methods are not significantly different from each other, though the MLE is slightly higher than the mean of the posterior estimation. See Figure C.21 in Appendix C.4.

(a) μ_2



(b) $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$

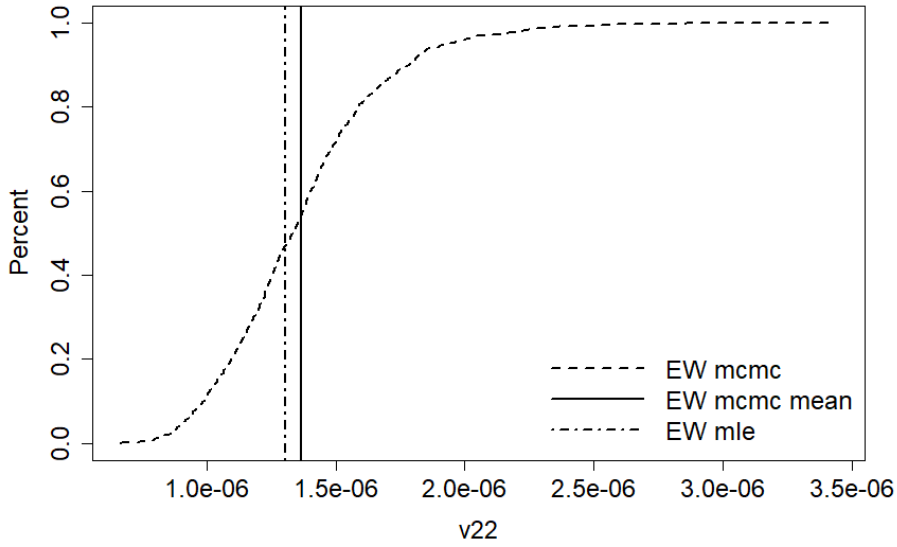


Figure 5.4: CDF: The posterior distribution of $\boldsymbol{\mu}_2$ (lower, dotted CDF) and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ (upper, dotted CDF) for the England and Wales data. The vertical dotted-dashed lines are the respected MLEs and the vertical solid line is the mean of the posterior distribution.

Without knowing the true parameter of the volatility of the England and Wales males, we cannot justify if the Bayesian method provides a much improved estimation for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ of the England and Wales males while in general the impact of the smoothness is not significant when the Poisson likelihood dominates the joint posterior distribution for large population. However, the Bayesian approach does allow us to investigate the parameter uncertainty and how it will affect the mortality projection.

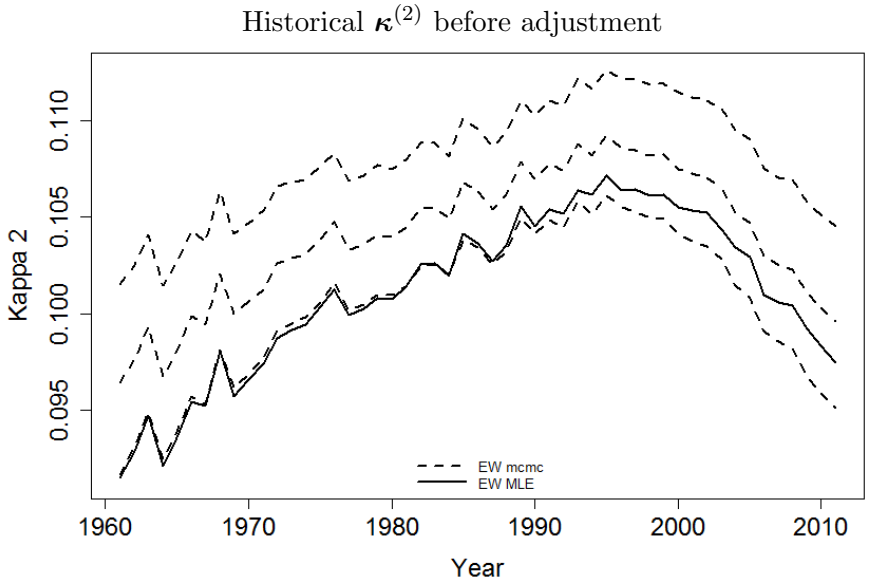
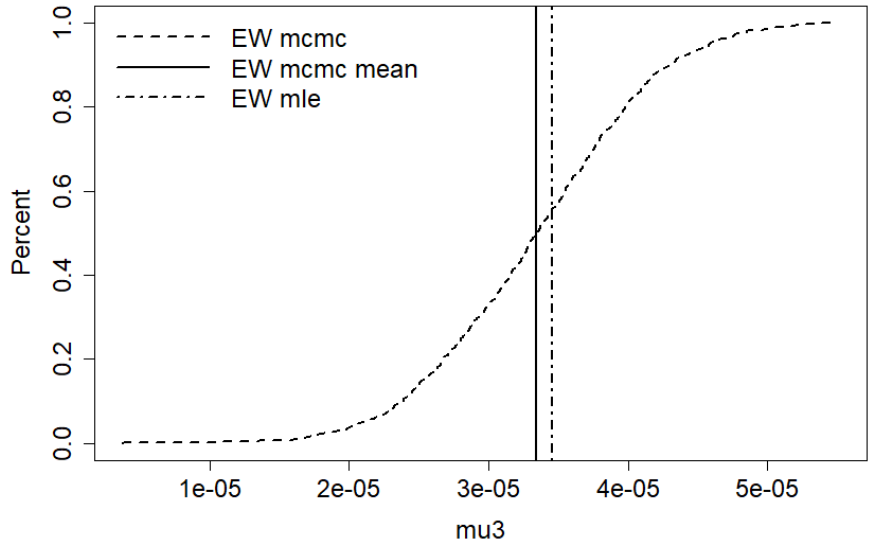


Figure 5.5: Credibility Intervals: The posterior distribution $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ (dashed lines) for the England and Wales males during year 1961-2011 aged 50-89 last birthday. The middle dashed line is the mean of $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$. The upper bound of CI is the 95% quantile of the posterior samples and the lower bound is 5% quantile. The solid line is the respect MLE $\hat{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$.

The posterior distribution $\tilde{\boldsymbol{\mu}}_1^{\mathrm{EW}}$ (dashed CDF) for the drift of the $\boldsymbol{\kappa}^{(1)}$ of the England and Wales data is demonstrated in Figure 5.2a, which has its posterior mean slightly on the right hand side of the respected MLE $\hat{\boldsymbol{\mu}}_1^{\mathrm{EW}}$ (vertical dotted-dashed line, derived according to Equation 3.8). See the characteristic statistics in Table 5.1. Consistently, a clockwise tilt for the trajectory of the posterior estimation $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ from the respected MLE $\hat{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ could be observed in Figure 5.3. Such tilt is a correspond to the anti-clockwise tile that could be observed in Figure 5.8 to the trajectory of the posterior estimation $\tilde{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$ (dashed lines) for the cohort effect of the England and Wales data from the respect MLE $\hat{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$ (solid line). We have discussed that the two-stage approach fails to generate a good fit to the short cohorts for the very early and late years of birth. The low level information involved in the few observations of the short cohorts greatly increases the volatility of the respected years of birth's estimation by pushing the MLEs much far away from the estimates for the other cohorts. On the other hand, the Bayesian approach balances such low ranked information with the posterior distribution. Recall that no identifiability constraints are explicitly applied to the posterior estimation of the

(a) μ_3



(b) $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$

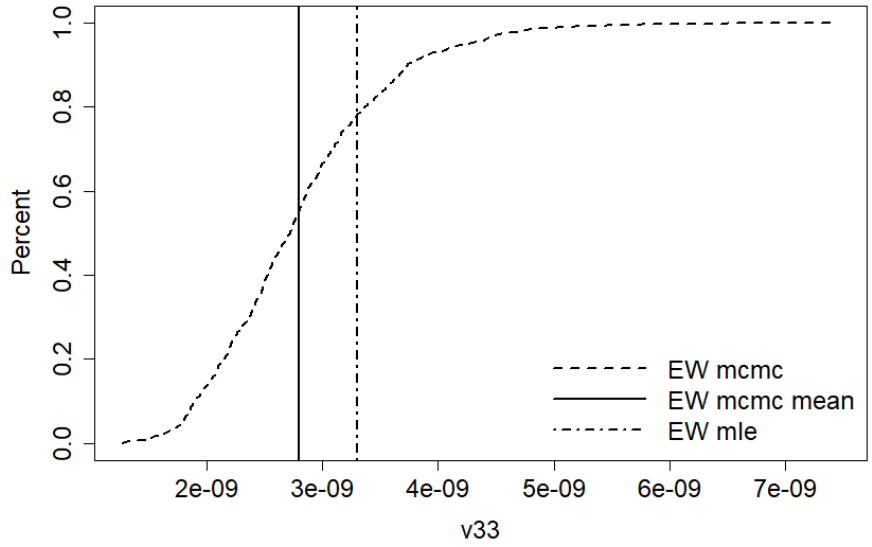


Figure 5.6: CDF: The posterior distribution of $\boldsymbol{\mu}_3$ (lower, dotted CDF) and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ (upper, dotted CDF) for the England and Wales data. The vertical dotted-dashed lines are the respected MLEs and the vertical solid line is the mean of the posterior distribution.

cohort effect. Therefore, the influence of the ARIMA restriction for the cohort effect anti-clockwiselly tilts the posterior estimation $\tilde{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$ such that the estimation for the short cohorts are relatively less volatile, i.e. not significantly higher or lower than the estimates of the other cohorts, compared with the respected MLEs. As a consequence, the trajectory of the $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ is clockwiselly tilted.

We demonstrate the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{\mathrm{EW}}$ (dashed CDF) in Figure 5.4a,

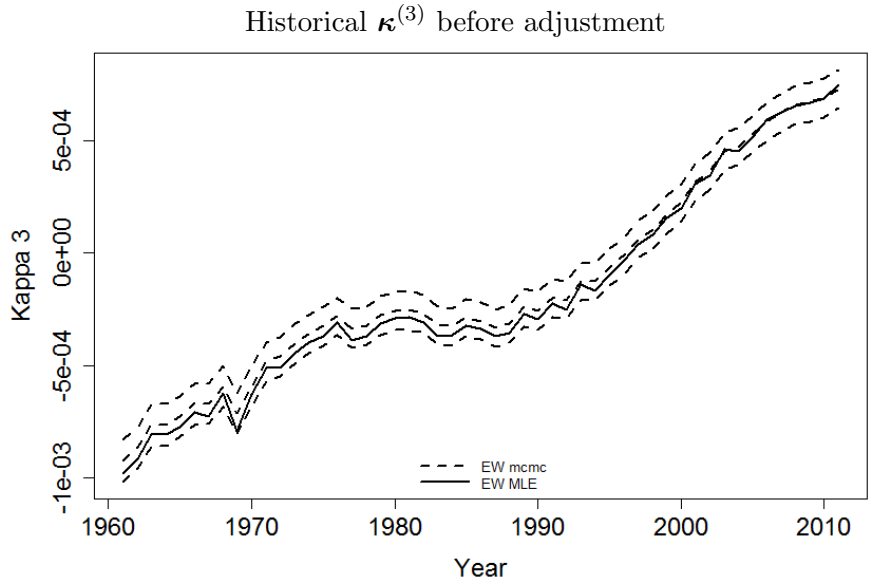


Figure 5.7: Credibility Intervals: The posterior distribution $\tilde{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ (dashed lines) for the England and Wales males during year 1961-2011 aged 50-89 last birthday. The middle dashed line is the mean of $\tilde{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$. The upper bound of CI is the 95% quantile of the posterior samples and the lower bound is 5% quantile. The solid line is the respect MLE $\hat{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$.

where the respected MLE $\hat{\boldsymbol{\mu}}_2^{\mathrm{EW}}$ (vertical dotted-dashed line) is roughly on the centre of the posterior distribution, through slightly higher than the median of $\tilde{\boldsymbol{\mu}}_2^{\mathrm{EW}}$. Consistent result could be observed in Figure 5.5 where the trajectory of the posterior distribution $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ (dashed lines) is in general similar with that of the respected MLE $\hat{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$. The posterior estimation $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ is shifted up from the MLE with the estimations for the years after 1985 (approximately) clockwiselly tilted from the respected MLEs of the same period.

Similar results could be observed for the posterior distribution $\tilde{\boldsymbol{\mu}}_3^{\mathrm{EW}}$ (dashed CDF) demonstrated in Figure 5.6a, who has its posterior median slightly less than the respected MLE $\hat{\boldsymbol{\mu}}_3^{\mathrm{EW}}$ (vertical dotted-dashed line). Consistently, a clockwise tilt could be observed on the posterior distribution $\tilde{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ before year 2000 (approximately) from the respected MLEs $\hat{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ of the same period. The trajectories of the $\tilde{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ and $\hat{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ are similar (despite the tilt) in general and the point estimates for years after 2000 are approximately the same with the respected posterior mean of the $\tilde{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ (central dashed line).

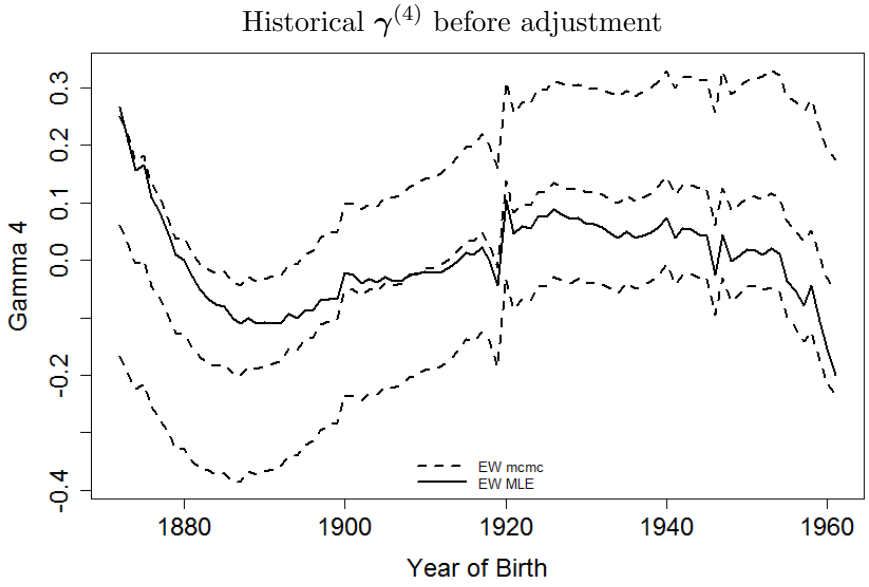


Figure 5.8: Credibility Intervals: The posterior distribution $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ (dashed lines) for the England and Wales males during year 1961-2011 aged 50-89 last birthday. The middle dashed line is the mean of $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$. The upper bound of CI is the 95% quantile of the posterior samples and the lower bound is 5% quantile. The solid line is the respect MLE $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$.

The posterior distribution $\tilde{\alpha}_\gamma^{\text{EW}}$ (upper) and $\tilde{\sigma}_\gamma^{\text{EW}}$ (lower) (solid CDFs) are demonstrated in Figure 5.9. The point estimates (dashed line) $\hat{\alpha}_\gamma^{\text{EW}}$ and $\hat{\sigma}_\gamma^{\text{EW}}$ are calculated by fitting the AR(1) model in Equation 5.37 to the MLE $\hat{\boldsymbol{\gamma}}^{(4)}$. The dotted lines are the mean of the MCMC samples. See the characteristic statistics in Table 5.2.

As may be expected, the posterior distribution $\tilde{\alpha}_\gamma^{\text{EW}}$ is left skewed with its posterior mean (vertical dotted line) slightly less than the respected MLE $\hat{\alpha}_\gamma^{\text{EW}}$ (vertical dashed line), implying the posterior estimation for the cohort effect is relatively less like the random walk process compared with the respected MLE. This is because we tend to push the estimation of α_γ away from one by using a beta-type prior distribution for α_γ. However, the posterior distribution $\tilde{\alpha}_\gamma^{\text{EW}}$ itself is still relatively quite close to one, more specifically over 90% samples drawn from the posterior distribution are greater than 0.9, which implies the estimated cohort effect is still far less like a standard AR(1) process. Consistently, the trajectory of the posterior distribution $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ in Figure 5.8 is approximately the same with the pattern of the respected MLE $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$ (despite the tilt). This is once again because the Poisson likelihood

		α_γ	σ_γ
MCMC	Mean	0.9578	0.0293
	SD	0.0252	0.0024
MCMC	Quantile		
	2.5%	0.8992	0.0251
	50%	0.9623	0.0291
	97.5%	0.9931	0.0343
MLE		0.9827	0.0293

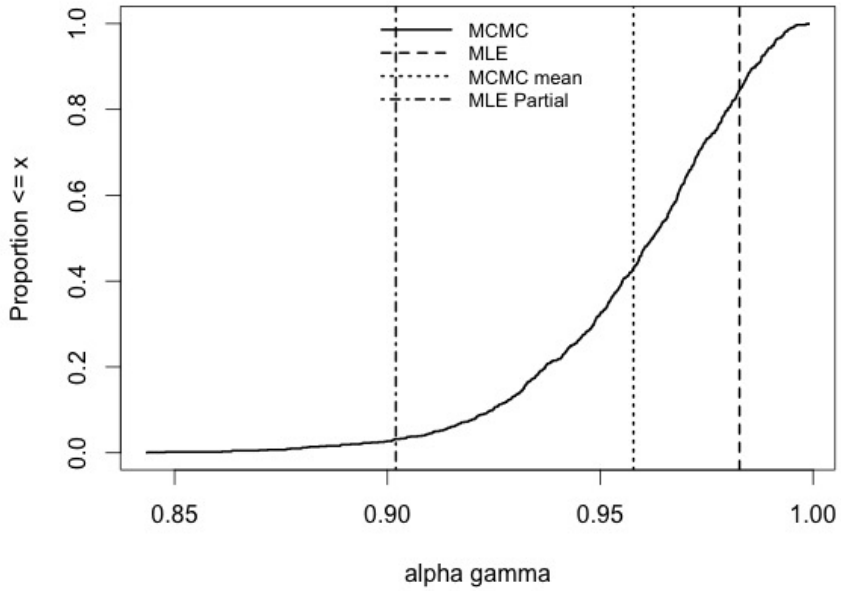
Table 5.2: Characteristic statistics of the posterior distribution $\tilde{\alpha}_\gamma^{\text{EW}}$ and $\tilde{\sigma}_\gamma^{\text{EW}}$ as well as the respected MLEs.

dominates the joint posterior distribution and the influence of the ARIMA model and the prior distribution for α_γ have relatively little impact on the trajectory of the $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ and the estimation $\tilde{\alpha}_\gamma^{\text{EW}}$. Therefore the restriction from the ARIMA model is not strong enough to make the $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ look more like the proposed zero mean-reverting model. Note that this does not conflict with our previous conclusion that the Bayesian approach provides a better fit for the short cohorts than the two-stage approach does.

The posterior distribution $\tilde{\sigma}_\gamma^{\text{EW}}$ is relatively symmetric around the respected MLE $\hat{\sigma}_\gamma^{\text{EW}}$. We therefore conclude that the two methods generate approximately the same estimation for the volatility of the cohort effect, more specifically the MLE $\hat{\sigma}_\gamma^{\text{EW}}$ is approximately equivalent to the mean of the posterior distribution $\tilde{\sigma}_\gamma^{\text{EW}}$. Consistently, the trajectory of $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ is as volatile as that of the respected MLE $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$ and no obvious smoothness could visually be observed for the $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$.

Recall that the point estimates for short cohorts at the beginning and the end

(a) α_γ



(b) σ_γ

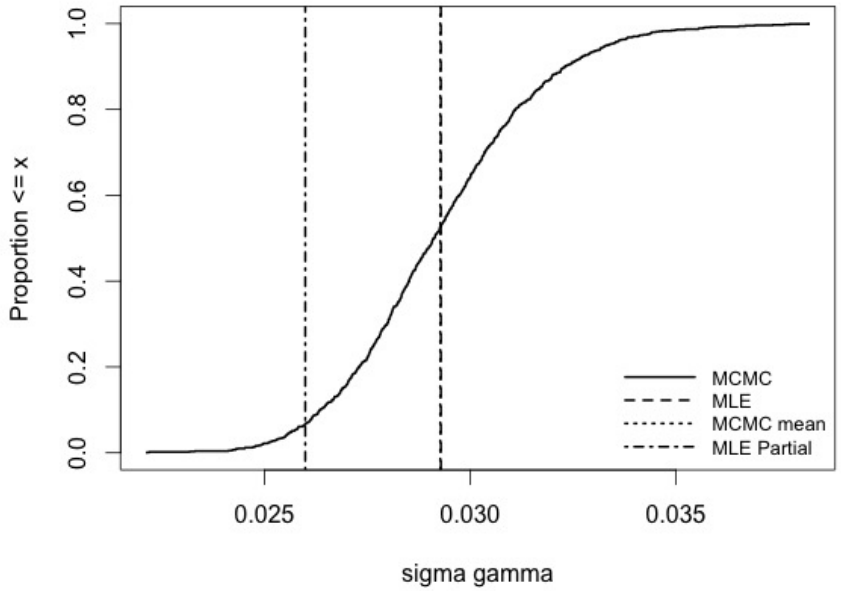


Figure 5.9: CDF of Bayesian estimates for α_γ (left plot) and σ_γ (right plot). The vertical dashed line is the point estimates for the parameters calculated by fitting (5.37) to the ML estimates $\hat{\boldsymbol{\gamma}}^{(4)}$. The dotted line is the mean of the posterior samples. The dot-dashed line is the point estimates for α_γ and σ_γ derived by fitting (5.37) to $\hat{\boldsymbol{\gamma}}^{(4)}$ with the first and last five cohorts removed. Note the MCMC mean and MLE for the σ_γ are overlapped in plot (b).

of the cohort process is not reliable as they are derived according to the low level information provided by only few observations. Although we balance these cohorts with a posterior distribution, it is not appropriate to compare the Bayesian estimation of α_γ and σ_γ directly with point estimates given the entire cohort process.

Alternatively, we remove five cohorts equally from the beginning and the end of the process. Then we work out the estimation for α_γ and σ_γ, denoted as $\hat{\alpha}^*_\gamma$ and $\hat{\sigma}^*_\gamma$ respectively, by fitting the $AR(1)$ model to the estimates of the rest of the cohorts. See the dot-dashed lines in Figure 5.9a-5.9b. Unsurprisingly both the new point estimates for α_γ and σ_γ are lower than the respected MCMC mean. The reduction is consistent with Figure 3.8b and 3.8c, which show that the coefficient and the variance of the time series process is reduced by removing 10 cohorts in total.

Fitted Death Rates

In this section, we investigate the fitted mortality rates given our Bayesian model. We start with the comparison of the fitted death rates for selected ages with respect to years given the ML estimates $\hat{\theta}_1^{\text{EW}}$ and the Bayesian estimation $\tilde{\theta}_1^{\text{EW}}$ of θ_1, denoted as $m(\hat{\theta}_1^{\text{EW}}, t, x)$ and $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ respectively. Denote as $m(\tilde{\theta}_1^{(k),\text{EW}}, t, x)$ the k^{th} observation of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ for $k = 1, \ldots, N_2$. The fitted rates are calculated according to Equation (2.4) and (2.5), given $\hat{\theta}_1^{\text{EW}}$ and $\tilde{\theta}_1^{(k),\text{EW}}$ respectively.

In Table 5.3, we calculated the MCMC mean, median and credibility intervals (CI) of fitted empirical death rates $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ for the generation aged 50 in year 1972 and compared with the crude rates as well as the fitted rates $m(\hat{\theta}_1^{\text{EW}}, t, x)$ given the MLE $\hat{\theta}_1^{\text{EW}}$. It is worth to notice that for this particular cohort year $c = 1922$, there is no significant difference between $m(\hat{\theta}_1^{\text{EW}}, t, x)$ and the mean of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$, that is both methods provide a similar level of fit for this cohort year.

We plotted the log-scaled fitted deaths rates given the MLE (dashed line), $m(\hat{\theta}_1^{\text{EW}}, t, x)$, the credibility intervals of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ (dotted line) and the crude rates (solid line) with respect to the years at age 50, 60, 80 and 89 in Figure 5.10. The upper bound of the credibility interval is 95% quantile of the $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ and the lower bound is 5% quantile. Error bars with widths equivalent to 1.96 times of the

Age	Year	Crude	Mean	Median	90% CI	MLE
50	1972	0.00748	0.00738	0.00738	[0.00729,0.00747]	0.00737
51	1973	0.00802	0.00812	0.00812	[0.00803,0.00821]	0.00811
52	1974	0.00916	0.00899	0.00899	[0.00890,0.00908]	0.00900
53	1975	0.00989	0.00983	0.00983	[0.00974,0.00992]	0.00981
54	1976	0.01097	0.01098	0.01098	[0.01089,0.01107]	0.01101
55	1977	0.01190	0.01181	0.01181	[0.01172,0.01190]	0.01179
56	1978	0.01296	0.01316	0.01316	[0.01306,0.01325]	0.01316
57	1979	0.01451	0.01446	0.01446	[0.01437,0.01456]	0.01448
58	1980	0.01582	0.01554	0.01554	[0.01544,0.01564]	0.01556
59	1981	0.01719	0.01670	0.01670	[0.01660,0.01680]	0.01671
60	1982	0.01851	0.01815	0.01815	[0.01804,0.01826]	0.01814
61	1983	0.01972	0.01980	0.01980	[0.01969,0.01991]	0.01980
62	1984	0.02150	0.02121	0.02121	[0.02109,0.02133]	0.02122
63	1985	0.02354	0.02356	0.02356	[0.02343,0.02369]	0.02356
64	1986	0.02498	0.02548	0.02548	[0.02534,0.02561]	0.02549
65	1987	0.02714	0.02723	0.02723	[0.02708,0.02738]	0.02726
66	1988	0.02969	0.02949	0.02949	[0.02933,0.02965]	0.02953
67	1989	0.03201	0.03162	0.03162	[0.03144,0.03179]	0.03157
68	1990	0.03400	0.03410	0.03410	[0.03392,0.03428]	0.03414
69	1991	0.03667	0.03688	0.03688	[0.03668,0.03708]	0.03686
70	1992	0.03935	0.03946	0.03946	[0.03925,0.03967]	0.03950
71	1993	0.04433	0.04371	0.04371	[0.04348,0.04394]	0.04368
72	1994	0.04449	0.04557	0.04557	[0.04532,0.04581]	0.04558
73	1995	0.05029	0.05042	0.05042	[0.05016,0.05068]	0.05044
74	1996	0.05345	0.05346	0.05346	[0.05319,0.05374]	0.05344
75	1997	0.05635	0.05707	0.05707	[0.05678,0.05736]	0.05703
76	1998	0.06146	0.06189	0.06189	[0.06157,0.06220]	0.06187
77	1999	0.06667	0.06716	0.06716	[0.06682,0.06750]	0.06712
78	2000	0.07028	0.07028	0.07028	[0.06993,0.07064]	0.07027
79	2001	0.07454	0.07516	0.07516	[0.07478,0.07554]	0.07510
80	2002	0.08159	0.08186	0.08186	[0.08145,0.08228]	0.08186
81	2003	0.08951	0.08931	0.08931	[0.08885,0.08977]	0.08922
82	2004	0.09282	0.09329	0.09329	[0.09278,0.09379]	0.09325
83	2005	0.10022	0.10108	0.10108	[0.10051,0.10165]	0.10110
84	2006	0.10790	0.10773	0.10773	[0.10708,0.10838]	0.10768
85	2007	0.11715	0.11751	0.11751	[0.11676,0.11826]	0.11754
86	2008	0.13191	0.12938	0.12938	[0.12851,0.13025]	0.12944
87	2009	0.14058	0.13731	0.13731	[0.13631,0.13832]	0.13737
88	2010	0.14784	0.14971	0.14971	[0.14851,0.15091]	0.14973
89	2011	0.16264	0.16181	0.16181	[0.16032,0.16331]	0.16225

Table 5.3: Empirical death rates for generation aged 50 in year 1972, including the crude rates, mean, median, 90% credibility intervals of MCMC and the MLE. Note that the upper and lower bound of the sixth column is the 95% and 5% quantile of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ respectively.

standard error[12] are added to the crude rates.

It is clear that both MCMC and the MLE provide a good fit to the crude rates with mid-ages (e.g. age 60, 80), which satisfies our expectation. However both methods failed to catch some of the extra variation involved in higher (age 89) and younger (age 50) ages due to the smaller sized deaths counts compared with other ages. By investigating the CI of the MCMC and the Poisson error bar for the crude rates at these two ages (50 and 89), we notice that although they overlap each other for most of the years, there are some years when the MLE and CI of MCMC stay outside of the error bar which suggests a poor fit, for example in year 2000 for age 50 with the corresponding cohort year 1950. This result is consistent with Figure 5.8, which shows there is significant variation to the years around cohort 1950 given both the MLE and MCMC methods[13].

In particular, we find that the mean of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ for all the selected ages are almost identical to the fitted rates $m(\hat{\theta}_1^{\text{EW}}, t, x)$, given the empirical estimates $\hat{\theta}_1^{\text{EW}}$. The smoothness incurred by adding the random walk and ARIMA models in our Bayesian approach for the benchmark population is not significant and hard to identify, which implies that the Bayesian estimations of the latent parameters, given a large exposure, is still produced that look like the Poisson model. This supports our expectation that the impact of time series models is little when the population is large and the Poisson likelihood dominates the posterior distribution.

However, we can still have an insight on the impact of smoothness on the trajectories of the death rates by calculating the following ratio: the empirical fitted rates given the latent parameters' MLE $\hat{\theta}_1^{\text{EW}}$ over the mean of the fitted rates given

[12]The standard error of the Poisson assumption for the death rates can be calculated by taking the square root of $\frac{\text{crude rate}}{\text{exposure}}$.

[13]Recall a Poisson log-normal CBD model could be used to fit these ages by allowing for over-dispersion.

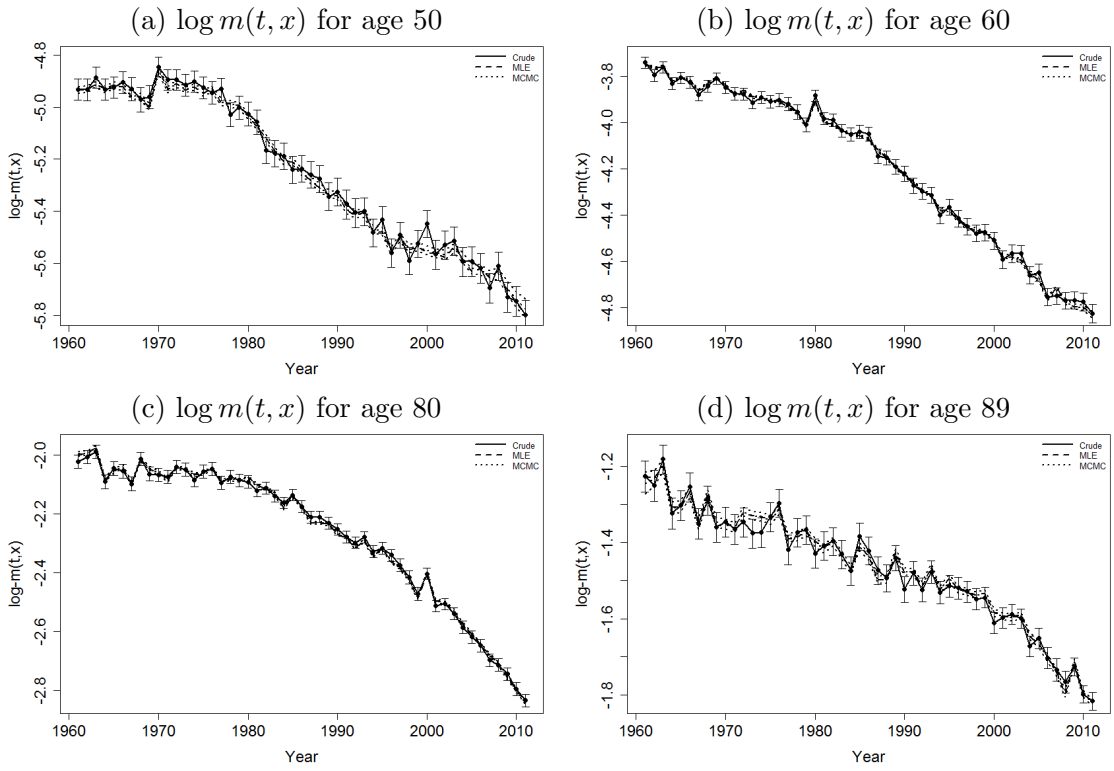


Figure 5.10: Compare the deaths rates between: the fitted deaths rates given the empirical estimates with maximum likelihood method; the 90% confidence interval of fitted deaths rates given samples from posterior distribution (or the MCMC); crude death rates with Poisson error bar, with respected to year range in the dataset for selected ages (50, 60, 80 89). The error bar is set to 1.96 times of the standard error $(\sqrt{\frac{m(\hat{\theta}^{\mathrm{EW}},t,x)}{E^{\mathrm{EW}}(t,x)}})$.

the Bayesian estimation $\tilde{\theta}_1^{\mathrm{EW}}$, more specifically:

$$r(t,x) = \frac{m(\hat{\theta}_1^{\mathrm{EW}}, t, x)}{\frac{1}{N_2}\sum_{k=1}^{N_2} m(\tilde{\theta}_1^{(k),\mathrm{EW}}, t, x)},$$

where $\tilde{\theta}_1^{(k),\mathrm{EW}}$ is the k^{th} observation of $\tilde{\theta}_1^{\mathrm{EW}}$ and the denominator is the posterior mean of the fitted rates.

We plotted the heat plot of $r(t,x)$ with respect to year and age in Figure 5.11 and the plots of $r(t,x)$ with respect to period, age and cohort year are shown in Figure 5.12. Figure 5.12a shows the values of $r(t,x)$ with respect to year. Other than few extreme values at year 1960 and 2010, the rates roughly fluctuate around 1, which implies the difference is negligible between the two estimations. Figure 5.11a is the heat plot of $r(t,x)$ with respect to both year and age, where darker dots

represents rates higher valued. This plot shows the details of, given a particular calendar year t, how the ratios shown in Figure 5.12a are distributed with respect to age x. In general the ratios are not randomly distributed and there are obvious horizontal patterns for the ratio within the age ranges $50-70$ (mainly in dark grey) and $70-80$ (mainly in light grey) through all the calendar years. For example, there is few colour shifts within the pattern of the ratios during age $70-80$ and the dark red dots in Figure 5.11b show that the ratios are in general less than 1 and greater than 0.9975. Further there is a diagonal line with continuous red dots starting from year 1978 to 1995, followed by continuous green dots (greater than 1 and less than 1.0025). All these patterns support the argument that the MCMC is smoothing any variety on the period and cohort process.

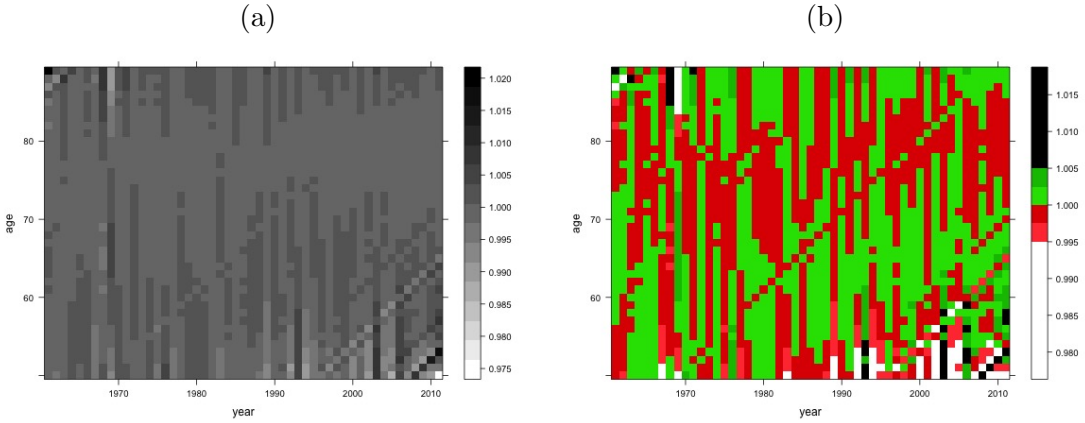


Figure 5.11: The heat plot of the ratio $r(t,x)$: historical fitted $m(t,x)$ given the MLE over the posterior mean, for all the observed ages and years. In plot (b), the red dots represents the ratio less than one but greater than 0.995; the green and black dots are the ratios on $[1, 1.005]$ and greater than 1.005 respectively. Ratios less than 0.995 are coloured white.

In Figure 5.12a, we can see that the ratio $r(t,x)$ is much wider distributed in certain years than the other with respect to the ages. In particular, $r(t,x)$ is right skewed in year 1968 and left tailed in year 1969. It is consistent with Figure 5.11b where most of the vertical dots for year 1968 are red coloured however, with black dots ($r(1968,x) > 1.005$) for ages over 85. Similarly in year 1969, the ratio $r(1969,x)$ is greater than 1 and less than 1.005 during age range 66 to 78, with white dots $r(1969,x) < 0.995$ for ages over 84. Figure 5.12d is the $r(t,x)$ with respect to the

year of birth. The green dots are the corresponding ratios with respect to cohort for year 1969 and the reds are for year 1968. We can see that the ratios are widely distributed for both periods with respect to the corresponding cohorts, which is consistent with Figure 5.3, 5.5 and 5.7 where the central MCMC estimations for the period effects at years 1968 and 1969 are greatly higher than the relative MLE. Thus question arises if one method fits better for a particular year through all the ages and we need to check the fit of the death rats to the historical crude rates for the two methods by calculating the standardized residuals.

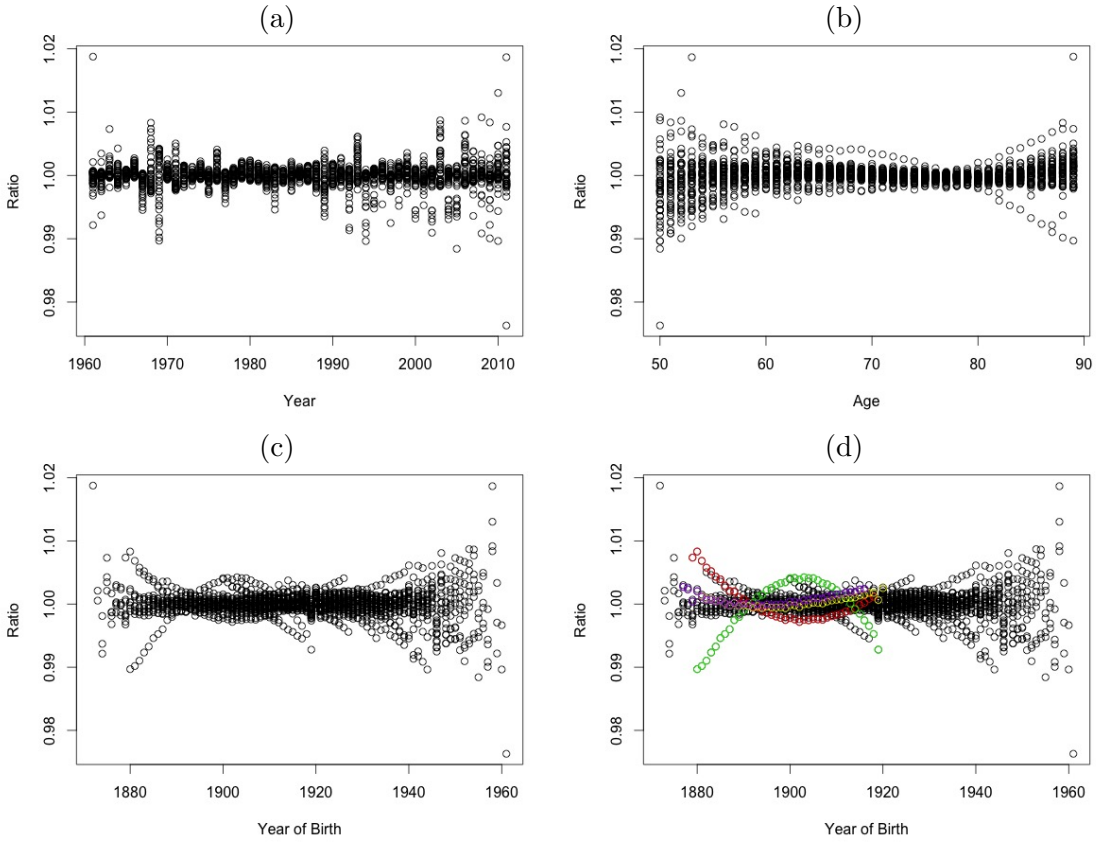


Figure 5.12: The ratio of historical fitted $m(t, x)$ of MLE over the posterior mean with respect to year, age and cohort.

In general, the residual is defined as the difference between the fitted and the observed values. We calculate the standardised residuals with the following formula:

$$z(t,x) = \frac{D(t,x) - E(t,x)\hat{m}(t,x)}{\sqrt{E(t,x)\hat{m}(t,x)}},$$

where $\hat{m}(t,x)$ can be either the fitted death rates given the MLE, $m(\hat{\theta}_1^{\text{EW}}, t, x)$, or the mean of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$. The embedded assumption for the deaths is that the death counts are independent Poisson random variables for each year and age. Thus the ideal method is expected to generate randomly distributed $i.i.d$ standard normal residuals. See Figure 5.13 for the heat plot of the residuals for the MCMC mean

(a) Residuals for MCMC mean (b) Residuals for ML estimates

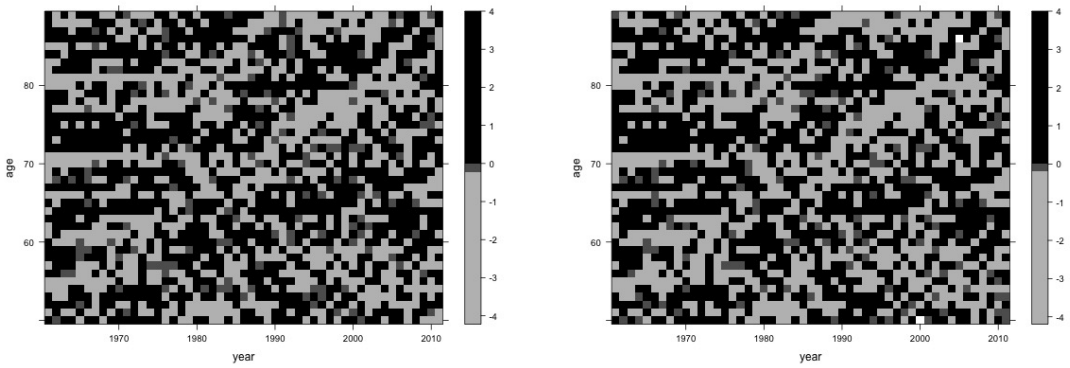


Figure 5.13: The standardized residuals of the MCMC mean (left) and ML method (right). The black dots represents a positive residuals and the grey ones are for negative residuals.

(left) and MLE (right). The residuals generated by both methods are reasonably randomly distributed with similar pattern, which suggests that Bayesian estimation fits the benchmark exposure as good as the MLE and both methods provide a good fit to our underlying dataset with the M7 model. This is consistent with the empirical work by Cairns et al. (2009).

Further, the residuals with respect to age, year and cohort are presented in Figure 5.14. The red dots represent the residuals of MLE method and the blacks are for the MCMC mean. Figure 5.14a-5.14c show that in general the red and black dots are distributed pair-wisely with no significant evidence that one method generate particularly larger residuals at certain years, ages or cohorts, even for years 1968 and 1989 (vertical lines in Figure 5.14b). This supports the argument that the MCMC mean fits the crude rates as good as the ML method. We plotted a scatter plot of the residuals of the MCMC mean vs. the ML method in Figure 5.14d, which indicates that there is a strong linear correlation between the residuals generated by

the two methods.

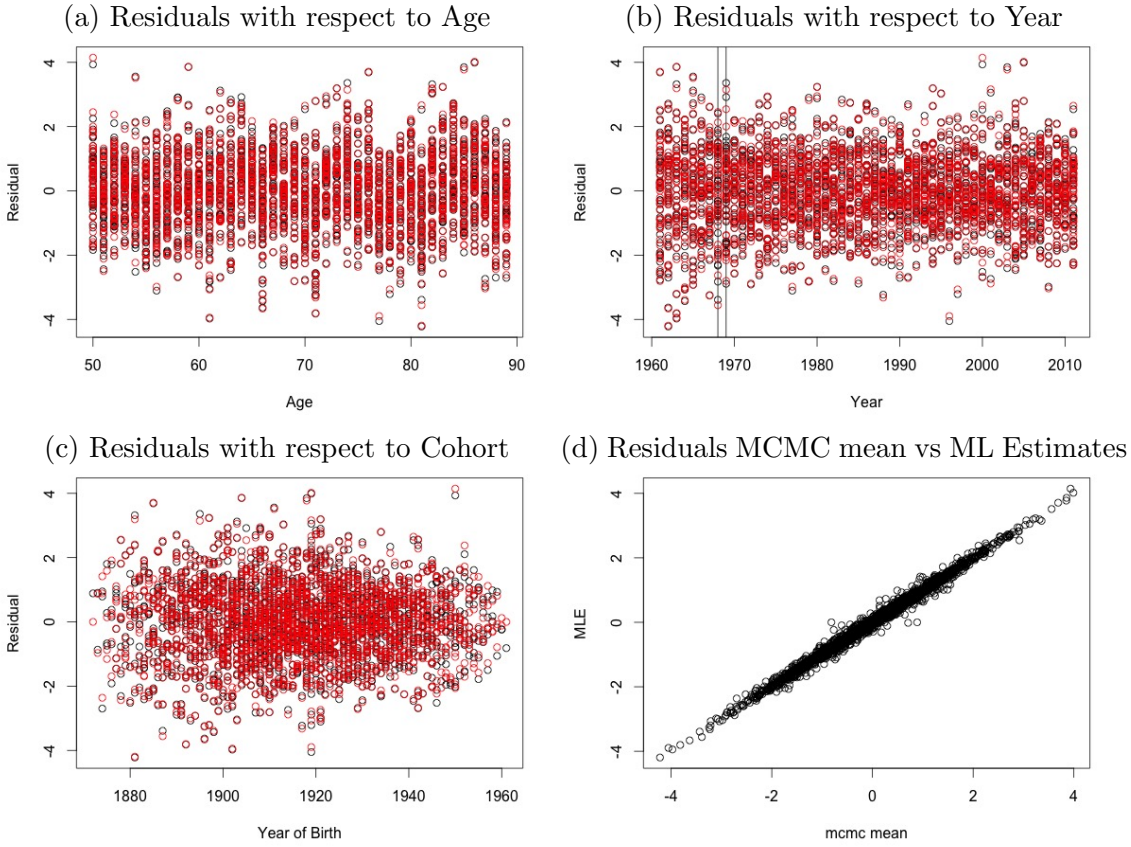


Figure 5.14: Compare the residuals of MCMC mean (black) and ML estimates (red) with respect to age (a), year (b), cohort (c). Figure (d) shows the correlation between the MCMC and ML residuals.

We can thus conclude that for our benchmark exposure there is not a particular year, age or cohort when one method provides significantly better fit to the crude death rates than the other and the residual pattern is almost the same given the two methods. Again this result satisfies our expectation since for large exposure the Poisson likelihood dominants the posterior distribution and the likelihood of the ARIMA models with impact little by smoothing any variety on period and cohort effects.

Projection

Denote as $\boldsymbol{m}'$ the future unknown but observable death rates. Recall that the posterior predictive distribution of $\boldsymbol{m}'$ conditional on the observed rates $\boldsymbol{m}$, $p(\boldsymbol{m}'|\boldsymbol{m})$ is defined as:

$$p(\boldsymbol{m}'|\boldsymbol{m}) = \int p(\boldsymbol{m}'|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{m})d\boldsymbol{\theta}.$$

Given the k^{th} historical posterior sample $\tilde{\boldsymbol{\theta}}_1^{(k)}$ for the latent parameters, we repeat the steps described in Section 5.4.5 to project $N_4 = 100$ sample paths for the period and cohort effects beginning with $\tilde{\boldsymbol{\kappa}}_{t_{n_y}}^{(k)}$ and $\tilde{\gamma}_{t_{n_y}-x_1}^{(4,k)}$ respectively for each $k = 1, \ldots, N_2$. Denote as $\tilde{\boldsymbol{\theta}}_1'$ the projected latent parameters with the k^{th} observation $\tilde{\boldsymbol{\theta}}_1^{\prime(k,l)}$ for the sample path $l = 1, \ldots, N_4$. Construct the projected sample death rates $\boldsymbol{m}^{\prime(k,l)}|\tilde{\boldsymbol{\theta}}_1^{\prime(k.l)}$ for the posterior predictive death rates $\boldsymbol{m}'|\tilde{\boldsymbol{\theta}}_1'$.

We project both the period and cohort effects fifty years forward. The prediction intervals for $\tilde{\boldsymbol{\theta}}_1^{\prime\text{EW}}$ (dashed lines) are plotted in Figure 5.15 compared with the prediction intervals of projection given the ML method (solid lines), denoted as $\hat{\boldsymbol{\theta}}_1^{\prime\text{EW}}$. The method for projecting the latent parameters given the two-stage approach is discussed in Section 3.3. Note that at this stage the MLEs of period effects are projected without allowance of the parameter uncertainty and in Figure 5.15d the projection for the cohort effect starts from the very last cohort year $c = t_{n_y} - x_1$. The point estimates of the parameters in the projecting model are derived by fitting $AR(1)$ to the entire $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$ without removing any short cohorts. The upper bound of the intervals is the 95% quantile of the samples and the lower bound is 5% quantile.

The variance of the MCMC projection for $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$ are much higher than the historical Bayesian estimates, due to the additional normal simulation randomness while simulating the sample paths for the latent parameters. We expect that the predictive intervals of the Bayesian projection are wider than the ML method since the uncertainty of $\tilde{\boldsymbol{\theta}}_1'$ includes both the simulation risk and the parameter uncer-

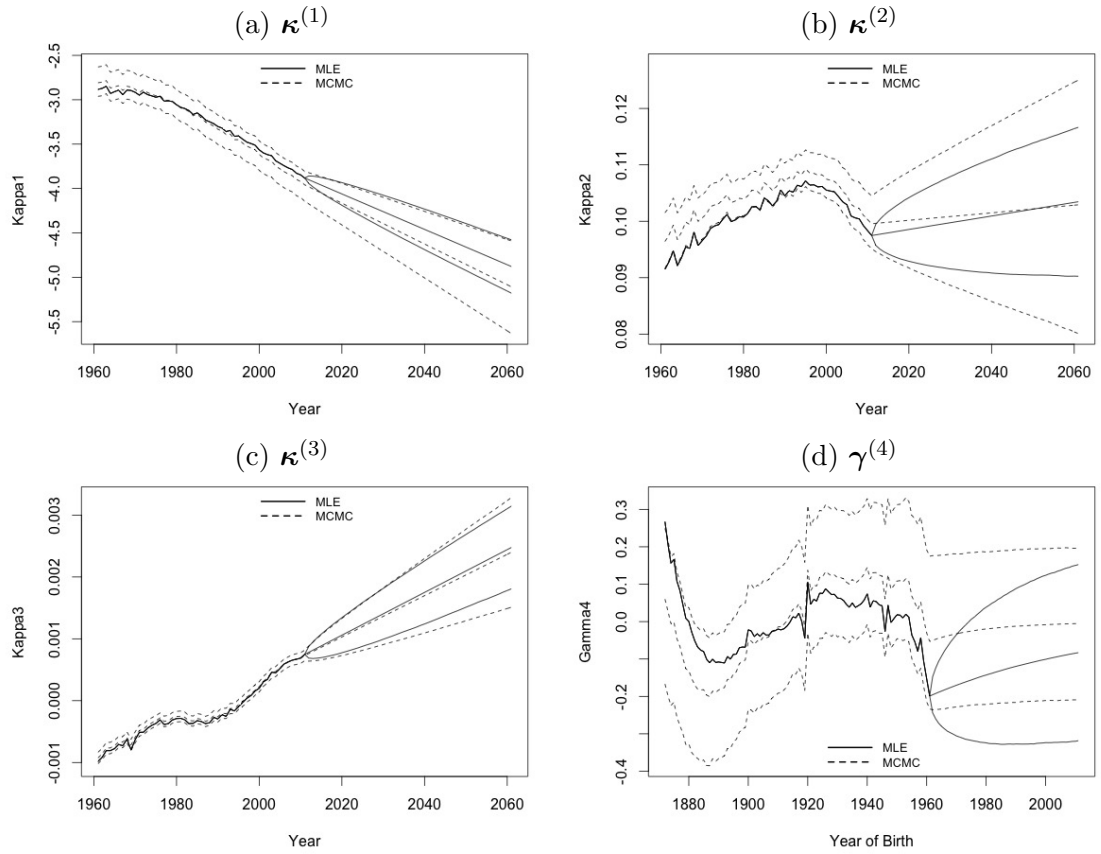


Figure 5.15: Comparison between the credibility intervals (dashed lines) and the confidence intervals (solid lines) for projecting the period and cohort effects for the males in England and Wales, given the Bayesian and MLE approach respectively. The upper bound of the interval is 95% quantile of the projected samples and the lower bound is 5% quantile. Both the period and cohort effects are projected fifty years forward.

tainty. As we expected, Figure 5.15a-5.15c show that the Bayesian projection for the period effects has higher variance than the ML method and the difference of the variance given the two methods increases as the year increases. As for the cohort effect $\boldsymbol{\gamma}^{(4)}$, the predictive interval of $\tilde{\boldsymbol{\theta}}_1'$ is wider than $\hat{\boldsymbol{\theta}}_1'$ and the beginning of the projection but the difference slumps rapidly for longer term projection. We can see that the width of the predictive intervals is approximately the same at the last year of projection. This result is consistent with Figure 5.9b, which shows that by fitting the $AR(1)$ model to all the historical ML estimates including the short cohorts the point estimate of the variance $\hat{\sigma}_\gamma^{\mathrm{EW}}$ is approximately the same with the mean of posterior samples, due to the Poisson likelihood's over-fit to the short cohorts. Thus it makes more sense to project the point estimates of $\boldsymbol{\gamma}$ given $\hat{\alpha}_\gamma^*$ and $\hat{\sigma}_\gamma^*$ derived by

fitting the $AR(1)$ to the cohorts with the first and last five cohorts removed, as we discussed previously. Note that in this case, $\hat{\gamma}^{(4)}_{t_{ny}-x_1-5}$ is the last empirical point estimate of the process and the projection for $\boldsymbol{\gamma}$ starts from $\gamma^{(4)}_{t_{ny}-x_1-4}$, see Figure 5.16. The Bayesian estimation $\tilde{\boldsymbol{\gamma}}^{(4)}$ is projected normally since we balance the impact of short cohorts with the ARIMA likelihood.

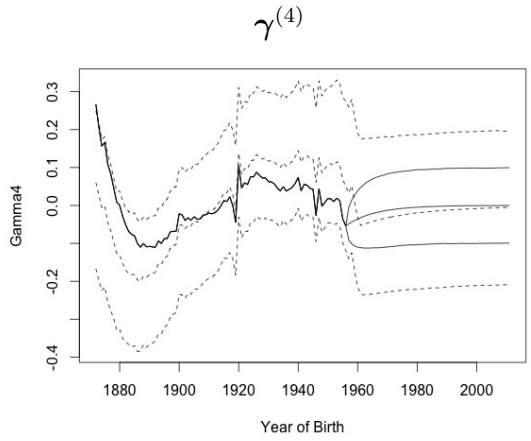


Figure 5.16: Comparison between the credibility intervals (dashed lines) and the confidence intervals (solid lines) for projecting the cohort effect for the males in England and Wales, given the Bayesian and MLE approach respectively. The upper bound of the interval is 95% quantile of the projected samples and the lower bound is 5% quantile. Both the period and cohort effects are projected fifty years forward. Note that the projection for MLE starts from cohort year $t_{n_y} - x_1 - 4$ given $\hat{\alpha}^*_\gamma$ and $\hat{\sigma}^*_\gamma$ derived by fitting the $AR(1)$ to the cohorts with the first and last five cohorts removed

Figure 5.16 shows that there is a clear zero mean reversion in the projection of $\boldsymbol{\gamma}^{(4)}$ for both MCMC and ML methods. The predictive intervals of MCMC is greater as we expect. We notice that the central projection for $\boldsymbol{\gamma}^{(4)}$ given the MCMC method is slightly lower than the mean of $\hat{\boldsymbol{\gamma}}'^{(4)}$ and the difference of the central projection between the two methods declines as the cohort year increases. This is because both the mean of the Bayesian estimation $\tilde{\gamma}^{(4)}_{t_{ny}-x_1}$ and the point estimate $\hat{\gamma}^{(4)}_{t_{ny}-x_1-5}$, i.e. the last empirical estimations, are negative with $\hat{\gamma}^{(4)}_{t_{ny}-x_1-5} > \tilde{\gamma}^{(4)}_{t_{ny}-x_1}$. Further the point estimate $\hat{\alpha}^*_\gamma$ for the coefficient of $AR(1)$ after removing the short cohorts is smaller than the mean of the Bayesian estimation $\tilde{\alpha}_\gamma$, implying a faster mean reversion for the ML projection.

We notice that the predictive intervals for $\boldsymbol{\kappa}^{(3)}$ given the MCMC method is not significantly wider than MLE, with respect to year, although we have discussed that the variance of $\tilde{\boldsymbol{\theta}}_1'$ is contributed by the parameter uncertainty and the normal future simulation noise and the latter is the only source of randomness for the ML projection. This result is consistent with Figure 5.6b, where the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ is left shifted from the point estimate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(3,3)$ with a smaller mean of volatility for $\boldsymbol{\kappa}^{(3)}$.

On the other hand, the predictive intervals for $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\kappa}^{(2)}$ are much wider than the projection given MLE due to the impact of parameter uncertainty according to Figure 5.2b and 5.4b where the volatility generated by ML is approximately the same with the mean of the posterior samples. In particular, MCMC generates a much lower central projection for $\boldsymbol{\kappa}^{(1)}$ than the ML method. As for $\boldsymbol{\kappa}^{(2)}$, we can see that the MCMC starts with a slightly higher central projection increases as time increases. However the gap between the two central projections diminishes and becomes eliminated approximately around year 2050. The trajectories of the projection is consistent with the distributions shown in Figure 5.2a and 5.4a, where the posterior distribution of $\boldsymbol{\mu}_1$ is left shifted from $\hat{\boldsymbol{\mu}}_1$ while $\tilde{\boldsymbol{\mu}}_2$ is approximately centred around the corresponding point estimates.

Given the projected period and cohort effects $\tilde{\boldsymbol{\theta}}_1' = \{\tilde{\boldsymbol{\theta}}_1'^{(k,l)}\}_{k=1,\ldots,N_2}^{l=1,\ldots,N_4}$ and $\hat{\boldsymbol{\theta}}_1' = \{\hat{\boldsymbol{\theta}}_1'^{(l)}\}_{l=1,\ldots,N_4}$, generated by MCMC and MLE respectively, construct the future death rates $\boldsymbol{m}'|\tilde{\boldsymbol{\theta}}_1'$ and $\boldsymbol{m}'|\hat{\boldsymbol{\theta}}_1'$. We plotted the predictive intervals of the projected rates for selected ages (50, 60, 80, and 89) with respect to year in Figure 5.17 for both the MCMC (dashed line) and the MLE (solid line) method, where the upper bound is 95% quantile of the simulated future samples and the lower bound is 5% quantile. Note that the future cohort used for calculating the $\boldsymbol{m}'|\hat{\boldsymbol{\theta}}_1'$ is projected from the cohort year $c = t_{n_y} - x_1 - 4$ given $\hat{\alpha}_\gamma^*$ and $\hat{\sigma}_\gamma^*$ as we discussed previously and the impact of this can be observed in Figure 5.17a-5.17b.

As we expected, the predictive intervals of the MCMC is much wider than its

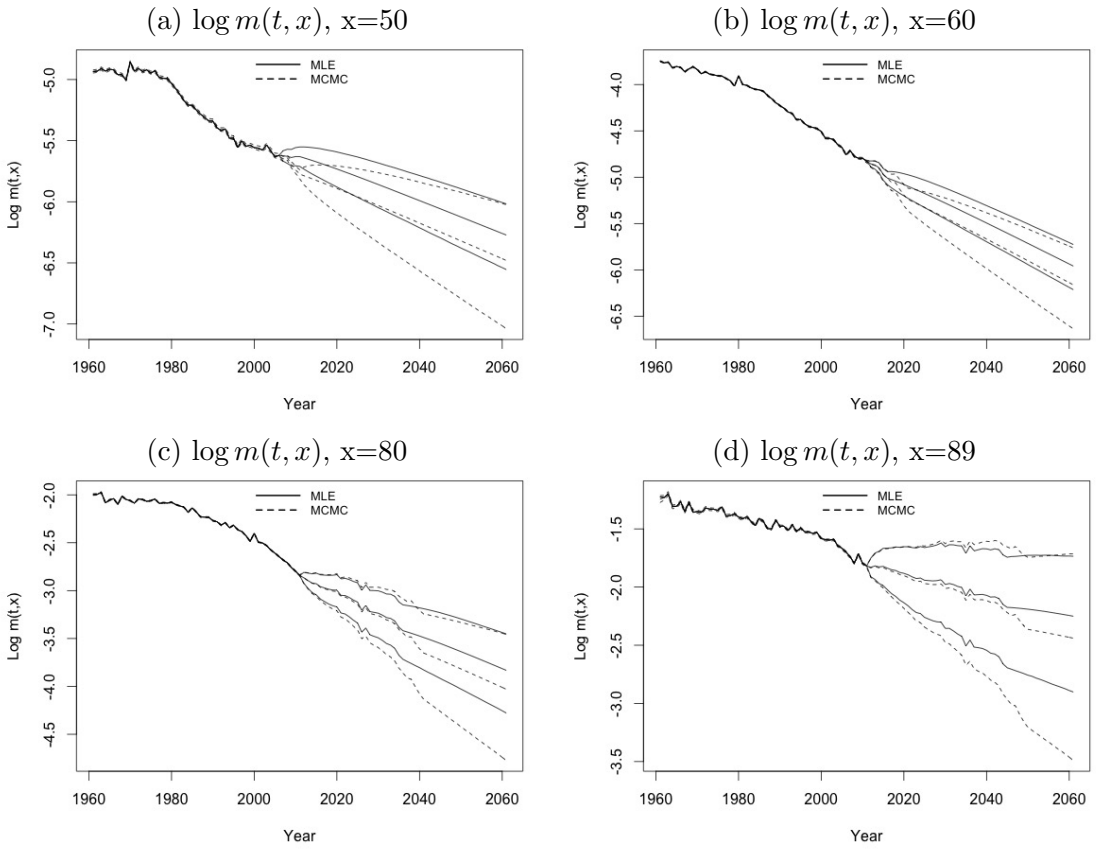


Figure 5.17: The predictive intervals of log-scaled projected death rates given MCMC (dashed) and ML method (solid) for selected ages (50,60,80,89) with respect to year. Note the upper bound is 95% quantile of the simulated samples and the lower bound is 5% quantile.

credibility intervals of the empirical fitted rates due to the normal randomness for simulating the future latent parameters. In general, the variance of $\boldsymbol{m}'|\tilde{\boldsymbol{\theta}}_1'$ is greater than $\boldsymbol{m}'|\hat{\boldsymbol{\theta}}_1'$ due to the parameter uncertainty and the difference is larger for longer term projection. In Figure 5.17c, we find that the projection is much more smoothed after year 2041 and 2036 for MCMC and MLE respectively since the cohort used for $\boldsymbol{m}'$ consists both the empirical estimation and the projection. It is worth to notice that the zigzag around year 2025 due to the variety of the cohort effect is smoothed by the time series prior, which supports the argument that MCMC smooths any noise in the period and cohort effect. Similar trajectory can be found for age 89 in Figure 5.17d.

We notice that at early ages, e.g. 50, 60, the central projection of the death rates generated by MCMC are lower than the ML's. This is mainly due to the central

projection of $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\gamma}^{(4)}$ in Figure 5.15a and 5.16. Given $x = 50$, the respected cohort years for the projection of $m(t_{n_y} + n', 50)$ given MCMC method ranged from cohort 1962 to 2011. Figure 5.16 shows that both the MCMC and ML method have the similar mean reverting trajectories with the central projection close to zero with respect to cohort year 1962-2011. In particular, we can see that the projection mean of MCMC is slightly lower than the ML method. Since we have observed that MCMC generated a lower central projection for $\boldsymbol{\kappa}^{(1)}$ and difference of the mean of the projection for $\boldsymbol{\kappa}^{(2)}$ and $\boldsymbol{\kappa}^{(3)}$ is not significant given the two methods, we can conclude that the trajectories of future $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\gamma}^{(4)}$ have significant impact on $\boldsymbol{m'}$. The explanation is the same for the pattern of $m'(t, x)$ at elder ages, e.g. 80 and 89, where the central projection is similar for both methods before a particular future year and then lower given the MCMC method.

We at last demonstrate the cohort mortality for the cohorts initially aged 55 (upper) and 65 (lower) in Figure 5.18. As expected, the prediction intervals are wider for the MCMC (black) at both ages due to the parameter uncertainty. Unsurprisingly, the central predictions based on the MCMC and the MLE are approximately the same along with the projecting horizon for each cohort.

5.5.3 Sensitivity Test

Theoretically the importance of the prior distribution diminishes as the population size increases. In this section, a sensitivity test is carried out to investigate the impact of choosing a stronger informative prior distribution on the posterior distribution. More specifically, we assume that the co-variance matrix follows the Inverse Wishart distribution with degree of freedom ν and the scale matrix $\boldsymbol{\Sigma}$, that is $\boldsymbol{V_\epsilon} \sim$ Inv-Wishart$(\nu, \boldsymbol{\Sigma})$ and the drift follows a uniform distribution. The values of ν and $\boldsymbol{V_\epsilon}$ are determined such that the mean of the prior distribution of $\boldsymbol{V_\epsilon}$ is the same with $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$. Note that given we have no prior knowledge of the true parameters for the benchmark population, we are not saying that the point estimate of $\boldsymbol{V_\epsilon}$,

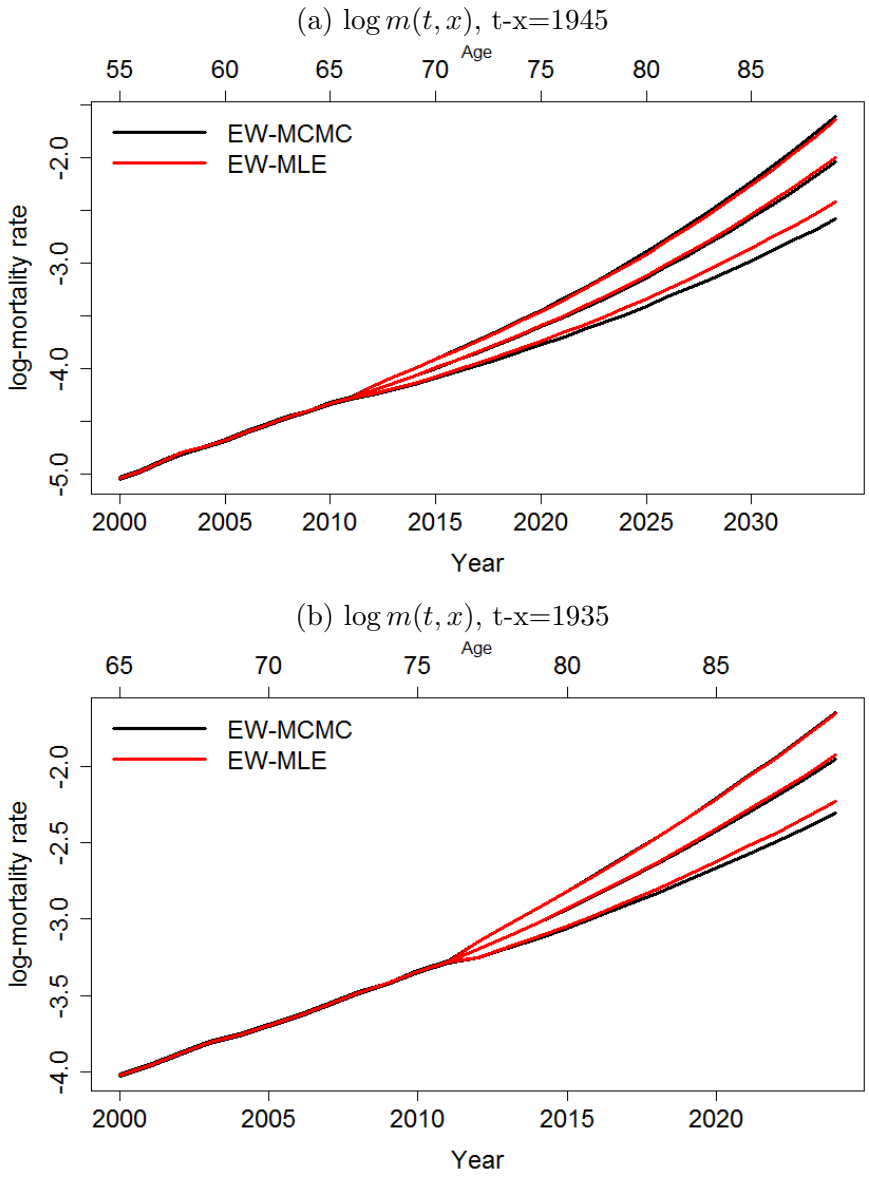


Figure 5.18: The predictive intervals of log-scaled cohort mortality rate of MCMC (black) and MLE (red) initially aged 55 (upper) and 65 (lower) in year 2000. Note the upper bound is 95% quantile of the simulated samples and the lower bound is 5% quantile.

given the MLE of the period effects, is an appropriate information for constructing an informative prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$. However, as we mentioned earlier, instead of determining a reasonable informative prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, the purpose of this test is to demonstrate an example of how the posterior distribution of $\boldsymbol{\theta}$, especially the hyper-parameters are affected by replacing the non-informative Jeffrey's prior to a stronger informative Inverse Wishart distribution. By selecting the values of the parameters for the prior distribution of the hyper-parameters to be consistent with

the simulated deaths scenarios of small population (w=0.01), the result of the sensitivity test enables us to learn how the population size affects the impact of the same informative prior on the posterior distribution. More specifically, we fix $\nu = 12$, $\boldsymbol{\Sigma} = 8\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ and $g = 3$ for α_γ.

We plotted the CDFs of the posterior distribution for the hyper-parameters, given both the non-informative (dashed curve) and informative (dotted curve) prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ in Figure 5.19-5.20. The vertical line is the point estimates for the hyper-parameters given the MLE of the latent parameters. As we expected, both the mean and the variance of the posterior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is slightly decreased due to the much stronger prior. In Figure 5.19a-5.19c, we can see that the variance of the posterior distribution of $\boldsymbol{\mu}$ is slightly decreased that is consistent with a smaller mean of the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$. The mean of $\boldsymbol{\mu}$ is barely affected by having a much stronger informative prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$.

Figure 5.20a shows that both the mean and the variance of the posterior for α_γ is a smaller by increasing the value of the parameter g to 3. This is because by fixing $g = 3$ we assume a much stronger beta-type prior for α_γ to push the estimation even further from one. We therefore conclude that the posterior distribution for α_γ is sensitive to the choice of the prior distribution. The distribution of σ_γ is hardly affected by using stronger prior, see Figure 5.20b.

However, for all the hyper-parameters the impact on the posterior distribution is relatively small referring to changing the prior distribution from non-informative to a much stronger informative. This supports our argument that the Poisson likelihood dominates the posterior distribution when the population is large and we may need not work especially hard to find a prior that reflects all the possible information.

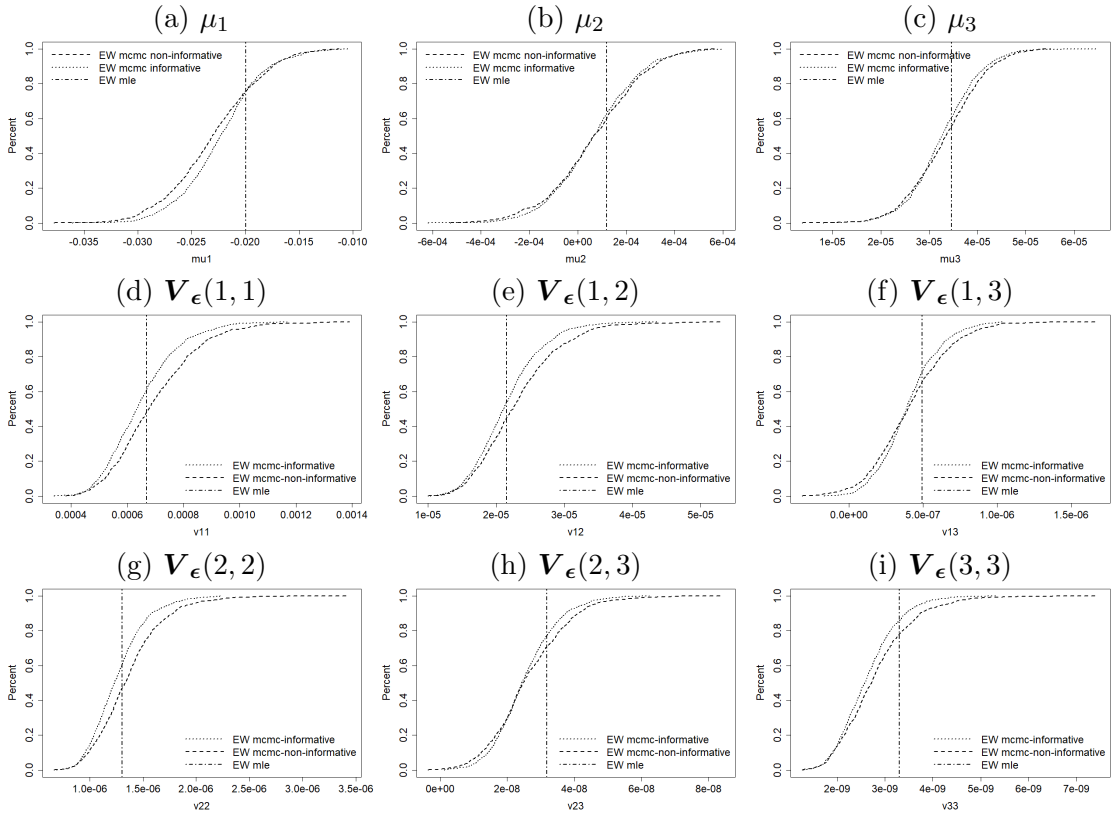


Figure 5.19: CDF of sensitivity test: The posterior distribution of $\boldsymbol{\mu}$ (first row) and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ (second and third row) for the males in England and Wales during year 1961-2011, aged 50-89 last birthday, given the non-informative Jeffrey's prior (dashed curve) and the informative Inverse Wishart prior (dotted curve) for the co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$. The vertical line is the corresponding point estimates generated by the MLE of the latent parameters.

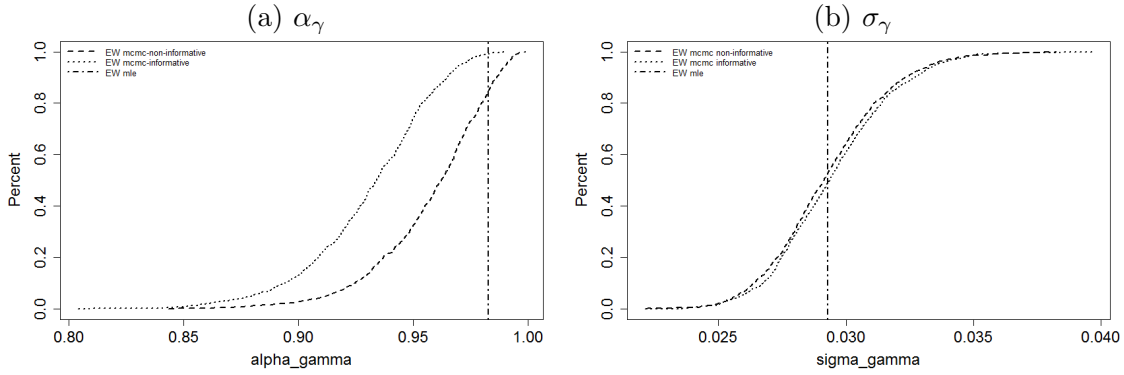


Figure 5.20: CDF of sensitivity test: The posterior distribution of α_γ (left) and σ_γ (right) for the males in England and Wales during year 1961-2011, aged 50-89 last birthday, given the non-informative Jeffrey's prior (dashed curve) and the informative Inverse Wishart prior (dotted curve) for the co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$. The vertical line is the corresponding point estimates generated by the MLE of the latent parameters.

5.5.4 Conclusion

In Section 5.5, we simulated the posterior distribution of $\boldsymbol{\theta}$ for the males in England and Wales during 1961-2011 aged 50-89 last birthday with Hamiltonian Monte Carlo updating algorithm. As we expected the MCMC smooths any variety of the period and cohort effects and the trajectories generated is slightly different from the MLE's since no identifiability constraints are adopted and the short cohorts are balanced by the ARIMA model for the cohort effects.

For the hyper-parameters in the random walk model, we notice that both of the posterior distributions for the drift and the co-variance matrix are approximately centred around the respected point estimates.

Similarly, the posterior distribution of the coefficient of the ARIMA model for the cohort effect is left shifted from the point estimate. However the point estimate of the volatility of the cohort process lies approximately in the centre of the posterior distribution. We re-fit the ARIMA model to the cohort process by removing the first and the last five short cohorts since the point estimates derived based on few observations are less reliable. As we expect, the new point estimates for the coefficient and the variance are much less than the MCMC mean. No change is made to the posterior distribution of the short cohorts as the low level information is balanced by the time series prior distribution.

By calculating the ratio of fitted rates given the two methods and the pattern of the residuals, we can conclude that both the mean of the MCMC and MLE provide a good and similar fit to the crude rates and there is no single year or age for which one method fits much better than the other. We can see the smoothness added by the MCMC to fitting the crude rates but the impact is quite small as the population is large. The trajectories of central projection of the latent parameters are different given the two methods due to the identifiability constraints in the fitting process and the differed estimation for the hyper-parameters, which result in lower central

projection for the death rates.

Without information of the true rates for the underlying benchmark population, we cannot determine if the mean of the posterior distribution provides a better fit to the point estimates. However, the posterior distribution allows us to study how the parameter uncertainty affects the parameter estimation and the projection and the impact of the time series prior for large population.

5.6 Summary

In this chapter, we discussed fitting the deaths data with a Bayesian model which combines the Poisson likelihood with the time series prior for the latent parameters. The period effects are assumed to follow a multi-variate random walk model while an $AR(1)$ model is allocated for the cohort effects. We determined the prior distribution of the hyper-parameters according to the population size. In particular, we selected an informative prior distribution for the volatility of the simulated deaths scenario such that the information of the benchmark population can be adopted to provide more accurate estimation. The posterior distribution of the parameters were simulated by the Hamiltonian Monte Carlo that provided an efficient convergence without the identifiability constraints. In the last section, we carried out an empirical study by fitting the Bayesian model to the males in England and Wales and compared the estimation with the two-stage approach.

Chapter 6

Fitting Small Population with Bayesian Approach

In the previous chapter, the empirical study shows that the impact of the time series prior is not significant when the population size is large and without any prior knowledge of the true rates of the reference population it is hard to decide which method provides a better fit and projection for the underlying population.

In this chapter, we simulated N_1 deaths scenarios $D^w(t,x) = \{D_j^w(t,x)\}_{j=1,\ldots,N_1}$ with corresponding exposure that is as small as one percent of the benchmark population by assuming $D^w(t,x) \sim \text{Pois}(m(\theta_{1,0},t,x)wE_0(t,x))$, where $w = 0.01$ is the weight for scaling down the benchmark exposure $E_0(t,x)$, such that the true parameters of $D^w(t,x)$ are the point estimates of the reference population. Recall that $\theta_{1,0}$ is a given parameter vector defined on page 28 and fixed to θ_1^{EW}, the relative parameter estimates of the England and Wales data. As described in Section 5.4.1 of Chapter 5, we adopt the information from the benchmark population to improve the estimation by employing a more informative prior distribution for the co-variance matrix, that is either the mean of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is fixed to $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ or the mode of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is fixed to $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ as a sensitivity test. We then fit the

Bayesian model to $D_j^w(t,x)$ for $j = 1,\ldots,N_1$ and draw $N_2 = 1000$ samples from the MCMC procedure by following the steps and the initial settings for $w = 0.01$ described in Section 5.4.4.

Section 5.4.4 also introduced five methods based on which the empirical study on the posterior samples is carried out. Let $\tilde{\boldsymbol{\theta}}^w = \{\tilde{\boldsymbol{\theta}}_j^{w,(k)}\}_{j=1,\ldots,N_1}^{k=1,\ldots,N_2}$ be the posterior estimation for $\boldsymbol{\theta}$ of $w = 0.01$, where $\tilde{\boldsymbol{\theta}}_j^{w,(k)}$ is the k^{th} posterior sample for the deaths scenario j. Recall the five methods described in Section 5.4.4:

Method 1 All the samples of $\tilde{\boldsymbol{\theta}}^w$ for $j = 1,\ldots,N_1$ and $k = 1,\ldots,N_2$.

Method 2 Choose a k^* from $1,\ldots,N_2$. For each $j = 1,\ldots,N_1$, draw the k^*th posterior sample $\{\tilde{\boldsymbol{\theta}}_j^{w,(k)}\}_{j=1,\ldots,N_1;k=k^*}$.

Method 3 For each $j = 1,\ldots,N_1$, calculate the mean of the corresponding posterior samples, defined as $\bar{\tilde{\boldsymbol{\theta}}}_j^w = \frac{1}{N_2}\sum_{k=1}^{N_2}\tilde{\boldsymbol{\theta}}_j^{w,(k)}$ for $j = 1,\ldots,N_1$.

Method 4 Study the posterior distribution given all the N_2 posterior samples of a randomly drawn deaths scenario $j^* \in N_1$, $\{\tilde{\boldsymbol{\theta}}_j^{w,(k)}\}_{j=j^*;k=1,\ldots,N_2}$.

Method 5 Study how much variability there is in the posterior distribution, by comparing 100 randomly drawn deaths scenarios. Denote as $\tilde{\boldsymbol{\theta}}_{j'}^w$ for $j' = j_1,\ldots,j_{100}$ the posterior estimation for the j'^{th} randomly selected death scenario. In practice, a practitioner may only have one underlying population. The reason for this method is that there is uncertainty on the posterior distribution itself driven by the sampling variation of the death counts. We want to investigate that uncertainty.

Method 6 As Method 4 but we fix the mode of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the relative true parameter.

Further, we carry out the sensitivity test (Method 6) by fitting the model to the deaths scenario in Method 4 and fixing the mode of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the true parameter instead of the mean.

6.1 Convergence

We start with investigating the convergence of the MCMC procedures. In Section 5.4.1, we discussed one of the motivations for adopting the informative prior distribution for the volatility of the small population. That is the MCMC procedure does not converge to a stationary distribution, given a flat likelihood $p(\boldsymbol{D}, \boldsymbol{E}|\boldsymbol{\theta})$ and a flat prior $p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ when the impact of the prior distribution is large and competing with the Poisson likelihood.

We illustrate this with a case study by running the MCMC for the simulated death scenario j^*, given the non-informative Jeffrey's prior for the joint density $p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$, without changing the other settings. Figures 6.1-6.2 compare the MCMC trajectories of the death scenario j^*, given the Jeffrey's prior for $p(\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$ (Figure 6.1) and the more informative Inverse Wishart for $p(\boldsymbol{V_\epsilon})$ (6.2) respectively. The other settings are the same. The charts include the period and cohort effects in selected years (1961, 2011) and years of birth (1901, 1931 and 1961) respectively, the drift of multi-variate random walk $\boldsymbol{\mu}$, the diagonal of the co-variance matrix $\boldsymbol{V_\epsilon}$, coefficient α_γ and the standard deviation σ_γ of $AR(1)$ model.

We can see that use of the Jeffrey's prior distribution for $\boldsymbol{V_\epsilon}$ causes problems for convergence of the MCMC even after 100000 iterations (more likely $\boldsymbol{V_\epsilon}(2,2) \to 0$ and $l(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E}) \to \infty$). More specifically the MCMC converges to a straight line for most of the parameters. On the other hand, adopting a stronger informative prior distribution with the external information from the benchmark population greatly improves the efficiency of the MCMC mixture. For prudence purpose, we checked the convergence of the MCMC for every simulated death scenario $j = 1, \ldots, N_1$.

Recall that the degrees of freedom ν of the Inverse Wishart distribution for $\boldsymbol{V_\epsilon}$ is fixed to 12 to guarantee a convergent MCMC for every simulated death scenario. We thus need to check if the prior distribution of $\boldsymbol{V_\epsilon}$ for small population $w = 0.01$ is too strong by comparing the following two distributions:

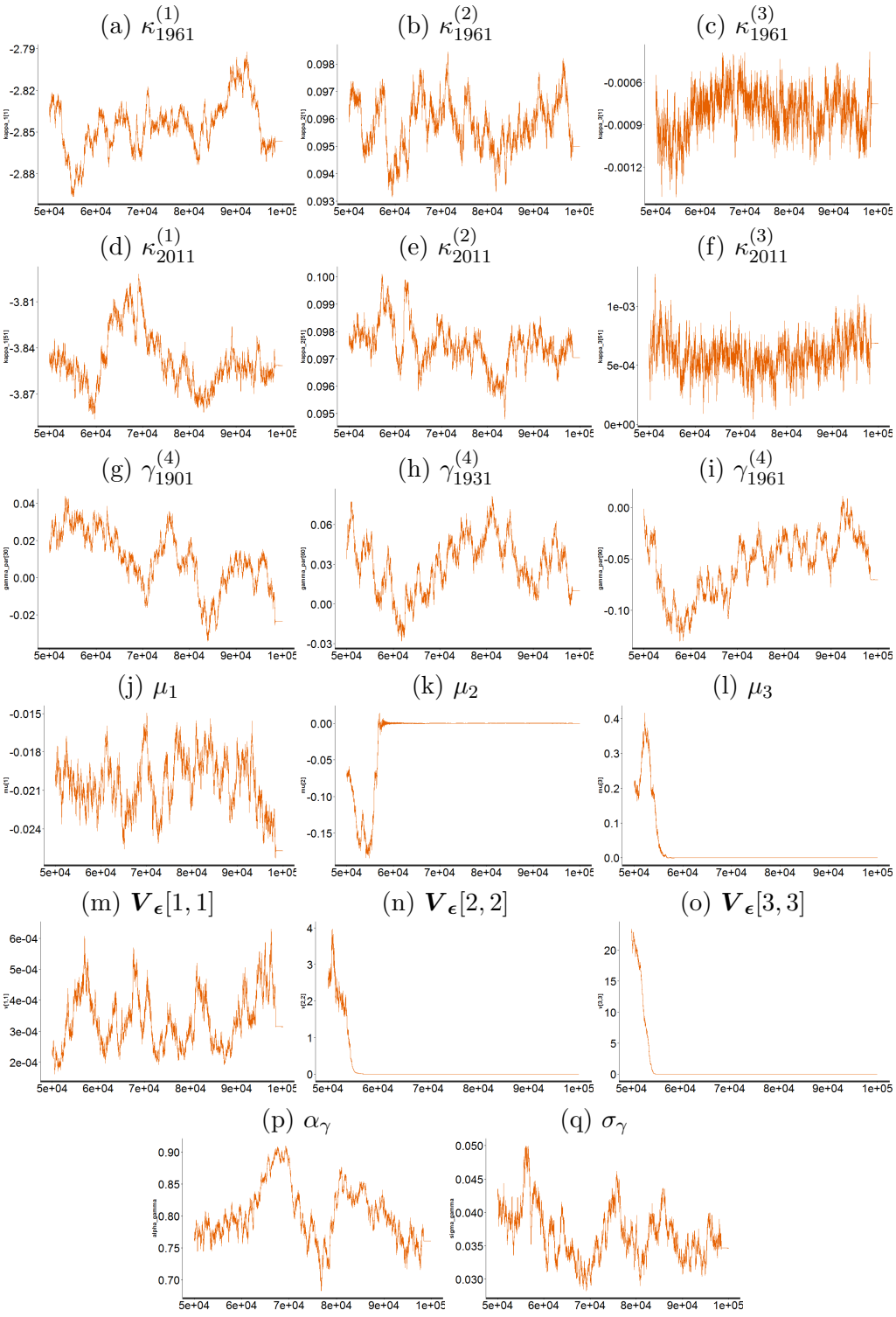


Figure 6.1: Trace plot for selected parameters, including the period and cohort effects with selected year (1961, 2011) and year of birth (1901,1931,1961) respectively, the drift of multi-variate random walk $\boldsymbol{\mu}$, the diagonal of co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, coefficient and standard deviation of AR(1) model α_γ, σ_γ. The data fitted is one of the simulated deaths scenario with exposure size one percent of the England and Wales data, males during year 1961 to 2011 aged 50 to 89 last birthday. The prior density for $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is a non-informative Jeffrey's prior.

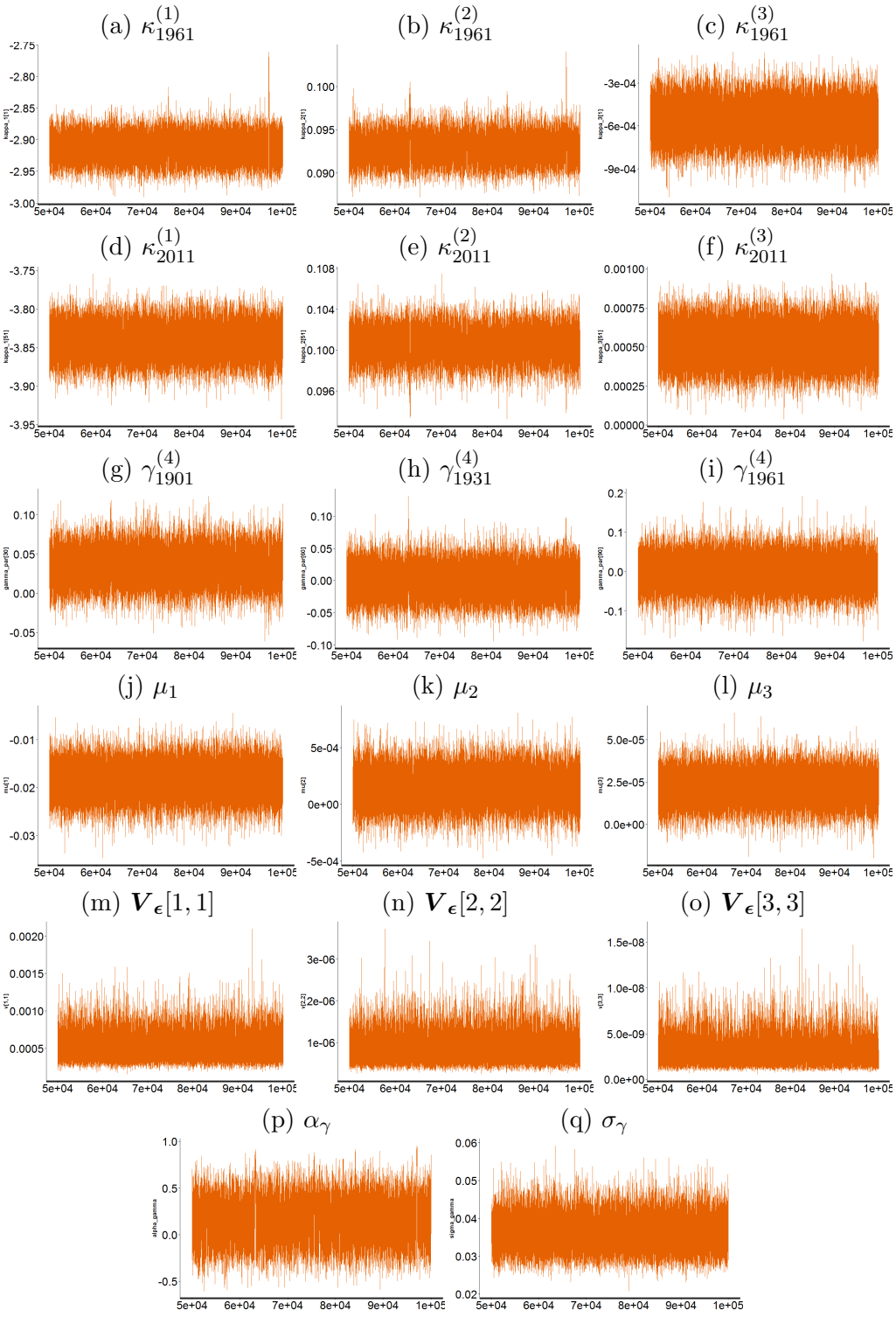


Figure 6.2: Trace plot for selected parameters, including the period and cohort effects with selected year (1961, 2011) and year of birth (1901,1931,1961) respectively, the drift of multi-variate random walk $\boldsymbol{\mu}$, the diagonal of co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, coefficient and standard deviation of AR(1) model α_γ, σ_γ. The data fitted is one of the simulated deaths scenarios with exposure size one percent of the England and Wales data, males during year 1961 to 2011 aged 50 to 89 last birthday. The prior density for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is the Inverse Wishart distribution with MCMC-Mean.

(a) Prior distribution for $\boldsymbol{V_\epsilon}$ for $w = 0.01$;

(b) Posterior distribution of $\boldsymbol{V_\epsilon}$ for England and Wales data, given the Jeffrey's prior.

We calculate the variance for distribution (a) and (b) and we find that the variance of (a) is approximately six times greater than (b). Thus (a) is much wider spread than (b), which implies that our choice of prior for $\boldsymbol{V_\epsilon}$ of simulated death scenarios is not too strong. Further the posterior distribution of $\boldsymbol{V_\epsilon}$ of England and Wales data given the Jeffrey's prior is approximately an Inverse Wishart distribution with much higher degree of freedom than the simulated death scenarios.

6.2 Compare with the Two-Stage Approach

For a small population, e.g. one percent of England and Wales population size, the time series prior competes with the Poisson likelihood so that the sampled period and cohort effects are more like they are from the proposed ARIMA models. In this section, we investigate the impact of the time series prior on the posterior distribution of all the parameters for a small population, especially the improvement of estimation for the volatility of the data.

6.2.1 Period Effect $\boldsymbol{\kappa}$ and the Hyper-Parameters $\boldsymbol{\mu}$, $\boldsymbol{V_\epsilon}$

This section studies the posterior distribution of the period effects $\boldsymbol{\kappa}$ and the corresponding hyper-parameters $\boldsymbol{\mu}$ and $\boldsymbol{V_\epsilon}$ for small population.

Recall that by simulating death samples from the Poisson model

$$D^w(t,x) \sim \text{Pois}(m(\theta_{1,0},t,x)wE_0(t,x)),$$

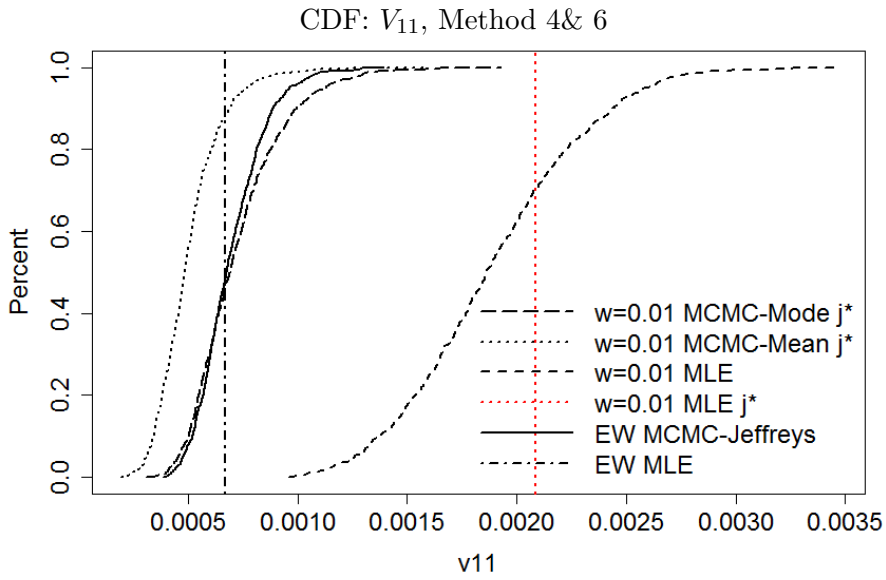


Figure 6.3: CDF: posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ of the simulated death scenario j^*, given the MCMC-Mean (Method 4, dotted curve) and the MCMC-Mode (Method 6, long-dashed curve). The solid curve and the vertical dot-dashed line are the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(1,1)$ and the true rate of the simulated death scenarios $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ respectively. The dashed curve is the CDF for $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(1,1)$, the finite-sample MLEs of all N_1 simulated death scenarios, where the red vertical line is $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j^*}(1,1)$, the relative MLE for death scenario j^*.

we ensure that the true parameters for the small population $w = 0.01$ are equivalent to the point estimates of the benchmark population, $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}^{\text{EW}}$.

For $\boldsymbol{\kappa}^{(1)}$, $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$

Figure 6.3 includes the following curves:

- $w = 0.01$ MCMC-Mode j^*: the posterior distribution of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ for the simulated death scenario j^*, given the prior setting of MCMC-Mode,
- $w = 0.01$ MCMC-Mean j^*: the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ of the simulated death scenario j^*, given the prior setting of MCMC-Mean,
- $w = 0.01$ MLE: the distribution $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(1,1)$ of the finite-sample MLEs of the N_1 simulated death scenarios,

- $w = 0.01$ MLE j^*: the relative MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j^*}(1,1)$ for the death scenario j^*,

- EW MCMC: the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$ of the England and Wales data,

- EW MLE: the MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$ of the England and Wales data.

