A Bayesian approach to modelling mortality, with applications to insurance

George Lindsay Cairns

A thesis for the degree of Doctor of Philosophy
completed in the School of Mathematics and Statistics,
submitted to College of Science and Engineering
University *of* Glasgow

May 2013

Abstract

The purpose of this research was to use Bayesian statistics to develop flexible mortality models that could be used to forecast human mortality rates. Several models were developed as extensions to existing mortality models, in particular the Lee-Carter mortality model and the age-period-cohort model, by including some of the following features: age-period and age-cohort interactions, random effects on mortality, measurement errors in population count and smoothing of the mortality rate surface. One expects mortality rates to change in a relatively smooth manner between neighbouring ages or between neighbouring years or neighbouring cohorts. The inclusion of random effects in some of the models captures additional fluctuations in these effects. This smoothing is incorporated in the models by ensuring that the age, period and cohort parameters of the models have a relatively smooth sequence which is achieved through the choice of the prior distribution of the parameters. Three different smoothing priors were employed: a random walk, a random walk on first differences of the parameters and an autoregressive model of order one on the first differences of the parameters. In any model only one form of smoothing was used. The choice of smoothing prior not only imposes different patterns of smoothing on the parameters but is seen to be very influential when making mortality forecasts.

The mortality models were fitted, using Bayesian methods, to population data for males and females from England and Wales. The fits of the models were analysed and compared using analysis of residuals, posterior predictive intervals for both in-sample and out-of-sample data and the Deviance Information Criterion. The models fitted the data better than did both the Lee-Carter model and the age-period-cohort model.

From the analysis undertaken, for any given age and calendar year, the preferred model based on the Deviance Information Criterion score, for male and female death counts was a Poisson model with the mean parameter equal to the number of lives exposed to risk of dying for that age in that calendar year multiplied by a mortality parameter. The logit of this mortality parameter was a function of the age, year (period) and cohort with additional interactions between the age and period parameters and between the age and cohort parameters. The form of parameter smoothing that suited the males was an autoregressive model of order one on the first differences of the parameters and that for the females was a random walk. Moreover, it was found useful to add Gaussian random effects to account for overdispersion caused by unobserved heterogeneity in the population mortality.

The research concluded by the application of a selection of these models to the provision of forecasts of period and cohort life expectancies as well as the numbers of centenarians for males and females in England and Wales. In addition, the thesis illustrated how Bayesian mortality models could be used to consider the impact of the new European Union solvency regulations for insurers (Solvency II) for longevity risk.

This research underlined the important role that Bayesian stochastic mortality models can play in considering longevity risk.

Declaration

I, George Lindsay Cairns, declare that this thesis titled, 'A Bayesian approach to modelling mortality, with applications to insurance' and the work presented in it is my own. I confirm that:

- This work was done wholly or mainly while studying for a research degree at Glasgow University.
- No part of this thesis has previously been submitted for a degree or any other qualification at Glasgow University or any other institution.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed: _____

Date: _____

Acknowledgements

Firstly I would like to thank my supervisors Agostino Nobile and Michael Titterington for their valuable experience, knowledge, patience and support over the last four years. None of this work would have been possible without their help and encouragement. I am also very grateful to my supervisor Vincent Macaulay for his valuable help and input in this project.

I gratefully acknowledge the financial support provided by the Engineering and Physical Sciences Research Council from the University of Glasgow.

I am also grateful to all the people in the department for their friendship and assistance through my 4 years of study.

Finally, I need to thank my wife Ann and my two sons Steven and Andrew who have happily put up with my absences as I worked on this project.

To my wife Ann and sons Steven and Andrew and my parents.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

In general, people in the U.K. are continuing to live longer (Office for National Statistics (2011a)). This is good news but has given rise to serious social and economic concerns.

The goal of this research is to develop new flexible mortality models that are capable of forecasting mortality that can assist in understanding these longevity concerns and therefore contribute to more informed decision making.

Increasing longevity has already brought some social and economic changes in the U.K., but the prospect of further increases to longevity continues to worry governments, companies, organisations and individuals (see Vojak, F. (2011)). These concerns include the following:

- social security and public health care for the elderly,
- restructuring of public pensions,
- closure or restructuring of final salary pension schemes in the private sector,
- falling living standards in retirement.

These fears are largely emanate out of the uncertainty regarding human longevity. For example, increasing longevity has already had a significant impact on private and public pension schemes in the U.K. where sponsors of pension schemes have seen pension costs increase not only as members live longer but more importantly

as the expectation of the members' lives increases. The consequence of these concerns has been that, in the U.K., most final salary schemes in the private sector have closed to new members, moving the longevity risk on to the individual new member as the replacement pension schemes do not offer a fixed pension at retirement but a cash amount that is converted to pension at the prevailing mortality at the time of retirement (Office for National Statistics (2012b) and National Association of Pension Funds (2010)). In the U.K. public sector there has been much negotiation about restructuring final salary pension schemes by a combination of increasing the retirement age, increasing member contributions and reducing the rate at which pension benefits accrue to scheme members. The U.K. governments have been introducing legislation to increase the age that the basic state pension is payable: legislation was agreed back in 1995 gradually to increase women's State Pension Age from 60 to 65 over a ten year period starting in April 2010. The Pensions Act 2011 speeds up that process so that women's State Pension Age will now reach 65 by November 2018 and then State Pension Age for both men and women will increase to 66 by October 2020. Other legislation increases State Pension Age to 68 by 2046 but the Government has announced that it wants this increase to happen faster (Department of Work and Pensions (2011)).

In recent years, the capital markets have introduced longevity bonds and other mortality bonds as an attempt to transfer some of the longevity risks to the market. So far these have not been a huge success, one of the reasons being the challenge of determining what a fair price might be for the risk transfer.

Throughout European Community, the governments are committed to introducing the Solvency II Directive, which is a new capital and risk framework for the European Insurance sector (European Parliament (2009)). The key quantitative capital requirement is defined in terms of holding sufficient capital to provide protection for the company to cover all losses which may occur in the following year with a probability of at least 99.5%. Longevity risk is one of the key risks: how does one estimate this required level of capital?

All of these interested groups can benefit from stochastic mortality models, which yield distributional statements on the probabilities of outcomes instead of pure projections of some scenarios. For this purpose, credible stochastic mortality models that help quantify this uncertainty are very valuable.

The last twenty years has been an exciting period in the development of stochastic mortality models. A benchmark model was developed by R. D. Lee and L. Carter, (Lee and Carter (1992)). Their aim was to develop a mortality model to make long-run forecasts of age-specific mortality in U.S.A.. Since then the model has been applied to many different populations including the U.K. and has been modified and extended in different ways by other researchers, such as, Wilmoth (1993), Brouhns et al. (2002), Renshaw and Haberman (2003b,a, 2006), Cairns et al. (2006) and Delwarde et al. (2007). Another benchmark model that has been used for many decades is the age-period-cohort model used by Holford (1983, 1991, 2006), Clayton and Schifflers (1987a,b).

This thesis extends further these developments of the Lee-Carter and age-period-cohort models using a Bayesian framework. Over the last decade a number of researchers have applied Bayesian analysis to the Lee-Carter model and simple variations of that model; see for example Czado et al. (2005) and Pedroza (2006). Over the past three decades researchers have applied Bayesian analysis to the age-period-cohort model; see for example Berzuini and Clayton (1994), Besag et al. (1995), Knorr-Held and Rainer (2001) and Schmid and Held (2007).

Bayesian analysis is a very flexible modelling paradigm and is ideally suited to dealing with multiple sources of uncertainty that arise in real world problems like human mortality. In Bayesian models the parameters are considered random variables and therefore these models can assist in assessing the variability of the parameters which is not easily done using a frequentist approach. Bayesian models can deal with complex non-linear structures, including where the variables are latent or where there is missing data; all these features can be dealt with within the one coherent model. Further, model comparison using Bayesian methods is not constrained

to nested models; they can compare widely different models with totally different structures.

A Bayesian approach is followed in this thesis using markov chain monte carlo methods to fit the Lee-Carter model and the main extensions to that model as well as the age-period-cohort model and also the new more complex models, using population data from England and Wales. Then, after demonstrating that the new models give an improved fit, the thesis considers the uncertainty of the fit and forecasting from these new mortality models. The thesis also compares the results of model forecasts with official forecasts by the Office for National Statistics, see ONS March 2012.

The thesis is organized as follows. Chapter 2 contains a description of the data and notation used. Chapter 3 reviews the literature covering the key mortality models that have been developed. Chapter 4 specifies the various models to be considered in this thesis. Chapter 5 discusses the fit of these models and compares the various models using the Deviance Information Criterion. It then considers the forecasting methodology and, for the better-fitting models, considers how well the model fits out of sample data. Chapter 6 applies a selection of these models to forecast life expectancy in future years as well forecasting the number of centenarians using population data from England and Wales. Concluding remarks are made in the final chapter.

Throughout this thesis, the data analysis and the production of the figures was done using the R package (R Development Core Team (2011)).

CHAPTER 2

Notation, Basic Terminology and Data

2.1 Notation and Basic Terminology

In mortality analysis there are a number of standard terms that are here defined as depending on age and calendar year, as mortality is assumed to change with time.

$q_{x,T}$ is the probability that a life aged exactly x in calendar year T dies before age $x+1$;

$S_x(t,T)$ is probability that a life aged exactly x in calendar year T survives t years to age $x+t$, this also known as the survivor function in survival analysis;

$\mu_{x,t}$ is force of mortality at age x and exact time t; this is often referred to as the hazard function in survival analysis;

$m_{x,T}$ is central mortality rate at age x as measured in calendar year T. This is the rate at which deaths are occurring over the year of age x to $x+1$ exact as measured in calendar year T, relative to the expected amount of time that a life, initially alive at age x will spend alive in that year of age as measured in calendar year T. The formula for $m_{x,T}$ is shown in equation 2.1.3. In this thesis I call it the mortality rate.

There are standard relationships between these mortality variables:

$$\hat{\mu}_{x+t,T} = -\frac{d}{dt}\log S_x(t,T). \tag{2.1.1}$$

$$q_{x,T} = \int_0^1 S_x(t,T)\mu_{x+t,T}dt. \tag{2.1.2}$$

$$m_{x,T} = \frac{q_{x,T}}{\int_0^1 S_x(t,T)dt}. \tag{2.1.3}$$

The derivations of these standard relationships are given in 'Notes on Survival Models' (Konstantopoulos (2006)).

The mortality rate $m_{x,T}$ is usually estimated from the observed data using:

$$m_{x,t} = \frac{d_{x,T}}{E_{x,T}} \tag{2.1.4}$$

where $d_{x,T}$ is number of deaths during calendar year T aged x last birthday, and $E_{x,T}$ is the exposure to risk of death of individuals aged x during calendar year T. The latter is usually approximated by the average of the population size aged x at the beginning of year T and population size aged x at the beginning of year $T+1$. See, for example, Heligman and Pollard (1980). As discussed later the data used in this thesis came from the Human Mortality Database (HMD) (www.mortality.org) where data are analysed by cohort, and because of this, $E_{x,T}$ is average of the population size as described except for a small correction that reflects the timing of deaths during the interval.

It is usual from census and other data sources that the numbers of deaths are available by sex, calendar year and age as well as the population size by age and sex at the date of the census. Therefore, the calculation of raw mortality rates is relatively straightforward using equation 2.1.4. The estimated probabilities of death and the estimated force of mortality can then be calculated from the raw mortality rates by making two relatively noncontentious assumptions: (1) the force of mortality is assumed constant over each age and calendar year , i.e. $\mu_{x+u,t+s} = \mu_{x,t}$ for integers x and t and $u, s \in (0,1)$; and (2) the size of the population at all ages remains constant over the calendar year. From equations 2.1.3 and 2.1.2, when the force of mortality is constant over each age and calendar year we can deduce that $\mu_{x,t} = m_{x,t}$. From

these assumptions one can derive the relationship that

$$q_{x,t} = 1 - \exp(-m_{x,t}). \tag{2.1.5}$$

Another basic demographic concept, the life expectancy of an individual aged x, is loosely defined as the future average lifetime of that individual. It can be calculated in two ways: by the '*period life expectancy*' or the '*cohort life expectancy*' methods.

Period life expectancy at a given age is the average number of years a person would live, if he or she experienced the particular age-specific mortality rates for that time period throughout his or her life. It makes no allowance for any later actual or projected changes in mortality. In practice, death rates are likely to change in the future so period life expectancy does not give the number of years someone could actually expect to live.

Despite this drawback, period life expectancies are a useful measure of mortality rates actually experienced over a given period and, for past years, provide an objective means of comparison of the trends in mortality over time. Official life tables, in the U.K. and in other countries, which relate to past years are generally period life tables for these reasons.

Cohort life expectancies are calculated using age-specific mortality rates which allow for known or projected changes in mortality in later years and are thus regarded as a more appropriate measure than period life expectancy of how long a person of a given age would be expected to live, on average.

For example, period life expectancy at age 65 in 2000 would be worked out using the mortality rate for age 65 in 2000, for age 66 in 2000, for age 67 in 2000, and so on. Cohort life expectancy at age 65 in 2000 would be worked out using the mortality rate for age 65 in 2000, for age 66 in 2001, for age 67 in 2002, and so on. For mortality in future years from the current year, cohort life expectancy incorporates an assumption on how future mortality will develop in these future years.

Based on the two assumptions noted above, one can derive mathematical expressions for the period life expectancy and cohort life expectancy at age x in year t:

Period life expectancy

$$e^{p}_{x,t} = \frac{[1-\exp(-m_{x,t})]}{m_{x,t}} + \sum_{s=1}^{\omega} \frac{[1-\exp(-m_{x+s,t})]}{m_{x+s,t}} \prod_{r=0}^{s-1}(1-q_{x+r,t}) \tag{2.1.6}$$

Cohort life expectancy

$$e^{c}_{x,t} = \frac{[1-\exp(-m_{x,t})]}{m_{x,t}} + \sum_{s=1}^{\omega} \frac{[1-\exp(-m_{x+s,t+s})]}{m_{x+s,t+s}} \prod_{r=0}^{s-1}(1-q_{x+r,t+r}) \tag{2.1.7}$$

where ω is the limit on lifespan, for this report assumed to be 120. Equations 2.1.5, 2.1.6 and 2.1.7 are derived in Appendix A.

2.2 Data

This thesis considers various mortality models in the analysis of male and female longevity and then how to forecast how longevity might develop in the future. The data used was age-specific and sex-specific population data from England and Wales.

Data were obtained from the Human Mortality Database (HMD), University of California, Berkeley (USA), and the Max Planck Institute for Demographic Research (Germany), available at www.mortality.org and www.humanmortality.de respectively (data downloaded on 17^{th} September 2010). The data used in this report concerned 'Total Population' of England and Wales for ages 60 to 100 and years 1960 to 2006. The ages below 60 were excluded as the focus of the thesis is on the longevity of those of pensionable age and the period excludes pre-war and immediately post-war censuses and so complications caused by war are avoided. All of the raw data used for the HMD came from the Office for National Statistics (ONS). However, the data have undergone some manipulation. These are described in detail in the Methods Protocol (Wilmoth et al. (2007)).

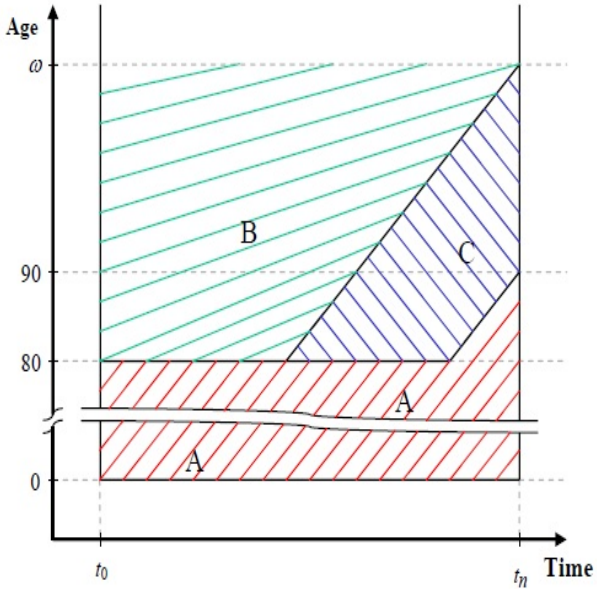
We now give a brief description of the key manipulations.

(1) Where there are individuals of unknown age, in either death or census counts,

the number of these individuals are distributed across the age range in proportion to the number of observed individuals in each age group.

(2) At ages 80 and older (except for those cohorts who are younger than age 90 at the end of the observation period), population size estimates are calculated using either the method of extinct cohorts or the survivor ratio method. For those cohorts who are younger than age 90 at the end of the observation period as well as ages under 80, population estimates are obtained by applying the method of intercensal survival. Figure 2.1 is taken from the Methods Protocol (Wilmoth et al. (2007)) and shows which method is used to estimate population sizes.

Figure 2.1: Method used for population size estimates. Area A population estimates by ONS official estimate or intercensal survival, area B estimates by extinct cohorts and area C estimates by survival ratio.



(3) The extinct cohort method relies on the cohort being extinct based on the cohort attaining a specific age by the end of the observation period t_n. To calculate this age, the HMD adopted a method proposed by Kannisto (1994) see also Andreev (2001). Then, for these extinct cohorts, the population size at age x at time t is estimated by summing all future deaths from time t for the cohort from age x at time t. This method assumes that there is no international migration after age x for the cohort in question, which is a reasonable assumption only for advanced ages.

(4) For non-extinct cohorts, the survivor ratio method uses a survival ratio based on extinct cohorts to estimate the population at age $\omega - 1$ for the non-extinct cohorts.

The populations for ages $\omega - 2, \ldots$ can be estimated by working backwards from the cohort population at age $\omega - 1$ by adding on the number of deaths in a similar manner to the extinct cohort method. In calculating the survivor ratio the HMD aggregates the previous 5 cohorts and the ratio is calculated using those in the same cohort who were alive 5 years earlier. One further adjustment was made to allow for the trend in mortality that may be declining, increasing or constant, by multiplying the ratio by a factor that is greater than, less than or equal to one respectively. The value of this factor is obtained by adjusting the population calculated by the survivor ratio method so that it equals the population census estimate where the estimate at a given age (In this case this age is 90) is believed to be accurate.

(5) Population estimates between census years are calculated by the method of intercensal survival. This simply involves subtracting death counts from the census count at the beginning of the period to obtain cohort population estimates for each succeeding year. Unfortunately, the final step of such a computation usually yields an estimate of cohort size at time $t + 10$ that differs from the number given by the corresponding census. Then, estimates of cohort size for intercensal years are found by subtracting, from the initial census count, both the observed death counts and an estimate of net migration/error. A similar approach is followed when the population is estimated after the last census year.