The same line type settings for the various distributions are applied for the other parameters unless otherwise noted. See the definition of MCMC-Mode and MCMC-Mean in Equation (5.33) and (5.32) respectively in Section 5.4.1, Chapter 5. The MLEs for each death scenario and England and Wales data are calculated based on Equation (3.9) in Section 3.3.1, Chapter 3.

The CDF of $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(1,1)$ (dashed curve), the finite-sample MLEs for the N_1 death scenarios, is significantly shifted to the right-hand side of the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$ (vertical dot-dashed line, the point estimate of the England and Wales data), which supports our conclusions in Chapter 3 for small populations, e.g. $w = 0.01$, such that the finite-sample MLE $(\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(1,1)|\hat{\boldsymbol{\kappa}}^{(1),w})$ conditional on the point estimate $\hat{\boldsymbol{\kappa}}^{(1),w}$ according to the two-stage approach is significantly over-estimated from the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$ due to the large sampling variation of the small population.

On the other hand, the CDFs of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ for the simulated death scenario j^* in Figure 6.3, conditional on both MCMC-Mode (long-dashed curve) and MCMC-Mean (dotted curve), are slightly right-tailed and lie much closer to the true parameter $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$, implying a much smoothed trajectory of $\boldsymbol{\kappa}^{(1)}$ of $w = 0.01$ generated by the Bayesian approach. More specifically, we can see that the posterior mean of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$, (given either the MCMC-Mode or MCMC-Mean), is significantly lower than the distribution of the finite-sample MLEs (and hence the point estimate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j^*}(1,1)$ of the death scenario j^*).

This is because the prior distribution plays an even more important role on the joint posterior distribution and competes the Poisson likelihood when the population size is small. Therefore, the significant sampling variation is balanced by the time

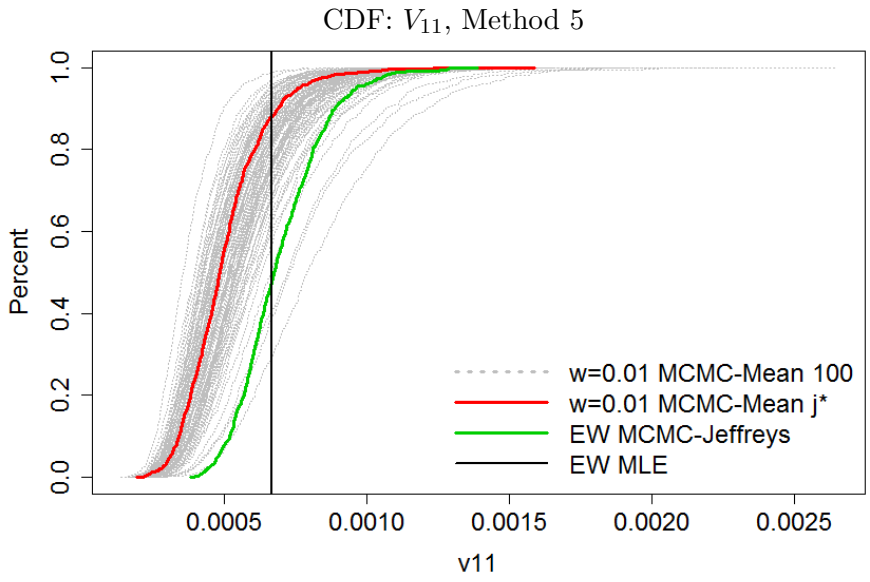


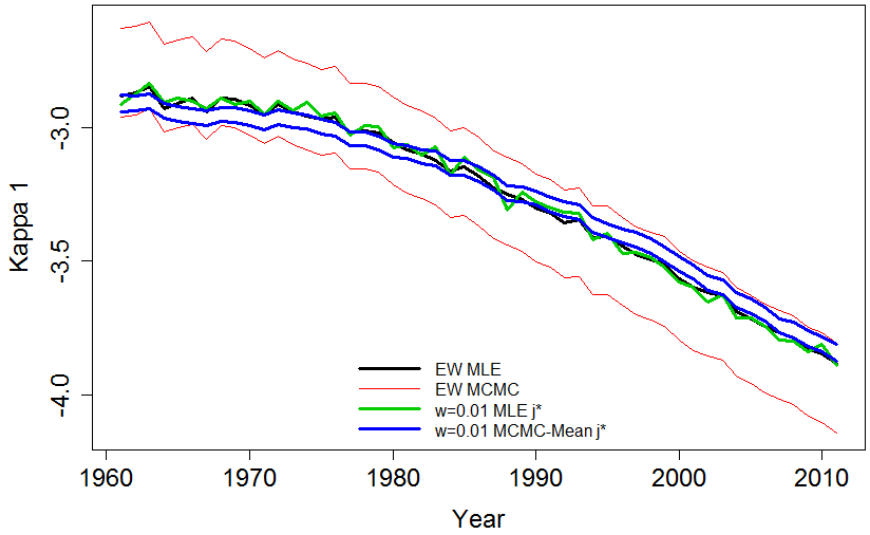
Figure 6.4: CDFs of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}(1,1)$ for $j' = j_1, \ldots, j_{100}$ (MCMC-Mean, dotted curves), the posterior distributions for 100 randomly selected simulated death scenarios. The red curve is the CDF of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ conditional on MCMC-Mean for death scenario j^*. The green curve and the vertical solid line are the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(1,1)$ and the MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ of the England and Wales data respectively.

series prior of the latent parameters and the estimation is more like it is produced from the proposed time series models.

In particular, the CDF of the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ conditional on the MCMC-Mode (long dashed curve) is shifted to the right hand side of the distribution conditional on the MCMC-Mean with a slightly higher posterior variance by having a doubled prior mean and variance for $\tilde{\boldsymbol{V}}(1,1)$, see Table 6.1 and 6.3 for the statistics of the posterior and prior distributions respectively. We can see that $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ conditional on the MCMC-Mode is centred around the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ and approximately the same with the posterior distribution of the England and Wales data (solid curve), see Table 6.1.

In Figure 6.5b, we could observe a relatively slightly less smoothed trajectory of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$, the posterior estimation of $\boldsymbol{\kappa}^{(1)}$ for the death scenario j^*, conditional on the MCMC-Mode (dashed lines) compared with the MCMC-Mean (solid lines) due to the higher volatility estimation given MCMC-Mode, although visually the

(a) Historical $\boldsymbol{\kappa}^{(1)}$, Method 4 MCMC-Mean



(b) Historical $\boldsymbol{\kappa}^{(1)}$, Method 4 MCMC-Mean versus Method 6 MCMC-Mode

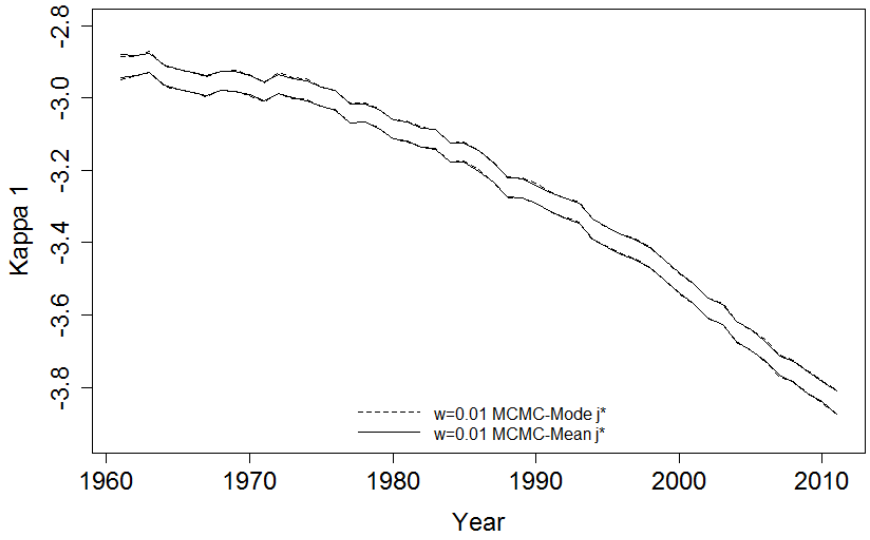


Figure 6.5: Credibility Intervals: posterior distribution of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ for the simulated death scenario j^*, given the MCMC-Mean (upper blue lines; lower solid lines) and the MCMC-Mode (lower dashed lines). The red lines and the black line in the upper figure are the posterior distribution $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ and the MLE $\hat{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ respectively for the England and Wales data. The green line is $\hat{\boldsymbol{\kappa}}_{j^*}^{(1),w}$, the MLE of death scenario j^*. Note that the upper and lower bound are the 95% and 5% quantile of the posterior distribution.

estimations of $\boldsymbol{\kappa}^{(1)}$ given the two conditions are approximately the same for the death scenario j^*. In Figure 6.5a, as we expect the credibility intervals of the posterior distribution $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ (MCMC-Mean, blue lines) for the death scenario j^* is much more smoothed compared with the trajectory of $\hat{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ (green line), the MLE of the death scenario j^*, which is consistent with our findings that the volatility of $\boldsymbol{\kappa}^{(1)}$ for the

small population j^* is much better estimated by the MCMC. We could therefore conclude that, based on Figure 6.5, MCMC-Mode (and MCMC-Mean) generates a less fluctuated estimation for $\boldsymbol{\kappa}^{(1)}$ compared with the two-stage approach.

				$\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$
Posterior	Mean	England and Wales	C	6.97E-04	1.36E-06	2.80E-09
		$w = 0.01$	A	7.26E-04	1.36E-06	4.85E-09
			B	5.04E-04	8.28E-07	2.87E-09
	Std Dev	England and Wales	C	1.54E-04	3.25E-07	0.77E-09
		$w = 0.01$	A	2.11E-04	4.10E-07	1.87E-09
			B	1.49E-04	2.72E-07	1.24E-09
	Skewness	England and Wales	C	0.8227	1.0830	1.0605
		$w = 0.01$	A	1.2121	1.0843	1.4038
			B	1.4013	1.2088	1.6301

Table 6.1: Characteristic statistics of: the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$, $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$, and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ for England and Wales data and selected death scenario with $w = 0.01$, during year 1961-2011, aged 50-89 last birthday. Note A = MCMC-Mode; B = MCMC-Mean; C = Jeffrey's Prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for modelling England and Wales data.

We can also see that the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$ (solid curve) for the England and Wales data is higher than the $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ for death scenario j^* conditional on MCMC-Mean, which indicates that the posterior estimation $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ for the first period effect $\tilde{\boldsymbol{\kappa}}^{(1)}$ of the England and Wales data is relatively more volatile compared the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ (MCMC-Mean). In Figure 6.5a we can see that $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ (MCMC-Mean, blue lines) is relatively more smoothed than the $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ (red lines). This is once again because the influence of the time series prior on the joint posterior distribution is relatively much smaller for the England and Wales data compared

with the smaller population $w = 0.01$. Therefore the estimation of the period effect $\tilde{\boldsymbol{\kappa}}^{(1)}$ for the England and Wales data is supposed to look more like MLEs generated according to the Poisson likelihood (which is also the reason $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(1,1)$ is approximately centred around the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$) with barely smoothness from the random walk model.

It is worth noticing that the uncertainty of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$, conditional on either MCMC-Mode and MCMC-Mean, is not significantly different from that of $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(1,1)$. We demonstrate the characteristic statistics (mean, standard deviation and skewness) of the posterior distributions in Table 6.1. The standard deviation of the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(1,1)$ (1.54×10^{-4}) is slightly greater than that of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ (MCMC-Mean 1.49×10^{-4}), which could be interpreted by the choice of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ as a non-informative prior is assigned for the England and Wales data. On the other hand, $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ has more variability conditional on MCMC-Mode than $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(1,1)$ does mainly due to the prior variance for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ is doubled for the MCMC-Mode, compared with the MCMC-Mean.

According to Method 5, we randomly selected 100 death scenarios and denote as $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}(1,1)$ for $j' = j_1, \ldots, j_{100}$ the posterior distribution (conditional on the MCMC-Mean) of the j'^{th} death scenario. In Figure 6.4, one hundred CDFs of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}$ where $j' = j_1, \ldots, j_{100}$ are plotted for all the randomly selected death scenarios respectively. By doing this, we demonstrate how much the distribution of the posterior distributions is driven by the sampling variation for the simulated death scenarios so that we could investigate if the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ for death scenario j^*, given either the MCMC-Mode or MCMC-Mean in Figure 6.3 is systematically the same with the true parameter $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ and EW's posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(1,1)$.

We can see that the influence of the sampling variation shifting the posterior distribution from one side to another and such shift could be quite large. Visually, most of the selected death scenarios have approximately 65% of their posterior samples less than the true parameter $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ and the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$

(MCMC-Mean, red curve) lies a bit to the left side of the centre of the distribution of the posterior distributions. The posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$ (green curve) also lies within the region of the distribution of the posterior distributions (a bit right to the centre though). It implies that the posterior estimation $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(1,1)$ of $w = 0.01$ conditional on the MCMC-Mean is only slightly less than the respective true parameter and there is no significant difference between the posterior distribution of the England and Wales data and the small population $w = 0.01$ (MCMC-Mean), given the influence of the sampling variation.

We have concluded that the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ conditional on MCMC-Mode for death scenario j^* is similar with $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$ and approximately centred around the true parameter $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$. We therefore expect that, without the need to re-running the MCMC for all the N_1 simulated death scenarios conditional on the MCMC-Mode, the distribution of the posterior distributions for all the simulated death scenarios conditional on MCMC-Mode will be shifted to the right hand side and centred around the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$ with the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$ approximately in the centre of the distribution. The sampling variation shifts the posterior distribution of the simulated death scenario from one side to another around the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$ and $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$, implying an unbiased estimation for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ of the simulated death scenarios.

According to Method 3, for each death scenario $j = 1, \ldots, N_1$ we worked out the posterior mean of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}$ and $\tilde{\boldsymbol{\mu}}^{w,j}$. We denoted as $\bar{\tilde{\boldsymbol{\mu}}}^{w} = \{\bar{\tilde{\boldsymbol{\mu}}}^{w,j}\}_{j=1,\ldots,N_1}$ and $\bar{\tilde{\boldsymbol{V}}}_{\epsilon}^{w} = \{\bar{\tilde{\boldsymbol{V}}}_{\epsilon}^{w,j}\}_{j=1,\ldots,N_1}$ the distributions of the posterior means for $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ respectively. Some characteristic statistics of $\bar{\tilde{\boldsymbol{\mu}}}^{w}$ and $\bar{\tilde{\boldsymbol{V}}}_{\epsilon}^{w}$ are calculated in Table 6.2, include the mean and the percentiles. It shows that due to the sampling variation shifting the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}(1,1)$ around 95% of the simulated death scenarios have their posterior means lie within the range of $(4.06 \times 10^{-4}, 7.31 \times 10^{-4})$.

One application is shown in Figure 6.6, which demonstrates the influence of the sampling variation on the standard deviations (upper) and the coefficient of skew-

Quantile	$\bar{\tilde{\boldsymbol{\mu}}}_1^w$	$\bar{\tilde{\boldsymbol{\mu}}}_2^w$	$\bar{\tilde{\boldsymbol{\mu}}}_3^w$	$\bar{\tilde{\boldsymbol{V}}}^w(1,1)$	$\bar{\tilde{\boldsymbol{V}}}^w(2,2)$	$\bar{\tilde{\boldsymbol{V}}}^w(3,3)$
97.5%	-1.84E-02	21.84E-05	2.48E-05	7.31E-04	11.48E-07	4.26E-09
50%	-1.93E-02	13.83E-05	1.90E-05	5.18E-04	8.78E-07	3.03E-09
2.5%	-2.02E-02	-0.11E-05	1.37E-05	4.06E-04	7.47E-07	2.46E-09
Mean	-1.93E-02	13.06E-05	1.90E-05	5.27E-04	8.96E-07	3.12E-09
	$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_2$	$\boldsymbol{\mu}_3$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$
True rate	-2.00E-02	11.98E-05	3.46E-05	6.70E-04	13.05E-07	3.30E-09

Table 6.2: The mean and 97.5%, 50%, 2.5% quantiles of the finite samples of the MCMC mean $\bar{\tilde{\boldsymbol{\mu}}}^w = \{\bar{\tilde{\boldsymbol{\mu}}}^{w,j}\}_{j=1,\ldots,N_1}$ and $\bar{\tilde{\boldsymbol{V}}}^w = \{\bar{\tilde{\boldsymbol{V}}}^{w,j}\}_{j=1,\ldots,N_1}$ for the 1000 death scenarios (Method 3). Note that the true rate is the MLE of the England and Wales data.

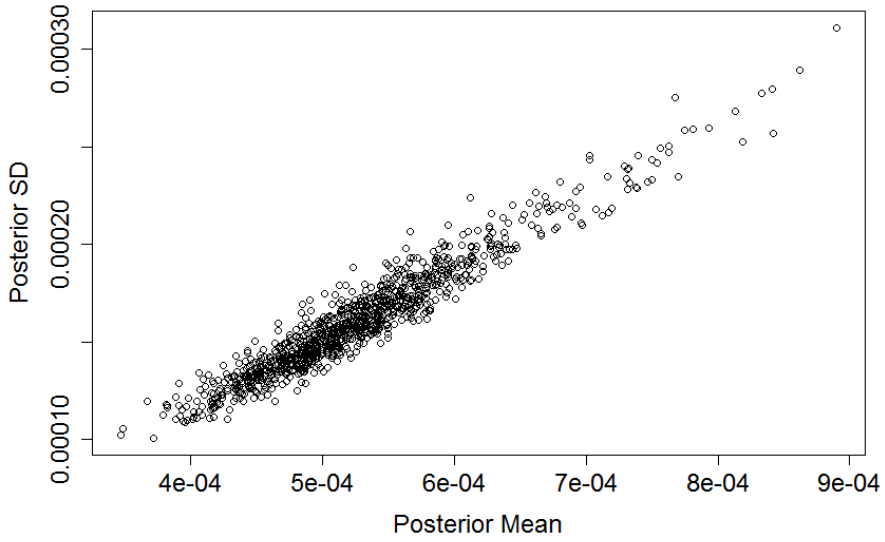
ness[1] (lower) of the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(1,1)$ for $w = 0.01$ according to the N_1 simulated death scenarios. As may be expected, Figure 6.6a shows that the posterior variance increases when the sampling variation shifts the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(1,1)$ away from zero, while there is no obvious trend between the skewness and the mean in Figure 6.6b. Therefore we conclude that while the sampling variation shifts the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(1,1)$ from one way to another, it has no obvious impact on the skewness of the posterior distributions. Similar conclusion of the influence of the sampling variation could also be made for $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(2,2)$ and $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(3,3)$, see Figure 6.15 and 6.24 respectively.

We therefore conclude that embedding the information of the benchmark population with an informative prior distribution for the volatility of the period effect $\boldsymbol{\kappa}^{(1)}$

[1]For random variable $\boldsymbol{X}$ with mean μ and standard deviation σ, the coefficient of skewness is defined as the third standardized moment:

$$E\left[\left(\frac{\boldsymbol{X}-\mu}{\sigma}\right)\right]$$

(a) Scatter Plot: Posterior Means vs SD



(b) Scatter Plot: Posterior Means vs Skewness

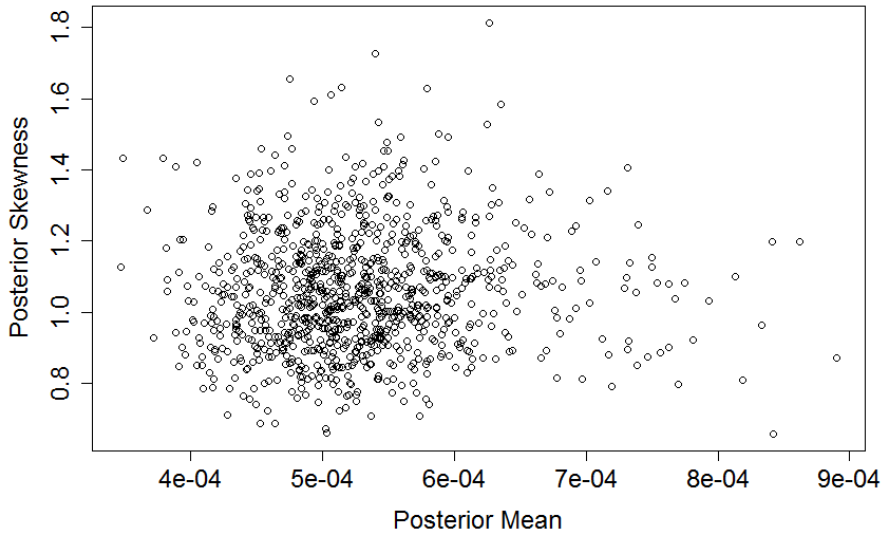


Figure 6.6: The scatter plot: the posterior mean vs. coefficient of skewness (lower); the posterior mean vs. standard deviation (upper) for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ of $w = 0.01$, based on the posterior distributions $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}$ for $j = 1, \ldots, N_1$. Note that the coefficient of skewness is calculated as the third standardized moment of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$.

provides a much improved estimation for the volatility and hence a more smoothed estimation for $\boldsymbol{\kappa}^{(1)}$, compared with the two-stage approach. We observed that the sampling variation shifts the posterior distribution from one side to another and according to the distribution of the posterior distributions of the simulated deaths we found that fixing the mode of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ to the corresponding MLE of the benchmark population generates an unbiased estimated volatility from the true rate for the simulated small populations.

				$\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$
Prior	Mean	England and Wales	C	NA	NA	NA
		$w = 0.01$	A	13.40E-04	2.61E-06	6.60E-09
			B	6.70E-04	1.31E-06	3.30E-09
	Std Dev	England and Wales	C	NA	NA	NA
		$w = 0.01$	A	7.73E-04	15.10E-06	3.81E-09
			B	3.87E-04	7.54E-07	1.91E-09

Table 6.3: The mean and the standard deviation of the prior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$, $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$, and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ for England and Wales data and $w = 0.01$, during year 1961-2011, aged 50-89 last birthday. Note A = MCMC-Mode; B = MCMC-Mean; C = Jeffrey's Prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for modelling England and Wales data.

These conclusions could be adopted by a pension scheme manager whose underlying population is only a small subset of a bigger population that can be well fitted by a certain stochastic model. The manager could use the information of the volatility of the large population to form an informative prior distribution for estimating the volatility of the small population with the aid of his/her prior knowledge about the relationship of the volatility between the large and smaller populations. The study of the influence of the sampling variation could reminds the manager that the posterior distribution could be shifted around from one side to another and such shift could be very large. In the next Chapter, we demonstrate the application with an empirical study by modelling the males in Scotland with the information of the England and Wales data.

For $\boldsymbol{\mu}_1$

Figure 6.7 includes the following curves:

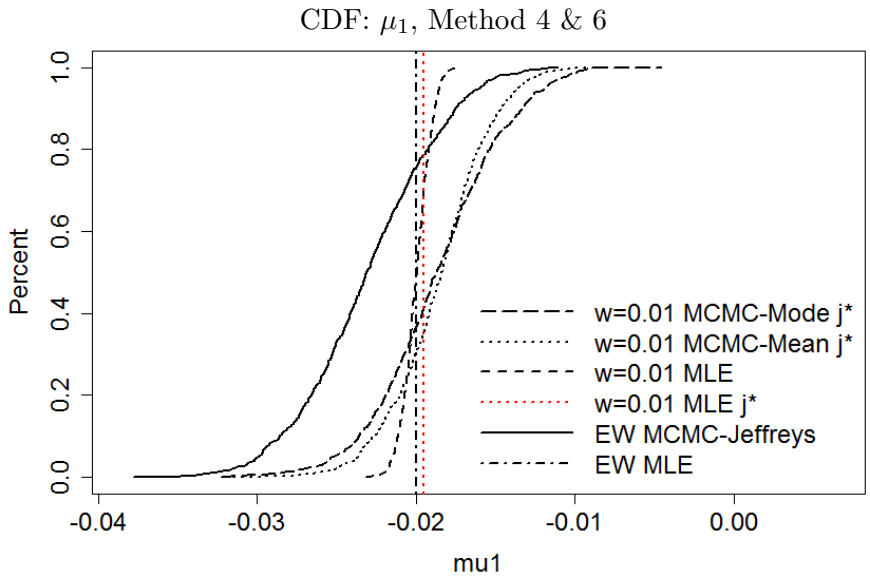


Figure 6.7: CDF: posterior distribution $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ of the simulated death scenario j^*, given the MCMC-Mean (Method 4, dotted curve) and the MCMC-Mode (Method 6, long-dashed curve). The solid curve and the vertical dot-dashed line are the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ and the true rate of the simulated death scenarios $\hat{\boldsymbol{\mu}}_1^{\text{EW}}$ respectively. The dashed curve is the CDF for $\hat{\boldsymbol{\mu}}_1^{w}$, the finite-sample MLEs of the N_1 simulated death scenarios. The red vertical line is $\hat{\boldsymbol{\mu}}_1^{w,j^*}$, the relative MLE for death scenario j^*.

- $w = 0.01$ MCMC-Mode j^*: the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ of the simulated death scenario j^*, given the prior setting of MCMC-Mode,

- $w = 0.01$ MCMC-Mean j^*: the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ of the simulated death scenario j^*, given the prior setting of MCMC-Mean,

- $w = 0.01$ MLE: the distribution $\hat{\boldsymbol{\mu}}_1^{w}$ of the finite-sample MLEs of the N_1 simulated death scenarios,

- $w = 0.01$ MLE j^*: the MLE $\hat{\boldsymbol{\mu}}_1^{w,j^*}$ for the simulated death scenario j^*,

- EW MCMC: the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ of the England and Wales data with Jeffreys prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$,

- EW MLE: the MLE $\hat{\boldsymbol{\mu}}_1^{\text{EW}}$ of the England and Wales data,

The MLEs for each death scenario and England and Wales data are calculated based on Equation (3.8) in Section 3.3.1, Chapter 3.

In Figure 6.7, we demonstrate the distribution $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ conditional on the MCMC-Mode (Method 6, long dashed CDF), which has approximately the same posterior mean and slightly higher posterior variance than the estimation conditional on the MCMC-Mean (Method 4, dotted CDF). The latter observation is consistent with our finding that the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(1,1)$ (MCMC-Mode) is shifted to the right hand side of MCMC-Mean and centred around the corresponding true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ in Figure 6.3. The reason for such consistency is shown in Figure 6.10 where we demonstrate the relationship between the median of $\sqrt{\tilde{\boldsymbol{V}}_{\epsilon}^{w}(1,1)}$ and the standard deviation of the $\tilde{\boldsymbol{\mu}}_1^w$ based on the corresponding distributions of the N_1 death scenarios. The linear relationship between the two implies that the posterior variance of $\tilde{\boldsymbol{\mu}}_1^w$ increases when the sampling variation shifts the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(1,1)$ higher. See the characteristic statistics in Table 6.4.

The posterior distribution $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$, conditional on both MCMC-Mode (long dashed CDF) and MCMC-Mean (dotted CDF), of the random walk drift $\boldsymbol{\mu}_1$ for the period effect $\boldsymbol{\kappa}^{(1)}$ of the death scenario j^* in Figure 6.7 is slightly shifted to the right hand side of both the true parameter $\hat{\boldsymbol{\mu}}_1^{\text{EW}}$ (vertical dotted-dashed line) and the relative MLE $\hat{\boldsymbol{\mu}}_1^{w,j^*}$ (dashed CDF) of the simulated death scenario j^*. A relatively larger right hand side shift can also be observed from the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ (solid CDF) of the England and Wales data to the $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ and the $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ has a relatively higher posterior variance compared with the variance of $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ (both MCMC-Mode and MCMC-Mean, see Table 6.4). Consistently, different level of anti-clockwise tiles can be observed in Figure 6.5a for the trajectory of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ (MCMC-Mean, blue lines) from the true parameter $\hat{\boldsymbol{\kappa}}^{(1),\text{EW}}$ (black line), posterior distribution $\tilde{\boldsymbol{\kappa}}^{(1),\text{EW}}$ of the England and Wales data (red lines) and the MLE $\hat{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ for the death scenario j^* (green line). The $\hat{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ is relatively more fluctuated around the true rate compared with what the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ is as we expected while $\tilde{\boldsymbol{\kappa}}^{(1),\text{EW}}$ clock-wisely tilts from $\hat{\boldsymbol{\kappa}}^{(1),\text{EW}}$ as we have already discussed in the previous chapter. A wider posterior fitted intervals can be observed for the England and Wales data compared with the death scenario j^*, which is consistent with both of our findings that the posterior variance of $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ is

slightly greater than that of the $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ and a higher posterior mean for the $\tilde{\boldsymbol{V}}_\epsilon^{\text{EW}}(1,1)$ of EW compared with the $\tilde{\boldsymbol{V}}_\epsilon^{w,j^*}(1,1)$ of the death scenario j^*. It is worth noticing that a wider fitted intervals for the $\boldsymbol{\kappa}^{(1)}$ does not necessarily mean (a) a more volatile drift or (b) on average a higher volatility of the $\boldsymbol{\kappa}^{(1)}$ separately as one could have a similar trajectory of $\tilde{\boldsymbol{\kappa}}^{(1)}$ with the $\tilde{\boldsymbol{\kappa}}^{(1),\text{EW}}$ by satisfying only one of the (a) and (b).

The reason of a narrower and less steep $\boldsymbol{\kappa}^{(1)}$ estimation for death scenario j^* compared with that of the England and Wales data (and not as steep as the true rate) in Figure 6.5a is because of the dominating influence of the time series prior for the latent parameters on the joint posterior distribution when the population size is small. Recall that the cohort effect $\boldsymbol{\gamma}^{(4)}$ is restricted to follow a zero mean-reverting AR(1) model. In Figure 6.38a, we demonstrate the credibility intervals of the posterior distribution $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ (MCMC-Mean, blue lines) of the death scenario j^*, which is, as we expect, almost horizontal and fluctuates around zero with relatively significant smoothness compared with $\hat{\boldsymbol{\gamma}}_{j^*}^{(4),w}$, the MLE of the death scenario j^* (green line). The entire trajectory of the $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ is clock-wisely tilted from the true parameter (black line) $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$ (excluding the MLEs for the short cohorts at the very early and late years) and looks exactly like a standard zero mean-reverting AR(1) process. On the other hand the distribution $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ of the England and Wales data (red lines) with a much steeper trajectory and wider credibility intervals looks relatively much less like an AR(1) process that reverts to zero compared with the death scenario j^*'s, as we expect. This is because for the small population $w = 0.01$, as we have mentioned the time series prior for the $\boldsymbol{\kappa}$s and $\boldsymbol{\gamma}$ dominates the posterior distribution and the strong restriction for the cohort effect from the AR(1) likelihood forces the cohort estimation to be more like the proposed AR(1) model that more strictly reverts to zero by horizontally tilting the trajectory and squeezing the posterior variance of the $\boldsymbol{\gamma}$'s estimation. On the other hand, such restriction is far less for the estimation $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ of the England and Wales data since the impact of the time series prior distributions are negligible compared with the Poisson likelihood when the population size is large. Therefore the trajectory of

$\tilde{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$ looks relatively far less like an AR(1) process with mean reverted to zero and in general has a steeper slope for the entire process and wider credibility intervals.

Correspondingly, the random walk likelihood for $\boldsymbol{\kappa}^{(1)}$ therefore tiles the estimation $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ of the death scenario j^* anti-clockwiselly from the true rate $\hat{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$. The strong influence of the likelihood for the $\boldsymbol{\kappa}^{(1)}$ given $w = 0.01$ restricts the estimation for $\boldsymbol{\kappa}^{(1)}$ more like a random walk process and hence relatively narrower credibility intervals for the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ compared with the $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$ of the England and Wales data when such restriction is much weaker when the influence of the random walk for $\boldsymbol{\kappa}^{(1)}$ is negligible on the joint posterior distribution for big population.

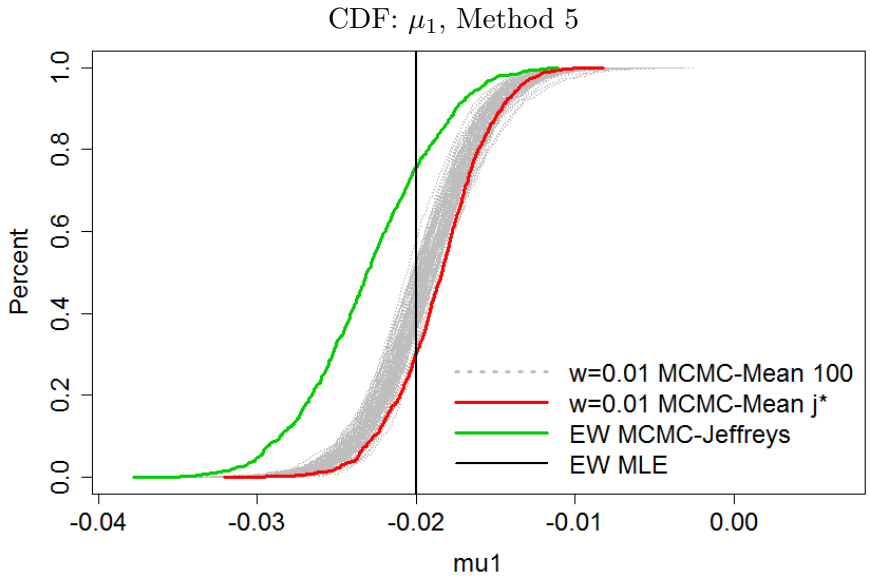


Figure 6.8: CDFs of $\tilde{\boldsymbol{\mu}}_1^{w,j'}$ for $j' = j_1, \ldots, j_{100}$ (MCMC-Mean, dotted curves), the posterior distributions for 100 randomly selected simulated death scenarios. The red curve is the CDF of $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ conditional on MCMC-Mean for death scenario j^*. The green curve and the vertical solid line are the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{\mathrm{EW}}$ and the MLE $\hat{\boldsymbol{\mu}}_1^{\mathrm{EW}}$ of the England and Wales data respectively.

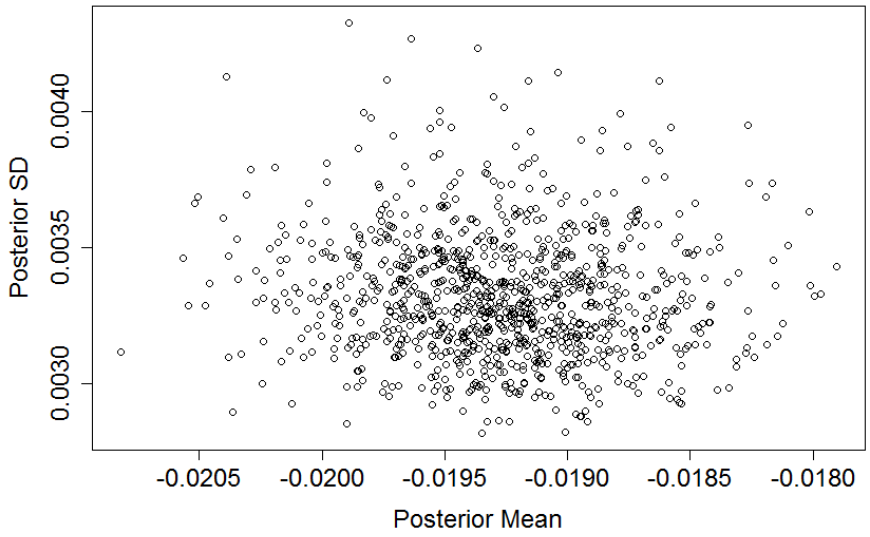
The influence of the sampling variation on the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{w,j'}$ for $j' = j_1, \ldots, j_{100}$ (MCMC-Mean, dotted curves) from one death scenario to another are demonstrated in Figure 6.8 according to the 100 selected death scenarios. Once again we can see that the sampling variation shifts the posterior distribution from one side to another with the true parameter $\hat{\boldsymbol{\mu}}_1^{\mathrm{EW}}$ (vertical solid line) slightly lower than the centre of the spread of the posterior distributions. The distribution $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$

of the death scenario j^* (MCMC-Mean, red CDF) is on the right tail of the spread implying that the shift due to the sampling variation could be large. Recall that we have concluded that the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ conditional on the MCMC-Mode is approximately the same with the distribution given the prior setting MCMC-Mean. We therefore would expect that without re-running the MCMC for all the N_1 simulated death scenarios conditional on the MCMC-Mode, the relative distribution of the posterior distributions based on Method 5 will be approximately the same with the spread demonstrated in Figure 6.8 conditional on the MCMC-Mean. The posterior distribution $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ for the England and Wales data (green CDF) is far on the left hand side of the spread as may be expected, which is consistent with the different level of the impact for the time series prior we just discussed on the joint posterior distribution. We therefore conclude that the estimation of $\tilde{\boldsymbol{\mu}}_1$ generated by the MCMC is not significantly different from the corresponding true rate, although the sampling variation shifts the posterior distribution from one side to another. On the other hand, the estimation for the small population is significantly higher than for the England and Wales data. See Table 6.2 for the characteristic statistics of the distribution of the posterior mean $\bar{\tilde{\boldsymbol{\mu}}}_1^w$ driven by the sampling variation.

In Figure 6.9 we demonstrate that, while the sampling variation shift the posterior distribution from one way to another, how it affects the standard deviation (upper) and the skewness (lower) of the posterior distribution based on N_1 simulated death scenarios' posterior standard deviations and skewness, conditional on MCMC-Mean. According to the patterns of the scatter plot, we could conclude that there is no obvious relationship between the posterior standard deviation/skewness and the sampling variation.

At last, we demonstrate the influence of the sampling variation on the distribution of the posterior distributions $\tilde{\kappa}_{t,j}^{(1),w}$ for $j = 1, \ldots, N_1$ in Figure 6.11 (MCMC-Mean, the dotted CDFs) for some selected years. As we expect, the sampling variation shifts the posterior distribution from one side to another. We conclude that

(a) Scatter Plot: Posterior Means vs SD



(b) Scatter Plot: Posterior Means vs Skewness

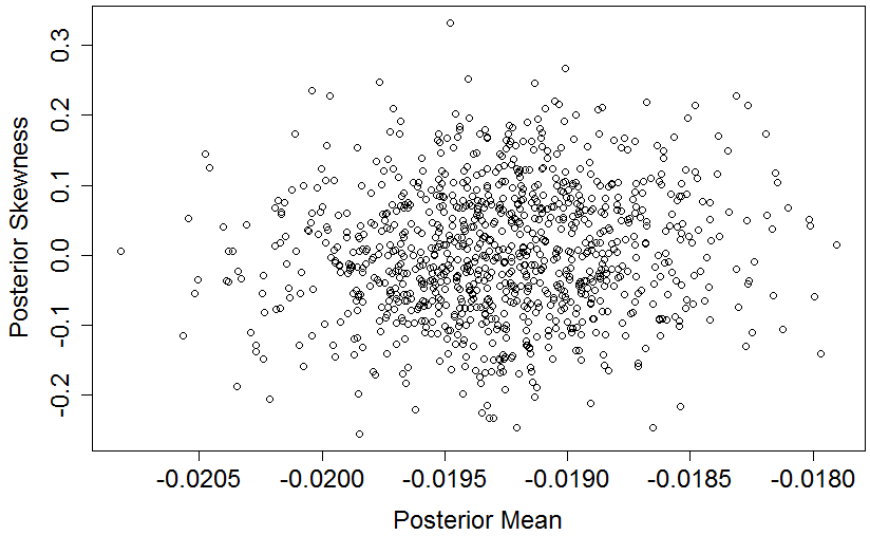


Figure 6.9: The scatter plot: the posterior mean vs. coefficient of skewness (lower); the posterior mean vs. standard deviation (upper) for $\boldsymbol{\mu}_1$ of $w = 0.01$, based on the posterior distributions $\tilde{\boldsymbol{\mu}}_1^{w,j}$ (MCMC-Mean) for $j = 1, \ldots, N_1$.

the spread of the distributions driven by the sampling variation is not significantly different from the respective true parameter (the vertical solid lines), although the shift due to the sampling variation could sometimes be large. The sampling variation shifting the the rest years of the $\boldsymbol{\kappa}^{(1)}$ are shown in Figure D.1 in the Appendix D.

				μ_1	μ_2	μ_3
Posterior	Mean	England and Wales	C	-2.30E-02	0.67E-04	3.34E-05
		$w = 0.01$	A	-1.86E-02	1.42E-04	2.14E-05
			B	-1.85E-02	1.56E-04	2.19E-05
	Std Dev	England and Wales	C	4.24E-03	1.90E-04	7.59E-06
		$w = 0.01$	A	3.98E-03	1.71E-04	10.37E-06
			B	3.11E-03	1.34E-04	8.35E-06
	Skewness	England and Wales	C	-0.0129	-0.1703	-0.0853
		$w = 0.01$	A	-0.0592	0.0861	0.0385
			B	-0.2166	-0.0843	-0.1480

Table 6.4: Characteristic statistics of: the posterior distribution of $\boldsymbol{\mu}$ for England and Wales data and selected death scenario with $w = 0.01$, during year 1961-2011, aged 50-89 last birthday. Note A = MCMC-Mode; B = MCMC-Mean; C = Jeffrey's Prior for $\boldsymbol{V_\epsilon}$ for modelling England and Wales data.

For $\boldsymbol{\kappa}^{(2)}$, $\boldsymbol{V_\epsilon}(2,2)$

Similar results can be observed for estimating $\boldsymbol{V_\epsilon}(2,2)$, the volatility of the period effect $\boldsymbol{\kappa}^{(2)}$ and the interpretations for the observations (unless otherwise discussed) can be referred to the discussions we did for $\boldsymbol{V_\epsilon}(1,1)$. Figure 6.12a demonstrates that the distribution of the finite-sample MLEs $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j^*}(2,2)$ (dashed CDF) for the N_1 simulated death scenarios is far on the right hand side of the respective true parameter $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$ (vertical dotted-dashed line) and such over-estimation is on scale relatively greater than the two-stage approach over-estimating the $\boldsymbol{V_\epsilon}(1,1)$ (see Figure 6.3), implying the MLE for the $\boldsymbol{V_\epsilon}(2,2)$ of the small population is relatively more sensitive to the greater sampling variation than for the $\boldsymbol{V_\epsilon}(1,1)$.

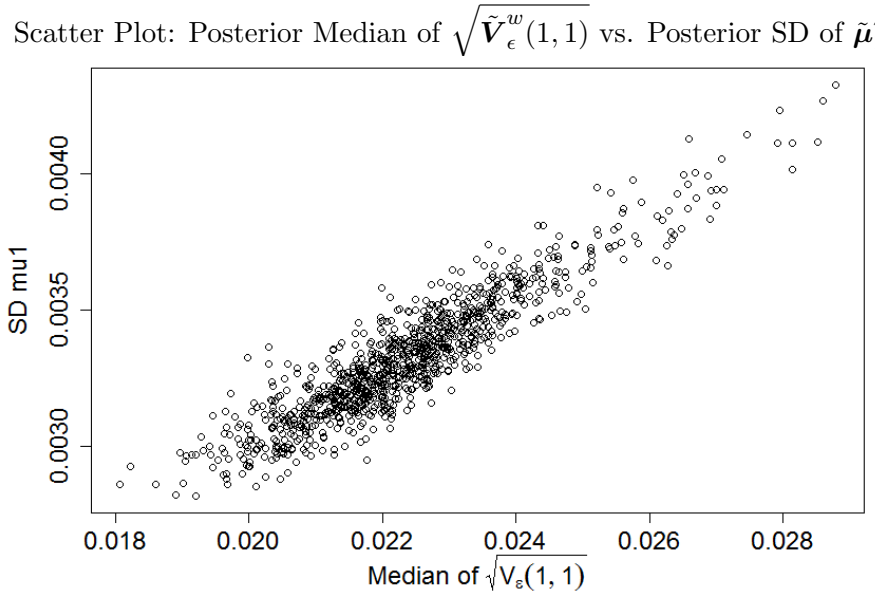


Figure 6.10: Scatter Plot: the posterior median of the $\sqrt{\tilde{\boldsymbol{V}}_{\epsilon}^{w}(1,1)}$ (horizontal axis) vs. the standard deviation of $\tilde{\boldsymbol{\mu}}_{1}^{w}$ (vertical axis) based on the corresponding posterior distributions of the N_1 death scenarios.

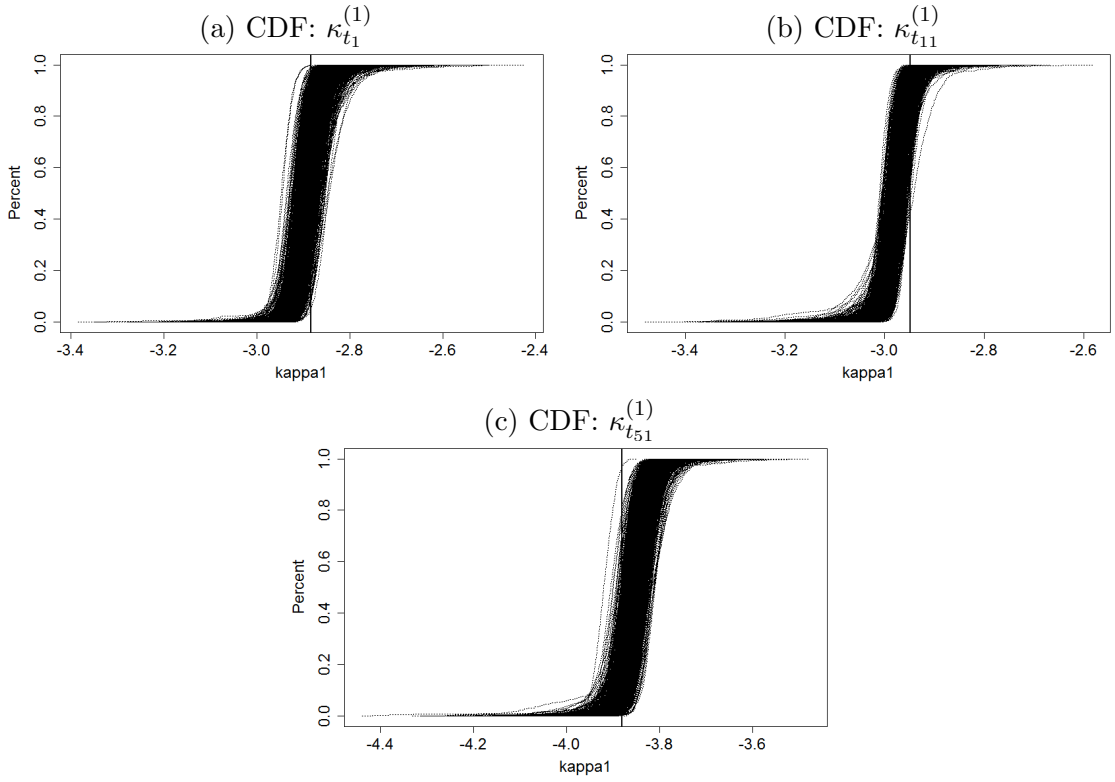
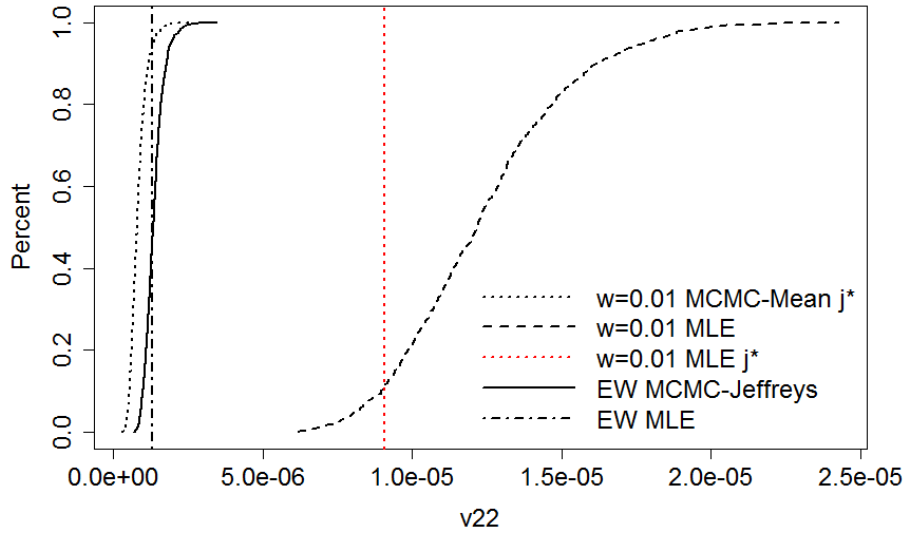


Figure 6.11: The CDFs of posterior distribution for $\boldsymbol{\kappa}^{(1)}$ with selected years according to Method 5. Note that each dotted curve represents the posterior distribution for one randomly selected death scenario and all the N_1 CDFs are plotted.

(a) CDF: V_{22}, Method 4, MCMC-Mean



(b) CDF: V_{22}, Method 4 MCMC-Mean & Method 6 MCMC-Mode

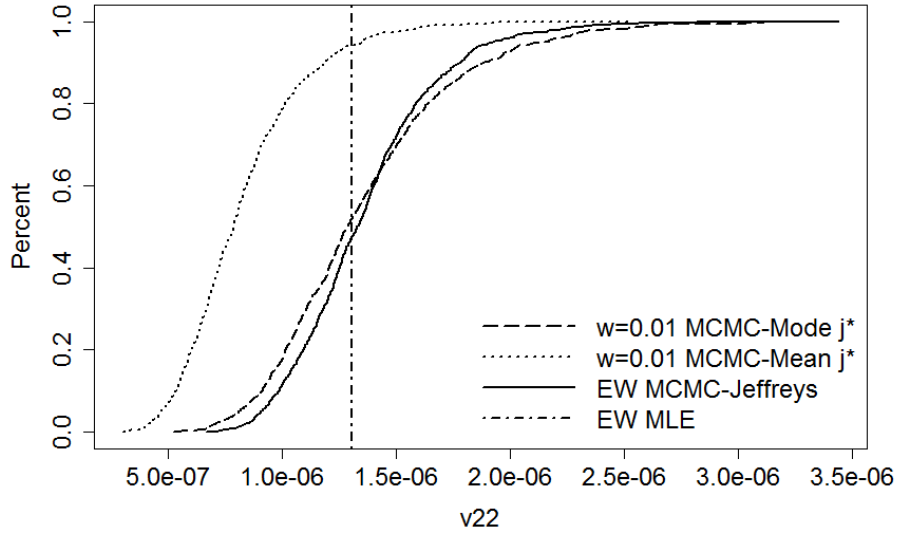
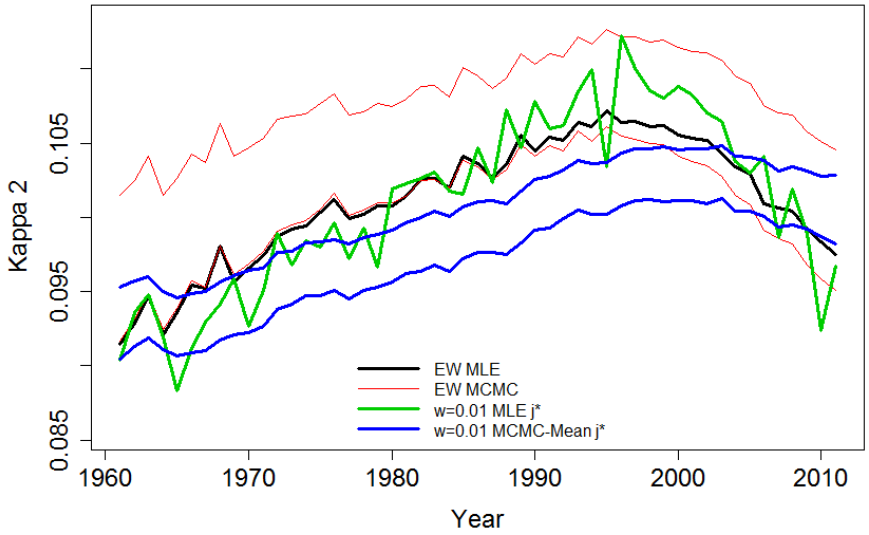


Figure 6.12: CDF: posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ of the simulated death scenario j^*, given the MCMC-Mean (upper and lower, dotted curve) and the MCMC-Mode (lower, long-dashed curve). The solid curve and the vertical dot-dashed line are the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(2,2)$ and the true rate of the simulated death scenarios $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$ respectively. The dashed curve (upper) is the CDF for $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(2,2)$, the finite-sample MLEs of the N_1 simulated death scenarios. The red vertical line is the relative MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j^*}(2,2)$ for the death scenario j^*.

We removed the CDF for the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$ in Figure 6.12b for a clearer demonstration of the posterior distributions, where we can see that on the contrast the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ for the death scenario j^* conditional on both MCMC-Mode (Method 6, long-dashed CDF) and the MCMC-Mean (Method 4, dotted CDF) provides a much improved estimation for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$.

(a) Historical $\boldsymbol{\kappa}^{(2)}$, Method 4 MCMC-Mean



(b) Historical $\boldsymbol{\kappa}^{(2)}$, Method 4 MCMC-Mean versus Method 6 MCMC-Mode

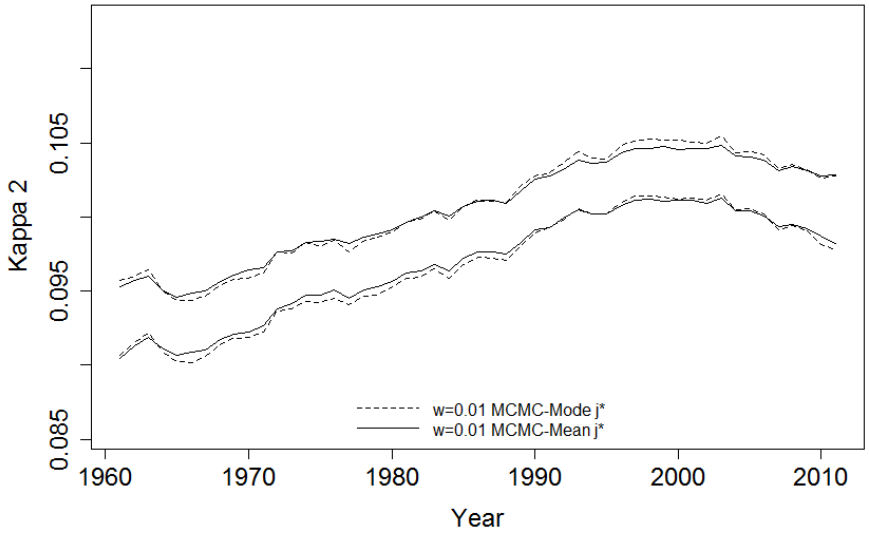


Figure 6.13: Credibility Interval: posterior distribution $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ of the simulated death scenario j^*, given the MCMC-Mean (Method 4, upper blue lines; lower solid lines) and the MCMC-Mode (Method 6, lower dashed lines). The red lines and the black line in the upper figure is posterior distribution $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ and the MLE $\hat{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ respectively for the England and Wales data. The green line is $\hat{\boldsymbol{\kappa}}_{j^*}^{(2),w}$, the MLE of death scenario j^*. Note that the upper and lower bound are the 95% and 5% quantile of the posterior distribution.

In particular, the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ conditional on the MCMC-Mode (long dashed CDF) in Figure 6.12b is shifted to the right hand side of the estimation $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ conditional on the MCMC-Mean (dotted CDF). The CDF of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ (MCMC-Mode) is approximately centred around the respective true parameter $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(2,2)$, implying an unbiased estimation for the volatility $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$

of the death scenario j^*. We can see that the CDF conditional on the MCMC-Mode for death scenario j^* is approximately the same with the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(2,2)$ of the England and Wales data, with a relatively higher posterior variance for the former distribution as may be expected. Some characteristic statistics for the prior and posterior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ are shown in Table 6.1 and 6.3 respectively with respect to the two populations and the types of the prior distributions.

On the other hand, the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ conditional on the MCMC-Mean lies slightly on the left hand side of the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$ though, implying a non-significant difference between the posterior estimation (MCMC-Mean) and the true parameter for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ of the death scenario j^*. We could also see that for the England and Wales data, the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(2,2)$ (solid CDF) is approximately centred around the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$ and higher than $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ conditional on the MCMC-Mean. The posterior variances of the two populations are visually of non-significant difference, conditional on both prior settings for death scenario j^*. See Table 6.1 for detailed characteristic statistics.

Consistently, we could observe relatively a bit more fluctuated trajectory of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ conditional on the MCMC-Mode (dashed lines) in Figure 6.13b compared with the estimation given MCMC-Mean (solid lines) due to a higher estimation for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ of death scenario j^*, though the two trajectories are approximately the same. In Figure 6.13a, a relatively much more smoothness could be observed from the trajectory of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ conditional on MCMC-Mean (blue lines) compared with the more fluctuated MLE $\hat{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ (green line). Visually the trajectory of the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ (MCMC-Mean) is also a bit more smoothed than the true parameter $\hat{\boldsymbol{\kappa}}^{(2),\text{EW}}$, which is consistent with our finding in Figure 6.12b that the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$ lies approximately on the 90% quantile of the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$. We therefore conclude that, based on Figure 6.13b, the estimation for the period effect $\boldsymbol{\kappa}^{(2)}$ generated by the MCMC (conditional on both MCMC-Mode and MCMC-Mean) is more smoothed than the MLE for the

death scenario j^* when the large Poisson noise is balanced by the random walk likelihood for death scenario j^*.

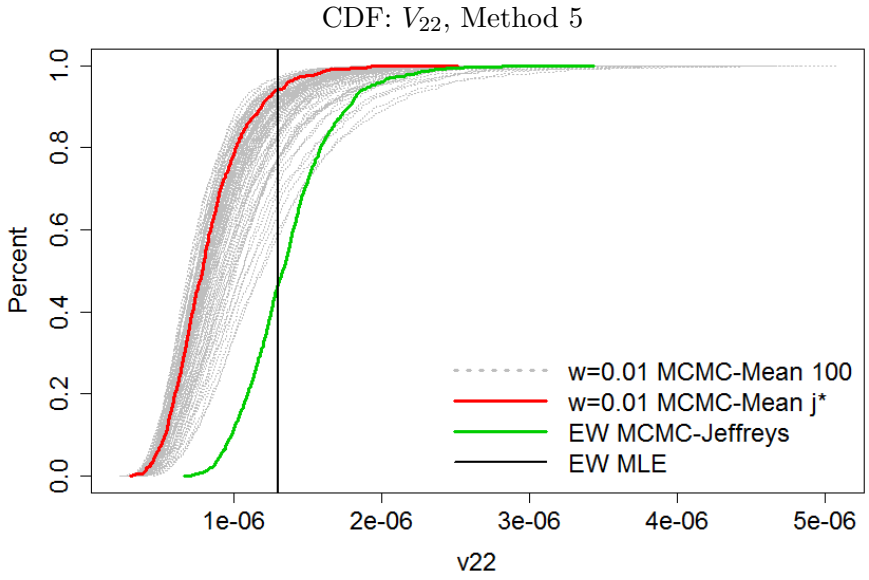


Figure 6.14: CDFs of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}(2,2)$ for $j' = j_1, \ldots, j_{100}$ (MCMC-Mean, dotted curves), the posterior distributions for 100 randomly selected simulated death scenarios. The red curve is the CDF of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ conditional on MCMC-Mean for death scenario j^*. The green curve and the vertical solid line are the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(2,2)$ and the MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$ of the England and Wales data respectively.

The spread of the posterior distributions $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}(2,2)$ conditional on the MCMC-Mean for $j' = j_1, \ldots, j_{100}$ (dotted CDFs) for the 100 selected death scenarios are demonstrated in Figure 6.14. See the characteristic statistics of the spread of the posterior means in Table 6.2. Unsurprisingly, the sampling variation shifts the posterior distribution from one side to another and we could see that the shift varies from one death scenario to another and could be very large.

Visually most of the distributions have their 90% quantiles lower than the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$, indicating that the posterior estimation conditional on the MCMC-Mean for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ of $w = 0.01$ is not significantly differed from the true rate. The posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(2,2)$ for the England and Wales data (green CDF) roughly lies on the right hand side of the spread of the distributions (MCMC-Mean), implying that the estimation with the prior setting of the MCMC-Mean for $w = 0.01$ is smaller than the relative estimation of the England and Wales data.

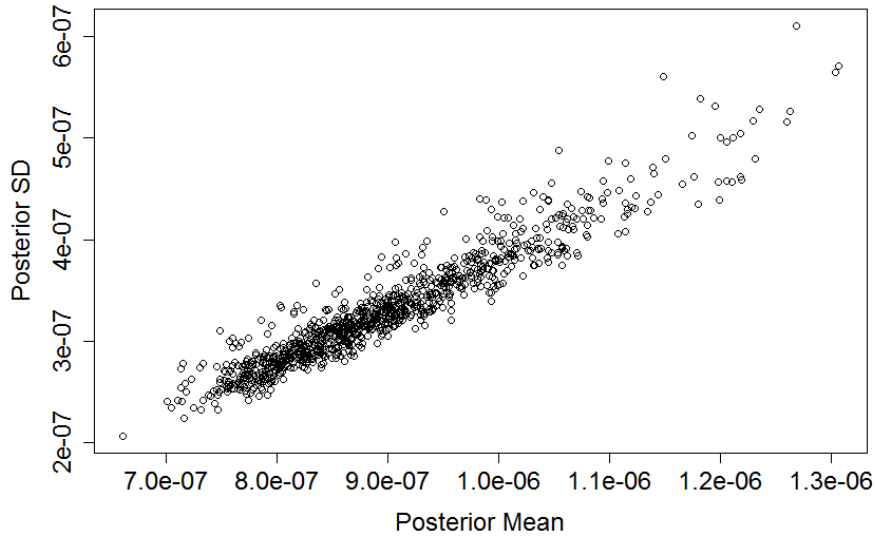
However, we have concluded that the prior setting MCMC-Mode shifts the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ to be centred around the true rate and approximately the same with the $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(2,2)$ of the England and Wales data. It is worth noticing that the $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ with the prior setting MCMC-Mean (red curve) is on the centre of the spread of distributions in Figure 6.14 ($\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}(2,2)$ is not symmetric). Therefore without re-running the MCMC for the N_1 death scenarios conditional on the MCMC-Mode, we expect that the spread of the posterior distributions will be shifted to the left just like the death scenario j^* with the sampling variation shifting the posterior distribution from one side to another centred around the true rate and the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(2,2)$ of the England and Wales data is approximately centre of the spread.

We therefore conclude that by understanding the shifting influence of the sampling variation, the MCMC conditional on the prior setting MCMC-Mode generates an unbiased estimator $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(2,2)$ for the volatility of the $\boldsymbol{\kappa}^{(2)}$ and there is no significant difference between the $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(2,2)$ and the $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(2,2)$ of the England and Wales data.

For $\boldsymbol{\mu}_2$

In Figure 6.16, we demonstrate the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ of the death scenario j^* conditional on the prior setting MCMC-Mode (long dashed CDF), which is of no obvious difference from the CDF conditional on the MCMC-Mean (dotted curve) as we expect, with approximately the same posterior mean and a slightly higher posterior variance due to the higher estimation $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(2,2)$ conditional on MCMC-Mode. Figure 6.19 demonstrates a relatively much weaker positive linear relationship between the median of $\sqrt{\tilde{\boldsymbol{V}}_{\epsilon}^{w}(2,2)}$ and the variance of $\tilde{\boldsymbol{\mu}}_2^{w}$, compared with Figure 6.10. See the characteristic statistics in Table 6.4, where the posterior mean of $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ is a bit increased from the MCMC-Mean to the MCMC-Mode. Consistently a relatively more steeper (negligible) trajectory of the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ conditional on the

(a) Scatter Plot: Posterior Means vs SD



(b) Scatter Plot: Posterior Means vs Skewness

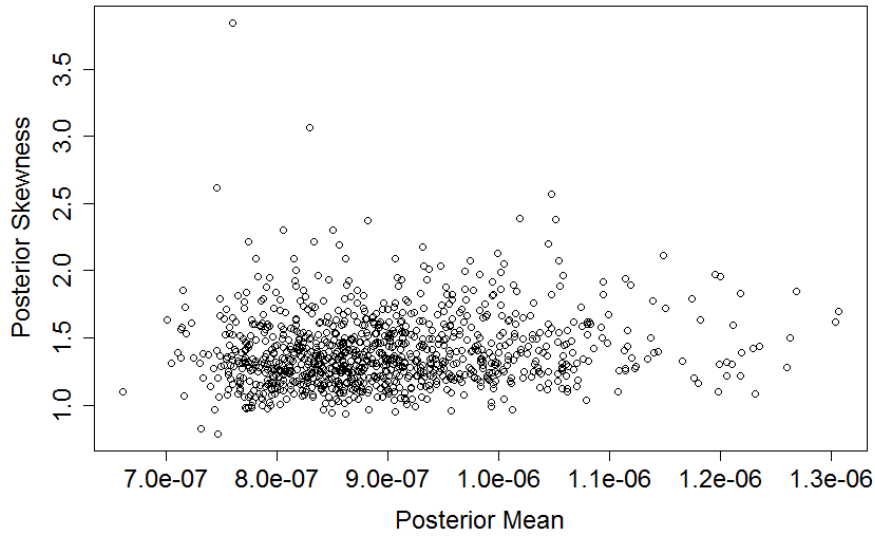


Figure 6.15: The scatter plot: the posterior mean vs. coefficient of skewness (lower); the posterior mean vs. standard deviation (upper) for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ of $w = 0.01$, based on the posterior distributions $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}$ for $j = 1, \ldots, N_1$.

MCMC-Mode (dashed lines) with barely no changes to the width of the credibility intervals can be observed in Figure 6.13b, compared with the trajectory conditional on the MCMC-Mean (solid lines).

In Figure 6.16, we could see that $\hat{\boldsymbol{\mu}}_2^{\mathrm{EW}}$, the relative true parameter for death scenario j^* lies approximately at 40% quantile of the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ conditional on the MCMC-Mode (long-dashed CDF) and MCMC-Mean (dotted CDF). We therefore conclude that there is no significant difference between the posterior

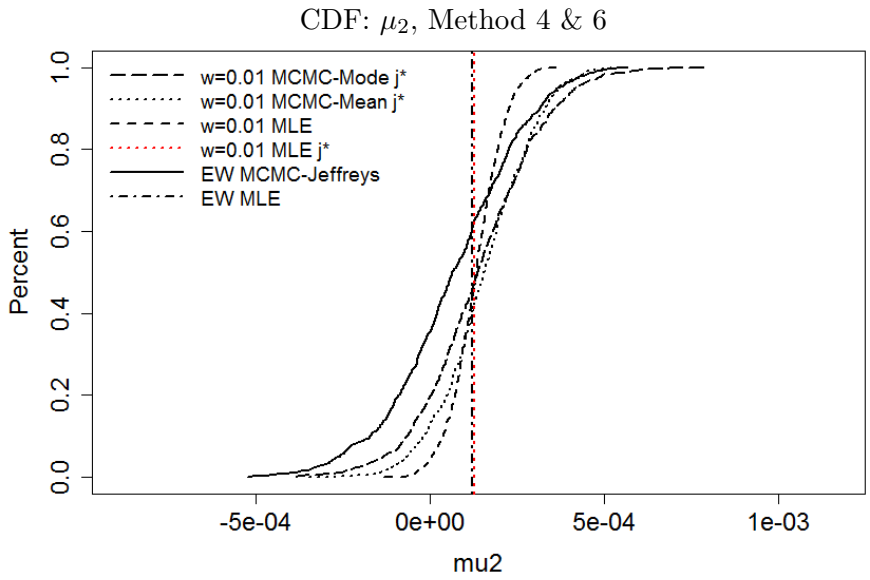


Figure 6.16: CDF: posterior distribution $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ of the simulated death scenario j^*, given the MCMC-Mean (Method 4, dotted curve) and the MCMC-Mode (Method 6, long-dashed curve). The solid curve and the vertical dot-dashed line are the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{\text{EW}}$ and the true rate of the simulated death scenarios $\hat{\boldsymbol{\mu}}_2^{\text{EW}}$ respectively. The dashed curve is the CDF for $\hat{\boldsymbol{\mu}}_2^{w}$, the finite-sample MLEs of the N_1 simulated death scenarios. The red vertical line is the relative MLE $\hat{\boldsymbol{\mu}}_2^{w,j^*}$ for the death scenario j^*.

estimation $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ (MCMC-Mode & MCMC-Mean) for $\boldsymbol{\mu}_2$ of the death scenario j^* and its relative true parameter.

On the other hand, the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{\text{EW}}$ for the England and Wales data (solid CDF) is on the left hand side of the CDFs for the $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ (MCMC-Mode & MCMC-Mean) with a little higher posterior variance, implying a smaller estimated drift for the $\boldsymbol{\kappa}^{(2)}$ of the England and Wales data. However, it is worth noticing that the gap between the posterior distributions of the $\boldsymbol{\mu}_2$ for the two populations is relatively much smaller compared with the difference between the $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ and $\tilde{\boldsymbol{\mu}}_1^{w,j^*}$ (MCMC-Mode & MCMC-Mean). See Table 6.4 for the characteristic statistics of the posterior distributions for the $\boldsymbol{\mu}_2$ of the two populations.

Recall the trajectories of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ conditional on the prior settings of MCMC-Mode and MCMC-Mean are approximately the same (Figure 6.13b). As expected, an anti-clockwise tile can be observed in Figure 6.13a for the estimation $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ conditional

on MCMC-Mean (blue lines) from its true rate, MLE $\hat{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ (black line) as well as the posterior distribution $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ (red lines) of the England and Wales data.

The credibility intervals of the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ is narrower than the $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ as we expect with the same reason we have discussed for the trajectory of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$. However, it is worth noticing that the difference of the width of the intervals for the two populations is also not as large as the difference between the credibility intervals for the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ and the $\tilde{\boldsymbol{\kappa}}^{(1),\mathrm{EW}}$. The reason for these observations is once again that without directly applying the identifiability constraints to the latent parameters, we restrict parameters with the time series models whose likelihood dominates the joint posterior distribution and restricts the latent parameter estimation to be more like the corresponding time series model. The random walk model for the $\boldsymbol{\kappa}^{(2)}$ therefore tiles the posterior estimation $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$, especially for those latest years (e.g. 1990-2011 approximately), such that it is restricted to be a random walk process with a positive drift. We thus have a steeper and relatively more linear trajectory of the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ than the shape of the $\hat{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$ and the $\tilde{\boldsymbol{\kappa}}^{(2),\mathrm{EW}}$.

However, the restriction from the random walk likelihood for the $\boldsymbol{\kappa}^{(2)}$ is not as strict as from the corresponding likelihood for the $\boldsymbol{\kappa}_1$ according to the smaller scaled value of the estimation for the $\boldsymbol{\kappa}^{(2)}$ due to the characteristic of the M7 model (i.e. the term $x - \bar{x}$ with the $\boldsymbol{\kappa}^{(2)}$). We therefore have the interval of the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ (MCMC-Mode & MCMC-Mean) that is relatively not much narrower than that of the England and Wales data as well as a relatively less strong tilt to the trajectory of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ and consistently a smaller difference between the $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ (MCMC-Mode & MCMC-Mean) and the $\tilde{\boldsymbol{\mu}}_2^{\mathrm{EW}}$.

We demonstrate the spread of the posterior distributions of the 100 selected death scenarios $\tilde{\boldsymbol{\mu}}_2^{w,j'}$ for $j' = j_1, \ldots, j_{100}$ (dotted CDFs) conditional on MCMC-Mean driven by the sampling variation in Figure 6.17. Unsurprisingly, the sampling variation shifts the posterior distribution from one side to another with the true rate $\hat{\boldsymbol{\mu}}_2^{\mathrm{EW}}$ (vertical solid line) on the centre of the spread of the distributions. The CDF

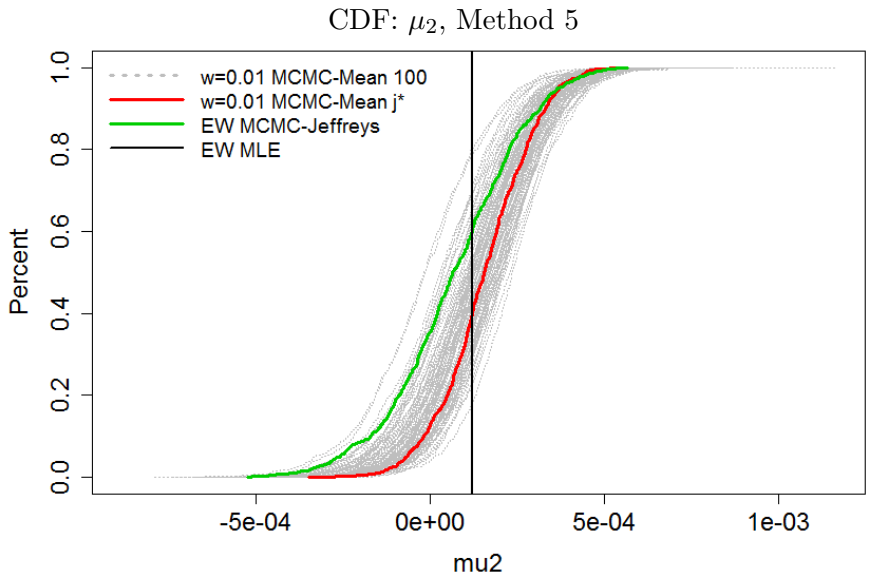


Figure 6.17: CDFs of $\tilde{\boldsymbol{\mu}}_2^{w,j'}$ for $j' = j_1, \ldots, j_{100}$ (MCMC-Mean, dotted curves), the posterior distributions for 100 randomly selected simulated death scenarios. The red curve is the CDF of $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ conditional on MCMC-Mean for death scenario j^*. The green curve and the vertical solid line are the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{\text{EW}}$ and the MLE $\hat{\boldsymbol{\mu}}_2^{\text{EW}}$ of the England and Wales data respectively.

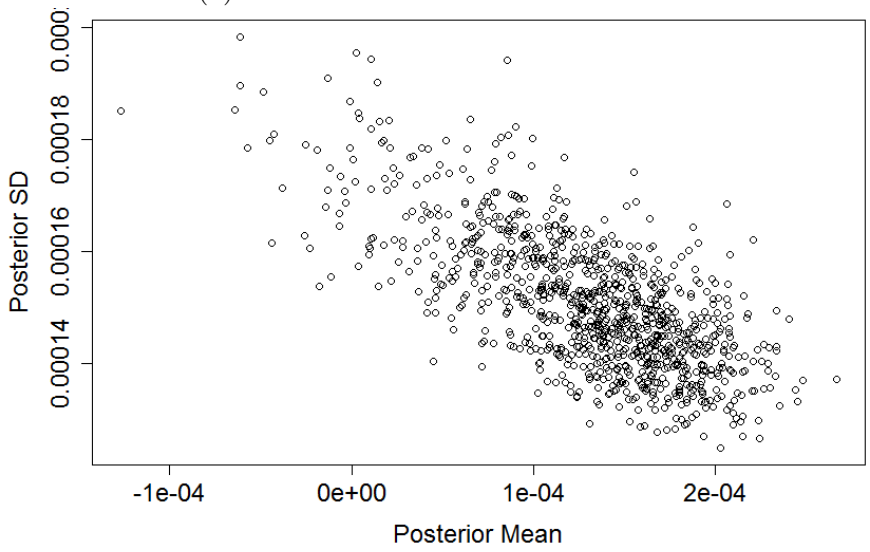
for the $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ conditional on the MCMC-Mean (red curve) for the death scenario j^* is slightly on the right hand side of the centre of the spread. The shifting effect of the sampling variation varies from one death scenario to another and as it is observed such shift could be very large.

Recall that the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{w,j^*}$ conditional on the MCMC-Mode is approximately the same with the distribution given the MCMC-Mean (Figure 6.16). We therefore would also expect that without re-running the MCMC for all the N_1 simulated death scenarios conditional on the MCMC-Mode, the relative distribution of the posterior distributions based on Method 5 will be similar with the spread demonstrated in Figure 6.17. The posterior distribution $\tilde{\boldsymbol{\mu}}_2^{\text{EW}}$ of the England and Wales data is a bit on the left hand side of the centre of the spread, implying a non-significant difference on the average level of the MCMC estimation between the benchmark population and the simulated death scenarios conditional on MCMC-Mean (and MCMC-Mode).

We therefore conclude that according to the sampling variation's shifting effect on

the posterior distribution of the simulated death scenarios, the posterior estimation $\tilde{\boldsymbol{\mu}}_2^w$ is not significantly differed from the respective true parameter and further, the posterior estimation $\tilde{\boldsymbol{\mu}}_2^{\mathrm{EW}}$ of the England and Wales data is approximately the same with the $\tilde{\boldsymbol{\mu}}_2^w$, given either the MCMC-Mode or MCMC-Mean.

(a) Scatter Plot: Posterior Means vs SD



(b) Scatter Plot: Posterior Means vs Skewness

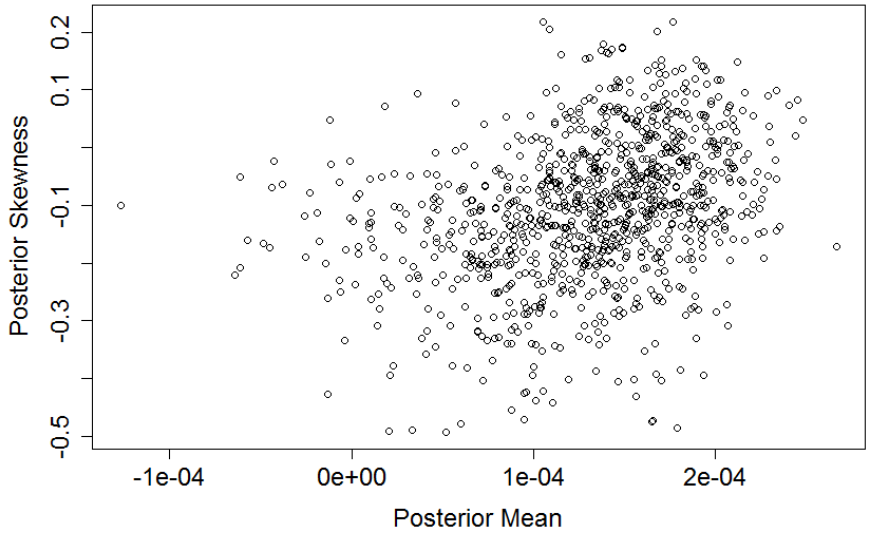


Figure 6.18: The scatter plot: the posterior mean vs. coefficient of skewness (lower); the posterior mean vs. standard deviation (upper) for $\boldsymbol{\mu}_2$ of $w = 0.01$, based on the posterior distributions $\tilde{\boldsymbol{\mu}}_2^{w,j}$ (MCMC-Mean) for $j = 1, \ldots, N_1$.

The characteristic statistics of the spread of the posterior means $\bar{\tilde{\boldsymbol{\mu}}}_2^w$ (MCMC-Mean) are shown in Table 6.2. We demonstrate the influence of the sampling variation on the standard deviation (upper) and the coefficient of skewness (lower) of the

posterior distribution based on the posterior distributions of the N_1 death scenarios in Figure 6.18. In general, the posterior standard deviation is approximately not linearly related to the corresponding mean when the mean of $\tilde{\boldsymbol{\mu}}_2^w$ (MCMC-Mean) is not shifted below 0.5×10^{-4} (roughly) in Figure 6.18a. On the other hand, no linear relationship could be observed between the mean and the skewness of the posterior distribution $\tilde{\boldsymbol{\mu}}_2^w$ (MCMC-Mean) in Figure 6.18b. We therefore could conclude that in general the sampling variation has no significant effect on the shape of the posterior distribution $\tilde{\boldsymbol{\mu}}_2^w$.

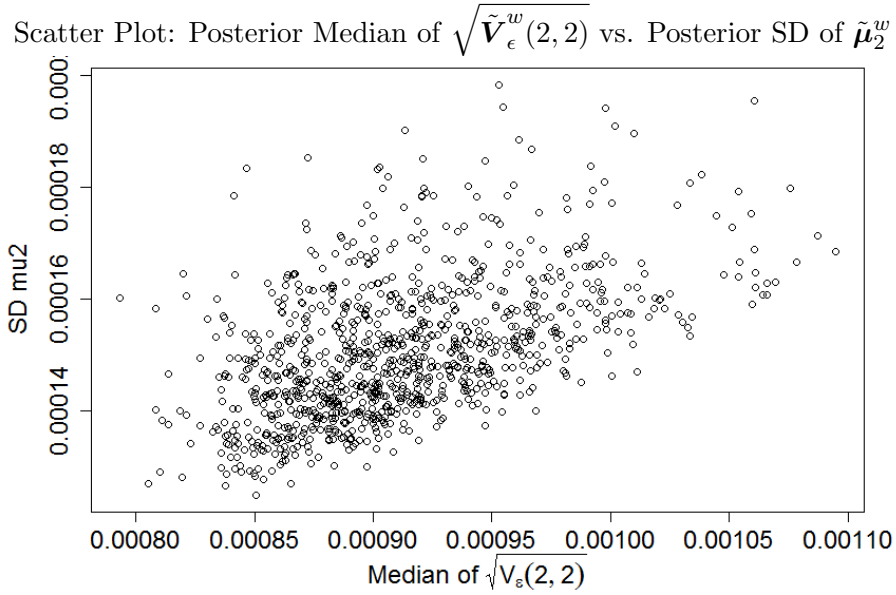


Figure 6.19: Scatter Plot: the posterior median of the $\sqrt{\tilde{\boldsymbol{V}}_\epsilon^w(2,2)}$ (horizontal axis) vs. the standard deviation of $\tilde{\boldsymbol{\mu}}_2^w$ (vertical axis) based on the corresponding posterior distributions (MCMC-Mean) of the N_1 death scenarios.

At last, we demonstrate the influence of the sampling variation shifting the posterior distribution of $\kappa_t^{(2)}$ (MCMC-Mean) with some years for the small populations (dotted CDFs) in Figure 6.20. The spread of the distributions in Figure 6.20b for the year 1991 is in general on the left hand side of the MLE $\hat{\kappa}_{t_{31}}^{(2),\mathrm{EW}}$ of the England and Wales data, which is consistent with our discussion that the estimation of the $\boldsymbol{\kappa}^{(2)}$ for the small population is restricted by the time series prior and hence more like a random walk process with a positive drift. The sampling variation shifting the other years of the $\boldsymbol{\kappa}^{(2)}$ are shown in Figure D.2 in the Appendix D.

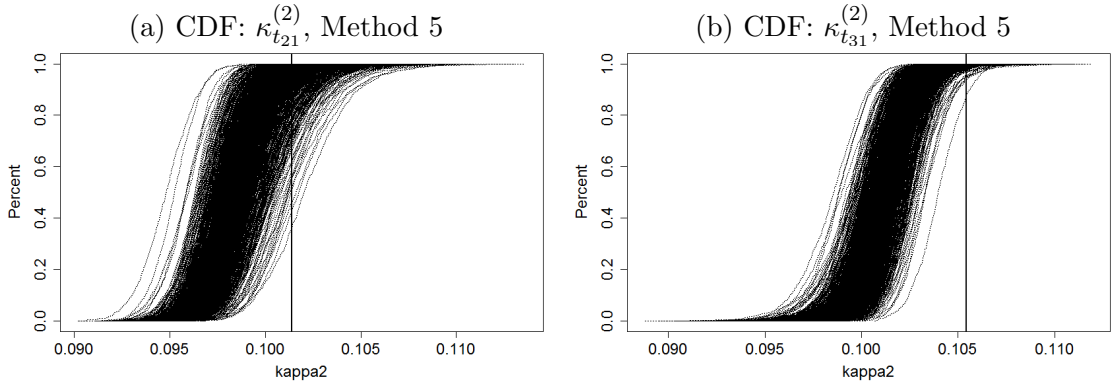


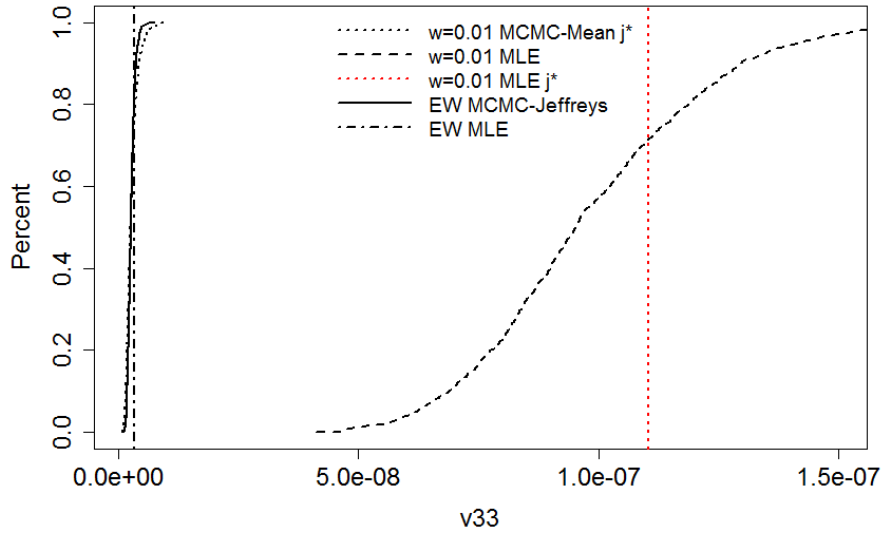
Figure 6.20: The CDF of posterior distribution for $\boldsymbol{\kappa}^{(2)}$ (MCMC-Mean) with selected years according to Method 5. Note that each dotted curve represents the posterior distribution for one randomly selected death scenario and all the N_1 CDFs are plotted.

For $\boldsymbol{\kappa}^{(3)}$, $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$

Similar results can be observed for the estimation of the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$, the volatility of the third period effect $\boldsymbol{\kappa}^{(3)}$ and once again the interpretations of the observations can be referred to the discussion of the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$, unless otherwise discussed. In Figure 6.21a, only part of the distribution of the finite-sample MLEs $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}(3,3)^{w}$ (dashed CDF) for the N_1 simulated death scenarios is demonstrated to illustrate that the two-stage approach greatly over-estimated the volatility of the $\boldsymbol{\kappa}^{(3)}$ and hence totally obscures the signal of the true volatility for the small population $w = 0.01$. It is worth noticing that the mean of the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{w}(3,3)$ (0.99×10^{-7}) is approximately 30 times greater than the respective true parameter $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{\mathrm{EW}}(3,3) = 3,3 \times 10^{-9}$ (vertical dot-dashed line), the MLE the England and Wales data, implying that estimating with the two-stage approach for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ is relatively most influenced by the large sampling variation, compared with the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$. The the characteristic statistics of the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{w}(3,3)$ in Table 6.5.

We removed the CDF for the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{\mathrm{EW}}(3,3)$ in Figure 6.21b for a clearer demonstration of the posterior distributions with respect to the population size. Unsurprisingly, the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ for the death scenario j^* conditional on both the MCMC-Mode (long-dashed CDF) and the MCMC-Mean (dotted CDF)

(a) CDF: V_{33}, Method 4, MCMC-Mean



(b) CDF: V_{33}, Method 4, MCMC-Mean versus Method 6, MCMC-Mode

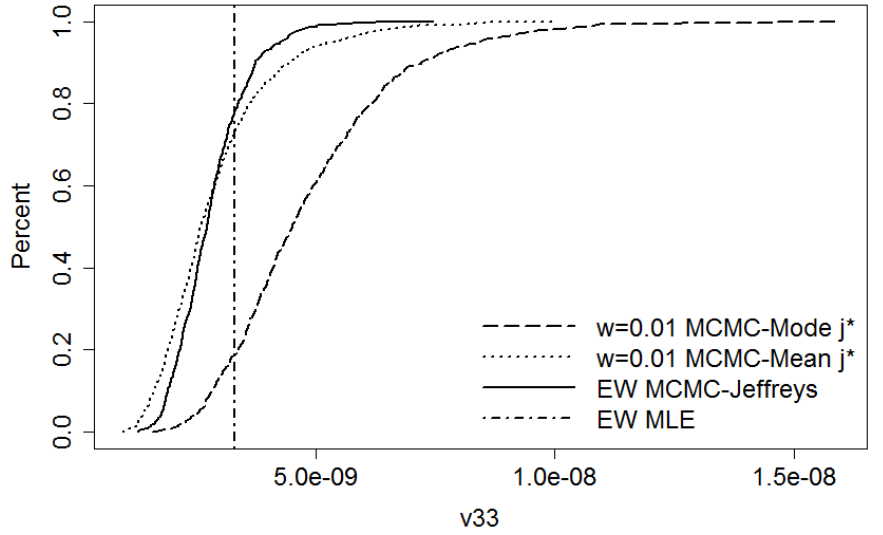


Figure 6.21: CDF: posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ of the simulated death scenario j^*, given the MCMC-Mean (upper and lower dotted curve) and the MCMC-Mode (lower, long-dashed curve). The solid curve and the vertical dot-dashed line are the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(3,3)$ and the true rate of the simulated death scenarios $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(3,3)$ respectively. The dashed curve is the CDF for $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(3,3)$, the finite-sample MLEs of the N_1 simulated death scenarios. The red vertical line is the relative MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j^*}(3,3)$ for the death scenario j^*.

provides a much improved estimation for the corresponding volatility, compared with the two-stage approach.

In particular, the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ conditional on the MCMC-Mode (long-dashed CDF) in Figure 6.21b has its posterior variance greater than

	$\hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{w}(3,3)$	$\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$
Mean	9.82×10^{-8}	0.29×10^{-8}
Standard Deviation	2.47×10^{-8}	0.12×10^{-8}
Quantile		
2.5%	5.70×10^{-8}	0.13×10^{-8}
50%	9.52×10^{-8}	0.26×10^{-8}
97.5%	15.21×10^{-8}	0.61×10^{-8}

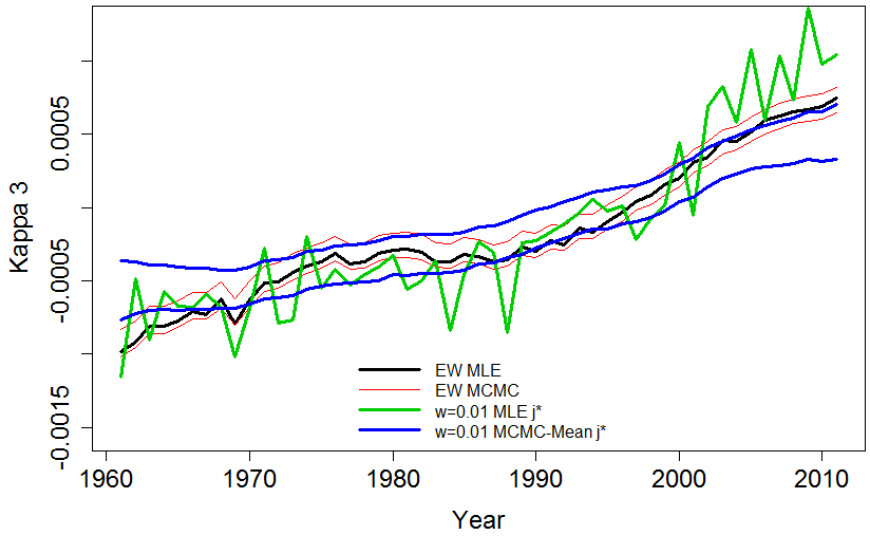
Table 6.5: The mean, standard deviation, 2.5%, 50%, and 97.5% quantiles of $\hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{w}(3,3)$ and $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ (MCMC-Mean).

both the $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ (MCMC-Mean) and $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(3,3)$ of the England and Wales data and its central estimation slightly on the right hand side of the true parameter, though the shift is not large. The true parameter $\hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{\mathrm{EW}}(3,3)$ is approximately on the 20% quantile of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ (MCMC-Mode). See characteristic statistics in Table 6.1. It implies that, as may be expected according to the discussion in the previous paragraph, the posterior distribution of the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ for the death scenario j^* is relatively more sensitive to the switch of the prior distribution for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$, compared with the posterior estimation for the volatilities of the other two period effects.

The true parameter is once again slightly on the right hand side of the centre of the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ conditional on (approximately 80% quantile), implying a non-significant difference. On the other hand, the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(3,3)$ (solid CDF) of the England and Wales data is on average at the same level with the distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ (MCMC-Mean), while the latter distribution is of a bit higher posterior variance. See the characteristic statistics of the posterior and the prior distributions of the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ in Table 6.1 and 6.3 respectively, with

respect to the population size.

(a) Historical $\boldsymbol{\kappa}^{(3)}$, Method 4, MCMC-Mean



(b) Historical $\boldsymbol{\kappa}^{(3)}$, Method 4, MCMC-mean versus Method 6, MCMC-Mode

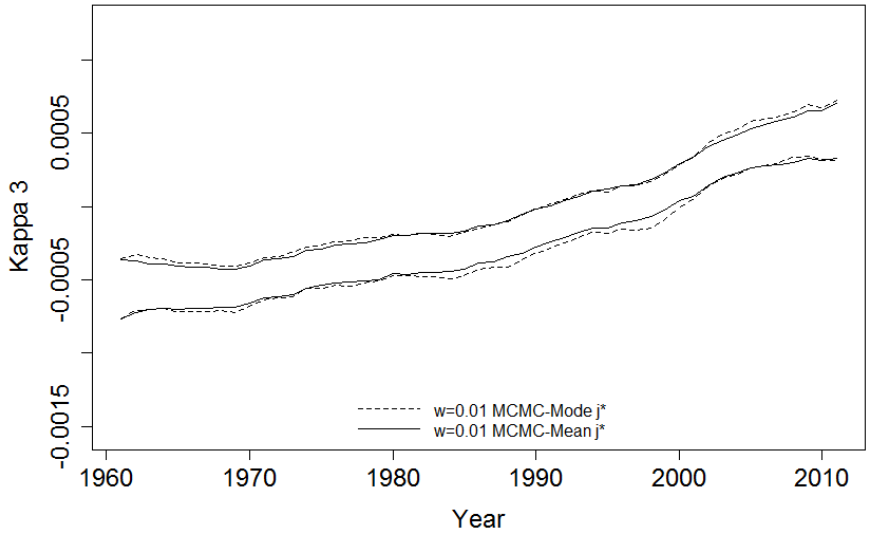


Figure 6.22: Credibility Interval: posterior distribution $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ of the simulated death scenario j^*, given the MCMC-Mean (upper blue lines; lower solid lines) and the MCMC-Mode (lower dashed lines). The red lines and the black line in the upper figure is posterior distribution $\tilde{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ and the MLE $\hat{\boldsymbol{\kappa}}^{(3),\mathrm{EW}}$ respectively for the England and Wales data. The green line is $\hat{\boldsymbol{\kappa}}_{j^*}^{(3),w}$, the MLE of death scenario j^*. Note that the upper and lower bound are the 95% and 5% quantile of the posterior distribution.

For the estimated $\boldsymbol{\kappa}^{(3)}$, Figure 6.22b shows that the trajectories of $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ conditional on MCMC-Mode (dashed lines) and MCMC-Mean (solid lines) for death scenario j^* are approximately the same. In Figure 6.22a, a good smoothness could be observed for $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ (MCMC-Mean, blue lines) compared with the heavily zigzagged

path of the point estimate $\hat{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ (green line) for the death scenario j^*. We could therefore conclude that the MCMC given both prior settings (MCMC-Mode & MCMC-Mean) generates much more smoothed estimation for $\boldsymbol{\kappa}^{(3)}$ than the MLE does for death scenario j^*. The path of the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ (MCMC-Mean, blue lines) is also a bit more smoothed than the respective true parameter $\hat{\boldsymbol{\kappa}}^{(3),\text{EW}}$ (black line), the MLE of the England and Wales data. The posterior distributions for the England and Wales data (red lines) and the death scenario j^* are approximately of the same level of the smoothness.

It is worth noticing that the credibility intervals for the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ for the death scenario j^* conditional on the both MCMC-Mode and MCMC-Mean is wider than the intervals of the $\tilde{\boldsymbol{\kappa}}^{(3),\text{EW}}$ for the England and Wales data. Recall that we have discussed the time series prior restricts the estimations of the three period effects to be a strict random walk process with the drift and such restriction is relatively weaker for estimating the $\boldsymbol{\kappa}^{(2)}$ compared with the $\boldsymbol{\kappa}^{(1)}$ due to the term $x-\bar{x}$ with the $\boldsymbol{\kappa}^{(2)}$. Similarly, due to the characteristic of the M7 model and hence the quadratic function of the age x with the $\boldsymbol{\kappa}^{(3)}$, the influence of the random walk likelihood for the $\boldsymbol{\kappa}^{(3)}$ on the joint posterior distribution is relatively not as strong as the random walk likelihood for the other two period effects (the scale of the value of the estimation for $\boldsymbol{\kappa}^{(3)}$ is relatively much smaller than the estimations of the other two effects due to the quadratic age function). Therefore the restriction on the posterior estimation of the $\boldsymbol{\kappa}^{(3)}$ and hence on the posterior variance of the $\tilde{\kappa}_{t,j^*}^{(3),w}$ such that it follows a random walk process is also not as strict as on the estimation of the $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\kappa}^{(2)}$. Eventually, $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ (MCMC-Mean) has its credibility intervals wider than the England and Wales data's.