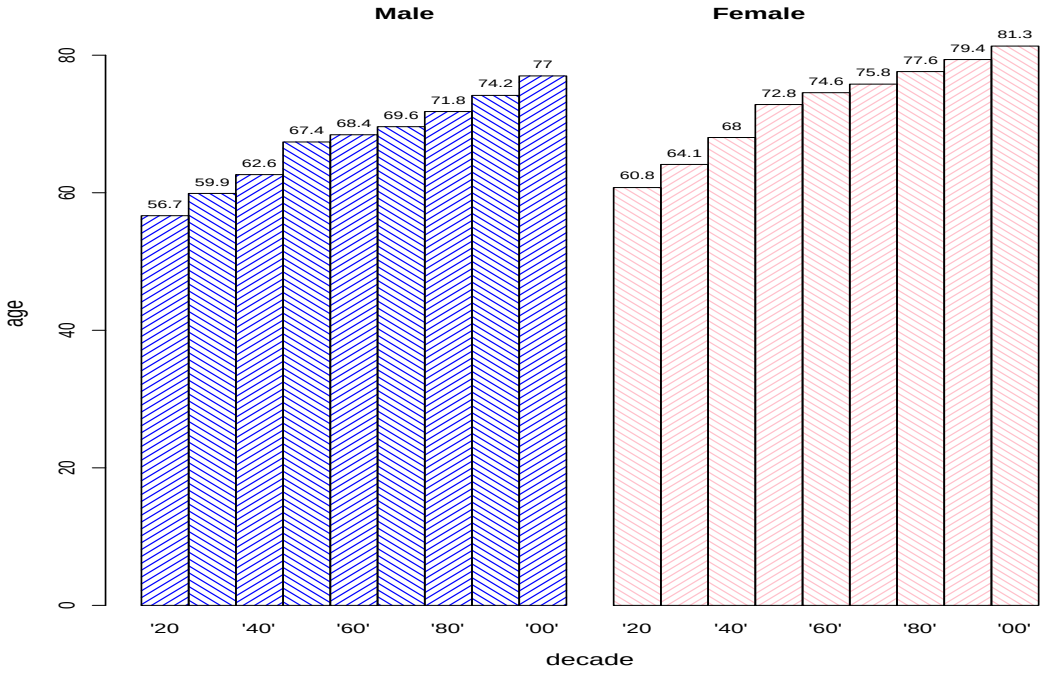
(6) The calculation of exposure or exposure-to-risk for some age-time interval, say age x in calendar year t, (denoted by $E_{x,t}$) is a count of the time that individuals of age x in year t, should they die, would be included in the death count $d_{x,t}$ for age x in year t. Estimates of the population exposed to the risk of death during some age-time interval, say age x in year t, is based on average annual population estimates for age x in year t and x in year $t + 1$, with a small correction that reflects the timing of deaths during the interval. This correction arises because the HMD works mainly with cohorts; many other demographers and actuaries working with population data use the average annual population estimates for age x in year t and x in year $t + 1$.

2.3 High Level Analysis of Longevity

England and Wales, like many other countries, has seen increasing longevity of its population; Figure 2.2 illustrates how the life expectancy at birth has steadily increased for both males and females in past decades. Life expectancy for a given decade is the life expectancy at birth averaged over each year in the decade.

Life expectancy figures from HMD show that over the 90 years shown there has been a significant increase in expected lifetime for both males and females. In the mid 1920s males lived on average 56.7 years and females 60.8 years from birth and by the mid 2000s these had increased to 77.0 years and 81.3 years for males and females respectively, an increase of 20 years.

Figure 2.2: Period life expectancy at birth (in years of age) for males and females in each of the decades from 1920s to 2000s. Figures from HMD.



The increase in life expectancy is broken down in a mosaic plot (Figure 2.3). The width of each column is proportional to the increase in life expectancy at birth over the previous decade relative to the overall increase in life expectancy at birth over the entire period 1920s to mid 2000s. Each column is then divided into sections by age groups, namely under 20, age 20 to under 40, etc. The length of the section

reflects the proportion of the increase in life expectancy over the previous decade contributed by that age group. Over some decades for some age groups there was a reduction in life expectancy. These reductions were set to zero and are indicated by the dotted lines. The mosaic plot shows that ages contributing to the increase in life expectancy have changed with time. Up to the 1960s the increase in expectation has been driven by the younger ages, the under 20s and 20-40, age groups, for both males and females. However, since the 1960s, the age groups 60 and over have progressively become the main age groups driving the increase in life expectancy.

This analysis indicates that the underlying drivers for mortality improvement are not static and have changed over time. This makes mortality projections more uncertain. There are many competing influences that may affect future mortality, on the positive side the development of medicines and medical treatments arising from genetic research and on the negative side obesity and sedentary life styles. Without a crystal ball, no one can judge the resultant effect; in this thesis the underlying assumption is that the overall forces of change continue although the individual reasons may differ. However, from the mosaic plots, the pattern of change in mortality is similar for both males and females but the changes occur at different rates. In addition, the mosaic plots indicate that both calendar year and age are factors that can help describe the changes in mortality rates. Age of course is the key factor affecting mortality rates.

Figures 2.4, 2.5 and 2.6 show the observed mortality rate (on a log scale) for each age and sex over the period of investigation. The blue and red lines in each panel denote male and female mortality respectively. The plots clearly show that mortality has been reducing with time for almost all but the very oldest ages and that there is a general pattern that the rate of decline has been reducing with age. There is also evidence of cohort effects. For example, males and females born around 1920 have a spike in mortality rate; and this spike is seen in each of the age plots from 60 up to 80, progressively moving to the right as the age increases. Note the vertical scale is the mortality rate (not the log mortality) and the scales are different in each panel.

Figures 2.7, 2.8 and 2.9 are plots of the logarithm of observed mortality rates for males and females *by cohort*. Again, the blue and red lines in each panel denote

Figure 2.3: Analysis of the change in period life expectancy by decade and age group. The width of each column is proportional to the change in life expectancy relative to the overall change in life expectancy from the 1920s to mid 2000s. The length of each block is proportional to the change in life expectancy for that age group corresponds to the overall change in life expectancy over the decade.

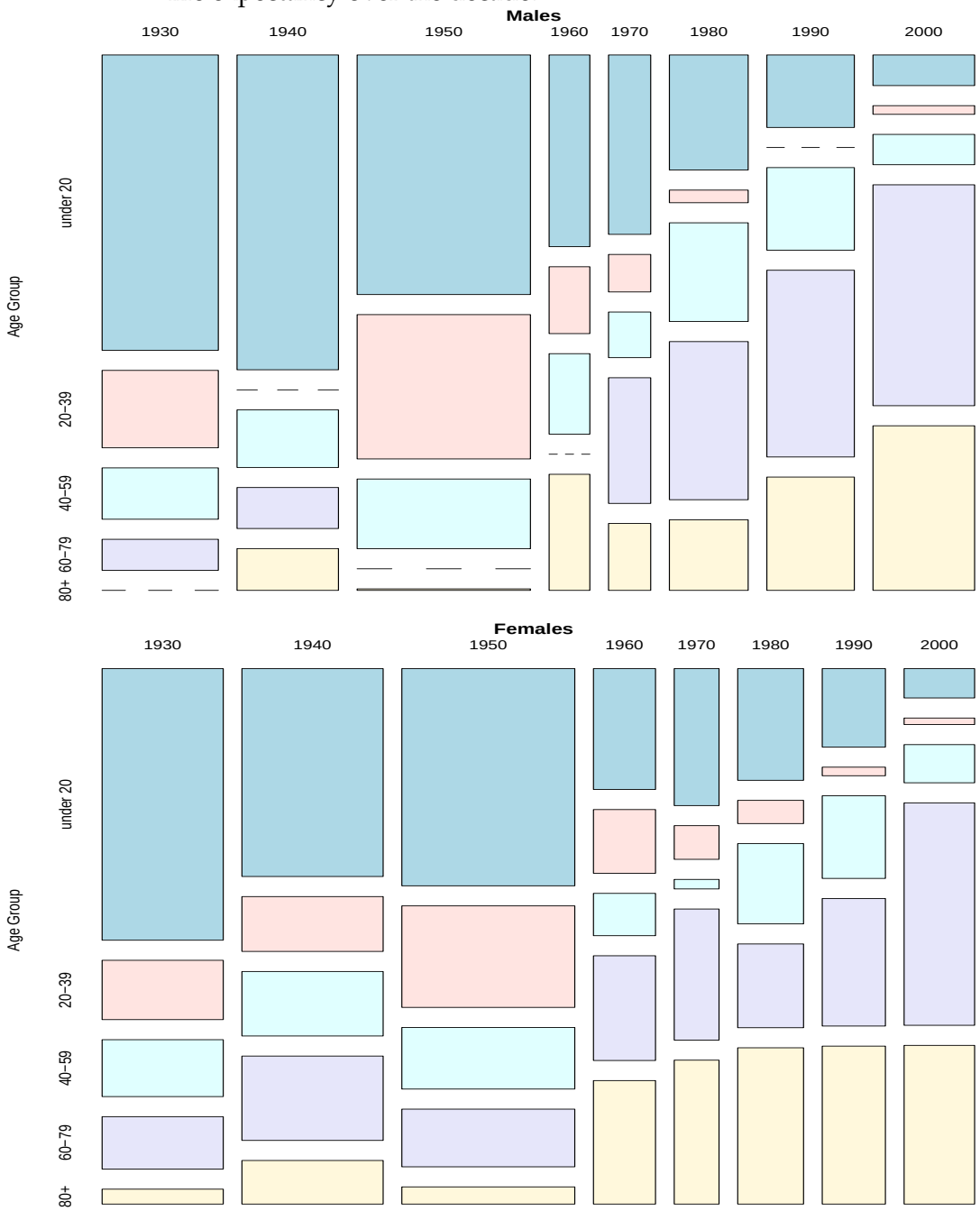


Figure 2.4: Panels of the observed male and female mortality rates (on a log scale) by age for ages 60 to 74. For each age, the blue line denotes male mortality and the red line denotes female mortality. Calendar year is along the x-axis and the y-axis corresponds to the mortality rate.

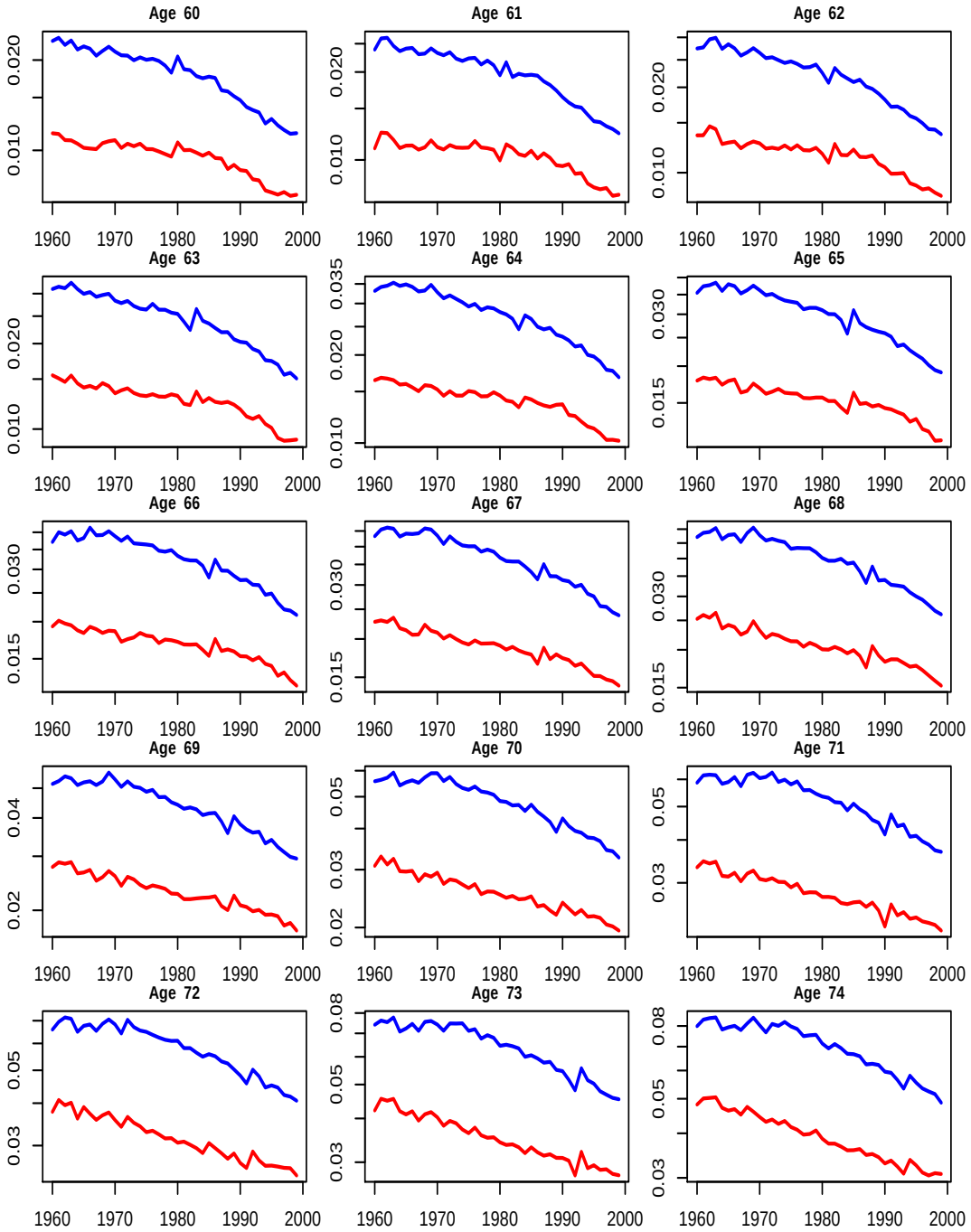


Figure 2.5: Panels of the observed male and female mortality rates (on a log scale) by age for ages 75 to 89. For each age, the blue line denotes male mortality and the red line denotes female mortality. Calendar year is along the x-axis and the y-axis corresponds to the mortality rate.

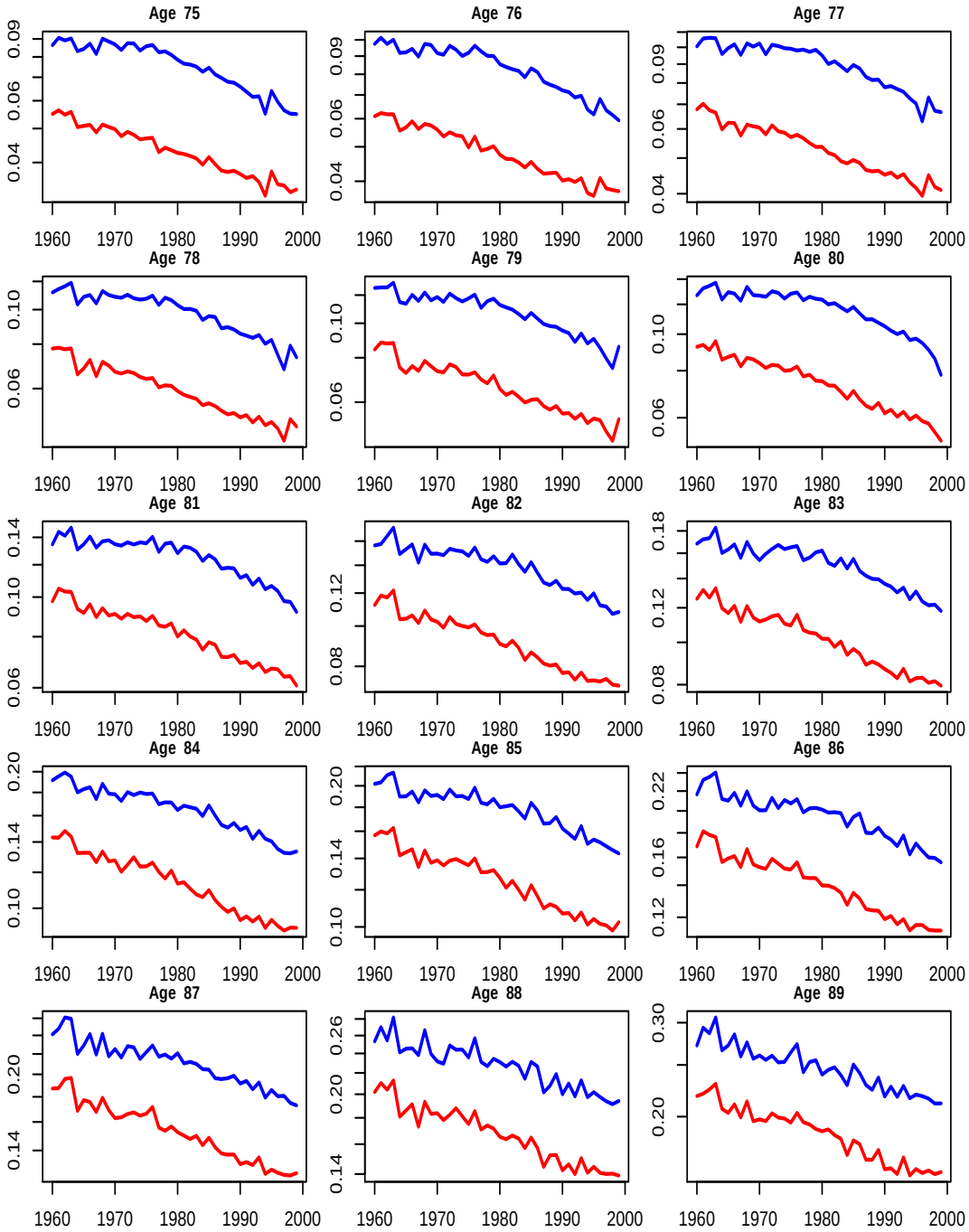
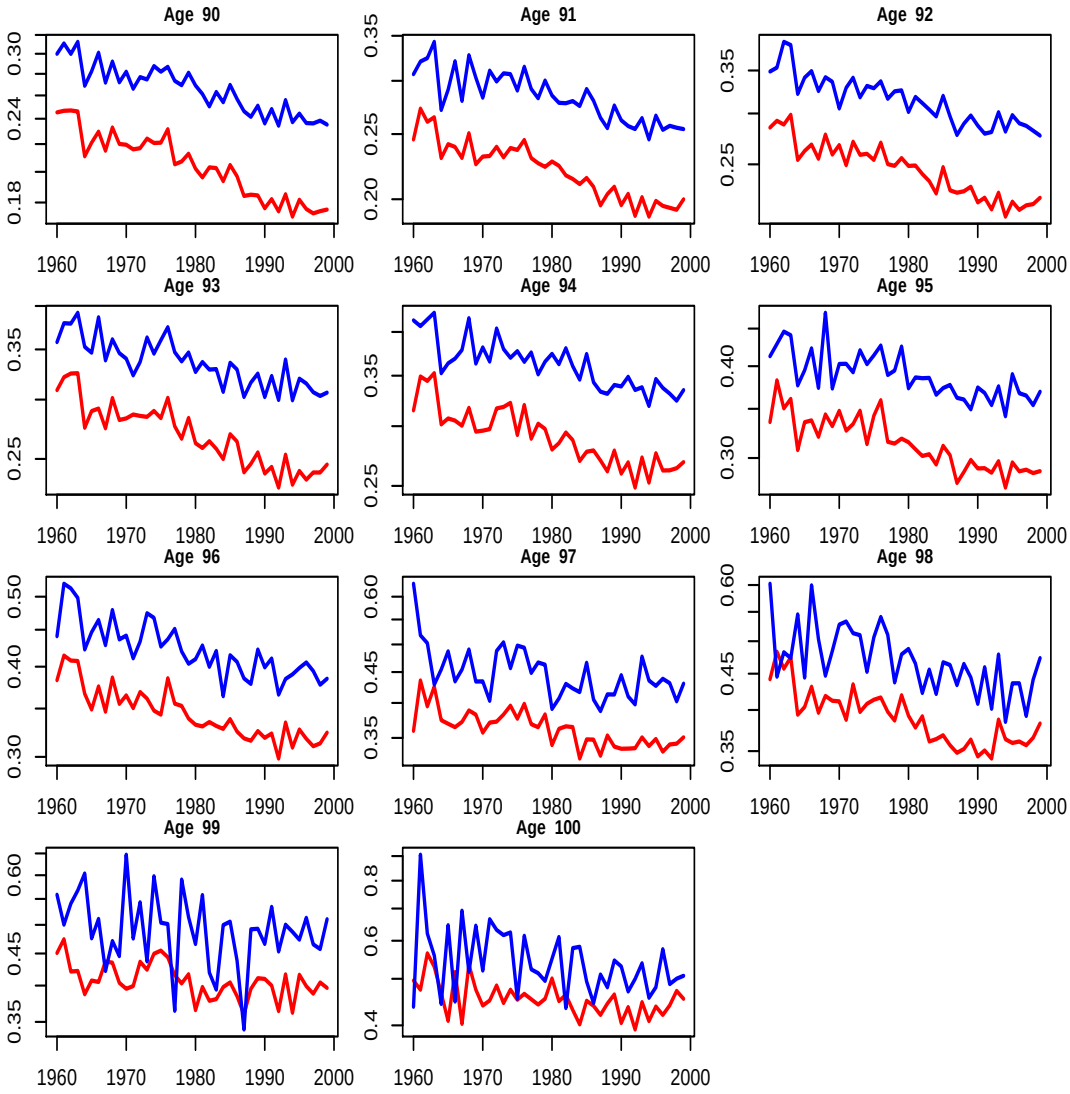