Further, Figure 6.22b shows that the prior setting MCMC-Mode (dashed lines) apparently introduces relatively more fluctuations to the trajectory of the estimation $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ than it does to the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ (Figure 6.5b) and $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$ (6.13b). The reason is once again due to the even weaker (compared with $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\kappa}^{(2)}$) influence of $\boldsymbol{\kappa}^{(3)}$ on

the joint posterior distribution and therefore a weaker restriction from the random walk likelihood on the estimation of $\boldsymbol{\kappa}^{(3)}$. The influence of the prior density for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ formed by the corresponding random walk likelihood is stronger, compared with the influence of the prior for the other two volatilities.

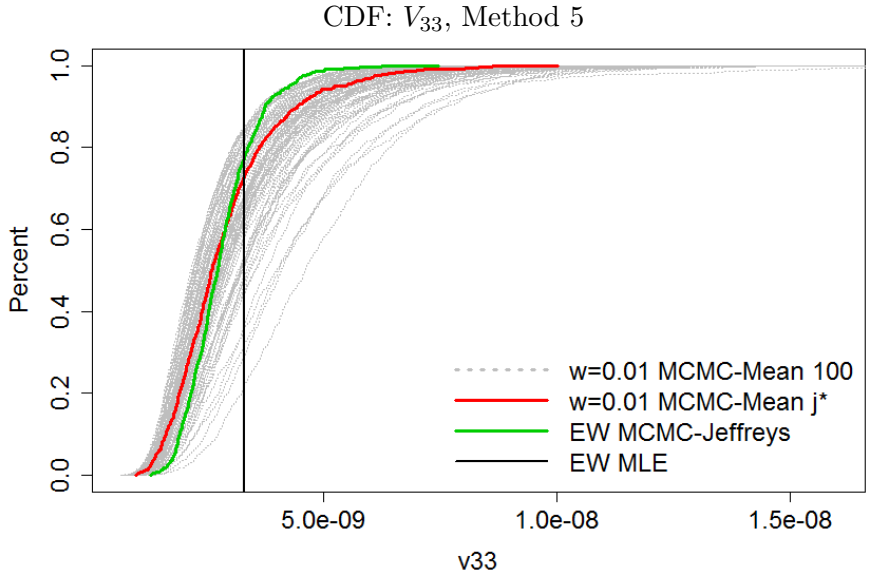


Figure 6.23: CDFs of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}(3,3)$ (MCMC-Mean) for $j' = j_1, \ldots, j_{100}$ (dotted curves), the posterior distributions for 100 randomly selected simulated death scenarios. The red curve is the CDF of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ conditional on MCMC-Mean for death scenario j^*. The green curve and the vertical solid line are the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(3,3)$ and the MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{\mathrm{EW}}(3,3)$ of the England and Wales data respectively.

In Figure 6.23, the spread of the posterior distributions $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}(3,3)$ (MCMC-Mean) for $j' = j_1, \ldots, j_{100}$ of the 100 selected death scenarios (dotted CDF) are demonstrated and unsurprisingly the influence of the sampling variation shifts the posterior distribution from one side to another. See Table 6.2 for the characteristic statistics of the distribution of the spread of the posterior means. We can observe that the level of such shift varies from one death scenario to another and could be very large according to the relative positions of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ (red CDF, MCMC-Mean) and the right edge of the spread of the distributions. The true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{\mathrm{EW}}(3,3)$ is a bit on the right hand side of the centre of the spread. We therefore conclude that given the sampling variation's shifting effect, the estimation for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ generated by the MCMC conditional on the MCMC-Mean is not significantly different

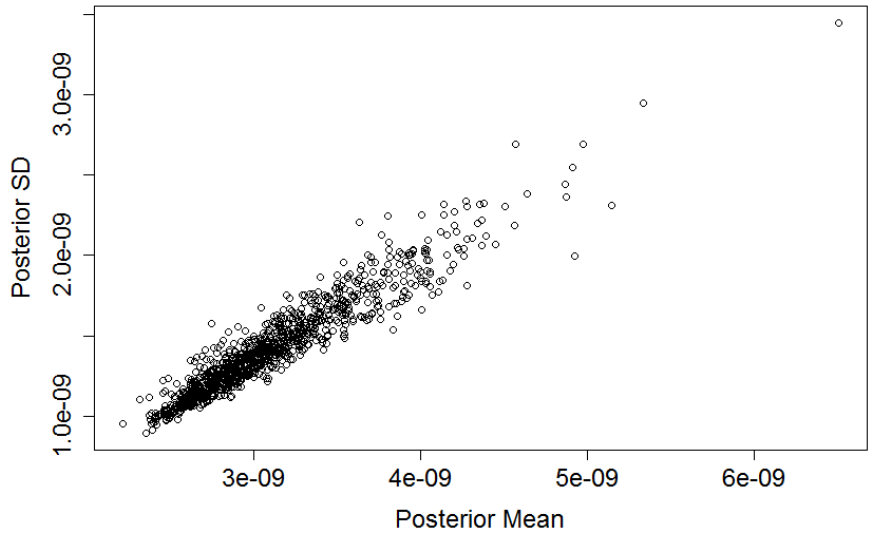
from the respective true parameter. Further, the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(3,3)$ (green CDF) of the England and Wales data is approximately in the centre of the spread of the posterior distributions with a similar posterior variance, which also implies a non-significant difference between the posterior estimation of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ for the two populations. Given that the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(3,3)$ is on the 20% quantile of the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(3,3)$ conditional on the prior setting MCMC-Mode, we expect that the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(3,3)$ will also be on the 20% quantiles of the posterior distributions of the most of the simulated death scenarios (without the need to re-run the MCMC with the alternative prior settings to all the N_1 death scenarios) with the sampling variation shifting the posterior distributions around. We therefore could also conclude that the posterior estimation for the $w = 0.01$ conditional on the MCMC-Mode is not significantly different from the true rate. On the other hand, we expect that the spread of the posterior distribution driven by the sampling variation will be on the right hand side of the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(3,3)$ of the England and Wales data, though the difference between the two is not significant either.

For $\boldsymbol{\mu}_3$

The posterior distribution $\tilde{\boldsymbol{\mu}}_3^{w,j^*}$ conditional on the MCMC-Mode is demonstrated in Figure 6.25 (long-dashed CDF), which is of approximately the same posterior mean with the MCMC-Mean (dotted CDF) and a slightly increase on the posterior variance over the latter's as may be expected. Figure 6.28 demonstrates a positive linear relationship between the level of $\sqrt{\tilde{\boldsymbol{V}}_{\epsilon}^{w}(3,3)}$ and the variance of $\tilde{\boldsymbol{\mu}}_3^{w}$. See the characteristic statistics in Table 6.1. Consistently, no obvious tilt could be visually spotted on the trajectory of the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ conditional on the MCMC-Mode (dashed lines) in Figure 6.22b.

In particular, the posterior distribution $\tilde{\boldsymbol{\mu}}_3^{w,j^*}$ (MCMC-Mode & MCMC-Mean) is on the left hand side of the posterior distribution $\tilde{\boldsymbol{\mu}}_3^{\mathrm{EW}}$ of the England and Wales data

(a) Scatter Plot: Posterior Means vs SD



(b) Scatter Plot: Posterior Means vs Skewness

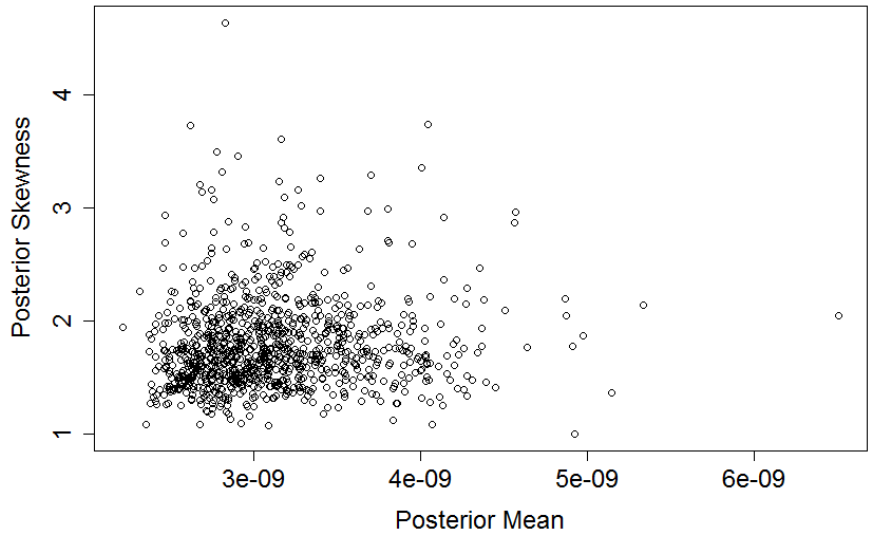


Figure 6.24: The scatter plot: the posterior mean vs. coefficient of skewness (lower); the posterior mean vs. standard deviation (upper) for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ of $w = 0.01$, based on the posterior distributions $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}$ (MCMC-Mean) for $j = 1, \ldots, N_1$.

(solid CDF), implying the posterior estimation generated by the MCMC is lower for the death scenario j^* given the underlying prior settings. On the other hand, the respective true parameter of the death scenario j^*, $\hat{\boldsymbol{\mu}}_3^{\mathrm{EW}}$ (vertical dot-dashed line) the MLE of the England and Wales data, is approximately on the 95% quantile of the distribution $\tilde{\boldsymbol{\mu}}_3^{w,j^*}$. We therefore conclude that the posterior estimation for the $\boldsymbol{\mu}_3$ of the death scenario j^* is less than the respective true parameter but the difference is not significant given the underlying two prior settings. See the characteristic

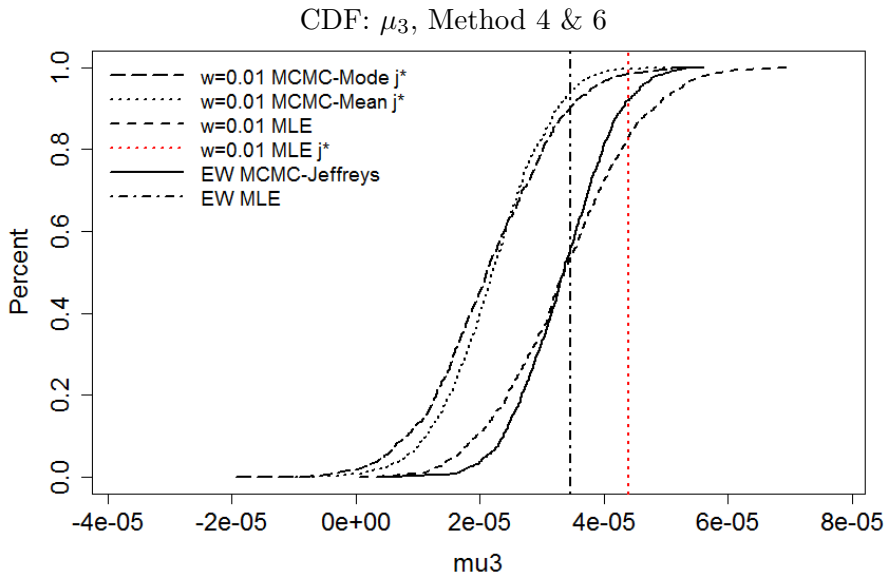


Figure 6.25: CDF: posterior distribution $\tilde{\boldsymbol{\mu}}_3^{w,j^*}$ of the simulated death scenario j^*, given the MCMC-Mean (dotted curve) and the MCMC-Mode (long-dashed curve). The solid curve and the vertical dot-dashed line are the posterior distribution $\tilde{\boldsymbol{\mu}}_3^{\text{EW}}$ and the true rate of the simulated death scenarios $\hat{\boldsymbol{\mu}}_3^{\text{EW}}$ respectively. The dashed curve is the CDF for $\hat{\boldsymbol{\mu}}_3^{w}$, the finite-sample MLEs of the N_1 simulated death scenarios. The red vertical line is the relative MLE $\hat{\boldsymbol{\mu}}_3^{w,j^*}$ for death scenario j^*.

statistics in Table 6.4. Consistently, a clock-wise tilt can be observed in Figure 6.22a for the trajectory of the posterior estimation $\tilde{\boldsymbol{\kappa}}_{j^*}^{(3),w}$ (blue lines) of the death scenario j^* (MCMC-Mean) from the paths of the MLE $\hat{\boldsymbol{\kappa}}^{(3),\text{EW}}$ (black line) and the posterior estimation $\tilde{\boldsymbol{\kappa}}^{(3),\text{EW}}$ (red lines) for the England and Wales data. Visually, the shape of the credibility intervals for the $\tilde{\boldsymbol{\kappa}}^{(3),w}$ (MCMC-Mode & MCMC-Mean) is more like a linear process than the other two estimations due to the random walk restriction effect. The reason for the clock-wise tilt and therefore a less steep estimation for the death scenario j^* is a response to the anti-clockwise tilts of the $\tilde{\boldsymbol{\kappa}}_{j^*}^{(1),w}$ and $\tilde{\boldsymbol{\kappa}}_{j^*}^{(2),w}$, and therefore a correspond to the tile of the $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ as may be expected. Similar correlation of the estimation between the three period effects can be observed in Figure 3.1, where the estimation of the $\kappa_t^{(1)}$ is negatively correlated to the estimated $\kappa_t^{(3)}$ at the very early and late years regardless with population size.

We demonstrate the spread of the posterior distributions $\tilde{\boldsymbol{\mu}}_3^{w,j'}$ for $j' = j_1, \ldots, j_{100}$ conditional on the MCMC-Mean driven by the sampling variation in Figure 6.26

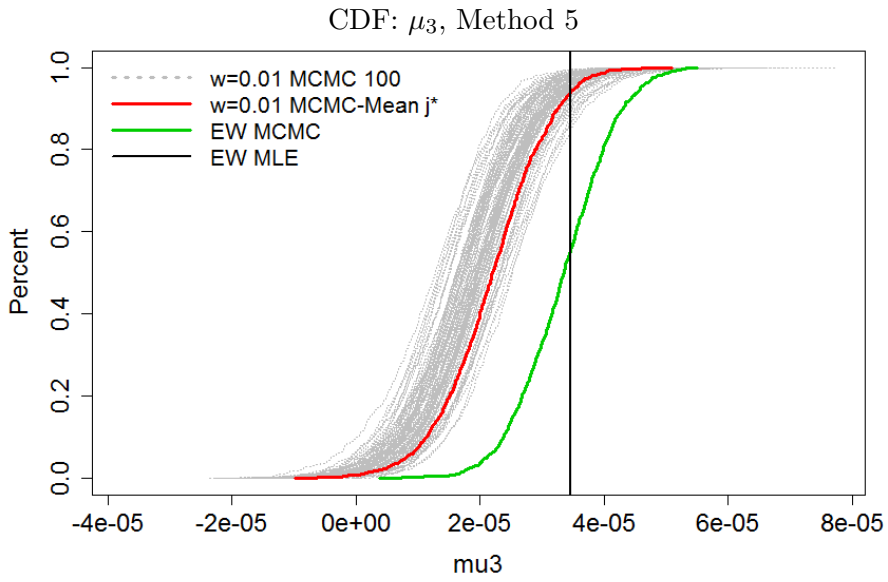


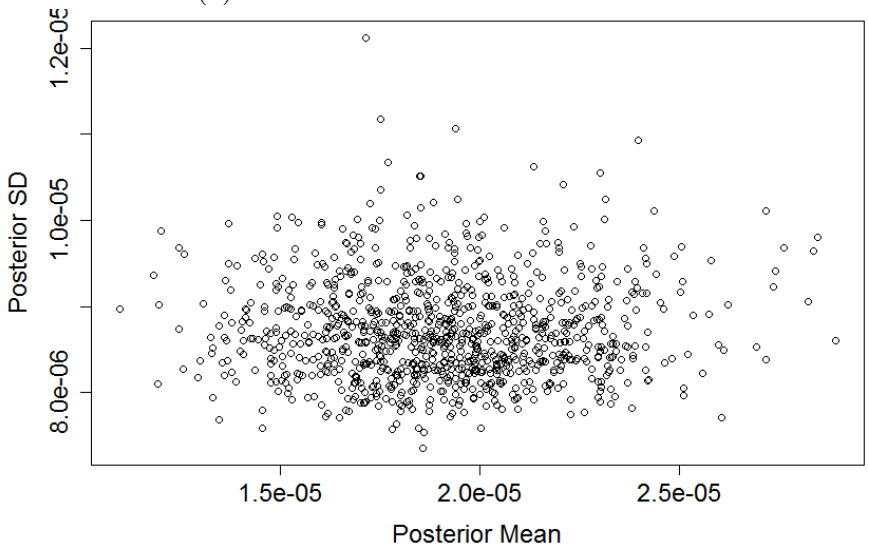
Figure 6.26: CDFs of $\tilde{\boldsymbol{\mu}}_3^{w,j'}$ (MCMC-Mean) for $j' = j_1, \ldots, j_{100}$ (dotted curves), the posterior distributions for 100 randomly selected simulated death scenarios. The red curve is the CDF of $\tilde{\boldsymbol{\mu}}_3^{w,j^*}$ conditional on MCMC-Mean for death scenario j^*. The green curve and the vertical solid line are the posterior distribution $\tilde{\boldsymbol{\mu}}_3^{\text{EW}}$ and the MLE $\hat{\boldsymbol{\mu}}_3^{\text{EW}}$ of the England and Wales data respectively.

(dotted CDFs), where the sampling variation shifts the posterior distribution from one side to another, visually with no obvious effect on the posterior variance. Such shifting effect varies from one death scenario to another and could be very large.

Recall that the posterior distribution $\tilde{\boldsymbol{\mu}}_3^{w,j^*}$ conditional on the MCMC-Mode is approximately the same with the distribution given the MCMC-Mean (Figure 6.25). We therefore would also expect that without re-running the MCMC for all the N_1 simulated death scenarios conditional on the MCMC-Mode, the relative distribution of the posterior distributions based on Method 5 will be similar with the spread demonstrated in Figure 6.26. The true parameter $\hat{\boldsymbol{\mu}}_3^{\text{EW}}$ (vertical line) is on the right tail of the spread of the posterior distributions and we therefore conclude that in order to correspond to the tilts on the estimation of the $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\kappa}^{(2)}$, the MCMC generates an estimation for the $\boldsymbol{\mu}_3$ of the $w = 0.01$ (given both underlying prior settings) that is smaller than the respective true parameter, though such reduction is not significant. We also conclude that the posterior estimation for the England and Wales data (green CDF) is higher than the posterior estimation for the small

population $w = 0.01$. See the characteristic statistics of the distribution of the posterior means $\bar{\tilde{\boldsymbol{\mu}}}_3^w$ according to the N_1 simulated samples in Table 6.2.

(a) Scatter Plot: Posterior Means vs SD



(b) Scatter Plot: Posterior Means vs Skewness

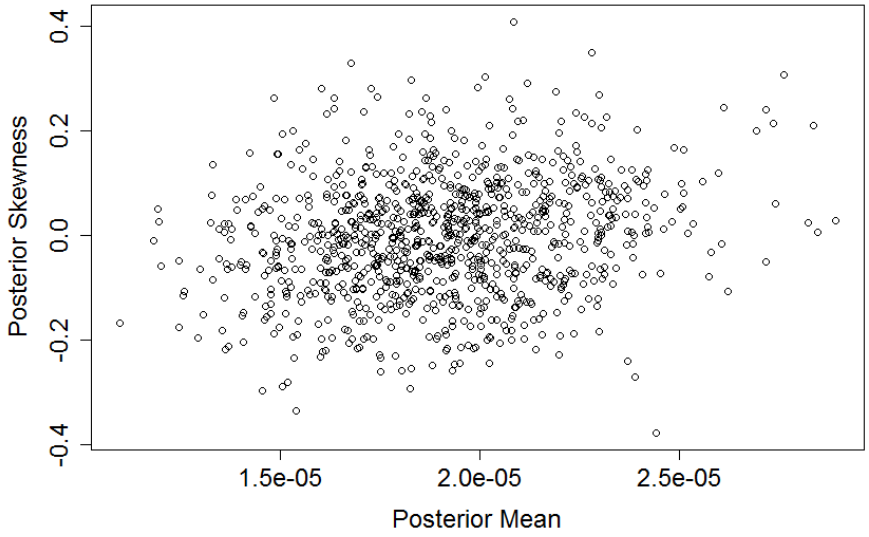


Figure 6.27: The scatter plot: the posterior mean vs. coefficient of skewness (lower); the posterior mean vs. standard deviation (upper) for $\boldsymbol{\mu}_3$ of $w = 0.01$, based on the posterior distributions $\tilde{\boldsymbol{\mu}}_3^{w,j}$ (MCMC-Mean) for $j = 1, \ldots, N_1$.

The influence of the sampling variation on the standard deviation and the coefficient of skewness of the posterior distribution is demonstrated in Figure 6.27 based on the posterior distributions of the N_1 death scenarios given MCMC-Mean. Unsurprisingly, no linear trend could be observed between: the mean and standard deviation; the mean and the skewness of the posterior distribution in the two plots.

We therefore could conclude that the sampling variation affects the posterior distribution $\tilde{\boldsymbol{\mu}}_3^w$ by shifting it from one way to another without significantly changing the uncertainty and the skewness.

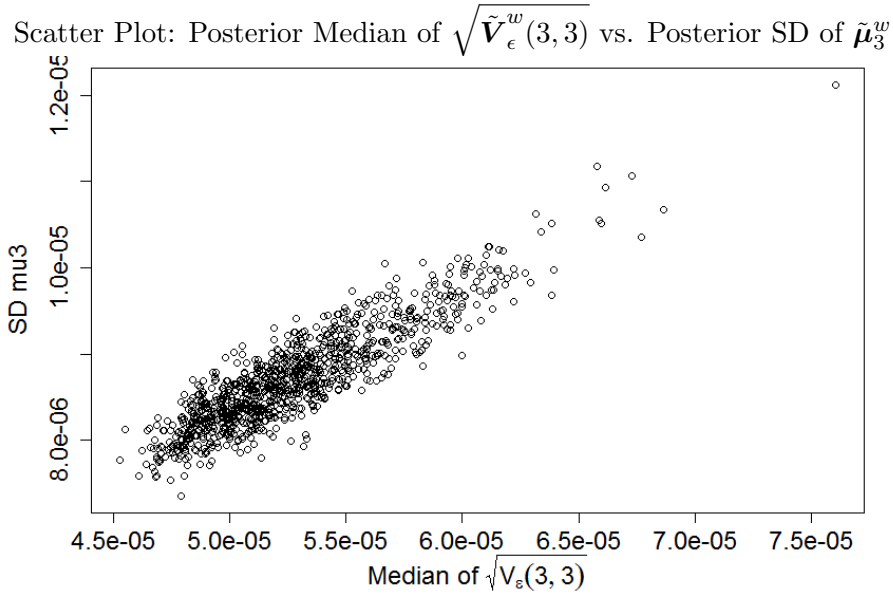


Figure 6.28: Scatter Plot: the posterior median of the $\sqrt{\tilde{\boldsymbol{V}}_\epsilon^w(3,3)}$ (horizontal axis) vs. the standard deviation of $\tilde{\boldsymbol{\mu}}_3^w$ (vertical axis) based on the corresponding posterior distributions of the N_1 death scenarios.

We at last demonstrate the spread of the posterior distributions $\tilde{\kappa}_{t,j}^{(3),w}$ (MCMC-Mean) for $j = 1, \ldots, N_1$ of all the N_1 death scenarios at some years in Figure 6.29 (dotted CDFs). The sampling variation shifts the posterior distribution from one way to another with no obvious influence on the posterior variance. For the year (t_{31}) in the mid of the fitted year range, the true rate $\hat{\kappa}_{t_{31}}^{(3),\mathrm{EW}}$ (vertical line) is on the centre of the spread of the distributions. On the other hand, the true rates $\hat{\kappa}_{t_1}^{(3),\mathrm{EW}}$ and the $\hat{\kappa}_{t_{51}}^{(3),\mathrm{EW}}$ of the first and the last years are on the left and the right tails of the corresponding spread of the posterior distributions. We therefore conclude that the posterior estimation of the small population $w = 0.01$ is statistically the same with the respective true parameter for the years in the middle of the empirical year range. The estimations for the earlier and the later years are not significantly different from the respective true parameter due to the influence of the time series models. Therefore such tilt and changes on the shape of the trajectories of the

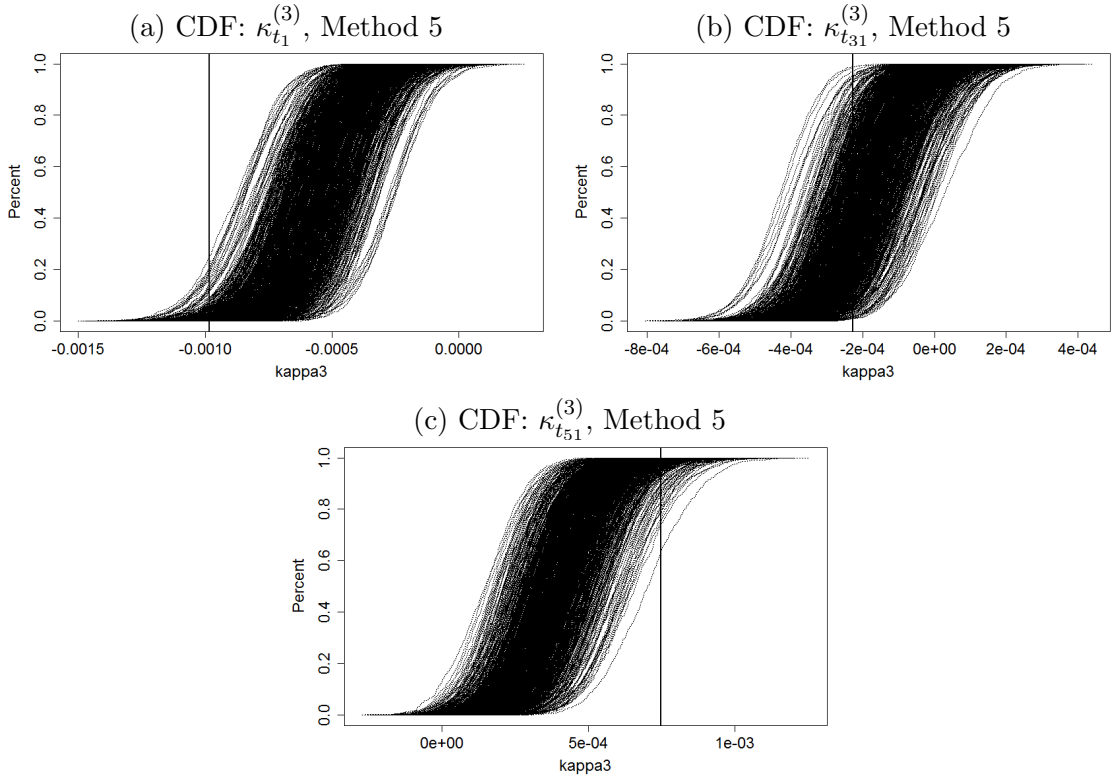


Figure 6.29: The CDF of posterior distribution for $\boldsymbol{\kappa}^{(3)}$ with selected years, given Method 5. Note that each dotted curve represents the posterior distribution for one randomly selected death scenario and all the N_1 CDFs are plotted.

latent-parameter estimations compared with the respective true parameters and the MCMC estimations of the England and Wales data are all expected when we restrict the latent parameters with the time series models for a better estimation of their volatilities. The influence of the sampling variation shifting the posterior distributions of the other years are demonstrated in Figure D.3 in Appendix D.

6.2.2 Cohort effect $\boldsymbol{\gamma}$ and the Hyper-Parameters σ_γ, α_γ

Recall that for the Bayesian approach, $\boldsymbol{\gamma}^{(4)}$ follows the AR(1) model with the formula:

$$\gamma_c^{(4)} = \alpha_\gamma \gamma_{c-1}^{(4)} + \epsilon_c, \text{for } c > t_1 - x_{n_a},$$

where $\epsilon_c \sim N(0, \sigma_\gamma^2)$. In previous subsection we have already discussed that due to the dominating impact of the time series prior on the joint posterior distribution, the

estimation for the cohort effect $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ of the death scenario j^* is strictly constrained to follow the zero mean-reverting AR(1) process demonstrated in Figure 6.38a (blue lines) with a narrower credibility intervals than the posterior estimation of the England and Wales data (red lines). In this subsection, we study the marginal posterior distributions for the hyper-parameters of the cohort effect and the influence of the sampling variation on the posterior distribution.

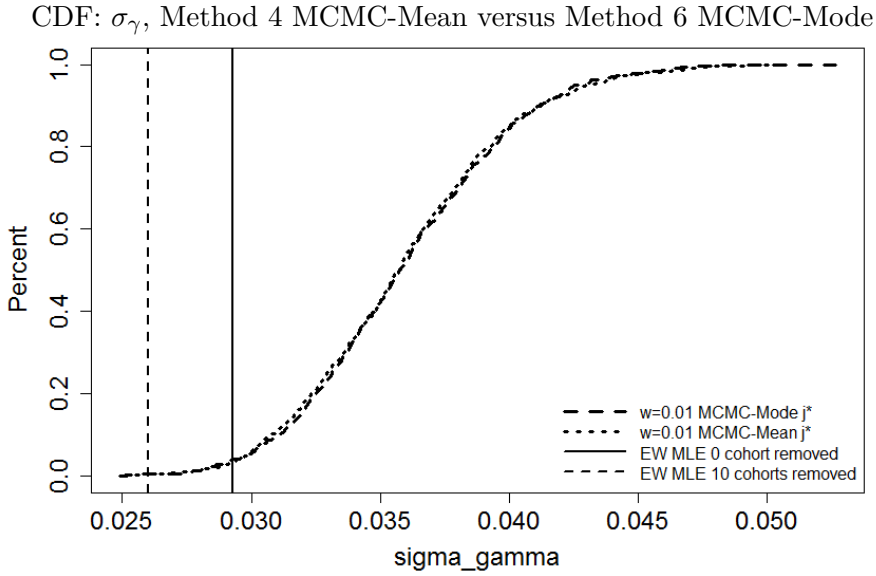


Figure 6.30: CDF: the posterior distribution of $\tilde{\sigma}_\gamma^{w,j^*}$ for death scenario j^*, given MCMC-Mode (long-dashed curve) and MCMC-Mean (dotted curve). The dashed vertical line is the point estimate of σ_γ for England and Wales data given 10 cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$. The solid vertical line is the true rate of $w = 0.01$, $\hat{\sigma}_\gamma^{\text{EW}}$, i.e. no cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$.

We demonstrate the posterior distribution $\tilde{\sigma}_\gamma^{w,j^*}$ for the death scenario j^* conditional on the prior setting MCMC-Mode for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ in Figure 6.30 (bold long-dashed CDF) and as may be expected no differences can be visually spotted on the posterior distributions $\tilde{\sigma}_\gamma^{w,j^*}$ and the $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ (Figure 6.38b) given the two prior settings for the volatility of the period effects. See the characteristic statistics of $\tilde{\sigma}_\gamma^{w,j^*}$ (MCMC-Mode) in Table 6.6.

Figure 6.31 includes the following curves:

- $w = 0.01$ MCMC-Mean j^*: the posterior distribution $\tilde{\sigma}_\gamma^{w,j^*}$ of the simulated death

scenario j^*, given the prior setting of MCMC-Mean,

- $w = 0.01$ MLE $(\cdot)$ cohorts removed: the distribution $\hat{\sigma}_\gamma^w$ of the finite-sample MLEs of the N_1 simulated death scenarios with $(\cdot) = 2$, 4, 6, 8 and 10 short cohorts removed.

- EW MCMC: the posterior distribution $\tilde{\sigma}_\gamma^{\text{EW}}$ of the England and Wales data,

- EW MLE $(\cdot)$ cohorts removed: the MLE $\hat{\sigma}_\gamma^{\text{EW}}$ of the England and Wales data with $(\cdot) = 0$ and 10 short cohorts removed.

Note that $\hat{\sigma}_\gamma^w$ and $\hat{\sigma}_\gamma^{\text{EW}}$ are calculated by fitting the Equation (5.37) in Section 5.4.2 of Chapter 5 to the corresponding MLEs of the cohort effect. See the discussion of removing short cohorts in Section 3.3.2 of Chapter 3.

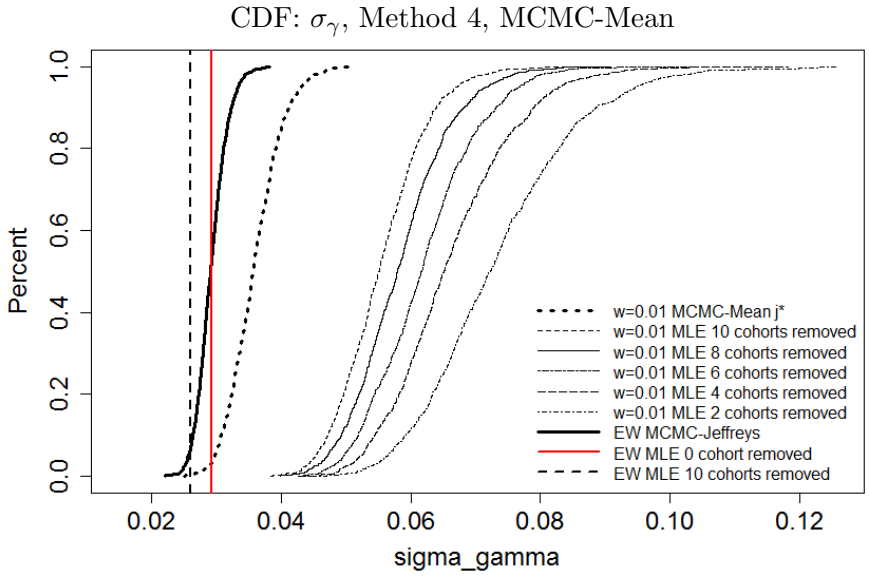


Figure 6.31: CDF: the posterior distribution of $\tilde{\sigma}_\gamma^{w,j^*}$ (MCMC-Mean, dotted curve) for death scenario j^*, the finite-sample point estimates $\hat{\sigma}_\gamma^w$ given selected number of cohorts at the beginning and the end of $\hat{\boldsymbol{\gamma}}^{(4),w}$ are removed for $w = 0.01$; posterior distribution $\tilde{\sigma}_\gamma^{\text{EW}}$ (solid curve). The dashed vertical line is the point estimate of σ_γ for England and Wales data given 10 cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$ (0.0260). The red solid vertical line is the true rate of $w = 0.01$, $\hat{\sigma}_\gamma^{\text{EW}}$ (0.0293), i.e. no cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$.

We demonstrate the posterior distribution $\tilde{\sigma}_\gamma^{w,j^*}$ (MCMC-Mean, bold dotted CDF) in Figure 6.31 for the death scenario j^* conditional on the prior setting

MCMC-Mean for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$. Recall that the low level information involved by the short cohorts with few observations for the very young and old years of birth push the relative MLEs much higher or lower than the MLEs for the rest cohorts, and such failure on estimating the short cohorts becomes even worse for the small populations, e.g. $w = 0.01$, 0.001. In Section 3.3.2 we investigated the distribution of the finite-sample MLEs for the parameters of the AR(1) model given $w = 1$ when the first and the last 1, 2, 3, 4, 5, 6 cohorts' estimates are removed. We concluded that for $w = 1$, removing the short cohorts significantly affects the MLE of the α_γ and has relatively smaller impact on estimating the σ_γ. Note that the AR(1) model adopted in Section 3.3.2 is different from the zero mean-reverting model applied for the joint posterior distribution. By following the idea of removing the shorts, we calculated the finite-sample MLEs $\hat{\alpha}_\gamma^w(\cdot)$ and $\hat{\sigma}_\gamma^w(\cdot)$ by fitting the AR(1) model in Equation 5.11 to the finite-sample MLEs $\hat{\boldsymbol{\gamma}}^{(4),w}$ without the first and the last 1, 2, 3, 4, 5 short cohorts, where $(\cdot)$ =1, 2, 3, 4, 5 is the number of short cohorts removed from the beginning and the end of the years of birth. The distributions of the finite-sample MLEs $\hat{\sigma}_\gamma^w(\cdot)$ are demonstrated in Figure 6.31 (groups of the CDFs in the right hand side of the plot), where unsurprisingly the volatility is smaller with more short cohorts removed. Note that the MLE of the cohort effect for some of the simulated death scenarios can not be modelled by a stationary AR(1) model without removing any short cohorts and therefore no demonstration is available for the distribution of the finite-sample MLEs $\hat{\alpha}_\gamma^w(\cdot)$ and $\hat{\sigma}_\gamma^w(\cdot)$ with $(\cdot) = 0$.

As may be expected, the $\hat{\sigma}_\gamma^w$ is significantly higher than the respective true parameter $\hat{\sigma}_\gamma^{\text{EW}}$ (vertical solid line) when as many as 10 short cohorts are removed (the first left fine-dashed CDF), implying that given relatively fewer short cohorts to be over-fitted, the two-stage approach still generates an over-estimation for the volatility of the cohort effect of the small population $w = 0.01$ because of the noise to the latent parameter estimation when the population size is small. On the other hand, the true rate $\hat{\alpha}_\gamma^{\text{EW}}$ is approximately on the 5% quantile of the posterior distribution $\tilde{\sigma}_\gamma^{w,j^*}$ of the death scenario j^* conditional on the MCMC-Mean, implying

	$\tilde{\sigma}_\gamma^{\text{EW}}$	$\tilde{\sigma}_\gamma^{w,j^*}$, A MCMC-Mode	$\tilde{\sigma}_\gamma^{w,j^*}$, B MCMC-Mean	$\hat{\sigma}_\gamma^{w}$, 10 Removed	$\hat{\sigma}_\gamma^{w}$, 2 Removed
Mean	0.0293	0.0360	0.0359	0.0554	0.0734
Mode	0.0284	0.0355	0.0356	0.0544	0.0679
SD	0.0024	0.0040	0.0041	0.0068	0.0120
Skewness	0.4091	0.3005	0.3748	0.5174	0.7413
Quantile					
2.5%	0.0251	0.0286	0.0289	0.0439	0.0537
50%	0.0291	0.0357	0.0357	0.0551	0.0723
97.5%	0.0343	0.0445	0.0447	0.0698	0.0996

Table 6.6: Characteristic statistics and 2.5%, 50% and 97.5% quantiles of: $\tilde{\sigma}_\gamma^{\text{EW}}$, $\tilde{\sigma}_\gamma^{w,j^*}$ given MCMC-Mode (A) and MCMC-Mean (B), $\hat{\sigma}_\gamma^{w}$ given 2 and 10 cohorts removed from $\hat{\boldsymbol{\gamma}}^{(4),w}$.

a non-significant difference between the MCMC estimation and the respective true parameter. Once again this is because the time series prior restricts the estimation for the cohort effect and its corresponding hyper-parameters such that the trajectory of the estimation is as smooth as a strict zero mean-reverting AR(1) process. The posterior distribution $\tilde{\sigma}_\gamma^{\text{EW}}$ of the England and Wales data (black CDF) is approximately centred around the corresponding MLE as we have discussed. Therefore we conclude that the posterior estimation of the cohort effect's volatility for the England and Wales data is not as high as for the death $j*$. Consistently, a better smoothed trajectory of the posterior distribution $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ for the death scenario j^* (blue lines) can be observed, compared with the path of the corresponding MLE $\hat{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ (green line), though $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ is not as smooth as the respective true parameter $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$ and the England and Wales data's posterior estimation $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ (red lines). See the characteristic statistics in Table 6.6.

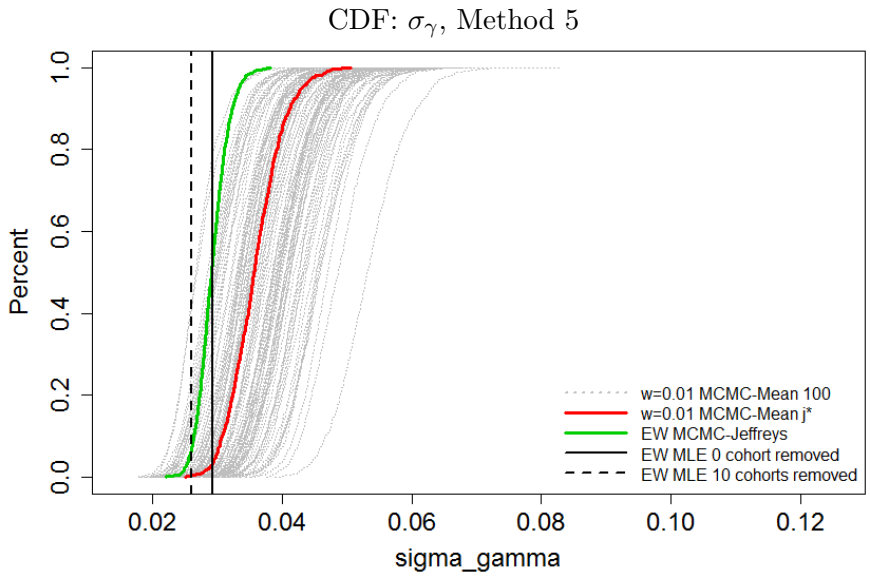
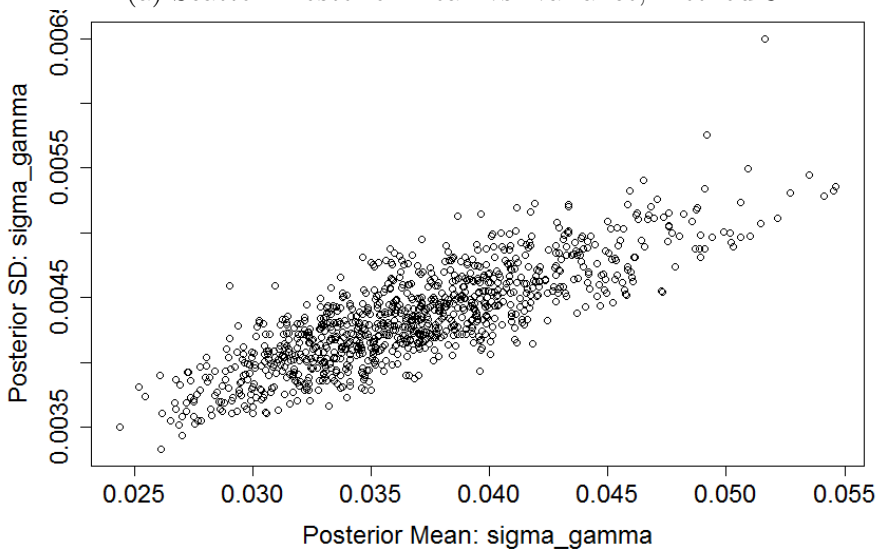


Figure 6.32: CDF according to Method 5: posterior distributions $\tilde{\sigma}_\gamma^{w,j'}$ (MCMC-Mean, dotted curves) for death scenario $j' = j_1, \ldots, j_{100}$. The dashed vertical line is the point estimate of σ_γ for England and Wales data given 10 cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$. The solid vertical line is the true rate of $w = 0.01$, $\hat{\sigma}_\gamma^{\text{EW}}$, i.e. no cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$.

We demonstrate the spread of the posterior distributions $\tilde{\sigma}_\gamma^{w,j'}$ (MCMC-Mean) for $j' = j_1, \ldots, j_{100}$ of the 100 selected death scenarios in Figure 6.32 (dotted CDFs), where the sampling variation shifts the posterior distribution from one way to another, visually with no obvious influence on the posterior variance. The gap between the posterior distribution $\tilde{\sigma}_\gamma^{w,j^*}$ (red CDF) and the right tail of the spread of the distributions indicates the shift of the sampling variation varies from one death scenario to another and could be very large.

We demonstrate the influence of the sampling variation on the posterior variance and the skewness in Figure 6.33. There is a very weak correlation between the distribution of the posterior variance and posterior mean (6.33a) according to the samples of the N_1 simulated death scenarios such that the posterior variance is higher when the further the sampling variation shifts the posterior distribution to the right hand side of the true rate. However, the increase on the posterior variance from one death scenario to another is relatively much smaller compared with the sampling variation shifting the posterior distribution. In particular, most of the simulated

(a) Scatter: Posterior mean vs. variance, Method 3



(b) Scatter: Posterior mean vs. Skewness, Method 3

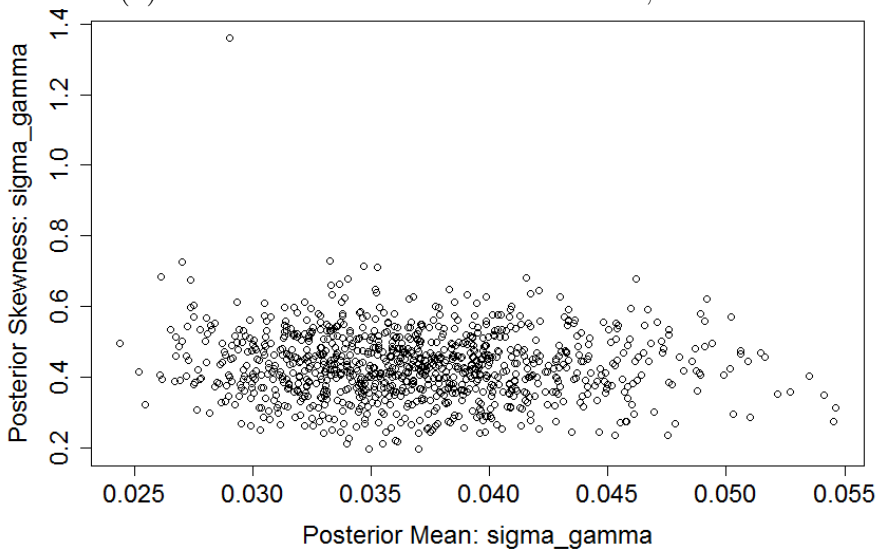


Figure 6.33: Scatter plot of: posterior mean vs. variance; posterior mean vs. skewness, of $\tilde{\sigma}_\gamma^w$ according to the posterior distributions of the N_1 death scenarios (MCMC-Mean).

death scenarios have their posterior means within the range of 0.03 to 0.04, while the corresponding variances relatively remain unchanged with only little increases. On the other hand, the level of skewness is almost invariant to the influence of the sampling variation since no obvious trend could be observed in Figure 6.33b between the mean and the skewness of the posterior distribution. We therefore conclude that as we expect the sampling variation shifts the posterior distribution around with little influence on the posterior variance and skewness.

The respective true parameter $\hat{\sigma}_\gamma^{\text{EW}}$ (vertical solid line) with no short cohorts removed and the posterior distribution $\tilde{\sigma}_\gamma^{\text{EW}}$ of the England and Wales data (green CDF) are a bit on the left hand side of the centre of the spread of the $\tilde{\sigma}_\gamma^{w,j'}$s. We therefore conclude that with the sampling variation shifting the posterior distribution, statistically the posterior estimation of the simulated death scenarios' volatility of the cohort effect is not significantly differed from its respective true parameter as well as the posterior estimation of the England and Wales data. See the characteristic statistics of the distribution of the posterior means $\bar{\tilde{\sigma}}_\gamma^w$ of the N_1 death scenarios in Table 6.7.

	α_γ	σ_γ
True Rate	0.9827	0.0293
Mean	0.4435	0.0368
Quantile		
97.5%	0.7799	0.0487
50%	0.4489	0.0365
2.5%	0.0597	0.0277

Table 6.7: The mean and 97.5%, 50%, 2.5% quantiles of the finite-sample MCMC mean $\bar{\tilde{\alpha}}_\gamma^w = \{\bar{\tilde{\alpha}}_{\gamma,j}^w\}_{j=1,\ldots,N_1}$ and $\bar{\tilde{\sigma}}_\gamma^w = \{\bar{\tilde{\sigma}}_{\gamma,j}^w\}_{j=1,\ldots,N_1}$ for the 1000 death scenarios according to Method 3. Note that the true rate is the point estimates for the England and Wales data derived from the point estimates of cohort effects without removing any cohort year from $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$.

Figure 6.34 includes the following curves:

- $w = 0.01$ MCMC-Mode j^*: the posterior distribution $\tilde{\alpha}_\gamma^{w,j^*}$ of the simulated death scenario j^*, given the prior setting of MCMC-Mode,
- $w = 0.01$ MCMC-Mean j^*: the posterior distribution $\tilde{\alpha}_\gamma^{w,j^*}$ of the simulated death scenario j^*, given the prior setting of MCMC-Mean,

	$\tilde{\alpha}_\gamma^{\text{EW}}$	$\tilde{\alpha}_\gamma^{w,j^*}$, A MCMC-Mode	$\tilde{\alpha}_\gamma^{w,j^*}$, B MCMC-Mean	$\hat{\alpha}_\gamma^{w}$, 10 Removed	$\hat{\alpha}_\gamma^{w}$, 2 Removed
Mean	0.9578	0.1409	0.1728	0.6125	0.5723
Mode	0.9692	0.1434	0.2329	0.7515	0.7320
SD	0.0252	0.2038	0.2207	0.2062	0.2433
Skewness	-1.005	0.0376	0.1110	-0.7542	-0.6820
Quantile					
2.5%	0.8992	-0.2357	-0.2482	0.1477	0.0233
50%	0.9623	0.1421	0.1691	0.6593	0.6193
97.5%	0.9931	0.5435	0.6196	0.9015	0.9258

Table 6.8: Characteristic statistics and 2.5%, 50% and 97.5% quantiles of: $\tilde{\alpha}_\gamma^{\text{EW}}$, $\tilde{\alpha}_\gamma^{w,j^*}$ given MCMC-Mode (A) and MCMC-Mean (B), $\hat{\alpha}_\gamma^{w}$ given 2 and 10 cohorts removed from $\hat{\boldsymbol{\gamma}}^{(4),w}$.

- $w = 0.01$ MLE $(\cdot)$ cohorts removed: the distribution $\hat{\alpha}_\gamma^{w}$ of the finite-sample MLEs of the N_1 simulated death scenarios with $(\cdot) = 2$, 4, 6, 8 and 10 short cohorts removed.

- EW MCMC: the posterior distribution $\tilde{\alpha}_\gamma^{\text{EW}}$ of the England and Wales data,

- EW MLE $(\cdot)$ cohorts removed: the MLE $\hat{\alpha}_\gamma^{\text{EW}}$ of the England and Wales data with $(\cdot) = 0$ and 10 short cohorts removed.

Note that $\hat{\alpha}_\gamma^{w}$ and $\hat{\alpha}_\gamma^{\text{EW}}$ are calculated by fitting the Equation (5.37) in Section 5.4.2 of Chapter 5 to the corresponding MLEs of the cohort effect.

We demonstrate the influence of the prior setting MCMC-Mode on the posterior distribution $\tilde{\alpha}_\gamma^{w,j^*}$ for death scenario j^* in Figure 6.34 (bold dashed CDF). As may be expected, no obvious changes could be observed on the posterior distribution by

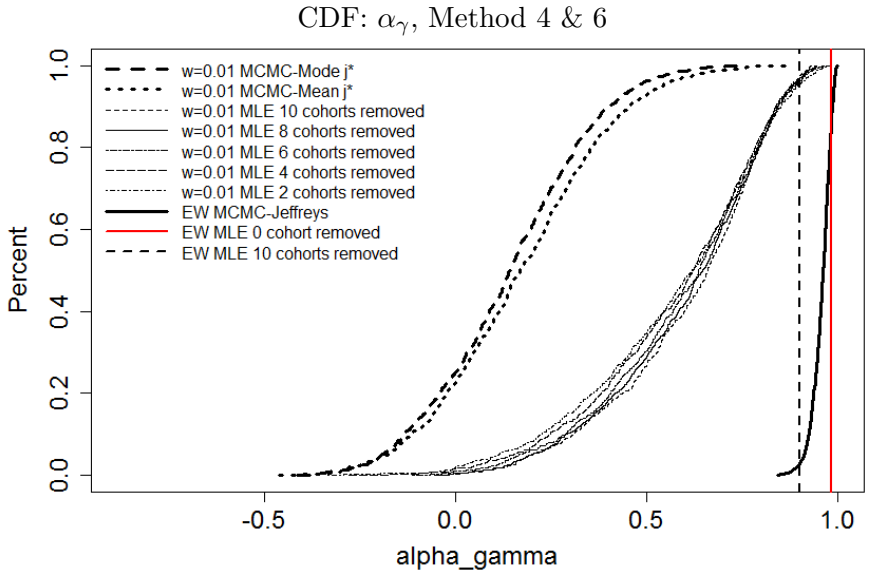


Figure 6.34: CDF: the posterior distribution $\tilde{\alpha}_\gamma^{w,j^*}$ for death scenario j^*, given MCMC-Mode (bold long-dashed curve) and MCMC-Mean (bold dotted curve); the finite-sample point estimates $\hat{\alpha}_\gamma^w$ given selected number of cohorts at the beginning and the end of $\hat{\boldsymbol{\gamma}}^{(4),w}$ are removed for $w = 0.01$. The dashed vertical line is the point estimate of α_γ for England and Wales data given 10 cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$. The red vertical line is the true rate of $w = 0.01$, $\hat{\alpha}_\gamma^{\text{EW}}$, i.e. no cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$.

alternating the prior setting of the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, though both the posterior mean and variance are relatively higher for the $\tilde{\alpha}_\gamma^{w,j^*}$ (bold dotted CDF) with MCMC-Mean. See characteristic statistics in Table 6.8. Recall we have observed that the posterior distributions $\tilde{\sigma}_\gamma^{w,j^*}$ conditional on MCMC-Mode and MCMC-Mean are approximately the same as well (Figure 6.30). Consistently, no apparent differences could be observed on the trajectories of $\tilde{\boldsymbol{\gamma}}_{j_*}^{(4),w}$ given the two prior settings in Figure 6.38b.

The $\tilde{\alpha}_\gamma^{w,j^*}$ (MCMC-Mode & MCMC-Mean) is on the left hand side of the MLE $\hat{\alpha}_\gamma^{\text{EW}}$ (no short cohort removed, vertical solid line) and the posterior distribution $\tilde{\alpha}_\gamma^{\text{EW}}$ for the England and Wales data. This is because we intended to push the estimation of α_γ away from 1 by assuming a beta formed prior distribution $p(\alpha_\gamma) \propto (1-\alpha_\gamma^2)^g$ so that the cohort estimation is more like an AR(1) process and it pushes the estimation even further away from 1 when the population size is smaller. It is also consistent with our previous discussion that the AR(1) likelihood restricts the estimation for the cohort effect to be more like a strict zero mean-reverting AR(1) process. Therefore,

as a correspond the $\tilde{\alpha}_{\gamma}^{w,j^*}$ (MCMC-Mode & MCMC-Mean) of the death scenario j^* is generated to ensure that the MCMC estimated cohort effect strictly fluctuates around zero.

Therefore we could observe that $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ in Figure 6.38a (blue lines) is much more like a zero mean-reverting process than the paths for the corresponding MLE $\hat{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ (green line), the MLE $\hat{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$ (black line) and the posterior estimation $\tilde{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$ (red lines) of the England and Wales data. Recall that the trajectory of MLE $\hat{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ for death scenario j^* is in general relatively closer to the relative true parameter $\hat{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$ than the $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ since only the Poisson likelihood is maximised for calculating the MLEs and $\hat{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ restricted by the same identifiability constraints with $\hat{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$.

Note that this does not mean that MCMC generates a biased estimation for neither the α_{γ} nor the $\boldsymbol{\gamma}^{(4)}$ from the respective true parameters since we expected that without other constraints on the latent parameters (e.g. identifiability constraints), our joint posterior distribution could adjust the estimation of the $\boldsymbol{\theta}_1$ according to the dominating influence of the time series prior when the population is small such that the posterior estimation $\tilde{\boldsymbol{\theta}}_1^w$ is more like the proposed time series models and eventually provides more improved estimations for the volatility of the $\boldsymbol{\theta}_1$.

The posterior variance of the $\tilde{\alpha}_{\gamma}^{w,j^*}$ (MCMC-Mode & MCMC-Mean) is greater than the $\tilde{\alpha}_{\gamma}^{\mathrm{EW}}$ for the England and Wales data. We demonstrate the spread of the posterior distributions $\tilde{\alpha}_{\gamma}^{w,j'}$ (MCMC-Mean) for $j' = j_1, \ldots, j_{100}$ of the 100 selected death scenarios in Figure 6.35 (dotted CDFs). The sampling variation not only shifts the posterior distribution from one way to another but also has an obvious impact on the variance and skewness of the posterior distribution.

We calculated the standard deviation, skewness the mean of every $\tilde{\alpha}_{\gamma}^{w,j}$ (MCMC-Mean) for $j = 1, \ldots, N_1$ and the influence of the sampling variation on the variance and skewness is demonstrated by scatter plots of the posterior SD vs. the posterior mean (upper) and mean vs. skewness (lower) in Figure 6.36. In Figure 6.36a the

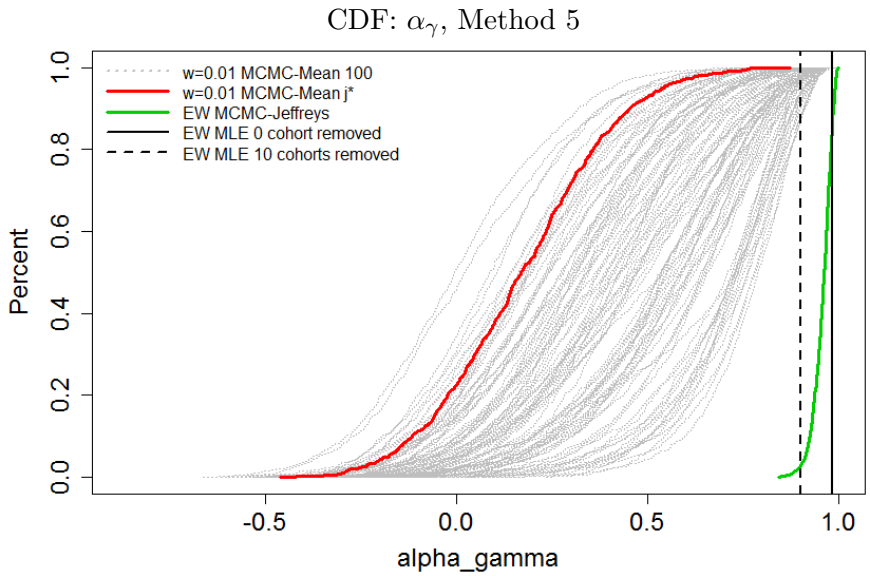
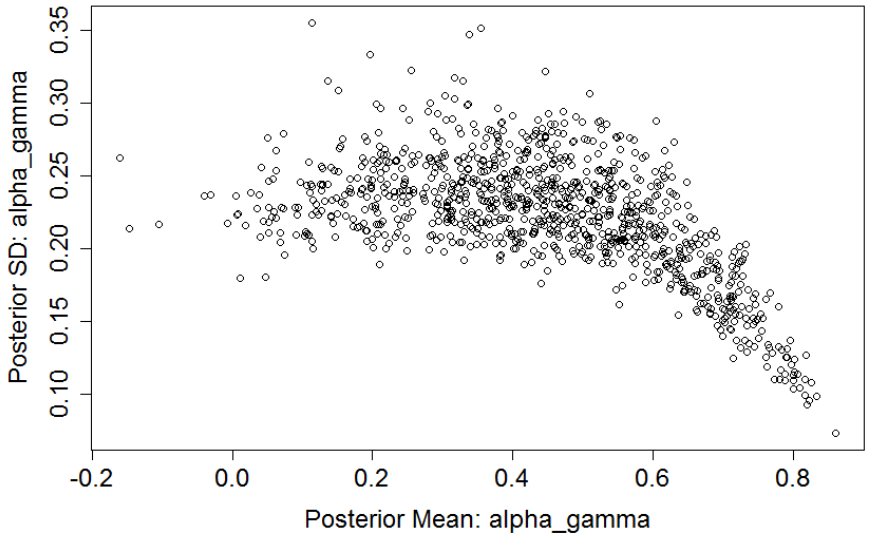


Figure 6.35: CDF according to Method 5: posterior distributions $\tilde{\alpha}_\gamma^{w,j'}$ (MCMC-Mean, dotted curves) for death scenario $j' = j_1, \ldots, j_{100}$. The red CDF is the posterior distribution $\tilde{\alpha}_\gamma^{w,j^*}$ (MCMC-Mean) for the death scenario j^*. The dashed vertical line is the point estimate of α_γ for England and Wales data given 10 cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$. The solid vertical line is the true rate of $w = 0.01$, $\hat{\alpha}_\gamma^{\text{EW}}$, i.e. no cohorts are removed from $\hat{\boldsymbol{\gamma}}^{\text{EW}}$.

sampling variation has no obvious influence on the posterior variance when the mean is not shifted to be above 0.6 (approximately). However, the posterior variance decreases after the sampling variation shifts the posterior distribution even closer to the MLE $\hat{\alpha}_\gamma^{\text{EW}}$. In Figure 6.36b a linear trend could be observed between the skewness and the mean of the posterior distribution. The reason for these findings are that when the sampling variation shifts the $\tilde{\alpha}_\gamma$ closer to 1, the trajectory of the $\tilde{\boldsymbol{\gamma}}^{(4)}$ is more like a random walk rather than a stationary zero mean-reverting AR(1) process. Therefore the AR(1) prior restricts the posterior variance to ensure that the MCMC estimation of the cohort effect follows its proposed AR(1) model. Thus, the posterior estimation $\tilde{\alpha}_\gamma^{\text{EW}}$ of the England and Wales data tails to the left side with much smaller posterior variance when the population is large and the estimated cohort effect is more affected by the dominating Poisson likelihood. See the characteristic statistics of the posterior distribution $\tilde{\alpha}_\gamma$ with respect to population size and the distribution of the posterior means $\bar{\tilde{\alpha}}_\gamma^w$ in Table 6.8 and 6.7 respectively.

We demonstrate the influence of the sampling variation on the stationary vari-

(a) Scatter: Posterior mean vs. variance, Method 3



(b) Scatter: Posterior mean vs. variance, Method 3

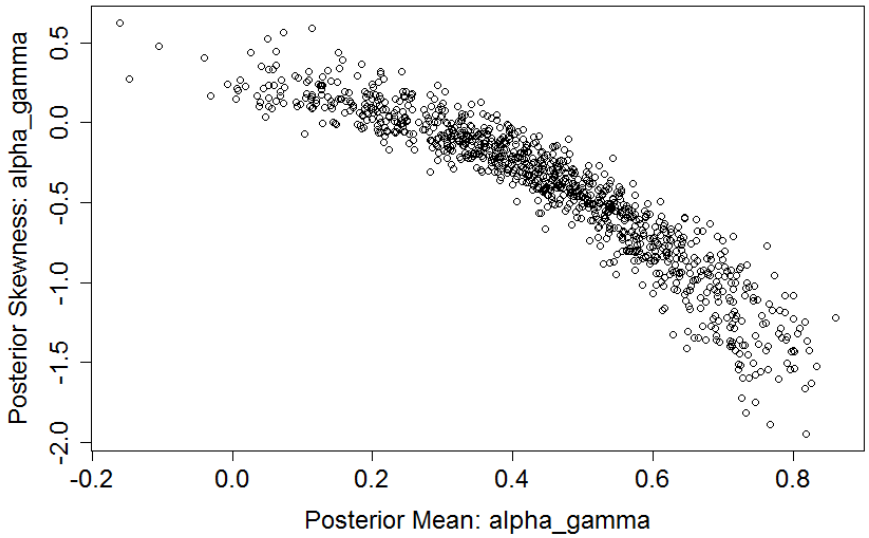


Figure 6.36: Scatter plot of: posterior mean vs. variance; posterior mean vs. skewness, of $\tilde{\alpha}_\gamma^w$ according to the posterior distributions of the N_1 death scenarios.

ance of the posterior distribution in Figure 6.37, based on the posterior distributions of $j' = j_1, \ldots, j_{100}$ selected death scenarios conditional on MCMC-Mean. Recall the formula for calculating the stationary variance:

$$\text{Var} = \frac{\sigma_\gamma^2}{(1-\alpha_\gamma^2)}. \tag{6.1}$$

We could see that unsurprisingly the sampling variation shifts the distribution of the stationary variance from one way to another (grey CDFs) and the size of shift

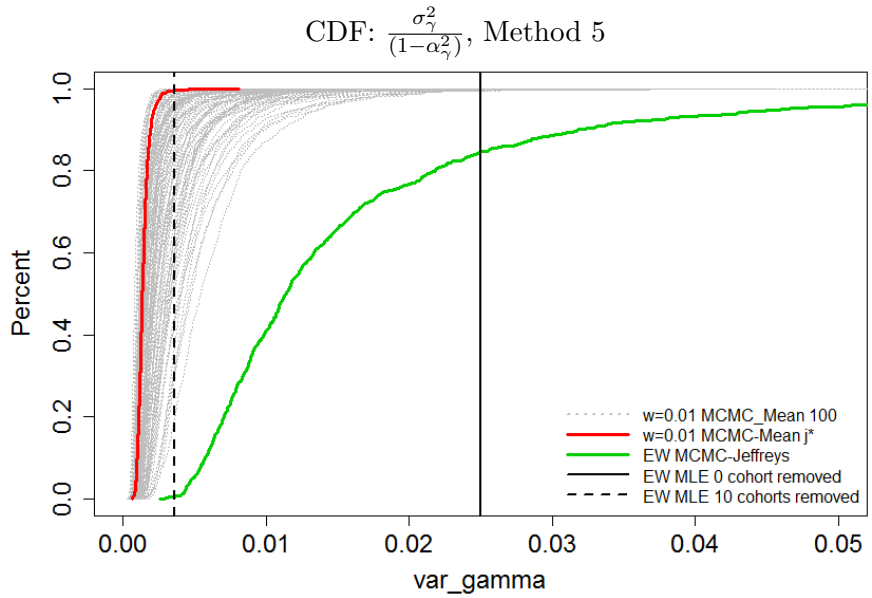
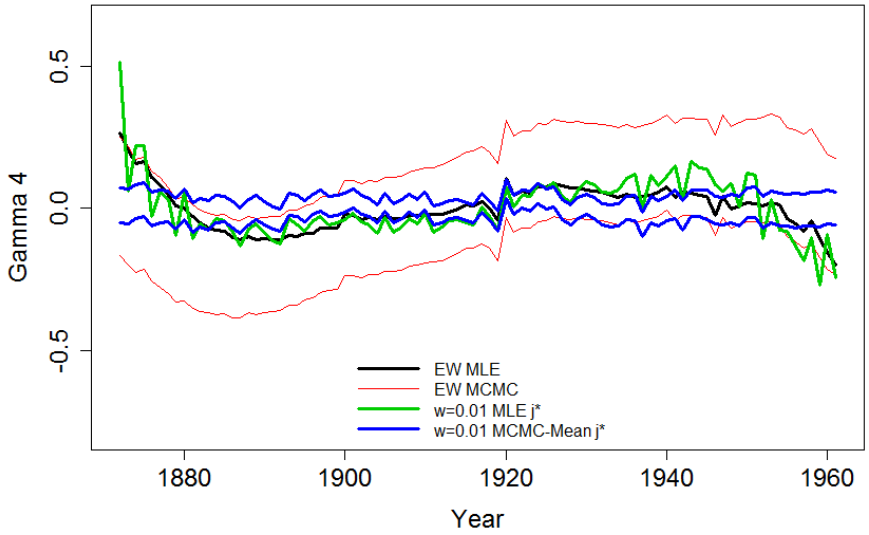


Figure 6.37: CDF according to Method 5: the variance of stationary AR(1) process $\frac{\sigma_\gamma^2}{(1-\alpha_\gamma^2)}$ for death scenario $j' = j_1, \ldots, j_{100}$ (grey curves), death scenario j^* (red curve), England and Wales data (green curve). The solid and dashed vertical lines are the stationary variance for the England and Wales data MLE with no short cohort and 10 short cohorts removed respectively.

varies from one death scenario to another and sometimes could be very large. The distribution of the England and Wales data MCMC (green CDF) is relatively higher and wider spread compared with the $w = 0.01$. The stationary variance based on the England and Wales data MLE (vertical solid line) is in general higher than the MCMC of the England and Wales data and $w = 0.01$ mainly because the MLE of α_γ for the England and Wales data is relatively higher than the respective posterior distribution of either the England and Wales or the $w = 0.01$. The reason for a higher estimated coefficient of England and Wales data MLE is discussed in the previous few paragraphs.

We at last demonstrate the spread of the posterior distributions $\tilde{\gamma}_{c,j}^{(4),w}$ (MCMC-Mean) for $j = 1, \ldots, N_1$ of all the N_1 simulated death scenarios at some years of birth in Figure 6.39 (dotted CDFs). Once again the sampling variation shifts the posterior distribution from one way to another and the size of the shift varies upon the death scenarios with relatively much smaller impact on the posterior variance. For te selected years, the MLEs $\hat{\gamma}_c^{(4),\text{EW}}$ of the England and Wales data for $c = 1882$,

(a) Historical $\boldsymbol{\gamma}^{(4)}$, Method 4



(b) Historical $\boldsymbol{\gamma}^{(4)}$, Method 6

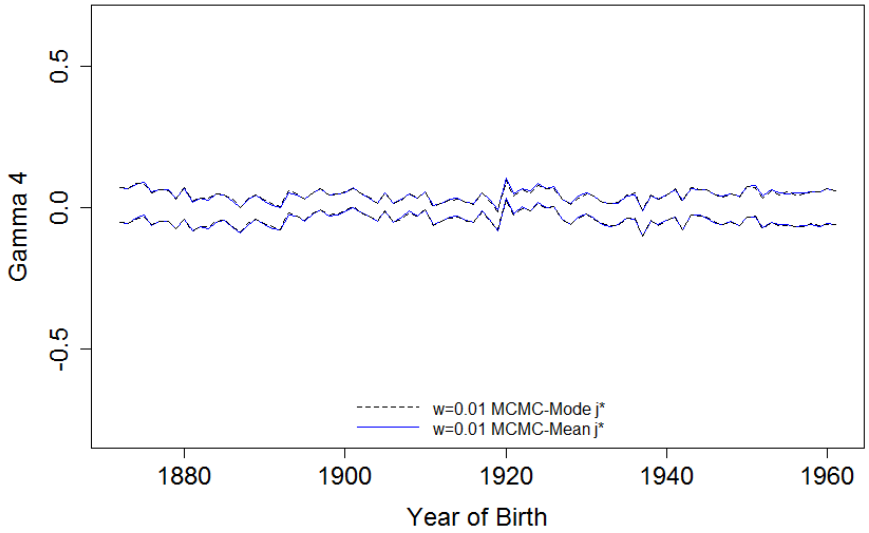


Figure 6.38: Credibility Interval: posterior distribution $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$ of the simulated death scenario j^*, given the MCMC-Mean (upper, blue lines; lower, blue solid lines) and the MCMC-Mode (lower, dashed lines). The red lines and the black line in the upper figure is posterior distribution $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$ and the MLE $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$ respectively for the England and Wales data. The green line is $\hat{\boldsymbol{\gamma}}_{j^*}^{(4),w}$, the MLE of death scenario j^*. Note that the upper and lower bound are the 95% and 5% quantile of the posterior distribution.

1912, and 1952 are approximately on the centre of the respective spread of the distributions. The sampling variation shifting the posterior distributions of the other years are demonstrated in Figure D.4 in Appendix D.

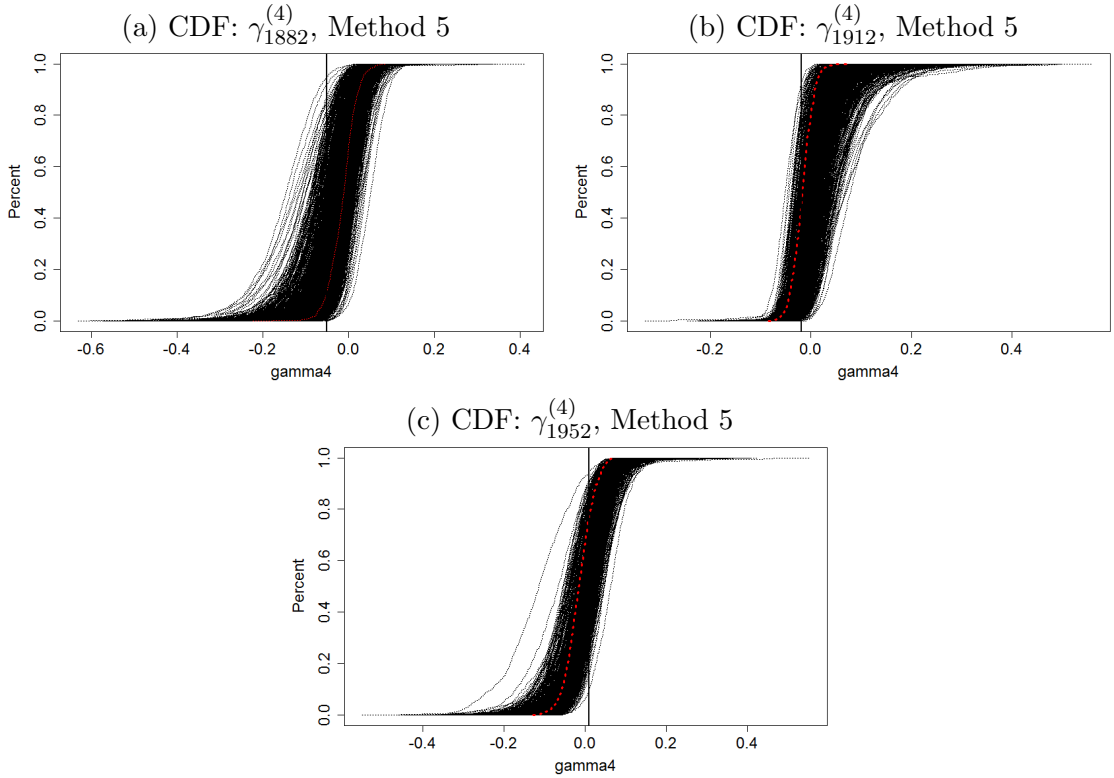


Figure 6.39: The CDF of posterior distribution for $\boldsymbol{\gamma}^{(4)}$ with selected years of birth 1882, 1912 and 1952, according to Method 5. Note that each dotted curve represents the posterior distribution for one randomly selected death scenario and all the N_1 CDFs are plotted. The red curve is the CDF for the j^* death scenario.

6.2.3 Fitted Death Rate

We have found strong evidence for the latent parameters $\boldsymbol{\theta}_1$ to be smoothed by the time series prior functions. We now investigate to what extent the Bayesian approach can smooth the fitted death rates given a large sampling variation due to the small population size. A reminder of some of the notations we applied in this section are shown as follows:

$\tilde{\boldsymbol{\theta}}_1^{w=0.01}$ = the posterior estimation of the latent parameters for $w = 0.01$.

$\tilde{\boldsymbol{\theta}}_1^{w,j}$ = the posterior estimate of the latent parameters for the j^{th} death scenario, where $j = 1, \ldots, N_1$ and $w = 0.01$.

$\tilde{\boldsymbol{\theta}}_1^{w,j,(k)}$ = the k^{th} sample drawn for the $\tilde{\boldsymbol{\theta}}_1^{w,j}$ of the j^{th} death scenario, where $k = 1, \ldots, N_2$ and $w = 0.01$.

Denote as

$$\tilde{\theta}_1^{w} = (\tilde{\kappa}_t^{(1),w}, \tilde{\kappa}_t^{(2),w}, \tilde{\kappa}_t^{(3),w}, \tilde{\gamma}_c^{(4),w}),$$

$$\tilde{\theta}_1^{w,j} = (\tilde{\kappa}_{t,j}^{(1),w}, \tilde{\kappa}_{t,j}^{(2),w}, \tilde{\kappa}_{t,j}^{(3),w}, \tilde{\gamma}_c^{(4),w,j}),$$

$$\tilde{\theta}_1^{w,j,(k)} = (\tilde{\kappa}_{t,j}^{(1),w,(k)}, \tilde{\kappa}_{t,j}^{(2),w,(k)}, \tilde{\kappa}_{t,j}^{(3),w,(k)}, \tilde{\gamma}_c^{(4),w,j,(k)})$$

Denote as

$$m(\tilde{\theta}_1^{w=0.01}, t, x) = \{m(\tilde{\theta}_1^{w,j}, t, x)\}_{j=1,\ldots,N_1},$$

the fitted death rate of year t and age x, conditional on the posterior estimation $\tilde{\theta}_1^{w=0.01}$, where

$$m(\tilde{\theta}_1^{w,j}, t, x) = \{m(\tilde{\theta}_1^{w,j,(k)}, t, x)\}_{k=1,\ldots,N_2}$$

is the k^{th} observation of the fitted death rate for the j^{th} death scenario, conditional on $\tilde{\theta}_1^{w,j,(k)}$, the k^{the} sample of the posterior estimate $\tilde{\theta}_1^{w,j}$ of the j^{th} death scenario. The $m(\tilde{\theta}_1^{w,j,(k)}, t, x)$ can be calculated according to Equation (2.5) and (2.4), more specifically,

$$\begin{aligned}
\text{logit } q(\tilde{\theta}_1^{w,j,(k)}, t, x) &= \tilde{\kappa}_{t,j}^{(1),w,(k)} + \tilde{\kappa}_{t,j}^{(2),w,(k)}(x - \bar{x}) \\
&\quad + \tilde{\kappa}_{t,j}^{(3),w,(k)}((x - \bar{x})^2 - \hat{\sigma}_x^2) + \tilde{\gamma}_c^{(4),w,j,(k)} \\
m(\tilde{\theta}_1^{w,j,(k)}, t, x) &= -\log(1 - q(\tilde{\theta}_1^{w,j,(k)}, t, x))
\end{aligned}$$

where $q(\tilde{\theta}_1^{w,j,(k)}, t, x)$ is the corresponding fitted mortality rate at year t and age x.

The study will be following the idea of Method 4 and 5, see Page 143. We will start with studying the impact of fixing the mode of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ on the fitted death rates. See Figure 6.40 for the comparison of the fitted rates $m(\tilde{\theta}_1^{w,j^*}, t, x)$, given the MCMC-Mean (red lines) and MCMC-Mode (black lines) for age 65. The upper and lower bounds are 95% and 5% quantile of the MCMC samples. Visually, the two choices of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ generate approximately the same

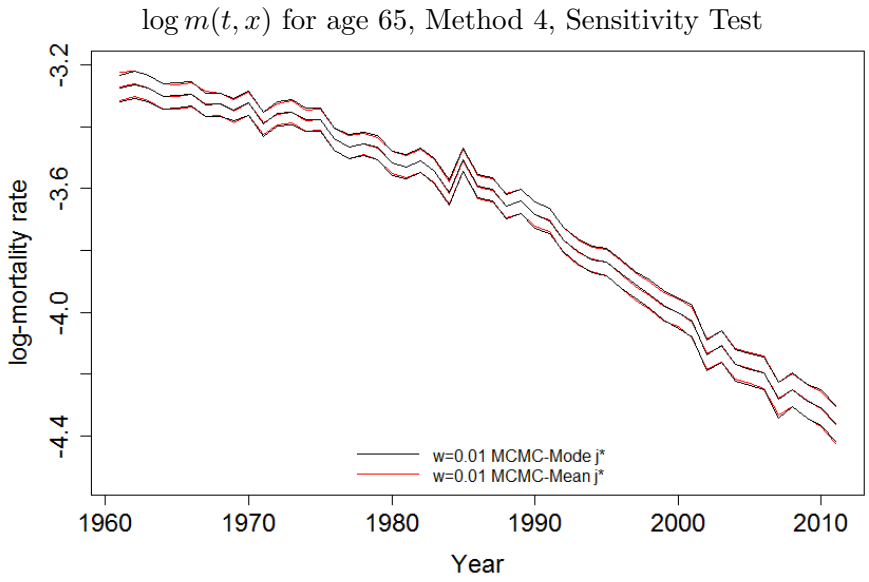


Figure 6.40: Comparison: the distribution of $m(\tilde{\theta}_1^{w,j^*}, t, x)$ given the MCMC-Mean (red lines) and the MCMC-Mode (black lines) at age 65 for the death scenario j^*. The upper and lower bound for the intervals are 95% and 5% quantiles of the MCMC samples.

fitness for age 65 as we expected. Similar conclusions can be inspected for the other ages (55, 75, 85) in Figure D.5 in Appendix D.

Recall the Method 4 and 5 in the context of the death rates as follows:

Method 4 Compare the distribution of the fitted death rate

$$m(\tilde{\theta}_1^{w,j^*}, t, x) = \{m(\tilde{\theta}_1^{w,j^*,(k)}, t, x)\}_{k=1,\ldots,N_2}$$

for the j^* death scenario given N_2 samples of the the posterior estimation $\tilde{\theta}_1^{w,j^*}$ with the fitted rate $m(\hat{\theta}_1^{w,j^*}, t, x)$ give the point estimate $\hat{\theta}_1^{w,j^*}$ for the death scenario j^*.

Method 5 For the 100 randomly selected death scenarios, calculate the fitted death rate

$$m(\tilde{\theta}_1^{w,j'}, t, x) = \{m(\tilde{\theta}_1^{w,j',(k)}, t, x)\}_{k=1,\ldots,N_2}$$

for each $j' = j_1, \ldots, j_{100}$, in order to generate an approximation of the "distribution" of the finite-sample distribution of the fitted death rate, denoted

as

$$m'(\tilde{\theta}_1^w, t, x) = \{m(\tilde{\theta}_1^{w,j'}, t, x)\}_{j'=j_1,\ldots,j_{100}},$$

conditional on the posterior estimation $\tilde{\theta}_1^{w,j'}$ of the selected death scenarios.

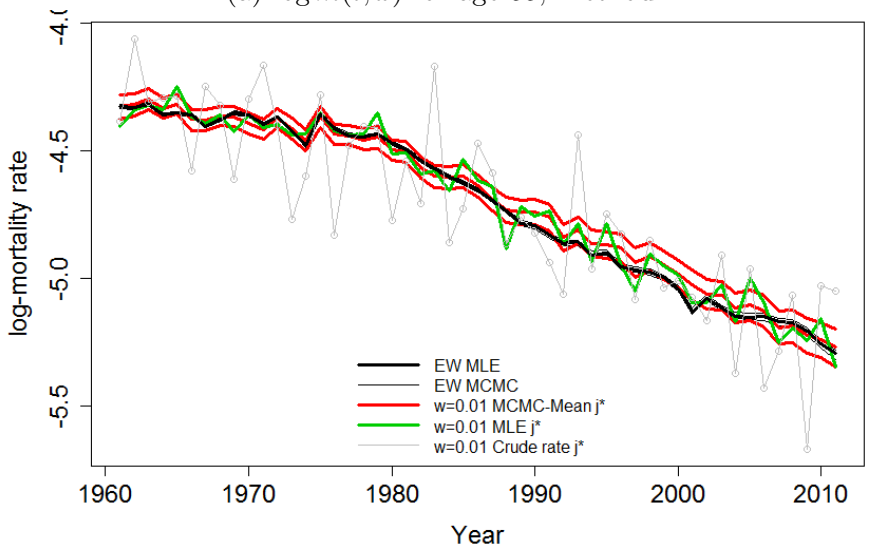
The distributions of the fitted rate according to Method 4 and 5 are plotted in Figure 6.41-6.48 for the selected ages (55, 65, 75, 85) with respect to the calendar year t. In particular, 6.41a-6.47a include the following lines:

- EW MLE: the log-scaled fitted death rates conditional on $\hat{\theta}_1^{\text{EW}}$ for the England and Wales data over the period at the selected ages.
- EW MCMC: the credibility intervals of the log-scaled fitted death rates conditional on $\tilde{\theta}_1^{\text{EW}}$ for the England and Wales data over the period at the selected ages.
- $w = 0.01$ MCMC-Mean j^* : the credibility intervals of the log-scaled fitted death rates conditional on $\tilde{\theta}_1^{w,j^*}$ for the death scenario j^* with the prior setting MCMC-Mean over the period at the selected ages.
- $w = 0.01$ MLE j^*: the log-scaled fitted death rates conditional on the MLE $\hat{\theta}_1^{w,j^*}$ for the death scenario j^* over the period at the selected ages.
- $w = 0.01$ Crude rate j^*: the log-scaled simulated crude death rates for the death scenario j^* over the period at the selected ages.

Note that all the upper and lower bounds are constructed by calculating the 95% and 5% quantiles respectively of the posterior distribution of the fitted rates.

In general, the distribution of $m(\tilde{\theta}_1^{w,j^*}, t, x)$ has greater variance than $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ at age $x = 55$ for $t = t_1, \ldots, t_{n_y}$. Same results could be observed for the other selected ages, see for example, age 65 (Figure 6.43a), 75 (Figure 6.45a), 85 (Figure 6.47a). This is consistent with the fact that the sampling variation of the death counts, and hence of the death rates are greatly increased by reducing the population size.

(a) $\log m(t,x)$ for age 55, Method 4



(b) Mean of $\log m(t,x)$ for age 55, Method 5

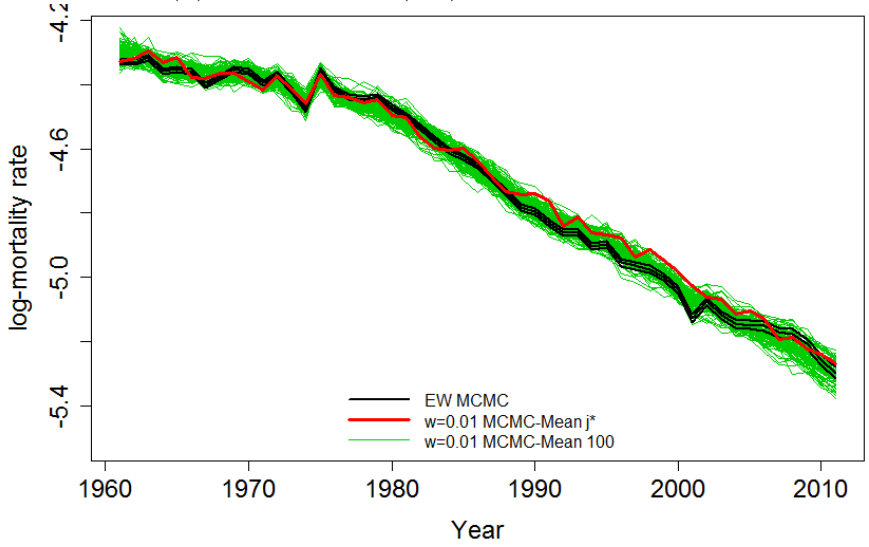


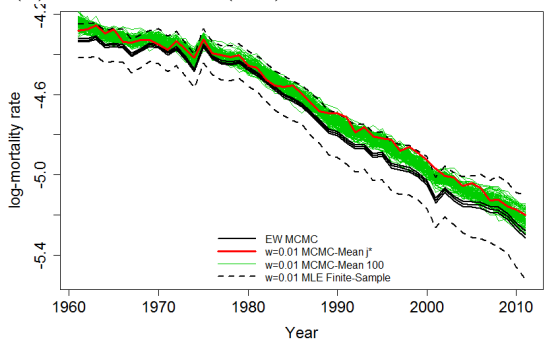
Figure 6.41: Upper: comparison of the distribution of the log-scaled fitted death rate at age 55 $m(\tilde{\theta}_1^{w,j^*},t,x)$ (red lines) given the posterior estimation $\tilde{\theta}_1^{w,j^*}$ with the fitted rate $m(\hat{\theta}_1^{w,j^*},t,x)$ given the point estimate $\hat{\theta}_1^{w,j^*}$ (green line) for death scenario j^*, referring to the true rate of the simulated death scenarios $m(\hat{\theta}_1^{\mathrm{EW}},t,x)$ (bold black solid line) conditional on the true rate $\hat{\theta}_1^{\mathrm{EW}}$. The fine black solid lines are the distribution of $m(\tilde{\theta}_1^{\mathrm{EW}},t,x)$ given the posterior estimation $\tilde{\theta}_1^{\mathrm{EW}}$ for the England and Wales data. The grey line is the crude simulated death rate for death scenario j^*. Lower: comparison of the distribution of the mean of $m'(\tilde{\theta}_1^{w},t,x)$ with the distribution of $m(\hat{\theta}_1^{w},t,x)$ given the finite-sample point estimate $\hat{\theta}_1^{w}$ (dashed lines). That is, each green line represents the mean of $m'(\tilde{\theta}_1^{w,j'},t,x)$ for $j'=j_1,\ldots,j_{100}$. The red line is the mean of $m(\tilde{\theta}_1^{w,j^*},t,x)$ for age 55. All the upper and lower bound are constructed by calculating the 95% and 5% quantiles of the corresponding samples.

Recall that the Bayesian approach provides as good fit to the crude England and Wales males' death rates as the MLE does, especially the impact of smoothness

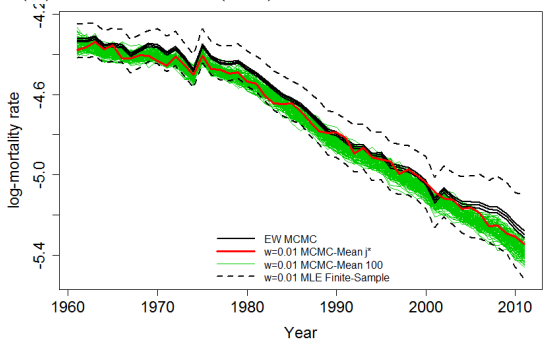
by adding the time series prior is not significant by having a large population and the Poisson likelihood dominates the posterior distribution. For simulated death scenario j^*, we can see that the trajectory of the fitted rate $m(\hat{\theta}_1^{w,j^*}, t, x)$ (green line) conditional on the MLE $\hat{\theta}_1^{w,j^*}$ is much more fluctuate than both the true rate $m(\hat{\theta}_1^{\text{EW}}, t, x)$ (bold black line) and the distribution of $m(\tilde{\theta}_1^{w,j^*}, t, x)$ (red lines) conditional on the true rate $\hat{\theta}_1^{\text{EW}}$ and $\tilde{\theta}_1^{w,j^*}$ respectively. More specifically, the fitted rate $m(\hat{\theta}_1^{w,j^*}, t, x)$ fails to capture most of the features of the pattern of the true rate $m(\hat{\theta}_1^{\text{EW}}, t, x)$ due to the significant variety, which implies that the fitted rate given $\hat{\theta}_1^{w,j^*}$ is significantly different from the true rate of the underlying dataset with an over-estimated volatility. Once again, this is because by scaling down the population size to $w = 0.01$, we greatly increased the sampling variation to the death counts $D_{t,x}^w$ and hence more uncertainty to $\frac{D_{t,x}^w}{wE_0(t,x)}$. Consistently the trajectory of the simulated crude death rates for death scenario j^* (grey line) over the period at each age is of much more fluctuation than the true rate. The maximum likelihood method tries to fit each crude death rate data that has such large sized variety (MLE's over-fitting problem), which inevitably results in significant noise to the parameter estimates and hence large variations to the fitted death rates that are too various to capture the features of the true rates. And as we have discussed, it is the significant noise to the parameter estimates produces a huge bias to the volatility estimation, which is one of the essential motivations of adopting the Bayesian approach. We notice that such fluctuation reduces when the age increases from 55 to 75 as more death data available at higher ages (see Figure 6.43a, 6.45a), while even more volatility can be observed for $m(\hat{\theta}_1^{w,j^*}, t, x)$ for $x = 85$ as fewer deaths counts are available at such a high age for calculating the MLE $\hat{\theta}_1^{w,j^*}$ (Figure 6.47a).

On the other hand, a significant smoothness can be observed on the trajectory of $m(\tilde{\theta}_1^{w,j^*}, t, x = 50)$ (red lines) given the posterior estimation $\tilde{\theta}_1^{w,j^*}$. Visually, the $m(\tilde{\theta}_1^{w,j^*}, t, x)$ captures most of the shapes of the trajectory of the true rate without over-estimating the volatility compared with $m(\hat{\theta}_1^{w,j^*}, t, x)$. In particular, the mean of the MCMC is approximately the same with the true rate with quite a good smooth-

(a) 95% of $\log m(t,x)$ for age 55, Method 5



(b) 5% of $\log m(t,x)$ for age 55, Method 5



(c) 95%-5% of $\log m(t,x)$ for age 55, Method 5

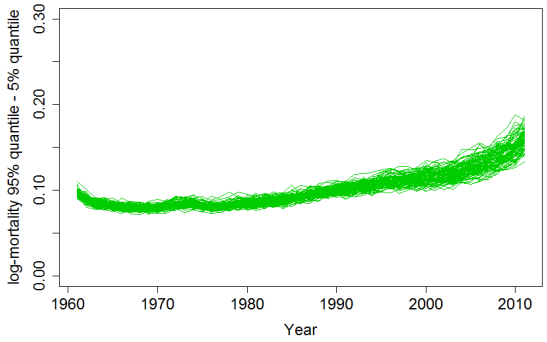
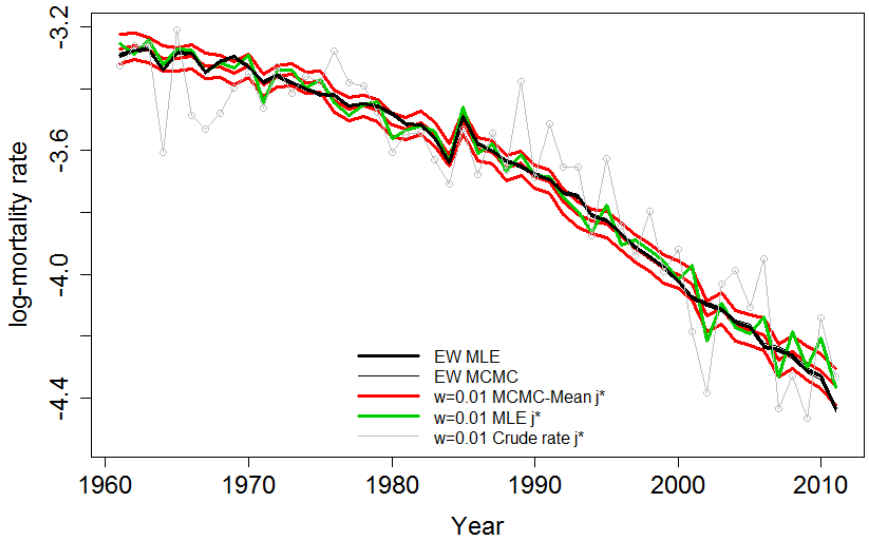


Figure 6.42: The distribution of the upper bound, lower bound and the difference between the two bounds of $m'(\tilde{\theta}_1^w, t, x)$ for age 55, according to Method 5. Note that each green line represents the credibility intervals for one randomly selected death scenario $j' = j_1, \ldots, j_{100}$. The red lines are the credibility intervals for $m(\tilde{\theta}_1^{w,j^*}, t, x)$ of the j^* death scenario.

ness when the crude death rates are of significant noise. The corresponding cohort years for age 55 is 1906 to 1956 including the cohort with only six observations at the very late cohort years. As we have discussed, the ML method over-fits the short cohorts and greatly pushes the corresponding point estimates away, which is one of the reasons for the large fluctuation to the corresponding fitted rates. The Bayesian approach, as we know, balances the short cohorts with wider posterior distribution as we can see the distribution of $m(\tilde{\theta}_1^{w,j^*}, t, x = 55)$ is much more smoothed than the MLE and the central MCMC fitted rates are more similar with the true rates for the respect late years. The posterior distribution, as we can see, is consistently slightly wider at the ending years than it is at the previous years.

Similar results can be found for the other ages as well. In particular, the MCMC still provides a specifically better fit at a higher age x =85 compared with MLE

(a) $\log m(t,x)$ for age 65, Method 4



(b) Mean of $\log m(t,x)$ for age 65, Method 5

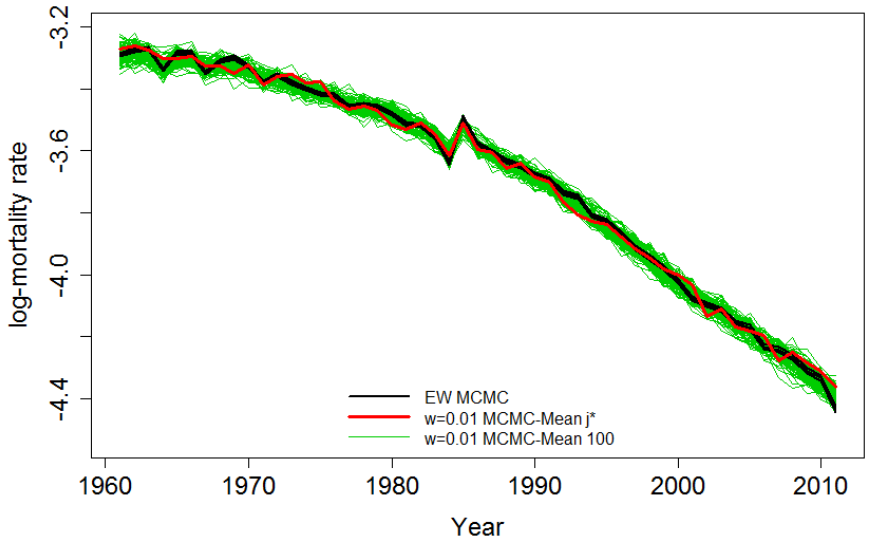


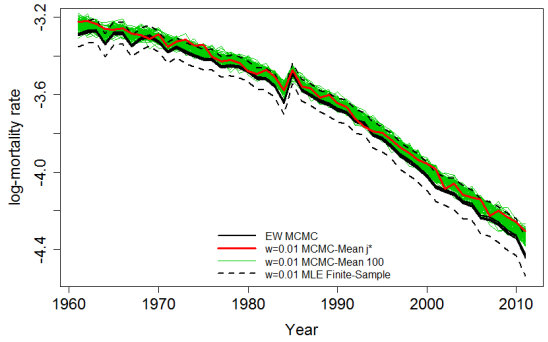
Figure 6.43: Upper: comparison of the distribution of the log-scaled fitted death rate at age 65 $m(\tilde{\theta}_1^{w,j^*},t,x)$ (red lines) given the posterior estimation $\tilde{\theta}_1^{w,j^*}$ with the fitted rate $m(\hat{\theta}_1^{w,j^*},t,x)$ given the point estimate $\hat{\theta}_1^{w,j^*}$ (green line) for death scenario j^*, referring to the true rate of the simulated death scenarios $m(\hat{\theta}_1^{\text{EW}},t,x)$ (bold black solid line) conditional on the true rate $\hat{\theta}_1^{\text{EW}}$. The fine black solid lines are the distribution of $m(\tilde{\theta}_1^{\text{EW}},t,x)$ given the posterior estimation $\tilde{\theta}_1^{\text{EW}}$ for the England and Wales data. The grey line is the crude simulated death rate for death scenario j^*. Lower: comparison of the distribution of the mean of $m'(\tilde{\theta}_1^{w},t,x)$ with the distribution of $m(\hat{\theta}_1^{w},t,x)$ given the finite-sample point estimate $\hat{\theta}_1^{w}$ (dashed lines). That is, each green line represents the mean of $m'(\tilde{\theta}_1^{w,j'},t,x)$ for $j'=j_1,\ldots,j_{100}$. The red line is the mean of $m(\tilde{\theta}_1^{w,j^*},t,x)$ for age 65. All the upper and lower bound are constructed by calculating the 95% and 5% quantiles of the corresponding samples.

in Figure 6.47a. The MLE provides an extremely poor fit for the simulated crude death scenario j^* at a higher age as we can see the fitted rate is significantly differed

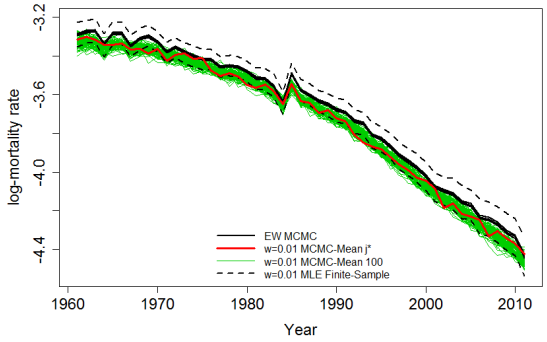
from the true rate with large variety, the size of which is greater compared with its fitness at younger ages. Once again this is because the MLE tries to fit the significantly fluctuated simulated crude death rates and therefore inevitably missed the information of the true rates for the death scenario. Such miss-fitting is amplified when the age is higher and therefore relatively even more fluctuation on the MLE fitted rates than the lower ages. However, a good smoothness could still be observed from the MCMC fitted rates and visually the trajectory is quite similar with the true rate with most of the true rates' pattern captured by the MCMC. In particular, the central MCMC is approximately the same with the true rate after year 1970 (roughly). For the years from 1961 to 1970 (i.e. includes the very early short cohorts), although the level of the MCMC mean is lower than the true rate, the MCMC fitted rate successfully captures some of the zigzags of the true rates while the MLE is totally differently from the true rates with significant volatility. We therefore could conclude that even at a relatively elder age, the MCMC still provides a good fit to the death counts and smoothness for the short cohorts.

It is worth noticing that the distribution of MCMC is not always centred around the true rate for every year. In other words, the mean of the MCMC can be lower or higher than the true rates for a range of continuous years given a certain age, even when both the true rates and the MCMC share the similar pattern. For example, Figure 6.41a shows that the central estimation of MCMC is slightly higher than the true rate approximately after year 1985-2005. This is mainly due to the significant impact of the time series prior on the joint posterior distribution that tries to add sufficient smoothness to the latent parameter estimation and compete with the Poisson likelihood. It thus tilts and shifts the latent parameter estimation to ensure that the processes of the period and cohort effects are sufficiently smoothed and looked like being generated from the time series process. Recall that the central estimation of MCMC for $\boldsymbol{\kappa}^{(1)}$ is anti-clockwise tilted around year 1980 (Figure 6.5a) and central MCMC for $\boldsymbol{\kappa}^{(3)}$ is higher than the respective true parameters approximately during year 1985-2000 (Figure 6.22a). Although the MCMC has its central estimation for

(a) 95% of $\log m(t,x)$ for age 65, Method 5



(b) 5% of $\log m(t,x)$ for age 65, Method 5



(c) 95%-5% of $\log m(t,x)$ for age 65, Method 5

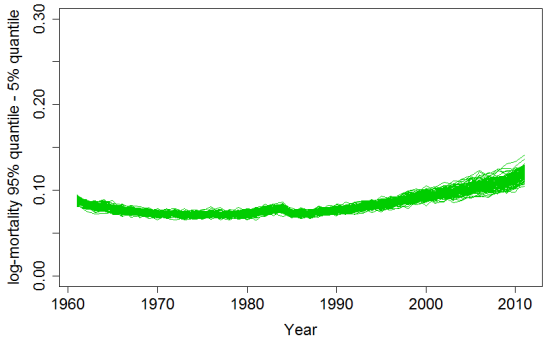
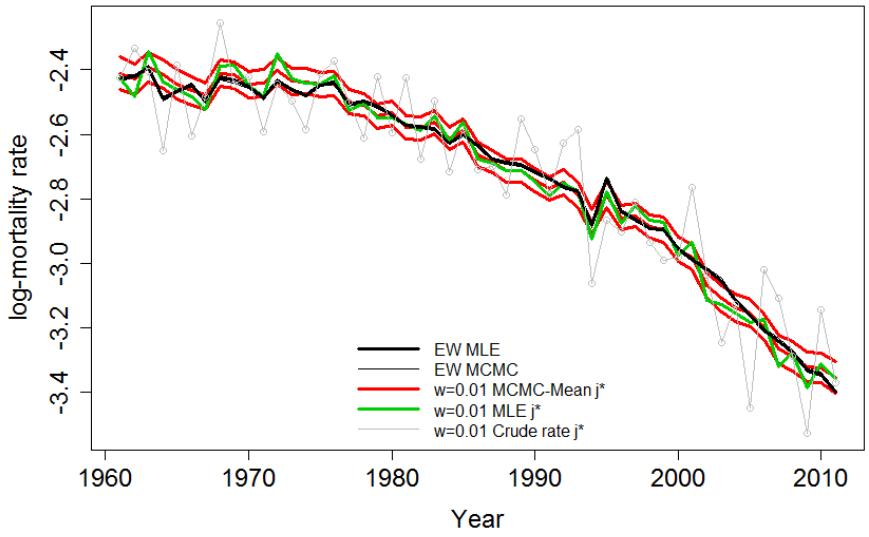


Figure 6.44: The distribution of the upper bound, lower bound and the difference between the two bounds of $m'(\tilde{\theta}_1^w, t, x)$ for age 65, according to Method 5. Note that each green line represents the credibility intervals for one randomly selected death scenario $j' = j_1, \ldots, j_{100}$. The red lines are the credibility intervals for $m(\tilde{\theta}_1^{w,j^*}, t, x)$ of the j^* death scenario.

$\boldsymbol{\kappa}^{(2)}$ less than the true rates from year 1980 to 2000 (approximately) (Figure 6.13a), the influence of which on the fitted rate is relatively too small compared with the $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\kappa}^{(3)}$ when the year increases, especially when the age is as young as 55 due to the parametric quadratic term for the age x. Further, the central estimation for the corresponding cohort years of age 55 is not significantly different from the true rate $\hat{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$ (Figure 6.38a) and the mean of MCMC for the very late short cohorts are even higher than the true rates. All of these eventually push up the MCMC fitted rates and hence a higher central estimation after year 1985 for age 55.

On the other hand, the central estimation is not always lower than the true rate for each year before 1985 since the impact of $\boldsymbol{\kappa}^{(1)}$ on the fitted rate is balanced by the other latent parameters, e.g. the corresponding cohort effect. For the other selected ages, there is no obvious trend that the distribution of MCMC fitted rates is higher

(a) $\log m(t,x)$ for age 75, Method 4



(b) Mean of $\log m(t,x)$ for age 75, Method 5

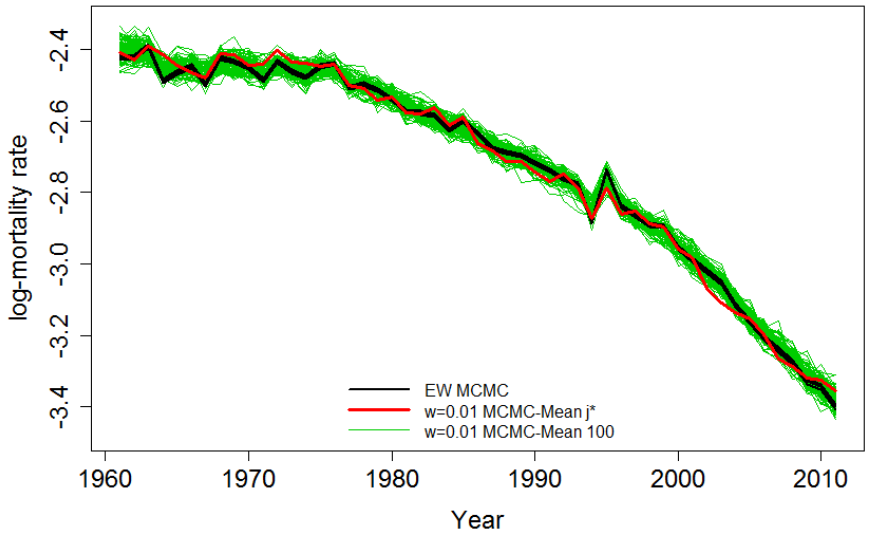
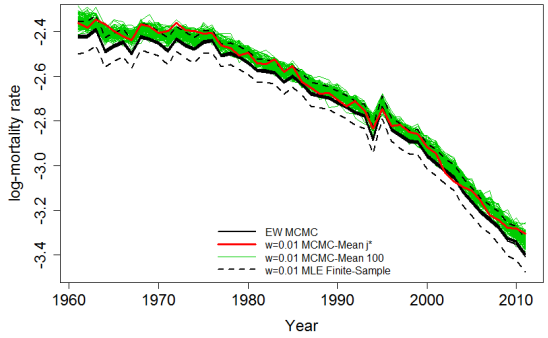


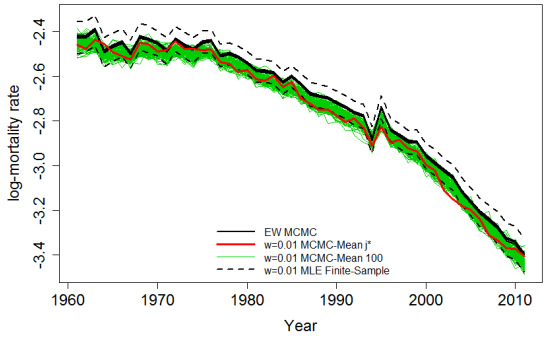
Figure 6.45: Upper: comparison of the distribution of the log-scaled fitted death rate at age 75 $m(\tilde{\theta}_1^{w,j^*},t,x)$ (red lines) given the posterior estimation $\tilde{\theta}_1^{w,j^*}$ with the fitted rate $m(\hat{\theta}_1^{w,j^*},t,x)$ given the point estimate $\hat{\theta}_1^{w,j^*}$ (green line) for death scenario j^*, referring to the true rate of the simulated death scenarios $m(\hat{\theta}_1^{\text{EW}},t,x)$ (bold black solid line) conditional on the true rate $\hat{\theta}_1^{\text{EW}}$. The fine black solid lines are the distribution of $m(\tilde{\theta}_1^{\text{EW}},t,x)$ given the posterior estimation $\tilde{\theta}_1^{\text{EW}}$ for the England and Wales data. The grey line is the crude simulated death rate for death scenario j^*. Lower: comparison of the distribution of the mean of $m'(\tilde{\theta}_1^{w},t,x)$ with the distribution of $m(\hat{\theta}_1^{w},t,x)$ given the finite-sample point estimate $\hat{\theta}_1^{w}$ (dashed lines). That is, each green line represents the mean of $m'(\tilde{\theta}_1^{w,j'},t,x)$ for $j'=j_1,\ldots,j_{100}$. The red line is the mean of $m(\tilde{\theta}_1^{w,j^*},t,x)$ for age 75. All the upper and lower bound are constructed by calculating the 95% and 5% quantiles of the corresponding samples.

or lower than the true rates for a range of continuous years since the impact of $\boldsymbol{\kappa}^{(2)}$ and $\boldsymbol{\kappa}^{(3)}$ increases as the age increases.

(a) 95% of $\log m(t,x)$ for age 75, Method 5



(b) 5% of $\log m(t,x)$ for age 75, Method 5



(c) 95%-5% of $\log m(t,x)$ for age 75, Method 5

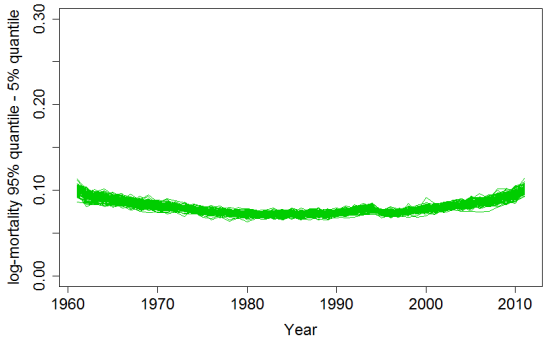
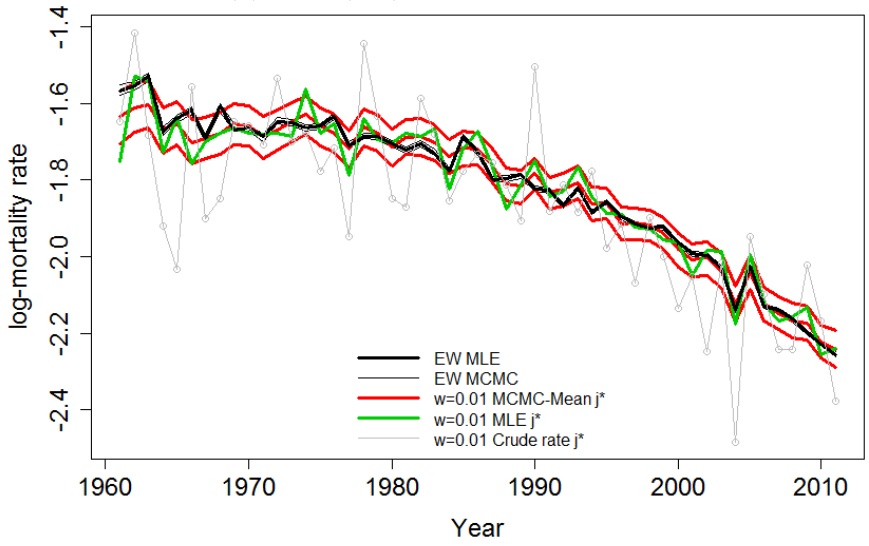


Figure 6.46: The distribution of the upper bound, lower bound and the difference between the two bounds of $m'(\tilde{\theta}_1^w, t, x)$ for age 75, according to Method 5. Note that each green line represents the credibility intervals for one randomly selected death scenario $j' = j_1, \ldots, j_{100}$. The red lines are the credibility intervals for $m(\tilde{\theta}_1^{w,j^*}, t, x)$ of the j^* death scenario.

A short conclusion could therefore be made that while fitting the small population with the Bayesian approach, we fitted time series models to the period and cohort effects according to our prior knowledge on the pattern of the latent parameters and an informative prior distribution employing prior information of the benchmark population to the volatility of the period effects. By combining the Poisson likelihood with the these prior settings for the parameters, we successfully generated much more smoothed estimation for the stochastic processes of the small population and hence sufficiently smoothed fitted rates that are not significantly different from the true rate at all the ages in the underlying dataset, including the relatively younger (x=55) and elder (x=85) ages.

We demonstrate the influence of the sampling variation on the distribution of the MCMC fitted rates according to the Method 3 and 5. By following the idea of

(a) $\log m(t,x)$ for age 85, Method 4



(b) Mean of $\log m(t,x)$ for age 85, Method 5

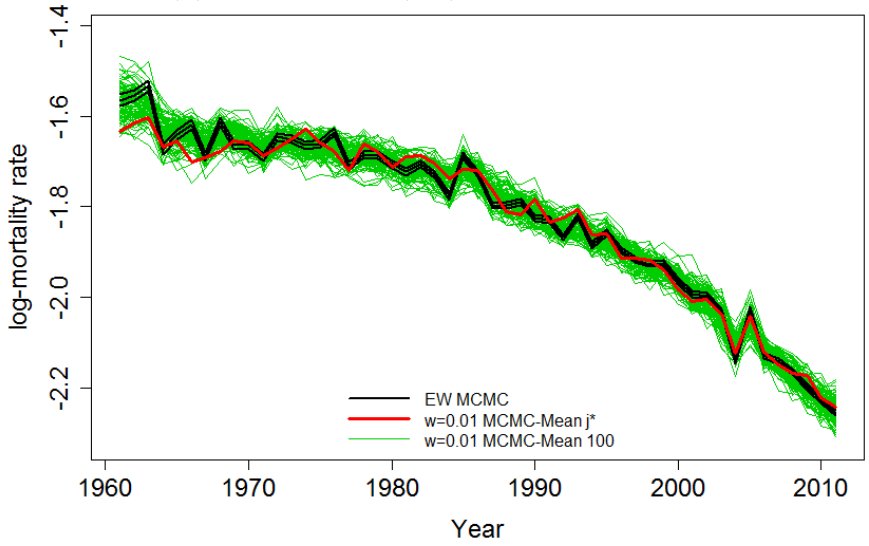


Figure 6.47: Upper: comparison of the distribution of the log-scaled fitted death rate at age 85 $m(\tilde{\theta}_1^{w,j^*},t,x)$ (red lines) given the posterior estimation $\tilde{\theta}_1^{w,j^*}$ with the fitted rate $m(\hat{\theta}_1^{w,j^*},t,x)$ given the point estimate $\hat{\theta}_1^{w,j^*}$ (green line) for death scenario j^*, referring to the true rate of the simulated death scenarios $m(\hat{\theta}_1^{\mathrm{EW}},t,x)$ (bold black solid line) conditional on the true rate $\hat{\theta}_1^{\mathrm{EW}}$. The fine black solid lines are the distribution of $m(\tilde{\theta}_1^{\mathrm{EW}},t,x)$ given the posterior estimation $\tilde{\theta}_1^{\mathrm{EW}}$ for the England and Wales data. The grey line is the crude simulated death rate for death scenario j^*. Lower: comparison of the distribution of the mean of $m'(\tilde{\theta}_1^{w},t,x)$ with the distribution of $m(\hat{\theta}_1^{w},t,x)$ given the finite-sample point estimate $\hat{\theta}_1^{w}$ (dashed lines). That is, each green line represents the mean of $m'(\tilde{\theta}_1^{w,j'},t,x)$ for $j'=j_1,\ldots,j_{100}$. The red line is the mean of $m(\tilde{\theta}_1^{w,j^*},t,x)$ for age 85. All the upper and lower bound are constructed by calculating the 95% and 5% quantiles of the corresponding samples.

Method 5, 100 death scenarios are randomly selected. For each of the j' deaths, the fitted rate $m(\tilde{\theta}_1^{w,j',(k)},t,x)$ is calculated for the k^{th} MCMC sample. By follow-

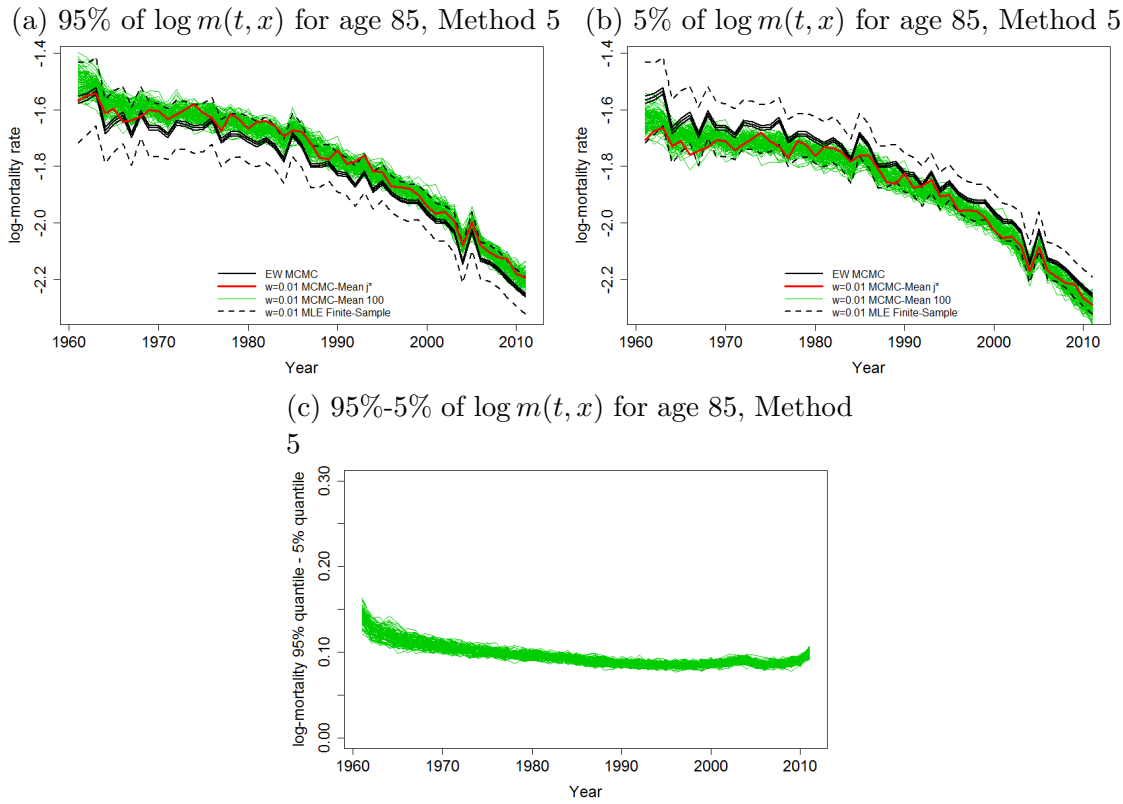


Figure 6.48: The distribution of the upper bound, lower bound and the difference between the two bounds of $m'(\tilde{\theta}_1^w, t, x)$ for age 85, according to Method 5. Note that each green line represents the credibility intervals for one randomly selected death scenario $j' = j_1, \ldots, j_{100}$. The red lines are the credibility intervals for $m(\tilde{\theta}_1^{w,j^*}, t, x)$ of the j^* death scenario.

ing the idea of Method 3, the central fit, i.e. the mean of the MCMC fitted rate $\bar{m}(\tilde{\theta}_1^{w,j'}, t, x) = \frac{1}{N_2} \sum_{k=1}^{N_2} m(\tilde{\theta}_1^{w,j',(k)}, t, x)$, is worked out for $j' = j_1, \ldots, j_{100}$. We then demonstrate the trajectories of these 100 central fitted rates for a certain age through the year range together in one plot.

For example, the trajectories of the central fitness $\bar{m}(\tilde{\theta}_1^{w,j'}, t, x = 50)$ for $j' = j_1, \ldots, j_{100}$ at age 55 for the year range 1961-2011 are plotted in Figure 6.41b (green lines). The red line is again the central fitness of death scenario j^* to demonstrate its position on the distribution of the MCMC means. The green lines shows the distribution of the central fitness of the MCMC fitted rates at age 50. One can of course plot all the N_1 death scenarios but we find the sample size 100 is sufficient to reflect how the sampling variation affects the central MCMC fitted rates. The uncertainty of the distribution for the central MCMC fitted rates is driven by the

Poisson randomness, i.e. sampling variation. Unsurprisingly, the sampling variation shifts the central MCMC fitted rates up and down around the respective true parameters. The position of the central MCMC for the death scenario j^* implies that such shift varies from one death scenario to another and therefore the influence of the sampling variation on the posterior distribution could sometimes be very large.

In general, the width of the distribution is relatively stable, especially for age 55, 65 (Figure 6.43b) and 75 (Figure 6.45b). For the elder age 85, the distribution is slightly wider at the beginning years due to the influence of the short cohorts (Figure 6.47b). By referring to the true rates and EW MCMC, we can see that the pattern of the distribution of the central fitted rates is approximately the same with both the true rates and EW MCMC and most of the patterns (zigzags) of the true rates are captured by the MCMC method. Visually, there is no single year at which the distribution of the central fitted rates is significantly lower or higher than the true rate at age 55, even for the rates after year 1985. We have discussed that the mean of $m(\tilde{\theta}_1^{w,j^*}, t, x)$ for death scenario j^* is higher than the true rate after year 1985 while Figure 6.41b suggests that this is mainly because of the Poisson randomness.

Similar findings could be inspected for the other ages and therefore we could conclude that there are no significant differences between the true rate and the central MCMC fitted rates when the Poisson randomness could shift the posterior distribution up and down.

Once again, in practice we may not have the information of 100 death samples. Thus the plots of the distribution of the central fitted rate conditional on the MCMC tells us that the distribution of the fitted rate of one death sample given the posterior distribution of the latent parameters can be shifted up and down due to the variation of the sample and such shift can be either large or small.

As a complementary, we demonstrate the influence of the sampling variation on the posterior distribution by calculating the 5% (upper right) and 95% (upper left)

quantiles of the MCMC fitted rates for each of the 100 selected death scenarios. For example, the distributions of the upper (left) and the lower (right) bounds of the MCMC fitted rates at age 55 are plotted in Figure 6.42. It is worth noticing that for all the selected ages: 65 (Figure 6.44); 75 (Figure 6.46) and 85 (Figure 6.48), the distributions of the quantiles are approximately within the confidence intervals of the finite-sample fitted rates given the MLEs.

We at last demonstrate the influence of the sampling variation on the differences between the two quantiles based on the posterior distributions of the selected 100 death scenarios in Figure 6.42c-6.48c for age 55, 65, 75 and 85 respectively. As we expected, the widths of the differences are relatively stable throughout the years for most of the ages with slightly wider distribution at the last and early few years of age 55 and 85 respectively due to the short cohorts. We could see that for each age, the spread of the quantile difference is relatively much narrower in scale compared with the relative spread of either the 95% or the 5% quantile over the period, which implies that the sampling variation mainly shifts the posterior distribution of the fitted rates up and down without significantly affecting the variance of the posterior distribution for the fitted rates.

6.2.4 Summary

In this section, we carried out a comparison for the estimation of $\boldsymbol{\theta}$ for the small population $w = 0.01$ whose true rate is the point estimates of the benchmark population, between the two-stage and the Bayesian approach. The finds are summarised as follows.

For the volatility of the period effects $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, the Bayesian method that combines the two stages into one smoothed the period effect $\boldsymbol{\kappa}^{(i)}$, $i = 1, 2, 3$ by providing a much improved estimation $\tilde{\boldsymbol{V}}_{\epsilon}^{w}$ for the volatility compared with the point estimate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}$ and the uncertainty of the former is also greatly smaller than that of the latter.

The sensitivity test indicates that given an Inverse Wishart prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ and the underlying prior settings, the posterior estimation for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is sensitive to the level of the prior mean and variance for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ and we found that fixing the mode of $\boldsymbol{V}'_{\boldsymbol{\epsilon}}$s prior to the respective point estimate of the benchmark population provides a better estimation than the MCMC-Mean. More specifically, the MCMC-Mode results in an approximately unbiased estimation for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ and the estimation for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ is only slightly higher than the true rate with non-significant difference.

For the drift of the period effects $\boldsymbol{\mu}$, the Bayesian estimation $\tilde{\boldsymbol{\mu}}_i^w$ is approximately centred around the true rate for $i = 1, 2$. The $\tilde{\boldsymbol{\mu}}_3^w$ is shifted to the left hand side of the true rate and the point estimate $\hat{\boldsymbol{\mu}}_3^w$. The impact of $\boldsymbol{\mu}_3$ on the mortality improvement from year $t-1$ to t is relatively small compared to the impact of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ do, especially for ages that are not too young or too old. Thus having on average a lower Bayesian estimation $\tilde{\boldsymbol{\mu}}_3^w$ implies that the mortality improvement for small population $w = 0.01$ given the Bayesian approach is not as large as the one given the two-stage approach and this is particularly true for the mortality improvement at very young and very high ages. The result of the sensitivity test indicates that increasing the prior mean and variance of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ has non-significant impact on the posterior estimation $\tilde{\boldsymbol{\mu}}_i^w$ for $i = 1, 2, 3$.

For the period effect $\boldsymbol{\kappa}^{(i)}$ for $i = 1, 2, 3$, we can see the process is smoothed by the Bayesian approach, with significantly smaller estimation variance. The trajectory of the posterior estimation $\boldsymbol{\kappa}^{(i),w}$ for $i = 1, 2, 3$ are different from both the true rate and the finite-sample point estimate since no identifiability constraints were applied neither during the MCMC algorithm nor to the samples drawn from the posterior distribution. The posterior estimation $\tilde{\kappa}_t^{(i),w}$ for $i = 1, 2, 3$ itself is not sensitive to increasing the mean and the variance of the $\boldsymbol{V}'_{\boldsymbol{\epsilon}}$s prior distribution.

For the volatility of the cohort effect σ_γ, the Bayesian approach once again provides a much improved estimation with a much smaller estimation uncertainty than that of the finite-sample point estimate. The posterior estimation is barely

affected by the sensitivity test.

The coefficient α_γ of the AR(1) model assumed for the cohort process, by being assumed to follow a strong beta prior in order to push the estimation away from 1, the Bayesian estimation is significantly lower than the both of the estimate of the England and Wales data and the finite-sample point estimate $\hat{\alpha}_\gamma^w$ as we expected. Similarly, the posterior distribution is not significantly affected by the sensitivity test.

At last, for the cohort effect $\boldsymbol{\gamma}^{(4)}$, we can see that the Bayesian approach provides sufficient smoothness to the entire process. The posterior variance is much smaller than that of the finite-sample point estimates as the large sampling variation is balanced by the prior distribution. This is particularly true for the estimation of the cohorts at the very early and very late of the years of birth. The Bayesian approach addressed the over-fitting problem by balancing the low level information of the short cohorts with the time series prior. The trajectory of the Bayesian estimation is significantly different from that the point estimate of England and Wales data and the finite-sample estimates for $w = 0.01$. This reflects the influence of the strong beta distribution for the α_γ and once again emphasis the impact of prior distribution competes with the Poisson likelihood that makes the cohort estimation look like generated by the time series model. The different pattern of the trajectories is also because no identifiability constraints were explicitly applied for the Bayesian method. In the end, the posterior distribution is hardly affected by the sensitivity test as we expected.

As for the fitted rate, all the findings satisfy our expectation of the impact of prior distribution for modelling a small population. More specifically, unlike the two-stage approach obscuring the true signal of the population due to the large sampling variation to the death counts, over-estimating the volatility of the stochastic process and producing the latent parameter estimation with large uncertainty, the Bayesian method balances the large sampling variation with the time series prior and the

informative prior distribution for the volatility so that most of the true information of the benchmark population can be involved to generate a much smoothed estimation for each stochastic latent process with smaller uncertainty and as we can see the fitted rates conditional on the these estimation successfully capture most of the features of the true rates.

6.2.5 Projection

We simulate the posterior predictive distribution of the death rate for the small population $w = 0.01$ by following the steps described in Section 5.4.5.

Projecting $\boldsymbol{\theta}$ for the Small Population

In this section, the latent parameter $\boldsymbol{\theta}_1$ of the small population is projected. Consistent notation system with projecting the England and Wales data is adopted for $w = 0.01$. More specifically, we define as:

$\tilde{\boldsymbol{\theta}}_1^{'\,w}$ = the projected latent parameters for $w = 0.01$.

$\tilde{\boldsymbol{\theta}}_1^{'\,w,j}$ = the projected latent parameters for the j^{th} death scenario, where $j = 1, \ldots, N_1$ and $w = 0.01$.

$\tilde{\boldsymbol{\theta}}_1^{'\,w,j,(k)}$ = projection for the k^{th} sample drawn for the $\tilde{\boldsymbol{\theta}}_1^{w,j}$ of the j^{th} death scenario, where $k = 1, \ldots, N_2$ and $w = 0.01$.

$\tilde{\boldsymbol{\theta}}_1^{'\,w,j,(k,l)}$ = the l^{th} sample path of the projection for the k^{th} sample of the j^{th} death scenario, where $l = 1, \ldots, N_4$.

Correspondingly we denote as:

$$\tilde{\theta}_1^{'\,w} = (\tilde{\kappa}_{t'}^{'\,(1),w}, \tilde{\kappa}_{t'}^{'\,(2),w}, \tilde{\kappa}_{t'}^{'\,(3),w}, \tilde{\gamma}_{c'}^{'\,(4),w}),$$

$$\tilde{\theta}_1^{'\,w,j} = (\tilde{\kappa}_{t',j}^{'\,(1),w}, \tilde{\kappa}_{t',j}^{'\,(2),w}, \tilde{\kappa}_{t',j}^{'\,(3),w}, \tilde{\gamma}_{c',j}^{'\,(4),w}),$$

$$\tilde{\theta}_1^{'\,w,j,(k)} = (\tilde{\kappa}_{t',j}^{'\,(1),w,(k)}, \tilde{\kappa}_{t',j}^{'\,(2),w,(k)}, \tilde{\kappa}_{t',j}^{'\,(3),w,(k)}, \tilde{\gamma}_{c',j}^{'\,(4),w,(k)})$$

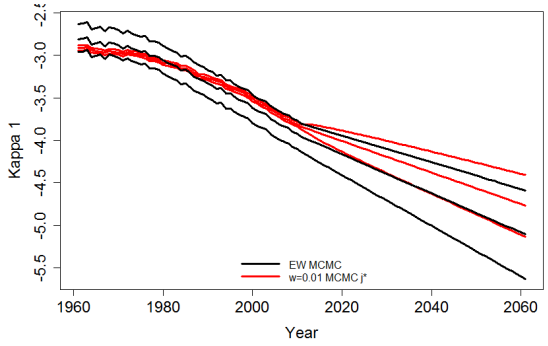
$$\tilde{\theta}_1^{'\,w,j,(k,l)} = (\tilde{\kappa}_{t',j}^{'\,(1),w,(k,l)}, \tilde{\kappa}_{t',j}^{'\,(2),w,(k,l)}, \tilde{\kappa}_{t',j}^{'\,(3),w,(k,l)}, \tilde{\gamma}_{c',j}^{'\,(4),w,(k,l)}),$$

where t' and c' are the projected years and years of birth respectively.

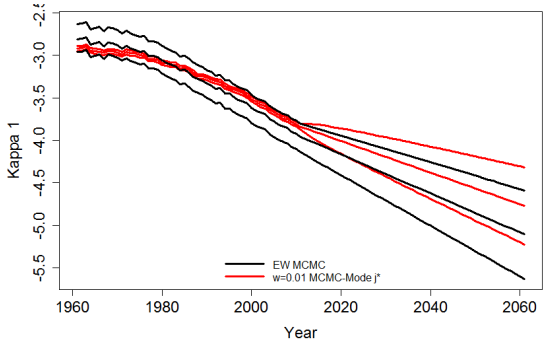
The study of the latent parameter projection will be focusing on the idea of Method 4 and 5 by comparing the distribution of $\tilde{\theta}_1^{'\,w,j^*}$ with $\tilde{\theta}_1^{'\,\text{EW}}$ of the England and Wales data for the death scenario j^* (Method 4) and comparing the distribution of the projection, in particular the central projection, of j^* with the projection of 100 selected death scenarios. The corresponding credibility intervals and the distribution of central projections are plotted in Figure 6.49-6.52 for $\kappa_{t'}^{'\,(1)}$, $\kappa_{t'}^{'\,(2)}$, $\kappa_{t'}^{'\,(3)}$ and $\gamma_{c'}^{'\,(4)}$ respectively. Note that the term "central projection" represents the posterior mean of the projection, unless otherwise discussed. The upper and lower bounds at each year or year of birth are the 95% and 5% simulated samples of all the sample paths of the all the empirical posterior samples for the death scenario j^*. Note that $\boldsymbol{\theta}_1$ is projected forward by fifty years.

Recall that the uncertainty of the posterior projection in one single death scenario consists of the parameter uncertainty and the normal randomness of the N_3 simulated sample paths for the empirical posterior sample k of that particular death scenario. For the period effect $\kappa_{t'}^{(i)}$ where $i = 1, 2, 3$, in general we can see that the spread of the upper bound and the lower bound increases as we project further years due to the random walk model for all the three period effects of both the England and Wales and $w = 0.01$ and the latter has smaller projection variance at each projected year t for $i = 1, 2$ (Figure 6.49a and 6.50a), which is consistent with our findings (see Figure 6.3 and 6.12a), that the posterior estimation $\tilde{\boldsymbol{V}}_\epsilon^w(1,1)$ and $\tilde{\boldsymbol{V}}_\epsilon^w(2,2)$ of $w = 0.01$ is on average smaller than that of the $\tilde{\boldsymbol{V}}_\epsilon^{\text{EW}}(1,1)$ and $\tilde{\boldsymbol{V}}_\epsilon^{\text{EW}}(2,2)$. Similar for $i = 3$, the variance of $\tilde{\kappa}_{t',j^*}^{'\,(3),w}$ is approximately the same with that of the $\tilde{\kappa}_{t'}^{'\,(3),\text{EW}}$ (Figure 6.51a) since the posterior distribution $\tilde{\boldsymbol{V}}_\epsilon^w(3,3)$ is roughly the same

(a) Historical & projected $\boldsymbol{\kappa}^{(1)}$, Method 4, MCMC-Mean



(b) Historical & projected $\boldsymbol{\kappa}^{(1)}$, Method 4, MCMC-Mode



(c) MCMC Mean of $\boldsymbol{\kappa}^{(1)}$, Method 5

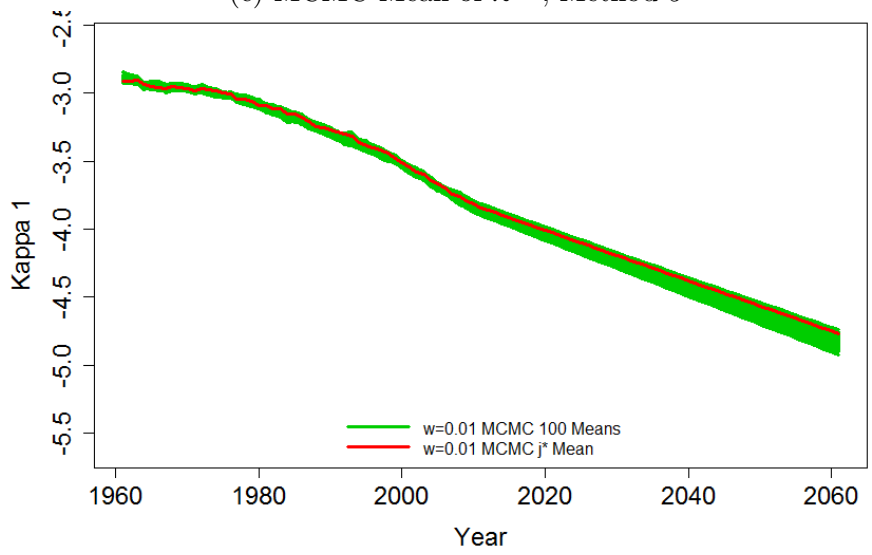
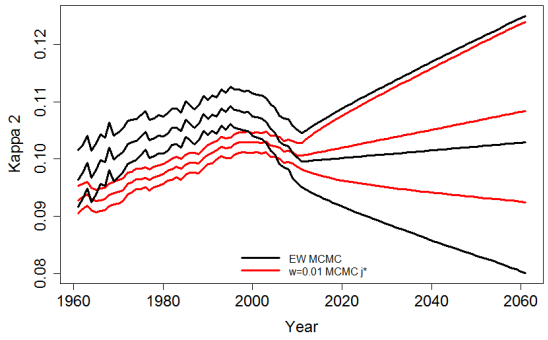


Figure 6.49: Upper: Credibility intervals of MCMC for the empirical and fifty-year forward projected $\boldsymbol{\kappa}^{(1)}$ for England and Wales (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (left) and MCMC-Mode (right). Lower: Central (posterior mean) MCMC estimations and projections for 100 selected death scenarios (green lines) and the death scenario j^*. Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples.

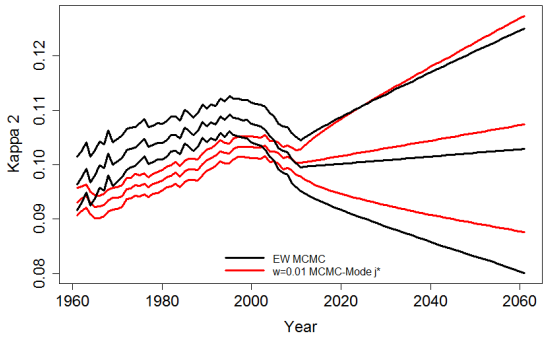
with the England and Wales data. Further, the central projection of $\tilde{\kappa}_{t',j^*}^{'\,(i),w}$ is higher than that of $\tilde{\kappa}_{t'}^{'\,(i),\mathrm{EW}}$ for $i = 1, 2$ since the corresponding posterior estimation $\tilde{\boldsymbol{\mu}}_i^{w,j^*}$ is on average higher than $\tilde{\boldsymbol{\mu}}_i^{\mathrm{EW}}$ (see Figure 6.7 and 6.16). Similarly, as the posterior estimation of $w = 0.01$ for $\boldsymbol{\mu}_3$ is lower than that of the England and Wales data, the small population $w = 0.01$ has a lower central projection of $\kappa_{t'}^{'\,(3)}$ compared with the England and Wales data.

The credibility interval of $\tilde{\kappa}_{t',j^*}^{'\,(i),w}$ given the MCMC-Mode is plotted in Figures 6.49b-6.51b for $i = 1, 2, 3$ respectively. By having a higher estimation of the volatility

(a) Historical & projected $\boldsymbol{\kappa}^{(2)}$, Method 4, MCMC-Mean



(b) Historical & projected $\boldsymbol{\kappa}^{(2)}$, Method 4, MCMC-Mode



(c) MCMC Mean of $\boldsymbol{\kappa}^{(2)}$, Method 5

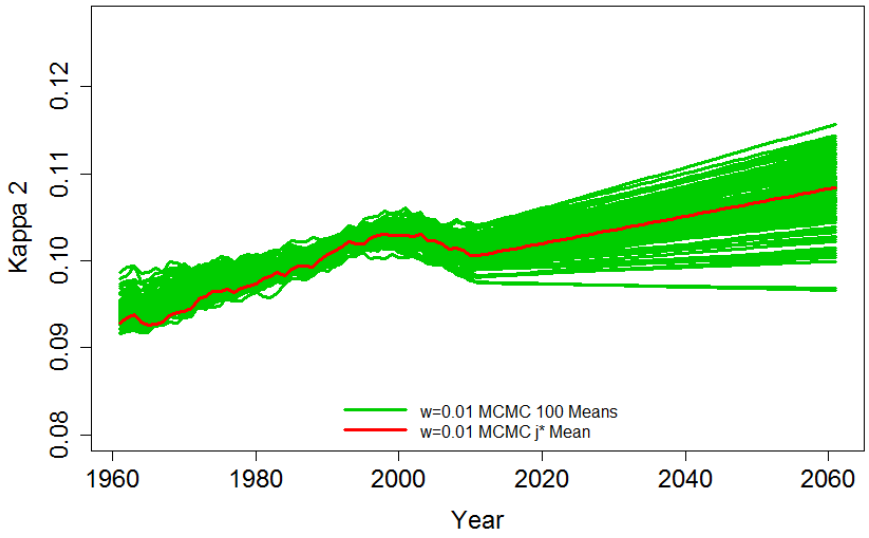


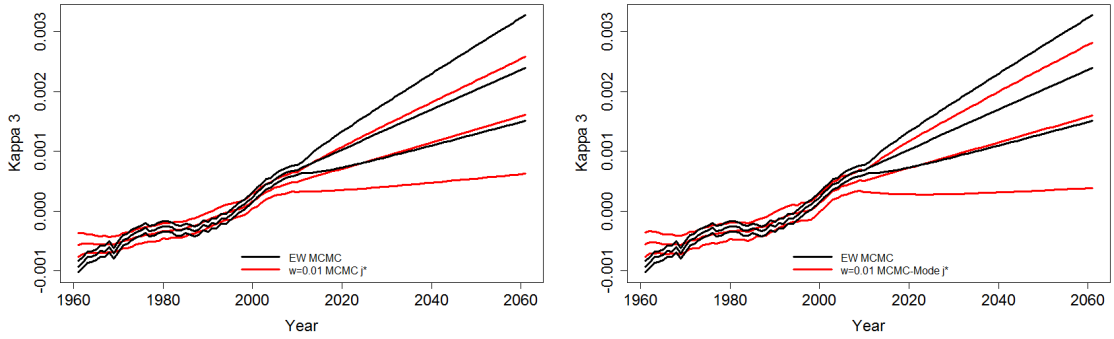
Figure 6.50: Upper: Credibility intervals of MCMC for the empirical and fifty-year forward projected $\boldsymbol{\kappa}^{(2)}$ for England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (left) and MCMC-Mode (right). Lower: Central MCMC estimations and projections for 100 selected death scenarios (green lines) and the death scenario j^*. Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples.

for the three period effects given the MCMC-Mode, the uncertainty of the projection at each year t' for all the three period effects are much higher compared with the uncertainty given the MCMC-Mean. In particular, by having a higher prior and posterior mean of $\tilde{\boldsymbol{V}}$, the spread of $\tilde{\kappa}_{t',j^*}^{'\,(1),w}$ for $w = 0.01$ is approximately the same with that of $\tilde{\kappa}_{t'}^{'\,(1),\mathrm{EW}}$ for the England and Wales roughly after year 2020 while the small population has a significantly higher projection uncertainty than the England and Wales does for $i = 3$ at each projecting year t'. These imply that the level of the estimation for the volatility of $\boldsymbol{\kappa}^{(i)}$ is a main deterministic factor of its projection

uncertainty.

(a) Historical & projected $\boldsymbol{\kappa}^{(3)}$, Method 4, MCMC-Mean

(b) Historical & projected $\boldsymbol{\kappa}^{(3)}$, Method 4, MCMC-Mode



(c) MCMC Mean of $\boldsymbol{\kappa}^{(3)}$, Method 5

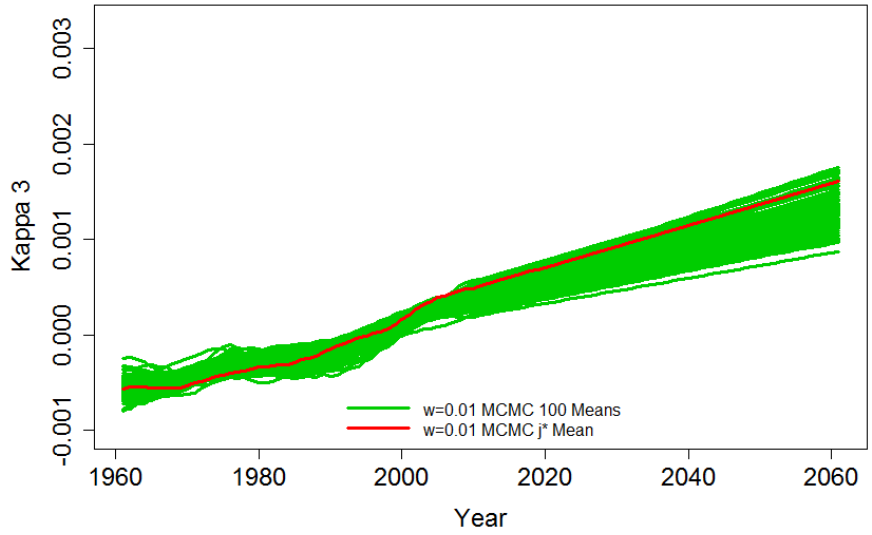


Figure 6.51: Upper: Credibility intervals of MCMC for the empirical and fifty-year forward projected $\boldsymbol{\kappa}^{(3)}$ for England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (left) and MCMC-Mode (right). Lower: Central MCMC estimations and projections for 100 selected death scenarios (green lines) and the death scenario j^*. Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples.

As for the cohort effect $\boldsymbol{\gamma}^{(4)}$, in general the spread of the projection for both the England and Wales and the small population $w = 0.01$ is relatively stable through the projected year of birth c' (Figure 6.52a) with both of the central projection converge to zero , which is consistent with the feature of the zero mean-revering AR(1) model. In particular, we can see that the spread of the projection $\tilde{\gamma}_{c',j^*}^{\prime\,(4),w}$ is much narrower than that of the $\tilde{\gamma}_{c'}^{\prime\,(4),\mathrm{EW}}$ for the England and Wales data. The central projection of the $w = 0.01$ reverts much faster that that of the England and

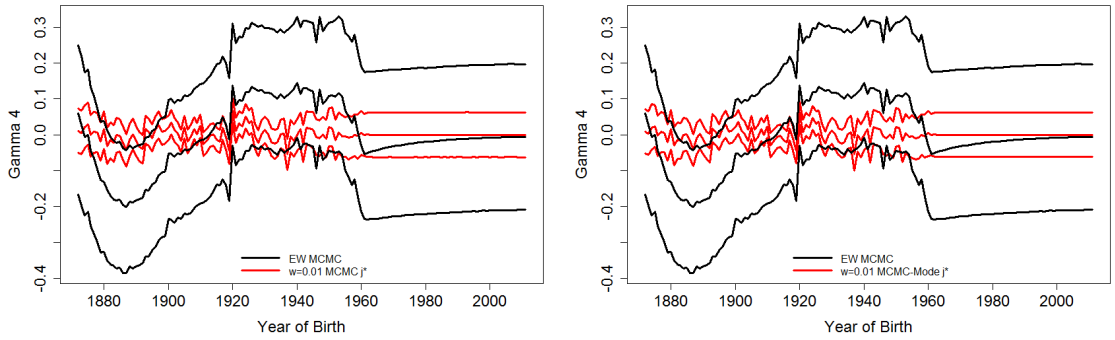
Wales data does. This is because the posterior estimation of the coefficient of the AR(1) model α_γ for $w = 0.01$ is significantly lower than the MCMC for England and Wales data, which makes the cohort process of $w = 0.01$ look more like an $AR(1)$ process. It is worth noticing that the posterior distribution of the volatility for the cohort process $\tilde{\sigma}_\gamma^{w,j^*}$ is slightly greater than the $\tilde{\sigma}_\gamma^{\text{EW}}$, which implies that the impact of having a higher estimation on the volatility on the uncertainty of projection for $\gamma_c^{(4)}$ is competed over by the significantly lower estimation for the coefficient of the AR(1) model. Figure 6.52b shows that changing the prior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ has barely any impact on the projection of the cohort effect as we expected.

We will then study the influence of sampling variation on the distribution of projection for the latent parameters by investigating the posterior predictive distribution of $\boldsymbol{\theta}_1$ of 100 selected death scenarios according to the Method 5. For all the three period effects and the cohort effect, we expect that the spread of the distribution of the projection varies from one death scenario to another at the same projected year t' or year of birth c'. This is because the sampling variation shifts the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, α_γ and σ_γ from one way to another and probably multiplies a constant to the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ and α_γ from one death scenario to another, according to the distribution of the corresponding posterior distributions of 100 death scenarios, see Figure 6.4, 6.14, 6.23, 6.35 and 6.32.

We are especially interested in the distribution of the central projection of the 100 selected death scenarios, driven by the sampling variation. For each selected death scenario $j' = j_1, \ldots, j_{100}$, we calculated the mean of $\tilde{\kappa}_{t',j'}^{'\,(i),w}$ and $\tilde{\gamma}_{c',j'}^{'\,(4),w}$ for $i = 1, 2, 3$. The 100 central projections and the corresponding central empirical estimations are plotted in Figure 6.49c-6.52c (green lines). The red lien is the central estimation & projection for the death scenario j^*. Note that the spread of the fans of the projection is only driven by the sampling variation. In practice term, we may have only one death sample with some sampling variation that causes our projection lower or higher than the MCMC projection of the benchmark population.

(a) Historical & projected $\boldsymbol{\gamma}^{(4)}$, Method 4, MCMC-Mean

(b) Historical & projected $\boldsymbol{\gamma}^{(4)}$, Method 4, MCMC-Mode



(c) MCMC Mean of $\boldsymbol{\gamma}^{(4)}$, Method 5

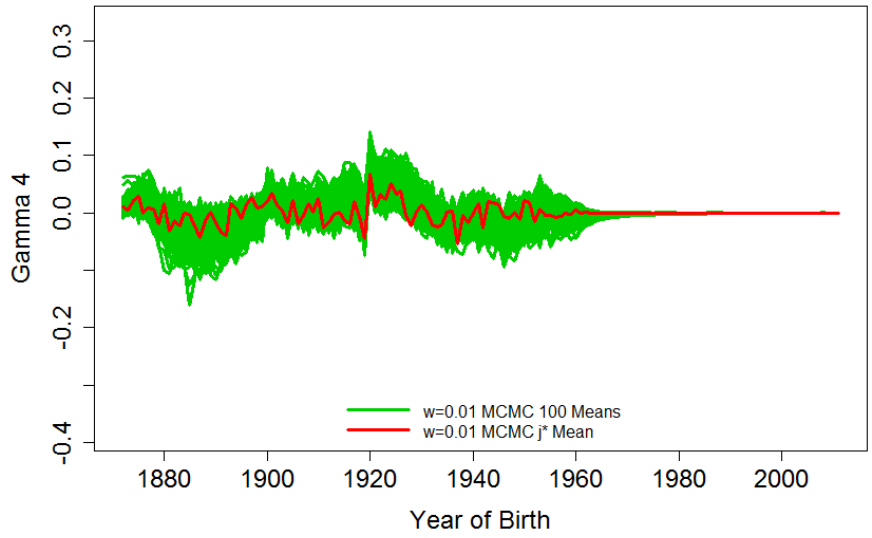


Figure 6.52: Upper: Credibility intervals of MCMC for the empirical and fifty-year forward projected $\boldsymbol{\gamma}^{(4)}$ for England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (left) and MCMC-Mode (right). Lower: Central MCMC estimations and projections for 100 selected death scenarios (green lines) and the death scenario j^*. Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples.

The randomness of the death sample could have meant that the projection could be shifted up and down and according to these plots the shift could be either big or small that is determined by how the sampling variation influencing the posterior distribution of the hyper-parameters, more specifically the drift and the coefficient of the random walk model and the $AR(1)$ model respectively. In particular, the spread of the central projection for the 100 death scenarios' period effects is determined by how the sampling variation affects/shifts the random walk drift $\boldsymbol{\mu}$ from one death scenario to another.

For $i = 1$, the shift to the posterior distribution of $\boldsymbol{\mu}_1$ due to the sampling variation is relatively small with no obvious impact on the variance of the distribution (Figure 6.8). Thus we can see that the spread of the green fan in Figure 6.49c for the distribution of the central projection of $\tilde{\kappa}_{t'}^{'(1)}$ given 100 death scenarios is relatively stable throughout the projecting year t' though slightly widened as t' increases and the width of the projection is not significantly greater compared with that of the empirical estimation as the varieties of both distributions are driven by the sampling variation. For $i = 2$ (Figure 6.50c), the spread of the central projection for $\tilde{\kappa}_{t'}^{'(2)}$ widens as the projecting year t' increase and the distribution of projection is wider than the empirical central estimation since the corresponding random walk drift $\boldsymbol{\mu}_2$ is shifted relatively more from on death scenario to another by the sampling variation. Similar interpretation applies to the distribution of central projection for $\tilde{\kappa}_{t'}^{'(3)}$ (Figure 6.51c).

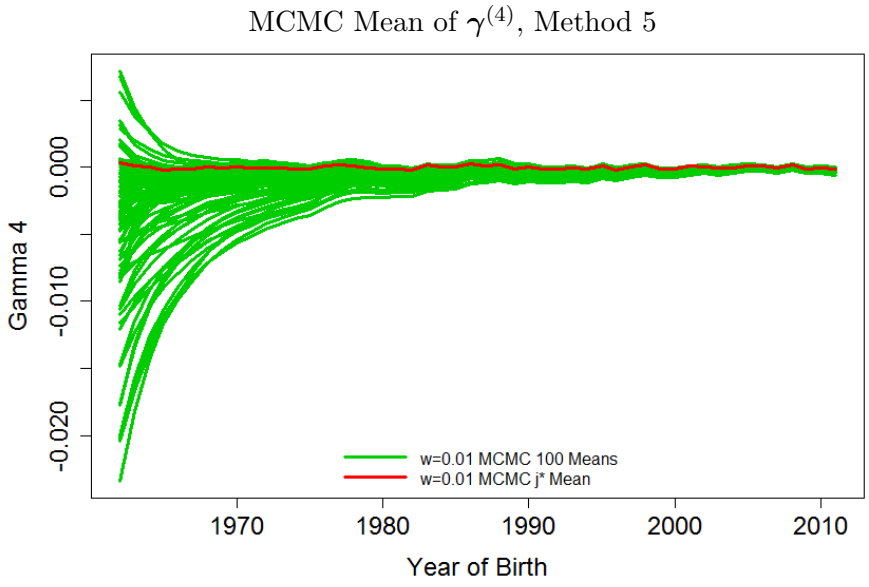


Figure 6.53: Central MCMC projections for 100 selected death scenarios (green lines) and the death scenario j^*, conditional on MCMC-Mean.

For $\tilde{\gamma}_{c'}^{'(4)}$, the distribution of the central projection is relatively much narrower than the empirical central estimation (Figure 6.52c). This is mainly due to the feature of the zero mean-reverting AR(1) model such that the means of the projection for each death sample converges to zero. The influence of the sampling variation

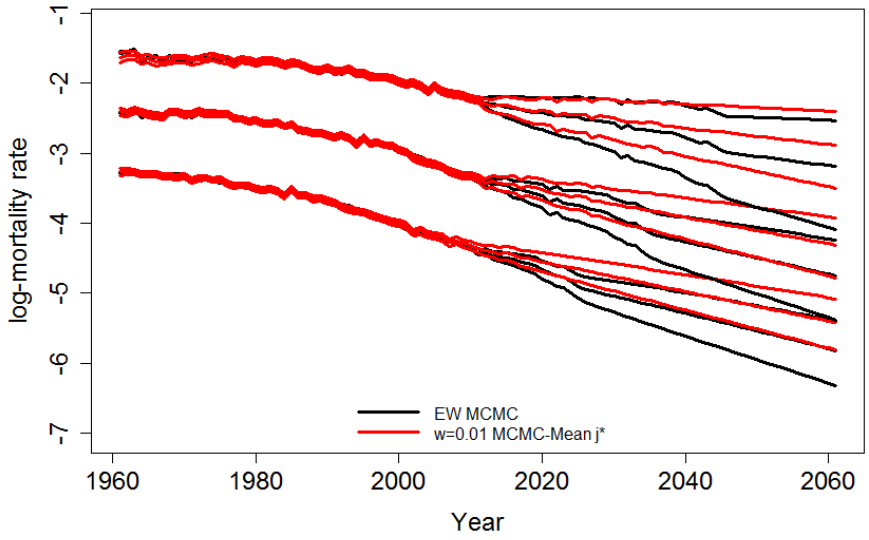
shifting the posterior distribution of α_γ for one death scenario to another is to affect how fast that the central projection converges to zero. This is consistent with Figure 6.53 where only the central projections are plotted and we can see that the mean of $\tilde{\gamma}^{'}_{c',j'}{}^{(4),w}$ converges to zero for each of the $j' = j_1, \ldots, j_{100}$ in varied paces. Note that we could observe slightly non-smooth (wave shaped) exponential convergence to zero. This is because only 100 sample paths are simulated for each death scenario.

Projecting the Death Rate

The projected death rates at projecting year t' aged x are then calculated conditional on he projected latent parameters. Denote as $m'(\tilde{\theta}^{'}_1{}^{w,j,(k,l)}, t', x)$ the l^{th} sample path of the projected rate of the k^{th} posterior sample of the death scenario j. We also denote as $m'(\tilde{\theta}^{'}_1{}^{w,j}, t', x)$ the distribution of the projected death rates for the death scenario j at year t' aged x. Once again our study will follow the idea of Method 4 and 5 by firstly comparing the $m'(\tilde{\theta}^{'}_1{}^{w,j^*}, t', x)$ of the death scenario j^* with the $m'(\tilde{\theta}^{'\text{EW}}_1, t', x)$ of the England and Wales data and then focusing on the distribution of the central projection according to the means of $m'(\tilde{\theta}^{'}_1{}^{w,j'}, t', x)$ of 100 selected death scenarios at a range of selected ages (55, 65, 75, 85 and 89). It is worth noticing that the prediction interval at age x and year t' differs from one death scenario to another since the sampling variation shifts the posterior distribution $\tilde{\boldsymbol{V}}_\epsilon$ from one way to another. Thus given a particular death sample, we could end up with a prediction interval that is either greater or smaller than the intervals of death scenario j^* and England and Wales data.

In Figure 6.54, we demonstrate the impact of age x on the level (upper) and the uncertainty (lower) of the log-scaled projected death rates. The red and black lines in Figure 6.54a are the credibility intervals of the $m'(\tilde{\theta}^{'}_1{}^{w,j^*}, t', x)$ and $m'(\tilde{\theta}^{'}_1{}^{\text{EW}}, t', x)$ respectively at age 55 (bottom), 75 (mid) and 85 (top). Note the upper and lower bounds of the intervals are the 95% and 5% quantiles of the samples of all the simulated sample paths of all the posterior samples for death scenario j^* and England

(a) $\log m(t,x)$ for age 65, 75, and 85, Method 4



(b) $\log m(t,x)$: 95% quantile - 5% quantile, at Year 2061, Method 4

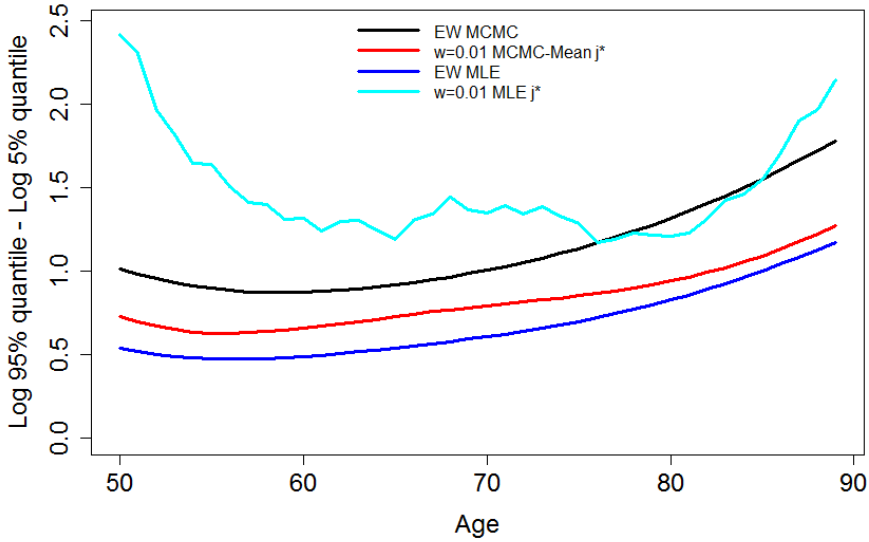


Figure 6.54: Upper: 90% credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 65, 75 and 85 for England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean. Lower: The difference between upper and lower bound at year 2061 for the age range of 50-89. Note that the upper and lower bounds are the log-scaled 5% and 95% quantiles respectively of the samples.

and Wales data respectively. We can see that the central projection for the log death rates is significantly increased for higher ages as expected.

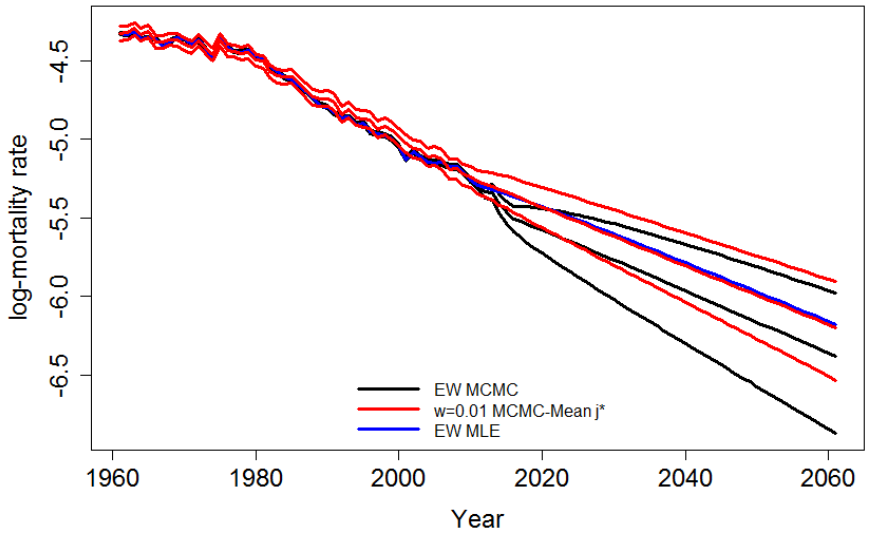
In Figure 6.54b, we plotted the log(95%-quantile) less log(5%-quantile) for year 2061 as a function of age x for the death scenario j^* MCMC-Mean(red line) and MLE (light blue), the England and Wales data MCMC (black line) and MLE (blue

line). The light blue curve is much less smoothed since we only simulated 100 sample paths for the j^* MLE and is much higher than the curve for EW MLE due to over-estimating the volatility of the period effects compared with the true rates. By providing a much improved estimation for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, we can observe that the gap between the prediction intervals of j^* MCMC and EW MLE , which is mainly driven by the parameter uncertainty, is much narrower.

The shapes of the two curves are similar for the two populations given MCMC while the England and Wales data in general has a wider prediction interval at each age x compared with the death scenario j^*, since the estimation of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ for j^* is smaller than that for the England and Wales data due to the non-informative prior distribution we choose for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ of the England and Wales data. More specifically, the uncertainty of the prediction for both populations only starts to increase approximately after age 60 while it is slightly decreasing for younger ages. It is also worth noticing that the increase of the uncertainty from younger to the elder ages is greater for the England and Wales data than that of for the death scenario j^*, which is similar with our finding in Figure 3.12 for the projection of the finite-sample MLE that the prediction interval for the smaller populations (w=0.01, 0.001) seems to be less affected by the age x. Further, we notice that the difference of the prediction intervals between the two populations in year 2061 is not stable with respect to the age x and as we can see the gap only increases after about age 70 while it is slightly decreasing for younger ages.

In order for a better visual inspection, we then plotted the prediction intervals for both populations aged 55, 65, 75, 85 and 89 separately in Figure 6.55-6.59 respectively as a function of year, given the MCMC-Mean (upper) and MCMC-Mode (lower). The blue lines are the mean of the fitted ($m(\hat{\theta}_1^{\text{EW}}, t, x)$) and forecast ($m'(\hat{\theta}_1'^{\text{EW}}, t', x)$) log-scaled death rates for the England and Wales data MLE. Unsurprisingly, the uncertainty about future mortality rates increases as the forecast horizon increases. In general, the prediction intervals of MCMC are much wider

(a) $\log m(t,x)$ for age 55, Method 4, MCMC-Mean



(b) $\log m(t,x)$ for age 55, Method 4, MCMC-Mode

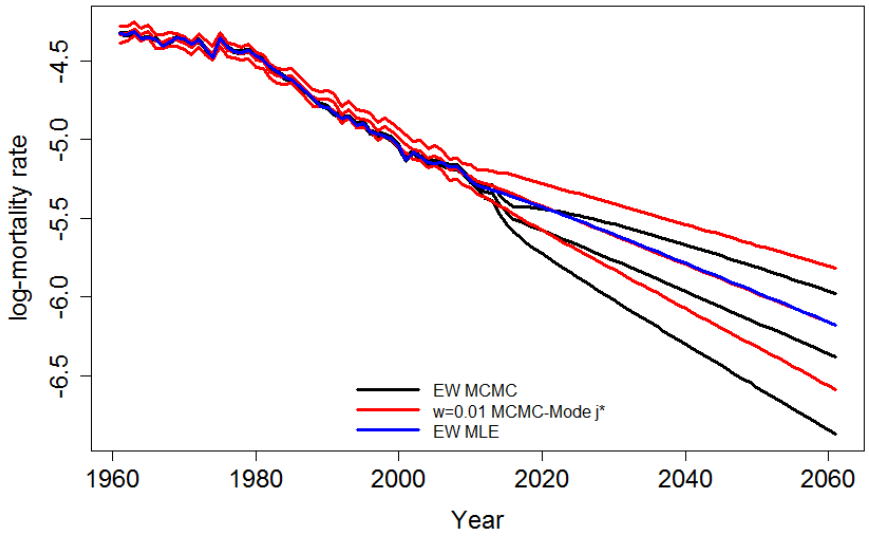
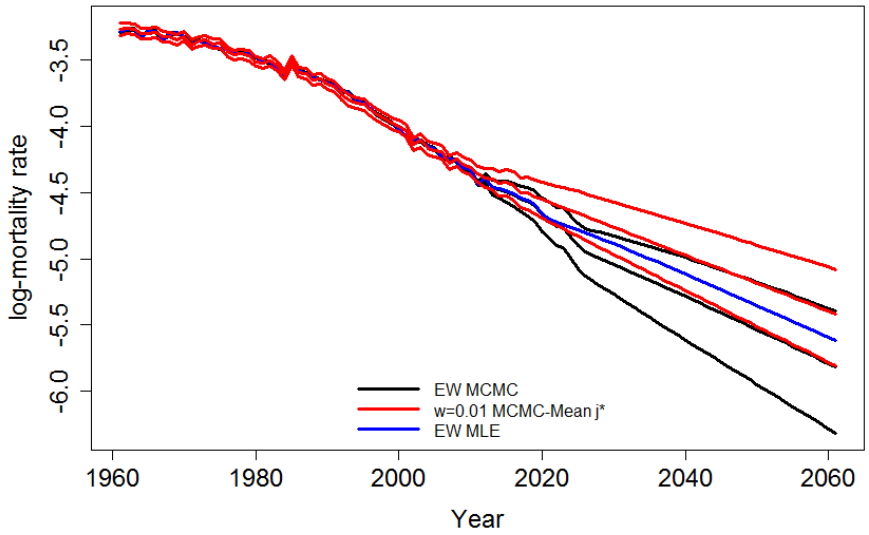


Figure 6.55: Credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 55 for the England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (upper) and MCMC-Mode (lower). Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples. The central black and red lines are the log-scaled mean of the death rates for the relative death data. The blue line is the central fitted and projected death rates for the England and Wales data MLE.

than the empirical fitted rates at all the ages for each projected year t' since the uncertainty of the former is driven by both the parameter uncertainty and the the normal randomness for simulating the predicted sample paths while the uncertainty of the fitting intervals only consists the parameter uncertainty. The zigzag of the prediction intervals for both populations during the early forecast horizon is driven

by the cohort effect (e.g. Figure 6.56a-6.59a). Note that for age 55, such fluctuation could be observed on the fitted period, though not observable during the forecasts since the corresponding cohort effect only includes the projected years of birth such that the zigzag is relatively too small to be inspected.

(a) $\log m(t,x)$ for age 65, Method 4, MCMC-Mean



(b) $\log m(t,x)$ for age 65, Method 4, MCMC-Mode

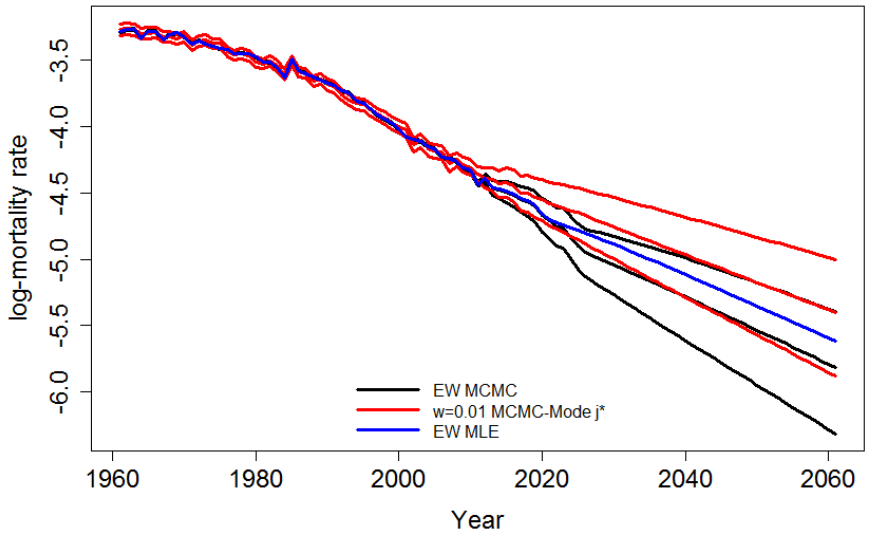


Figure 6.56: Credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 65 for the England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (upper) and MCMC-Mode (lower). Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples. The central black and red lines are the log-scaled mean of the death rates for the relative death data. The blue line is the central fitted and projected death rates for the England and Wales data MLE.

Recall that the death scenario j^* has a higher estimation for both $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$

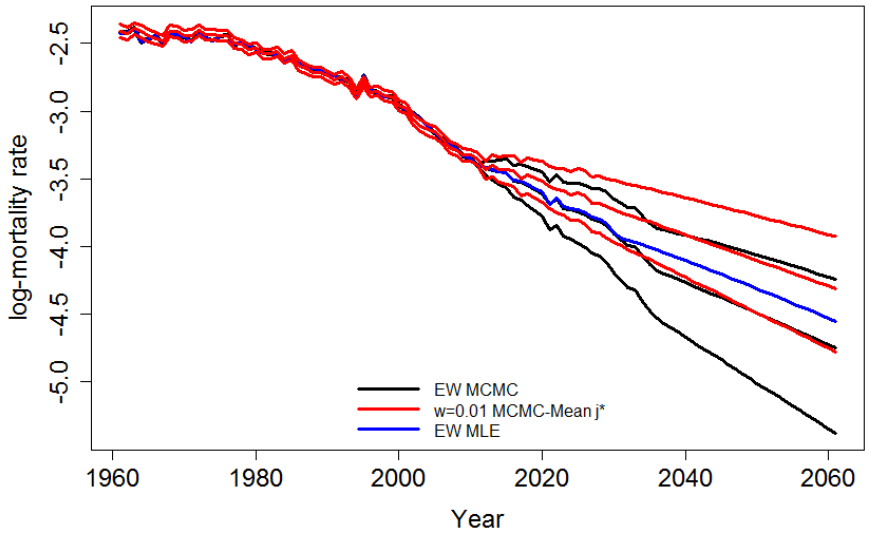
while the estimation for $\boldsymbol{\mu}_3$ is smaller compared with the England and Wales data. The central prediction at year t' can be written as a linear function of $E[\tilde{\boldsymbol{\mu}}_1]$, $E[\tilde{\boldsymbol{\mu}}_2]$ and $E[\tilde{\boldsymbol{\mu}}_3]$. More specifically,

$$\begin{aligned} E[\text{logit } q'(t',x)] &= E[\text{logit } q(t_{n_y},x)] + E[\tilde{\boldsymbol{\mu}}_1](t'-t_{n_y}) + E[\tilde{\boldsymbol{\mu}}_2](x-\bar{x})(t'-t_{n_y}) \\ &\quad + E[\tilde{\boldsymbol{\mu}}_3]((x-\bar{x})^2-\hat{\sigma}_x^2)(t'-t_{n_y}) \\ &\quad + E[\tilde{\gamma}^{(4)}_{t'-x}] - E[\tilde{\gamma}^{(4)}_{t_{n_y}-x}] \end{aligned} \tag{6.2}$$

which is also a quadratic function of age x at year t'. Thus given the value of $E[\tilde{\boldsymbol{\mu}}_1]$ is significantly higher than the $E[\tilde{\boldsymbol{\mu}}_2]$ and $E[\tilde{\boldsymbol{\mu}}_3]$ for both populations (Table 6.4), the level of the central prediction at year t' and age x is mainly affected by the level of $E[\tilde{\boldsymbol{\mu}}_1]$. Based on Figure 6.52a, the influence of the cohort effect is smaller when the projected cohort years are applied for the forecasts due to the zero mean-reverting effect. Unsurprisingly, we find that the central prediction for the England and Wales data MCMC is in general no greater than that for the death scenario j^* as expected through the predict horizon for each age selected. For example, the central projection at age 65 (Figure 6.56a) of the death scenario j^* is greater than that of the England and Wales data MCMC and the gap increases along the forecasting horizon since: the death rates is mainly affected by the level of $E[\tilde{\boldsymbol{\mu}}_1]$ at age 65; the influence of $E[\tilde{\boldsymbol{\mu}}_3]$ could be negligible due to the quadratic age term. The corresponding cohort year range for the forecasts is 1947-1996, which explains the trajectory drops around the forecast year 2025-2026 for the England and Wales data. The forecasts are then barely affected by the cohort effect when the projected cohort effect reverts to zero for both populations.

On the other hand, however, Equation (6.2) also implies that the influence of $E[\tilde{\boldsymbol{\mu}}_2]$ and $E[\tilde{\boldsymbol{\mu}}_3]$ increases as the age approaches the youngest and eldest ages, which is particularly true for $\boldsymbol{\mu}_3$ due to the quadratic term $((x-\bar{x})^2-\hat{\sigma}_x^2)$. Moreover, more estimated empirical cohort years are used for the forecasts when the age is

(a) $\log m(t,x)$ for age 75, Method 4, MCMC-Mean



(b) $\log m(t,x)$ for age 75, Method 4, MCMC-Mode

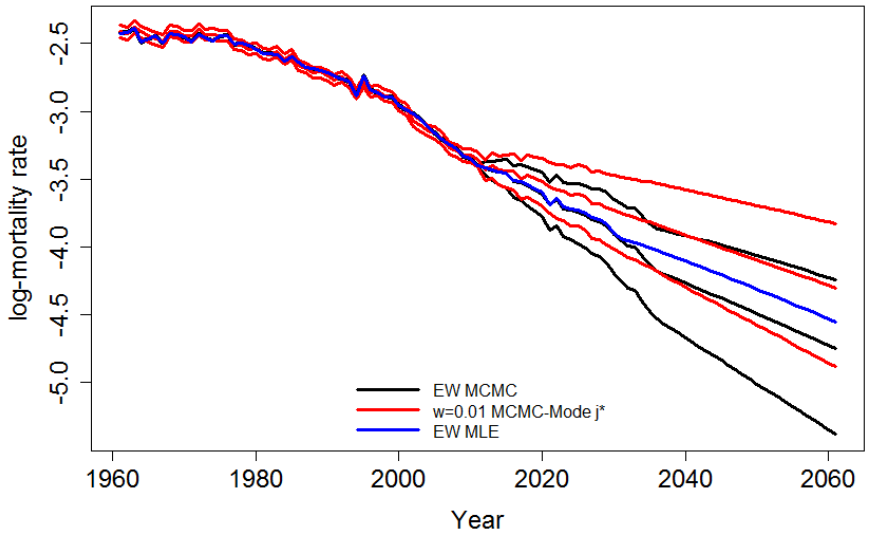


Figure 6.57: Credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 75 for the England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (upper) and MCMC-Mode (lower). Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples. The central black and red lines are the log-scaled mean of the death rates for the relative death data. The blue line is the central fitted and projected death rates for the England and Wales data MLE.

higher. For example, Figure 6.59a shows that at age 89 the central prediction of death scenario j^* is approximately the same with that of the England and Wales data during the first several years of the forecasting horizon and then a drop to the England and Wales data MCMC central projection. This is driven by two reasons. Firstly, the lower mean of $\boldsymbol{\mu}_3$ of the death scenario j^* compared with the England

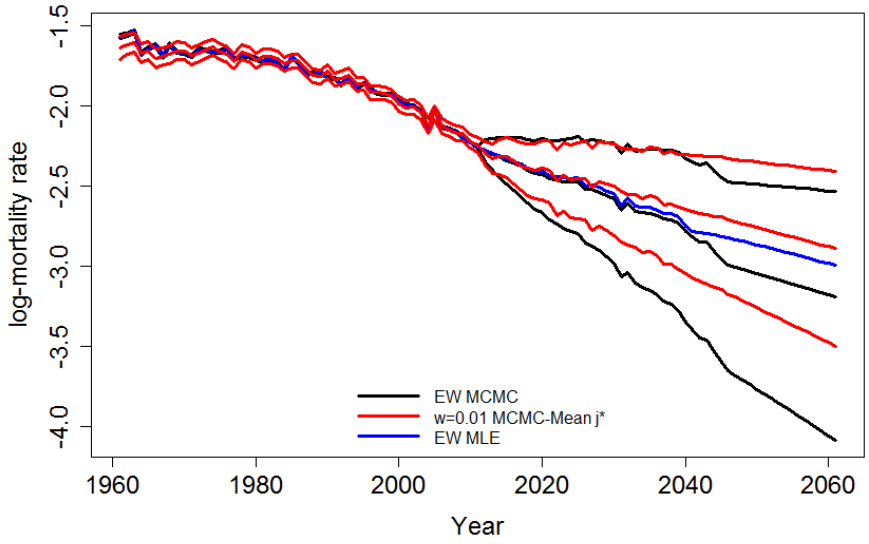
and Wales data MCMC pushes its central red line closer to the black when the impact of $\boldsymbol{\mu}_3$ is the most significant at the "eldest" age and competing with $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Secondly, the term $E[\tilde{\gamma}^{(4)}_{t'-x}] - E[\tilde{\gamma}^{(4)}_{t_{ny}-x}]$ in Equation (6.2) of the death scenario j^* is approximately equivalent to zero through the forecast horizon since the respective estimated cohort years fluctuate around zero with small volatility while the quantity of the England and Wales data MCMC is greater than zero along with $t'-x$ (Figure 6.52a), which gradually pushes up the central black line along with t'. However, the influence of the cohort effect becomes smaller when elder projected cohort years are used for forecasting due to the zero mean-reverting effect, in other words relatively much smaller values of $E[\tilde{\gamma}^{(4)}_{t'-x}] - E[\tilde{\gamma}^{(4)}_{t_{ny}-x}]$. Thus in total, the effect of increasing the central prediction of the death scenario j^* due to the higher $E[\tilde{\boldsymbol{\mu}}_1^{w,j^*}]$ and $E[\tilde{\boldsymbol{\mu}}_2^{w,j^*}]$ is offset by a smaller $E[\tilde{\boldsymbol{\mu}}_3^{w,j^*}]$ and relatively greater $E[\tilde{\gamma}^{(4),\text{EW}}_{t'-x}] - E[\tilde{\gamma}^{(4),\text{EW}}_{t_{ny}-x}]$. The relationship between the patterns of the central predictions for the other ages of the two populations can also be interpreted by the influence of $\boldsymbol{\mu}$ according to the level of the $|x-\bar{x}|$ and the cohort years employed for the forecasting.

Recall that the hyper-parameter estimates (of the period effects)[2] of the projecting model for $m'(\hat{\theta}_1'^{\,\text{EW}}, t', x)$ are the true rates of the death scenario j^*. It is worth noticing that the central forecast of $m'(\tilde{\theta}_1'^{\,w,j^*}, t', x)$ is only slightly higher than that of $m'(\hat{\theta}^{\text{EW}}, t', x)$ and the difference is relatively much smaller compared with the one between the MCMC of j^* and England and Wales data. In particular, the central projection of the death scenario j^* at age 55 is almost identical to that of the England and Wales data MLE (blue line) in Figure 6.55a. Similar results could also be observed in Figure 6.59a at age 89. For the central projection at all the selected ages, we could see that the difference of the central projection stays relatively unchanged through the forecasting horizon after the cohort-caused zigzags for a particular age x.

We have discussed that the prediction interval of death scenario j^* is narrower

[2]Note that the hyper-parameter MLEs for the cohort effect is calculated without the short cohorts estimates

(a) $\log m(t,x)$ for age 85, Method 4, MCMC-Mean



(b) $\log m(t,x)$ for age 85, Method 4, MCMC-Mode

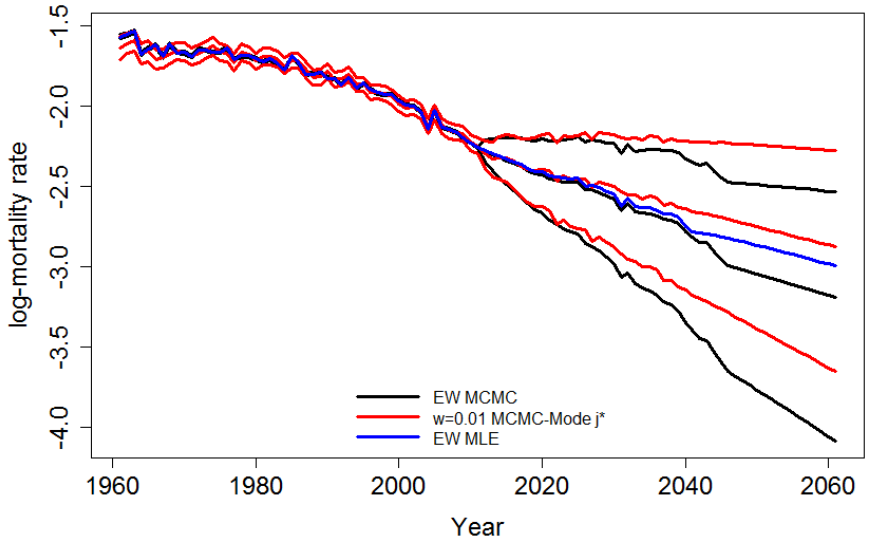
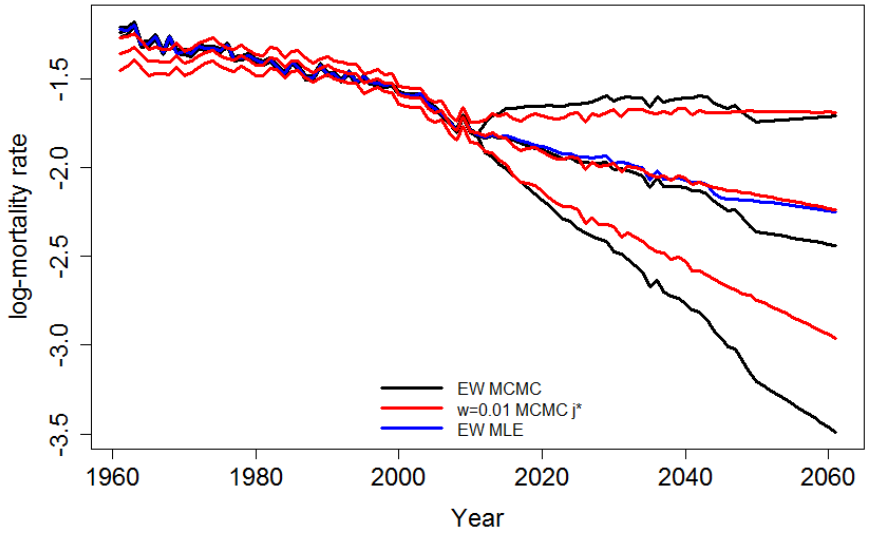


Figure 6.58: Credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 85 for the England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (upper) and MCMC-Mode (lower). Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples. The central black and red lines are the log-scaled mean of the death rates for the relative death data. The blue line is the central fitted and projected death rates for the England and Wales data MLE.

than that of the England and Wales data at age x year t' due to the choice of the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ on the estimation $\tilde{\boldsymbol{V}}_{\epsilon}$, more specifically by fixing the mean of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$. Recall that fixing the mode of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ provides approximately unbiased estimates for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ by increasing the prior mean and variance for the co-variance matrix. Figure 6.55b-

6.59b shows that the prediction interval of death scenario j^* is approximately as wide as that of the England and Wales data along with the forecast horizon for an age level x.

(a) $\log m(t,x)$ for age 89, Method 4, MCMC-Mean



(b) $\log m(t,x)$ for age 89, Method 4, MCMC-Mode

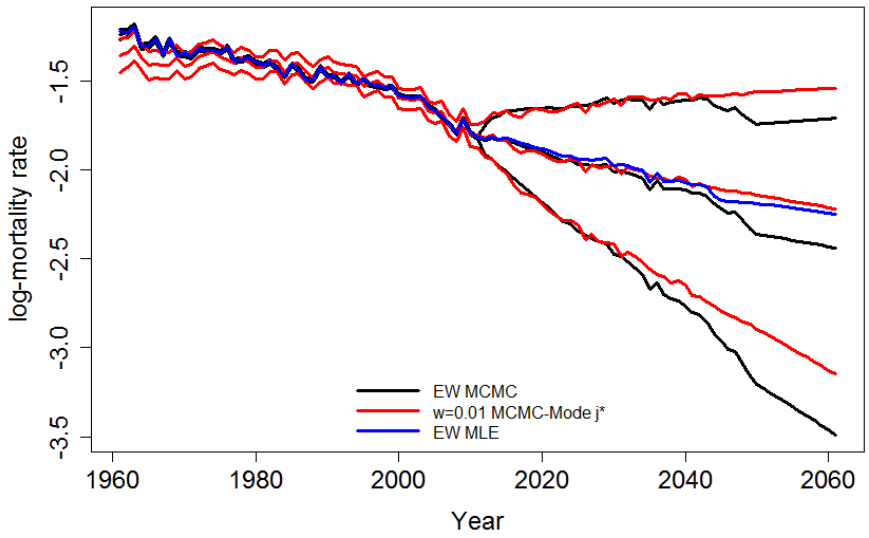


Figure 6.59: Credibility intervals of MCMC for the fitted and fifty-year forward projected death rates at age 89 for the England and Wales data (black lines) and the death scenario j^* for $w = 0.01$ (red lines), given MCMC-Mean (upper) and MCMC-Mode (lower). Note that the upper and lower bounds are the 5% and 95% quantiles respectively of the samples. The central black and red lines are the log-scaled mean of the death rates for the relative death data. The blue line is the central fitted and projected death rates for the England and Wales data MLE.

We plotted the central predictions of 100 death scenarios in Figure 6.60 (green lines) for age 55, 65, 75, 85 and 89. The red and black lines are the central predictions

of MCMC for death scenario j^* and the England and Wales data respectively. The blue lines are for the central fitted and forecasts for England and Wales data MLE. In general, the spread of the 100 central predictions is not significantly wider than that of the central fitted rates (in particular at age 55, 65, 75 and 85), although the former increase along with the forecast horizon, since the varieties of distributions for the central fitted rate and the prediction are both driven by the sampling variation only. Together with Figure 6.61 and 6.62 that include the distributions of the upper and lower bounds respectively of the prediction intervals of 100 death scenarios, we can conclude that the sampling variation shifts the prediction intervals up and down at all the ages for year t'. It implies that given a death sample, the sampling variation could shift the prediction intervals vertically for a particular age level and such shift could be either big or small. For example, the prediction intervals (e.g. the central prediction) of the death scenario j^* at age 85 is shifted up by the sampling variation and lie above the median of the distribution.

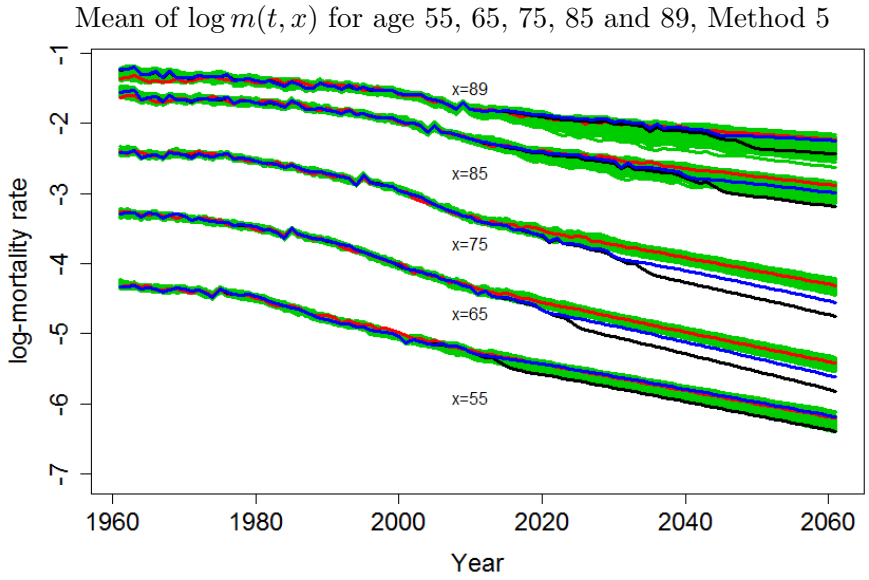


Figure 6.60: Central (log-scaled posterior mean) MCMC fitted and projected rates for 100 selected death scenarios (green lines), the death scenario j^* (red line) and the England and Wales data (black line) at age 55, 65, 75, 85 and 89. The blue line is the central fitted and projected death rates for the England and Wales data MLE.

Further in Figure 6.60, we find that the distribution of central prediction driven by the sampling variation is higher than the central prediction of the England and

Wales data MCMC at age 65 and 75 while there is no significant difference between the two at the other ages (same with the lower and upper bounds). However, we can also see that the central projection of England and Wales data MLE (blue line) is approximately within the distribution of the central prediction for each age level (slightly lower for England and Wales data MLE at age 75), which implies that MCMC generates approximately the same central forecast for the small population $w = 0.01$ with the central prediction generated by the true rates of the small population. Figure 6.61-6.62 shows the distributions of the upper and lower bounds of the forecasts for the selected age levels. Once again the sampling variation shifts the boundaries of the prediction intervals up and down and such shift varies from one death scenario to another. The upper and lower boundaries (blue lines) for the EW-MLE are within the spread of the green lines for the small populations which implies non-significant differed volatilities of forecasts between the EW-MLE and the MCMC of small population at all the ages.

We can thus conclude that forecast distribution given the MCMC method for one simulated death scenario is approximately the same with the distribution generated by the true rates of population with the prediction intervals of the former distribution slightly wider through the forecast horizon due to the parameter uncertainty. In the real circumstance, it is unlikely that we are given the true rates of the underlying small population and our findings mainly imply that the sampling variation shifts the predictive distribution up and down and the information of larger population can be adopted to improve the mortality projection if there are evidences that the two populations share some common characteristics.

We can also see that the impact of sampling variation shifting the prediction intervals varies with regard to the age level. For example, Figure 6.61 shows that there is relatively more variety to the distribution of the central prediction at very young and old ages as expected due to the term $x - \bar{x}$ and $((x - \bar{x})^2 - \hat{\sigma}_x^2)$, implying that the variety of the central mortality improvement rate driven by the sampling

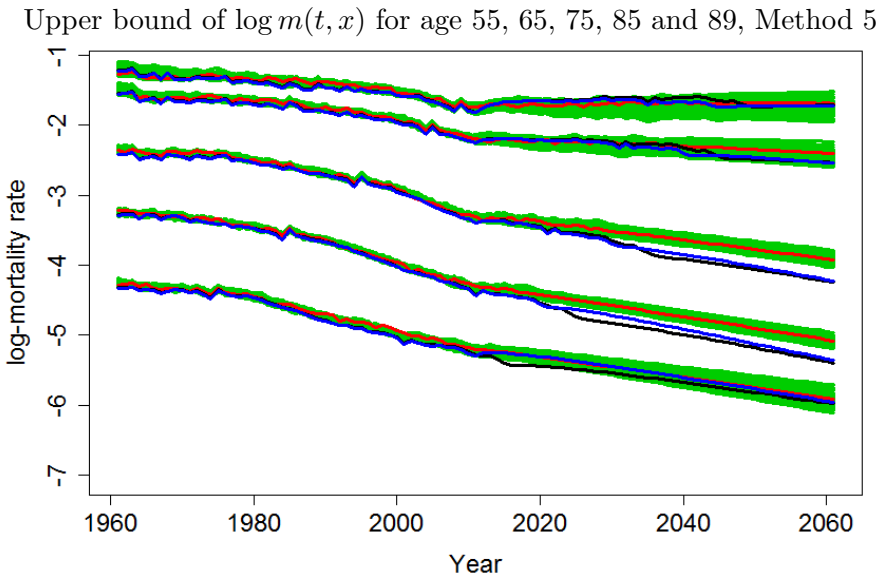


Figure 6.61: Upper bound (95% quantile) of the MCMC fitted and projected rates for 100 selected death scenarios (green lines), the death scenario j^* (red line) and the England and Wales data (black line) at age 55, 65, 75, 85 and 89. The blue lines are for the England and Wales data MLE.

variation is more like a quadratic function of age level x.

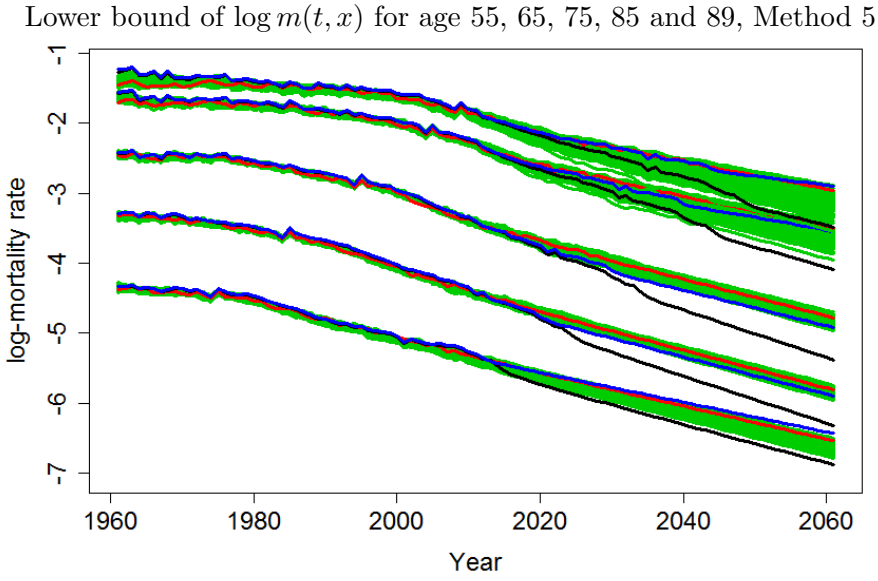


Figure 6.62: Lower bound (5% quantile) of the MCMC fitted and projected rates for 100 selected death scenarios (green lines), the death scenario j^* (red line) and the England and Wales data (black line) at age 55, 65, 75, 85 and 89. The blue lines are for the England and Wales data MLE.