Figure 2.6: Panels of the observed male and female mortality rates (on a log scale) by age for ages 90 to 100. For each age, the blue line denotes male mortality and the red line denotes female mortality. Calendar year is along the x-axis and the y-axis corresponds to the mortality rate.



male and female mortality, respectively. Across the plots the scale of the y-axis is the same; the scale is in mortality not log-mortality. The x-axis is age from 60 to 100. Some plots are not complete, because the whole cohort is left or right censored by the data extract. As time passes the age of the cohort increases and this is the main cause of mortality rate increasing with time. The plots of log-mortality appear roughly linear but the scale (as it accommodates ages from 60 up to 100) obscures the fluctuations in mortality. A closer examination of the plots reveals that the slope of the line changes gradually by cohort.

The accuracies of some of the approximations that are made are based on the assumption that the population size is constant for neighbouring cohorts. Figure 2.10 is a plot of the number of male and female births over the 80 cohorts covered by the data. The plots show that there has been a large variation in cohort size over the period but, apart from the years around the First World War where there is a marked change, change has occurred gradually and the sizes of the neighbouring cohorts are sufficiently similar not to be considered to invalidate the assumption. The filled points highlighted show the most extreme change in cohort size.

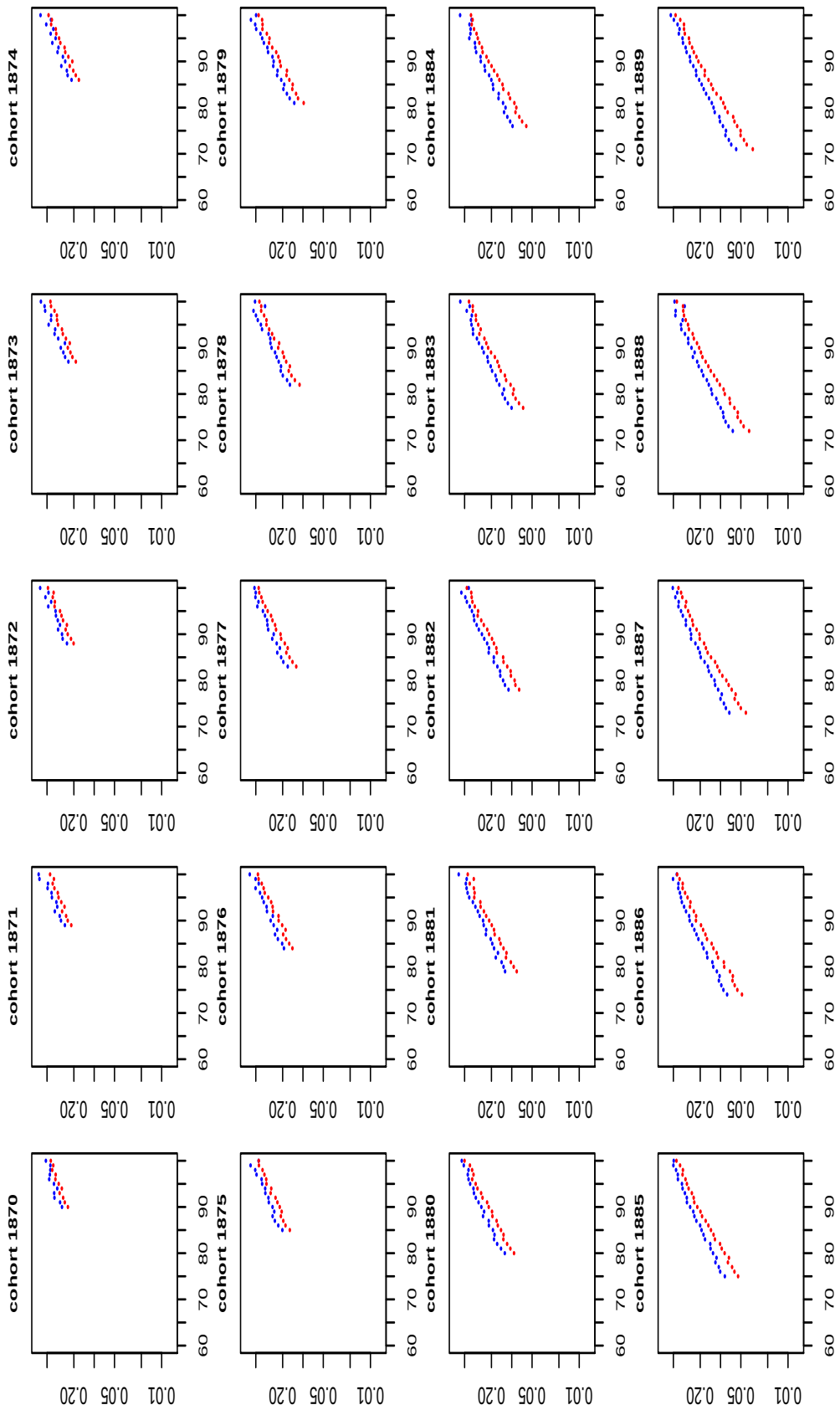


Figure 2.7: Plots of the observed male and female mortality rates (on a log scale) by age and cohort (year of birth), from 1870 to 1889. The blue points denote males and the red females. The x-axis is age and the y-axis corresponds to the mortality rate.