Define as IR the central logit-mortality improvement rate conditional on the death $\boldsymbol{D}$. Let $IR_j(\tilde{\theta}_1^{'\,w,j})$ be the MCMC central logit-mortality improvement rate

for the death scenario j conditional on the death $\boldsymbol{D}_j^w$ and can be written as:

$$E\Big[\frac{\text{logit } q'(\tilde{\theta}_1^{'\,w,j}, t_{n_y}+50, x) - \text{logit } q(\tilde{\theta}_1^{w,j}, t_{n_y}, x)}{50}\Big|\boldsymbol{D}_j^w\Big] \;=\; IR_j(\tilde{\theta}_1^{'\,w,j}), \quad (6.3)$$

which is nearly equal to

$$\begin{aligned} IR_j(\tilde{\theta}_1^{'\,w,j}) \;&\doteqdot\; E[\tilde{\boldsymbol{\mu}}_1^{w,j}|\boldsymbol{D}_j^w] + E[\tilde{\boldsymbol{\mu}}_2^{w,j}|\boldsymbol{D}_j^w](x-\bar{x}) \\ &\quad + E[\tilde{\boldsymbol{\mu}}_3^{w,j}|\boldsymbol{D}_j^w]((x-\bar{x})^2 - \hat{\sigma}_x^2). \end{aligned} \quad (6.4)$$

Note that $E[\tilde{\boldsymbol{\mu}}_i^{w,j}|\boldsymbol{D}_j^w]$ for $i = 1, 2, 3$ and $E[\text{logit } q(\tilde{\theta}_1^{w,j}, t_{n_y}, x)|\boldsymbol{D}_j^w]$ are the MCMC mean of the corresponding quantities given N_2 drawn posterior samples of death scenario j. $E[\text{logit } q(\tilde{\theta}_1^{'\,w,j}, t_{n_y}+50, x)]$ is the mean of all the simulated forecasting sample paths of death scenario j. The formulas for the central improvement of $w = 0.01$-MLE, EW-MCMC and EW-MLE can be derived similarly. The mean and the variance of IR are plotted as a function of age x in Figure 6.63-6.64. Note that the improvement rate (i.e. the slope of the central projection) is negative and a grater scaled rate means a steeper slope of the central prediction and hence a larger mortality improvement.

The trajectories of the curves indicate that the mean of the IR is a quadratic function of the age x and the age $x_{\min}$ at which the mean achieves the lowest value differs with regard to the population size and modelling methodology. For example, the mean of $IR(\tilde{\theta}_1^{'\,w})$ given N_1 death scenarios is

$$E[IR(\tilde{\theta}_1^{'\,w})] \doteqdot E[\bar{\tilde{\boldsymbol{\mu}}}_1^w|\boldsymbol{D}^w] + E[\bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w](x-\bar{x}) + E[\bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w]((x-\bar{x})^2 - \hat{\sigma}_x^2),$$

which achieves the minimum at age $x_{\min}$

$$x_{\min} = \bar{x} - \frac{E[\bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w]}{2E[\bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w]}.$$

The mean for the $w = 0.01$-MLE, EW-MCMC and EW-MLE and the age for the minimum mean can be derived similarly.

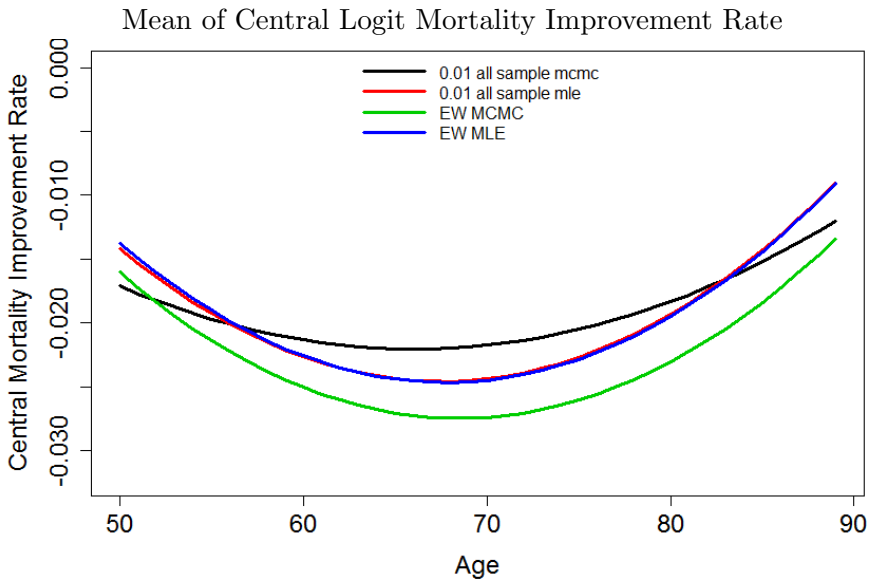


Figure 6.63: The mean of the logit-scaled mortality improvement with respect to age x, given MCMC and MLE of England and Wales data (green and blue lines) and 1000 death scenarios (black and red lines).

Unsurprisingly, the $E[IR(\tilde{\theta}_1'^{\,\text{EW}})]$ of the EW-MCMC (green line) is greater than the $E[IR(\hat{\theta}_1'^{\,\text{EW}})]$ of the EW-MLE while the latter is slightly greater than the $E[IR(\tilde{\theta}_1'^{\,w})]$ at most ages. More specifically, we can see that $E[IR(\tilde{\theta}_1'^{\,w})]$ (black curve) and $E[IR(\hat{\theta}_1'^{\,\text{EW}})]$ (blue curve) cross each other at very young and old ages, which is consistent with Figure 6.60, where the central prediction of EW-MLE lies on the top of the green fan at age 55 and 89. The $E[IR(\tilde{\theta}_1'^{\,w})]$ (black curve) is in general lower than the $E[IR(\tilde{\theta}_1'^{\,\text{EW}})]$ (green curve), except for the earliest ages. More specifically, the gap between the two is more like a quadratic function of the age level, implying that the central prediction of $w = 0.01$-MCMC is on average steeper compared with the EW-MCMC at early ages (e.g. age 50) while relatively flatter at the other ages and the size of the gap between the central predictions of the two populations should follow a quadratic function of the age level, which is consistent with Figure 6.60.

In Figure 6.64, the approximate U-shaped curve indicates that the standard deviation of the central mortality improvement $SD[IR(\tilde{\theta}_1'^{\,w})]$ for $w = 0.01$-MCMC

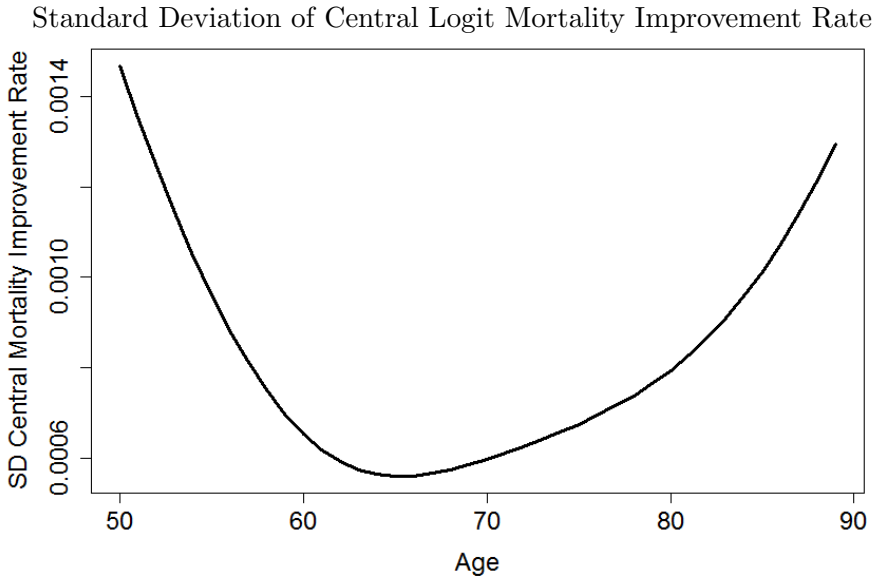


Figure 6.64: The standard deviation of 1000 death scenarios' central logit-scaled mortality improvement rate with respect to age x.

has some features of a quadratic function of age x. More specifically, the shape of the curve can be interpreted by the formula for $Var[IR(\tilde{\theta}_1'^{\,w})]$ with respect to age x and age $x_{\min}$ at which the lowest variance is achieved can be derived as follows:

$$\begin{aligned}
\mathrm{Var}[IR(\tilde{\theta}_1'^{\,w})] \;\doteqdot\; & \mathrm{Var}[\bar{\tilde{\boldsymbol{\mu}}}_1^w|\boldsymbol{D}^w] + \hat{\sigma}_x^2\mathrm{Var}[\bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w] \\
& + ((x-\bar{x})^2 - \hat{\sigma}_x^2)\mathrm{Var}[\bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w] \\
& + ((x-\bar{x})^2 - \hat{\sigma}_x^2)\mathrm{Cov}[\bar{\tilde{\boldsymbol{\mu}}}_1^w|\boldsymbol{D}^w, \bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w] \\
& + ((x-\bar{x})^2 - \hat{\sigma}_x^2)^2\mathrm{Var}[\bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w] \\
& + (x-\bar{x})\mathrm{Cov}[\bar{\tilde{\boldsymbol{\mu}}}_1^w|\boldsymbol{D}^w, \bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w] \\
& + (x-\bar{x})((x-\bar{x})^2 - \hat{\sigma}_x^2)\mathrm{Cov}[\bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w, \bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w], \qquad (6.5)
\end{aligned}$$

which is a quartic function of age x and more like a quadratic function of $((x-\bar{x})^2 - \hat{\sigma}_x^2)$ when the impact of the last two rows are relatively much smaller.[3] Thus

[3] $\mathrm{Cov}[E[\tilde{\boldsymbol{\mu}}_1^j], E[\tilde{\boldsymbol{\mu}}_2^j]] = 2.61 \times 10^{-9}$, $\mathrm{Cov}[E[\tilde{\boldsymbol{\mu}}_1^j], E[\tilde{\boldsymbol{\mu}}_3^j]] = 1.26 \times 10^{-11}$, $\mathrm{Cov}[E[\tilde{\boldsymbol{\mu}}_2^j], E[\tilde{\boldsymbol{\mu}}_3^j]] = -3.54 \times 10^{-11}$

the variance of the $Var[IR(\tilde{\theta}_1'^{\,w})]$ is the minimum when

$$(x_{\min} - \bar{x})^2 - \hat{\sigma}_x^2 \cong -\frac{\text{Var}[\bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w] + \text{Cov}[\bar{\tilde{\boldsymbol{\mu}}}_1^w|\boldsymbol{D}^w, \bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w]}{2\text{Var}[\bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w]}.$$

The above equation implies that $(x_{\min} - \bar{x})^2 - \hat{\sigma}_x^2 < 0$. Given $\text{Cov}[\bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w, \bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w] < 0$, the last two terms of Equation 6.2.5 have to be negative for such that the $Var[IR(\tilde{\theta}_1'^{\,w})]$ is the smallest. Thus we have $x_{\min} - \bar{x} < 0$ and

$$x_{\min} = \bar{x} - \sqrt{\hat{\sigma}_x^2 - \frac{\text{Var}[\bar{\tilde{\boldsymbol{\mu}}}_2^w|\boldsymbol{D}^w] + \text{Cov}[\bar{\tilde{\boldsymbol{\mu}}}_1^w|\boldsymbol{D}^w, \bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w]}{2\text{Var}[\bar{\tilde{\boldsymbol{\mu}}}_3^w|\boldsymbol{D}^w]}} \tag{6.6}$$

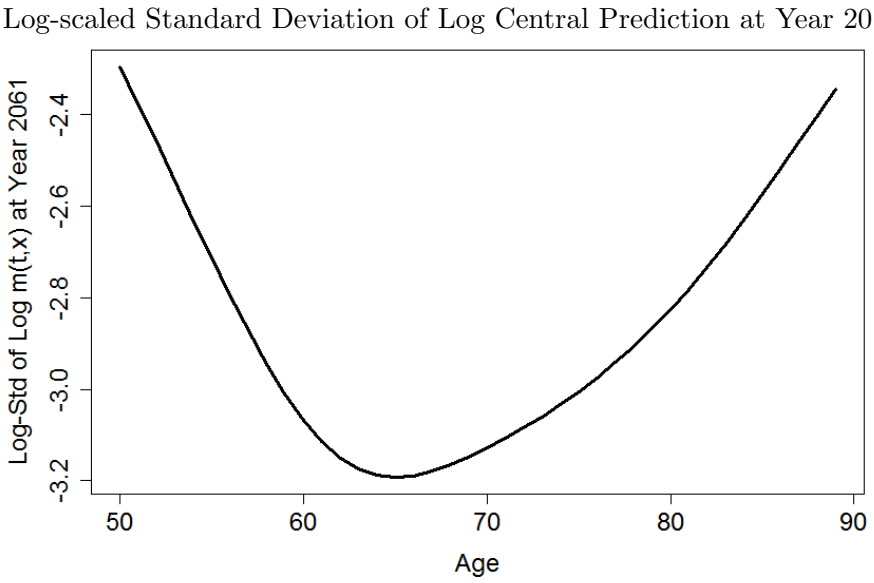


Figure 6.65: The log-scaled standard deviation of the distribution of log central prediction according to 1000 death scenarios as a function of age x at year 2061.

We calculated the log-scaled standard deviation of the distribution of log central prediction according to the 1000 death scenarios at year 2061 for age 50-89 and plotted in Figure 6.65 as a function of age x. We can see that the uncertainty of the central prediction driven by the sampling variation is also approximately a quadratic function of age x centred around about age 65, where the standard deviation is lowest. The prove of Figure 6.65 is similar with Figure 6.64

Impact on Annuity

In the previous section, we projected the MCMC mortality rates for the simulated death scenarios of population size $w = 0.01$ and found that the sampling variation shifted the distribution of the prediction up and down from one death scenario to another. Comparisons were carried out between the projected rates of the small population and England and Wales. In this section, we will study the financial implication of the findings in the last section by calculating the annuity price.

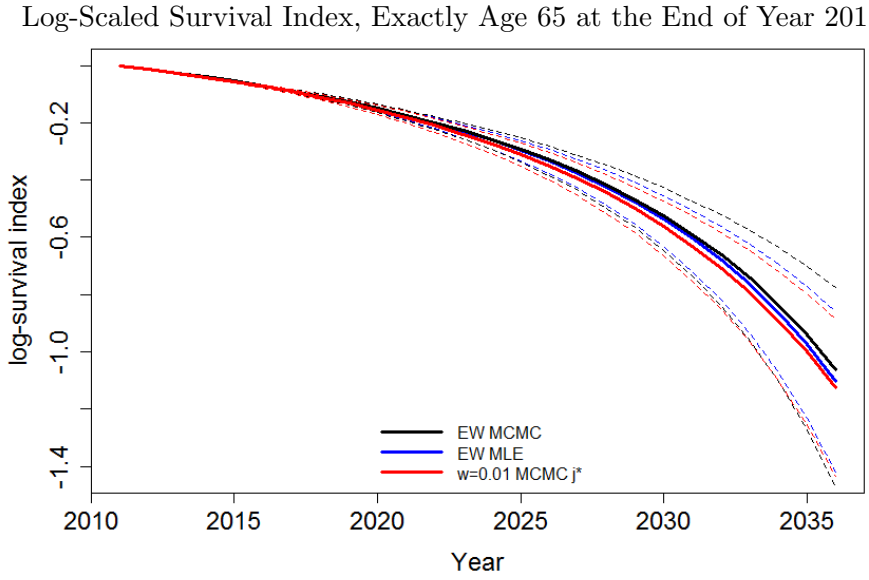


Figure 6.66: Survival Index of individual aged 65 exact at the beginning of year 2012 for EW-MCMC (black), EW-MLE (blue) and j^*-MCMC (red). The thin dashed lines are the 95% and 5% quantiles.

Similar with Section 4.4, we will consider the following annuities and discuss the effect of population size, sampling variation and the fitting method (MLE or MCMC) on their values:

- A temporary annuity of £1 per annum payable annually in arrears to a life now aged 65 exactly, starting at the beginning of year 2012 with term of 25 years. Its expected present value is calculated as:

$$a_{65:\overline{25}|} = \sum_{j=1}^{25} S(T+j, 65)v^j,$$

		$i = 4\%$	$i = 2\%$	$i = 1\%$	$i = 0.5\%$	$i = 0\%$
	SD	0.2684	0.3875	0.4686	0.5161	0.5690
	Mean	12.2631	14.8394	16.4503	17.3556	18.3365
	99.5%	12.9166	15.7830	17.5923	18.6135	19.7221
	97.5%	12.7669	15.5667	17.3301	18.3250	19.4055
	95%	12.6918	15.4573	17.1980	18.1789	19.2444
EW MCMC	Median	12.2705	14.8498	16.4626	17.3692	18.3513
	5%	11.8081	14.1836	15.6587	16.4838	17.3746
	2.5%	11.7137	14.0469	15.4927	16.3007	17.1734
	0.5%	11.5252	13.7773	15.1673	15.9443	16.7816
	SD	0.2331	0.3316	0.3983	0.4372	0.4805
	Mean	12.1174	14.6351	16.2068	17.0893	18.045
	99.5%	12.7194	15.4911	17.2364	18.2201	19.2852
	97.5%	12.5699	15.2792	16.9819	17.9403	18.9809
	95%	12.4957	15.1736	16.854	17.8002	18.8263
j^* MCMC	Median	12.1201	14.6384	16.2108	17.0939	18.0498
	5%	11.7326	14.0881	15.55	16.368	17.2523
	2.5%	11.6498	13.9713	15.4108	16.216	17.0857
	0.5%	11.4772	13.7240	15.1128	15.8893	16.7274

Table 6.9: The characteristic statistics of the price of the 25-year term annuity for an individual aged exactly 65 at the beginning of year 2012 with respect to different levels of interest rates for England and Wales data (MCMC) and death scenario j^* (MCMC)

where T corresponds to the start of year 2012.

- An annuity of £1 per annum payable annually in arrears to a life now aged 55 exactly, deferred for 10 years, starting at the beginning of year 2012 with term of 25 years. Its expected present value is:

$$ {}_{10|}a_{55:\overline{25|}} = \sum_{j=11}^{35} S(T+j, 55)v^j $$

where v is the discount factor and $S(T+t, x)$ is the survival index of an individual aged x exactly at the start of year T. Various levels of interest rates are studied (i = 4%, 2% 1%, 0.5% and 0%) since the interest rates are generally kept at a relatively low level in England and Wales. The annuity price are calculated for each sample path. For example, the annuity price of each simulated sample path for one drawn posterior sample of the death scenario j or England and Wales data are calculated and therefore there are $N_2 \times N_4$ samples of the distribution of the annuity price. The characteristic statistics of the distribution of the annuity price for EW-MCMC&MLE, j^*-MCMC&MLE are calculated in Table 6.9-6.10 (temporary annuity) and 6.11-6.12 (deferred annuity). The distribution of the survival index for the corresponding cohort year are demonstrated in Figure 6.66 and 6.67 respectively.

Unsurprisingly, the annuity price decreases given a smaller interest rate and as we may expect that for both types of the annuity, given either the MCMC or MLE, the price is more sensitive to the interest cut when the rate is relatively higher. For instance, the mean of the temporary annuity price for EW-MCMC is boosted by 2.5763 (1.6109) from 12.2631 to 14.8394 (14.8394 to 16.4503) when the interest rate halved from 4% to 2% (2% to 1%) while such increase on price is relatively even smaller for the interest rate less than 1%. The effect of the interest rate varies with respect to the type of the annuity and the deferred annuity is relatively more sensitive to the reduction of the interest rate since more years are discounted back for the deferred annuity compared with the temporary annuity. For instance, the

		$i = 4\%$	$i = 2\%$	$i = 1\%$	$i = 0.5\%$	$i = 0\%$
	SD	0.2149	0.3096	0.3739	0.4116	0.4536
	Mean	12.2166	14.772	16.3686	17.2656	18.2371
	99.5%	12.7379	15.5243	17.2773	18.2661	19.3379
	97.5%	12.6236	15.3583	17.0767	18.045	19.0969
	95%	12.5609	15.2675	16.9674	17.9252	18.9643
EW MLE	Median	12.2212	14.7787	16.3768	17.2745	18.2471
	5%	11.8558	14.2527	15.7414	16.5751	17.4762
	2.5%	11.7821	14.1483	15.6169	16.4389	17.3268
	0.5%	11.6422	13.9439	15.3704	16.1664	17.0261
	SD	0.3396	0.4828	0.5801	0.637	0.7004
	Mean	12.2052	14.7441	16.3283	17.2176	18.1805
	99.5%	12.8860	15.681	17.4512	18.449	19.5323
	97.5%	12.8079	15.5847	17.3272	18.3082	19.3724
	95%	12.7176	15.4653	17.1889	18.159	19.2113
j^* MLE	Median	12.2396	14.7826	16.3724	17.2666	18.2388
	5%	11.7310	14.0623	15.5043	16.32	17.1935
	2.5%	11.6967	14.0276	15.4574	16.2515	17.1055
	0.5%	11.2761	13.3953	14.6943	15.4166	16.1934

Table 6.10: The characteristic statistics of the price of the 25-year term annuity for an individual aged exactly 65 at the beginning of year 2012 with respect to different levels of interest rates for England and Wales data (MLE) and death scenario j^* (MLE)

mean of the deferred annuity price for EW-MCMC is raised by 3.9775 given the interest rate is sliced from 4% to 2%.

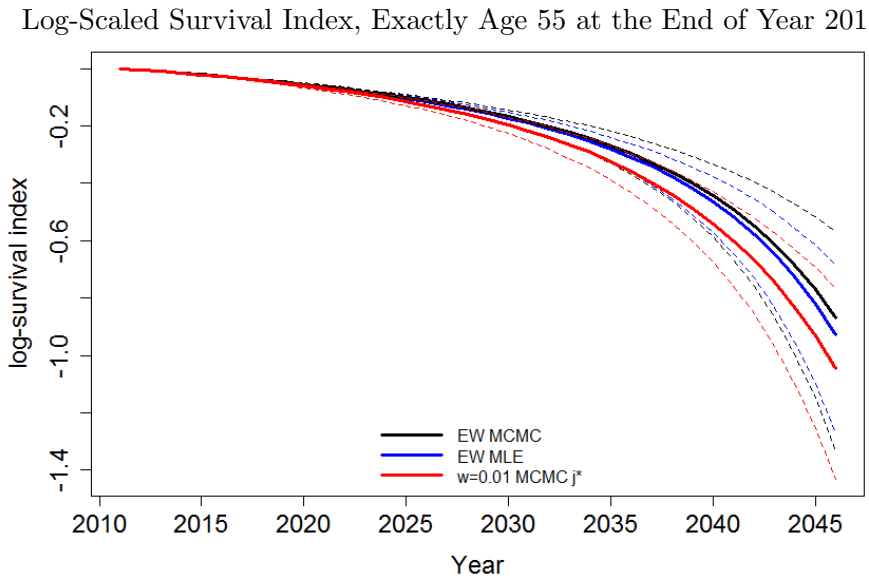


Figure 6.67: Survival Index of individual aged 55 exact at the beginning of year 2012 for EW-MCMC (black), EW-MLE (blue) and j^*-MCMC (red). The thin dashed lines are the 95% and 5% quantiles.

For both types of the annuities, unsurprisingly the mean price of j^*-MCMC is less than that of the EW-MCMC but greater than the EW-MLE's, which is consistent with our findings with regard to the central mortality projection as well as the corresponding two survival indices for the two types of annuities in Figure 6.66 and 6.67. In particular, for each level of interest rate, as we may expected the level of the differences of the mean price between EW-MCMC, j^*-MCMC and EW-MLE is relatively higher for the deferred annuity than the corresponding differences of the temporary annuity, which is consistent with the survival index plots where the gap between the central survival indices are relatively smaller for the temporary annuity since ten more years are projected for the deferred annuity and the annuity starts ten years later than the temporary annuity. However, the difference is relatively less significant compared with the impact of the interest rate, especially when a cut of the interest rate from 4% to 2% results in a boost of mean price of deferred annuity by 3.9775, 3.7517, 3.9222 for EW-MCMC, j^*-MCMC and EW-MLE respectively. Therefore, having different levels of central forecast mortality rates has a relatively

		$i = 4\%$	$i = 2\%$	$i = 1\%$	$i = 0.5\%$	$i = 0\%$
	SD	0.2570	0.4419	0.5848	0.6742	0.7785
	Mean	8.2744	12.2519	15.0503	16.7223	18.6117
	99.5%	8.8696	13.2745	16.4041	18.2822	20.4139
	97.5%	8.7388	13.0502	16.1073	17.9418	20.0205
	95%	8.6717	12.9344	15.9535	17.7637	19.8143
EW MCMC	Median	8.2878	12.2748	15.0803	16.757	18.6517
	5%	7.8287	11.4849	14.0352	15.552	17.2597
	2.5%	7.7315	11.3205	13.8186	15.3016	16.9728
	0.5%	7.5442	10.9975	13.3931	14.8143	16.4075
	SD	0.2321	0.3918	0.514	0.5901	0.6786
	Mean	7.9272	11.6789	14.3072	15.874	17.6418
	99.5%	8.5059	12.6575	15.5926	17.3524	19.3443
	97.5%	8.3676	12.4246	15.2862	16.9987	18.9359
	95%	8.2983	12.3071	15.1313	16.8197	18.7303
j^* MCMC	Median	7.9321	11.6873	14.3185	15.8868	17.6563
	5%	7.5398	11.0244	13.4475	14.8867	16.5062
	2.5%	7.4518	10.8775	13.2571	14.6705	16.2567
	0.5%	7.2765	10.5796	12.867	14.221	15.7426

Table 6.11: The characteristic statistics of the price of the 25-year term annuity deferred by 10 years for an individual aged exactly 55 at the beginning of year 2012 with respect to different levels of interest rates for England and Wales data (MCMC) and death scenario j^* (MCMC)

non-significant impact on the annuity prices.

We can also see that the uncertainties of the price distributions for the EW-MCMC, j^*-MCMC and EW-MLE are higher given the deferred annuity and once again this is because the annuity starts ten years later for the deferred annuity and therefore the survival index at the beginning of annuity-paying year (annuity is paid in arrears), unlike the survival index of temporary annuity payment starts from 1 with no uncertainty, is not equivalent to 1 with randomness from the previous ten-year projection. For both annuities, we notice that the uncertainty of the price for j^*-MCMC is greater than EW-MLE's and smaller than the EW-MCMC's as expected but the differences are relatively small, especially for the temporary annuity.

We calculated the longevity risk given the price distribution for both types of annuities with respect to EW/j^* and MCMC/MLE. The longevity risk $\mathrm{R}_{99.5\%}$ at 99.5% level is defined as:

$$\mathrm{R}_{99.5\%} = \left(\frac{a_{99.5\%}}{a_{50\%}} - 1\right) \times 100,$$

where $a_{.\%}$ is the $(\cdot)$ quantile of the random variable a for the annuity price. See Table 6.13 and 6.14 for the longevity risk given temporary and deferred annuity respectively. As expected the longevity risk increases when the interest rate goes to zero for the both types of annuities. The risk in general is relatively higher for the deferred annuity at the same interest level since the total number of years of the product is ten years longer and the payment starts later than the temporary, therefore cheaper on average with more uncertainty.

For the temporary annuity, the impact of changing the population size and modelling method on the risk is not significant. As may be expected, the risk given EW-MCMC is the highest for each interest rate level because of the wider projected mortality rates due to the choice of the prior distribution for the volatility of period effects and relatively much smaller impact of the time series prior on the

		$i = 4\%$	$i = 2\%$	$i = 1\%$	$i = 0.5\%$	$i = 0\%$
	SD	0.1942	0.3321	0.4384	0.5049	0.5824
	Mean	8.1928	12.115	14.8712	16.517	18.3759
	99.5%	8.6510	12.8974	15.903	17.7043	19.7447
	97.5%	8.5553	12.7348	15.6885	17.4574	19.4603
	95%	8.4995	12.6398	15.5642	17.3151	19.2966
EW MLE	Median	8.2000	12.1275	14.8878	16.5361	18.3975
	5%	7.8614	11.5489	14.1235	15.6561	17.3825
	2.5%	7.7956	11.4341	13.9722	15.4818	17.181
	0.5%	7.6588	11.2021	13.6681	15.1336	16.7819
	SD	0.3395	0.5676	0.7417	0.8501	0.976
	Mean	8.2539	12.2039	14.9787	16.6352	18.5059
	99.5%	8.9397	13.3498	16.4793	18.3565	20.4833
	97.5%	8.8783	13.2525	16.3454	18.1994	20.2993
	95%	8.7909	13.1125	16.1718	18.0057	20.0827
j^* MLE	Median	8.3106	12.2702	15.0663	16.7233	18.593
	5%	7.6536	11.2106	13.6899	15.1642	16.8243
	2.5%	7.6283	11.158	13.6029	15.0498	16.6766
	0.5%	7.3804	10.7316	13.0509	14.4245	15.9671

Table 6.12: The characteristic statistics of the price of the 25-year term annuity deferred by 10 years for an individual aged exactly 55 at the beginning of year 2012 with respect to different levels of interest rates for England and Wales data (MLE) and death scenario j^* (MLE)

	$i = 4\%$	$i = 2\%$	$i = 1\%$	$i = 0.5\%$	$i = 0\%$
EW-MCMC	5.27	6.28	6.86	7.16	7.47
j^*-MCMC	4.95	5.83	6.33	6.59	6.84
EW-MLE	4.24	5.04	5.50	5.74	5.98
j^*-MLE	5.12	6.08	6.59	6.85	7.09

Table 6.13: The longevity risk (in percentage) of the temporary annuity.

joint posterior distribution. The risk given j^*-MCMC is relatively closer to the risk given EW-MLE than given the j^*-MLE and the differences between both the j^*-MCMC & EW-MLE and j^*-MLE & EW-MLE are relatively stable when the interest rate reduces to zero. j^*-MLE has relatively higher longevity risk (6.08) than the j^*-MCMC does mainly due to the over-estimated MLE for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ given the two-stage approach.

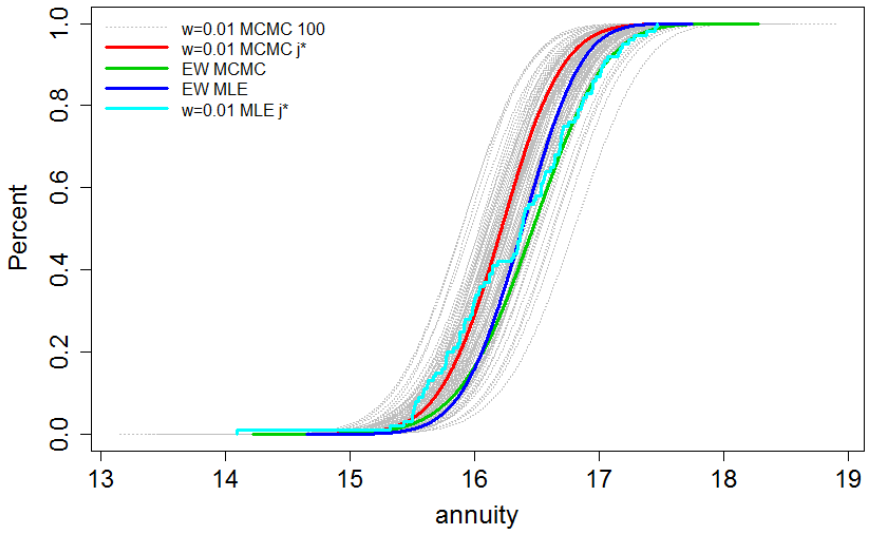
	$i = 4\%$	$i = 2\%$	$i = 1\%$	$i = 0.5\%$	$i = 0\%$
EW-MCMC	7.02	8.14	8.78	9.1	9.45
j^*-MCMC	7.23	8.3	8.9	9.23	9.56
EW-MLE	5.50	6.35	6.82	7.06	7.32
j^*-MLE	7.57	8.8	9.38	9.77	10.17

Table 6.14: The longevity risk (in percentage) of the deferred annuity.

On the other hand, the risk of the deferred annuity is more sensitive to the distribution of the mortality forecast. In particular, the risk given j^*-MLE is much higher than the others at each level of the interest rate due to the significantly higher uncertainty of price's distribution when its median price is similar with the other three's. In the contrast of the temporary annuity, the j^*-MCMC has a high risk than the EW-MCMC, implying that the difference of the average level of the

annuity is relatively greater than the difference of the uncertainty of the two price distributions. This is consistent with our findings according to the survival indices in Figure 6.67. Similar reason for a relatively larger gap between the risk of j^*-MCMC and EW-MLE than the temporary annuity.

(a) CDF: Temporary Annuity price



(b) CDF: Deferred Annuity price

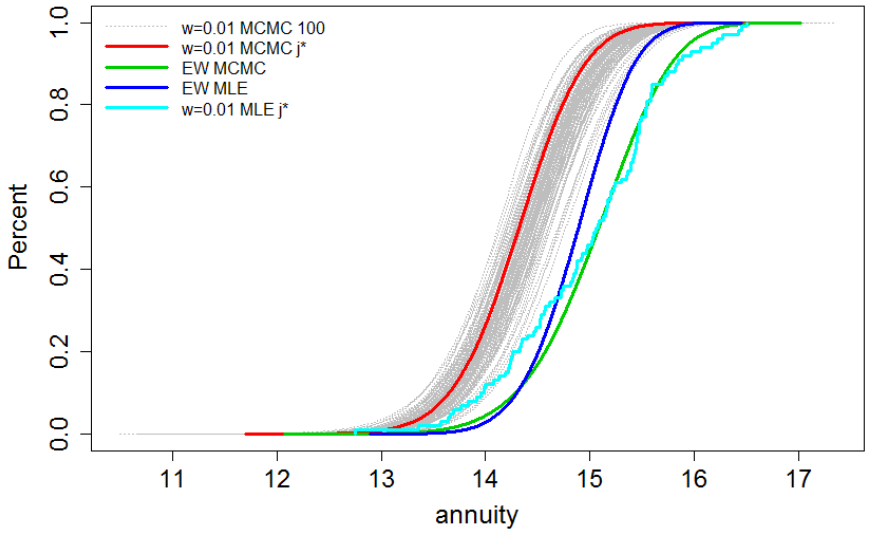


Figure 6.68: The CDF of the temporary (upper) and deferred (lower) annuity at interest rate 1%. Each of the grey curve is the distribution in one of the 100 random death scenarios of $w = 0.01$ and the red curve is for death scenario j^*. The blue and green curves are for the EW-MLE and EW-MCMC respectively. The light blue curve is for the j^*-MLE, which is not as smooth as the others since only 100 sample paths are simulated for j^*-MLE.

We have already found that the impact of sampling variation shifts the distribution of mortality projection up and down from one death scenario to another.

In Figure 6.68, we investigate how the sampling variation affect the distribution of annuity price of both types by plotting the CDFs of the 100 death scenarios (black dotted curves) with interest rate of 1%[4]. As may be expected, for both types the sampling variation shifts the annuity around from one death scenario to another. For the temporary annuity, the distribution of j'-MCMC is shifted and approximately centred around the CDF for the EW-MLE (blue curve), implying that the distribution j^*-MCMC is shifted to the left hand side of the EW-MLE by the sampling variation and there is no significant difference between the annuity price of death scenario j'-MCMC and EW-MLE.

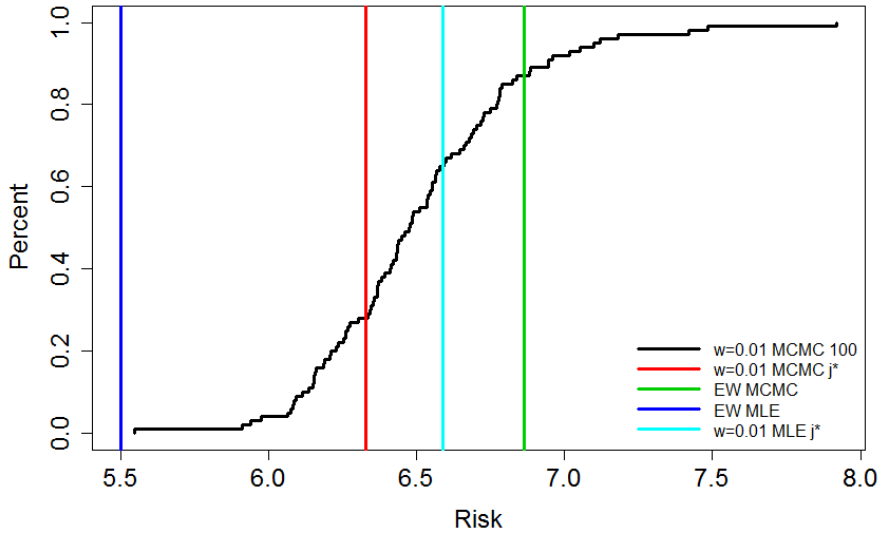
On the other hand, for the deferred annuity the distribution of EW-MLE lies in the right tail and is in general slightly higher than the CDFs of the j'-MCMCs'. This is consistent with our findings that the gap between the central projection increases along the forecasting horizon and therefore more impact on the annuity pricing. The price of deferred annuity is more influenced by the mortality rates compared with the temporary annuity since the payment are deferred by ten years and thus ten more years are projected.

At last, we calculated the longevity risk for the 100 death scenarios of interest rate 1%, see the CDF in Figure 6.69 for both types[5]. We can see that the for both types the longevity risk of small population given MCMC is greater than the EW-MLE mainly because the uncertainty of the survival index is approximately the same for the two while EW-MLE has a higher central survival index and hence a higher median of the annuity price. In particular, the gap of the risk between the two is relatively greater for the deferred annuity as may be expected when the annuity is influenced more by the survival index.

[4]See the CDFs given other level of interest rates in Figure D.6-D.9in Appendix D

[5]See the plots for the other level of interest rates in Figure D.11-D.13 in Appendix D

(a) CDF: Longevity Risk given a Temporary Annuity



(b) CDF: Longevity Risk given a Deferred Annuity

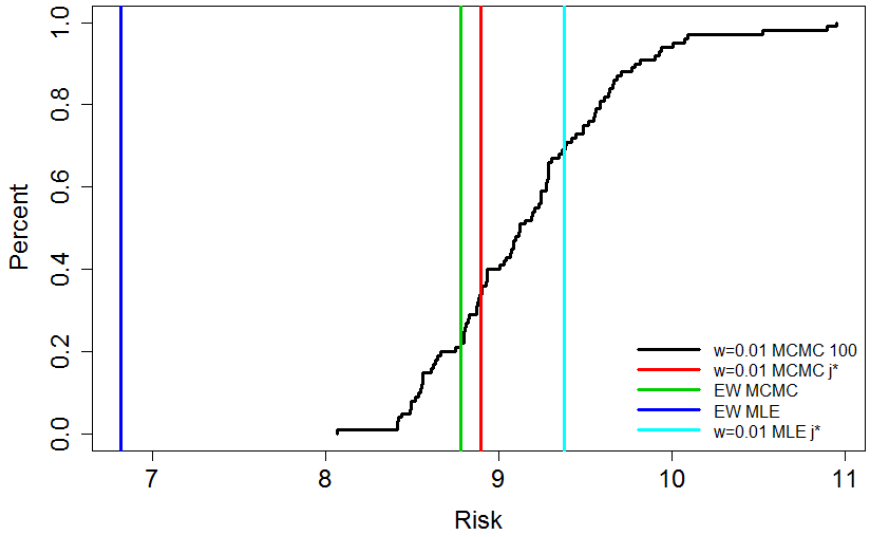


Figure 6.69: Longevity risk given temporary (upper) and deferred (lower) annuity of interest rate 1%. The black curve is the distributions of 100 random death scenarios with $w = 0.01$ and the red line is for death scenario j^*. The blue and green line are for the EW-MLE and EW-MCMC respectively. The light blue line is for the j^*-MLE.

6.3 Summary

In this chapter, we found that the mortality projection of the small population based on our MCMC method is not significantly different from the "true" projection. We adopted the Bayesian approach by combining the fitting and forecasting stages into one for the simulated death scenario j^* with exposure equivalent to one

percent of the benchmark population whose true rates are the point estimates of the benchmark. A relatively strong prior distribution with the information of the true rates is adopted for the volatility of the period effects. In particular, we fixed the mean and the mode (sensitivity test) of the prior distribution for the volatility to the true volatility parameter. Then a comparison is carried our for the fitting and forecasting with respect to the modelling method and the population size, i.e. the EW-MLE (true rate), EW-MCMC, j^*-MCMC and j^*-MLE. The influence of the sampling variation on the posterior distribution of fitting and forecasting are then studied by re-simulating N_1 death scenarios, each of which is fitted and projected by the Bayesian model.

When the population size is small, the impact of the time series prior, i.e. the random walk and AR likelihood which form the prior density, dominates the posterior distribution and competes with the Poisson likelihood such that the estimated period and cohort effects are more like the proposed time series process. Therefore, we observed that for the cohort process, the trajectory of the posterior distribution of j^*-MCMC is much more smoothed by the AR(1) likelihood compared with the trajectory of j^*-MLE. The credibility interval based on j^*-MCMC is narrower than the EW-MCMC credibility interval, tilted from the true rate EW-MLE and more strictly centred around zero like a zero mean-reverting AR(1) process. Correspondingly, the three period effects are tilted. The MCMC trajectories for each of the period effects are much more smoothed than that of the respective MLE. The credibility intervals of the MCMC for the first and the second period effects are narrower than the relevant intervals of the MCMC for the England and Wales data. Unlike the two-stage approach significantly over-estimating the volatility of the period and cohort effects, the MCMC could provide greatly improved estimation for the volatility, based on our prior beliefs about the location of the true parameter. In our experiment, given the underlying prior distribution (Inverse-Wishart) we applied for the volatility of the period effect, it turns out that allocating the mode, rather than the mean, of the prior to our prior believe about the location of the true

parameter generates a better estimation. We observed that the sampling variation shifts the level of the posterior distributions of the latent and hyper parameters up and down from one death scenario to another with relatively non-significant impact on the posterior variance.

The fitted rates given the j^*-MCMC are much more smoothed by the time series prior compared with the j^*-MLE and therefore the respective trajectory of MCMC is quite similar to the trajectory of the fitted rates given EW-MLE, especially at the very old and young ages when the j^*-MLE completely failed to capture most of the features of the true rates. The sampling variation shifts the posterior distribution of the fitted rates up and down from one death scenario to another without significantly changing the posterior variance. We observed that the true rate lies within the "distribution" of the central fitted rates driven by the sampling variation for all of the ages. It implies that the sampling variation could shift the distribution (in particular the mean) higher or lower than the true rate but there is not significant difference between the true rates and the central fitted rates given MCMC.

We projected the mortality rate forward by 50 years and observed that the prediction intervals given j^*-MCMC is only slightly higher than that of the EW-MLE and narrower than the EW-MCMC because of the lower volatility estimation given MCMC-Mean. At age 50, the central projection of EW-MLE is higher than the EW-MCMC and j^*-MCMC since the projection of the cohort effect starts five years earlier for the EW-MLE, given the five very early and late short cohorts have been removed. The central projection of j^*-MCMC and EW-MCMC crossed each other when approximately 73 years are projected for the death rates. For the other ages, the central projection given EW-MLE is higher than the EW-MCMC while lower than the j^*-MCMC. The sampling variation shifts the prediction intervals up and down without significantly changing the widths and the degree of the shift (e.g. for the central prediction) varies with respect to the age level. We thus studied the distribution of the central mortality improvement rate driven by the sampling varia-

tion based on the simulated death scenarios. The mean of the central improvement rate is a strict quadratic function of $x - \bar{x}$ while the variance is approximately a quadratic function of $(x - \bar{x})^2$. We also worked out the ages at which the mean and variance are the lowest. We observed that the widths of the predictive intervals given j^*-MCMC is not significantly greater than the intervals given EW-MLE, whose central prediction is also within the distribution of the central prediction driven by the sampling variation for most of the ages. The true central prediction is only lower than the small population at the ages when the variance of the central mortality improvement rate is very low.

We studied the financial implication by calculating two different types of annuities (temporary and deferred annuities) based on the projections of England and Wales data and the simulated death scenarios, given both the MCMC and MLE.. The influence of the mortality projection impacts more on the deferred annuity due to the longer projection horizon (extra 10 years).

In practice, as we have discussed that we are not given the information of the true rates of the underlying population and we do not deal with as many as 1000 populations. However, if a small pension scheme manager has solid prior knowledge that his/her underlying population shares some characteristics with a large population (for example, the volatility of the small population is approximately the same or follows a function of the volatility for the big population), then the estimation of the volatility for the small population could be improved by adopting the information of the big population with an informative prior distribution, and we understand that the impact of the sampling variation is to shift the estimation and the projection only.

Chapter 7

Empirical Case Study

In the previous chapter, we applied the Bayesian approach by combining the fitting and forecasting stages into one for modelling a one percent sized subset of the benchmark population with its true rate equivalent to the point estimate of the benchmark.

7.1 Introduction

In this chapter, we will be carrying out a case study by fitting our Bayesian model to a population that is not a subset of the benchmark population. We select Scotland as our small population and once again England and Wales (EW) to be our benchmark population. The reason of the choices is that Scotland is not a subset of England and Wales, however both are subsets of the UK population and adjacent to each other in geography. The last two features imply that the two populations could share some common characteristics and therefore the information of the England and Wales could be employed for modelling the Scotland mortality with an informative prior distribution. Therefore it is reasonable to assume that there should be no significant difference between the volatility of the two populations and we will use

the information of the volatility of England and Wales as our prior knowledge for estimating the volatility of Scotland.

7.2 Data

The datasets for the case study are the males in UK, England and Wales, Scotland from the Human Mortality Database (HMD). Due to the data quality issue of the Scottish dataset before the year 1971, we will fit the corresponding dataset for each population during the years 1971 to 2011 instead and once again aged from 50 to 89 last birthday. Recall that the Scottish deaths counts are not strict integers during certain years according to the HMD, however the raw datasets are perfect integers in a yearly format for each age level. The reason of this is not interpreted on any document available to the users. Therefore, similar with Chapter 4 we will directly use the raw data without making any adjustment for the Scotland. The datasets for the other two populations are used in a normal way.

7.3 Modelling

Denote as $\hat{\boldsymbol{\theta}}_1^{(\cdot)}$ the MLEs of the latent parameters for a population $(\cdot)$, where $(\cdot)$ = EW, UK, ST for the England and Wales, UK and Scotland respectively. The corresponding MLEs for the hyper-parameters, more specifically the drift $\boldsymbol{\mu}$ and volatility $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ of the random walk model for the period effects and the coefficient α_γ and volatility σ_γ of the zero mean-reverting $AR(1)$ model for the cohort effect, are then calculated conditional on the $\hat{\boldsymbol{\theta}}_1^{(\cdot)}$ according to Equations (3.8), (3.9) and (3.17) (see page 41, 48) respectively.

While modelling the Scotland population with the Bayesian method, similar with Chapter 6, a more informative prior distribution (Inv-Wishart($\nu = 12, \Sigma$))

is assumed for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ and the information of the England and Wales is adopted by choosing the values of ν and Σ such that the mean of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ is fixed to the corresponding MLE of the England and Wales. The sensitivity test includes fixing the mean of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the MLE of the UK, the mode of the prior of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the MLEs of both England and Wales and the UK. To be more specific, the alternative prior distributions for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ are listed as follows:

A-UK: $\text{Mode}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{UK}}$,

A-EW: $\text{Mode}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$,

B-EW: $\text{E}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$,

B-UK: $\text{E}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{UK}}$.

The reason for setting A/B-UK is that both the England and Wales and the Scotland are the subset populations of the UK and therefore Scotland is expected to have the same volatility with the UK and we expect that there would be no significant difference on the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ between either A-UK and A-EW or B-UK and B-EW.

Once again, we simulate N_1 death scenarios $D^w(t,x) = \{D_j^w(t,x)\}_{j=1,\ldots,N_1}$ with the Poisson model:

$$D^w(t,x) \sim \text{Pois}(m(\theta_{1,0},t,x)wE_0(t,x)),$$

where the benchmark exposure is the England and Wales and the true rate of $D^w(t,x)$ is fixed to the England and Wales point estimates, that is $\theta_{1,0} = \hat{\theta}_1^{\text{EW}}$. The weight w is fixed to 0.1 since the Scottish population size is approximately ten percent of the England and Wales. Similar with the $p(\boldsymbol{V}_{\boldsymbol{\epsilon}})$ of $w = 0.01$, the prior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ of $w = 0.1$ is assumed to follow a Inverse Wishart distribution with the degree of freedom $\nu = 12$ and the scale matrix Σ is determined such that

the $\mathrm{E}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}$ (B-EW). The purpose of re-simulating death scenarios is to investigate how the sampling variation would shift the posterior distribution of the population sized $w = 0.1$. Therefore we can tell if the posterior distribution of the volatility of the Scotland is shifted away from the corresponding point estimate of England and Wales only by the sampling variation.

As for the joint posterior distribution of the England and Wales, in addition to the non-informative Jeffrey's prior distribution for the joint distributions of $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, the prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ are listed as follows:

C-JR $p(\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}) \propto |\boldsymbol{V}_{\boldsymbol{\epsilon}}|^{-2}$,

A-EW $\mathrm{Mode}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}$, $p(\boldsymbol{V}_{\boldsymbol{\epsilon}}) \sim \text{Inv-Wishart}(\nu = 12, \Sigma)$,

B-EW $\mathrm{E}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}$, $p(\boldsymbol{V}_{\boldsymbol{\epsilon}}) \sim \text{Inv-Wishart}(\nu = 12, \Sigma)$.

Note that the degrees of freedom ν are fixed to 12 for consistency purposes. Once again, we are not claiming that A-EW and B-EW that re-use the estimation of the EW are good prior distributions for the posterior distribution of the England and Wales and the main reason is to be consistent with the prior distribution $p(\boldsymbol{V}_{\boldsymbol{\epsilon}})$ of the simulated death scenarios with $w = 0.1$.

The prior distributions for the rest of the parameters are following exactly the same settings introduced in Chapter 6 for the small population $w = 0.01$. We simulated the joint posterior distribution with MCMC. For each population given varied prior settings, N_2 posterior samples are drawn from the posterior distribution after the MCMC process achieves stationary.[1]

[1]We demonstrate the trace plots of the stationary MCMC for selected latent and hyper-parameters of Scotland given varied prior settings in Figure E.1-E.3 in Appendix E

7.4 Results

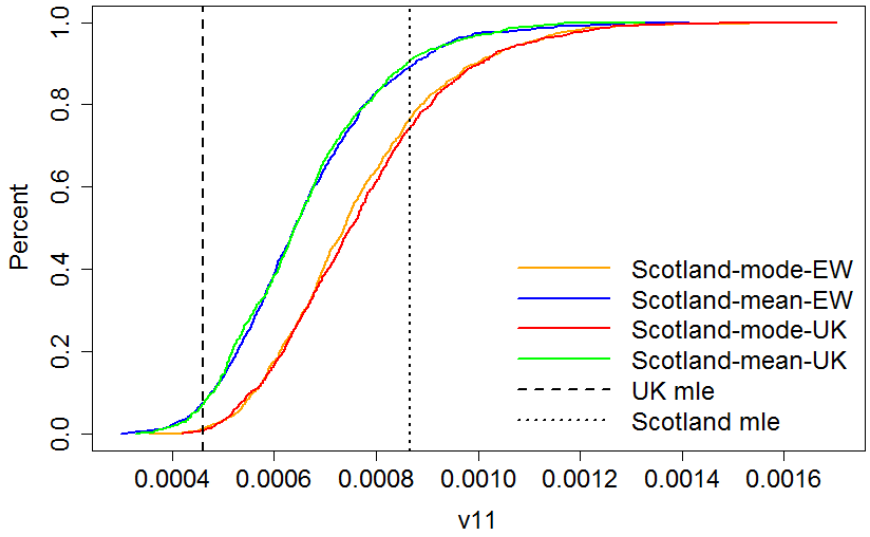
We will start with the impact of the varied prior settings of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ on the posterior distribution of the hyper-parameters of Scotland. In Figure 7.1-7.2, we plotted the CDFs of the hyper-parameters for Scotland, including $\tilde{\boldsymbol{\mu}}_1^{\text{SL}}$, $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$ for the random walk model of the period effects and $\tilde{\alpha}_{\gamma}^{\text{SL}}$, $\tilde{\sigma}_{\gamma}^{\text{SL}}$ for the AR(1) model of the cohort effect, given A-EW (orange), A-UK (red), B-EW (blue) and B-UK (green).[2]

As may be expected, the posterior distributions $\tilde{\boldsymbol{\mu}}^{\text{SL}}$, $\tilde{\alpha}_{\gamma}^{\text{SL}}$ and $\tilde{\sigma}_{\gamma}^{\text{SL}}$ are not significantly influenced by the different candidate $p(\boldsymbol{V}_{\boldsymbol{\epsilon}})$. As for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, unsurprisingly Figure 7.1a shows that both A-EW and A-UK push the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$ to the right hand side of the respective distributions given B-EW and B-UK due to the higher prior mean and variance. We can see that the green curve (B-UK) and the blue one (B-EW) approximately cover each other, same with the red (A-UK) and the orange (A-EW), which imply that the underlying way of using the information of volatility of either the England and Wales or the UK does not significantly change the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$. Similar results can be observed for the other entries of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$.

We will then focus on assessing the performance of our Bayesian model for the parameter estimation, especially the volatility $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ of the random walk model. We will firstly compare the posterior distribution with the MLE of England & Wales and Scotland. The closer the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}$ to the MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$ would imply that the two populations do have a similar volatility and using the information could improve the estimation. On the other hand, we could end up with $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}$ lying between $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{SL}}$ and $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$. Recall that in Chapter 6 we concluded that the impact of the sampling variation is to shift the posterior distribution from one side to another and such shift could be very large. We will therefore investigate if it is the sampling variation that shifts the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}$ by following the Method 5 adopted in Chapter 6 and comparing the

[2]See Figure E.4-E.6 in Appendix E for the posterior distributions of the other hyper-parameters.

(a) CDF: $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{SL}}(1,1)$



(b) CDF: $\tilde{\boldsymbol{\mu}}_1^{\mathrm{SL}}$

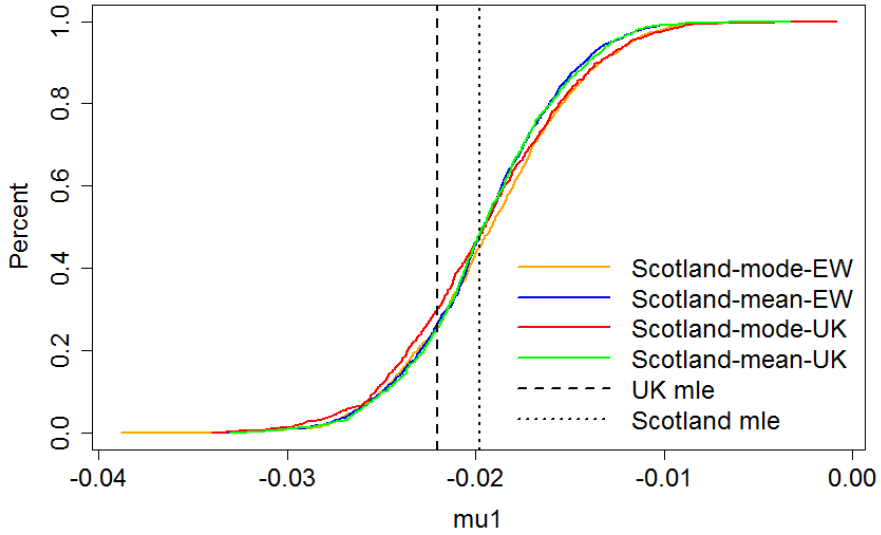
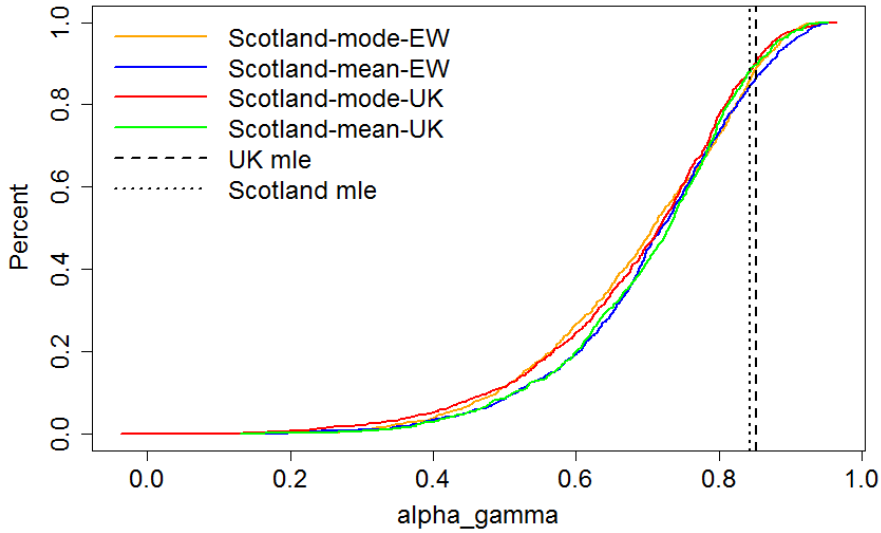


Figure 7.1: CDF: the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ and $\boldsymbol{\mu}_1$ for the Scotland, given varied prior settings for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$, e.g. A-EW (orange), A-UK (red), B-EW (blue) and B-UK (green). The dashed vertical line is the point estimate for the UK and the dotted line is for the Scotland.

distribution of the posterior distributions of 350 selected simulated death scenarios of $w = 0.01$ with the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{SL}}$. Figure 7.3-7.4 and 7.6-7.8 include the CDFs of $\tilde{\boldsymbol{V}}_{\epsilon}$ and $\tilde{\boldsymbol{\mu}}$ (upper) for both England and Wales data and $\tilde{\boldsymbol{V}}_{\epsilon}^{w}$ and $\tilde{\boldsymbol{\mu}}^{w}$ for N_1 simulated death scenarios (lower).

In Figure 7.3a, the Scotland MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{SL}}(1,1)$ (vertical dotted line) is greater than the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$ of the England and Wales data (vertical dashed line) and lies

(a) CDF: $\tilde{\alpha}_\gamma^{\text{SL}}$



(b) CDF: $\tilde{\sigma}_\gamma^{\text{SL}}$

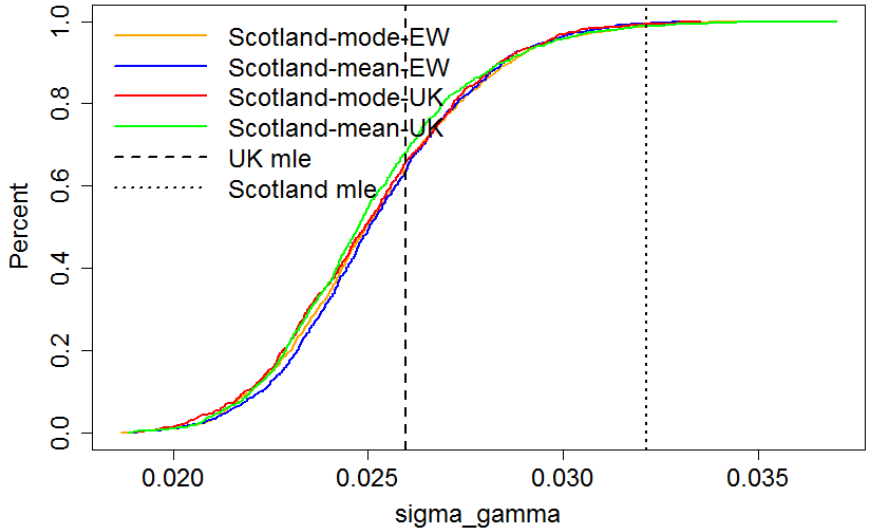


Figure 7.2: CDF: the posterior distribution of $\tilde{\alpha}_\gamma^{\text{SL}}$ and $\tilde{\sigma}_\gamma^{\text{SL}}$ for Scotland, given varied prior settings for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$, e.g. A-EW (orange), A-UK (red), B-EW (blue) and B-UK (green). The dashed vertical line is the point estimate for the UK and the dotted line is for the Scotland.

above 95% quantile of $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w=0.1}(1,1)$ (purple curve), the CDF of the finite sample MLE of the N_1 death scenarios. The latter implies that the volatility of the period effect $\boldsymbol{\kappa}^{(1)}$ of the Scotland is not the same but higher than the MLE of the England and Wales data $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ and therefore the period effect $\boldsymbol{\kappa}^{(1)}$ could be relatively more volatile for Scotland than for the England and Wales data. The algorithm of this implication is that $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ is the true rate of the simulated death scenarios of $w = 0.1$ and therefore the point estimate of the Scotland should be no less or

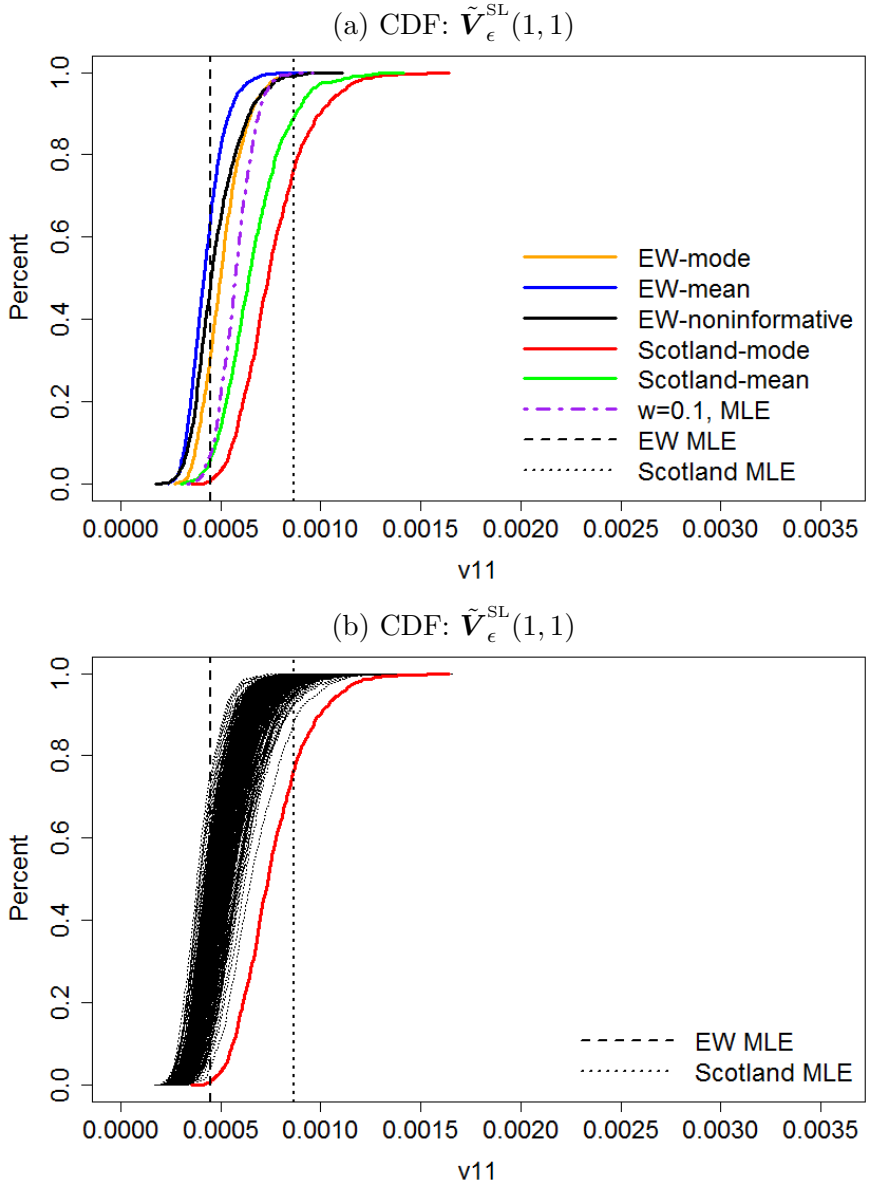


Figure 7.3: Upper: the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ for Scotland, given A-EW (red) and B-EW (green); for England and Wales given A-EW (orange), B-EW (blue) and C-JR (black). The purple dotted curve is the CDF of the finite sample MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(1,1)$ for $w = 0.1$. The vertical dashed and dotted lines are the corresponding MLEs of the England and Wales and Scotland respectively. Lower: CDF of 350 selected simulated death scenarios with $w = 0.1$, given A-EW. The red curve is the CDF for Scotland given A-EW.

greater than most of the finite-sample MLEs $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(1,1)$ if the Scottish volatility is not significantly differed from the England and Wales data MLE. Note that the finite sample MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w=0.1}(1,1)$ is relatively not significantly higher than the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$ compared with our findings for $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w=0.01}(1,1)$ in Chapter 6 since 10% of the England and Wales is relatively a big population and the impact of the sampling

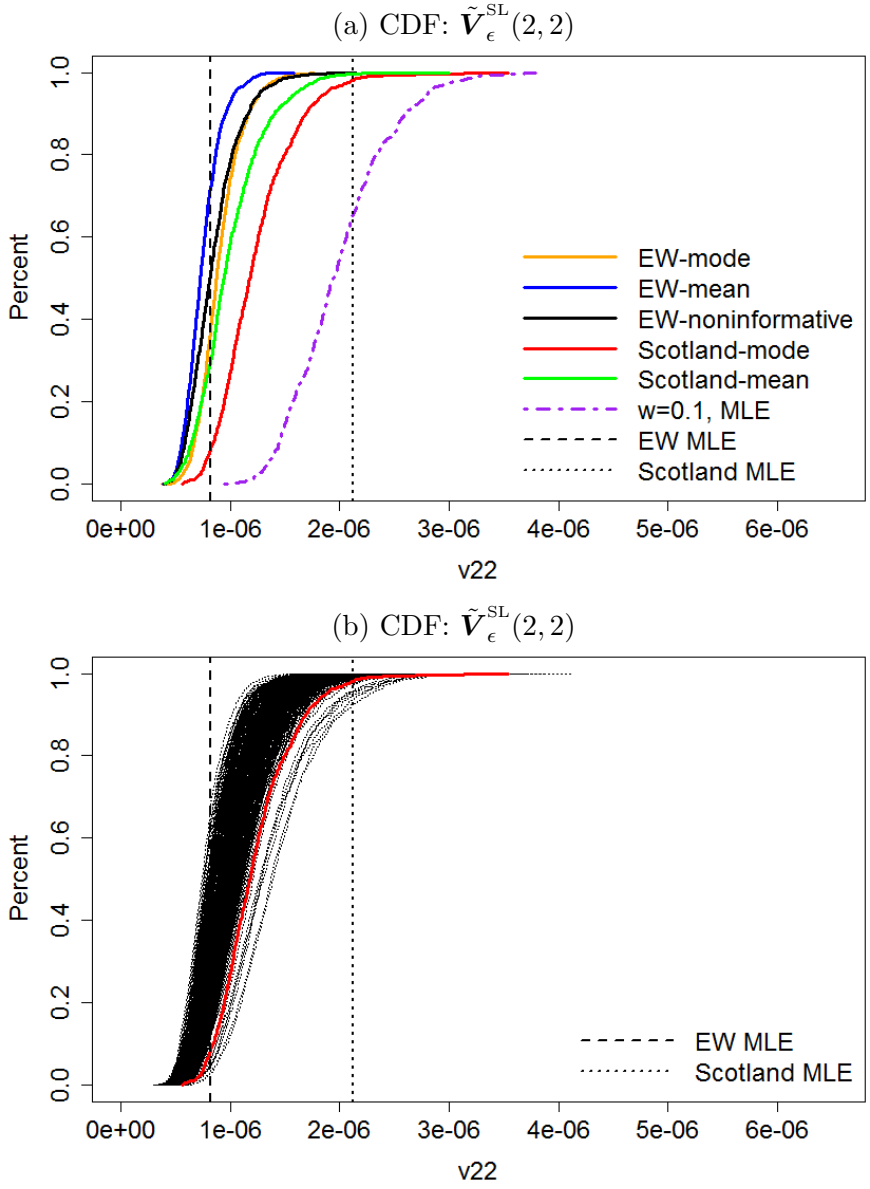


Figure 7.4: Upper: the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ for Scotland, given A-EW (red) and B-EW (green); for England and Wales given A-EW (orange), B-EW (blue) and C-JR (black). The purple dotted curve is the CDF of the finite sample MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(2,2)$ for $w = 0.1$. The vertical dashed and dotted lines are the corresponding MLEs of the England and Wales and Scotland respectively. Lower: CDF of 350 selected simulated death scenarios with $w = 0.1$, given A-EW. The red curve is the CDF for Scotland given A-EW.

variation over-estimating the volatility of $\boldsymbol{\kappa}^{(1)}$ is not so strong. Unsurprisingly, for both populations the $\tilde{\boldsymbol{V}}_{\epsilon}(1,1)$ given A-EW is higher than that given the B-EW. We can see that $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(1,1)$ given C-JR (black curve) lies between the A-EW (orange curve) and B-EW (blue curve).

The distribution of $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$, given both the A-EW (red curve) and B-EW (green curve), lies on the right hand side of the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(1,1)$, indicating a higher posterior estimation for the volatility of the period effect $\boldsymbol{\kappa}^{(1)}$. It is worth noticing that the point estimates $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{SL}}(1,1)$ and $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ lie on the right and left tails of the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$ (A-EW & B-EW) respectively. We therefore need to check if it is the sampling variation that shifts the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$ away from the point estimate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$. We will compare the posterior distribution of the Scotland with the distribution of N_1 posterior distributions of the simulated death scenarios whose true rate is the point estimate of the England and Wales data. Recall that the sampling variation shifts these distributions from one side to another. Therefore the algorithm of this check is that if the distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$ is no higher or lower than, for example 95% and 5% respectively, of the N_1 death scenarios, there is no significant difference the (true) volatility of $\boldsymbol{\kappa}^{(1)}$ of the Scotland and the point estimate of England and Wales data, otherwise different.

In Figure 7.3b, we plotted the CDFs of the N_1 death scenarios and once again we can observe the influence of the sampling variation shifting the posterior distribution from one side to another approximately around the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$ of the death scenarios. The distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$ given A-EW is above the 95% quantile of the distribution of the posterior distributions driven by the sampling variation. We therefore conclude that the volatility for $\boldsymbol{\kappa}^{(1)}$ of Scotland is statistically not the same but slightly higher than the respective point estimate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$. Visually, comparing the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$ given B-EW with the distribution of the posterior distributions does not change our conclusion. The reason we only study the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$ given A-EW in Figure 7.3b is to be consistent with our conclusion in Chapter 6 that fixing the mode of the prior of $\tilde{\boldsymbol{V}}_{\epsilon}$ to the true rate for the simulated death scenarios provides an unbiased estimate of the volatility of the period effects.

For $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$, both the MLE of the Scotland $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{SL}}(2,2)$ and the CDF of the finite sample MLEs of the simulated death scenarios $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(2,2)$ are relatively shifted

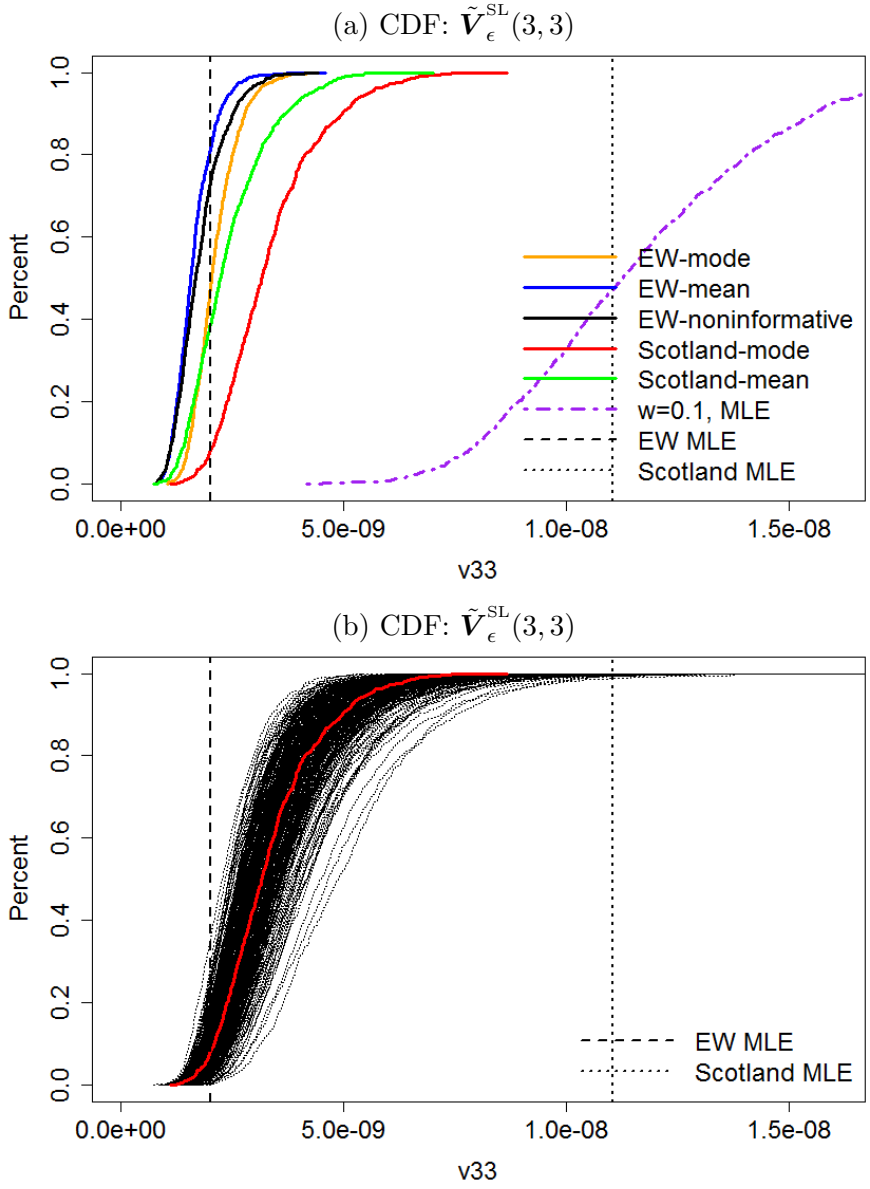


Figure 7.5: Upper: the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ for Scotland, given A-EW (red) and B-EW (green); for England and Wales given A-EW (orange), B-EW (blue) and C-JR (black). The purple dotted curve is the CDF of the finite sample MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(3,3)$ for $w = 0.1$. The vertical dashed and dotted lines are the corresponding MLEs of the England and Wales and Scotland respectively. Lower: CDF of 350 selected simulated death scenarios with $w = 0.1$, given A-EW. The red curve is the CDF for Scotland given A-EW.

more to the right hand side of the MLE of the England and Wales data $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(2,2)$, compared with the scale of the shift for $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{SL}}(1,1)$ and $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(2,2)$ respectively from $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(1,1)$. It implies that the MLE of the volatility for $\boldsymbol{\kappa}^{(2)}$ is more sensitive to an increase of the sampling variation due to the drop of population size. The MLE for the Scotland $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{SL}}(2,2)$ lies approximately on the 60% quantile of the finite sample

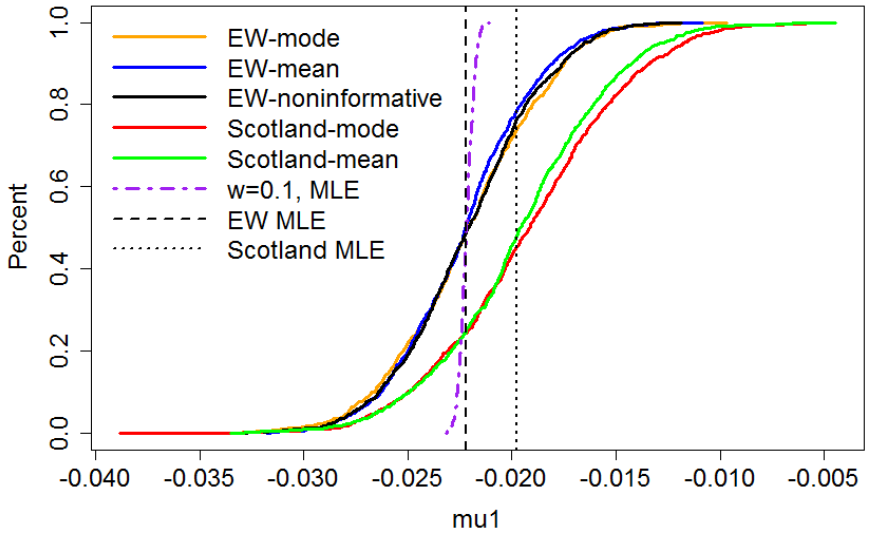
MLEs of the simulated death scenarios $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$. Therefore we could conclude that statistically there is no significant difference between the Scottish volatility and the respective MLE of the England and Wales data given the underlying model, according to the same algorithm we discussed with $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$. In other words, the two populations have a similar level of volatility for $\boldsymbol{\kappa}^{(2)}$.

Visually, the posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(2,2)$ given both the A-EW and B-EW is significantly lower than the corresponding MLE of Scotland $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{SL}}(2,2)$ and stays relatively much closer to the point and posterior estimate of the England and Wales data, compared with $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$. In particular, the EW MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$ is higher than approximately 40% quantile of $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(2,2)$ given B-EW and 15% quantile of $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(2,2)$ with A-EW. It implies that the volatility of period effect $\boldsymbol{\kappa}^{(2)}$ of the Scotland could be the same with the MLE of the England and Wales data given the underlying model.

Therefore we plotted the CDFs of the N_1 simulated death scenarios $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(2,2)$ in Figure 7.4b. Once again the sampling variation shifts the posterior distribution from one side to another around the true rate $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(2,2)$. The posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(2,2)$ lies within the distribution of the N_1 posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(2,2)$, which supports our conclusion that it is the sampling variation shifts the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(2,2)$ away from the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(3,3)$ while statistically there is no significant difference between the volatility of the period effect $\boldsymbol{\kappa}^{(2)}$ of the Scotland and the respective MLE of the England and Wales data given the underlying model.

As for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$, we can observe an even further shift on both the MLE of the Scotland $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{SL}}(3,3)$ and the finite sample MLEs of the simulated death scenarios $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(3,3)$ to the right hand side of the England and Wales' MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(3,3)$. It implies that the MLE of $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}(3,3)$ is relatively more sensitive to the increased sampling variation due to the smaller population size than the volatilities of the other two period effects. The $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(3,3)$ is visually centred around the point estimate of the Scotland $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{SL}}(3,3)$ and is relatively much wider spread than $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w}(1,1)$ and

(a) CDF: $\tilde{\boldsymbol{\mu}}_1^{\mathrm{SL}}$



(b) CDF: $\tilde{\boldsymbol{\mu}}_1^{\mathrm{SL}}$

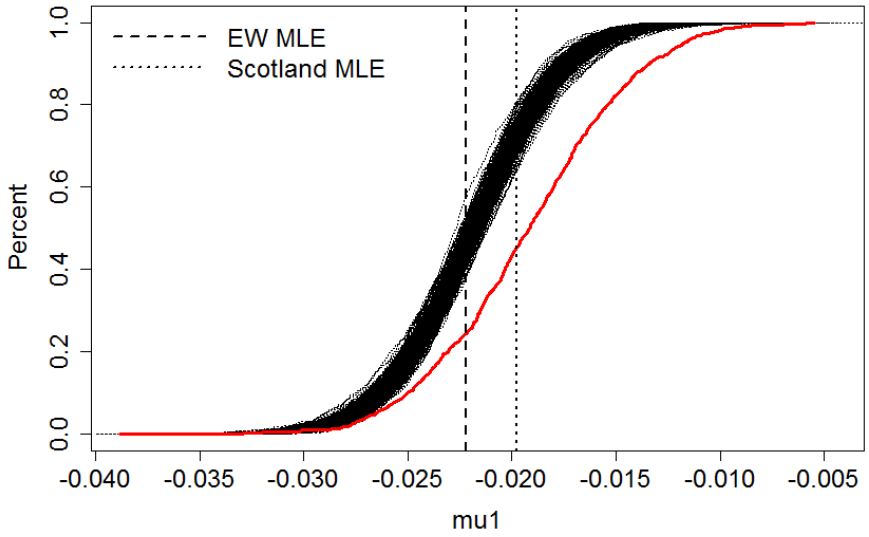


Figure 7.6: Upper: the posterior distribution of $\boldsymbol{\mu}_1$ for the Scotland, given A-EW (red) and B-EW (green); for England and Wales given A-EW (orange), B-EW (blue) and C-JR (black). The purple dotted curve is the CDFs of the finite sample MLE $\hat{\boldsymbol{\mu}}_1^w$ for $w = 0.1$. The vertical dashed and dotted lines are the corresponding MLEs of the England and Wales and Scotland respectively. Lower: CDF of 350 selected simulated death scenarios with $w = 0.1$, given A-EW. The red curve is the CDF for Scotland given A-EW.

$\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{w}(1,1)$. We therefore conclude that there is no significant difference between the level of the volatility of the period effect $\boldsymbol{\kappa}^{(3)}$ for the Scotland and the England and Wales' corresponding MLE.

The posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{SL}}(3,3)$ (A&B-EW) is far less than the corresponding MLE as may be expected and yet slightly higher than the MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}_{\mathrm{EW}}}(3,3)$ and the

posterior distribution $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(3,3)$ of the England and Wales data. In particular, the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(3,3)$ (B-EW) is approximately centred around the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{\text{EW}}(3,3)$ while once again the latter is higher than at least 15% quantile of the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(3,3)$ (A-EW). We therefore need to check if the sampling variation shifts the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(3,3)$ away from $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{\text{EW}}(3,3)$. The CDFs of the posterior distributions of the N_1 simulated death scenarios $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(3,3)$ are plotted in Figure 7.5b, where the posterior distribution of the Scotland $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(3,3)$ (A-EW) is in the centre of the N_1 CDFs. We therefore conclude that the volatility of the $\boldsymbol{\kappa}^{(3)}$ of the Scotland is not significantly different from the MLE $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}{}^{\text{EW}}(3,3)$ of the England and Wales data.

In Figure 7.6a for the drift $\boldsymbol{\mu}_1$, as may be expected the MLE of the Scotland $\hat{\boldsymbol{\mu}}_1^{\text{SL}}$ is higher than the $\hat{\boldsymbol{\mu}}_1^{\text{EW}}$ of the England and Wales data. The CDF of the finite sample MLEs of the N_1 simulated death scenarios (purple curve) $\hat{\boldsymbol{\mu}}_1^{w}$ is centred around the true rate $\hat{\boldsymbol{\mu}}_1^{\text{EW}}$ and we can see that $\hat{\boldsymbol{\mu}}_1^{\text{SL}}$ is higher than the maximum value of the $\hat{\boldsymbol{\mu}}_1^{w}$, which strongly indicates that the $\boldsymbol{\mu}_1$ for the Scotland is significantly higher than the MLE for the England and Wales data.

As may be expected, for both populations the posterior distributions of $\boldsymbol{\mu}_1$ are centred around the corresponding MLE of the same population and not much influenced by changing the prior distribution of the volatility $\boldsymbol{V}_{\boldsymbol{\epsilon}}$. We can see that the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{\text{SL}}$ is shifted to the right hand side of the $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ and the $\hat{\boldsymbol{\mu}}_1^{\text{EW}}$ and both populations have similar posterior variance. In Figure 7.6b we plotted the CDFs of the posterior distributions for the N_1 simulated death scenarios $\tilde{\boldsymbol{\mu}}_1^{w}$ and the sampling variation shifts the posterior distribution from one side to another around the true rate $\hat{\boldsymbol{\mu}}_1^{\text{EW}}$. The red curve is the CDF of the posterior distribution for the Scotland $\tilde{\boldsymbol{\mu}}_1^{\text{SL}}$ given A-EW. We find that $\tilde{\boldsymbol{\mu}}_1^{\text{SL}}$ is significantly higher than $\tilde{\boldsymbol{\mu}}_1^{w}$, the distribution of N_1 posterior distributions. We therefore conclude that the drift of the period effect $\boldsymbol{\kappa}^{(1)}$ of the Scotland is significantly higher than $\hat{\boldsymbol{\mu}}_1^{\text{EW}}$, the corresponding MLE of the England and Wales data.

In Figure 7.7a, the MLE for the drift of $\boldsymbol{\kappa}^{(2)}$ of the Scotland $\hat{\boldsymbol{\mu}}_2^{\text{SL}}$ is at approx-

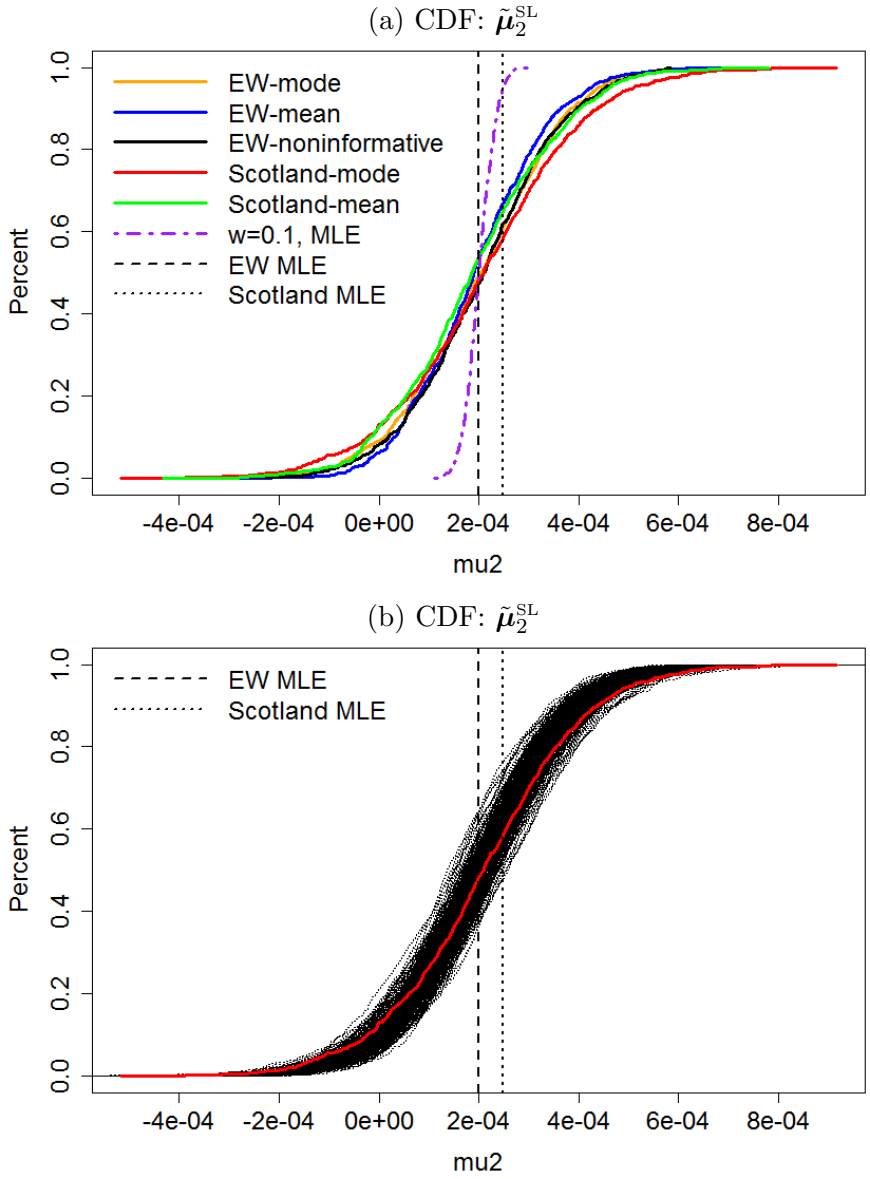


Figure 7.7: Upper: the posterior distribution of $\boldsymbol{\mu}_2$ for the Scotland, given A-EW (red) and B-EW (green); for England and Wales given A-EW (orange), B-EW (blue) and C-JR (black). The purple dotted curve is the CDFs of the finite sample MLE $\hat{\boldsymbol{\mu}}_2^w$ for $w = 0.1$. The vertical dashed and dotted lines are the corresponding MLEs of the England and Wales and Scotland respectively. Lower: CDF of 350 selected simulated death scenarios with $w = 0.1$, given A-EW. The red curve is the CDF for Scotland given A-EW.

imately 90% quantile of the CDF for the finite sample MLEs of the N_1 simulated death scenarios $\hat{\boldsymbol{\mu}}_2^w$. We therefore conclude that the drift $\boldsymbol{\mu}_2$ for the Scotland is not significantly higher than $\hat{\boldsymbol{\mu}}_2^{\mathrm{EW}}$, the MLE for the England and Wales data.

We can see that for both populations, the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{\mathrm{EW}}$ and $\tilde{\boldsymbol{\mu}}_2^{\mathrm{SL}}$ are approximately the same and centred around the MLE of the England and Wales

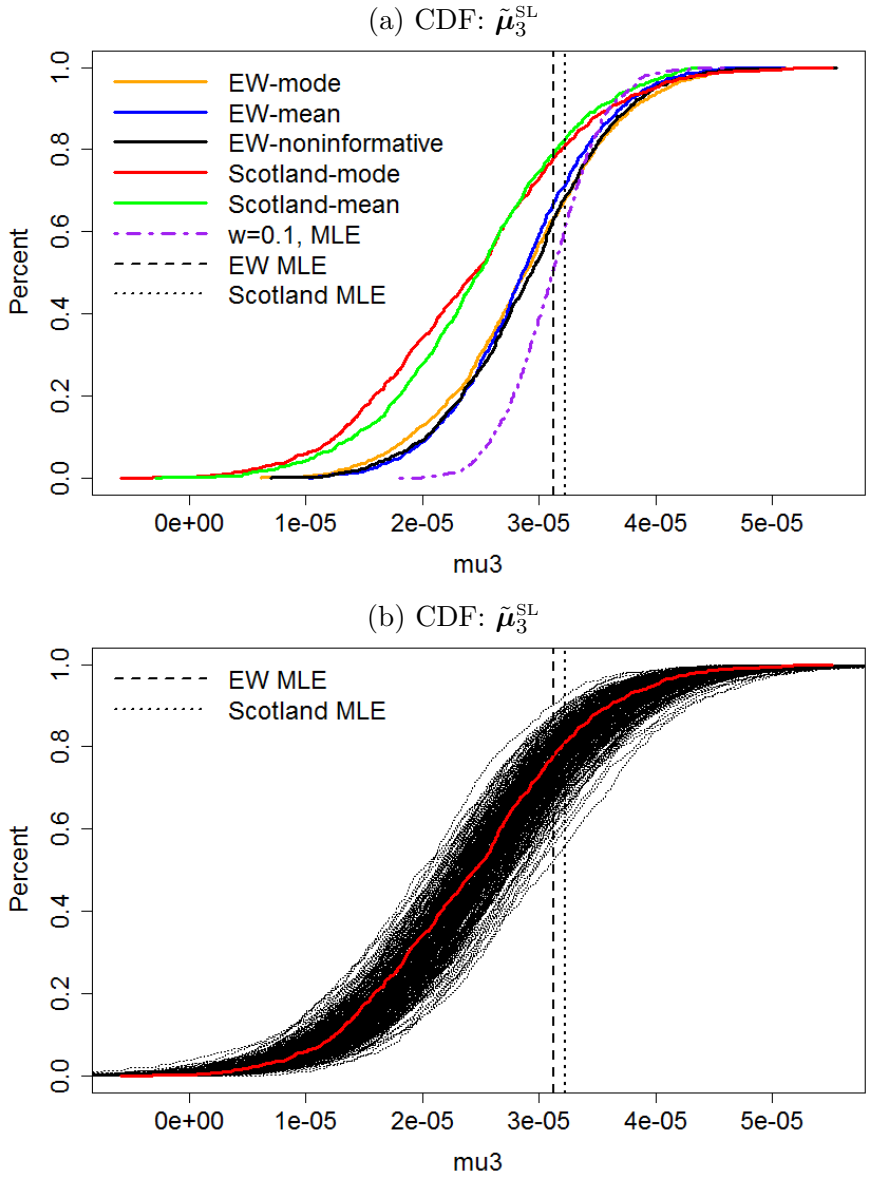


Figure 7.8: Upper: the posterior distribution of $\boldsymbol{\mu}_3$ for the Scotland, given A-EW (red) and B-EW (green); for England and Wales given A-EW (orange), B-EW (blue) and C-JR (black). The purple dotted curve is the CDFs of the finite sample MLE $\hat{\boldsymbol{\mu}}_3^w$ for $w = 0.1$. The vertical dashed and dotted lines are the corresponding MLEs of the England and Wales and Scotland respectively. Lower: CDF of 350 selected simulated death scenarios with $w = 0.1$, given A-EW. The red curve is the CDF for Scotland given A-EW.

data with similar posterior variance, regardless the prior settings for the volatility. We then plotted $\tilde{\boldsymbol{\mu}}_2^w$, the distribution of the posterior distributions for N_1 simulated death scenarios in Figure 7.7b, where once again the sampling variation shifts the distributions around the true rate $\hat{\boldsymbol{\mu}}_2^{EW}$. The red curve is the CDF for the posterior distribution $\tilde{\boldsymbol{\mu}}_2^{SL}$ (A-EW), which is in the middle of the $\tilde{\boldsymbol{\mu}}_2^w$. We therefore conclude

that the drift of $\boldsymbol{\kappa}^{(2)}$ for the Scotland is statistically the same with the MLE of the England and Wales data.

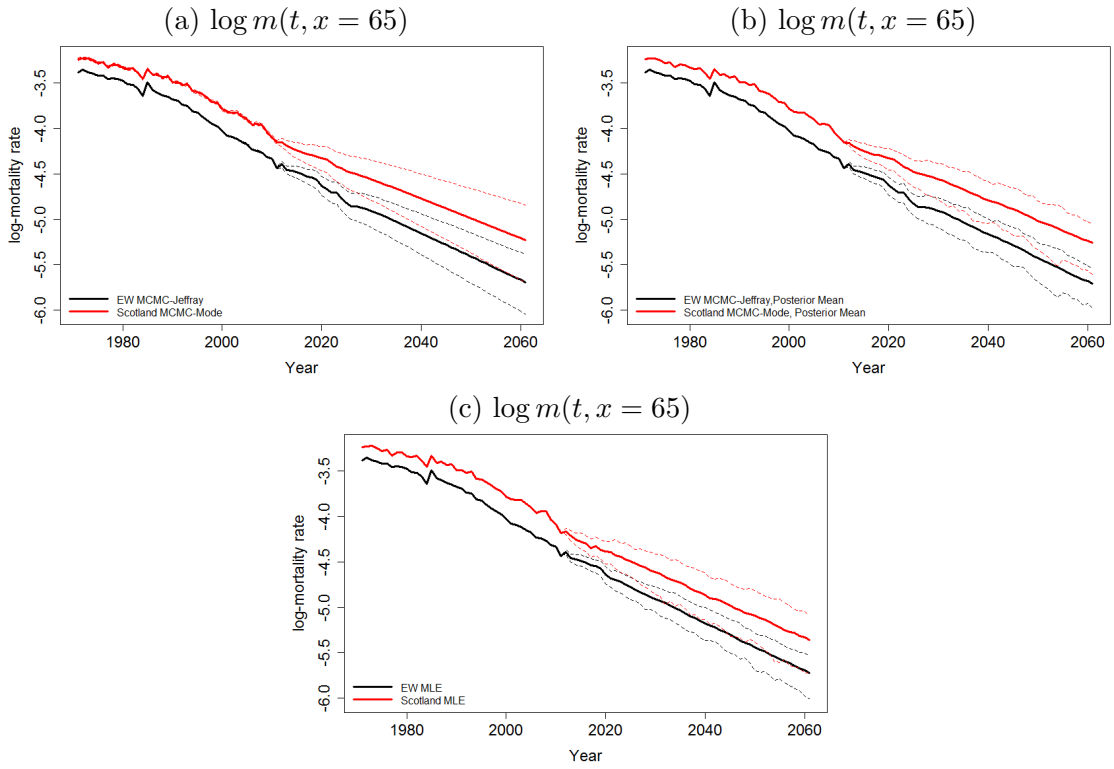


Figure 7.9: Log-scaled $m(t, x)$ at age 65 for England and Wales (black) and Scotland (red). (a) Scotland-PU vs. EW-PU; (b) Scotland-PC (Mean) vs. EW-PC (Mean); (c) Scotland-MLE vs. EW-MLE. The dashed lines are the 95% (upper) and 5% (lower) quantiles respectively.

In Figure 7.8a, the MLE of the drift $\tilde{\boldsymbol{\mu}}_3$ of the Scotland is relatively much closer to the point estimate of the England and Wales data, compared with the Scottish point estimates of the other two drifts. Visually, $\hat{\boldsymbol{\mu}}_3^{\mathrm{SL}}$ is approximately at the 60% quantile of the distribution of the finite sample MLEs of the N_1 simulated death scenarios. It indicates that the drift $\boldsymbol{\mu}_3$ of the Scotland and the respective MLE of the England and Wales data are statistically the same.