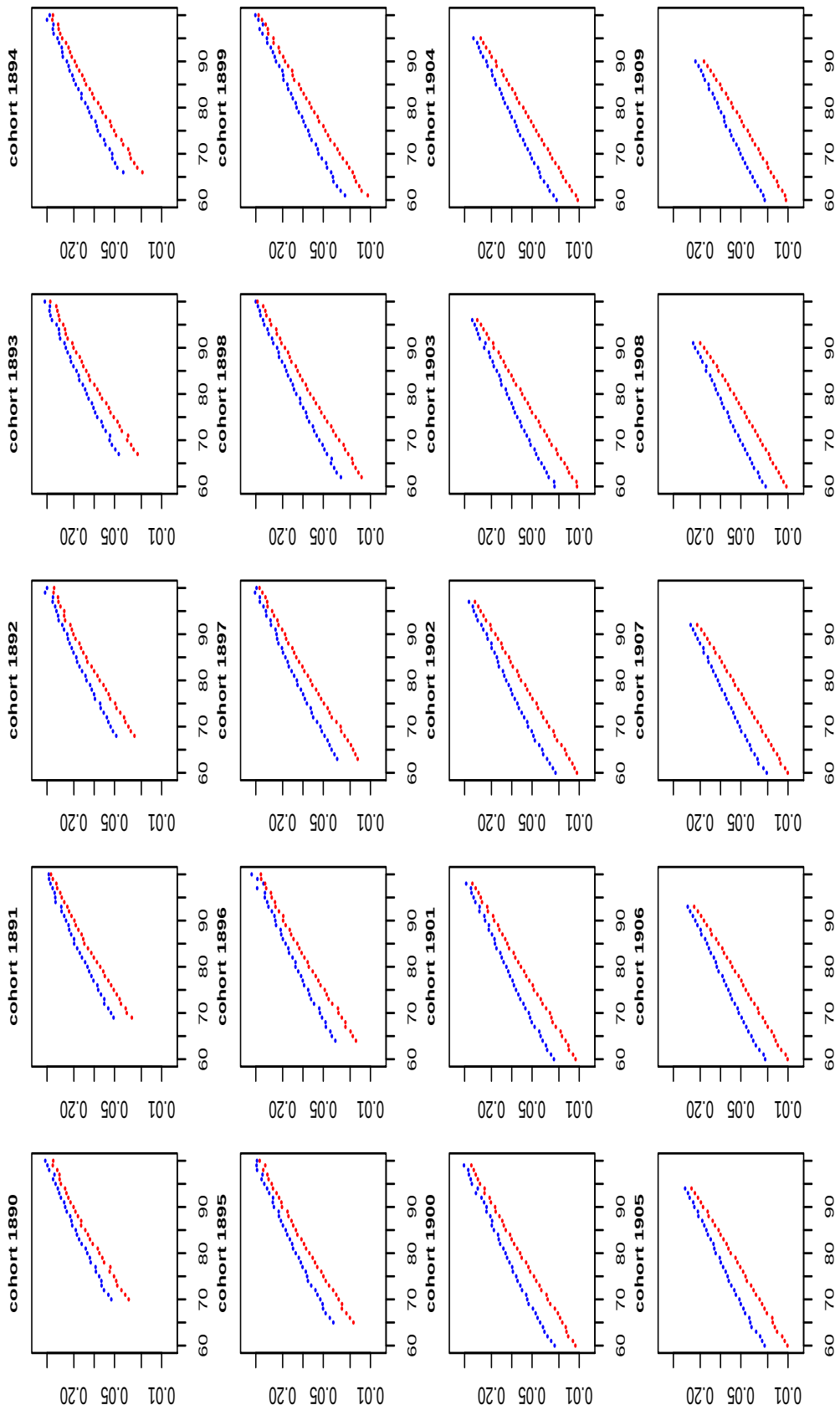


Figure 2.8: Plots of the observed male and female mortality rates (on a log scale) by age and cohort (year of birth), from 1890 to 1909. The blue points denote males and the red females. The x-axis is age and the y-axis corresponds to the mortality rate.

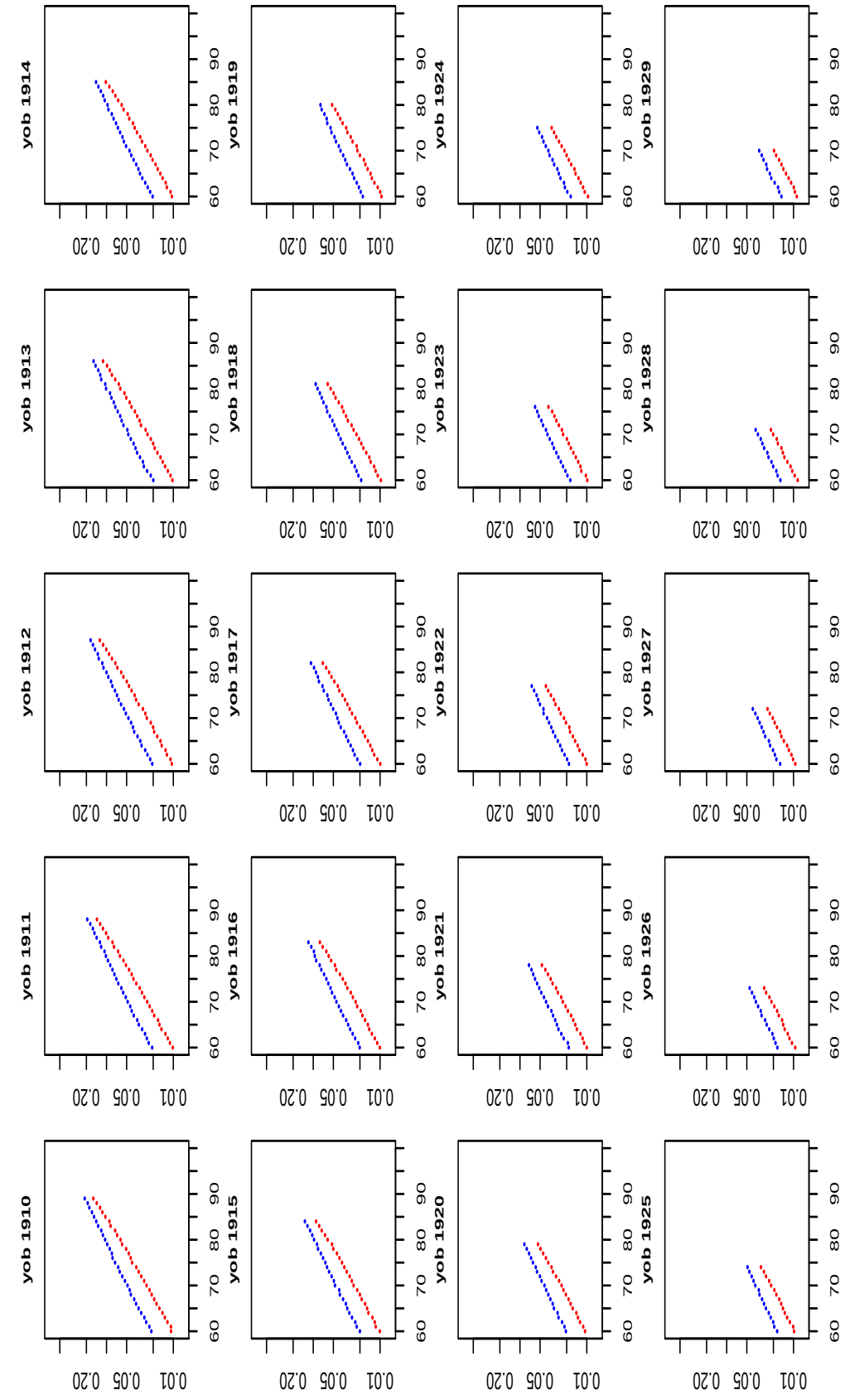
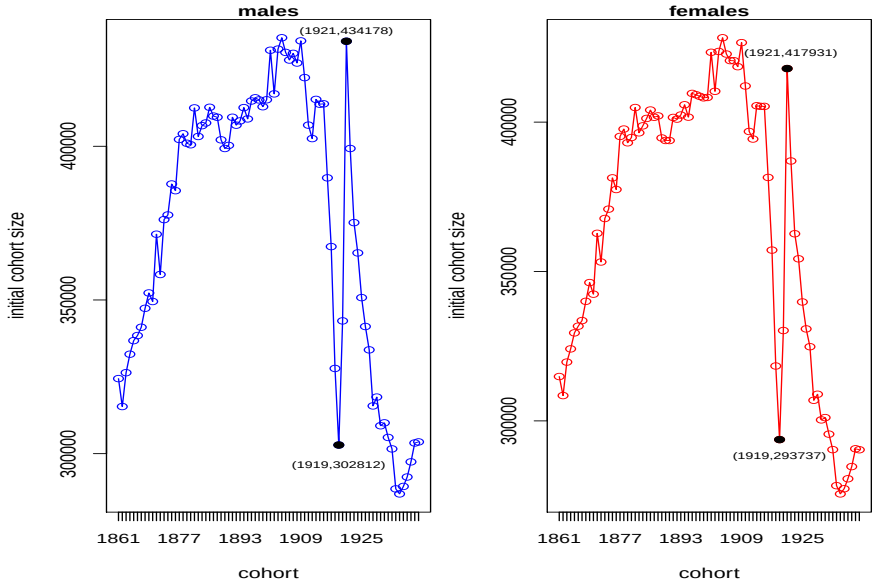


Figure 2.9: Plots of the observed male and female mortality rates (on a log scale) by age and cohort (year of birth), from 1910 to 1929. The blue points denote males and the red females. The x-axis is age and the y-axis corresponds to the mortality rate.

Figure 2.10: Plots of the numbers of males and females at birth in the years 1861 to 1940. Calendar year is along the x-axis and the y-axis corresponds to the number of births.



CHAPTER 3

Review of Mortality Modelling in the Literature

Up until the 1980's mortality modelling was largely focused on producing mathematical functions to fit observed mortality rates. However, over the last 20 to 30 years the development of stochastic mortality models has been very rapid in terms of both structure and statistical techniques used to fit the models. This section reviews many of the key changes that have taken place in the modelling of mortality in the last 100 years.