We can see that the posterior distribution $\tilde{\boldsymbol{\mu}}_3^{\mathrm{SL}}$ (A-EW & B-EW) of the Scotland is shifted to the left hand side of the posterior distribution $\tilde{\boldsymbol{\mu}}_3^{\mathrm{EW}}$ and the MLE $\hat{\boldsymbol{\mu}}_3^{\mathrm{EW}}$ with slightly higher posterior variance. In Figure 7.8b, we plotted $\tilde{\boldsymbol{\mu}}_3^{w}$, the distribution of the posterior distributions of N_1 simulated death scenarios. We can see that sampling variation shifts the posterior distribution from one side to another

and $\tilde{\boldsymbol{\mu}}_3^w$ is in general slightly lower than the MLE $\hat{\boldsymbol{\mu}}_3^{\text{EW}}$ of the England and Wales data. The posterior distribution $\tilde{\boldsymbol{\mu}}_3^{\text{SL}}$ (A-EW) of the Scotland (red curve) is in the centre of the $\tilde{\boldsymbol{\mu}}_3^w$ and we therefore could conclude that the Scotland has its drift of $\boldsymbol{\kappa}^{(3)}$ statistically the same with the respective MLE of the England and Wales data. Some characteristic statistics for the prior and posterior distribution of the $\boldsymbol{\mu}$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for both England and Wales and Scotland data are calculated in Table 7.1-7.3.

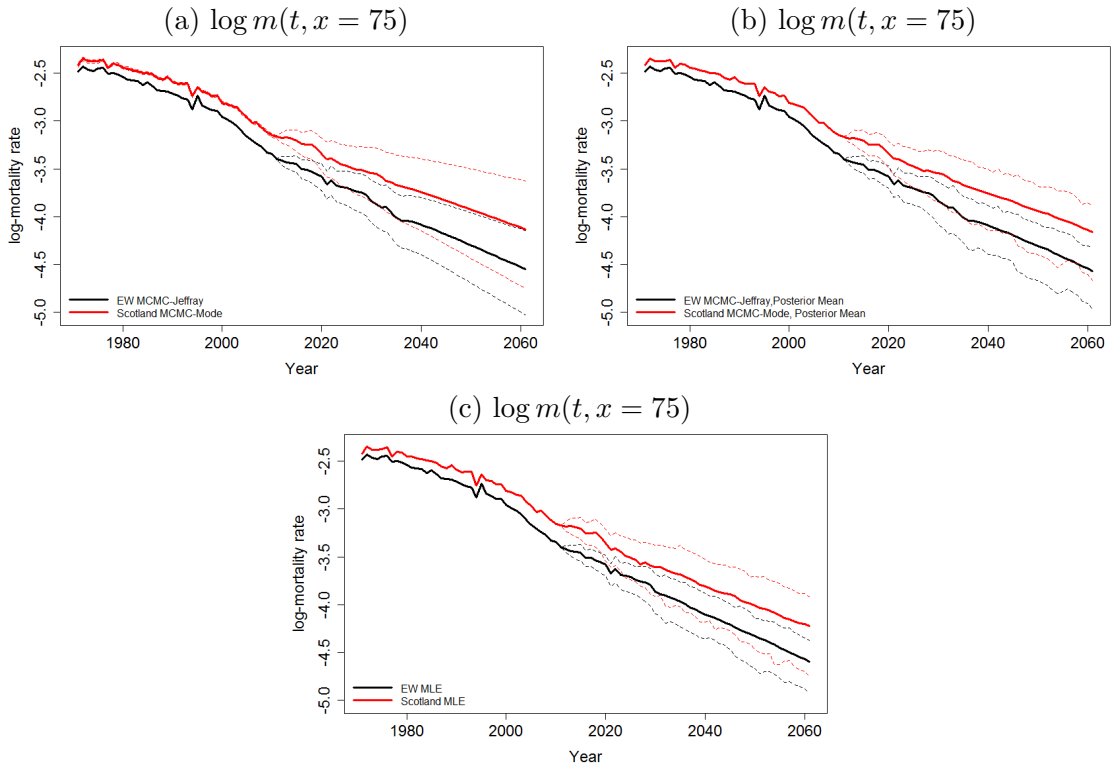


Figure 7.10: Log-scaled $m(t, x)$ at age 75 for England and Wales (black) and Scotland (red). (a) Scotland-PU vs. EW-PU; (b) Scotland-PC (Mean) vs. EW-PC (Mean); (c) Scotland-MLE vs. EW-MLE. The dashed lines are the 95% (upper) and 5% (lower) quantiles respectively.

At last, for both populations, the death rate is projected forward by fifty years, denoted as $m'(t', x)$, conditional on:

$\tilde{\boldsymbol{\theta}}_1'|\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2$ = the posterior predictive distribution of all the latent parameters conditional on the posterior distributions of all the latent $\tilde{\boldsymbol{\theta}}_1$ and the hyper-parameters $\tilde{\boldsymbol{\theta}}_2$. The method of generating the posterior predictive distribution is introduced in Chapter 5. Denote as $m'(\tilde{\boldsymbol{\theta}}_1', t', x)$ the projected death rate with full parameter uncertainty.

$\bar{\tilde{\boldsymbol{\theta}}}_1'|\bar{\tilde{\boldsymbol{\theta}}}_1, \bar{\tilde{\boldsymbol{\theta}}}_2$ = projected latent parameters conditional on $\bar{\tilde{\boldsymbol{\theta}}}_1$ and $\bar{\tilde{\boldsymbol{\theta}}}_2$ the posterior mean of the latent and hyper-parameters respectively, which are calculated according to the method introduced in Chapter 6. Therefore the projected death rate $m'(\bar{\tilde{\boldsymbol{\theta}}}_1', t', x)$ conditional on $\bar{\tilde{\boldsymbol{\theta}}}_1'$ does not include any parameter uncertainty.

$\hat{\boldsymbol{\theta}}_1'|\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2$ = projected latent parameters conditional on the MLEs. In other words, the projected death rate $m(\hat{\boldsymbol{\theta}}_1', t', x)$ is calculated by the usual two-stage approach.

We denote as $(\cdot)$-PU/ PC (Mean) /MLE for the projected death rate with parameter uncertainty; without parameter uncertainty; two-stage approach respectively, where $(\cdot)$=England and Wales (EW) and Scotland (SL). In Figure 7.9-7.11, we plotted the fan charts for the log-scaled fitted and projected mortality rates conditional on PU (a), PC (b) and MLE (c) at age 65, 75, and 85 for England and Wales (black lines) and the Scotland data (red lines). Note that the dashed lines are the 5% and 95% quantiles of the distributions.

In general, an obvious empirical cohort (estimated cohort effect) impact on the mortality projection can be observed during the earlier forecasting horizon for both populations at the selected ages. The Scotland has a higher level of the fitted rate and central mortality forecasts than the England and Wales at the selected ages conditional on all of the three conditions (PU, PC and MLE). For the convenience purpose, we denote as $\Delta m^{(\cdot)}(t, x)$ and $\Delta m'^{(\cdot)}(t', x)$ the difference of the fitted and central projected log-scaled death rates at year t and t' respectively, aged x between the England and Wales and the Scotland, i.e. the gap between the solid red and black lines. Note that $(\cdot)$ =PU, PC and MLE and for the condition PU, $\Delta m^{\text{PU}}(t, x)$ represents the difference of the central fitted rate.

In particular, for each selected age level x, $\Delta m^{(\cdot)}(t, x)$ is smaller and remains relatively stable during the fitting horizon compared with $\Delta m'^{(\cdot)}(t', x)$ of the same x for all the three conditions. As may be expected, the differences of the central

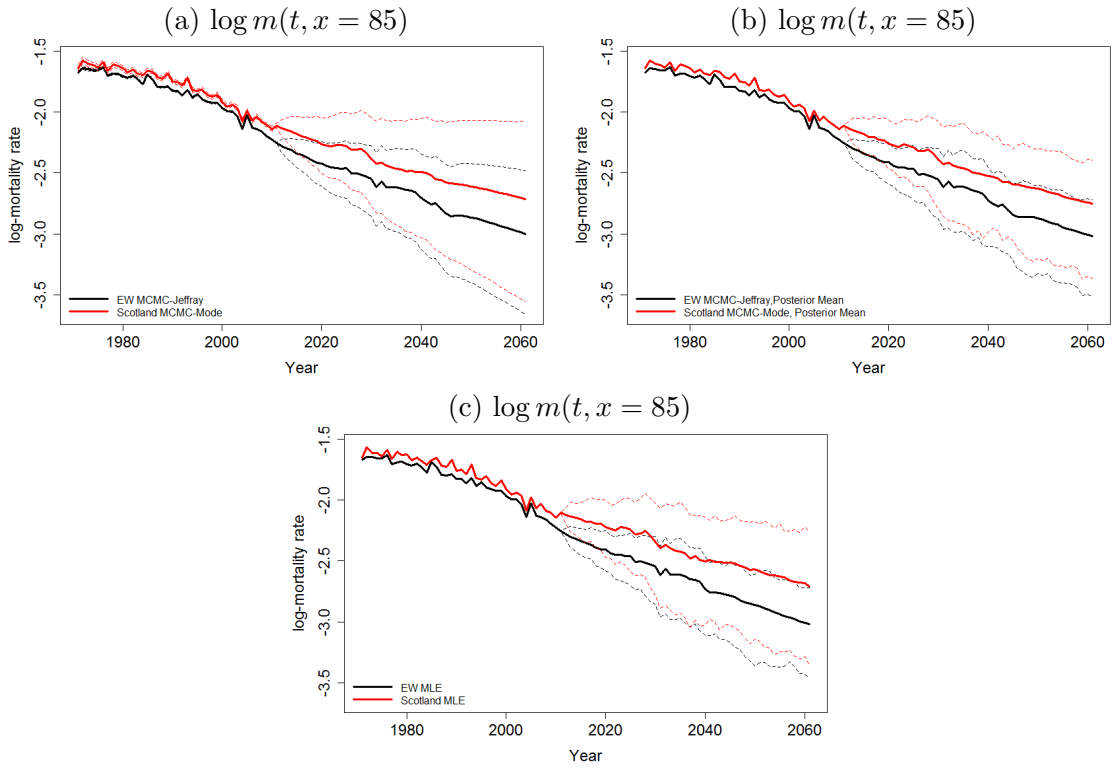


Figure 7.11: Log-scaled $m(t, x)$ at age 85 for England and Wales (black) and Scotland (red). (a) Scotland-PU vs. EW-PU; (b) Scotland-PC (Mean) vs. EW-PC (Mean); (c) Scotland-MLE vs. EW-MLE. The dashed lines are the 95% (upper) and 5% (lower) quantiles respectively.

projection increase through the forecasting horizon, especially without the impact of the cohort effect. This finding is consistent with our conclusion that the MLE of $\boldsymbol{\mu}_1$ for Scotland is significantly higher than the corresponding point estimate of England and Wales data and the posterior distribution $\tilde{\boldsymbol{\mu}}_1^{\text{SL}}$ for Scotland is shifted to the right hand side of $\tilde{\boldsymbol{\mu}}_1^{\text{EW}}$ for England and Wales data, see Equation (6.2) for the central prediction at year t' in Chapter 6. We can also see that $\Delta m^{(\cdot)}(t, x)$ becomes smaller at year t when the age increases while $\Delta m'^{(\cdot)}(t', x)$ remains relatively stable for the forecasting year t' at higher ages, especially during the forecasting horizon with no impact of empirical cohort effect.

Unsurprisingly, for both populations the prediction intervals increases through the forecasting horizon at each selected age level. As may be expected, the prediction intervals is wider at a higher age level. We can see that the Scottish projection have more uncertainty through the forecasting horizon compared with the projection of

England and Wales data given all the three conditions due to a higher estimation for the volatility of period effects for the Scotland.

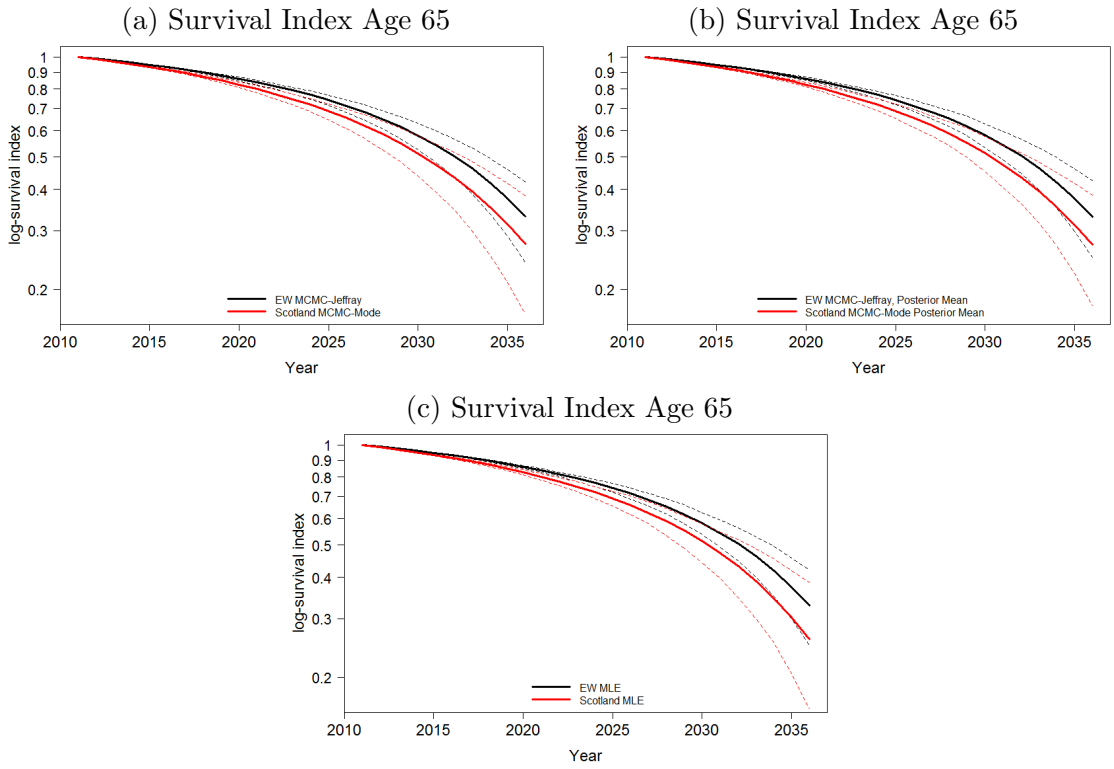


Figure 7.12: Log-scaled survival index at age 65 for England and Wales (black) and Scotland (red). (a) Scotland-PU vs. EW-PU; (c) Scotland-PC (Mean) vs. EW-PC (Mean); (c) Scotland-MLE vs. EW-MLE. The dashed lines are the 95% (upper) and 5% (lower) quantiles respectively.

We will then focus on the impact of parameter uncertainty (PU) on the mortality forecasting. Figure 7.9a and 7.9b show that for both populations, the prediction intervals are wider for the condition PU compared with PC through the forecasting horizon at age 65 since the randomness of the projection given PU includes both the normal randomness and the parameter uncertainty while the uncertainty is driven by only the normal randomness for the projection given the condition PC. Same results can be observed for the other ages. Note that the prediction intervals of PC and MLE are less smoothed compared with the intervals given PU since only 100 sample paths are simulated for these two conditions. It is worth noticing that the differences of the projection uncertainty between the condition PU and PC are relatively larger at a higher age level for both populations. As may be expected, both

PU and PC generate approximately the same central forecasts with similar degree of smoothness through the forecasting horizon and this is true for both populations at all the ages. The prediction interval given PU is wider than that given MLE. This is once again due to the additional parameter uncertainty and the our conclusion that the volatility of $\boldsymbol{\kappa}^{(1)}$ of Scotland is relatively not very significantly over-estimated by the two-stage approach (see Figure 7.3a) and therefore $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{SL}}(1,1)$ is not significantly higher than the $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(1,1)$.

Figure 7.9b and 7.9c shows that the central forecasts does not vary too much from the condition PC to MLE, as may be expected, while the prediction interval given MLE for Scotland is wider than that given the SL-PC while on the other hand the two conditions generate similar projection uncertainties for the England and Wales data at age 65, as we expect. Same results can be observed for the other ages. This is once again because of the two-stage approach over-estimating the volatility of the period effects for the relatively smaller sized population (Scotland) due to the larger sampling variation, while on the other hand we have concluded that there is no significant difference between the volatility of $\boldsymbol{\kappa}^{(2)}$ and $\boldsymbol{\kappa}^{(3)}$ of Scotland and the respective point estimates of the England and Wales data and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$ is only slightly higher than the $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(1,1)$. The Bayesian approach provides a better estimation for the volatility of period effects for the Scotland compared with the two-stage approach. It is also worth noticing that for all the ages, the central forecasts is slightly more smoothed given both the PU and PC for both populations compared with the condition MLE.

At last, we plotted the survival index for individuals aged exactly 65 at the beginning of year 2012 given the three conditions for England and Wales data (black) and the Scotland (red) in Figure 7.12. The dashed lines are the 5% and 95% quantiles of the projection. In general, Scotland has a lower central index with higher uncertainty compared with the England and Wales data given all the three conditions, which is consistent with our findings for the mortality projection. For both populations,

				$\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$
Prior	Mean	England and Wales	A	8.95E-04	1.64E-06	4.02E-09
			B	4.47E-04	8.18E-07	2.01E-09
			C	NA	NA	NA
		Scotland	A	8.95E-04	1.64E-06	4.02E-09
			B	4.47E-04	8.18E-07	2.01E-09
	Std Dev	England and Wales	A	5.17E-04	9.44E-07	2.32E-09
			B	2.58E-04	4.72E-07	1.16E-09
			C	NA	NA	NA
		Scotland	A	5.17E-04	9.44E-07	2.32E-09
			B	2.58E-04	4.72E-07	1.16E-09

Table 7.1: The mean and the standard deviation of the prior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$, $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$, and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ for England and Wales and Scotland data, during year 1971-2011, aged 50-89 last birthday. Note A = Fixing the mode of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the point estimate of England and Wales data $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$; B = Fixing the mean of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the poinrt estimate of England and Wales data $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$; C = Jeffrey's Prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for modelling England and Wales data.

as may be expected the central survival index is approximately the same given the conditions PU and PC, while on the other hand the survival index given condition PU has a greater uncertainty compared with the index given PC due to the additional parameter uncertainty. It is worth noticing that for the England and Wales data, the distribution of survival index given the two-stage approach (Figure 7.12c) is approximately the same with the Bayesian approach (Figure 7.12a and 7.12b) due to a large population while the uncertainty of the Scotland given the condition MLE is wider than the other two conditions since the volatilities of the period effects are over-estimated. It also implies that the impact of such over-estimation on the uncertainty survival index over-weighs the influence of the parameter uncertainty.

				μ_1	μ_2	μ_3
Posterior	Mean	England and Wales	A	-2.21E-02	2.00E-04	2.87E-05
			B	-2.23E-02	1.95E-04	2.86E-05
			C	-2.21E-02	2.06E-04	2.90E-05
		Scotland	A	-1.92E-02	2.09E-04	2.42E-05
			B	-1.95E-02	1.92E-04	2.46E-05
	Std Dev	England and Wales	A	3.59E-03	1.53E-04	7.44E-06
			B	3.23E-03	1.34E-04	6.43E-06
			C	3.39E-03	1.47E-04	6.90E-06
		Scotland	A	4.50E-03	1.86E-04	9.36E-06
			B	4.16E-03	1.63E-04	8.12E-06

Table 7.2: The mean and the standard deviation of the posterior distribution of $\boldsymbol{\mu}$ for England and Wales and Scotland data, during year 1971-2011, aged 50-89 last birthday. Note A = Fixing the mode of the prior for $\boldsymbol{V_\epsilon}$ to the point estimate of England and Wales data $\hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{\text{EW}}$; B = Fixing the mean of the prior for $\boldsymbol{V_\epsilon}$ to the poinrt estimate of England and Wales data $\hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{\text{EW}}$; C = Jeffrey's Prior for $\boldsymbol{V_\epsilon}$ for modelling England and Wales data.

7.5 Summary

In this Chapter, we carried out an empirical study with our Bayesian model based on the Scotland males data. Information of the England and Wales datasets are used to elicit prior distributions for the volatility of the random walk model for the period effects of the Scotland data. To summarise, we find the England and Wales and the Scotland have similar volatilities for the second and third period effects. The first period effect is relatively more volatile for the Scotland data. The Scotland data has a higher level of mortality forecasts than the England and Wales, given the underlying models. The Scottish projections have more uncertainty through the

				$\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$	$\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$
Posterior	Mean	England and Wales	A	5.08E-04	9.01E-07	2.12E-09
			B	4.26E-04	7.47E-07	1.64E-09
			C	4.75E-04	8.48E-07	1.77E-09
		Scotland	A	7.61E-04	1.23E-06	3.35E-09
			B	6.61E-04	9.94E-07	2.41E-09
	Std Dev	England and Wales	A	1.07E-04	2.01E-07	5.11E-10
			B	9.09E-05	1.62E-07	4.50E-10
			C	1.23E-04	2.35E-07	5.54E-10
		Scotland	A	1.76E-04	3.58E-07	1.14E-09
			B	1.61E-04	3.11E-07	9.14E-10

Table 7.3: The mean and the standard deviation of the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,1)$, $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$, and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ for England and Wales and Scotland data, during year 1971-2011, aged 50-89 last birthday. Note A = Fixing the mode of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the point estimate of England and Wales data $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$; B = Fixing the mean of the prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ to the poinrt estimate of England and Wales data $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$; C = Jeffrey's Prior for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for modelling England and Wales data.

forecasting horizon than the England and Wales data.

Chapter 8

Summary

Stochastic mortality models are widely used as risk management tools in the insurance and pensions industry with the main application being the generation of plausible scenarios for future mortality rates. When new models have been developed the objective was mostly to improve the goodness of fit of the model to mortality data observed in relatively large populations. In contrast, actuaries often face the problem of modelling the mortality experience of much smaller populations, for example, the members of a pension scheme.

This thesis covers our research about modelling small populations during my PhD period with my supervisors. Our research could benefit the industries (e.g. life insurance, pension scheme) and academics that often model the mortality data of smaller populations.

As use of the two-stage approach is widespread (perhaps because of its relative simplicity) we have, in the first stage of our project, attempted the first systematic analysis of the impact of population size on parameter estimates and forecasts using the two-stage approach. In this way, users of the two-stage approach will be better informed about its limitations as well as understanding about how the likelihood ratio test might be used to exploit data from large populations. To summarise

the main findings, we find that the size of a population has a significant effect on the uncertainty about the estimated parameters and mortality projections. In particular, we found that there exists a bias in the estimated co-variance matrix of the random walk fitted to the period effects when the size of the underlying population is small. As a consequence, prediction intervals are rather wide for small populations even when parameter uncertainty is ignored.

In the following stage of our project, we have demonstrated to the users of the stochastic mortality models (e.g. manager of a small pension scheme) how the information of a larger population could be embedded for parameter estimation and forecasts performed with Bayesian modelling, to what extent the parameter estimation could be improved compared with the two-stage approach and the financial implication, such as annuity price and longevity risk, given the two modelling methodologies. The users are informed how the importance of the prior information takes over the parameter estimation of a much smaller population and in what way the sampling variation affects the parameter estimation and mortality forecasts. To summarise the main findings, we find that our Bayesian model and the methodology of using the information of large referencing population provide an improved estimation for the volatility of small populations. The projections based on the small populations are not "significantly" different from the "true" projections (based on the larger reference population). When the population is small, the prior distributions, in particular the time series prior for the latent parameters, dominate the likelihood.

For future research, we will apply the Bayesian method for modelling the county-based Japanese mortality data. We will pursue our research on developing a methodology with the advantages of both the two-stage and Bayesian methods for modelling the mortality experience of small populations. Further research will be carried out for the Irish health and critical illness data.

Appendix A

Notations Introduced in Each Chapter

A.1 Chapter 2

- $\boldsymbol{t} = (t_1, \ldots, t_{n_y})$, vector of calendar years in the underlying data with length n_y.

- $\boldsymbol{x} = (x_1, \ldots, x_{n_a})$, vector of ages in the underlying data with length n_a.

- $D(t, x)$, $E(t, x)$, the number of death counts and the respective central exposure to risk respectively, during calendar year $t = t_1, \ldots, t_{n_y}$ at age $x = x_1, \ldots, x_{n_a}$.

- $m(t, x)$, the crude death rate at age x during calendar year t.

- $q(t, x)$, the mortality rate which is the probability that a person aged exactly x will die during year t.

- $\boldsymbol{E} = \{E(t, x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, the $(t_{n_y} - t_1 + 1) \times (x_{n_a} - x_1 + 1)$ dimensioned matrix of exposure.

- $\boldsymbol{D} = \{D(t, x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, the corresponding matrix of death counts.

- $\kappa_t^{(i)}$, period effect in year $t = t_1, \ldots, t_{n_y}$ for each $i = 1, 2, 3$.

- $\boldsymbol{\kappa}^{(i)} = (\kappa_{t_1}^{(i)}, \ldots, \kappa_{t_{n_y}}^{(i)})$ for $i = 1, 2, 3$, is the vector of period parameters.

- $\boldsymbol{\kappa} = (\boldsymbol{\kappa}^{(1)}, \boldsymbol{\kappa}^{(2)}, \boldsymbol{\kappa}^{(3)})$.

- $\gamma_c^{(4)}$ is the cohort effect for the cohort born in year $c = t - x$.

- $\boldsymbol{\gamma}^{(4)} = (\gamma_{t_1 - x_{n_a}}^{(4)}, \ldots, \gamma_{t_{n_y} - x_1}^{(4)})$ is the vector of cohort parameters.

- $\bar{x} = \frac{1}{n_a} \sum_{x=x_1}^{x=x_{n_a}} x$, the mean of the age range.

- $\hat{\sigma}_x^2$ is the mean of $(x - \bar{x})^2$.

- $\theta_{1,t,x} = (\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_c^{(4)})$, the vector includes the period effects and cohort effects at year t and year of birth $c = t - x$ respectively.

- $\theta_1 = \theta_{1,t,x}$. For convenience purpose, we let $\theta_1 = \theta_{1,t,x}$ and these two notations are used inter-changeably.

- $m(\theta_1, t, x)$, $q(\theta_1, t, x)$, the death rate and mortality rate given the parameter vector θ_1 at age x year t.

- $\boldsymbol{q}(\boldsymbol{\theta}_1) = \{q(\theta_1, t, x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, the matrix of mortality rates.

- $\boldsymbol{\theta}_1 = (\boldsymbol{\kappa}^{(1)}, \boldsymbol{\kappa}^{(2)}, \boldsymbol{\kappa}^{(3)}, \boldsymbol{\gamma}^{(4)})$, the vector includes all the latent parameters with length $4n_y + n_a - 1$.

- $\boldsymbol{\theta}_{11} = \boldsymbol{\kappa}_{t=t_1} = (\kappa_{t_1}^{(1)}, \kappa_{t_1}^{(2)}, \kappa_{t_1}^{(3)})$, the vector of three period effects at year t_1.

- $\boldsymbol{\theta}_{12} = (\boldsymbol{\kappa}_{t_2}, \ldots, \boldsymbol{\kappa}_{t_{n_y}})$, where $\boldsymbol{\kappa}_t = (\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)})$ for $t = t_2, \ldots, t_{n_y}$.

- $\boldsymbol{\theta}_{13} = \gamma_{t_1 - x_{n_a}}^{(4)}$, the first cohort year in the data.

- $\boldsymbol{\theta}_{14} = (\gamma_{t_1 - x_{n_a}+1}^{(4)}, \ldots, \gamma_{t_{n_y} - x_1}^{(4)})$.

- $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}, \alpha_\gamma, \sigma_\gamma^2)$, the vector includes all the hyper-parameters.

- $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, the vector of complete parameters that consists of the sub-vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ including all the latent and hyper parameters respectively.

A.2 Chapter 3

- $\hat{\boldsymbol{\theta}}_1$, the maximum likelihood estimator (point estimator) of $\boldsymbol{\theta}_1$.

- $\boldsymbol{\theta}_{1,0} = \{\theta_{1,0}(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, a given parameter vector with length $4n_y + n_a - 1$, where $\theta_{1,0} = (\kappa_{t,0}^{(1)}, \kappa_{t,0}^{(2)}, \kappa_{t,0}^{(3)}, \gamma_{c,0}^{(4)})$ is a given parameter vector for simulating $D(t,x)$.

- $\theta_{1,0} = \theta_{1,0}(t,x)$. For convenience purpose, the notations $\theta_{1,0}$ and $\theta_{1,0}(t,x)$ are used inter-changeable.

- $\hat{\boldsymbol{\theta}}_1^{\text{EW}} = \{\hat{\theta}_1^{\text{EW}}\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, the estimated parameter vector $\boldsymbol{\theta}_1$ for the England and Wales data, where $\hat{\theta}_1^{\text{EW}}$ is the estimated parameter vector θ_1 for the England and Wales data in year t aged x.

- $\boldsymbol{E}_0 = \{E_0(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, the benchmark exposure, where $E_0(t,x)$ is the benchmark exposure in year t aged x.

- $w = 1, 0.1, 0.01, 0.001$, the weights by which we scale down the benchmark exposure.

- $\boldsymbol{E}^w = \{E^w(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, the exposure matrix for the weight w, where $E^w(t,x) = wE_0(t,x)$ is the exposure for the weight w in year t aged x.

- N_1, the total number of death scenarios we simulated.

- $\boldsymbol{D}^w = \{D^w(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$, the death matrix for the weight w, where $D^w(t,x)$ is the death counts for the weight w in year t aged x.

- $\boldsymbol{D}_j^w = \{D_j^w(t,x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$ for $j = 1,\ldots,N_1$, the j^{th} independently simulated death matrix for $\boldsymbol{D}^w$, where $D_j^w(t,x)$ is the j^{th} independently simulated death counts for $D^w(t,x)$ in year t aged x.

- $m(\theta_{1,0}, t, x)$, the death rate conditional on the given parameter vector $\theta_{1,0}$ in year t aged x.

- $\hat{\kappa}_{t,j}^{(i),w}$, the MLE for the i^{th} period effect $\kappa_t^{(i)}$ for the j^{th} simulated death scenario with weight $w = 1, 0.1, 0.01, 0.001$, where $i = 1, 2, 3$ and $j = 1, \ldots, N_1$.

- $\hat{\kappa}_t^{(i),w} = \{\hat{\kappa}_{t,j}^{(i),w}\}_{j=1,\ldots,N_1}$.

- $\hat{\gamma}_{c,j}^{(4),w}$, the MLE for the cohort effect $\gamma_c^{(4)}$ for the j^{th} simulated death scenario with weight $w = 1, 0.1, 0.01, 0.001$, where $c = t_1 - x_{n_a}, \ldots, t_{n_y} - x_1$ and $j = 1, \ldots, N_1$.

- $\hat{\gamma}_c^{(4),w} = \{\hat{\gamma}_{c,j}^{(4),w}\}_{j=1,\ldots,N_1}$.

- $\hat{\boldsymbol{\theta}}_{1,j}^{w} = (\hat{\kappa}_{t_1,j}^{(1),w}, \ldots, \hat{\kappa}_{t_{n_y},j}^{(1),w}, \hat{\kappa}_{t_1,j}^{(2),w}, \ldots, \hat{\kappa}_{t_{n_y},j}^{(2),w}, \hat{\kappa}_{t_1,j}^{(3),w}, \ldots, \hat{\kappa}_{t_{n_y},j}^{(3),w}, \hat{\gamma}_{t_1-x_{n_a},j}^{(4),w}, \ldots, \hat{\gamma}_{t_{n_y}-x_1,j}^{(4),w})$, the estimated parameter vector of MLEs for each simulated death scenario $j = 1, \ldots, N_1$ and each population size $w = 1, 0.1, 0.01, 0.001$.

- $\hat{\boldsymbol{\theta}}_1^{w} = \{\hat{\boldsymbol{\theta}}_{1,j}^{w}\}_{j=1,\ldots,N_1}$, where $w = 1, 0.1, 0.01, 0.001$.

- Δ^k, the k^{th} order difference.

- $\Delta\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1}$, where $\boldsymbol{\kappa}_t = (\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)})$.

- $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$, vector of the random walk drift with length 3.

- $\boldsymbol{V}_{\boldsymbol{\epsilon}} = \{V_\epsilon(i,k)\}_{i=1,2,3}^{k=1,2,3}$, the 3×3 co-variance matrix of the random walk drift independent of t.

- $\boldsymbol{\epsilon}_t \sim MVN(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\epsilon}})$, i.i.d multi-variate normal error term of the random walk drift.

- $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)$, the point estimator for $\boldsymbol{\mu}$.

- $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$, the posterior distribution for $\boldsymbol{\mu}$.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}} = \{\hat{V}_\epsilon(i,k)\}_{i=1,2,3}^{k=1,2,3}$, the point estimator for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$.

- $\tilde{\boldsymbol{V}}_{\epsilon} = \{\tilde{V}_\epsilon(i,k)\}_{i=1,2,3}^{k=1,2,3}$, the posterior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$.

- $\hat{\boldsymbol{\mu}}_j^{w} = (\hat{\mu}_{1,j}^{w}, \hat{\mu}_{2,j}^{w}, \hat{\mu}_{3,j}^{w})$, the point estimator for the $\boldsymbol{\mu}$ for the simulated death scenario $j = 1, \ldots, N_1$ with weight $w = 1, 0.1, 0.01, 0.001$.

- $\hat{\boldsymbol{\mu}}^{w} = \{\hat{\boldsymbol{\mu}}_j^{w}\}_{j=1,\ldots,N_1}$.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon},j}^{\ w} = \{\hat{V}_{\epsilon,j}^{w}(i,k)\}_{i=1,2,3}^{k=1,2,3}$, the point estimator for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the simulated death scenario $j = 1, \ldots, N_1$ with weight $w = 1, 0.1, 0.01, 0.001$.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ w} = \{\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon},j}^{\ w}\}_{j=1,\ldots,N_1}$.

- $\Delta\hat{\kappa}_{t,j}^{(i),w} = \hat{\kappa}_{t,j}^{(i),w} - \hat{\kappa}_{t-1,j}^{(i),w}$.

- $\tilde{\boldsymbol{V}}_{\boldsymbol{\epsilon},j}^{w}$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the death scenario $j = 1, \ldots, N_1$ with weight $w = 1, 0.1, 0.01, 0.001$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{w} = \{\tilde{\boldsymbol{V}}_{\epsilon,j}^{w}\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{\mu}}_{j}^{w}$, the posterior distribution of $\boldsymbol{\mu}$ for the death scenario $j = 1, \ldots, N_1$ with weight $w = 1, 0.1, 0.01, 0.001$.

- $\tilde{\boldsymbol{\mu}}^{w} = \{\tilde{\boldsymbol{\mu}}_{j}^{w}\}_{j=1,\ldots,N_1}$.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ \mathrm{EW}}$, the point estimator for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales data.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}$, the posterior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales.

- α_0, the mean reverting level of the AR(1) model for the cohort effect $\boldsymbol{\gamma}^{(4)}$.

- α_γ, the coefficient of the AR(1) model for the cohort effect.

- ϵ_c, the i.i.d error term of the AR(1) model independent of cohort year c.

- $\hat{\alpha}_0$, the point estimator for α_0.

- $\hat{\alpha}_0^w$, the point estimator for α_0 for the simulated death scenarios with weight $w = 1, 0.1, 0.01, 0.001$.

- $\hat{\alpha}_\gamma$, the point estimator for α_γ.

- $\hat{\alpha}_\gamma^w$, the point estimator for α_γ for the simulated death scenarios with weight w.

A.3 Chapter 4

- P_j^w, the p-value of a likelihood ratio test for the death scenario $j = 1, \ldots, N_1$.

- $P^w = \{P_j^w\}_{j=1,\ldots,N_1}$.

- Γ_j^w, the test statistic of a likelihood ratio test for the death scenario $j = 1, \ldots, N_1$.

- $\Gamma^w = \{\Gamma_j^w\}_{j=1,\ldots,N_1}$.

- $n_c = n_y + n_a - 1$, the total number of cohorts in a given dataset.

- α, the degree of freedom of a likelihood ratio test.

- $\theta_1^{(i)}(\lambda, t, x)$, a parameter vector with the i^{th} latent parameter shifted or scaled by λ in year t aged x. For convenient purpose, we use $\theta_1^{(i)}(\lambda)$ and $\theta_1^{(i)}(\lambda, t, x)$ inter-changeably.

- $\boldsymbol{\theta}_1^{(i)}(\lambda) = \{\theta_1^{(i)}(\lambda, t, x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$ for $i = 1, 2, 3, 4$.

- $\boldsymbol{\kappa}_0^{(i)} = \{\kappa_{t,0}^{(i)}\}_{t=t_1,\ldots,t_{n_y}}$, a given parameter vector for $\boldsymbol{\kappa}^{(i)}$ for $i = 1, 2, 3$ where $\kappa_{t,0}^{(i)}$ is defined in Appendix A.2.

- $\boldsymbol{\gamma}_0^{(4)} = \{\gamma_{c,0}^{(4)}\}_{c=t_1-x_{n_a},\ldots,t_{n_y}-x_1}$, a given parameter vector for $\boldsymbol{\gamma}^{(4)}$, where $\gamma_{c,0}^{(4)}$ is defined in Appendix A.2.

- $D_j^{w,(i)}(t, x)$, the j^{th} simulated death scenario with weight $w = 1, 0.1, 0.01$, conditional on $\theta_1^{(i)}(\lambda, t, x)$ in year t age x for $j = 1, \ldots, N_1$, $i = 1, 2, 3, 4$.

- $\boldsymbol{D}_j^{w,(i)} = \{D_j^{w,(i)}(t, x)\}_{t=t_1,\ldots,t_{n_y}}^{x=x_1,\ldots,x_{n_a}}$ for $j = 1, \ldots, N_1$.

- $\hat{\boldsymbol{\theta}}_{1,j}^{w,(i)}$, the relative MLE of $\boldsymbol{\theta}_1$ for $\boldsymbol{D}_j^{w,(i)}$.

- $P_j^{w,(i)} = P_j^{w,(i)}(\lambda)$, the p-value of the power test for death scenario $j = 1, \ldots, N_1$, $i = 1, 2, 3, 4$ and $w = 1, 0.1, 0.01$.

- $R^{w,(i)}(\lambda)$, the proportion of scenarios in which the null hypothesis is rejected among N_1 simulated scenarios for $i = 1, 2, 3, 4$ and $w = 1, 0.1, 0.01$.

- $p^{w,(i)}(\lambda)$, the (unknown) power of the LR test if alternative $\boldsymbol{\theta}_1^{(i)}(\lambda)$ with parameter λ is the true parameter set for the simulated death counts.

- $\lambda_{0.5}^{w,(i)}$, the value of λ that results in a power of 50% of the LR test for $i = 1, 2, 3, 4$ and $w = 1, 0.1, 0.01$.

- $\rho_{t,x}^{w,(i)}$, the ratio of death rate in year t aged x conditional on $\lambda_{0.5}^{w,(i)}$ to the relative death rate given $\theta_{1,0}$ for $i = 1, 2, 3, 4$ and $w = 1, 0.1, 0.01$.

- $S(T + t, x)$, the survival index for the probability of an individual aged x exactly at the start of year T, that will survive for the next t years.

- $v = (1 + i)^{-1}$, the discount factor.

- $\hat{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$, the vector of point estimator for $\boldsymbol{\gamma}^{(4)}$ for the England and Wales males' data.

- $\hat{\boldsymbol{\gamma}}^{(4),M}$, the vector of point estimator for $\boldsymbol{\gamma}^{(4)}$ for the England and Wales males' data.

- $\hat{\boldsymbol{\gamma}}^{(4),F}$, the vector of point estimator for $\boldsymbol{\gamma}^{(4)}$ for the England and Wales females' data.

- $\hat{\boldsymbol{\gamma}}^{(4),\mathrm{SL}}$, the vector of point estimator for $\boldsymbol{\gamma}^{(4)}$ for the Scotland males' data.

A.4 Chapter 5

- $p(\cdot|\cdot)$, a conditional density with its argument determined by the following context and similarly as $p(\cdot)$ the marginal density.
- $x = x_1, \ldots, x_{n_a}$, defined in Appendix A.1.
- $t = t_1, \ldots, t_{n_y}$, defined in Appendix A.1.
- $\boldsymbol{D}$, defined in Appendix A.1.
- $\boldsymbol{E}$, defined in Appendix A.1.
- $\boldsymbol{\theta}$, defined in Appendix A.1.
- $\tilde{\boldsymbol{\theta}}$, the posterior distribution for $\boldsymbol{\theta}$.
- $\tilde{\boldsymbol{\theta}}^{(k)}$, the k^{th} sample drawn from the posterior distribution for $\boldsymbol{\theta}$, where $k = 1, \ldots, N_2$.
- $\hat{\boldsymbol{\theta}}^{\text{EW}}$, the vector of MLEs for $\boldsymbol{\theta}$ for the England and Wales data.
- $\hat{\boldsymbol{\theta}}_1^{\text{EW}}$, the vector of MLEs for $\boldsymbol{\theta}_1$ for the England and Wales data.
- $\boldsymbol{\theta}_1$, defined in Appendix A.1.
- $\tilde{\boldsymbol{\theta}}_1^{(k)}$, the k^{th} posterior sample for $\boldsymbol{\theta}_1$.
- $\tilde{\boldsymbol{\theta}}_1'$, the projected $\tilde{\boldsymbol{\theta}}_1$.
- $\tilde{\boldsymbol{\theta}}_1^{'(k,l)}$, the projection for $\tilde{\boldsymbol{\theta}}_1^{(k)}$ for the sample path $l = 1, \ldots, N_4$.
- $\tilde{\boldsymbol{\theta}}_1^{'\text{EW}}$, the projected $\tilde{\boldsymbol{\theta}}_1^{\text{EW}}$ for the England and Wales data.
- $\hat{\boldsymbol{\theta}}_1^{'l}$ for $l = 1, \ldots, N_4$, the l^{th} sample path for the projected $\boldsymbol{\theta}_1$ conditional on $\hat{\boldsymbol{\theta}}_1$.
- $\hat{\boldsymbol{\theta}}_1' = \{\hat{\boldsymbol{\theta}}_1^{'l}\}_{l=1,\ldots,N_4}$.
- $\hat{\boldsymbol{\theta}}_1^{'\text{EW}}$, the projected $\hat{\boldsymbol{\theta}}_1^{\text{EW}}$ for the England and Wales data.

- $\hat{\boldsymbol{\theta}}_1^w$, defined in Appendix A.2.

- θ_1, defined in Appendix A.1.

- $\hat{\theta}_1^{\text{EW}}$, the vector of MLEs for θ_1 for the England and Wales data.

- $\tilde{\theta}_1^{\text{EW}}$, the posterior distribution for θ_1 for the England and Wales data.

- $\theta_{1,0}$, defined in Appendix A.2.

- $\boldsymbol{\theta}_{1,0}$, defined in Appendix A.2.

- $\boldsymbol{\theta}_{11}$, defined in Appendix A.1.

- $\boldsymbol{\theta}_{12}$, defined in Appendix A.1.

- $\boldsymbol{\theta}_{13}$, defined in Appendix A.1.

- $\boldsymbol{\theta}_{14}$, defined in Appendix A.1.

- $D(t,x)$, defined in Appendix A.1.

- $D_j^w(t,x)$, defined in Appendix A.2.

- $D^w(t,x) = \{D_j^w(t,x)\}_{j=1,\ldots,N_1}$.

- N_1, the total number of simulated death scenario.

- N_2, the number of samples drawn from the joint posterior distribution for each parameter for each death scenario.

- N_3, the total number of iterations we run for the MCMC.

- N_4, the number of sample projecting paths generated for the projection, starting from $\tilde{\boldsymbol{\kappa}}_{t_{n_y}}^{(k)}$ and $\tilde{\gamma}_{t_{n_y}-x_1}^{(4,k)}$ for $k = 1, \ldots, N_2$.

- $m(\theta_1, t, x)$, $q(\theta_1, t, x)$, defined in Appendix A.1.

- $m(\theta_{1,0}, t, x)$, $q(\theta_{1,0}, t, x)$, defined in Appendix A.2.

- $m(\hat{\theta}_1^{\text{EW}}, t, x)$, the fitted death rate conditional on $\hat{\theta}_1^{\text{EW}}$ at year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1 \ldots, x_{n_a}$ for the England and Wales data.

- $m(\tilde{\theta}_1^{\text{EW}}, t, x)$, the fitted death rate conditional on $\tilde{\theta}_1^{\text{EW}}$ at year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1 \ldots, x_{n_a}$ for the England and Wales data.

- $\hat{m}(t, x)$, either the fitted death rates given the MLE, $m(\hat{\theta}_1^{\text{EW}}, t, x)$, or the mean of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$.

- $m(\tilde{\theta}_1^{(k),\text{EW}}, t, x)$, the k^{th} observation of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$ for the England and Wale, where $k = 1, \ldots, N_2$.

- $\boldsymbol{m}$, the matrix of empirical death rates.

- $\boldsymbol{m}'$, the future unknown but observable death rates.

- $\boldsymbol{m}'^{(k,l)}|\tilde{\boldsymbol{\theta}}_1'^{(k,l)}$, the projected death rates conditional on $\tilde{\boldsymbol{\theta}}_1^{(k,l)}$.

- $\boldsymbol{m}'|\tilde{\boldsymbol{\theta}}_1' = \{\boldsymbol{m}'^{(k,l)}|\tilde{\boldsymbol{\theta}}_1'^{(k,l)}\}_{k=1,\ldots,N_2}^{l=1,\ldots,N_4}$.

- $\boldsymbol{m}'|\hat{\boldsymbol{\theta}}_1'$, projected death rates conditional on $\hat{\boldsymbol{\theta}}_1$.

- $r(t, x)$, the ratio of $m(\hat{\theta}_1^{\text{EW}}, t, x)$ to the mean of $m(\tilde{\theta}_1^{\text{EW}}, t, x)$.

- $z(t, x)$, the standardised residual in year $t = t_1, \ldots, t_{n_y}$ and age $x = x_1 \ldots, x_{n_a}$.

- $E(t, x)$, defined in Appendix A.1.

- $E_0(t, x)$, defined in Appendix A.2.

- $\gamma_c^{(4)}$ for $c = t - x$, defined in Appendix A.1.

- $\hat{\gamma}_c^{(4)}$, the MLE for $\gamma_c^{(4)}$.

- $\boldsymbol{\gamma}^{(4)}$, defined in Appendix A.1. For convenient purpose, we use $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^{(4)}$ interchangeably.

- $\tilde{\boldsymbol{\gamma}}^{(4)}$, the posterior distribution for the $\boldsymbol{\gamma}^{(4)}$.

- $\tilde{\boldsymbol{\gamma}}^{(4,k)}$ for $k = 1, \ldots, N_2$, the k^{th} posterior sample for $\boldsymbol{\gamma}^{(4)}$.

- $\hat{\boldsymbol{\gamma}}^{(4)}$, the MLE of $\boldsymbol{\gamma}^{(4)}$.

- $\hat{\boldsymbol{\gamma}}'^{(4)}$, the projected $\boldsymbol{\gamma}^{(4)}$ conditional on the $\hat{\boldsymbol{\gamma}}^{(4)}$.

- $\tilde{\gamma}^{(4,k)}_{t_{n_y}-x_1+n'}$ for $k = 1, \ldots, N_2$, the projected $\boldsymbol{\gamma}$ conditional on $\tilde{\boldsymbol{\gamma}}^{(4,k)}$ at year $t_{n_y} - x_1 + n'$.

- $\tilde{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$, the posterior distribution of $\boldsymbol{\gamma}^{(4)}$ for the England and Wales data.

- $\hat{\boldsymbol{\gamma}}^{(4),\mathrm{EW}}$, the MLE of $\boldsymbol{\gamma}^{(4)}$ for the England and Wales data.

- $\bar{x}$, defined in Appendix A.1.

- $\hat{\sigma}^w_x$, defined in Appendix A.1.

- $\kappa^{(i)}_t$ for $t = t_1, \ldots, t_{n_y}$ and $i = 1, 2, 3$, defined in Appendix A.1.

- $\boldsymbol{\kappa}^{(i)}$ for $i = 1, 2, 3$, defined in Appendix A.1.

- $\hat{\boldsymbol{\kappa}}^{(i),\mathrm{EW}}$ for $i = 1, 2, 3$, the MLE of $\boldsymbol{\kappa}^{(i)}$ for the England and Wales data.

- $\tilde{\boldsymbol{\kappa}}^{(i),\mathrm{EW}}$ for $i = 1, 2, 3$, the posterior distribution of $\boldsymbol{\kappa}^{(i)}$ for the England and Wales data.

- $\boldsymbol{\kappa}_t$ for $t = t_1, \ldots, t_{n_y}$, defined in Appendix A.1.

- $\boldsymbol{\kappa}$, defined in Appendix A.1.

- $\tilde{\boldsymbol{\kappa}}^{(k)}$, the k^{th} sample drawn from the posterior distribution for $\boldsymbol{\kappa}$, where $k = 1, \ldots, N_2$.

- $\tilde{\boldsymbol{\kappa}}^{(k)}_{t_{n_y}+n'}$ for $k = 1, \ldots, N_2$, the projected $\boldsymbol{\kappa}$ conditional on $\tilde{\boldsymbol{\kappa}}^{(k)}$ at year $t_{n_y} + n'$.

- $\boldsymbol{\mu}$, defined in Appendix A.2.

- $\boldsymbol{\mu}_i$, the i^{th} element of $\boldsymbol{\mu}$ for $i = 1, 2, 3$.

- $\tilde{\boldsymbol{\mu}}^{\mathrm{th}}$ for $k = 1, \ldots, N_2$, the k^{th} posterior sample for $\boldsymbol{\mu}$.

- $\tilde{\boldsymbol{\mu}}_i$, the posterior distribution of $\boldsymbol{\mu}_i$ for $i = 1, 2, 3$.

- $\tilde{\boldsymbol{\mu}}_i^{\text{EW}}$, the posterior distribution of $\boldsymbol{\mu}_i$ for the England and Wales data, where $i = 1, 2, 3$.

- $\hat{\boldsymbol{\mu}}_i^{\text{EW}}$, the MLE of $\boldsymbol{\mu}_i$ for the England and Wales data, where $i = 1, 2, 3$.

- $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, defined in Appendix A.2.

- $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$, the element on the i^{th} row and k^{th} column of the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for $i, k = 1, 2, 3$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{(k)}$ for $k = 1, \ldots, N_2$, the k^{th} posterior sample for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$, the matrix of the MLEs for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales data.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}(i, k)$, the MLE for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$ for the England and Wales data, where $i, k = 1, 2, 3$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}$, the posterior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales data.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(i, k)$, the posterior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$ for the England and Wales data, where $i, k = 1, 2, 3$.

- ν, the degree of freedom of the informative Inv-Wishart prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$.

- Σ, the scale matrix of the informative Inv-Wishart prior distribution for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$.

- ϵ_t, defined in Appendix A.2.

- α_γ, defined in Appendix A.2.

- $\tilde{\alpha}_\gamma^{(k)}$ for $k = 1, \ldots, N_2$, the k^{th} posterior sample for α_γ.

- $\tilde{\alpha}_\gamma^{\text{EW}}$, the posterior distribution of α_γ for the England and Wales data.

- $\hat{\alpha}_\gamma^*$, the MLE of α_γ, given five short cohorts equally removed from the beginning and the end of $\hat{\boldsymbol{\gamma}}^{(4)}$.

- $\hat{\alpha}_\gamma^{\text{EW}}$, the MLE of α_γ for the England and Wales data.

- g, a constant for the prior distribution for α_γ.

- ϵ_c, defined in Appendix A.2.

- σ_γ^2, the variance of ϵ_c.

- $\tilde{\sigma}_\gamma^{(k)}$ for $k = 1, \ldots, N_2$, the k^{th} posterior sample for σ_γ.

- $\tilde{\sigma}_\gamma^{\text{EW}}$, the posterior distribution of σ_γ for the England and Wales data.

- $\hat{\sigma}_\gamma^*$, the MLE of σ_γ, given five short cohorts equally removed from the beginning and the end of $\hat{\boldsymbol{\gamma}}^{(4)}$.

- $\hat{\sigma}_\gamma^{\text{EW}}$, the MLE of σ_γ for the England and Wales data.

- a_γ, the shape parameter of the prior distribution for σ_γ.

- b_γ, the scale parameter of the prior distribution for σ_γ.

- $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}, \alpha_\gamma, \sigma_\gamma^2)$, the parameter vector includes all the hyper-parameters.

- $\boldsymbol{\Delta}_t = \boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1}$ for $t = t_2, \ldots, t_{n_y}$, an i.i.d random variable independent of t such that $\boldsymbol{\Delta}_t | \boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}} \sim MVN(\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}})$.

- $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{t_2}, \ldots, \boldsymbol{\Delta}_{t_{n_y}})$.

- w, defined in Appendix A.2.

- $S_0 = \sum_{t=t_2}^{t_{n_y}} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})(\boldsymbol{\Delta}_t - \boldsymbol{\mu})^T$.

- $n' = 1, \ldots, N'$, the number of years projected.

A.5 Chapter 6

- $x = x_1, \ldots, x_{n_a}$, defined in Appendix A.1.

- $t = t_1, \ldots, t_{n_y}$, defined in Appendix A.1.

- t', the projected year.

- c', the projected year of birth.

- $\boldsymbol{D}$, defined in Appendix A.1.

- $\boldsymbol{D}_j^w$, the matrix for the j^{th} simulated death scenario, where $j = 1, \ldots, N_1$.

- $D_j^w(t, x)$, defined in Appendix A.2.

- $D^w(t, x) = \{D_j^w(t, x)\}_{j=1,\ldots,N_1}$, also denoted as $D_{t,x}^w$.

- $\boldsymbol{E}$, defined in Appendix A.1.

- $E_0(t, x)$, defined in Appendix A.2.

- $\boldsymbol{\theta}$, defined in Appendix A.1.

- $\boldsymbol{\theta}_0$, a given parameter vector for the full parameter $\boldsymbol{\theta}$.

- $\hat{\boldsymbol{\theta}}^{\text{EW}}$, defined in Appendix A.4.

- $\tilde{\boldsymbol{\theta}}_j^{w,(k)}$, the k^{th} posterior sample of the full parameter vector $\boldsymbol{\theta}$ for the simulated death scenario j, where $j = 1, \ldots, N_1$ and $k = 1, \ldots, N_2$.

- $\tilde{\boldsymbol{\theta}}^w = \{\tilde{\boldsymbol{\theta}}_j^{w,(k)}\}_{j=1,\ldots,N_1}^{k=1,\ldots,N_2}$

- $\bar{\tilde{\boldsymbol{\theta}}}_j^w = \frac{1}{N_2} \sum_{k=1}^{N_2} \tilde{\boldsymbol{\theta}}_j^{w,(k)}$, the mean of the posterior samples for the death scenario $j = 1, \ldots, N_1$.

- $\tilde{\boldsymbol{\theta}}_{j'}^w$ for $j' = j_1, \ldots, j_{100}$, the posterior distribution for the j'^{th} randomly selected death scenario.

- $\boldsymbol{\theta}_1$, defined in Appendix A.1.

- $\tilde{\boldsymbol{\theta}}_1$, the posterior distribution of $\boldsymbol{\theta}_1$.

- $\tilde{\boldsymbol{\theta}}_1^{w,j}$, the posterior distribution of $\boldsymbol{\theta}_1$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\tilde{\boldsymbol{\theta}}_1^{w} = \{\tilde{\boldsymbol{\theta}}_1^{w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{\theta}}_1^{w,j,(k)}$, the k^{th} posterior sample for $\tilde{\boldsymbol{\theta}}_1^{w,j}$, where $k = 1, \ldots, N_2$.

- $\tilde{\boldsymbol{\theta}}_1'$, the posterior distribution of the projected $\boldsymbol{\theta}_1$ conditional on $\tilde{\boldsymbol{\theta}}_1$.

- $\tilde{\boldsymbol{\theta}}_1^{'\,w,j}$, posterior distribution of the projected $\boldsymbol{\theta}_1$ conditional on $\tilde{\boldsymbol{\theta}}_1^{w,j}$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\tilde{\boldsymbol{\theta}}_1^{'\,w} = \{\tilde{\boldsymbol{\theta}}_1^{'\,w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{\theta}}_1^{'\,w,j,k} = \{\tilde{\boldsymbol{\theta}}_1^{'\,w,j,(k,l)}\}_{l=1,\ldots,N_4}$, the projection for the k^{th} sample drawn for the $\tilde{\boldsymbol{\theta}}_1^{w,j}$ of the j^{th} death scenario, where $k = 1, \ldots, N_2$.

- $\tilde{\boldsymbol{\theta}}_1^{'\,w,j,(k,l)}$, the l^{th} sample path of the projection for the k^{th} sample of the j^{th} death scenario, where $l = 1, \ldots, N_4$.

- $\tilde{\theta}_1^{w,j} = \tilde{\theta}_1^{w,j}(t,x) = (\tilde{\kappa}_{t,j}^{(1),w}, \tilde{\kappa}_{t,j}^{(2),w}, \tilde{\kappa}_{t,j}^{(3),w}, \tilde{\gamma}_c^{(4),w,j})$, the posterior distribution of θ_1 for the j^{th} simulated death scenario with weight w for $j = 1, \ldots, N_1$, in year $t = t_1, \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$.

- $\tilde{\theta}_1^{w,j,(k)}$, the k^{th} posterior sample for $\tilde{\theta}_1^{w,j}$, where $k = 1, \ldots, N_2$.

- $\tilde{\theta}_1^{w} = \{\tilde{\theta}_1^{w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\theta}_1^{w,j'}$, the posterior distribution for θ_1 for the j'^{th} simulated death scenario with weight w for $j' = j_1, \ldots, j_{100}$, in year $t = t_1, \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$.

- $\tilde{\theta}_1^{w,j',(k)}$, the k^{th} posterior sample for the $\tilde{\theta}_1^{w,j'}$.

- $\tilde{\theta}_1^{\text{EW}}$, the posterior distribution of θ_1 for the England and Wales data in year $t = t_1, \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$.

- $\tilde{\theta}_1'$, posterior distribution of the projected θ_1 conditional on $\tilde{\theta}_1$.

- $\tilde{\theta}_1'^{\,w,j}$, posterior distribution of the projected θ_1 conditional on $\tilde{\theta}_1^{w,j}$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\tilde{\theta}_1'^{\,w} = \{\tilde{\theta}_1'^{\,w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\theta}_1'^{\,w,j,(k)} = \{\tilde{\theta}_1'^{\,w,j,(k,l)}\}_{l=1,\ldots,N_4}$, the projection for the k^{th} sample drawn for the $\tilde{\theta}_1^{w,j}$ of the j^{th} death scenario, where $k = 1, \ldots, N_2$.

- $\tilde{\theta}_1'^{\,w,j,(k,l)}$, the l^{th} sample path of the projection for the k^{th} sample of the j^{th} death scenario, where $l = 1, \ldots, N_4$.

- $\tilde{\theta}_1'^{\,,w,j^*}$, the posterior distribution of the projected θ_1 conditional on $\tilde{\theta}_1^{w,j^*}$ for the death scenario j^* with weight w.

- $\tilde{\theta}_1'^{\,\text{EW}}$, the posterior distribution of the projected θ_1 conditional on the $\tilde{\theta}_1^{\text{EW}}$ for the England and Wales data.

- $\theta_{1,0}$, defined in Appendix A.2.

- $\hat{\theta}_1^{w,j}$, the MLE of θ_1 for the j^{th} simulated death scenario in year $t = t_1, \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$, where $j = 1, \ldots, N_1$.

- $\hat{\theta}_1^{w} = \{\hat{\theta}_1^{w,j}\}_{j=1,\ldots,N_1}$.

- $\hat{\theta}_1^{w,j^*}$, the MLE of θ_1 for the death scenario j^* in year $t = t_1, \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$.

- $\hat{\theta}_1^{\text{EW}}$, the MLE of θ_1 for the England and Wales data in year $t = t_1, \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$.

- $m(\theta_{1,0}, t, x)$, $q(\theta_{1,0}, t, x)$, defined in Appendix A.2.

- $m(\tilde{\theta}_1^{w,j}, t, x)$, the posterior distribution of the fitted death rate for the j^{th} simulated death scenario, conditional on $\tilde{\theta}_1^{w,j}$ in year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$ with weight w for $j = 1, \ldots, N_1$.

- $m(\tilde{\theta}_1^{w,j,(k)}, t, x)$, the k^{th} posterior sample for $m(\tilde{\theta}_1^{w,j}, t, x)$, where $k = 1, \ldots, N_2$.

- $m(\tilde{\theta}_1^{w}, t, x) = \{m(\tilde{\theta}_1^{w,j}, t, x)\}_{j=1,\ldots,N_1}$.

- $m(\tilde{\theta}_1^{w,j^*}, t, x)$, the posterior distribution of the fitted death rate for the death scenario j^*, conditional on $\tilde{\theta}_1^{w,j^*}$ in year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$ with weight w.

- $m(\tilde{\theta}_1^{w,j^*,(k)}, t, x)$, the k^{th} posterior sample for $m(\tilde{\theta}_1^{w,j^*}, t, x)$, where $k = 1, \ldots, N_2$.

- $m(\tilde{\theta}_1^{w,j'}, t, x)$, the posterior distribution of the fitted death rate for the j'^{th} simulated death scenario, conditional on $\tilde{\theta}_1^{w,j'}$ in year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$ with weight w for $j' = j_1, \ldots, j_{100}$.

- $m'(\tilde{\theta}_1^{w}, t, x) = \{m(\tilde{\theta}_1^{w,j'}, t, x)\}_{j'=j_1,\ldots,j_{100}}$.

- $m(\tilde{\theta}_1^{w,j',(k)}, t, x)$, the k^{th} posterior sample for $m(\tilde{\theta}_1^{w,j',(k)}, t, x)$, where $k = 1, \ldots, N_2$.

- $m(\tilde{\theta}_1^{\text{EW}}, t, x)$, the posterior distribution of the fitted death rate for the England and Wales data, conditional on $\tilde{\theta}_1^{\text{EW}}$, in year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$.

- $\bar{m}(\tilde{\theta}_1^{w,j'}, t, x) = \frac{1}{N_2} \sum_{k=1}^{N_2} m(\tilde{\theta}_1^{w,j',(k)}, t, x)$, the mean of $m(\tilde{\theta}_1^{w,j'}, t, x)$ for each $j' = j_1, \ldots, j_{100}$.

- $m(\hat{\theta}_1^{w,j^*}, t, x)$, the fitted death rate conditional on $\hat{\theta}_1^{w,j^*}$ in year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$ with weight w.

- $m(\hat{\theta}_1^{\text{EW}}, t, x)$, the fitted death rate conditional on $\hat{\theta}_1^{\text{EW}}$ in year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$.

- $m'(\tilde{\theta}_1^{'\,w,j,(k,l)}, t', x)$, the l^{th} sample path of the projected death rate of the k^{th} posterior sample of the death scenario j with weight w, where $j = 1, \ldots, N_1$, $k = 1, \ldots, N_2$ and $l = 1, \ldots, N_4$.

- $m'(\tilde{\theta}_1^{'\,w,j}, t', x)$, the posterior distribution of the projected death rate in year t' aged x for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $m'(\tilde{\theta}_1^{'\,w,j^*}, t', x)$, the posterior distribution of the projected death rate in year t' aged x for the death scenario j^* with weight w.

- $m'(\tilde{\theta}_1^{'\mathrm{EW}}, t', x)$, the posterior distribution of the projected death rate in year t' aged x for the England and Wales data.

- $m'(\tilde{\theta}_1^{'\,w,j'}, t', x)$, the posterior distribution of the projected death rate in year t' aged x for the $j^{'\,\mathrm{th}}$ simulated death scenario with weight w, where $j' = j_1, \ldots, j_{100}$.

- $m'(\hat{\theta}_1^{'\,\mathrm{EW}}, t', x)$, the distribution of the projected death rate in year t' aged x, conditional on $\hat{\theta}_1^{'\,\mathrm{EW}}$ for the England and Wales data.

- $q(\tilde{\theta}_1^{w,j}, t, x)$, the posterior distribution of the fitted mortality rate for the j^{th} simulated death scenario, conditional on $\tilde{\theta}_1^{w,j}$ in year $t = t_1 \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$ with weight w for $j = 1, \ldots, N_1$.

- $q(\tilde{\theta}_1^{w,j,(k)}, t, x)$, the k^{th} posterior sample for $q(\tilde{\theta}_1^{w,j}, t, x)$, where $k = 1, \ldots, N_2$.

- $q(\tilde{\theta}_1^{w}, t, x) = \{q(\tilde{\theta}_1^{w,j}, t, x)\}_{j=1,\ldots,N_1}$.

- $q'(t', x)$, the projected mortality rate in year t' aged x.

- $q'(\tilde{\theta}_1^{'\,w,j}, t', x)$, the posterior distribution of the projected mortality rate in year t' aged x for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- IR, the central logit-mortality improvement rate conditional on the death $\boldsymbol{D}$.

- $IR_j(\tilde{\theta}_1^{'\,w,j})$, the MCMC central logit-mortality improvement rate for the death scenario j conditional on the death $\boldsymbol{D}_j^w$, where $j = 1, \ldots, N_1$.

- $IR(\tilde{\theta}_1^{'\,w}) = \{IR_j(\tilde{\theta}_1^{'\,w,j})\}_{j=1,\ldots,N_1}$.

- $IR(\tilde{\theta}_1^{'\,\mathrm{EW}})$, the MCMC central logit-mortality improvement rate conditional on $\tilde{\theta}_1^{'\,\mathrm{EW}}$ for the England and Wales data.

- $IR(\hat{\theta}_1^{'\,\mathrm{EW}})$, the central logit-mortality improvement rate conditional on $\hat{\theta}_1^{'\,\mathrm{EW}}$ for the England and Wales data.

- $a_{\cdot\%}$, the $(\cdot)$ quantile of the random variable a for the annuity price.

- $\mathrm{R}_{\cdot\%}$, the longevity risk at $\cdot\%$ level.

- $\boldsymbol{\kappa}$, defined in Appendix A.1.

- $\boldsymbol{\kappa}^{(i)}$ for $i = 1, 2, 3$, defined in Appendix A.1.

- $\tilde{\boldsymbol{\kappa}}^{(i)}$ for $i = 1, 2, 3$, the posterior distribution of $\boldsymbol{\kappa}^{(i)}$.

- $\tilde{\boldsymbol{\kappa}}^{(i),w}_{j^*}$ for $i = 1, 2, 3$, the posterior distribution of $\boldsymbol{\kappa}^{(i)}$ for the death scenario j^* with weight w.

- $\tilde{\boldsymbol{\kappa}}^{(i),\mathrm{EW}}$ for $i = 1, 2, 3$, the posterior distribution of $\boldsymbol{\kappa}^{(i)}$ for the England and Wales data.

- $\hat{\boldsymbol{\kappa}}^{(i),w}$ for $i = 1, 2, 3$, the MLE of $\boldsymbol{\kappa}^{(i)}$ for the simulated death scenarios with weight w.

- $\hat{\boldsymbol{\kappa}}^{(i),w}_{j^*}$ for $i = 1, 2, 3$, the MLE of $\boldsymbol{\kappa}^{(i)}$ for the death scenario j^* with weight w.

- $\hat{\boldsymbol{\kappa}}^{(i),\mathrm{EW}}$ for $i = 1, 2, 3$, the MLE of $\boldsymbol{\kappa}^{(i)}$ for the England and Wales data.

- $\kappa^{(i)}_t$, defined in Appendix A.1.

- $\kappa^{'}_{t'}{}^{(i)}$, the projected period effect at time t' for $i = 1, 2, 3$.

- $\tilde{\kappa}^{(i)}_t$, the posterior distribution of $\kappa^{(i)}_t$ for $t = t_1, \ldots, t_{n_y}$, $i = 1, 2, 3$.

- $\tilde{\kappa}^{(i),w}_{t,j}$ for $j = 1, \ldots, N_1$, the posterior distribution of $\kappa^{(i)}_t$ for the j^{th} simulated death scenario with weight w, where $i = 1, 2, 3$.

- $\tilde{\kappa}^{(i),w,(k)}_{t,j}$, the k^{th} posterior sample for $\tilde{\kappa}^{(i),w}_{t,j}$, where $k = 1, \ldots, N_2$.

- $\tilde{\kappa}^{(i),w}_t = \{\tilde{\kappa}^{(i),w}_{t,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\kappa}^{(i),w}_{t,j^*}$ for $j = 1, \ldots, N_1$, the posterior distribution of $\kappa^{(i)}_t$ for the death scenario j^* with weight w, where $i = 1, 2, 3$.

- $\hat{\kappa}_t^{(i),\mathrm{EW}}$ for $i = 1, 2, 3$, the MLE for $\kappa_t^{(i)}$ for the England and Wales, where $t = t_1, \ldots, t_{n_y}$.

- $\tilde{\kappa}_{t'}^{'\,(i)}$, the posterior distribution of the projected $\tilde{\kappa}_t^{(i)}$ in year t' for $i = 1, 2, 3$.

- $\tilde{\kappa}_{t',j}^{'\,(i),w}$, the posterior distribution of the projected $\tilde{\kappa}_{t,j}^{(i),w}$ for the j^{th} simulated death scenario with weight w for $i = 1, 2, 3$ in year t'.

- $\tilde{\kappa}_{t'}^{'\,(i),w} = \{\tilde{\kappa}_{t',j}^{'\,(i),w}\}_{j=1,\ldots,N_1}$.

- $\tilde{\kappa}_{t',j}^{'\,(i),w,(k)} = \{\tilde{\kappa}_{t',j}^{'\,(i),w,(k,l)}\}_{l=1,\ldots,N_4}$, the projection for the k^{th} sample drawn for the $\tilde{\kappa}_{t,j}^{(i),w}$ of the j^{th} death scenario, where $k = 1, \ldots, N_2$ and $i = 1, 2, 3$.

- $\tilde{\kappa}_{t',j}^{'\,(i),w,(k,l)}$, the l^{th} sample path of $\tilde{\kappa}_{t',j}^{'\,(i),w,(k)}$, where $l = 1, \ldots, N_4$ and $i = 1, 2, 3$.

- $\tilde{\kappa}_{t',j^*}^{'\,(i),w}$ for $i = 1, 2, 3$, the posterior distribution of the projected $\tilde{\kappa}_{t,j^*}^{(i),w}$ for the death scenario j^* with weight w.

- $\tilde{\kappa}_{t'}^{'\,(i),\mathrm{EW}}$ for $i = 1, 2, 3$, the posterior distribution of the projected $\tilde{\kappa}_t^{(i),\mathrm{EW}}$ for the England and Wales data.

- $\tilde{\kappa}_{t',j'}^{'\,(i),w}$, the posterior distribution of the projected $\tilde{\kappa}_{t,j'}^{(i),w}$ for the $j'^{\text{ th}}$ simulated death scenario with weight w in year t', where $i = 1, 2, 3$ and $j' = j_1, \ldots, j_{100}$.

- $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, defined in Appendix A.2.

- $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$, defined in Appendix A.4.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}$, the matrix of the MLEs for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales data.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(i, k)$, the MLE for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$ for the England and Wales data, where $i, k = 1, 2, 3$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(i, k)$ for $i, k = 1, 2, 3$, defined in Appendix A.4.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}} = \{\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{EW}}(i, k)\}_{i,k=1,2,3}$.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j} = \{\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j}(i, k)\}_{i,k=1,2,3}$, the point estimator for the $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the simulated death scenario $j = 1, \ldots, N_1$ with weight w.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ w} = \{\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ w,j}\}_{j=1,\ldots,N_1}$.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ w,j^*}(i,k)$, the MLE of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i.k)$ for the simulated death scenario j^* with weight w, where $i,k = 1,2,3$.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ w}(i,k) = \{\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\ w,j}(i,k)\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{V}}_{\epsilon}$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$.

- $\bar{\tilde{\boldsymbol{V}}}_{\epsilon}$, the posterior mean of $\tilde{\boldsymbol{V}}_{\epsilon}$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}$ for $j = 1,\ldots,N_1$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the j^{th} simulated death scenario with weight w.

- $\bar{\tilde{\boldsymbol{V}}}_{\epsilon}^{w,j}$ for $j = 1,\ldots,N_1$, the posterior mean of $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}$.

- $\bar{\tilde{\boldsymbol{V}}}_{\epsilon}^{w} = \{\bar{\tilde{\boldsymbol{V}}}_{\epsilon}^{w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}(i,k)$ for $j = 1,\ldots,N_1$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i,k)$ for the j^{th} simulated death scenario with weight w, where $i,k = 1,2,3$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{w}(i,k) = \{\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}(i,k)\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j^*}(i,k)$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i.k)$ for the simulated death scenario j^* with weight w, where $i,k = 1,2,3$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j'}(i,k)$ for $j' = 1,\ldots,j_{100}$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i,k)$ for the j'^{th} death scenario with weight w, where $i,k = 1,2,3$.

- $\boldsymbol{\mu}$, defined in Appendix A.2.

- $\tilde{\boldsymbol{\mu}}$, the posterior distribution of $\boldsymbol{\mu}$.

- $\bar{\tilde{\boldsymbol{\mu}}}$, the posterior mean of $\tilde{\boldsymbol{\mu}}$.

- $\tilde{\boldsymbol{\mu}}^{w,j}$ for $j = 1,\ldots,N_1$, the posterior distribution of $\boldsymbol{\mu}$ for the j^{th} simulated death scenario with weight w.

- $\bar{\tilde{\boldsymbol{\mu}}}^{w,j}$ for $j = 1,\ldots,N_1$, the posterior mean of $\tilde{\boldsymbol{\mu}}^{w,j}$.

- $\bar{\tilde{\boldsymbol{\mu}}}^{w} = \{\bar{\tilde{\boldsymbol{\mu}}}^{w,j}\}_{j=1,\ldots,N_1}$.

- $\boldsymbol{\mu}_i$ for $i = 1, 2, 3$, defined in Appendix A.4.

- $\tilde{\boldsymbol{\mu}}_i$ for $i = 1, 2, 3$, the posterior distribution of $\boldsymbol{\mu}_i$.

- $\tilde{\boldsymbol{\mu}}_i^{w,j}$ for $j = 1, \ldots, N_1$, the posterior distribution of $\boldsymbol{\mu}_i$ for the j^{th} simulated death scenario, where $i = 1, 2, 3$.

- $\tilde{\boldsymbol{\mu}}_i^{w} = \{\tilde{\boldsymbol{\mu}}_i^{w,j}\}_{j=1,\ldots,N_1}$.

- $\bar{\tilde{\boldsymbol{\mu}}}_i^{w,j}$ for $j = 1, \ldots, N_1$, the posterior mean of $\tilde{\boldsymbol{\mu}}_i^{w,j}$ for $i = 1, 2, 3$.

- $\bar{\tilde{\boldsymbol{\mu}}}_i^{w} = \{\bar{\tilde{\boldsymbol{\mu}}}_i^{w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{\mu}}_i^{w,j^*}$, the posterior distribution of $\boldsymbol{\mu}_i$ for the simulated death scenario j^* with weight w, where $i = 1, 2, 3$.

- $\tilde{\boldsymbol{\mu}}_i^{\text{EW}}$, the posterior distribution of $\boldsymbol{\mu}_i$ for the England and Wales data, where $i = 1, 2, 3$.

- $\tilde{\boldsymbol{\mu}}_i^{w,j'}$ for $j' = j_1, \ldots, j_{100}$, the posterior distribution of $\boldsymbol{\mu}_i$ for the j'^{th} death scenario with weight w, where $i, k = 1, 2, 3$.

- $\hat{\boldsymbol{\mu}}_i^{w,j}$ for $j = 1, \ldots, N_1$, the MLE of $\boldsymbol{\mu}_i$ for the j^{th} death scenario with weight w, where $i = 1, 2, 3$.

- $\hat{\boldsymbol{\mu}}_i^{w} = \{\hat{\boldsymbol{\mu}}_i^{w,j}\}_{j=1,\ldots,N_1}$.

- $\hat{\boldsymbol{\mu}}_i^{w,j^*}$, the MLE of $\boldsymbol{\mu}_i$ for the death scenario j^* with weight w, where $i = 1, 2, 3$.

- $\hat{\boldsymbol{\mu}}_i^{\text{EW}}$, the MLE of $\boldsymbol{\mu}_i$ for the England and Wales data, where $i = 1, 2, 3$.

- $\boldsymbol{\gamma}^{(4)}$, defined in Appendix A.1.

- $\boldsymbol{\gamma}$, used inter-changeably with $\boldsymbol{\gamma}^{(4)}$.

- $\tilde{\boldsymbol{\gamma}}^{(4)}$, the posterior distribution of $\boldsymbol{\gamma}^{(4)}$.

- $\tilde{\boldsymbol{\gamma}}_j^{(4),w}$, the posterior distribution of $\boldsymbol{\gamma}^{(4)}$ for the j^{th} simulated death scenario with weight w.

- $\tilde{\boldsymbol{\gamma}}^{(4),w} = \{\tilde{\boldsymbol{\gamma}}_j^{(4),w}\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{\gamma}}_{j^*}^{(4),w}$, the posterior distribution of $\boldsymbol{\gamma}^{(4)}$ for the death scenario j^* with weight w.

- $\tilde{\boldsymbol{\gamma}}^{(4),\text{EW}}$, the posterior distribution of $\boldsymbol{\gamma}^{(4)}$ for the England and Wales data.

- $\hat{\boldsymbol{\gamma}}^{(4)}$, the MLE of $\boldsymbol{\gamma}^{(4)}$.

- $\hat{\boldsymbol{\gamma}}_j^{(4),w}$ for $j = 1, \ldots, N_1$, the MLE of $\boldsymbol{\gamma}^{(4)}$ for the j^{th} simulated death scenario with weight w.

- $\hat{\boldsymbol{\gamma}}^{(4),w} = \{\hat{\boldsymbol{\gamma}}_j^{(4),w}\}_{j=1,\ldots,N_1}$.

- $\hat{\boldsymbol{\gamma}}_{j^*}^{(4),w}$, the MLE of $\boldsymbol{\gamma}^{(4)}$ for the death scenario j^* with weight w.

- $\hat{\boldsymbol{\gamma}}^{(4),\text{EW}}$, the MLE of $\boldsymbol{\gamma}^{(4)}$ for the England and Wales data.

- $\gamma_c^{(4)}$, defined in Appendix A.1.

- $\gamma_{c'}^{'(4)}$, the projected cohort effect for the cohort year c'.

- $\tilde{\gamma}_{c,j}^{(4),w}$, the posterior distribution of $\gamma_c^{(4)}$ of the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$ and $c = t - x$.

- $\tilde{\gamma}_{c,j}^{(4),w,(k)}$, the k^{th} posterior sample for $\tilde{\gamma}_{c,j}^{(4),w}$, where $k = 1, \ldots, N_2$.

- $\tilde{\gamma}_c^{(4),w} = \{\tilde{\gamma}_{c,j}^{(4),w}\}_{j=1,\ldots,N_1}$.

- $\hat{\gamma}_c^{(4),\text{EW}}$, the MLE of $\gamma_c^{(4)}$ for the England and Wales data.

- $\tilde{\gamma}_{c'}^{'(4)}$, the posterior distribution of the projected $\tilde{\gamma}_c^{(4)}$ in the year of birth c.

- $\tilde{\gamma}_{c',j}^{'(4),w}$, the posterior distribution of the projected $\tilde{\gamma}_{c,j}^{(4),w}$ for the j^{th} simulated death scenario with weight w in the year of birth c'.

- $\tilde{\gamma}_{c'}^{'(4),w} = \{\tilde{\gamma}_{c',j}^{'(4),w}\}_{j=1,\ldots,N_1}$.

- $\tilde{\gamma}^{'\,(4),w,(k)}_{c',j} = \{\tilde{\gamma}^{'\,(4),w,(k,l)}_{c',j}\}_{l=1,\ldots,N_4}$, the projection for the k^{th} sample drawn for the $\tilde{\gamma}^{(4),w}_{c,j}$ of the j^{th} death scenario, where $k = 1, \ldots, N_2$ and $i = 1, 2, 3$.

- $\tilde{\gamma}^{'\,(4),w,(k,l)}_{c',j}$, the l^{th} sample path of $\tilde{\gamma}^{'\,(4),w,(k)}_{c',j}$, where $l = 1, \ldots, N_4$ and $i = 1, 2, 3$.

- $\tilde{\gamma}^{'\,(4),w}_{c',j^*}$, the posterior distribution of the projected $\tilde{\gamma}^{(4),w}_{c,j^*}$ for the death scenario j^* with weight w in the year of birth c'.

- $\tilde{\gamma}^{'\,(4),\text{EW}}_{c'}$, the posterior distribution of the projected $\tilde{\gamma}^{(4),\text{EW}}_{c}$ for the England and Wales data in the year of birth c'.

- $\tilde{\gamma}^{'\,(4),w}_{c',j'}$, the posterior distribution of the projected $\tilde{\gamma}^{(4),w}_{c,j'}$ for the $j^{'\text{ th}}$ simulated death scenario with weight w in the year of birth c', where $j' = j_1, \ldots, j_{100}$.

- α_γ, defined in Appendix A.2.

- $\tilde{\alpha}_\gamma$, the posterior distribution of α_γ.

- $\tilde{\alpha}^{w,j}_\gamma$, the posterior distribution of α_γ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\tilde{\alpha}^{w}_\gamma = \{\tilde{\alpha}^{w,j}_\gamma\}_{j=1,\ldots,N_1}$.

- $\tilde{\alpha}^{w,j^*}_\gamma$, the posterior distribution of α_γ for the death scenario j^* with weight w.

- $\tilde{\alpha}^{\text{EW}}_\gamma$, the posterior distribution of α_γ for the England and Wales data.

- $\tilde{\alpha}^{w,j'}_\gamma$, the posterior distribution of α_γ for the j'^{th} death scenario with weight w, where $j' = j_1, \ldots, j_{100}$.

- $\bar{\tilde{\alpha}}^{w,j}_\gamma$, the posterior mean of $\tilde{\alpha}^{w,j}_\gamma$ with weight w for $j = 1, \ldots, N_1$.

- $\bar{\tilde{\alpha}}^{w}_\gamma = \{\bar{\tilde{\alpha}}^{w,j}_\gamma\}_{j=1,\ldots,N_1}$.

- $\hat{\alpha}_\gamma$, the MLE of α_γ.

- $\hat{\alpha}^{w,j}_\gamma$, the MLE of α_γ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\hat{\alpha}_\gamma^w = \{\hat{\alpha}_\gamma^{w,j}\}_{j=1,\ldots,N_1}$.

- $\hat{\alpha}_\gamma^{w,j^*}$, the MLE of α_γ for the death scenario j^* with weight w.

- $\hat{\alpha}_\gamma^{\mathrm{EW}}$, the MLE of α_γ for the England and Wales data.

- g, a constant for the prior distribution of α_γ.

- σ_γ, defined in Appendix A.2.

- $\tilde{\sigma}_\gamma$, the posterior distribution of σ_γ.

- $\tilde{\sigma}_\gamma^{w,j}$, the posterior distribution of σ_γ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\tilde{\sigma}_\gamma^w = \{\tilde{\sigma}_\gamma^{w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\sigma}_\gamma^{w,j^*}$, the posterior distribution of σ_γ for the death scenario j^* with weight w.

- $\tilde{\sigma}_\gamma^{\mathrm{EW}}$, the posterior distribution of σ_γ for the England and Wales data.

- $\tilde{\sigma}_\gamma^{w,j'}$, the posterior distribution of σ_γ for the j'^{th} death scenario with weight w, where $j' = j_1, \ldots, j_{100}$.

- $\bar{\tilde{\sigma}}_\gamma^{w,j}$, the posterior mean of $\tilde{\sigma}_\gamma^{w,j}$ with weight w for $j = 1, \ldots, N_1$.

- $\bar{\tilde{\sigma}}_\gamma^w = \{\bar{\tilde{\sigma}}_\gamma^{w,j}\}_{j=1,\ldots,N_1}$.

- $\hat{\sigma}_\gamma^{w,j}$, the MLE of σ_γ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\hat{\sigma}_\gamma^w = \{\hat{\sigma}_\gamma^{w,j}\}_{j=1,\ldots,N_1}$.

- $\hat{\sigma}_\gamma^{w,j^*}$, the MLE of σ_γ for the death scenario j^* with weight w.

- $\hat{\sigma}_\gamma^{\mathrm{EW}}$, the MLE of σ_γ for the England and Wales data.

- ϵ_c, defined in Appendix A.2.

- $(\cdot) = 1, 2, 3, 4, 5$, the number of short cohorts removed from the beginning and the end of the years of birth.

- N_1, defined in Appendix A.4.

- N_2, defined in Appendix A.4.

- w, defined in Appendix A.2.

- j^*, a randomly selected number from $1, \ldots, N_1$.

- $j' = j_1, \ldots, j_{100}$, 100 randomly selected number from $1, \ldots, N_1$.

A.6 Chapter 7

- $D_j^w(t,x)$, the death counts in year $t = t_1, \ldots, t_{n_y}$ aged $x = x_1, \ldots, x_{n_a}$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $D^w(t,x) = \{D_j^w(t,x)\}_{j=1,\ldots,N_1}$.

- $E_0(t,x)$, defined in A.2.

- $\boldsymbol{\theta}_1$, defined in A.1.

- $\hat{\boldsymbol{\theta}}_1$, the MLE of $\boldsymbol{\theta}_1$.

- $\hat{\boldsymbol{\theta}}_1^{(\cdot)}$, the MLEs of the latent parameter $\boldsymbol{\theta}$ for a specified population $(\cdot)$, where $(\cdot) =$ EW, UK, ST for the data of England and Wales, UK and Scotland respectively.

- $\hat{\boldsymbol{\theta}}_1'$, distribution of the projected $\boldsymbol{\theta}_1$, conditional on the $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$.

- $\tilde{\boldsymbol{\theta}}_1$, the posterior distribution of $\boldsymbol{\theta}_1$.

- $\tilde{\boldsymbol{\theta}}_1'$, the posterior distribution of the projected $\boldsymbol{\theta}_1$, conditional on the $\tilde{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_2$.

- $\bar{\tilde{\boldsymbol{\theta}}}_1$, the posterior mean of $\tilde{\boldsymbol{\theta}}_1$.

- $\bar{\tilde{\boldsymbol{\theta}}}_1'$, the distribution of the projected $\boldsymbol{\theta}_1$, conditional on the $\bar{\tilde{\boldsymbol{\theta}}}_1$ and $\bar{\tilde{\boldsymbol{\theta}}}_2$. Note that $\bar{\tilde{\boldsymbol{\theta}}}_1'$ includes no parameter uncertainty.

- $\boldsymbol{\theta}_2$, defined in A.1.

- $\tilde{\boldsymbol{\theta}}_2$, the posterior distribution of $\boldsymbol{\theta}_2$.

- $\bar{\tilde{\boldsymbol{\theta}}}_2$, the posterior mean of $\tilde{\boldsymbol{\theta}}_2$.

- $\hat{\boldsymbol{\theta}}_2$, the MLE of $\boldsymbol{\theta}_2$.

- θ_1, defined in A.1.

- $\theta_{1,0}$, defined in Appendix A.2.

- $\hat{\theta}_1$, the MLE of θ_1.

- $\hat{\theta}_1^{\mathrm{EW}}$, the MLE of θ_1 for the England and Wales data.

- $m(\theta_{1,0}, t, x)$, defined in Appendix A.2.

- $\Delta m^{(\cdot)}(t, x)$, the difference of the fitted death rate at year$t$ aged x between the data of England and Wales and the Scotland , where $(\cdot)$ = PU, PC and MLE. Note that $\Delta m^{(\cdot)}(t, x)$ represents the difference of the central fitted rate if $(\cdot)$ = PU.

- $m'(t', x)$, the projected death rate in year t' aged x.

- $m'(\tilde{\boldsymbol{\theta}}_1', t', x)$, the posterior distribution of the projected death rate conditional on $\tilde{\boldsymbol{\theta}}_1'$.

- $m'(\bar{\tilde{\boldsymbol{\theta}}}_1', t', x)$, the distribution of the projected death rate conditional on $\bar{\tilde{\boldsymbol{\theta}}}_1'$. Note that $m'(\bar{\tilde{\boldsymbol{\theta}}}_1', t', x)$ includes no parameter uncertainty.

- $m'(\hat{\boldsymbol{\theta}}_1', t', x)$, the distribution of the projected death rate conditional on $\hat{\boldsymbol{\theta}}_1'$.

- $\Delta m'^{(\cdot)}(t', x)$, the difference of the central projected death rate in year t' aged x between the England and Wales and the Scotland , where $(\cdot)$ = PU, PC and MLE.

- $\boldsymbol{\kappa}^{(i)}$ for $i = 1, 2, 3$, defined in A.1.

- $\boldsymbol{\mu}$, defined in A.2.

- $\tilde{\boldsymbol{\mu}}$, the posterior distribution of $\boldsymbol{\mu}$.

- $\tilde{\boldsymbol{\mu}}^{w,j}$, the posterior distribution of $\boldsymbol{\mu}$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\tilde{\boldsymbol{\mu}}^{w} = \{\tilde{\boldsymbol{\mu}}^{w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{\mu}}^{\mathrm{SL}}$, the posterior distribution of $\boldsymbol{\mu}$ for the Scotland data.

- $\boldsymbol{\mu}_i$ for $i = 1, 2, 3$, defined in Appendix A.4.

- $\tilde{\boldsymbol{\mu}}_i$ for $i = 1, 2, 3$, the posterior distribution of $\boldsymbol{\mu}_i$.

- $\tilde{\boldsymbol{\mu}}_i^{\mathrm{SL}}$, the posterior distribution of $\boldsymbol{\mu}_i$ for the Scotland data, where $i = 1, 2, 3$.

- $\tilde{\boldsymbol{\mu}}_i^{\mathrm{EW}}$, the posterior distribution of $\boldsymbol{\mu}_i$ for the England and Wales data, where $i = 1, 2, 3$.

- $\tilde{\boldsymbol{\mu}}_i^{w,j}$ for $i = 1, 2, 3$, the posterior distribution of $\boldsymbol{\mu}_i$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\tilde{\boldsymbol{\mu}}_i^{w} = \{\tilde{\boldsymbol{\mu}}_i^{w,j}\}_{j=1,\ldots,N_1}$.

- $\hat{\boldsymbol{\mu}}_i$ for $i = 1, 2, 3$, the MLE of $\boldsymbol{\mu}_i$.

- $\hat{\boldsymbol{\mu}}_i^{\mathrm{SL}}$ for $i = 1, 2, 3$, the MLE of $\boldsymbol{\mu}_1$ for the Scotland data.

- $\hat{\boldsymbol{\mu}}_i^{\mathrm{EW}}$ for $i = 1, 2, 3$, the MLE of $\boldsymbol{\mu}_1$ for the England and Wales data.

- $\hat{\boldsymbol{\mu}}_i^{w,j}$ for $i = 1, 2, 3$, the MLE of $\boldsymbol{\mu}_i$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\hat{\boldsymbol{\mu}}_i^{w} = \{\hat{\boldsymbol{\mu}}_i^{w,j}\}_{j=1,\ldots,N_1}$.

- $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, defined in A.2.

- $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$ for $i, k = 1, 2, 3$, defined in Appendix A.4.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}(i, k)$ for $i, k = 1, 2, 3$, the MLE of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{SL}}(i, k)$ for $i, k = 1, 2, 3$, the MLE of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$ for the Scotland data.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}(i, k)$ for $i, k = 1, 2, 3$, the MLE of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$ for the England and Wales data.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{UK}}$, the MLE of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the UK data.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{EW}}$, the MLE of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales data.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\mathrm{SL}}$, the MLE of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the Scotland data.

- $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{w,j}(i, k)$ for $i, k = 1, 2, 3$, the MLE of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i, k)$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{w}(i,k) = \{\hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{w,j}(i,k)\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{V}}_{\epsilon}$, posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the j^{th} simulated death scenario with weight w, where $j = 1, \ldots, N_1$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{w} = \{\tilde{\boldsymbol{V}}_{\epsilon}^{w,j}\}_{j=1,\ldots,N_1}$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the Scotland data.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales data.

- $\tilde{\boldsymbol{V}}_{\epsilon}(i,k)$ for $i,k = 1,2,3$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i,k)$.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{SL}}(i,k)$ for $i,k = 1,2,3$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i,k)$ for the Scotland data.

- $\tilde{\boldsymbol{V}}_{\epsilon}^{\text{EW}}(i,k)$ for $i,k = 1,2,3$, the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(i,k)$ for the England and Wales data.

- A-UK, the prior setting for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the UK data such that $\text{Mode}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{\text{UK}}$.

- A-EW, the prior setting for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales data such that $\text{Mode}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{\text{EW}}$.

- B-EW, the prior setting for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the England and Wales data such that $\text{E}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{\text{EW}}$.

- B-UK, the prior setting for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$ for the UK data such that $\text{E}[\boldsymbol{V}_{\boldsymbol{\epsilon}}] = \hat{\boldsymbol{V}_{\boldsymbol{\epsilon}}}^{\text{UK}}$.

- ν, defined in A.4.

- Σ, defined in A.4.

- α_γ, defined in A.2.

- $\tilde{\alpha}_\gamma$, the posterior distribution of α_γ.

- $\tilde{\alpha}_\gamma^{\text{SL}}$, the posterior distribution of α_γ for the Scotland data.

- σ_γ, defined in A.2.

- $\tilde{\sigma}_\gamma$, the posterior distribution of σ_γ.

- $\tilde{\sigma}_\gamma^{\mathrm{SL}}$, the posterior distribution of σ_γ for the Scotland data.

- w, defined in A.2.

- N_2, defined in A.4.

- N_1, defined in A.4.

- t', defined in A.5.

- PU, with parameter uncertainty.

- PC, without parameter uncertainty.

Appendix B

For Chapter 3

B.1 Applying Identifiability Constraints for Parameter Estimation

This section describes the procedure for applying the following identifiability constraints introduced in (2.6).

$$\sum_{c\in C}\gamma_c^{(4)}=0, \qquad \sum_{c\in C}c\gamma_c^{(4)}=0, \qquad \sum_{c\in C}c^2\gamma_c^{(4)}=0$$

where $C=(t_1-x_{n_a},\ldots,t_{n_y}-x_1)$ is the set of all years of birth in a given dataset.

Let $\boldsymbol{s}$ be a matrix with 3×3 dimensions. Denote as $s_{i,j}$ the entry in $\boldsymbol{s}$ for $i,j=1,2,3$. Let $n_c=n_y+n_a-1$ and

$$\begin{aligned}
m &= \frac{1}{n_c}\sum_{c=t_1-x_{n_a}}^{t_{n_y}-x_1}(c+\bar{x})\\
s_{1,1} &= n_c\\
s_{1,2} &= \sum_{c=t_1-x_{n_a}}^{t_{n_y}-x_1}(c+\bar{x}-m)=s_{2,1}\\
s_{1,3} &= \sum_{c=t_1-x_{n_a}}^{t_{n_y}-x_1}(c+\bar{x}-m)^2=s_{3,1}=s_{2,2}\\
s_{2,3} &= \sum_{c=t_1-x_{n_a}}^{t_{n_y}-x_1}(c+\bar{x}-m)^3=s_{3,2}\\
s_{3,3} &= \sum_{c=t_1-x_{n_a}}^{t_{n_y}-x_1}(c+\bar{x}-m)^4.
\end{aligned}$$

Let $\boldsymbol{u}=(u_1,u_2,u_3)^T$ be a vector with length 3. Its entries can be expressed as:

$$u_i=\sum_{c=t_1-x_{n_a}}^{t_{n_y}-x_1}\gamma_c^{(4)}(c+\bar{x}-m)^{i-1}$$

for $i = 1, 2, 3$.

Let $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)^T$ be cross product of $\boldsymbol{s}^{-1}$ and $\boldsymbol{u}$, that is $\boldsymbol{\phi} = \boldsymbol{s}^{-1} \times \boldsymbol{u}$. We introduce three constants $\{\Phi(i)\}_{i=1,2,3}$ such that:

$$\begin{aligned}
\Phi(1) &= \phi_1 - \phi_2 m + \phi_3 m^2 \\
\Phi(2) &= \phi_2 - 2m\phi_3 \\
\Phi(3) &= \phi_3
\end{aligned}$$

The parameter vector θ_1 can be adjusted as follows:

$$\begin{aligned}
\tilde{\gamma}_c^{(4)} | \gamma_c^{(4)}, \{\Phi\}_{i=1,2,3} &= \gamma_c^{(4)} - \Phi(1) - \Phi(2)(xc + \bar{x}) - \Phi(3)(c + \bar{x})^2 \\
\tilde{\kappa}_t^{(1)} | \kappa_t^{(1)}, \{\Phi\}_{i=1,2,3} &= \kappa_t^{(1)} + \Phi(1) + \Phi(2)t + \Phi(3)(t^2 + \hat{\sigma}^2) \\
\tilde{\kappa}_t^{(2)} | \kappa_t^{(2)}, \{\Phi\}_{i=1,2,3} &= \kappa_t^{(2)} - \Phi(2) - 2t\Phi(3) \\
\tilde{\kappa}_t^{(3)} | \kappa_t^{(3)}, \{\Phi\}_{i=1,2,3} &= \kappa_t^{(3)} + \Phi(3).
\end{aligned}$$

The impact of the identifiability constraints in our study can be removed by calculating the following quantities of the point estimates: $\Delta^3 \tilde{\kappa}_t^{(1)}$, $\Delta^2 \tilde{\kappa}_t^{(2)}$, $\Delta \tilde{\kappa}_t^{(3)}$ and $\Delta^3 \tilde{\gamma}_c^{(4)}$, where Δ^k represents the k^{th} order difference. For example, we can remove the adjustment $\Phi(3)$ for $\kappa_{t^{(3)}}$ by:

$$\Delta \tilde{\kappa}_t^{(3),w} = \tilde{\kappa}_t^{(3)} - \tilde{\kappa}_{t-1}^{(3)} = \kappa_t^{(3)} + \Phi(3) - (\kappa_{t-1}^{(3)} + \Phi(3)) = \kappa_t^{(3)} - \kappa_{t-1}^{(3)}.$$

Same idea applies to the other period and cohort effects.

Appendix C

For Chapter 5

C.1 Metropolis-Hastings Algorithm and Gibbs Sampler

C.1.1 Review of the Metropolis-Hastings Algorithm

This section gives a short review of the Metropolis-Hastings (MH) algorithm and the Gibbs sampler. Let:

- $\boldsymbol{\theta} = \{\theta_k\}_{k=1,\ldots,N}$ = the vector of latent and hyper-parameters to be updated with length N.
- $\theta_j(i)$ = the value of the parameter θ_j at iteration i, given the values of other parameters $\theta_1(i+1), \ldots, \theta_{j-1}(i+1), \theta_{j+1}(i), \ldots, \theta_N(i)$.
- $\boldsymbol{y}$ = observed data.
- $p(\boldsymbol{\theta}|\boldsymbol{y})$ = the posterior density of $\boldsymbol{\theta}$ given $\boldsymbol{y}$.

The steps for updating the state from i to $i+1$ are as follows:

1. Generate a candidate $\hat{\theta}_j(i)$ from a proposal distribution with density $f(\theta_j|\theta_j(i))$.

2. Calculate the probability of acceptance defined as

$$\alpha(\theta_j(i), \hat{\theta}_j(i)) = \min\left(1, \frac{p(\hat{\theta}_j(i)|\boldsymbol{y}) f(\theta_j(i)|\hat{\theta}_j(i))}{p(\theta_j(i)|\boldsymbol{y}) f(\hat{\theta}_j(i)|\theta_j(i))}\right) \tag{C.1}$$

3. Draw a sample u from a uniform distribution $U(0,1)$. If $u \leq \alpha(\theta_j(i), \hat{\theta}_j(i))$, the candidate $\hat{\theta}_j(i)$ is accepted to replace the current $\theta_j(i)$, that is $\theta_j(i+1) = \hat{\theta}_j(i)$. Otherwise the Markov Chain does not update and $\theta_j(i+1) = \theta_j(i)$.

C.1.2 Case Study

Metropolis-Hastings algorithm can be low efficiency on convergence due to the random walk behaviour, especially for models with large number of correlated parameters. This section demonstrates an empirical study to show how the MH procedure causes us problems on achieving a stationary posterior distribution. The underlying dataset is still the males in England and Wales during 1961-2011, aged from 50-89 last birthday.

For the purpose of simplicity, the model used for the illustration is the Lee-Carter model (M1, Lee & Carter (1992)),

$$\log m(t,x) = \alpha_x + \beta_x \kappa_t,$$

and we assume that $D(t,x) \sim \text{Pois}(m(t,x)E(t,x))$. We fitted the M1 model to the underlying dataset and maximised the Poisson likelihood for the empirical point estimates of the latent parameters, denoted as $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\kappa}}$.

Period Effect

We model the period effect $\boldsymbol{\kappa} = (\kappa_{t_2}, \ldots, \kappa_{t_{ny}})^T$ we a simple random walk model:

$$\kappa_t - \kappa_{t-1} = \theta + \epsilon_t, \text{for } t \geq t_2 \tag{C.2}$$

with drift θ and $i.i.d$ error term $\epsilon_t \sim N(0, \sigma_\kappa^2)$. We fix $\kappa_{t_1} = 0$ and $\theta = -1$ to incorporate the identifiability constraints. It leads to $\kappa_t - \kappa_{t-1} \sim N(\theta, \sigma_\kappa^2)$ with its density $p(\kappa_t | \kappa_{t-1}, \theta, \sigma_\kappa^2)$:

$$p(\kappa_t | \kappa_{t-1}, \theta, \sigma_\kappa^2) \propto \frac{1}{\sigma_\kappa} \exp\left\{ -\frac{(\kappa_t - \kappa_{t-1} - \theta)^2}{2\sigma_\kappa^2} \right\} \tag{C.3}$$

We assume that the prior of σ_κ^2 is $\sigma_\kappa^2 \sim$ Inv- Gamma(a_κ, b_κ), where $a_\kappa = 2.0001$ and $b_\kappa = (a_\kappa - 1)\hat{\sigma}_\kappa^2$. The $\hat{\sigma}_\kappa^2$ is the variance of $\hat{\kappa}_t - \hat{\kappa}_{t-1}$ for $t \geq t_2$.

Denote as

$$\boldsymbol{\kappa}_{-t} = (\kappa_{t_1}, \ldots, \kappa_{t-1}, \kappa_{t+1}, \ldots, \kappa_{t_{n_y}})^T \tag{C.4}$$

the vector of period parameters excluding κ_t.

Denote as

$$\boldsymbol{D}_t = (D_{x_1,t}, D_{x_2,t}, \ldots, D_{x_{n_a},t})^T \tag{C.5}$$

the vector of the number of deaths for all the age levels in year t. Similarly we can define

$$D_{-t} = (\boldsymbol{D_{t_1}}, \ldots, \boldsymbol{D_{t-1}}, \boldsymbol{D_{t+1}}, \ldots, \boldsymbol{D}_{t_{n_y}})^T \tag{C.6}$$

to be the matrix of death without year t. Denote as

$$p(\kappa_t | \boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \boldsymbol{D},\ \sigma_\kappa^2, \sigma_\beta^2, \theta) \tag{C.7}$$

the conditional posterior of κ_t, given all the other parameters. Thus we have

$$p(\kappa_t | \boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \boldsymbol{D},\ \sigma_\kappa^2, \sigma_\beta^2, \theta) = \frac{p(\boldsymbol{\kappa},\ \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \boldsymbol{D},\ \sigma_\kappa^2, \sigma_\beta^2, \theta)}{p(\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \boldsymbol{D},\ \sigma_\kappa^2, \sigma_\beta^2, \theta)} \tag{C.8}$$

$$= \frac{p(\kappa_{t_{n_y}}, \boldsymbol{D}_{t_{n_y}} | \boldsymbol{\kappa_{-t_{n_y}}},\ \boldsymbol{D_{-t_{n_y}}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \theta) p(\boldsymbol{\kappa_{-t_{n_y}}},\ \boldsymbol{D_{-t_{n_y}}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \theta)}{p(\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \boldsymbol{D},\ \sigma_\kappa^2, \sigma_\beta^2, \theta)}$$

Then repeat the above iteration for $p(\boldsymbol{\kappa_{-n_y}},\ \boldsymbol{D_{-n_y}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \theta)$, so that the con-

ditional posterior can be re-written as

$$
\begin{aligned}
p(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \ \boldsymbol{\beta}, \ \boldsymbol{D}, \ \sigma^2_\kappa, \sigma^2_\beta, \theta) &= \frac{p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) p(\kappa_1, \boldsymbol{D}_1|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta)}{p(\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma^2_\kappa, \sigma^2_\beta, \theta)} \\
&\quad \times \prod_{t=2}^{n_y} p(\kappa_t, \boldsymbol{D}_t|\kappa_1, \ldots \kappa_{t-1}, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_{t-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) \\
&\quad \propto p(\kappa_1, \boldsymbol{D}_1|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) \\
&\quad \times \prod_{t=2}^{n_y} p(\kappa_t, \boldsymbol{D}_t|\kappa_1, \ldots \kappa_{t-1}, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_{t-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta)
\end{aligned} \tag{C.9}
$$

Note that $\boldsymbol{D}_t$ is mutually independent given the other parameters and a function of $\boldsymbol{\kappa}, \boldsymbol{\beta}, \boldsymbol{\alpha}$ only. Thus the second part of Equation (C.9) can be re-written as the product of

$$
\begin{aligned}
& p(\kappa_t, \boldsymbol{D}_t|\kappa_1, \ldots \kappa_{t-1}, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_{t-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) \\
&= p(\boldsymbol{D}_t|\kappa_1, ..., \kappa_t, \boldsymbol{D}_1, ..., \boldsymbol{D}_{t-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) \\
&\times p(\kappa_t|\kappa_1, ..., \kappa_{t-1}, \boldsymbol{D}_1, ..., \boldsymbol{D}_{t-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) \\
&= p(\boldsymbol{D}_t|\kappa_t, \boldsymbol{\alpha}, \boldsymbol{\beta}) p(\kappa_t|\kappa_1, ..., \kappa_{t-1}, \boldsymbol{D}_1, ..., \boldsymbol{D}_{t-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) \\
&= p(\boldsymbol{D}_t|\kappa_t, \boldsymbol{\alpha}, \boldsymbol{\beta}) p(\kappa_t|\kappa_{t-1}, \sigma^2_\kappa, \theta)
\end{aligned} \tag{C.10}
$$

Then the first part of Equation (C.9) can be written as

$$
\begin{aligned}
p(\kappa_1, \boldsymbol{D}_1|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) &= p(\boldsymbol{D}_1|\kappa_1, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) \\
&\quad \times p(\kappa_1|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\kappa, \sigma^2_\beta, \theta) \\
&\quad = p(\boldsymbol{D}_1|\kappa_1, \boldsymbol{\alpha}, \boldsymbol{\beta}) p(\kappa_1|\sigma^2_\kappa, \theta)
\end{aligned} \tag{C.11}
$$

Combining the two parts together we have the conditional posterior density for κ_t, given all the other parameters

$$p(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) \propto p(\boldsymbol{D}_1|\kappa_1, \boldsymbol{\alpha}, \boldsymbol{\beta})p(\kappa_1|\sigma_\kappa^2, \theta) \times \prod_{t=2}^{n_y} p(\boldsymbol{D}_t|\kappa_t, \boldsymbol{\alpha}, \boldsymbol{\beta})p(\kappa_t|\kappa_{t-1}, \sigma_\kappa^2, \theta) \quad \text{(C.12)}$$

Then the conditional log-posterior for κ_t is

$$\log p(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) = \text{Constant} + \log p(\boldsymbol{D}_1|\kappa_1, \boldsymbol{\alpha}, \boldsymbol{\beta}) + \log p(\kappa_1|\sigma_\kappa^2, \theta) + \sum_{t=2}^{n_y} \{\log p(\boldsymbol{D}_t|\kappa_t, \boldsymbol{\alpha}, \boldsymbol{\beta}) + \log p(\kappa_t|\kappa_{t-1}, \sigma_\kappa^2, \theta)\}$$

Note that the posterior can be simplified with respect to the value of t.

For $t = t_{n_y}$

$$p(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) \propto p(\boldsymbol{D}_{n_y}|\kappa_{n_y}, \boldsymbol{\alpha}, \boldsymbol{\beta})p(\kappa_{n_y}|\kappa_{n_y-1}, \sigma_\kappa^2, \theta) \quad \text{(C.13)}$$

$$\log p(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) = \text{Constant} + \sum_{x=x_1}^{x_{n_a}} \{D_{n_y,x} \log m_{n_y,x} - E_{n_y,x} m_{n_y,x}\} - \frac{(\kappa_{n_y} - \kappa_{n_y-1} - \theta)^2}{2\sigma_\kappa^2} \quad \text{(C.14)}$$

For the rest of $t > 1$,

$$p(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) \propto p(\boldsymbol{D}_t|\kappa_t, \boldsymbol{\alpha}, \boldsymbol{\beta})p(\kappa_t|\kappa_{t-1}, \sigma_\kappa^2, \theta)p(\kappa_{t+1}|\kappa_t, \sigma_\kappa^2, \theta) \quad \text{(C.15)}$$

$$\log p(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) = \text{Constant} + \sum_{x=x_1}^{x_{n_a}} \{D_{t,x} \log m_{t,x} - E_{t,x} m_{t,x}\} - \frac{(\kappa_t - \kappa_{t-1} - \theta)^2}{2\sigma_\kappa^2} - \frac{(\kappa_{t+1} - \kappa_t - \theta)^2}{2\sigma_\kappa^2} \quad \text{(C.16)}$$

Denote as $\kappa_{t,i}$ the i^{th} iteration of κ_t, and about to update to the $i+1$ iteration, given we have already updated the $i+1$ iteration for the previous $\kappa_1, \ldots, \kappa_{t-1}$, denoted as $\kappa_{1,i+1}, \ldots, \kappa_{t-1,i+1}$. The steps for updating are as follows

1 Generate a candidate $\hat{\kappa}_{t,i}$ from a proposal distribution $N(\kappa_{t,i}, \sigma_t^2)$.

2 Calculate the probability of acceptance defined as

$$\alpha(\kappa_{t,i}, \hat{\kappa}_{t,i}) = \min\left(1, \frac{p(\hat{\kappa}_{t,i}|\boldsymbol{\kappa}_{-t,i}, \boldsymbol{D}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \theta)}{p(\kappa_{t,i}|\boldsymbol{\kappa}_{-t,i}, \boldsymbol{D}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \theta)}\right) \tag{C.17}$$

where $\boldsymbol{\kappa}_{-t,i} = (\kappa_{1,i+1}, ..., \kappa_{t-1,i+1}, \kappa_{t+1,i}, ..., \kappa_{n_y,i})^T$. Note we denote as α the probability as standard practice.

3 Draw a sample u from $U(0,1)$. If $u \leq \alpha(\kappa_{t,i}, \hat{\kappa}_{t,i})$, we set $\kappa_{t,i+1} = \hat{\kappa}_{t,i}$. If $u > \alpha(\kappa_{t,i}, \hat{\kappa}_{t,i})$, the Markov Chain does not move and we set $\kappa_{t,i+1} = \kappa_{t,i}$.

Note that we set $\kappa_1 = 0$ and $\theta = -1$ to replace the usual identifiability constraints for a more convenient MCMC process.

Age Effect $\boldsymbol{\beta}$

Czado et al. (2005) modelled the age effect $\boldsymbol{\beta} = (\beta_{x_1}, \ldots, \beta_{x_{n_a}})^T$ with a multi-variate normal distribution such that $\boldsymbol{\beta} \sim MVN(\boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I}_{n_a})$ by assuming no mortality improvements for the population under study. This assumption is reasonable since the data is expected to appropriately transform the prior distribution in case that the improvement does exist. Similarly in this example, a multi-variate normal distribution with slight modification is employed . More specifically,

$$\boldsymbol{\beta} \sim MVN(\boldsymbol{0}, \sigma_\beta^2 \boldsymbol{R}), \tag{C.18}$$

where $\sigma_\beta^2 \boldsymbol{R}$ is a n_a dimensioned co-variance matrix. In particular, each element of

$\boldsymbol{R} = \{r_{i,j}\}$ for $i = j = 1, \ldots, n_a$ can be written as

$$r_{i,j} = \rho^{|i-j|}, \text{for } |\rho| < 1 \tag{C.19}$$

to ensure that the correlation between β_i and β_j is smaller for a larger value of $i-j$. The prior distribution for the hyper-parameter $\sigma_\beta^2 \sim$ Inv- Gamma(a_β, b_β), where $a_\beta = 2.0001$ and $b_\beta = (a_\beta - 1)\hat{\sigma}_\beta^2$. The $\hat{\sigma}_\beta^2$ is the variance of $\hat{\boldsymbol{\beta}}$.

Define

$$\boldsymbol{\beta}_{-x} = (\beta_1, ..., \beta_{x-1}, \beta_{x+1}, ..., \beta_{n_a})^T \tag{C.20}$$

and

$$\boldsymbol{D}_{-x} = (\boldsymbol{D}_1, ..., \boldsymbol{D}_{x-1}, \boldsymbol{D}_{x+1}, ..., \boldsymbol{D}_{n_a}) \tag{C.21}$$

where $\boldsymbol{D}_x = (D_{x,t_1}, ..., D_{x,t_{n_y}})^T$. Similarly, we derive the conditional posterior for β_x as

$$\begin{aligned} p(\beta_x|\boldsymbol{\beta}_{-x}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) &\propto p(\boldsymbol{D}_x|\beta_x, \boldsymbol{\alpha}, \boldsymbol{\kappa}) p(\beta_x|\boldsymbol{\beta}_{-x}, \sigma_\beta, \rho) \\ &\propto p(\boldsymbol{D}_x|\beta_x, \boldsymbol{\alpha}, \boldsymbol{\kappa}) p(\boldsymbol{\beta}, \sigma_\beta, \rho) \\ &\propto p(\boldsymbol{D}_x|\beta_x, \boldsymbol{\alpha}, \boldsymbol{\kappa}) p(\boldsymbol{\beta}|\sigma_\beta, \rho) \end{aligned} \tag{C.22}$$

$$\log p(\beta_x|\boldsymbol{\beta}_{-x}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) = \text{Constant} + \sum_{t=1}^{n_y} (D_{t,x} \log m_{t,x} - E_{t,x} m_{t,x}) \tag{C.23}$$

$$-\frac{1}{2}\boldsymbol{\beta}^T(\sigma_\beta^2 \boldsymbol{R})^{-1}\boldsymbol{\beta} - \frac{1}{2}\log(|\sigma_\beta^2 \boldsymbol{R}|) \tag{C.24}$$

given that $\boldsymbol{D}$ is independent and a function of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}$ only.

Denote as $\beta_{x,i}$ the i^{th} iteration of β_x and we are about to update to the $i+1$ iteration, given we have already update the $i+1$ iteration for the previous $\beta_1, ..., \beta_{x-1}$, denoted as $\beta_{1,i+1}, ..., \beta_{x-1,i+1}$. The steps for updating are as follows

1 Draw a candidate $\hat{\beta}_{x,i}$ from a proposal distribution $N(\beta_{x,i}, \sigma_x^2)$.

2 Calculate the probability of acceptance defined as

$$\alpha(\beta_{x,i}, \hat{\beta}_{x,i}) = \min\left(1, \frac{p(\hat{\beta}_{x,i}|\boldsymbol{\beta}_{-x,i}, \boldsymbol{D}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \theta)}{p(\beta_{x,i}|\boldsymbol{\beta}_{-x,i}, \boldsymbol{D}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \theta)}\right) \tag{C.25}$$

where $\boldsymbol{\beta}_{-x,i} = (\beta_{1,i+1}, ..., \beta_{x-1,i+1}, \beta_{x+1,i}, ..., \beta_{n_a,i})^T$

3 Draw a sample u from $U(0,1)$, If $u \leq \alpha(\beta_{x,i}, \hat{\beta}_{x,i})$, we set $\beta_{x,i+1} = \hat{\beta}_{x,i}$, else we have $\beta_{x,i+1} = \beta_{x,i}$

Age Effect $\boldsymbol{\alpha}$

Czado et al. (2005) modelled the prior of α_x as

$$e_x \sim \text{Gamma}(a_x, b_x), \tag{C.26}$$

where $e_x = \exp \alpha_x$. The constants are assumed to be known and fixed to $b_x = 0.0001$ and $a_x = b_x \exp \hat{\alpha}_x$.

Since $\boldsymbol{e}$ only depends on $\boldsymbol{\kappa}, \boldsymbol{\beta}, \boldsymbol{D}$, the conditional posterior of $e_x = \exp \alpha_x$ given all the other parameters is

$$\begin{aligned} p(e_x|\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \theta) &= p(e_x|\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}) \\ &\propto p(\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}|e_x)p(e_x) \end{aligned} \tag{C.27}$$

Note that

$$p(\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}|\boldsymbol{e}) = L_1(\boldsymbol{e}|\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}) \tag{C.28}$$

if we consider L_1 is a likelihood function of $\boldsymbol{e}$ only, where

$$\begin{aligned} L_1(\boldsymbol{e}|\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{D}) &= \prod_{t,x} \exp\left(D_{t,x}(\alpha_x+\beta_x\kappa_t) - E_{t,x}\exp(\alpha_x+\beta_x\kappa_t)\right) \\ &= \prod_{x} \exp\left(\sum_t D_{t,x}(\alpha_x+\beta_x\kappa_t) - \sum_t E_{t,x}\exp(\alpha_x+\beta_x\kappa_t)\right) \\ &\propto \prod_x e_x^{D_{x\cdot}} \exp\left(-e_x c_x\right) \end{aligned} \tag{C.29}$$

where

$$c_x = \sum_t E_{t,x}\exp(\beta_x\kappa_t);\ D_{x\cdot} = \sum_t D_{t,x} \tag{C.30}$$

Thus the conditional posterior can be re-written as

$$\begin{aligned} p(e_x|\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{D},\sigma_\kappa^2,\sigma_\beta^2,\theta) &\propto p(\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{D}|e_x)p(e_x) \\ &\propto e_x^{D_{x\cdot}}\exp(-e_x c_x)e_x^{a_x-1}\exp\left(-\frac{e_x}{b_x}\right) \\ &\propto e_x^{D_{x\cdot}+a_x-1}\exp\left(-(c_x+\frac{1}{b_x})e_x\right) \end{aligned} \tag{C.31}$$

which also follows gamma distribution

$$e_x|\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{D},\sigma_\kappa^2,\sigma_\beta^2,\theta \sim Gamma\left(D_{x\cdot}+a_x, \frac{1}{c_x+\frac{1}{b_x}}\right) \tag{C.32}$$

Thus we draw e_x from above distribution for each x.

Simulate $\boldsymbol{\beta}$, σ_β^2 and Fix the Rest Parameters

The first experiment is to update only the $\boldsymbol{\beta}$ and σ_β^2. The other parameters are fixed as follows:

- $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$

- $\boldsymbol{\kappa} = \hat{\boldsymbol{\kappa}}$

- $\sigma_\kappa^2 = \hat{\sigma}_\kappa^2$

- $\rho = 0,\ 0.95$.

The conditional posterior distribution of σ_β^2 on the other parameters is

$$p(\sigma_\beta^2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \theta) \propto p(\boldsymbol{\beta}|\sigma_\beta^2)p(\sigma_\beta^2) \tag{C.33}$$

$$p(\boldsymbol{\beta}|\sigma_\beta^2) \propto |\sigma_\beta^2 \boldsymbol{R}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^T(\sigma_\beta^2\boldsymbol{R})^{-1}\boldsymbol{\beta}\right) \tag{C.34}$$

$$\propto (\sigma_\beta^2)^{-\frac{n_a}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^T(\sigma_\beta^2\boldsymbol{R})^{-1}\boldsymbol{\beta}\right) \tag{C.35}$$

$$p(\sigma_\beta^2) \propto (\sigma_\beta^2)^{-a_\beta - 1} \exp\left(-\frac{b_\beta}{\sigma_\beta^2}\right) \tag{C.36}$$

$$p(\sigma_\beta^2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \theta) \propto (\sigma_\beta^2)^{-(a_\beta + \frac{n_a}{2} + 1)} \exp\left(-\frac{1}{\sigma_\beta^2}\left(b_\beta + \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{R}^{-1}\boldsymbol{\beta}\right)\right) \tag{C.37}$$

which is an inverse gamma distribution, denoted as

$$\sigma_\beta^2|\boldsymbol{\Phi}^c(\sigma_\beta^2) \sim \text{Inverse gamma}\left(a_\beta + \frac{n_a}{2}, b_\beta + \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{R}^{-1}\boldsymbol{\beta}\right) \tag{C.38}$$

We plotted the trace plots of the MCMC for selected parameters, given $\rho = 0$ and 0.95 in Figure C.1-C.2. We can see that MH provides a good converged MCMC when only the $\boldsymbol{\beta}$ and σ_β are updated and changing the value of ρ does not have significant effect on the posterior distribution of $\boldsymbol{\beta}$ and σ_β.

Simulate $\boldsymbol{\beta}$, σ_β^2 and ρ and Fix the Rest Parameters

In this experiment, we update $\boldsymbol{\beta}$, σ_β^2 and ρ and leave the rest parameters fixed to the point estimates. The prior density for ρ follows a beta distribution. More specifically,

$$p(\rho) \propto (1-\rho)^{a_\rho}(1+\rho)^{b_\rho}.$$

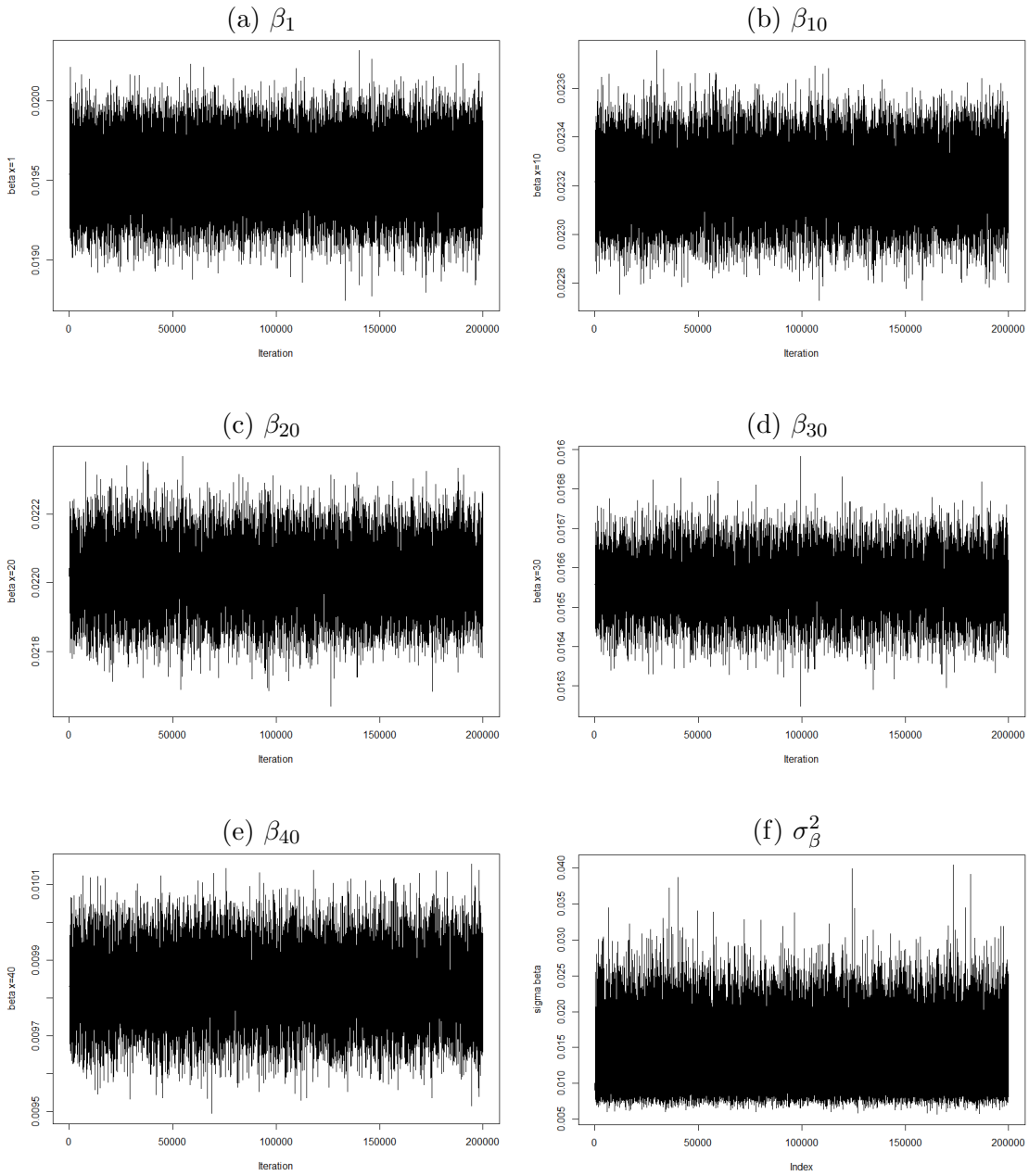


Figure C.1: The trace plots for β_x with selected ages (50, 59, 69, 79, 89) and σ^2_β. Only the $\boldsymbol{\beta}$ and σ^2_β are updated, given $\rho = 0$

The conditional posterior distribution of ρ given the other parameters is

$$p(\rho|\boldsymbol{\Phi}^c(\rho)) \propto |\boldsymbol{R}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^T(\sigma^2_\beta \boldsymbol{R})^{-1}\boldsymbol{\beta}\right)(1-\rho)^{a_\rho}(1+\rho)^{b_\rho}.$$

It is not easy to identify the distribution of $\rho|\boldsymbol{\Phi}^c(\rho)$, we applied the M-H sampling. Denote as ρ_i the iteration i of ρ and about to update to the $i+1$ iteration. The

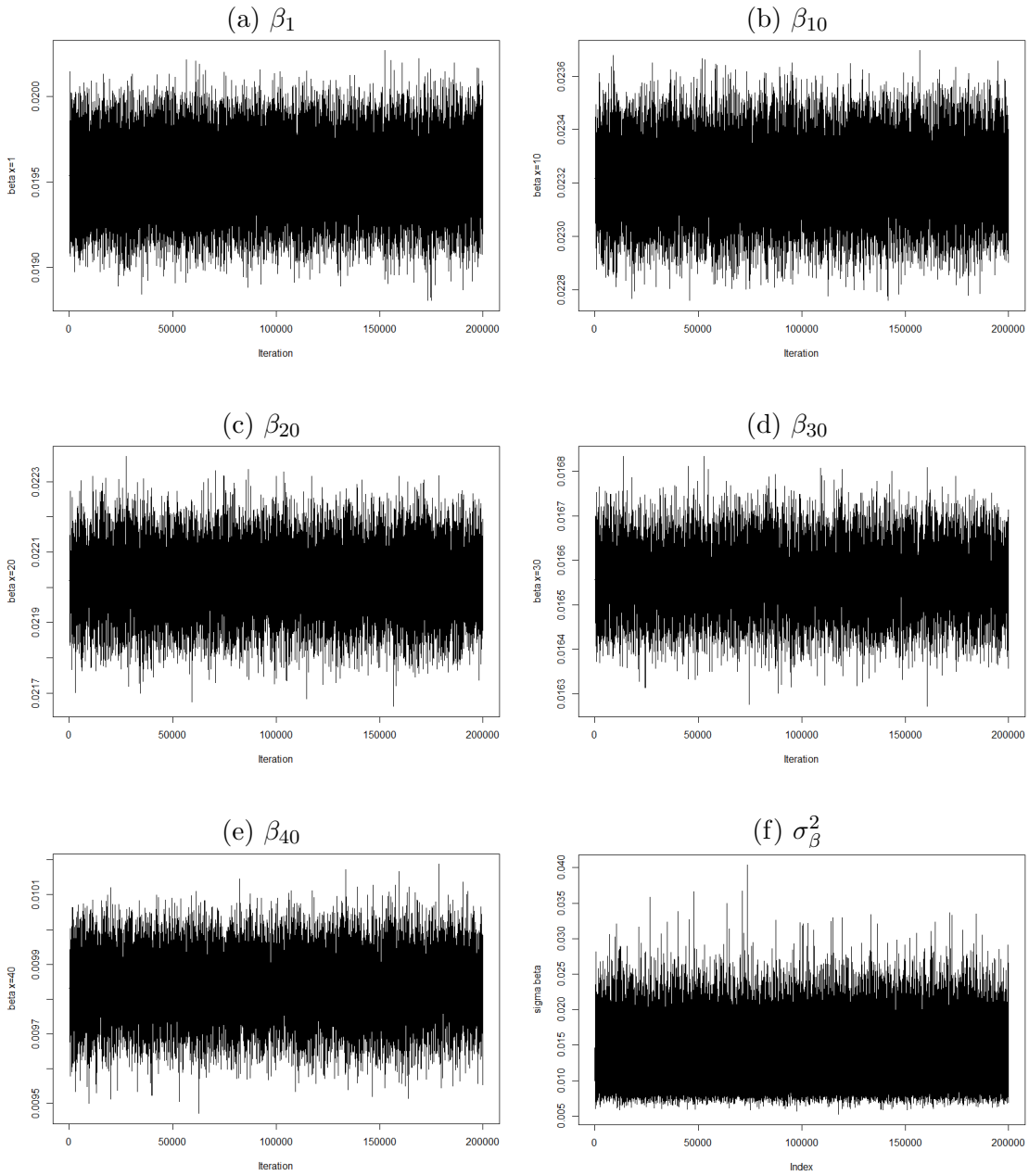


Figure C.2: The trace plots for β_x with selected ages (50, 59, 69, 79, 89) and σ^2_β. Only the $\boldsymbol{\beta}$ and σ^2_β are updated, given $\rho = 0.95$

steps for updating are as follows

1. Generate a candidate $\hat{\rho}_i$ from a proposal distribution $N(\rho_i, \sigma^2_{\rho^*})$ truncated in interval $(-1, 1)$, with density denoted as $q(.|\rho_i)$. Note that the truncated normal proposal distribution is not symmetric.

2. Calculate the probability of acceptance defined as

$$\alpha(\rho_i, \hat{\rho}_i) = \min\left(1, \frac{p(\hat{\rho}_i|\mathbf{\Phi}^c(\rho))q(\rho_i|\hat{\rho}_i)}{p(\rho_i|\mathbf{\Phi}^c(\rho))q(\hat{\rho}_i|\rho_i)}\right) \tag{C.39}$$

3. Draw a sample u from $U(0,1)$. If $u \leq \alpha(\rho_i, \hat{\rho}_i)$, we set $\rho_{i+1} = \hat{\rho}_i$. If $u > \alpha(\rho_i, \hat{\rho}_i)$, the Markov Chain does not move and we set $\rho_{i+1} = \rho_i$

The trace plots for the MCMC of selected parameters, given $a_\rho = b_\rho = 1$, is plotted in Figure C.3. The following adjustments are then made, such that the proposal distribution for ρ is altered from the truncated normal distribution in (-1,1) to a regular normal $N(\rho_i, \sigma^2)$ for the ith iteration, where

$$\sigma^2 = (\sigma^2_{\rho^*})^{\exp[\text{sign}(1-\sigma^2_{\rho^*})(\rho_i^2-1)]},$$

so that the closer ρ_i to 1 or -1, the smaller the variance of the proposal distribution. The $\sigma^2_{\rho^*}$ is a chosen constant that can be either greater than one or less than one. Since the variance depends on the size of mean, the probability of acceptance is still defined as

$$\alpha(\rho_i, \hat{\rho}_i) = \min\left(1, \frac{p(\hat{\rho}_i|\Phi^c(\rho))q(\rho_i|\hat{\rho}_i)}{p(\rho_i|\Phi^c(\rho))q(\hat{\rho}_i|\rho_i)}\right) \tag{C.40}$$

since $q(\rho_i|\hat{\rho}_i) \neq q(\hat{\rho}_i|\rho_i)$ given different variance. We set $a_\rho = b_\rho = 2$ and the trace plots for the MCMC are shown in Figure C.4. We can see that MH provides a good convergence for the MCMC and neither altering the proposal distribution or applying a stronger prior for ρ has a significant effect on the posterior distribution.

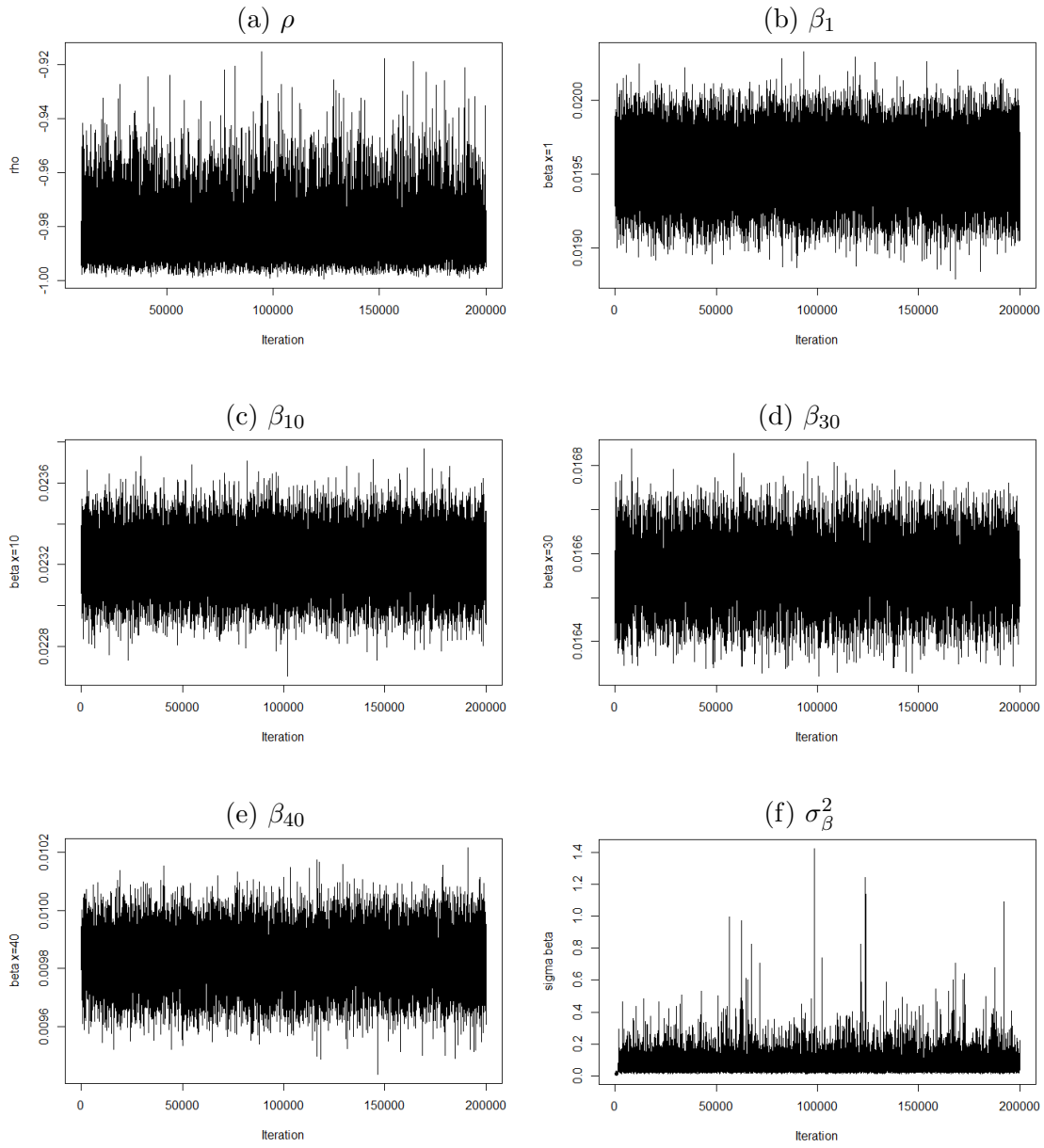


Figure C.3: The trace plots for β_x with selected ages (50, 59, 79, 89), σ_β^2 and ρ. Only the $\boldsymbol{\beta}$, σ_β^2 and ρ are updated, given $a_\rho = b_\rho = 1$. Note that the iteration from 10000 to the end are shown for ρ.

Simulate $\boldsymbol{\beta}$, σ_β^2 and $\boldsymbol{\kappa}$ and Fix the Rest Parameters

Denote as $\boldsymbol{\kappa}(i+1)$ and $\boldsymbol{\beta}(i+1)$ the updated value from iteration i to $i+1$ for $\boldsymbol{\kappa}$ and $\boldsymbol{\beta}$ respectively. We simulate $\Theta \sim U(1-\epsilon, 1+\epsilon)$ and make the following adjustment

$$\tilde{\kappa}_t(i+1) = \Theta\kappa_t(i+1)$$

$$\tilde{\beta}_x(i+1) = \frac{1}{\Theta}\beta_x(i+1) \text{for each} i.$$

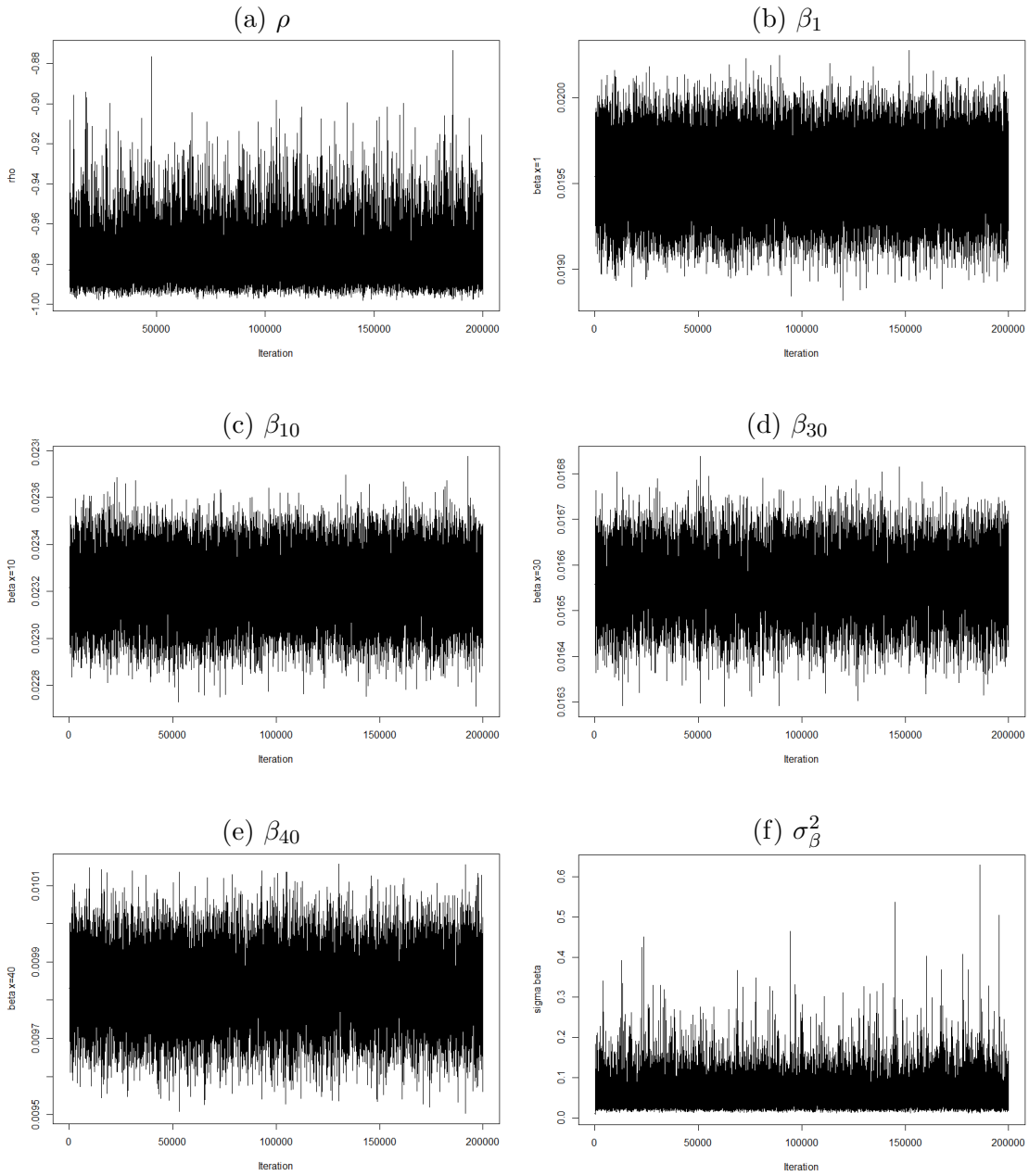


Figure C.4: The trace plots for β_x with selected ages (50, 59, 79, 89), σ^2_β and ρ. Only the $\boldsymbol{\beta}$, σ^2_β and ρ are updated, given $a_\rho = b_\rho = 2$. Note that the iteration from 10000 to the end are shown for ρ. Here we chose a stronger prior for the ρ and the variance of its proposal distribution depends on the value of mean.

In this experiment, we update only the $\boldsymbol{\beta}$, σ^2_β and $\boldsymbol{\kappa}$, given $\rho = 0$ and $\epsilon = 0.01$. The other parameters are then fixed to the respected point estimates. See Figure C.5 for the trace plots of the MCMC. We can see that MH has a problem on achieving the stationary distribution when updating $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ together as they are highly correlated according to the model structure.

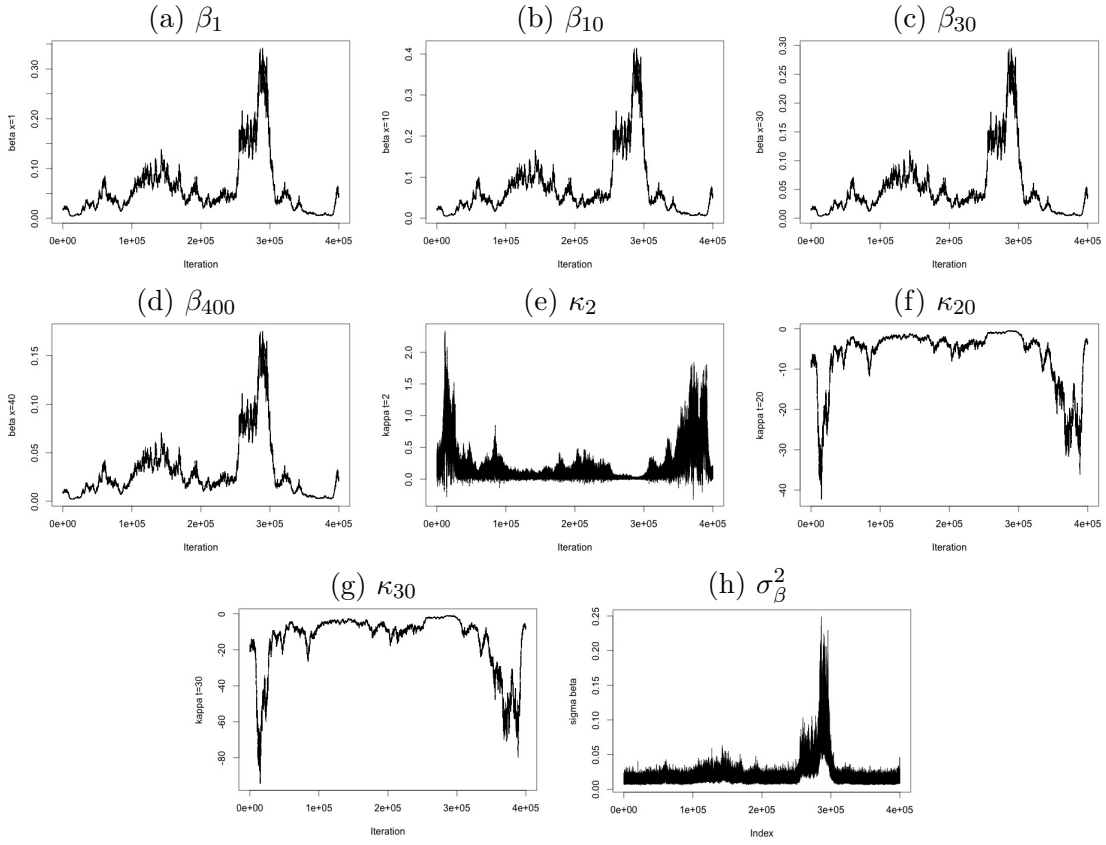


Figure C.5: The pattern of the updated β_x and κ_t, and σ^2_β, as well as the acceptance probability of κ_t and β_x. $\boldsymbol{\beta}$, σ^2_β, $\boldsymbol{\kappa}$ are updated with rest fixed, given $\rho = 0$.

Simulate $\boldsymbol{\beta}$, σ^2_β, $\boldsymbol{\kappa}$ and σ^2_κ

In this experiment, $\boldsymbol{\beta}$, σ^2_β, $\boldsymbol{\kappa}$ and σ^2_κ are updated together. For iteration i, in addition to the normal way of updating κ_t one by one with the MH method, the entire $\boldsymbol{\kappa}$ is then updated given the updated κ_t for $t = t_2, \ldots, t_{n_y}$ by following the steps listed below:

- At iteration i, follow the usual MH to update κ_t, denoted as $\kappa'_t(i)$. Let $\boldsymbol{\kappa}'(i) = \{\kappa'_t(i)\}_{t=t_2,\ldots,t_{n_y}}$ be the vector of updated κ_t.

- Draw a candidate from the proposal distribution $MVN(\boldsymbol{\kappa}'(i), \boldsymbol{V})$, where for $i > 1$,

$$\boldsymbol{V}_{i,j} = \begin{cases} 0, \text{ if } i \neq j, \\ \sigma^2_t (i-1)^g, \text{ if } i = j \end{cases} \tag{C.41}$$

and $\boldsymbol{V}_{1,1} = 0$, $g = 0$, 0.5 and 1. The σ_t^2 is assumed to be independent and fixed to 0.0007 for simplicity purpose.

- Calculate the acceptance probability and determine if the Markov Chain moves. Denote as $\boldsymbol{\kappa}(i+1)$ the value of $\boldsymbol{\kappa}$ at iteration $i+1$.

See Figure C.6-C.8 for the trace plots, given $g = 0$, 1 and 0.5 respectively. We also updated the $\boldsymbol{\kappa}$ entirely without firstly updating each κ_t separately. See Figure C.9-C.11 for $g = 0$, 1 and 0.5 respectively.

In addition to updating $\boldsymbol{\kappa}$ and $\boldsymbol{\beta}$ entirely, the following adjustments are made for each iteration:

- For iteration i, given the updated values $\boldsymbol{\kappa}(i+1)$ and $\boldsymbol{\beta}(i+1)$, simulate $\Theta \sim U(1-\epsilon, \frac{1}{1-\epsilon})$. Note that κ_t and β_x are not individually updated.
- Define as

$$\tilde{\boldsymbol{\kappa}}(i+1) = \Theta\boldsymbol{\kappa}(i+1); \tag{C.42}$$

$$\tilde{\boldsymbol{\beta}}(i+1) = \frac{\boldsymbol{\beta}(i+1)}{\Theta} \tag{C.43}$$

- For iteration $i+1$, draw the candidate from the proposed distribution conditional on $\tilde{\boldsymbol{\kappa}}(i+1)$ and $\tilde{\boldsymbol{\beta}}(i+1)$ respectively.

See Figure C.12-C.14 for $g = 0$, 1 and 0.5 respectively.

We then assume that the entries of the co-variance in $\boldsymbol{V}$ is non-zero and follows that:

$$\boldsymbol{V}_{i,j} = \sqrt{\boldsymbol{V}_{i,i}\boldsymbol{V}_{j,j}}r^{|i-j|}, \text{if } i \neq j$$

where $0 < r < 1$. See FigureC.15-C.17 for the trace plots given $r = 0.99$ and $g = 0$, 1 and 0.5 respectively.

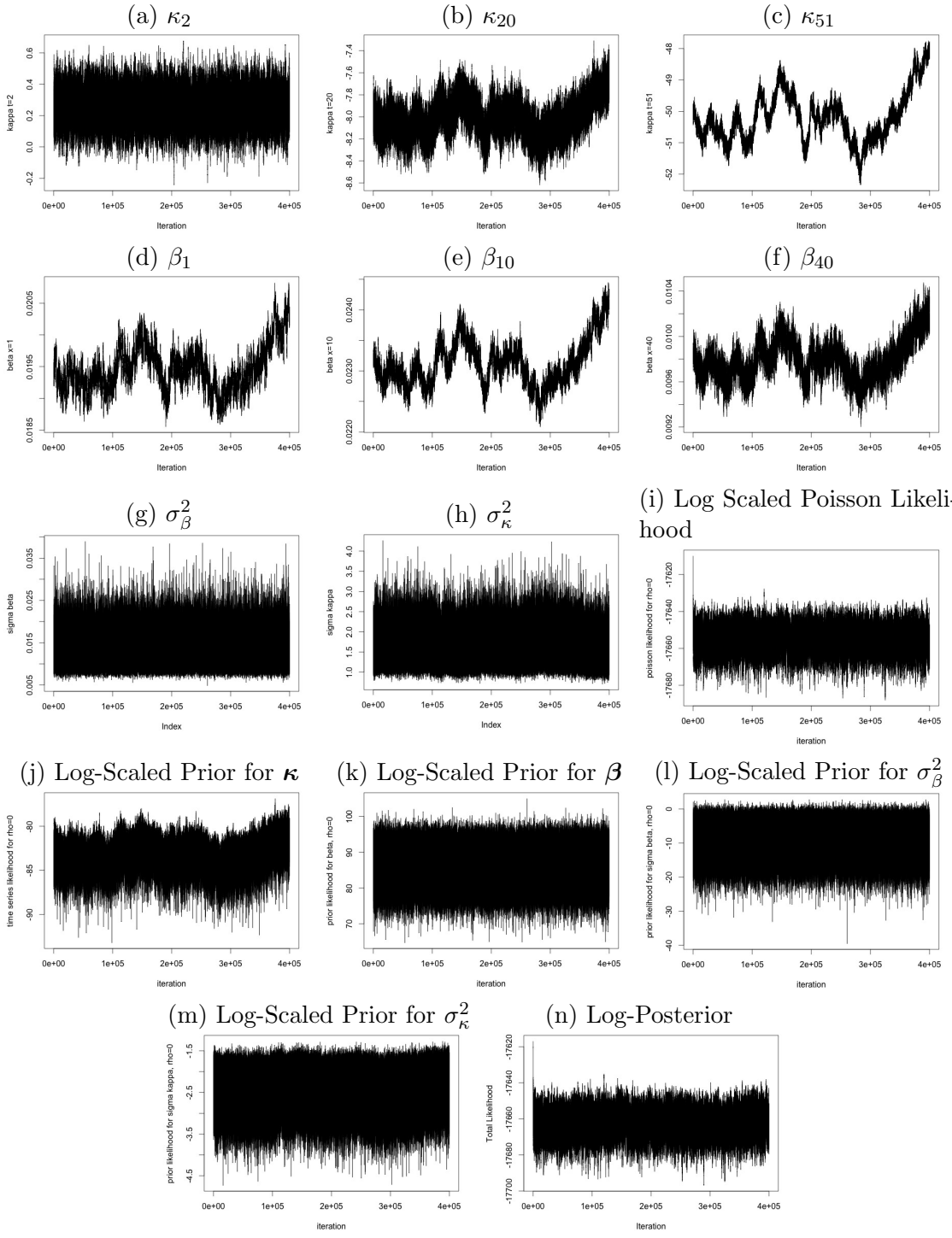


Figure C.6: The trace plots of MCMC for κ_t, β_x of selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0,\ \rho = 0$, while $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \sigma_\beta,\ \sigma_\kappa^2$ are updated. The trajectories of the log-scaled Poisson likelihood and the prior distribution for the $\boldsymbol{\theta}$ are also presented..

Our last experiment for updating $\boldsymbol{\kappa}$, $\boldsymbol{\beta}$, σ_κ and σ_β is to make adjustment to the initial values of $\boldsymbol{\kappa}$ and $\boldsymbol{\beta}$ before the updating started. The adjustments are that

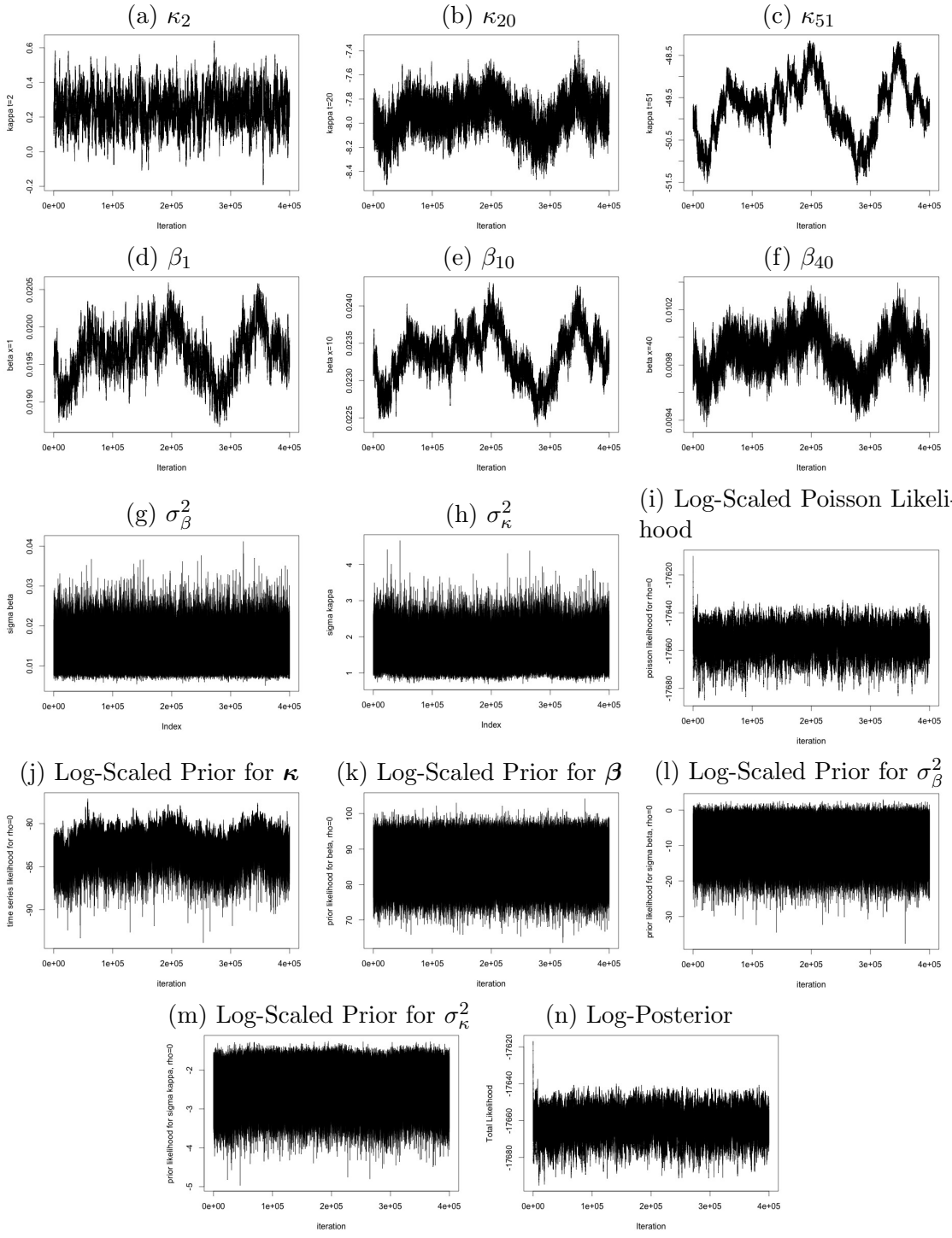


Figure C.7: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ^2_κ, σ^2_β, given $g = 1,\ \rho = 0$, while $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \sigma_\beta,\ \sigma^2_\kappa$ are updated. The trajectories of the log-scaled Poisson likelihood and the prior distribution for the $\boldsymbol{\theta}$ are also presented.

other than starting with the point estimates $\hat{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\beta}}$, denote as $\tilde{\boldsymbol{\kappa}}$ and $\tilde{\boldsymbol{\beta}}$ such that

$$\begin{aligned} \tilde{\boldsymbol{\kappa}} &= \hat{\boldsymbol{\kappa}} b \\ \tilde{\boldsymbol{\beta}} &= \frac{\hat{\boldsymbol{\beta}}}{b}, \end{aligned} \tag{C.44}$$

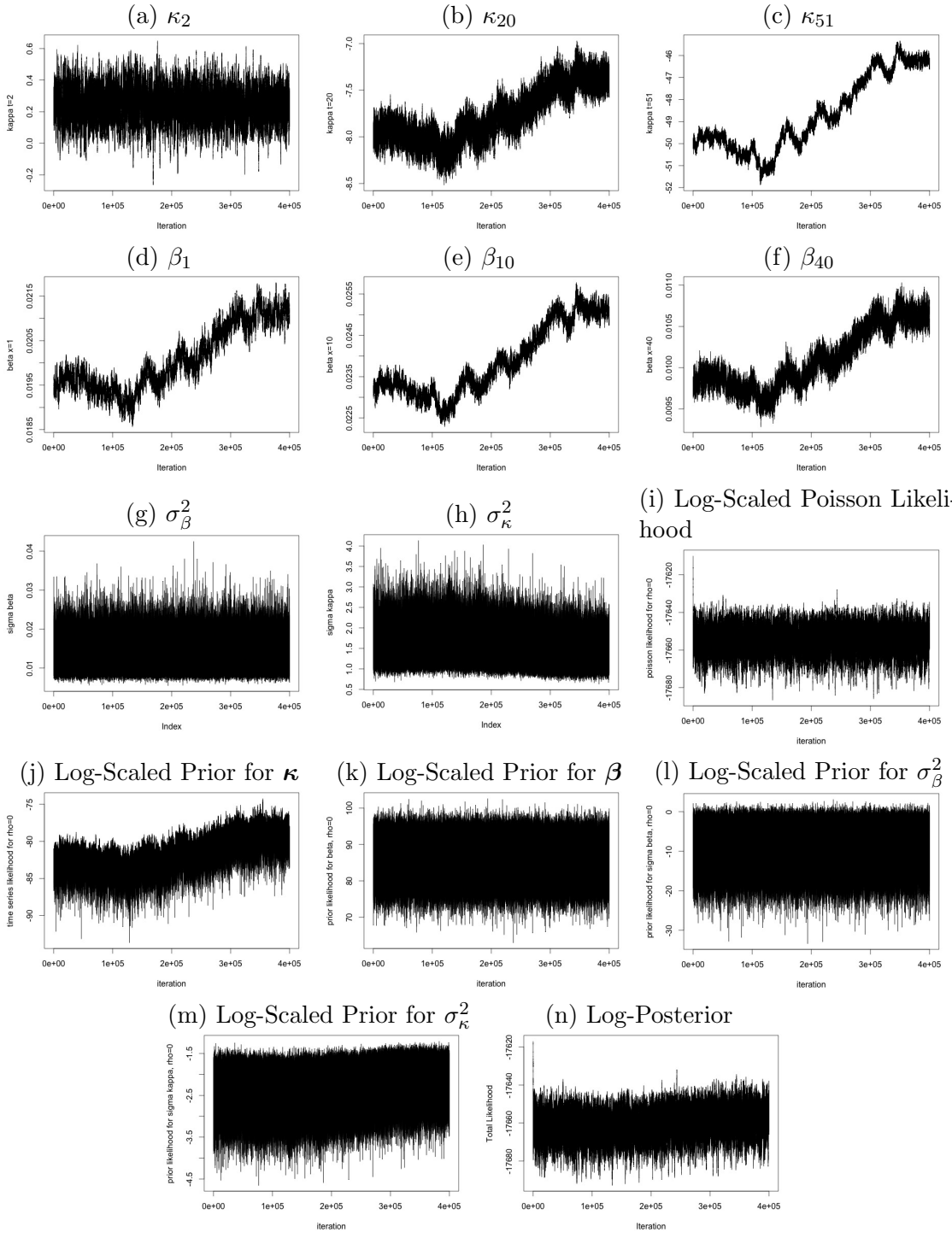


Figure C.8: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ^2_κ, σ^2_β, given $g = 0.5,\ \rho = 0$, while $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \sigma_\beta,\ \sigma^2_\kappa$ are updated. The trajectories of the log-scaled Poisson likelihood, the prior and the posterior for $\boldsymbol{\theta}$ are also presented.

where $b = -46/\hat{\kappa}_{t_{n_y}}$. We fixed the constant $g = 0.5$, $\rho = 0$. We ran 4 million iterations and saved every 100th iteration. The trace plots are shown in Figure C.18 and C.19 for $r = 0$ and 0.9 respectively.

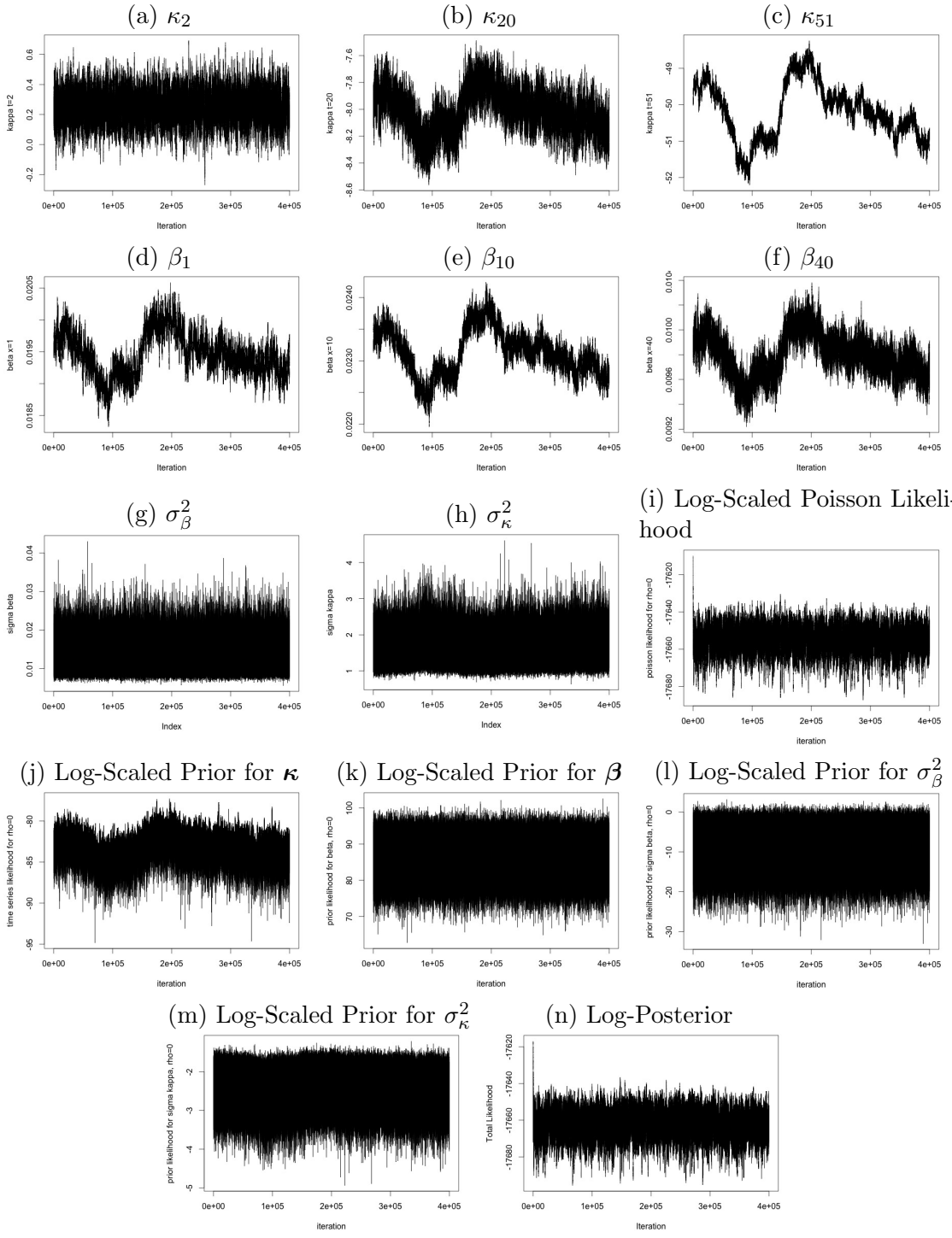


Figure C.9: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0,\ \rho = 0$ while $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \sigma_\beta,\ \sigma_\kappa^2$ are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that κ_t is not updated individually for $t = t_1, \ldots, t_{n_y}$.

Simulate All the Parameters with Hamiltonian Monte Carlo

We can see that the efficiency of the MH algorithm is low for achieving convergence if the correlated $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ are updated due to the random walk behaviour. This

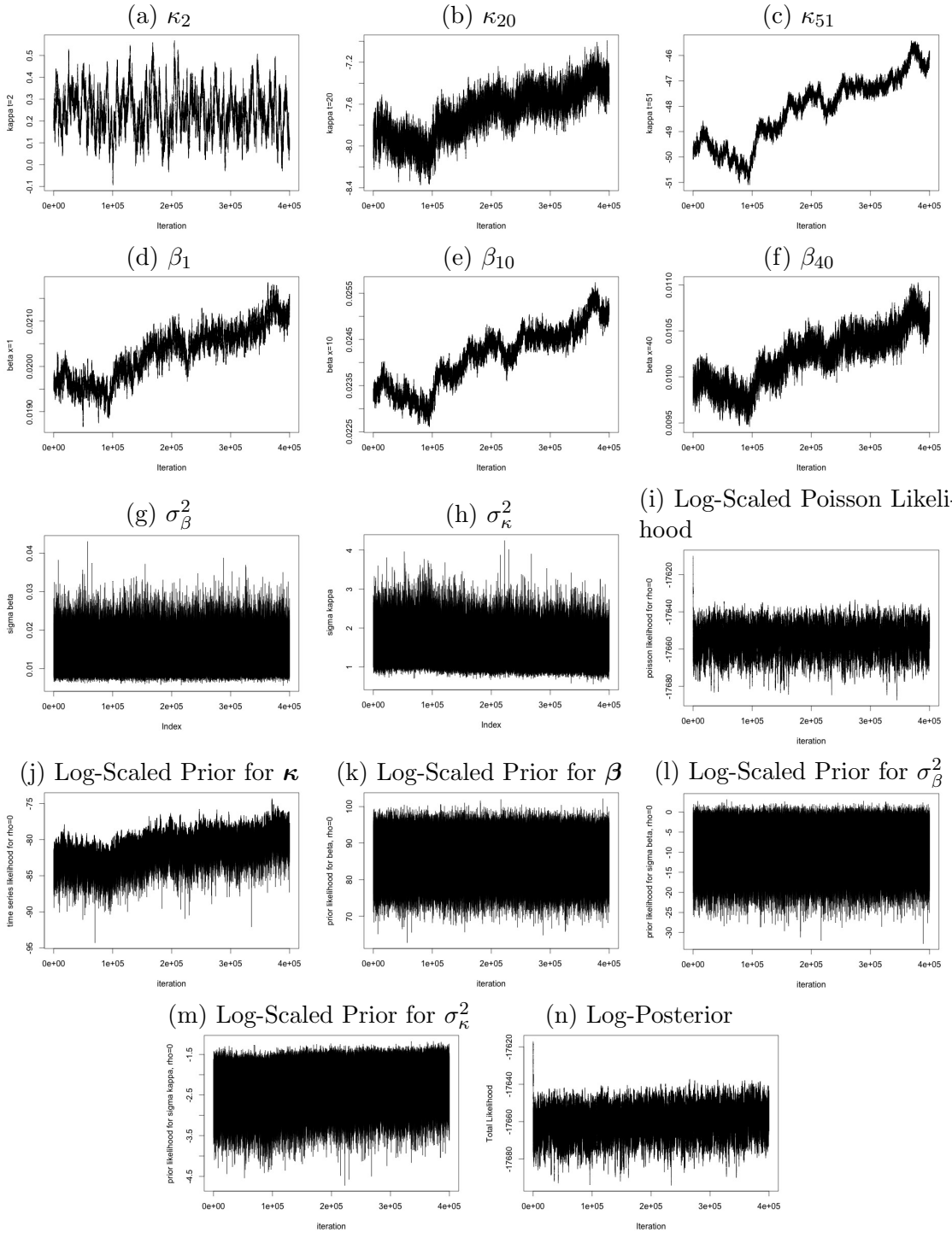


Figure C.10:]
The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 1,\ \rho = 0$, while $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \sigma_\beta,\ \sigma_\kappa^2$ are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that κ_t is not updated individually for $t = t_1, \ldots, t_{n_y}$

section updates all the parameters $\boldsymbol{\theta}$ with the Hamiltonian Monte Carlo. The prior distribution for $\boldsymbol{\theta}$ is the same with MH method without any adjustment. Recall:

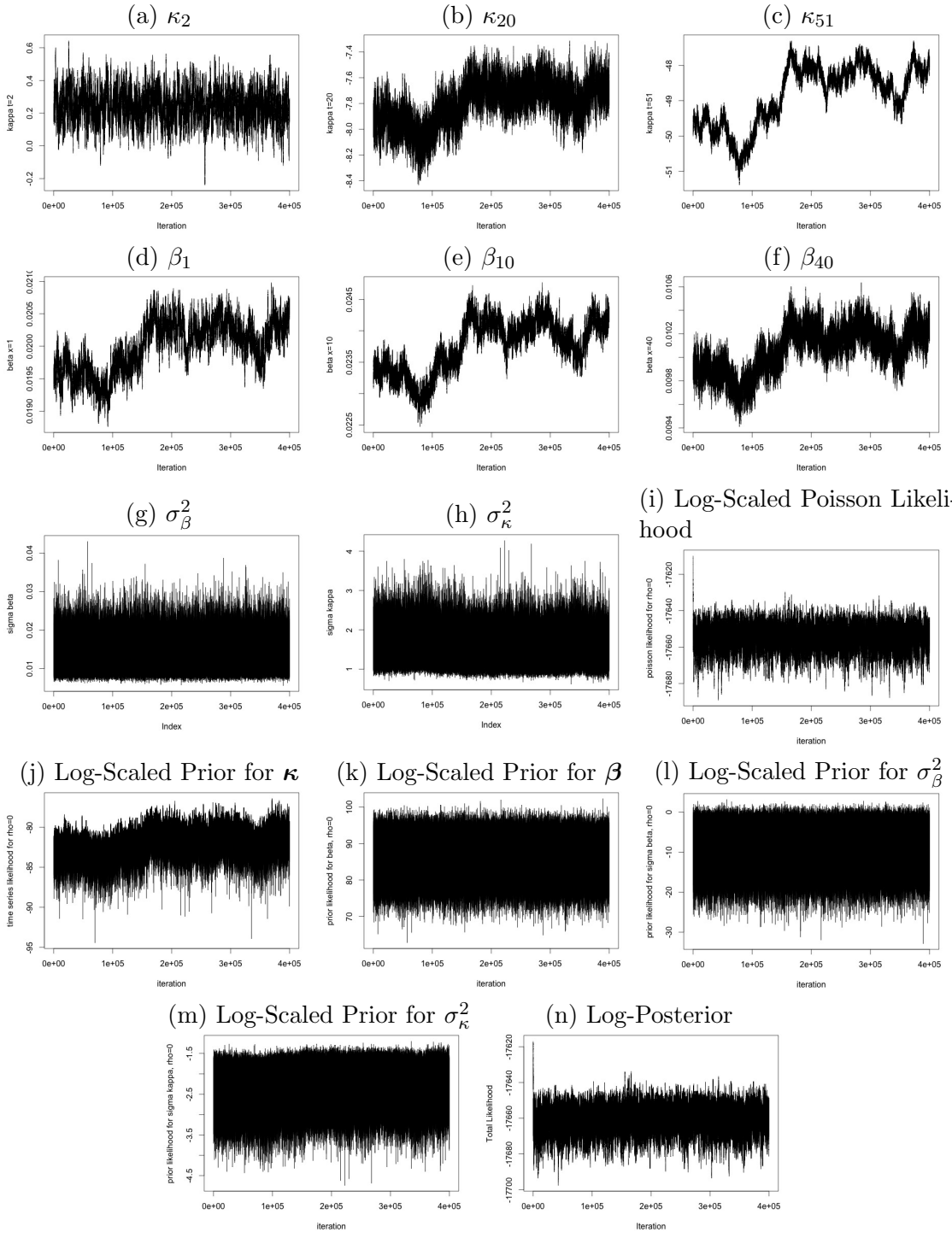


Figure C.11: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0.5,\ \rho = 0$, while $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \sigma_\beta,\ \sigma_\kappa^2$ are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not updated individually for $t = t_1, \ldots, t_{n_y}$.

- $\kappa_t - \kappa_{t-1} \sim N(\theta, \sigma_\kappa^2)$,

- $\boldsymbol{\beta} \sim MVN(\mathbf{0}, \sigma_\beta^2 \boldsymbol{R})$,

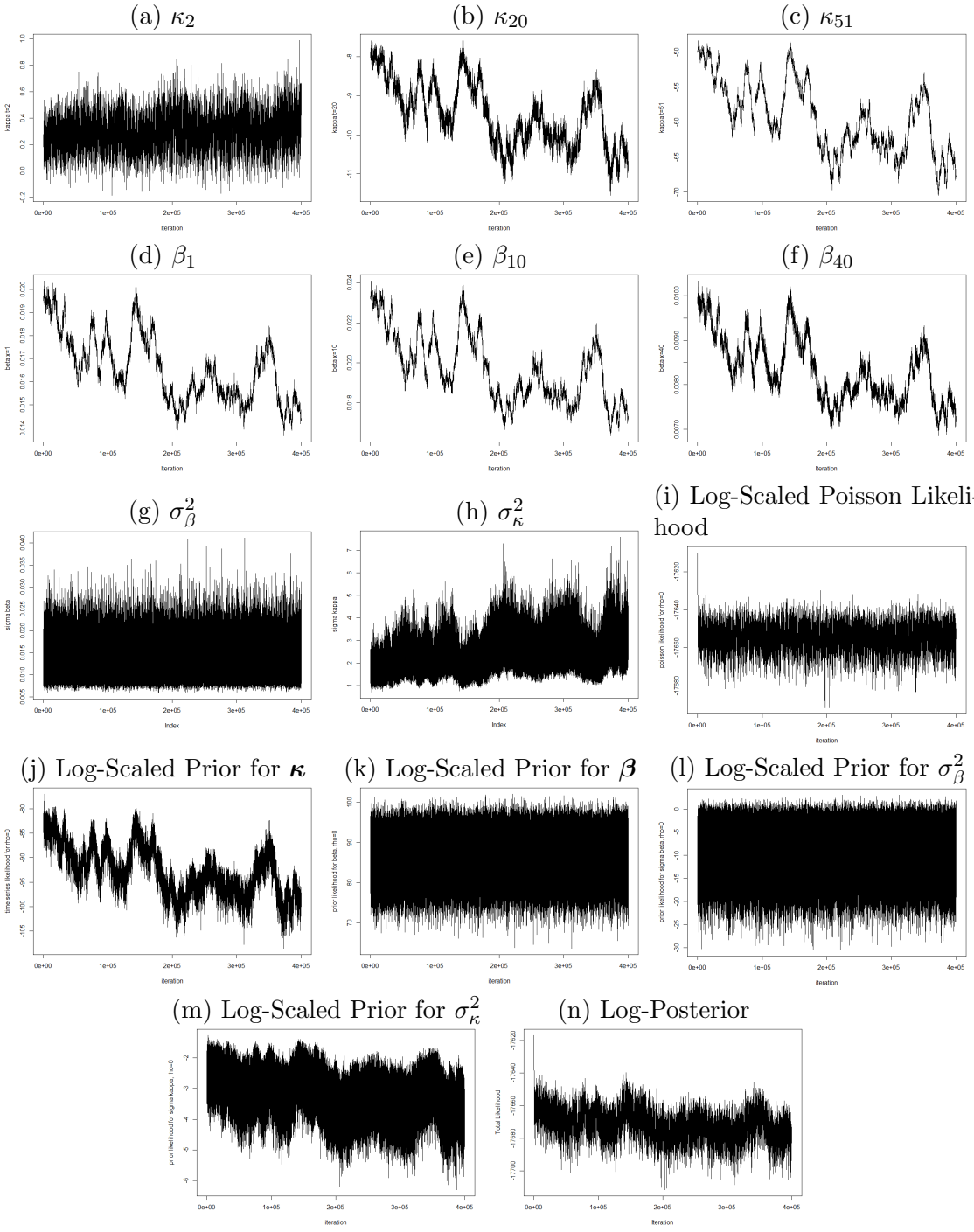


Figure C.12: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0,\ \rho = 0$, while $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \sigma_\beta,\ \sigma_\kappa^2$ are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not individually updated for $t = t_1, \ldots, t_{n_y}$ and the adjustment to κ_t and β_x is adopted.

- $\sigma_\beta^2 \sim$ Inv-Gamma(a_β, b_β),

- $\rho \sim (1-\rho)^{a_\rho}(1+\rho)^{b_\rho}$,

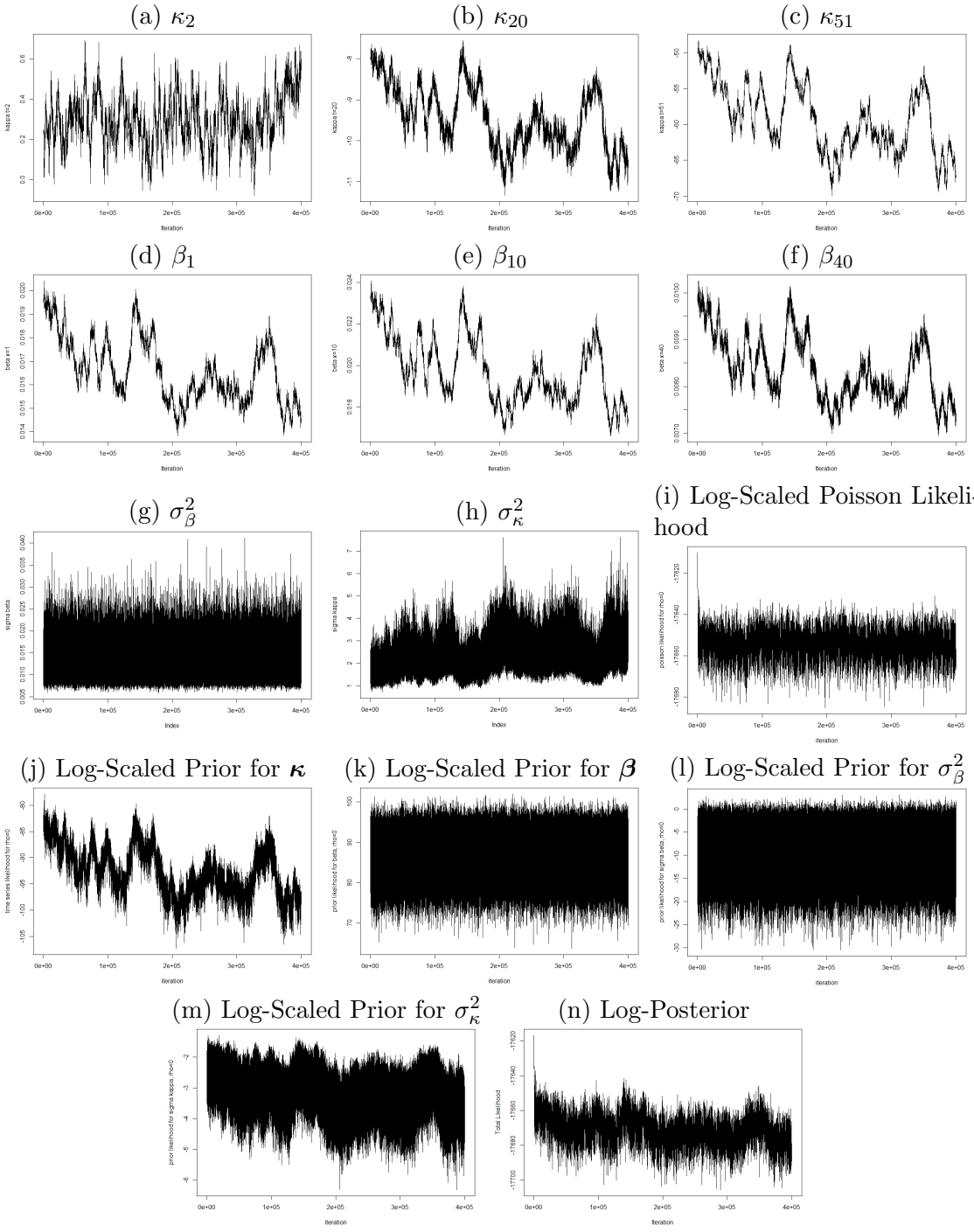


Figure C.13: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 1,\ \rho = 0$, while $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \sigma_\beta,\ \sigma_\kappa^2$ are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not individually updated for $t = t_1, \ldots, t_{n_y}$ and the adjustment to κ_t and β_x is adopted.

- $e_x \sim \text{Gamma}(a_x, b_x)$,

- $\sigma_\kappa^2 \sim \text{Inv-Gamma}(a_\kappa, b_\kappa)$,

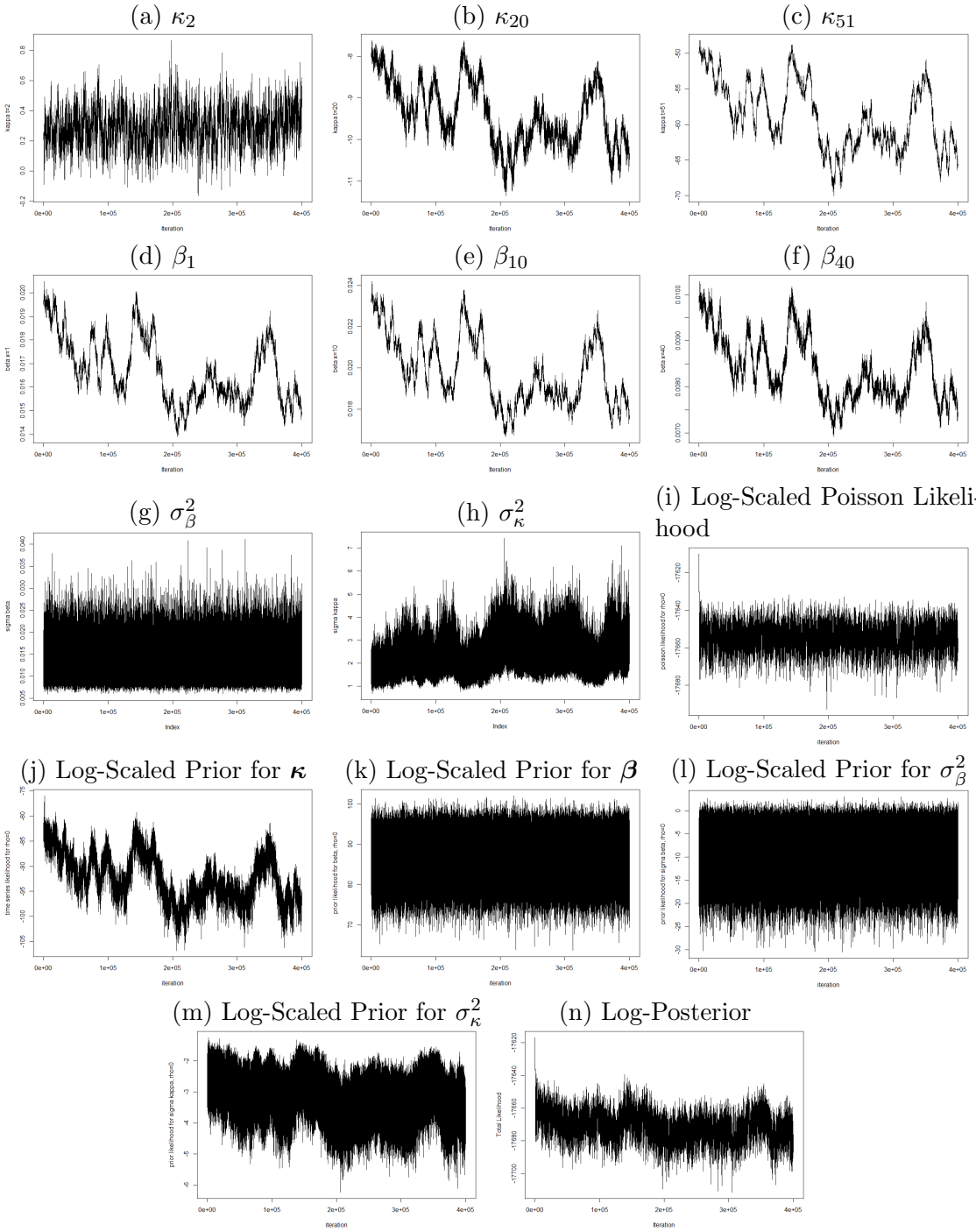


Figure C.14: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0.5$, $\rho = 0$, while $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, σ_β, σ_κ^2 are updated. The trajectories of the log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not individually updated for $t = t_1, \ldots, t_{n_y}$ and the adjustment to κ_t and β_x is adopted.

where $\theta = -1$, $\kappa_{t_1} = 0$, $\boldsymbol{R}_{i,j} = \rho^{i-j}$, $|\rho| < 1$, $e_x = \exp(\alpha_x)$. Similarly with the MH method, the latent parameters start from the respected point estimates and initial value for the hyper-parameters is the point estimate given the corresponding

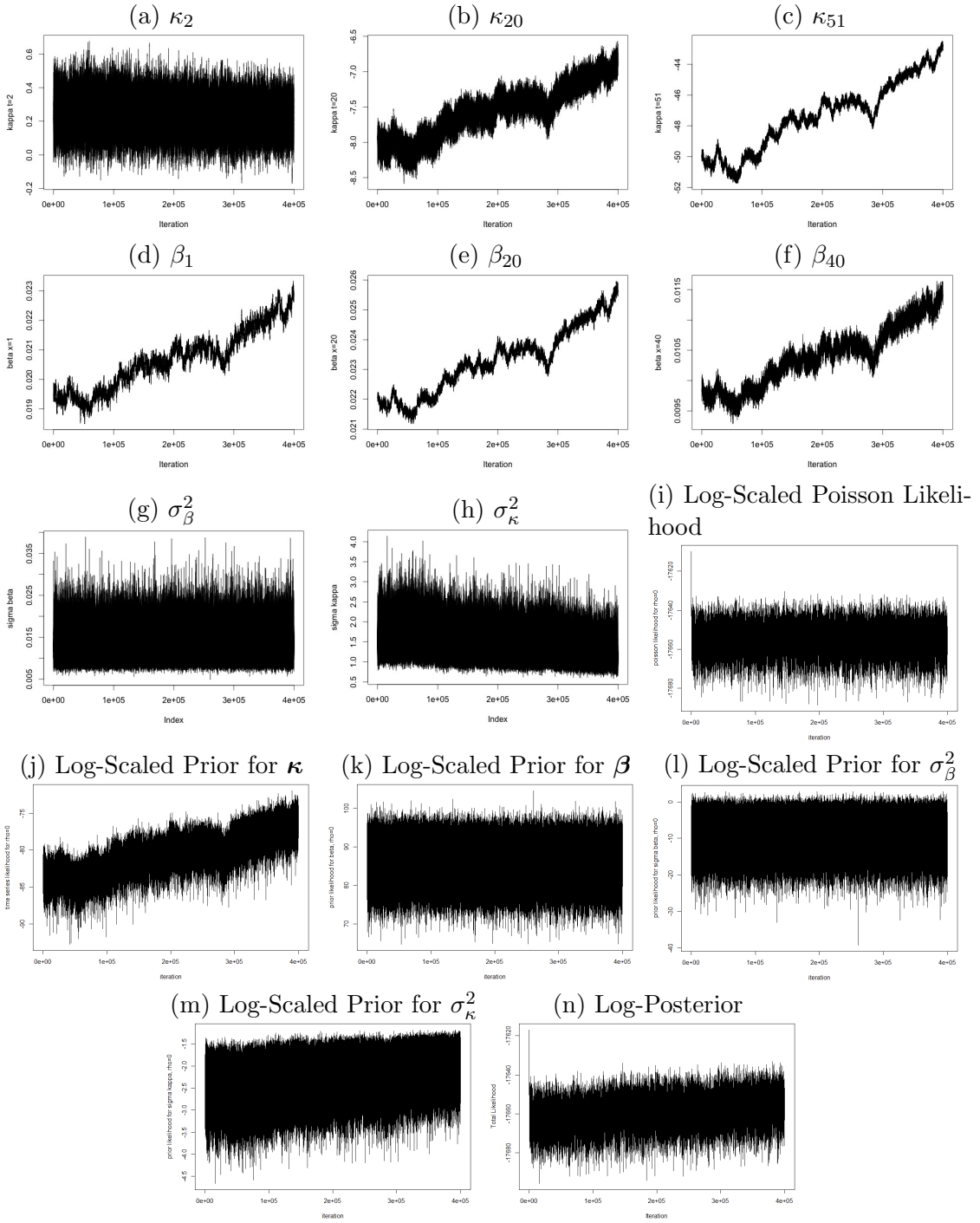


Figure C.15: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0$, $\rho = 0$ and non-zero co-variance, while $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, σ_β, σ_κ^2 are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not individually updated for $t = t_1, \ldots, t_{n_y}$ and the adjustment to κ_t and β_x is adopted.

latent parameter estimation. The setting of the parameters for the hyper-parameter densities are

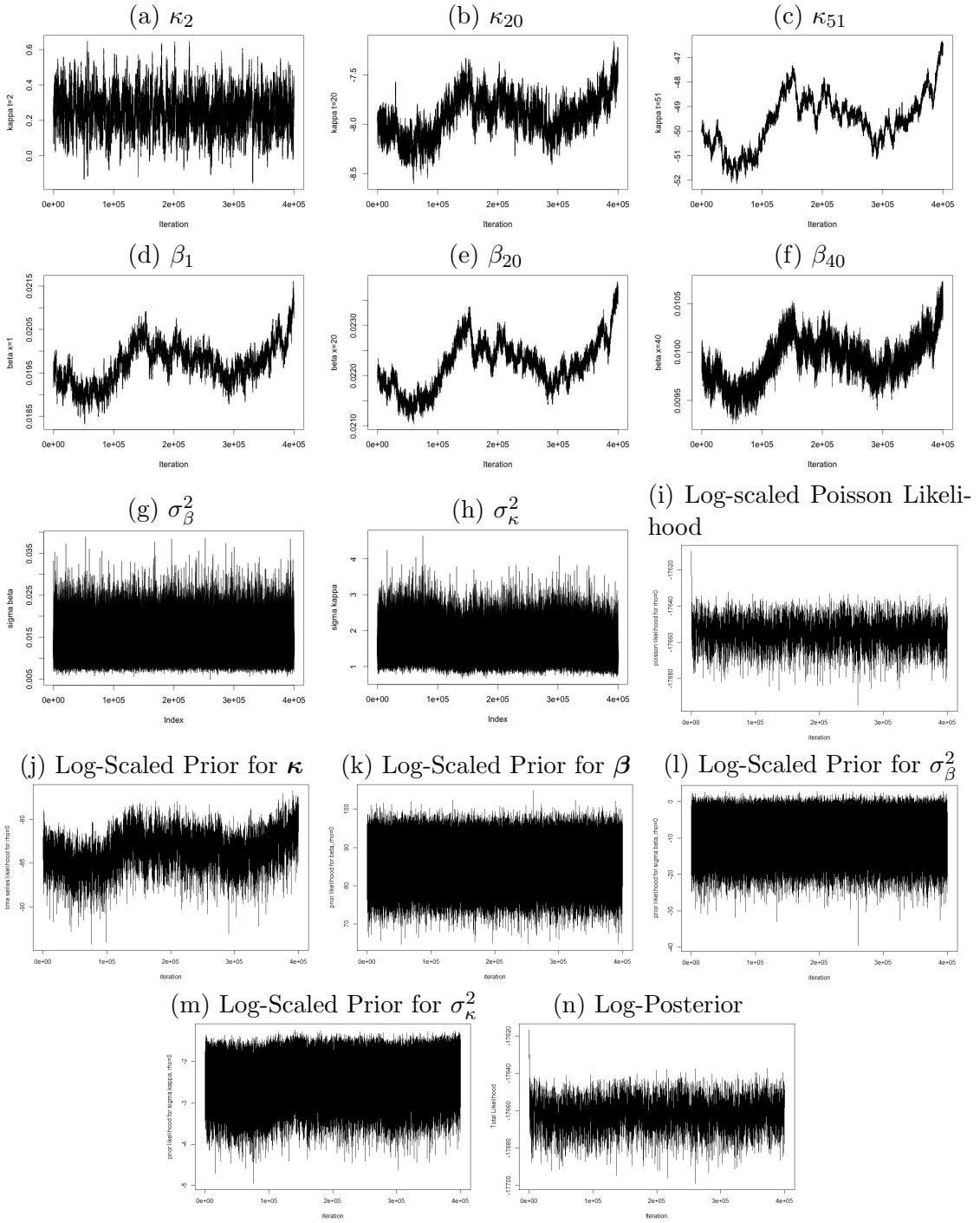


Figure C.16: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 1$, $\rho = 0$ and non-zero co-variance, while $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, σ_β, σ_κ^2 are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not individually updated for $t = t_1, \ldots, t_{n_y}$ and the adjustment to κ_t and β_x is adopted.