When discussing the various models the following notation is used: the data comprise n_a age groups and n_y calendar years of data, $d_{x,t}$ is the observed number of individuals aged x at death in calendar year t and $E_{x,t}$ is the number of individuals exposed to risk of death aged x in calendar year t. The raw mortality rate is therefore $\hat{m}_{x,t} = d_{x,t}/E_{x,t}$. Putting the elements into matrix form, D, E, M, are $(n_a \times n_y)$ matrices whose $(x,t)^{th}$ entries are $d_{x,t}$, $E_{x,t}$ and $\hat{m}_{x,t}$, respectively.

The key issue with many of the mortality models is parameter identifiability; that is, two or more different values of parameters correspond to the same statistical distribution of the number of deaths and the data alone cannot determine which set of parameters generated the data. Suppose θ_1 and θ_2 are two values of the parameter θ such that the parameter θ is non-identifiable i.e., $f(y|\theta_1) = f(y|\theta_2)$ for all data y. To identify parameters investigators have introduced various parameter constraints

that are highlighted as each model is discussed.

3.1 The Perks Model

From empirical studies of life tables to describe the 'law of mortality', Perks proposed in 1932 what is now referred to as a logistic model (Perks (1932) Beard (1971)) for μ_x, the force of mortality at age x. The model covers only adult mortality:

$$\mu_x = A + \frac{BC^x}{1 + DC^x} \,. \tag{3.1.1}$$

Makeham's law and Gompertz's law (see Spurgeon (2011)) are special cases of Perks' model corresponding to $D = 0$ and $A = D = 0$, respectively. Beard (Beard (1959)) showed that Perks' model could be derived from a theoretical approach as shown below, with the following assumptions: (a) accidental deaths A (assumed to occur at a constant rate at all ages), (b) an upper limit to the rate of mortality F, and (c) a progression of time.

From (c) Gompertz's law for age x is given by

$$\frac{d\mu_x}{dx} = \lambda\mu_x, \quad \Rightarrow \quad \mu_x = B\exp[\lambda x] \,.$$

From (a) and (c) Makeham's law for age x is given by

$$\frac{d\mu_x}{dx} = \lambda(\mu_x - A), \quad \Rightarrow \quad \mu_x = A + B\exp[\lambda x] \,.$$

From (a), (b) and (c) Perks' law for age x is given by

$$\frac{d\mu_x}{dx} = \lambda(\mu_x - A)\frac{(F - \mu_x)}{(F - A)}, \quad \Rightarrow \quad \mu_x = A + \frac{(F - A)D\exp[\lambda x]}{1 + D\exp[\lambda x]} \,.$$

Perks' (logistic) relation can therefore be expressed as stating that the rate of change of μ_x is proportional to the product of its value in excess of the rate of accidental deaths and the amount by which it falls short of its upper limiting value.

3.2 The Heligman and Pollard Model

Heligman and Pollard (1980) proposed a descriptive 'law of mortality' for q_x, the probability of dying at age x. The full model has three terms and eight parameters, and covers the whole of the age range:

$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + D\exp[-E(\log x - \log F)^2] + GH^x \,. \tag{3.2.1}$$

This can be re-expressed as

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

This shows that the model is continuous and allows q_x to take values between zero and one only. The eight parameters $A - H$ are estimated from values of q_x derived from observed data.

The model contains three terms, each representing a distinct component of mortality. The first, a rapidly declining exponential, reflects the fall in mortality during the early childhood. The parameter A is approximately q_1 and C reflects the improvement in mortality of an infant in the earliest years. B is a shift to the infant's age.

The second term reflects the 'accident hump' and appears either as a distinct hump in the mortality curve or at least as a flattening out of the mortality rates, generally between ages 10 and 40. The accident term has three parameters: F indicating the location, E representing the width and D the severity of the hump.

The third term in the model is the well-known Gompertz exponential, which reflects the geometric rise in mortality at the adult ages, and is generally considered to represent the ageing process. The parameter G represents the base level of adult mortality, while H reflects the rate of increase of that mortality.

Heligman and Pollard estimate the parameters of the curve by least squares, using Gauss-Newton iteration (Heligman and Pollard (1980)). The function minimized was

$$S^2 = \sum_{x=0}^{85} \left[\frac{q_x}{\hat{q}_x} - 1 \right]^2,$$

where q_x is the fitted value at age x and $\hat{q}_x$ is the observed mortality rate. The observed rates above age 85 were excluded from the calculation because they appeared to be less reliable.

For ages above 60, on which this report focuses, the first two terms can be neglected and the equation reduces to

$$q_x = \frac{GH^x}{1 + GH^x}$$

or, re-expressed in a linear form,

$$\text{logit}(q_x) = \log(G) + x \log(H) \, .$$

The Heligman and Pollard model does not naturally extend over time to produce forecasts but was developed for use in the graduation of mortality rates, i.e., to produce mortality rates that progress smoothly from age to age and, at the same time, reflect accurately the underlying mortality pattern.

3.3 The Lee-Carter Model

In 1992, the Lee-Carter model was introduced by R. D. Lee and L. Carter, (Lee and Carter (1992)), to model the central mortality rate based on age x and year t:

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{3.3.1}$$

where α_x represents the general shape of the age-specific mortality profile, κ_t represents the underlying time trend, β_x modulates how the time factor applies to individual ages, enabling some ages to improve whilst others may decline, and $\epsilon_{x,t}$ is an error term, assumed to follow independent $N(0, \sigma^2)$ distributions for each x, t.

Lee and Carter applied singular value decomposition to the log of the empirical

mortality rates, following a two-stage process to fit the model. First the age-related parameters α_x were estimated as the row means of the matrix $\log M$. The β_x and κ_t parameters were estimated as the left and right singular vectors corresponding to the leading singular value of the matrix whose $(x,t)^{th}$ element is $[\log m_{x,t} - \alpha_x]$. Further, to identify β_x and κ_t, two constraints were imposed by ensuring the two sets of parameters sum to one and zero, respectively. The second stage adjusted the κ_t so that the number of deaths in calendar year t from the model gave the actual number of deaths. This was achieved by solving for κ_t the equations

$$\sum_{x=0}^{\omega} d_{x,t} = \sum_{x=0}^{\omega} E_{x,t} \exp(\alpha_x + \beta_x \kappa_t) \quad \forall\, t\,,$$

where ω is the maximum age limit set as 110.

The Box-Jenkins time series methodology (Chatfield (2003)) of identification, estimation and diagnosis was applied to the derived κ_t values, and it was decided that a random walk with drift was satisfactory with one adjustment, namely the inclusion of an indicator variable flu to allow for the impact of the Spanish flu of 1918. Future mortality rates were generated by projecting the κ_t factor using the following process:

$$\kappa_t = \kappa_{t-1} - 0.365 + 5.24 flu + u_t\,.$$

The disturbance u_t is Gaussian with zero mean and constant variance. When considering the mortality of the 85+ age group it was necessary for Lee and Carter to disaggregate this group into 5-year age groups up to 110. Data now exist for these age groups but there remain concerns about measures of mortality at these older ages. The disaggregation was performed using the Coale-Kisker method, which assumes that the exponential rate of increase of mortality at very old ages is not constant, as stipulated by the classical Gompertz model, but declines linearly. Empirical studies on a few populations have demonstrated this feature; see for example Wilmoth (1995), who also gives details of the Coale-Kisker model.

The Lee-Carter model is a bilinear model that gave a good fit to the US data over a wide range of ages. However, the model has a number of weaknesses: there is no

smoothness imposed across ages, whereas one would expect mortality at contiguous ages to be similar. The fitting process was done in two stages. Wilmoth Wilmoth (1993) presented a weighted least squares Lee-Carter solution in a one-step process to determine the parameters. In the Lee-Carter model, error terms were assumed to be normal with constant variance at all ages. This is unrealistic as the logarithm of the observed mortality rate is much more variable at older ages than at younger ages. Part of this variability is due to the smaller number of deaths at older ages. The adoption of a Poisson model allows for heteroscedasticity. Brouhns et al. (2002) and Renshaw and Haberman (2003b) developed a maximum likelihood approach to estimate the parameters based on the assumption that the number of deaths by age and calendar year followed a Poisson distribution with a parameter similar to the structural part of the Lee-Carter model. These Poisson log-bilinear models are described in the next two sections.

3.4 Pedroza - Lee-Carter Model as a Bayesian State Space Model

Pedroza (2006) reformulated the Lee-Carter method as a state-space model:

$$\log \hat{m}_t = \alpha + \beta\kappa_t + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N_{n_a}(0, \sigma_\epsilon^2 \mathbb{I}_{n_a}), \tag{3.4.1}$$

$$\kappa_t = \kappa_{t-1} + \theta + u_t, \quad u_t \overset{i.i.d.}{\sim} N(0, \sigma_\kappa^2), \tag{3.4.2}$$

where $\log \hat{m}_t$ is the random $n_a \times 1$ vector of log-mortality rates in year t for the range of ages, i.e., $\{\log m_{x,t}, x : 1, \ldots, n_a\}$. Similarly, α and β are $n_a \times 1$ vectors, i.e. $\alpha = \{\alpha_1, \ldots, \alpha_{n_a}\}'$ and $\beta = \{\beta_1, \ldots, \beta_{n_a}\}'$. The model assumes that the observations of $\log m_t$ are independent and identically distributed with common variance σ_ϵ^2. The period parameters κ_t are assumed to follow a random walk with drift model and ϵ_t and u_t are assumed to be independent. The multivariate normal model for the log-mortality rates provides a joint distribution for the n_a age groups at any given point in time. For comparison purposes the same constraints as in the Lee-Carter model were applied to the β and κ parameters.