- $a_\rho = b_\rho = 1$,

- $a_\kappa = a_\beta = 2.0001$,

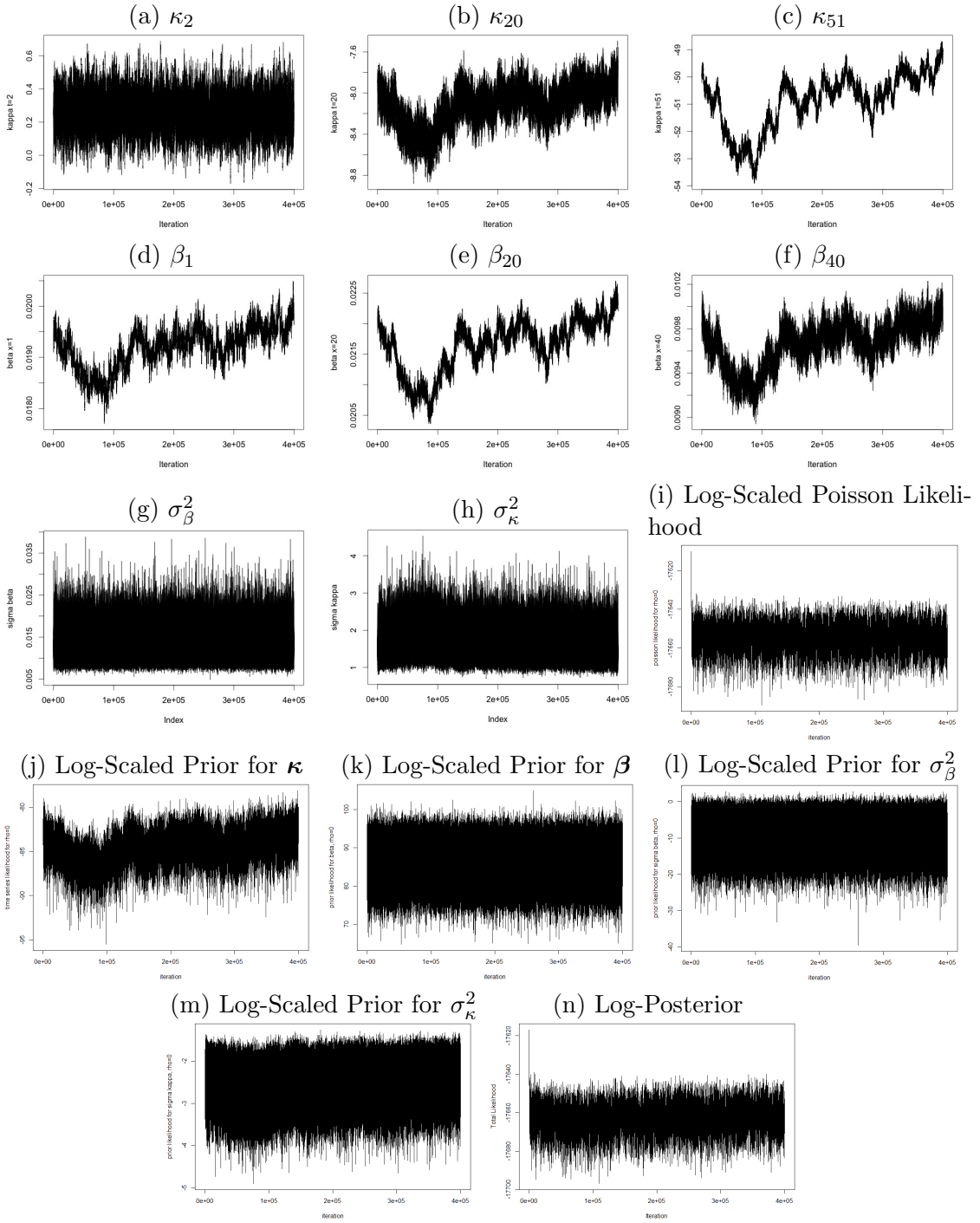


Figure C.17: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0.5$, $\rho = 0$ and non-zero co-variance, while $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, σ_β, σ_κ^2 are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not individually updated for $t = t_1, \ldots, t_{n_y}$ and the adjustment to κ_t and β_x is adopted.

- $b_{\boldsymbol{\kappa}} = (a_{\boldsymbol{\kappa}} - 1)\hat{\sigma_{\boldsymbol{\kappa}}}^2$,

- $b_{\boldsymbol{\kappa}} = (a_\beta - 1)\hat{\sigma_\beta}^2$,

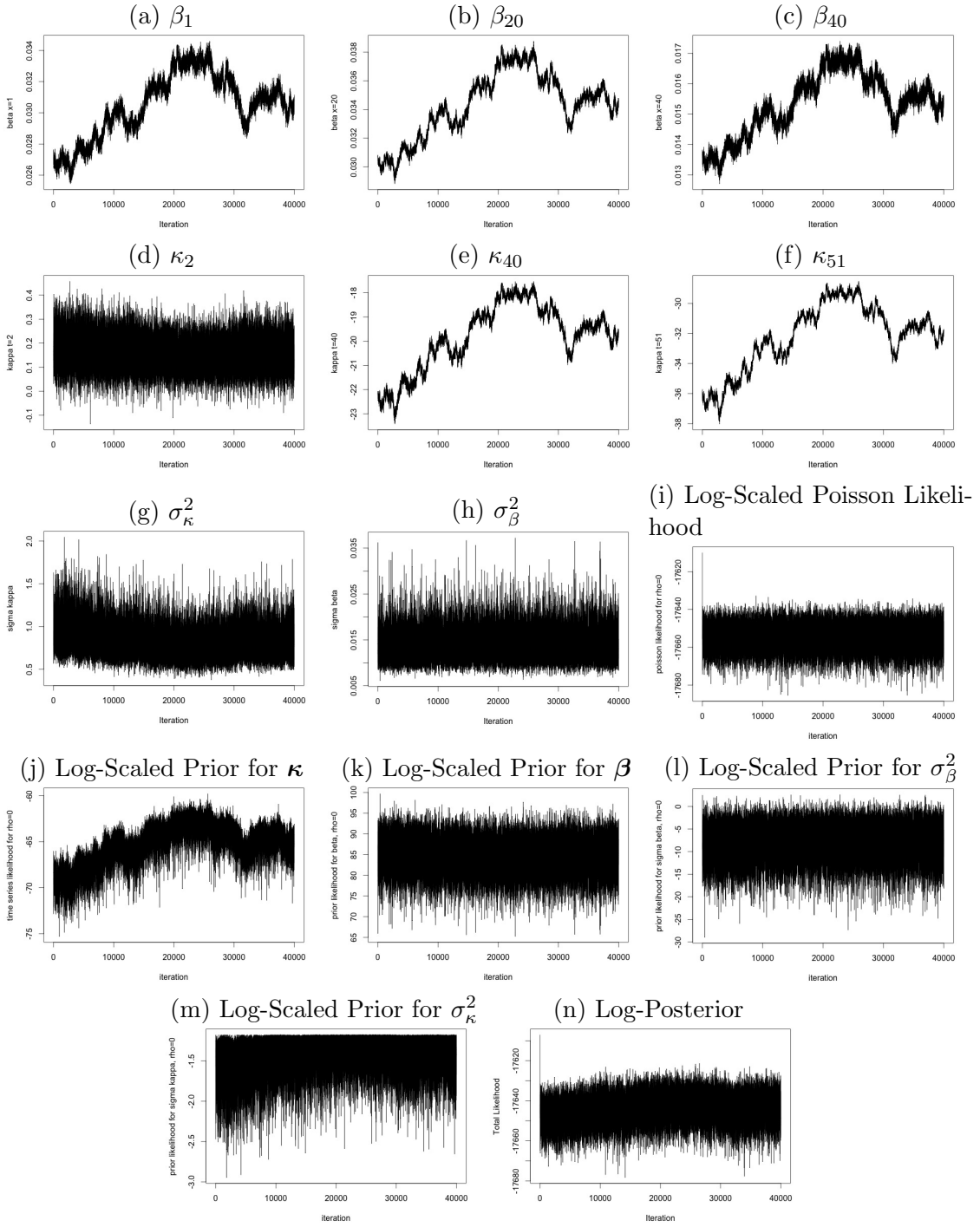


Figure C.18: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0.5$, $\rho = 0$ and $r = 0$ for non-zero co-variance, while $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, σ_β, σ_κ^2 are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not individually updated for $t = t_1, \ldots, t_{n_y}$ and the adjustment to κ_t and β_x is adopted.

- $b_x = 0.0001$,

- $a_x = b_x \times \exp \hat{\alpha}_x$.

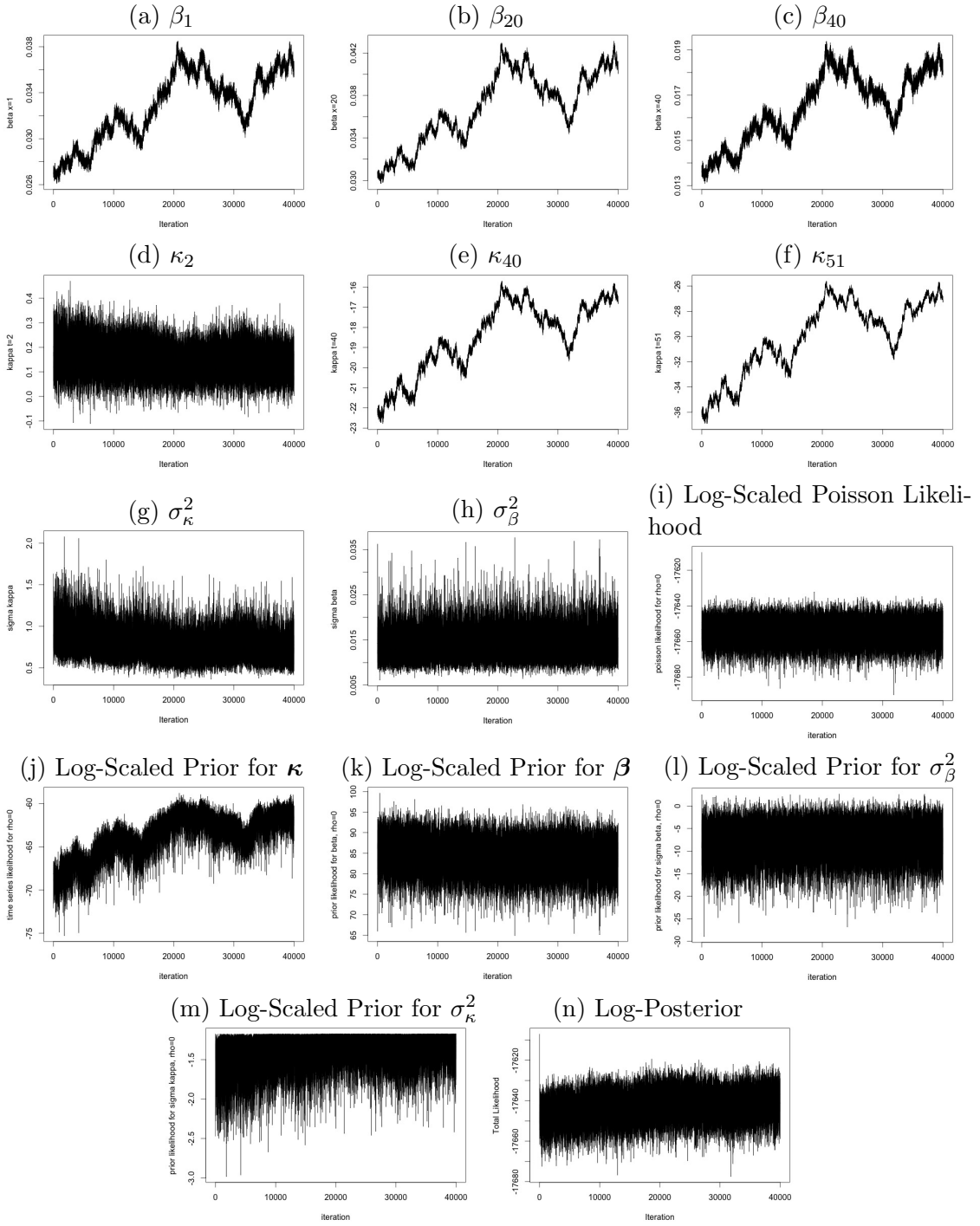


Figure C.19: The trace plots of MCMC for κ_t, β_x for selected t and x respectively and σ_κ^2, σ_β^2, given $g = 0.5$, $\rho = 0$ and $r = 0.9$ for non-zero co-variance, while $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, σ_β, σ_κ^2 are updated. The trajectories of log-scaled Poisson likelihood, the prior and the posterior of $\boldsymbol{\theta}$ are also presented. Note that the κ_t is not individually updated for $t = t_1, \ldots, t_{n_y}$ and the adjustment to κ_t and β_x is adopted.

The trace plots are shown in Figure

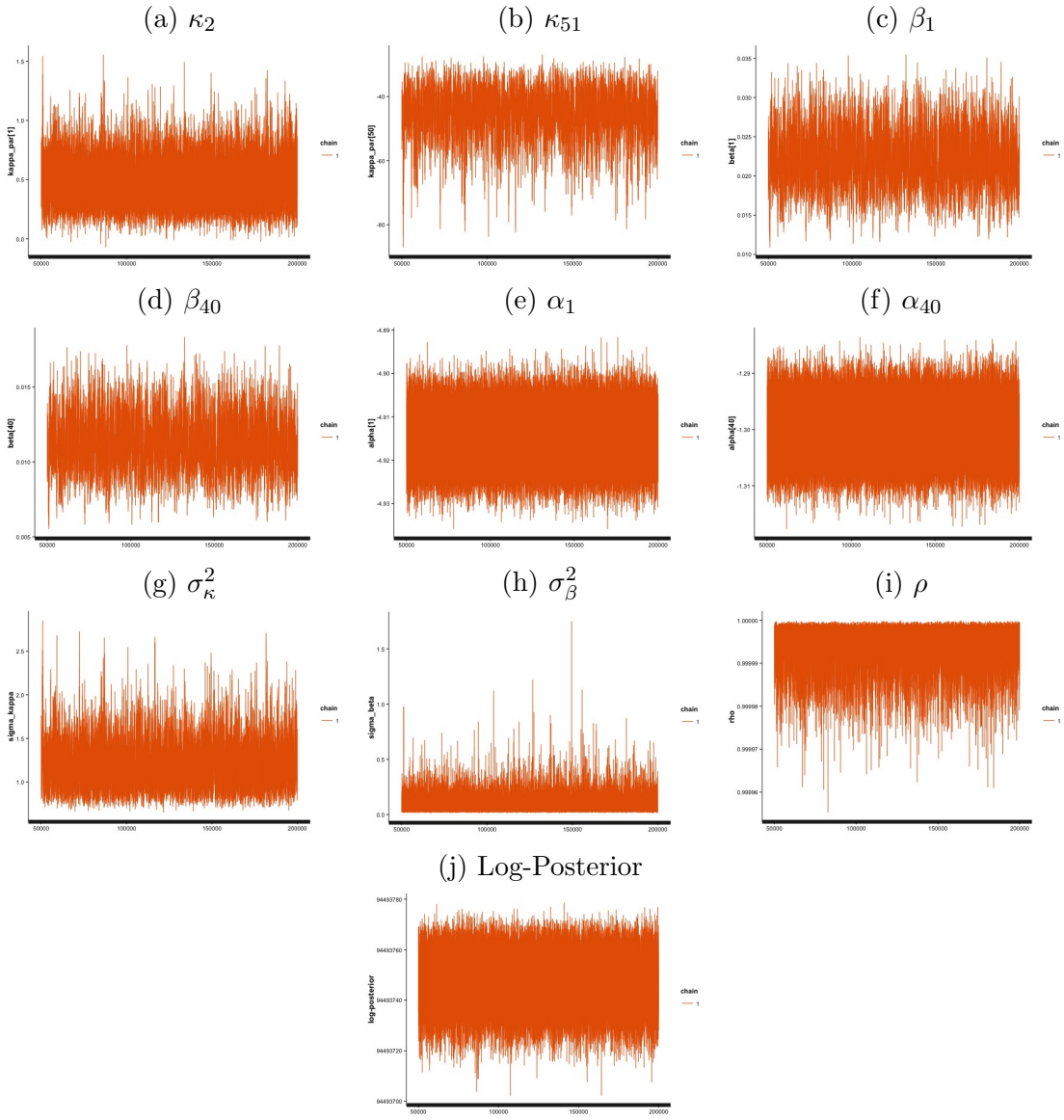


Figure C.20: The trace plots of the latent parameters with selected ages and years and the hyper-parameters.

C.2 Hamiltonian Monte Carlo

C.2.1 Energy Function

In physics, Energy function describes the total energy of a certain system with respect to the state of the system.

C.2.2 Potential Energy

In physics, potential energy is energy which results in position or configuration.

C.3 The Impact of Identifiability Constraints on Posterior Distribution

Appendix B.1 describes the procedure for applying the identifiability constraints for calculating the maximum likelihood estimates referring to the two-stage approach. This section shows how applying the constraints will affect the posterior distribution of the drift of $\kappa_t^{(1)}$, i.e. μ_1.

(5.29) gives the density of $p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V_\epsilon})$. Recall

$$p(\boldsymbol{\kappa}|\boldsymbol{\mu}, \boldsymbol{V_\epsilon}) \propto |\boldsymbol{V_\epsilon}|^{-\frac{n_y-1}{2}} \exp\Big(-\frac{1}{2}\sum_{t=t_2}^{t_{n_y}} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})^T \boldsymbol{V_\epsilon}^{-1} (\boldsymbol{\Delta}_t - \boldsymbol{\mu})\Big).$$

Same adjustment is made to $\boldsymbol{\kappa}$ as described in Appendix B.1 and denote as $\tilde{\boldsymbol{\kappa}}$ the adjusted period effects and $\tilde{\boldsymbol{\mu}}(t)$ the adjusted drift. Note that $\tilde{\boldsymbol{\mu}}(t)$ is not necessarily dependent on t.

We firstly prove that the posterior distribution of co-variance matrix $\boldsymbol{V_\epsilon}$ will not

be changed due to the adjustment.

$$\tilde{\kappa}_t^{(1)} - \tilde{\kappa}_{t-1}^{(1)} = \kappa_t^{(1)} - \kappa_{t-1}^{(1)} + \Phi(2) + \Phi(3)(2t-1) \tag{C.45}$$

$$\tilde{\kappa}_t^{(2)} - \tilde{\kappa}_{t-1}^{(2)} = \kappa_t^{(2)} - \kappa_{t-1}^{(2)} + 2\Phi(3) \tag{C.46}$$

$$\tilde{\kappa}_t^{(3)} - \tilde{\kappa}_{t-1}^{(3)} = \kappa_t^{(3)} - \kappa_{t-1}^{(3)} \tag{C.47}$$

Equation (C.45)-(C.47) show that the distribution of the co-variance matrix is the same after the adjustment.

Then the following condition needs to be satisfied in order to guarantee that the posterior distribution of period effects remain unchanged after the adjustment:

$$\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - \boldsymbol{\mu} = \tilde{\boldsymbol{\kappa}}_t - \tilde{\boldsymbol{\kappa}}_{t-1} - \tilde{\boldsymbol{\mu}}(t). \tag{C.48}$$

In particular, the adjusted drift $\tilde{\mu}_1(t)$ for $\tilde{\kappa}_t^{(1)}$ can be written as:

$$\tilde{\mu}_1(t) = \mu_1 + \Phi_2 + \Phi_3(2t-1),$$

which is now a function of t. It implies that by applying the constraints to the cohort effects and making appropriate adjustments to the period effects, the posterior distribution of μ_1 depends on the choice of the value of t.

C.4 England and Wales MCMC

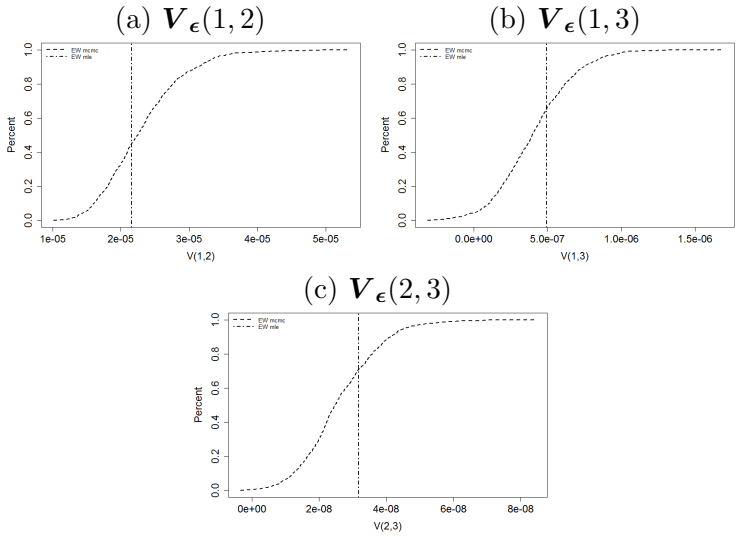
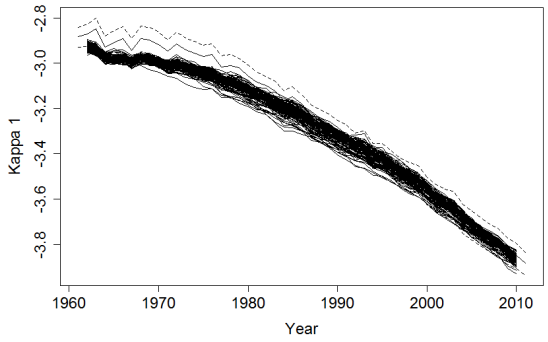


Figure C.21: CDF: The posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,2)$, $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,3)$ and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,3)$ (dashed CDFs). The vertical line is the corresponding empirical ML estimates of the benchmark exposure, the males in England and Wales.

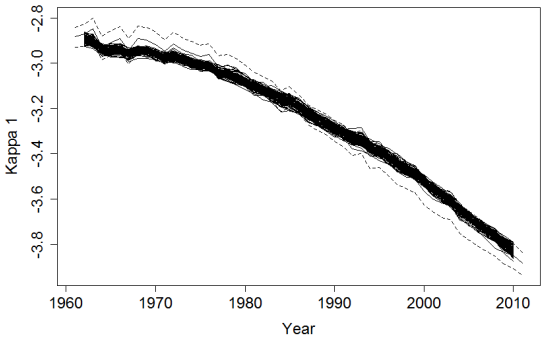
Appendix D

For Chapter 6

(a) Historical $\boldsymbol{\kappa}^{(1)}$, Method 5 Lower Bound



(b) Historical $\boldsymbol{\kappa}^{(1)}$, Method 5 Mean



(c) Historical $\boldsymbol{\kappa}^{(1)}$, Method 5 Upper Bound

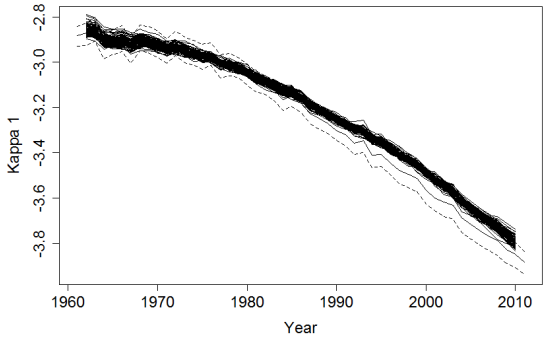
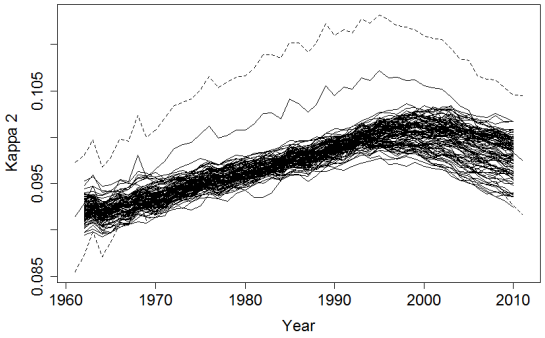
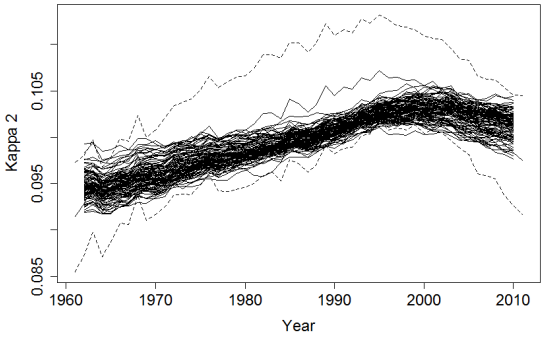


Figure D.1: The upper bound, mean and the lower bound of the credibility intervals of the posterior distribution for the 100 random death scenarios respectively according to the Method 5. For instance, we calculated the 5% quantile of the MCMC samples for each of the 100 death scenarios and plotted all these one hundred values in Figure (a).

(a) Historical $\boldsymbol{\kappa}^{(2)}$, Method 5 Lower Bound



(b) Historical $\boldsymbol{\kappa}^{(2)}$, Method 5 Mean



(c) Historical $\boldsymbol{\kappa}^{(2)}$, Method 5 Upper Bound

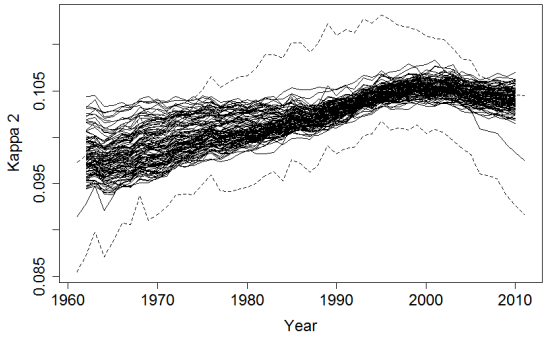
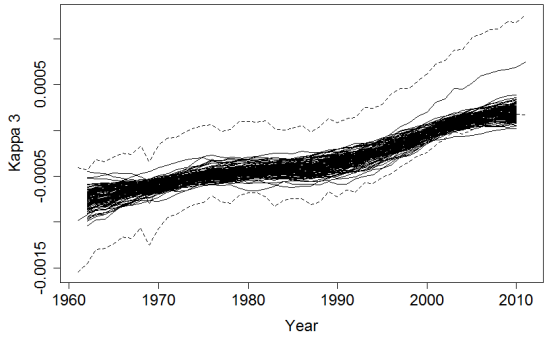
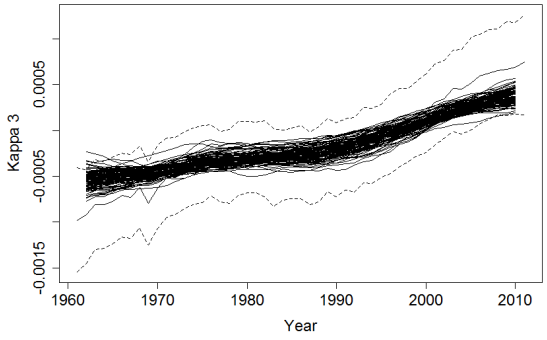


Figure D.2: The upper bound, mean and the lower bound of the credibility intervals of the posterior distribution for the 100 random death scenarios respectively according to the Method 5. For instance, we calculated the 5% quantile of the MCMC samples for each of the 100 death scenarios and plotted all these one hundred values in Figure (a).

(a) Historical $\boldsymbol{\kappa}^{(3)}$, Method 5 Lower Bound



(b) Historical $\boldsymbol{\kappa}^{(3)}$, Method 5 Mean



(c) Historical $\boldsymbol{\kappa}^{(3)}$, Method 5 Upper Bound

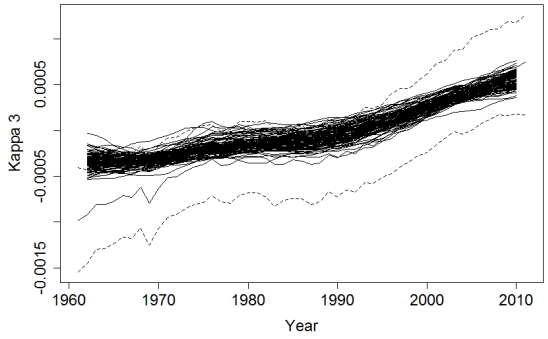
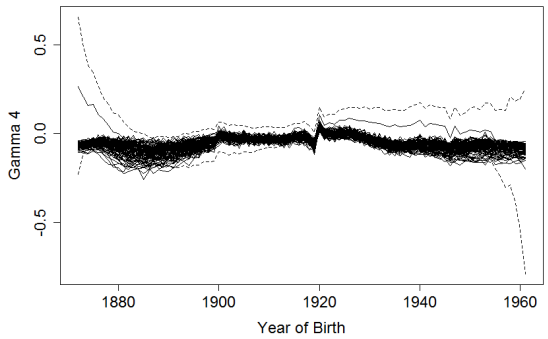
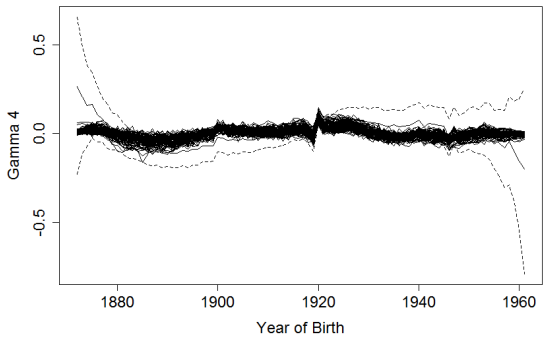


Figure D.3: The upper bound, mean and the lower bound of the credibility intervals of the posterior distribution for the 100 random death scenarios respectively according to the Method 5. For instance, we calculated the 5% quantile of the MCMC samples for each of the 100 death scenarios and plotted all these one hundred values in Figure (a).

(a) Historical $\boldsymbol{\gamma}^{(4)}$, Method 5 Lower Bound



(b) Historical $\boldsymbol{\gamma}^{(4)}$, Method 5 Mean



(c) Historical $\boldsymbol{\gamma}^{(4)}$, Method 5 Upper Bound

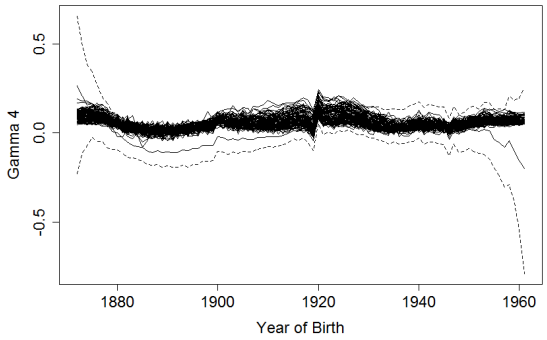
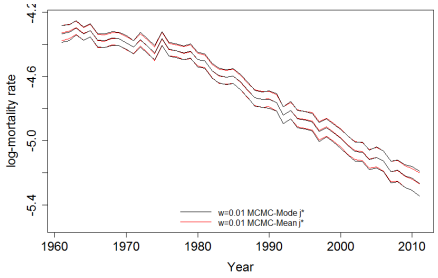
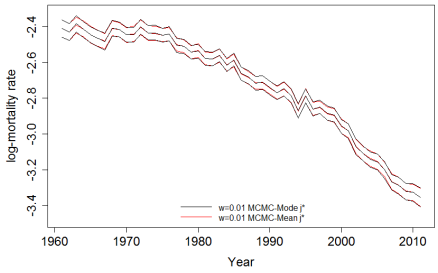


Figure D.4: Upper bound, mean and the lower bound of the credibility intervals of the posterior distribution for the 100 random death scenarios respectively, according to Method 5. For instance, we calculated the 5% quantile of the MCMC samples for each of the 100 death scenarios and plotted all these one hundred values in Figure (a). The dashed lines are the confidence intervals of the point estimates of $w = 0.01$ and the solid line is the estimate of England and Wales.

(a) $\log m(t,x)$ for age 55, Method 4, Sensitivity Test



(b) $\log m(t,x)$ for age 75, Method 4, Sensitivity Test



(c) $\log m(t,x)$ for age 85, Method 4, Sensitivity Test

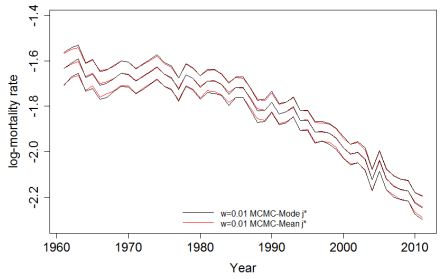
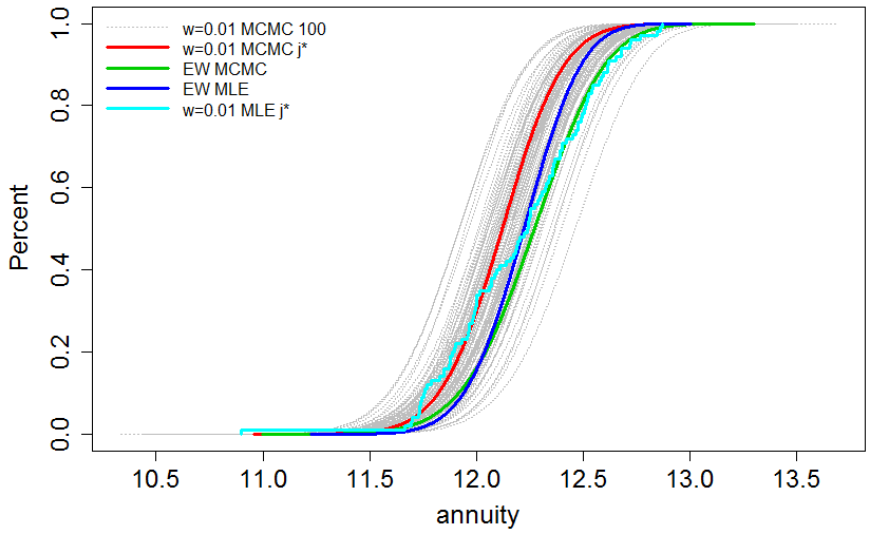


Figure D.5: Comparison: the distribution of $m(\tilde{\theta}_1^{w,j^*}, t, x)$ given the MCMC-Mean (red lines) and the MCMC-Mode (black lines) at age 55, 75 and 85 for the death scenario j^*. The upper and lower bound for the intervals are 95% and 5% quantiles of the MCMC samples.

(a) CDF: Temporary Annuity price



(b) CDF: Deferred Annuity price

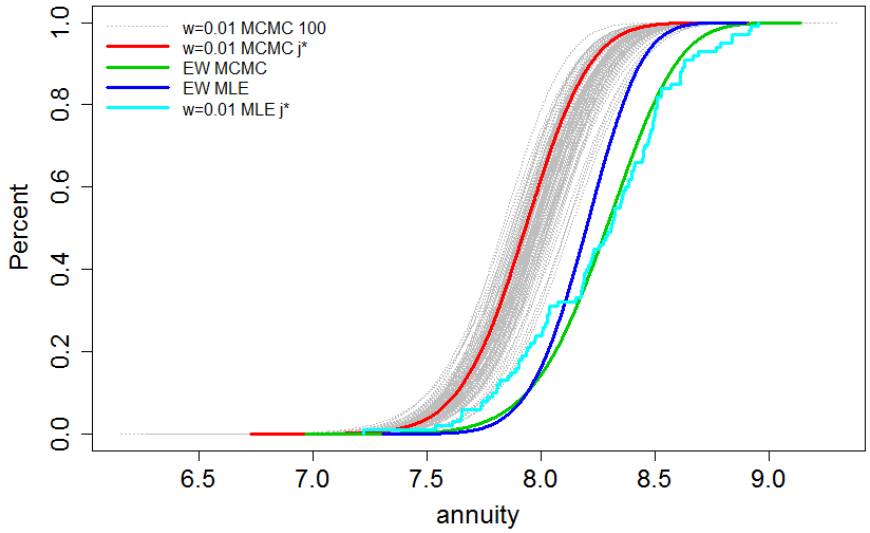
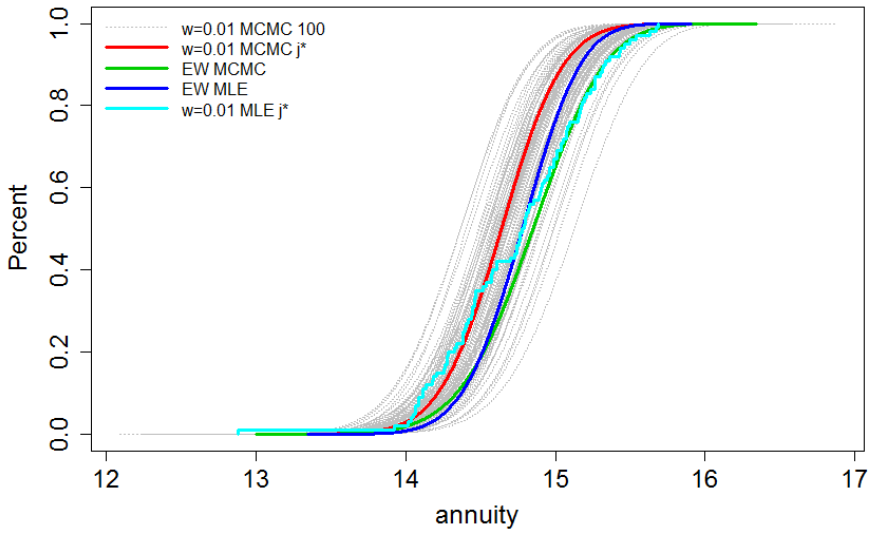


Figure D.6: The CDF of the temporary (upper) and deferred (lower) annuity of interest rate 4%. Each of the black curve is the distributions of one of the 100 random death scenarios of $w = 0.01$ and the red curve is for death scenario j^*. The blue and green curves are for the EW-MLE and EW-MCMC respectively. The light blue curve is for the j^*-MLE, which is not as smooth as the others since only 100 sample paths are simulated for j^*-MLE.

(a) CDF: Temporary Annuity price



(b) CDF: Deferred Annuity price

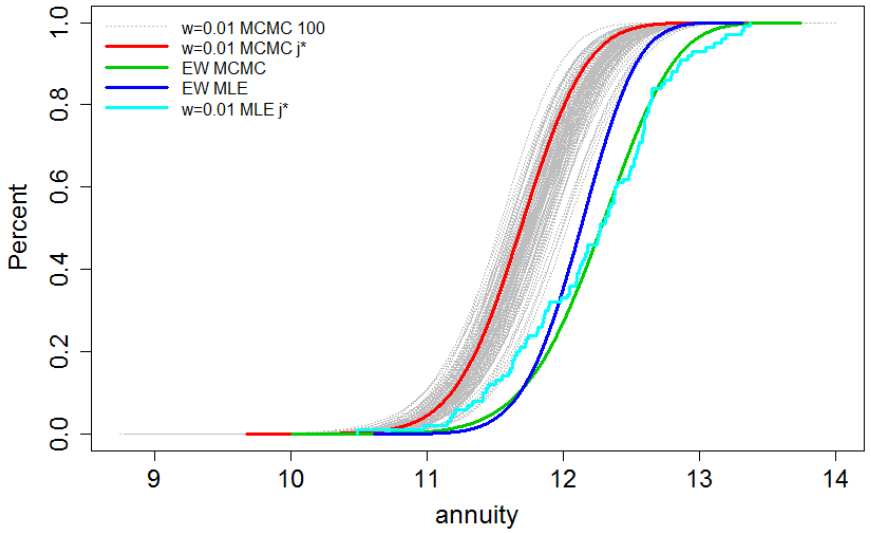
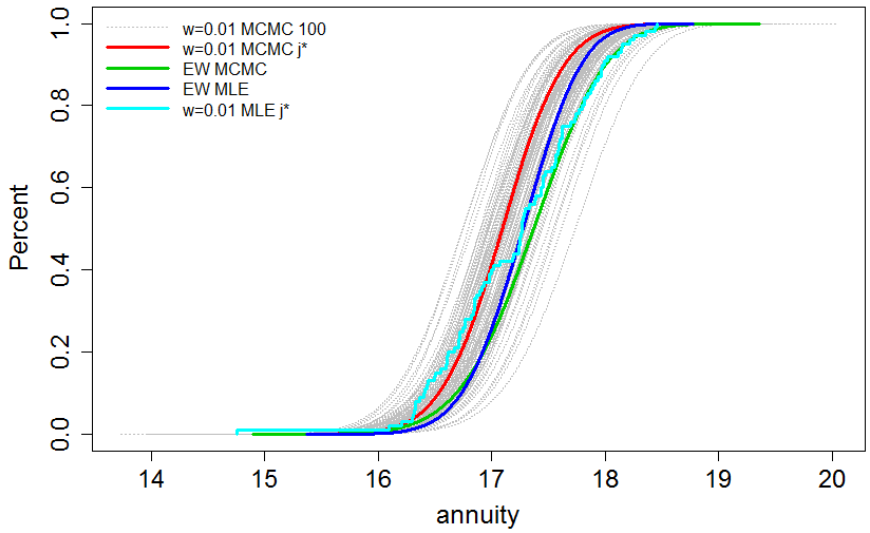


Figure D.7: The CDF of the temporary (upper) and deferred (lower) annuity of interest rate 2%. Each of the black curve is the distributions of one of the 100 random death scenarios of $w = 0.01$ and the red curve is for death scenario j^*. The blue and green curves are for the EW-MLE and EW-MCMC respectively. The light blue curve is for the j^*-MLE, which is not as smooth as the others since only 100 sample paths are simulated for j^*-MLE.

(a) CDF: Temporary Annuity price



(b) CDF: Deferred Annuity price

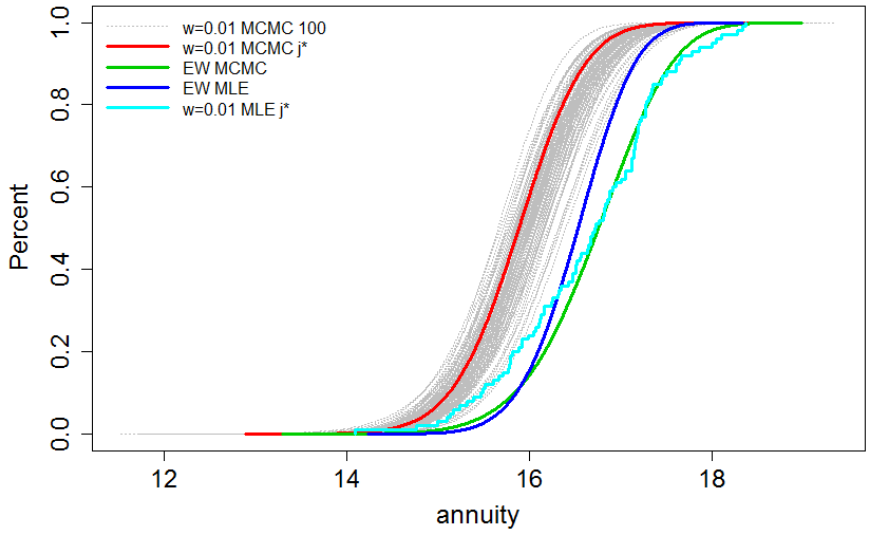
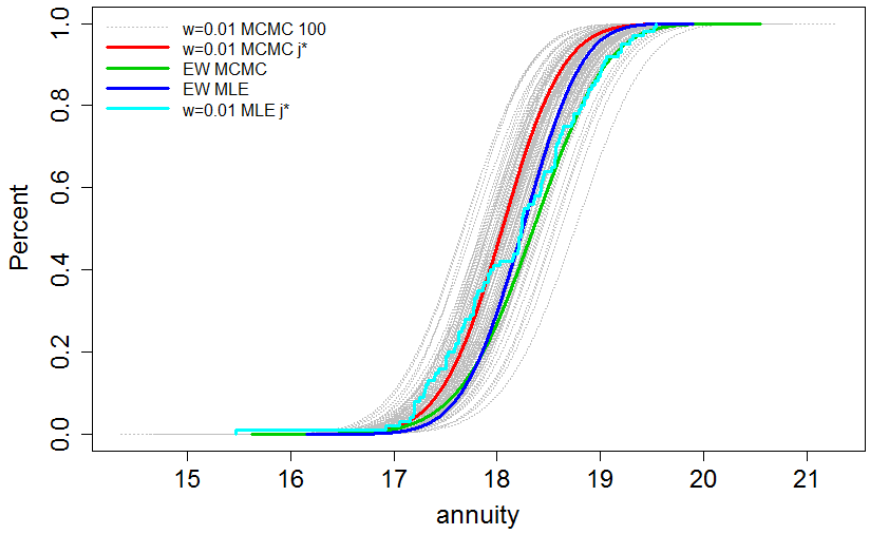


Figure D.8: The CDF of the temporary (upper) and deferred (lower) annuity of interest rate 0.5%. Each of the black curve is the distributions of one of the 100 random death scenarios of $w = 0.01$ and the red curve is for death scenario j^*. The blue and green curves are for the EW-MLE and EW-MCMC respectively. The light blue curve is for the j^*-MLE, which is not as smooth as the others since only 100 sample paths are simulated for j^*-MLE.

(a) CDF: Temporary Annuity price



(b) CDF: Deferred Annuity price

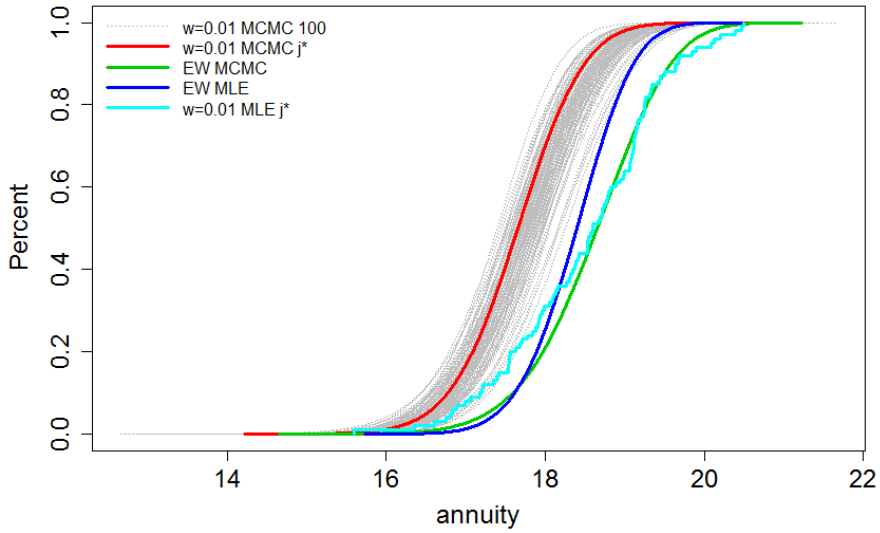
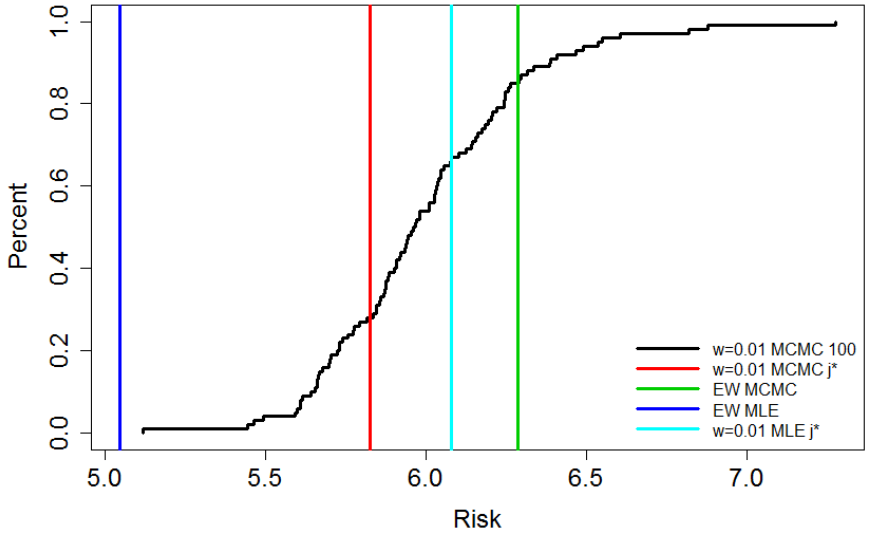


Figure D.9: The CDF of the temporary (upper) and deferred (lower) annuity of interest rate 0%. Each of the black curve is the distributions of one of the 100 random death scenarios of $w = 0.01$ and the red curve is for death scenario j^*. The blue and green curves are for the EW-MLE and EW-MCMC respectively. The light blue curve is for the j^*-MLE, which is not as smooth as the others since only 100 sample paths are simulated for j^*-MLE.

(a) CDF: Longevity Risk given a Temporary Annuity



(b) CDF: Longevity Risk given a Deferred Annuity

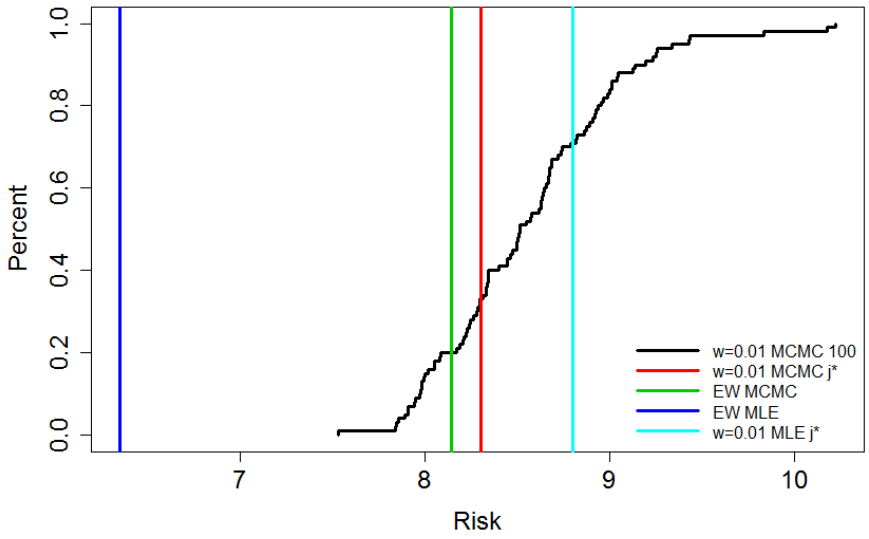
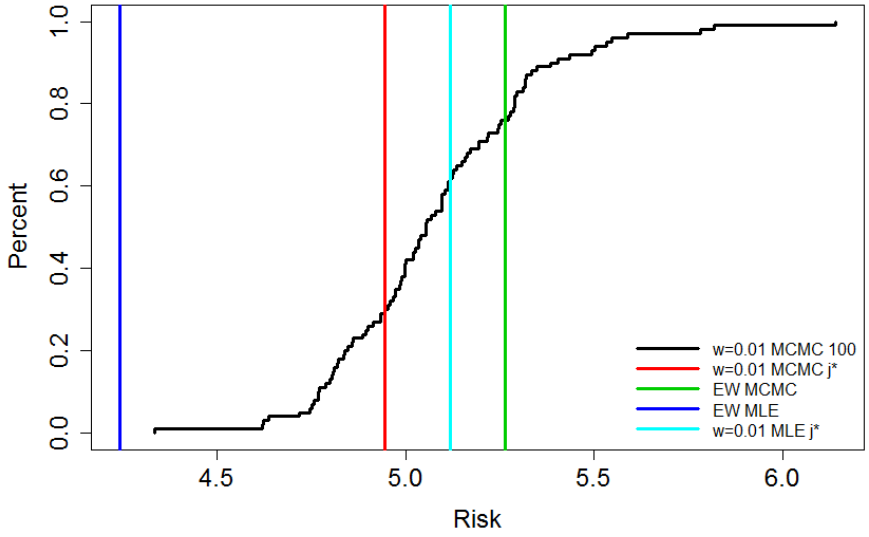


Figure D.10: The CDF of the longevity risk given temporary (upper) and deferred (lower) annuity of interest rate 2%. The black curve is the distributions of 100 random death scenarios of $w = 0.01$ and the red line is for death scenario j^*. The blue and green line are for the EW-MLE and EW-MCMC respectively. The light blue line is for the j^*-MLE.

(a) CDF: Longevity Risk given a Temporary Annuity



(b) CDF: Longevity Risk given a Deferred Annuity

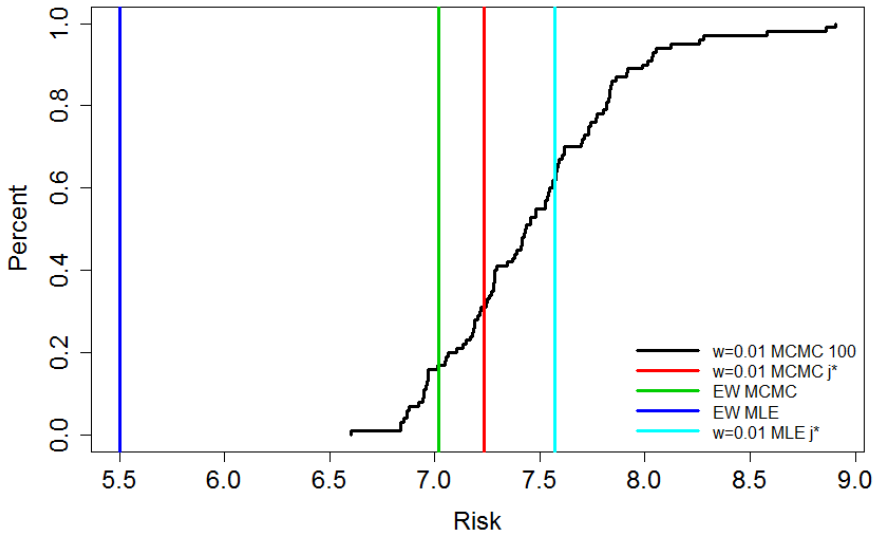
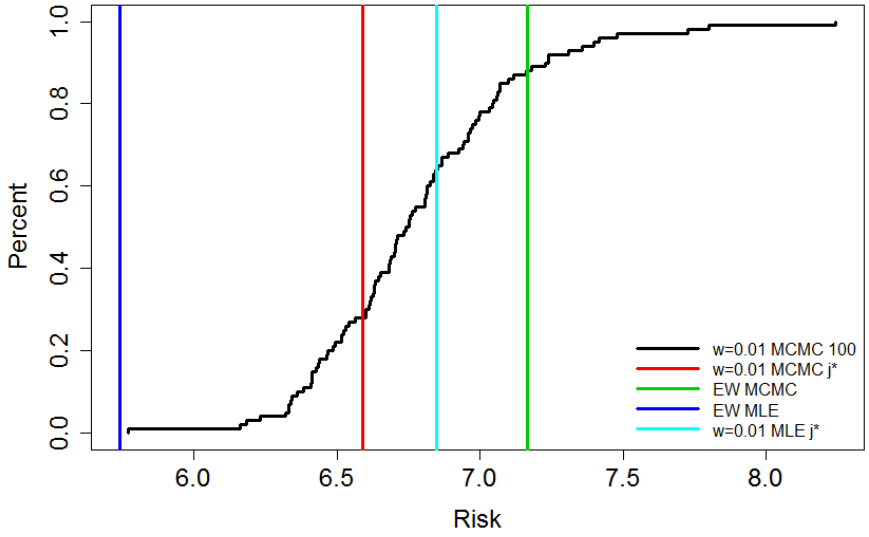


Figure D.11: The CDF of the longevity risk given temporary (upper) and deferred (lower) annuity of interest rate 4%. The black curve is the distributions of 100 random death scenarios of $w = 0.01$ and the red line is for death scenario j^*. The blue and green line are for the EW-MLE and EW-MCMC respectively. The light blue line is for the j^*-MLE.

(a) CDF: Longevity Risk given a Temporary Annuity



(b) CDF: Longevity Risk given a Deferred Annuity

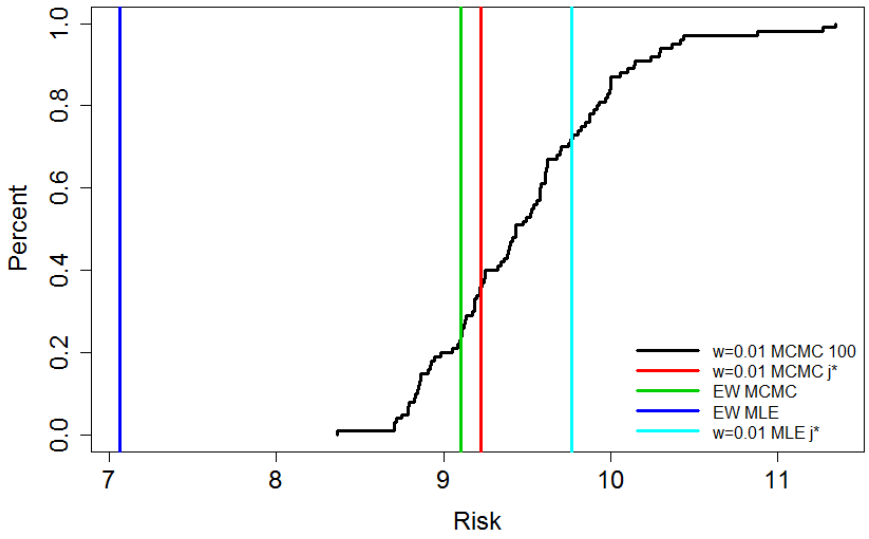
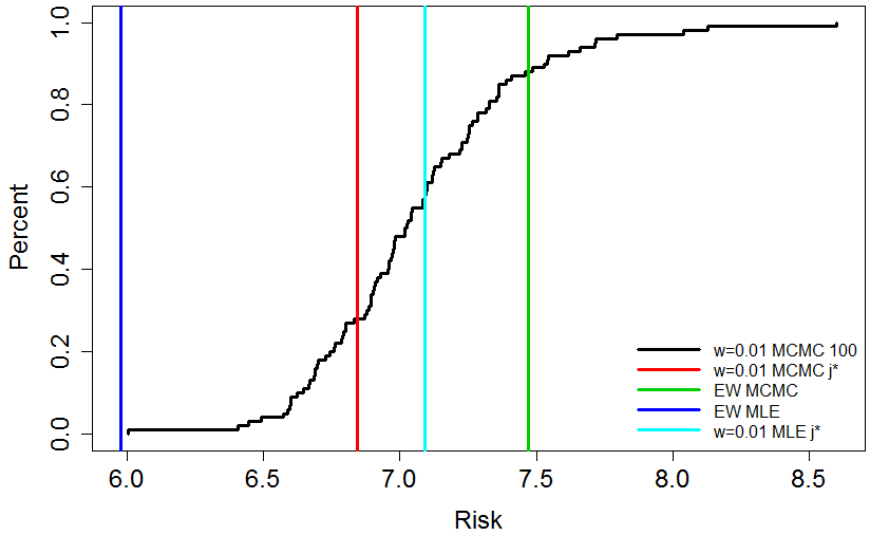


Figure D.12: The CDF of the longevity risk given temporary (upper) and deferred (lower) annuity of interest rate 0.5%. The black curve is the distributions of 100 random death scenarios of $w = 0.01$ and the red line is for death scenario j^*. The blue and green line are for the EW-MLE and EW-MCMC respectively. The light blue line is for the j^*-MLE.

(a) CDF: Longevity Risk given a Temporary Annuity



(b) CDF: Longevity Risk given a Deferred Annuity

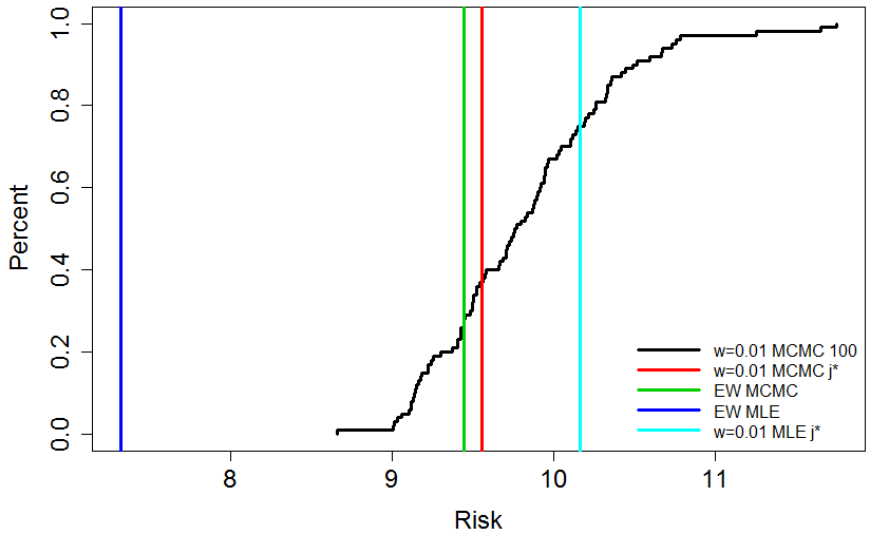


Figure D.13:]
The CDF of the longevity risk given temporary (upper) and deferred (lower) annuity of interest rate 0%. The black curve is the distributions of 100 random death scenarios of $w = 0.01$ and the red line is for death scenario j^*. The blue and green line are for the EW-MLE and EW-MCMC respectively. The light blue line is for the j^*-MLE.

Appendix E

For Chapter 7

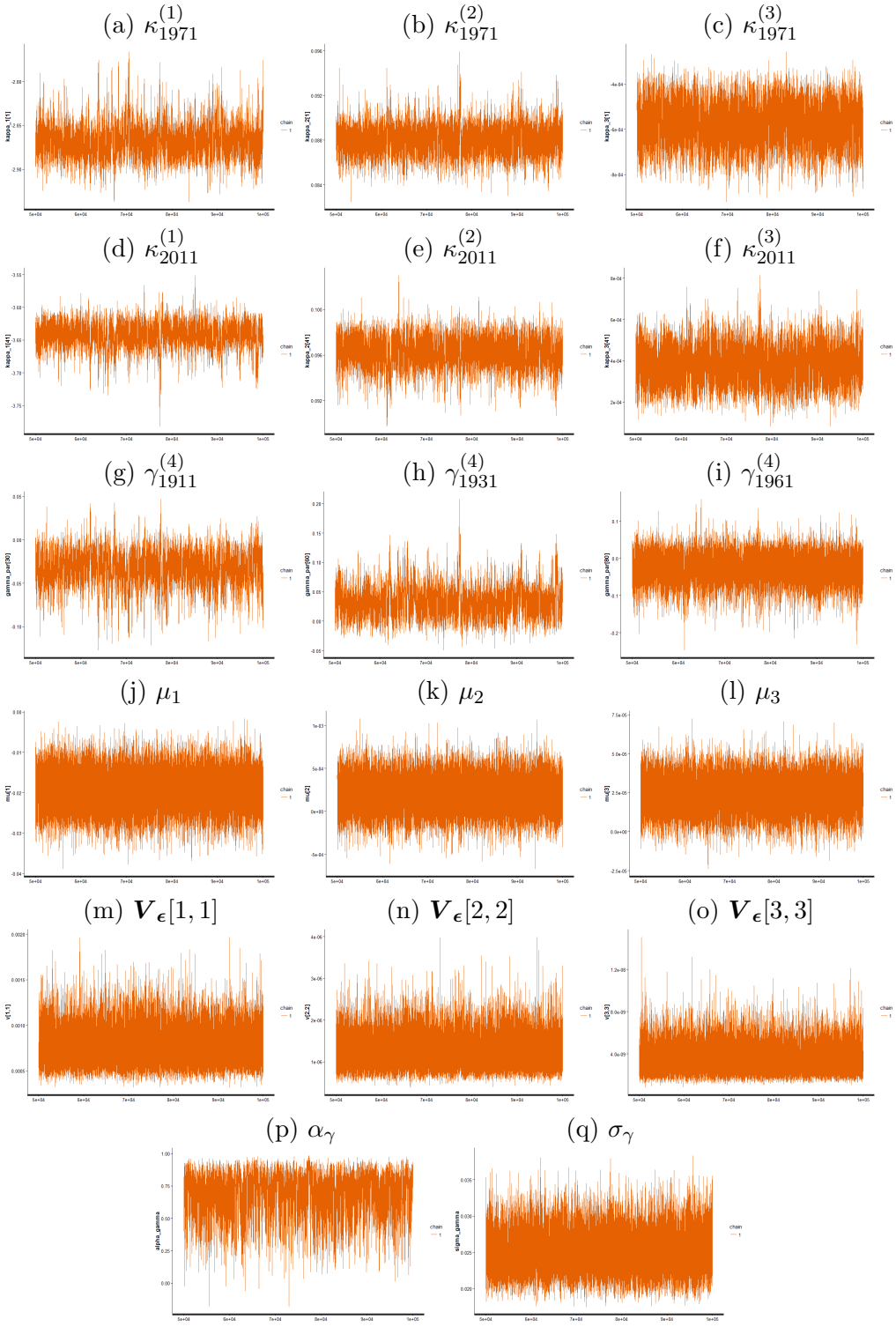


Figure E.1: Scotland B-EW: Trace plot for selected parameters, including the period and cohort effects with selected year (1971, 2011) and year of birth (1911,1931,1961) respectively, the drift of multi-variate random walk $\boldsymbol{\mu}$, the diagonal of co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, coefficient and standard deviation of AR(1) model α_γ, σ_γ.

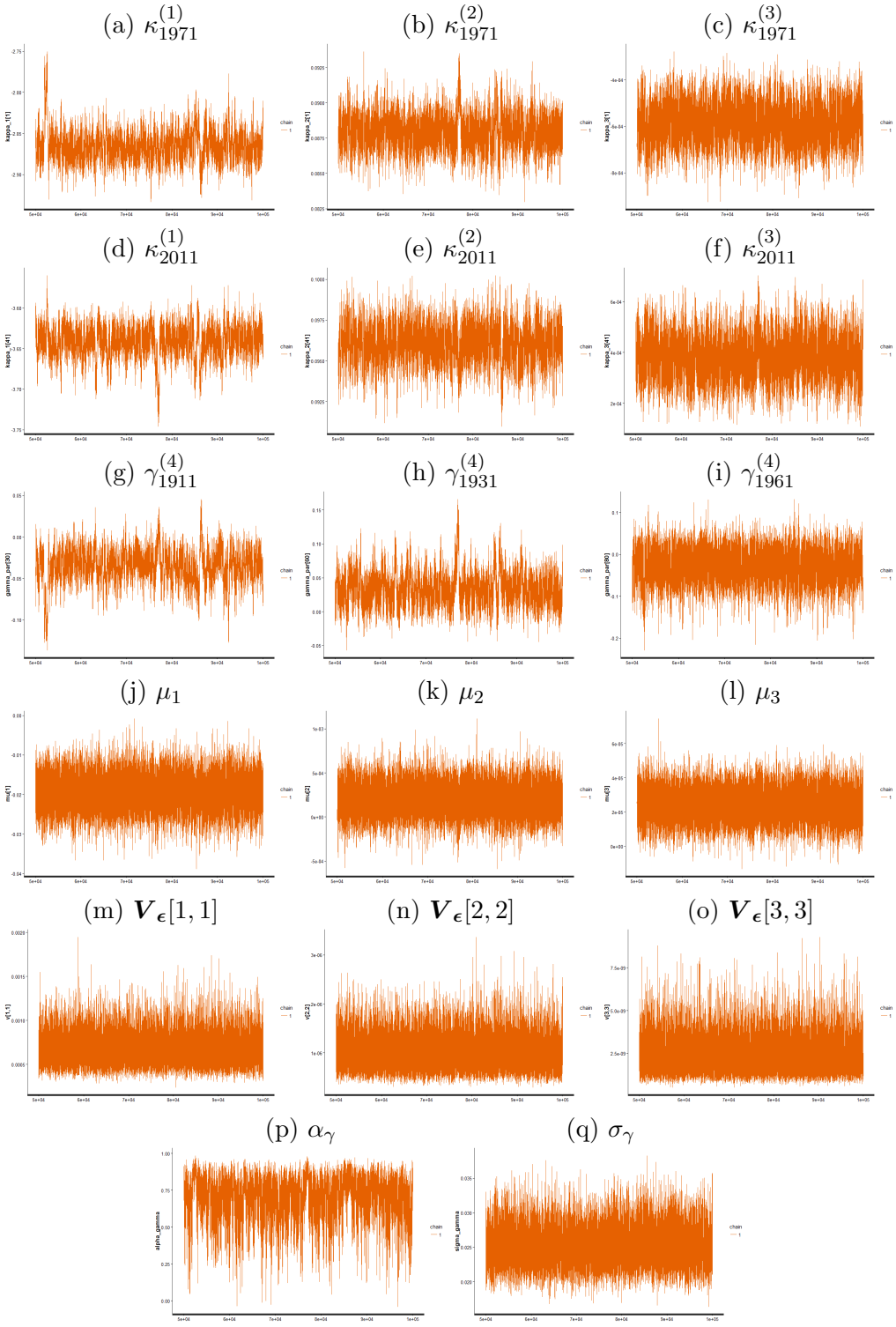


Figure E.2: Scotland B-UK: Trace plot for selected parameters, including the period and cohort effects with selected year (1971, 2011) and year of birth (1911,1931,1961) respectively, the drift of multi-variate random walk $\boldsymbol{\mu}$, the diagonal of co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, coefficient and standard deviation of AR(1) model α_γ, σ_γ.

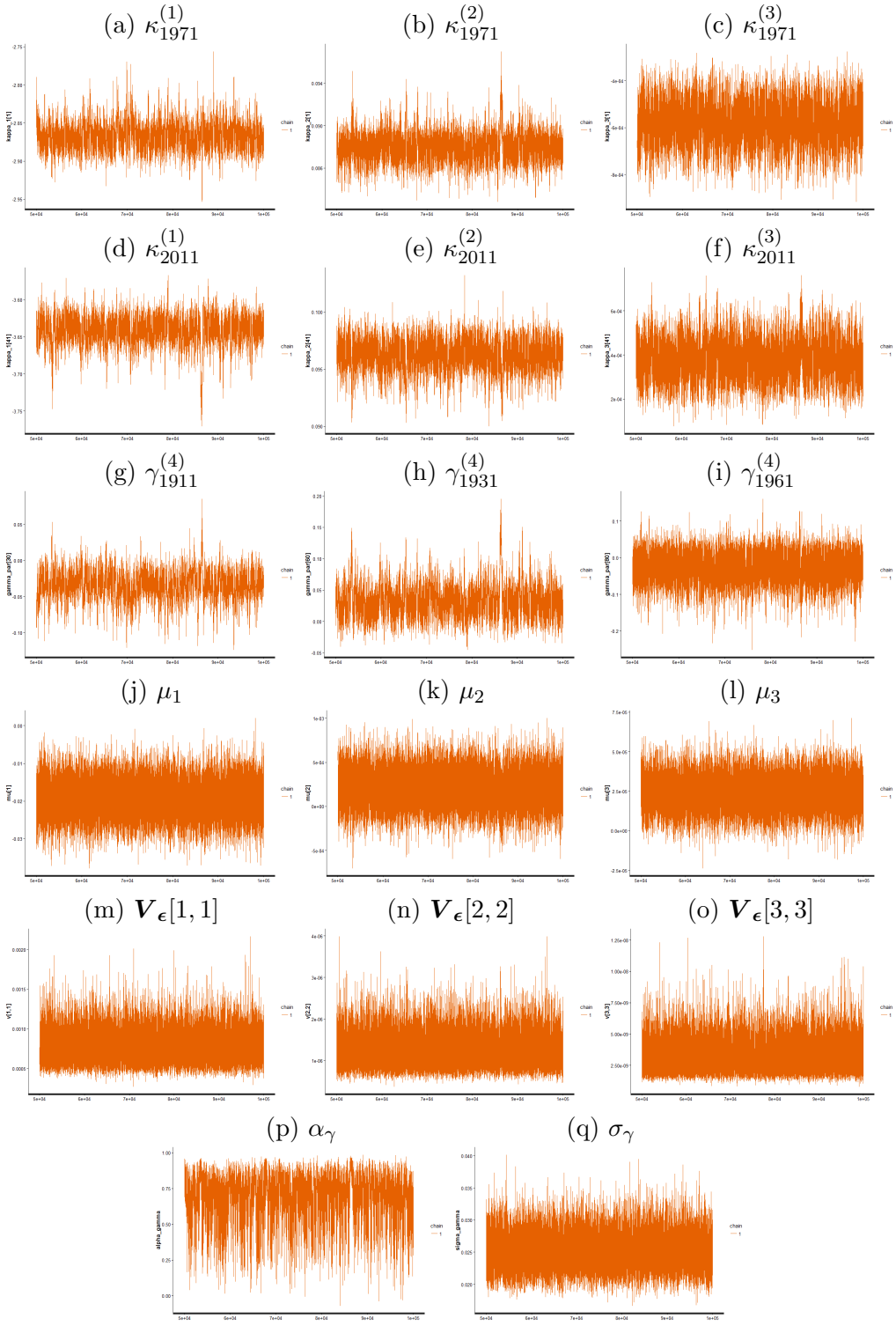
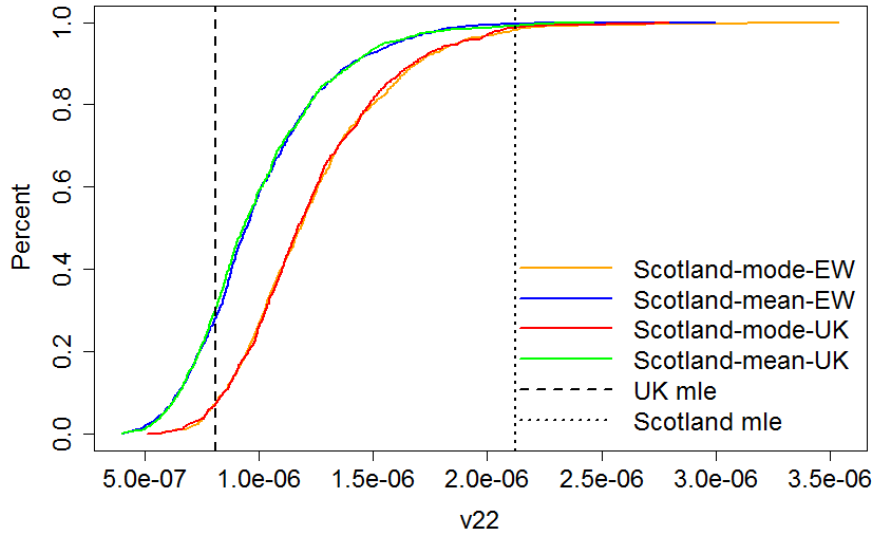


Figure E.3: Scotland A-UK: Trace plot for selected parameters, including the period and cohort effects with selected year (1971, 2011) and year of birth (1911,1931,1961) respectively, the drift of multi-variate random walk $\boldsymbol{\mu}$, the diagonal of co-variance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, coefficient and standard deviation of AR(1) model α_γ, σ_γ.

(a) CDF: $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{SL}}(2,2)$



(b) CDF: $\tilde{\boldsymbol{\mu}}_{2}^{\mathrm{SL}}$

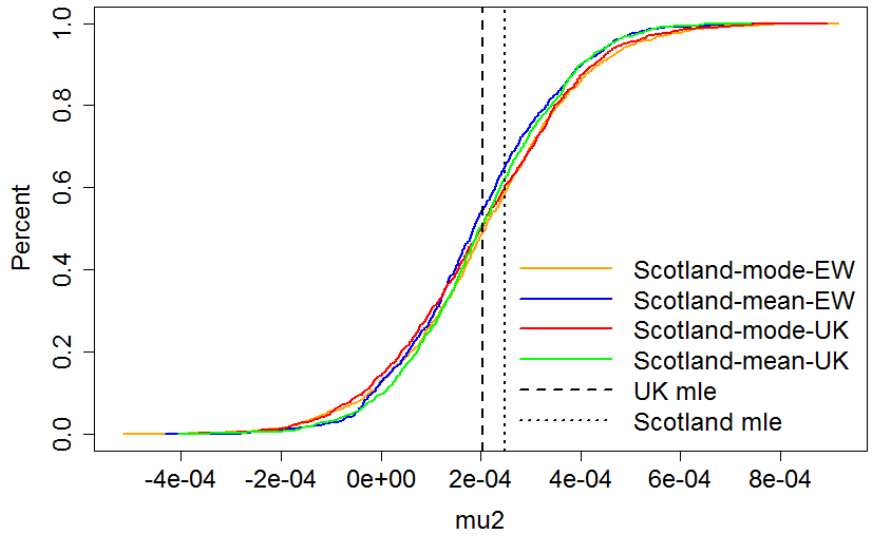
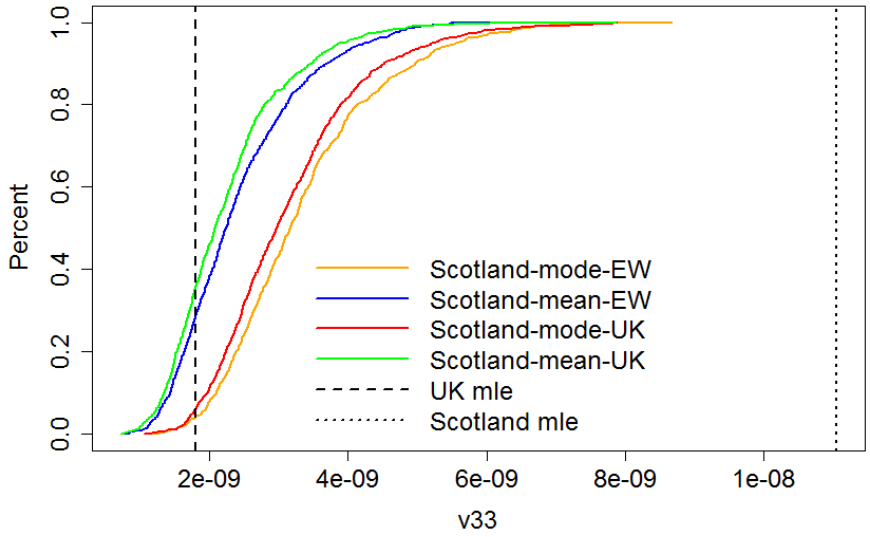


Figure E.4: CDF: the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$ and $\boldsymbol{\mu}_2$ for the Scotland, given varied prior settings for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,2)$, e.g. A-EW (orange), A-UK (red), B-EW (blue) and B-UK (green). The dashed vertical line is the point estimate for the UK and the dotted line is for the Scotland.

(a) CDF: $\tilde{\boldsymbol{V}}_{\epsilon}^{\mathrm{SL}}(3,3)$



(b) CDF: $\tilde{\boldsymbol{\mu}}_3^{\mathrm{SL}}$

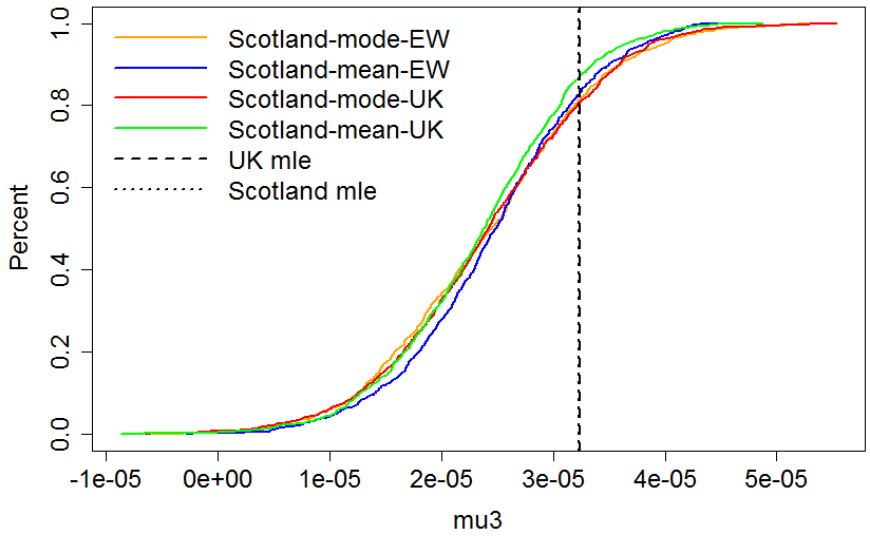


Figure E.5: CDF: the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$ and $\boldsymbol{\mu}_3$ for the Scotland, given varied prior settings for $\boldsymbol{V}_{\boldsymbol{\epsilon}}(3,3)$, e.g. A-EW (orange), A-UK (red), B-EW (blue) and B-UK (green). The dashed vertical line is the point estimate for the UK and the dotted line is for the Scotland.

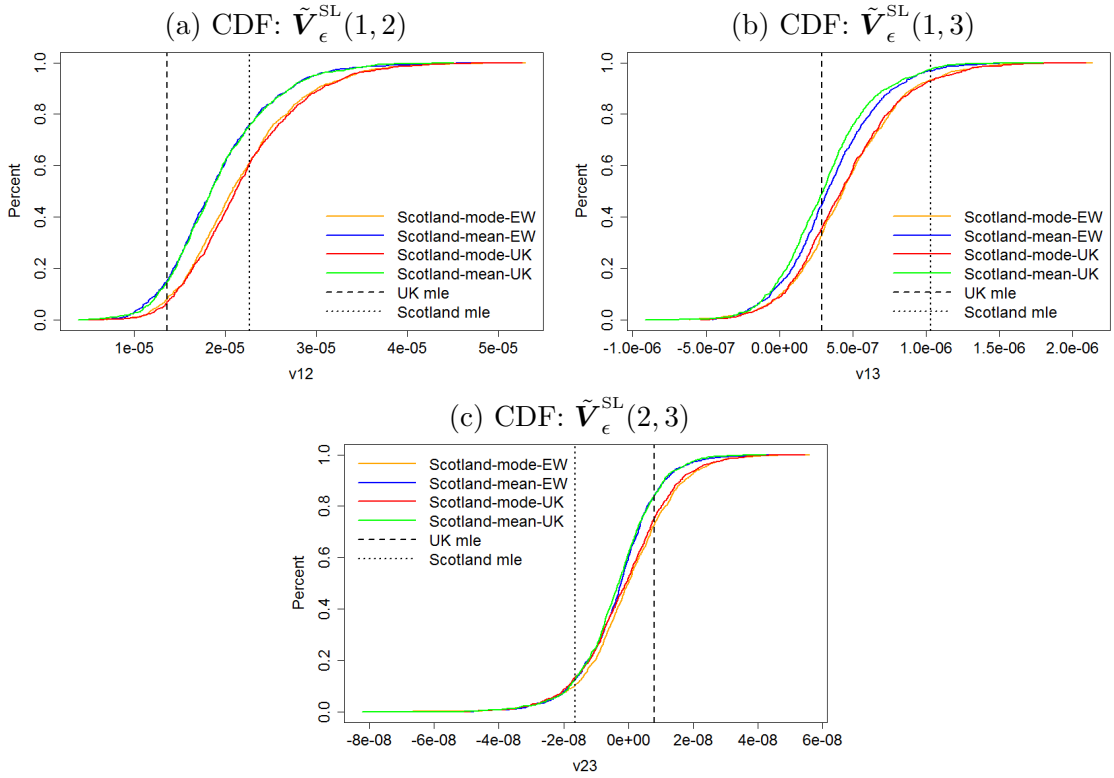


Figure E.6: CDF: the posterior distribution of $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,2)$, $\boldsymbol{V}_{\boldsymbol{\epsilon}}(1,3)$, and $\boldsymbol{V}_{\boldsymbol{\epsilon}}(2,3)$ for the Scotland, given varied prior settings for $\boldsymbol{V}_{\boldsymbol{\epsilon}}$, e.g. A-EW (orange), A-UK (red), B-EW (blue) and B-UK (green). The dashed vertical line is the point estimate for the UK and the dotted line is for the Scotland.

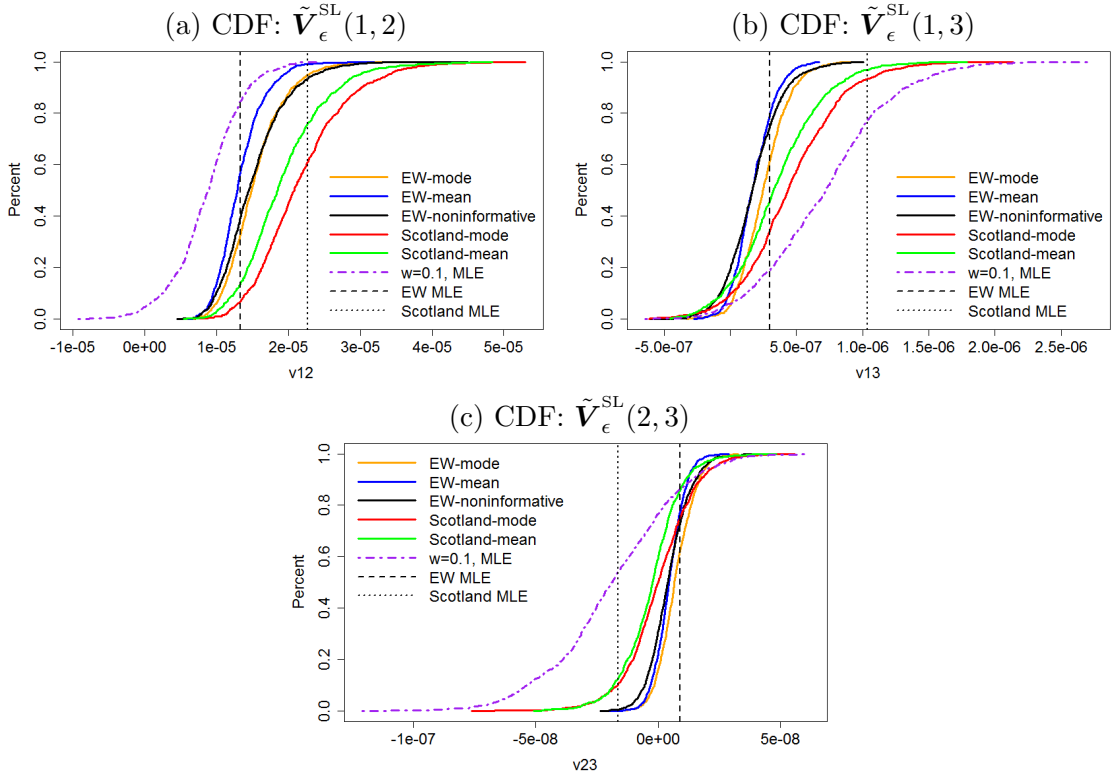


Figure E.7: CDF: the posterior distribution of $\tilde{\boldsymbol{V}}_\epsilon^{\mathrm{SL}}(1,2)$, $\tilde{\boldsymbol{V}}_\epsilon^{\mathrm{SL}}(1,3)$, and $\tilde{\boldsymbol{V}}_\epsilon^{\mathrm{SL}}(2,3)$ for the Scotland, given A-EW (red) and B-EW (green); for England and Wales given A-EW (orange), B-EW (blue) and C-JR (black). The purple dotted curves are the CDFs of the finite sample MLE of the corresponding parameters for $w = 0.1$. The vertical dashed and dotted lines are the corresponding MLEs of the England and Wales and Scotland respectively.

Bibliography

Alder, B. J. & Wainwright, T. (1959), 'Studies in molecular dynamics. i. general method', *The Journal of Chemical Physics* **31**(2), 459–466.

Andersen, H. C. (1980), 'Molecular dynamics simulations at constant pressure and/or temperature', *The Journal of chemical physics* **72**(4), 2384–2393.

Anderson, B. D. & Moore, J. B. (1979), 'Optimal filtering. 1979'.

Andreev, K. (2002), *Evolution of the Danish Population from 1835 to 2000*, Vol. 9, University Press of Southern Denmark.

Anisimova, M., Bielawski, J. P. & Yang, Z. (2001), 'Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution', *Molecular biology and evolution* **18**, 1585–1592.

Booth, H., Hyndman, R. J., Tickle, L., De Jong, P. et al. (2006), Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions, Technical report, Monash University, Department of Econometrics and Business Statistics.

Booth, H., Maindonald, J. & Smith, L. (2002), 'Applying lee-carter under conditions of variable mortality decline', *Population studies* **56**(3), 325–336.

Booth, H., Maindonald, J., Smith, L. et al. (2001), 'Age-time interactions in mortality projection: Applying lee-carter to australia'.

Booth, H. & Tickle, L. (2008), 'Mortality modelling and forecasting: A review of methods', *Annals of actuarial science* **3**(1-2), 3–43.

Box, G. E. & Tiao, G. C. (2011), *Bayesian inference in statistical analysis*, Vol. 40, John Wiley & Sons.

Brouhns, N., Denuit, M. & Van Keilegom, I. (2005), 'Bootstrapping the Poisson log-bilinear model for mortality forecasting', *Scandinavian Actuarial Journal* **2005**(3), 212–224.

Cairns, A. J. G., Blake, D. & Dowd, K. (2006a), 'A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration', *Journal of Risk and Insurance* **73**(4), 687–718.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D. & Khalaf-Allah, M. (2011a), 'Mortality density forecasts: An analysis of six stochastic mortality models', *Insurance: Mathematics and Economics* **48**(3), 355–367.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. & Balevich, I. (2009), 'A quantitative comparison of stochastic mortality models using data from england and wales and the united states', *North American Actuarial Journal* **13**(1), 1–35.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D. & Khalaf-Allah, M. (2011b), 'Bayesian stochastic mortality modelling for two populations', *Astin Bulletin* **41**(01), 29–59.

Carriere, J. F. (1992), 'Parametric models for life tables', *Transactions of the Society of Actuaries* **44**, 77–99.

Carter, C. K. & Kohn, R. (1994), 'On gibbs sampling for state space models', *Biometrika* **81**(3), 541–553.

Coale, A. J. & McNeil, D. R. (1972), 'The distribution by age of the frequency of first marriage in a female cohort', *Journal of the American Statistical Association* **67**(340), 743–749.

Cowles, M. K. & Carlin, B. P. (1996), 'Markov chain monte carlo convergence diagnostics: a comparative review', *Journal of the American Statistical Association* **91**(434), 883–904.

Currie, I. (2006), 'Smoothing and forecasting mortality rates with p-splines', *URL http://www. macs. hw. ac. uk/˜ iain/research/talks/Mortality. pdf* .

Currie, I. D. (2016), 'On fitting generalized linear and non-linear models of mortality', *Scandinavian Actuarial Journal* **2016**(4), 356–383.

Currie, I. D., Durban, M. & Eilers, P. H. (2004), 'Smoothing and forecasting mortality rates', *Statistical modelling* **4**(4), 279–298.

Czado, C., Delwarde, A. & Denuit, M. (2005), 'Bayesian poisson log-bilinear mortality projections', *Insurance: Mathematics and Economics* **36**(3), 260–284.

De Almeida, A. M. O. (1990), *Hamiltonian systems: chaos and quantization*, Cambridge University Press.

Dellaportas, P., Smith, A. F. & Stavropoulos, P. (2001), 'Bayesian analysis of mortality data', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **164**(2), 275–291.

Ellis, P. D. (2010), *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*, Cambridge University Press.

Forfar, D. O. (2004), 'Mortality laws', *Wiley StatsRef: Statistics Reference Online* .

Frühwirth-Schnatter, S. (1994), 'Data augmentation and dynamic linear models', *Journal of Time Series Analysis* **15**(2), 183–202.

Gage, T. B. & Mode, C. J. (1993), 'Some laws of mortality: how well do they fit?', *Human Biology* pp. 445–461.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2014), *Bayesian data analysis*, Vol. 2, Taylor & Francis.

Gelman, A. & Rubin, D. B. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical science* pp. 457–472.

Gelman, A. et al. (2006), 'Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper)', *Bayesian analysis* **1**(3), 515–534.

Geman, S. & Geman, D. (1984), 'Stochastic relaxation, gibbs distributions, and the bayesian restoration of images', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 721–741.

Gilks, W., Richardson, S. & Spiegelhalter, D. (1996), *Markov chain monte carlo in practice*, CHAPMAN & HALL.

Girosi, F. & King, G. (2003), 'Demographic forecasting', *Monograph* .

Gompertz, B. (1825), 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies', *Philosophical transactions of the Royal Society of London* **115**, 513–583.

Hannerz, H. (1999), *Methodology and applications of a new law of mortality*, Department of Statistics, Lund University.

Harrison, J. & West, M. (1999), *Bayesian Forecasting & Dynamic Models*, Springer.

Hartmann, M. (1987), 'Past and recent attempts to model mortality at all ages', *Journal of Official Statistics* **3**(1), 19–36.

Heligman, L. & Pollard, J. H. (1980), 'The age pattern of mortality', *Journal of the Institute of Actuaries* **107**(01), 49–80.

Hyndman, R. J. & Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: a functional data approach', *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

Jarner, S. F. & Kryger, E. M. (2011), 'Modelling adult mortality in small populations: The saint model', *Astin Bulletin* **41**(02), 377–418.

Kendall, M., Stuart, A. & Ord, J. (1987), 'Kendall's advanced theory of statistics'.

Keyfitz, N. (1991), 'Experiments in the projection of mortality', *Canadian Studies in Population* **18**(2), 1–17.

Kleinow, T. & Richards, S. J. (2016), 'Parameter risk in time-series mortality forecasts', Working paper, Heriot-Watt University.

Klugman, S. A. (1989), 'Bayesian modelling of mortality catastrophes', *Insurance: Mathematics and Economics* **8**(3), 159–164.

Kogure, A., Kitsukawa, K. & Kurachi, Y. (2009), 'A bayesian comparison of models for changing mortalities toward evaluating longevity risk in japan', *Asia-Pacific Journal of Risk and Insurance* **3**(2).

Kogure, A. & Kurachi, Y. (2010), 'A bayesian approach to pricing longevity risk based on risk-neutral predictive distributions', *Insurance: Mathematics and Economics* **46**(1), 162–172.

Kostaki, A. (1988), *Contributions to the methodology and application of the Heligman-Pollard formula*, Department of Statistics, University of Lund.

Lee, R. D. & Carter, L. R. (1992), 'Modeling and forecasting us mortality', *Journal of the American statistical association* **87**(419), 659–671.

Lee, R. & Miller, T. (2001), 'Evaluating the performance of the lee-carter method for forecasting mortality', *Demography* **38**(4), 537–549.

Leonard, T. & Hsu, J. S. (1999), *Bayesian methods: an analysis for statisticians and interdisciplinary researchers*, Vol. 5, Cambridge University Press.

Little, R. J. & Rubin, D. B. (2014), *Statistical analysis with missing data*, John Wiley & Sons.

Liu, X. & Braun, W. J. (2011), 'Investigating mortality uncertainty using the block bootstrap', *Journal of Probability and Statistics* **2010**.

Makeham, W. M. (1860), 'On the law of mortality and the construction of annuity tables', *The Assurance Magazine, and Journal of the Institute of Actuaries* **8**(6), 301–310.

McNown, R. & Rogers, A. (1992), 'Forecasting cause-specific mortality using time series methods', *International Journal of Forecasting* **8**(3), 413–432.

McNown, R., Rogers, A. & Little, J. (1995), 'Simplicity and complexity in extrapolatwe population forecasting models', *Mathematical Population Studies* **5**(3), 235–257.

Mode, C. J. & Busby, R. C. (1982), 'An eight-parameter model of human mortalitythe single decrement case', *Bulletin of Mathematical Biology* **44**(5), 647–659.

Mode, C. J. & Jacobson, M. E. (1984), 'A parametric algorithm for computing model period and cohort human survival functions', *International journal of bio-medical computing* **15**(5), 341–356.

Neal, R. M. (1993), 'Probabilistic inference using markov chain monte carlo methods'.

Nielsen, B. & Nielsen, J. P. (2014), 'Identification and forecasting in mortality models', *The Scientific World Journal* **2014**.

Nydick, S. W. (2012), 'The wishart and inverse wishart distributions', *J. Stat* .

Pedroza, C. (2002), *Bayesian hierarchical time series modeling of mortality rates*, Harvard University.

Pedroza, C. (2006), 'A bayesian forecasting model: predicting us male mortality', *Biostatistics* **7**(4), 530–550.

Plat, R. (2009), 'On stochastic mortality modeling', *Insurance: Mathematics and Economics* **45**(3), 393–404.

Pollard, J. H., Benjamin, B. & Soliman, A. (1987), 'Projection of age-specific mortality rates.', *Population Bulletin of the United Nations* (21/22), 55–69.

Raftery, A. E., Lewis, S. et al. (1992), 'How many iterations in the gibbs sampler', *Bayesian statistics* **4**(2), 763–773.

Rao, C. R. (2009), *Linear statistical inference and its applications*, Vol. 22, John Wiley & Sons.

Reichmuth, W. & Sarferaz, S. (2008), *Bayesian Demographic Modeling and Forecasting: An Application to U.S. Mortality*, Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät.

Renshaw, A. E. & Haberman, S. (2003), 'Lee–carter mortality forecasting with age-specific enhancement', *Insurance: Mathematics and Economics* **33**(2), 255–272.

Renshaw, A. E. & Haberman, S. (2006), 'A cohort-based extension to the lee–carter model for mortality reduction factors', *Insurance: Mathematics and Economics* **38**(3), 556–570.

Rogers, A. & Gard, K. (1990), 'Applications of the heligman/pollard model mortality schedule.', *Population Bulletin of the United Nations* (30), 79–105.

Rogers, A. & Planck, F. (1983), 'Model: a general program for estimating parametrized model schedules of fertility mortality migration and marital and labor force status transitions.'.

Schafer, J. L. (1997), *Analysis of incomplete multivariate data*, CRC press.

Schafer, J. L. & Olsen, M. K. (1998), 'Multiple imputation for multivariate missing-data problems: A data analyst's perspective', *Multivariate behavioral research* **33**(4), 545–571.

Searle, S. R. (1971), 'Linear models. new york: J', *Wiley& Sons* .

Siler, W. (1983), 'Parameters of mortality in human populations with widely varying life spans', *Statistics in medicine* **2**(3), 373–380.

Spall, J. C. (1988), *Bayesian analysis of time series and dynamic models*, Dekker.

Tabeau, E. (2001), A review of demographic forecasting models for mortality, *in* 'Forecasting Mortality in Developed Countries', Springer, pp. 1–32.

Tabeau, E., Willekens, F. & van Poppel, F. (2002), Parameterisation as a tool in analysing age, period and cohort effects on mortality: a case study of the netherlands, *in* 'The Life Table', Springer, pp. 141–169.

Wilks, S. S. (1938), 'The large-sample distribution of the likelihood ratio for testing composite hypotheses', *The Annals of Mathematical Statistics* **9**(1), 60–62.

Willekens, F. & Baydar, N. (1984), 'Age-period-cohort models for forecasting fertility.'.

Willets, R. (2004), 'The cohort effect: insights and explanations', *British Actuarial Journal* **10**(04), 833–877.

Wilmoth, J. R. (1993), Computational methods for fitting and extrapolating the lee-carter model of mortality change, Technical report, Technical report, Department of Demography, University of California, Berkeley.

Yashin, A. I., Iachine, I. A. & Begun, A. S. (2000), 'Mortality modeling: A review', *Mathematical Population Studies* **8**(4), 305–332.