Pedroza applied a Bayesian framework and used MCMC to estimate the parameters. Noninformative priors were used, which, according to Pedroza, lead to results comparable to those derived from Lee-Carter's original method. Uniform prior distributions are assumed for α_x, β_x and θ, i.e., $p(\alpha_x, \beta_x, \theta) \propto 1$. Priors for the variance parameters were also noninformative: $p(\sigma_\kappa^2) \propto 1/\sigma_\kappa$, $p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2$. The prior distribution for the starting point κ_0 is specified as $N(5, 10)$, the authors considered a variance of 10 to be large enough to consider this a 'vague' prior.

The full conditionals of σ_ϵ^2, σ_κ^2 and θ as derived by Pedrosa are

$$\sigma_\kappa^2|\cdots \sim Inv-Gamma\left(\frac{n_y-1}{2}, \frac{1}{2}\sum_{t=1}^{n_y}(\kappa_t-\kappa_{t-1}-\theta)^2\right), \tag{3.4.3}$$

$$\sigma_\epsilon^2|\cdots \sim Inv-Gamma\left(\frac{n_y n_a}{2}, \frac{1}{2}\sum_{x=1}^{n_a}\sum_{t=1}^{n_y}(\log m_{x,t}-\alpha_x-\beta_x\kappa_t)^2\right), \tag{3.4.4}$$

$$\theta|\cdots \sim N\left(\frac{\kappa_{n_y}-\kappa_0}{n_y}, \frac{\sigma_\kappa^2}{n_y}\right). \tag{3.4.5}$$

The joint full conditional distribution of α_x and β_x is obtained by regressions for each age x, α_x and β_x of $\log m_{x,t}$ on κ. Let $X = (1_{n_y\times 1}, \kappa)$. Then,

$$\alpha_x, \beta_x|\cdots \sim N\left((X'X)^{-1}X'\log m_t, \sigma_\epsilon^2(X'X)^{-1}\right). \tag{3.4.6}$$

To make draws for κ, the Kalman filter was used to estimate and forecast κ using the following identity:

$$p(\kappa_1,\ldots,\kappa_{n_y}|m) = p(\kappa_{n_y}|m)p(\kappa_{n_y-1}|m,\kappa_{n_y})\ldots p(\kappa_1|m,\kappa_2,\ldots,\kappa_{n_y}). \tag{3.4.7}$$

Let $M_{n_y}^*$ be the observed log-mortality data up to time n_y. Then run the Kalman filter with updating equations 3.4.8 for $t = 1,\ldots,n_y$ and store values Q_t and R_t up

to time n_y, $R_0 = \tilde{\sigma}_\kappa^2 \mathbb{I}_{n_y}$ where $\tilde{\sigma}_\kappa^2$ is the current draw of σ_κ^2.

$$\begin{aligned} \nu_t &= \log m_t - \alpha - \beta a_t, \\ Q_t &= \beta R_t \beta' + \sigma_\epsilon^2 \mathbb{I}_{n_a}, \\ K_t &= R_t \beta' Q_t^{-1}, \\ a_{t+1} &= a_t + \theta + K_t \nu_t, \\ R_{t+1} &= R_t(1 - K_t) + \sigma_\kappa^2. \end{aligned} \tag{3.4.8}$$

Next, sample the state vector as follows: (a) Sample κ_{n_y} from $(\kappa_{n_y}|M_{n_y}^*) \sim N(a_{n_y}, Q_{n_y})$, then (b) for each $t = n_y - 1, n_y - 2, \ldots, 1, 0$, sample κ_t from $\kappa_t|\kappa_{t+1}, M_{n_y}^* \sim N(h_t, H_t)$ where $h_t = a_t + B_t(\kappa_{t+1} - a_{t+1})$, $H_t = Q_t - B_t R_{t+1} B_t$ and $B_t = Q_t R_{t+1}^{-1}$.

3.5 Poisson Log-bilinear Model

Brouhns et al. (2002) proposed a Poisson log-bilinear model and applied it to Belgian population statistics. The model was set out in terms of the force of mortality, with $\mu_{x,t}$ as the force of mortality at age x during calendar year t. The assumption of the force of mortality being constant for a given age in a particular calendar year results in the force of mortality being equal to the central mortality rate (see section 2). The model is

$$\begin{aligned} d_{x,t} &\sim Poi(E_{x,t}\mu_{x,t}), \\ \log \mu_{x,t} &= \alpha_x + \beta_x \kappa_t \, . \end{aligned} \tag{3.5.1}$$

The parameters α_x, β_x and κ_t have the same meaning as in the Lee-Carter model.

Brouhns et al. use maximum likelihood to estimate the parameters, the log-likelihood being

$$\ell(\alpha, \beta, \kappa) = \sum_{x,t} [d_{x,t}(\alpha_x + \beta_x \kappa_t) - E_{x,t} \exp\{\alpha_x + \beta_x \kappa_t\}] + constant.$$

Brouhns et al. obtained the maximum likelihood estimates using the Newton-Raphson algorithm, updating one set of parameters whilst holding the other two sets at the

current values, e.g. update α_x's while keeping β_x's and κ_t's unchanged. This process is repeated until there is convergence of the estimates of all parameters. The updating steps are easily derived after obtaining the first and second partial derivatives of the log-likelihood function:

$$\alpha_x^{r+1} = \alpha_x^r - \frac{\sum_t (\hat{d}_{x,t} - d_{x,t}^r)}{-\sum_t d_{x,t}^r}, \quad \beta_x^{r+1} = \beta_x^r, \quad \kappa_t^{r+1} = \kappa_t^r,$$

$$\kappa_x^{r+2} = \kappa_x^{r+1} - \frac{\sum_x (\hat{d}_{x,t} - d_{x,t}^{r+1})\beta_x^{r+1}}{-\sum_x d_{x,t}^r (\beta_x^{r+1})^2}, \quad \alpha_x^{r+2} = \alpha_x^{r+1}, \quad \beta_t^{r+2} = \beta_t^{r+1},$$

$$\beta_x^{r+3} = \beta_x^{r+2} - \frac{\sum_t (\hat{d}_{x,t} - d_{x,t}^{r+2})\kappa_t^{r+2}}{-\sum_t d_{x,t}^{r+2} (\kappa_t^{r+2})^2}, \quad \alpha_x^{r+3} = \alpha_x^{r+2}, \quad \kappa_t^{r+3} = \kappa_t^{r+2},$$

where $d_{x,t}^r = E_{x,t} \exp\{\alpha_x^r + \beta_x^r \kappa_t^r\}$ and r is the iteration number. After updating the κ_t parameters, the location constraint was applied, for example by deducting the average of the κ_t^r parameters, $\bar{\kappa}^r$, from each κ_t^r parameter. This requires $\beta^r \bar{\kappa}^r$ to be added to each of the α_x^r parameters. After updating β_x^r, Brouhns et al. applied the scaling constraint, not the sum to one as used in Lee-Carter but a scaling factor that gave $\beta_1^r = 1$. This requires the κ_t^r parameters to be scaled by the reciprocal of this scaling factor.

Forecasting was done by fitting an ARIMA process to the κ_t parameters obtained by maximum likelihood. The Box-Jenkins methodology of identification, estimation and diagnosis was applied to the Belgian population and it was decided that an ARIMA(0,1,1) process was satisfactory for both males and females but with different trends for μ and moving average parameter ϕ, i.e.,

$$\kappa_t - \kappa_{t-1} = \mu + \epsilon_t + \phi \epsilon_{t-1} \,.$$

3.6 Poisson Log-bilinear Model - 2

Czado et al. (2005) present a Bayesian formulation of the Poisson log-bilinear model in terms of $\mu_{x,t}$, the force of mortality at age x in year t, rather than the central mor-

tality rate $m_{x,t}$. The period variable is assumed to follow an autoregressive process of order one with a linear trend:

$$\kappa_t - \phi_1 - \phi_2 t = \rho(\kappa_{t-1} - \phi_1 - \phi_2(t-1)) + \epsilon_t, \;\; \epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\kappa^2). \tag{3.6.1}$$

The parameters α, β, κ, $\phi = (\phi_1, \phi_2)'$, ρ and σ_κ^2 are estimated using MCMC methods, specifically the Gibbs sampler and Metropolis-Hastings (MH) algorithm, to generate samples from the posterior.

The prior on the linear trend parameters is given by $\phi \sim N_2(\phi_0, \Sigma_{\phi_0})$, where ϕ_0 and Σ_{ϕ_0} are constants, $\phi_0 = 0$ and $\Sigma_{\phi_0} = 10\mathbb{I}_2$. The prior on the autoregressive parameter ρ is described by $N(\rho_0, \sigma_{\rho_0}^2)$ truncated to the interval (0,1) and the prior on σ_κ^{-2} is described by $Gamma(c, f)$ where $\rho_0 = 0$, $\sigma_{\rho_0}^2 = 1$, $c = 1$ and $f = 0.001$. The prior distribution for β is described by $N_{n_a}(0, \sigma_\beta^2 \mathbb{I}_{n_a})$ and σ_β^{-2} is described by $Gamma(g, h)$ where $g = 1$ and $h = 0.001$.

For the vector of parameters, α, Czado et al. transform this vector to the vector containing the exponentials of α i.e., $\widetilde{\alpha} = \exp(\alpha)$ as the full conditional of $\widetilde{\alpha}$ can be expressed as a standard distribution. The prior on each element of $\widetilde{\alpha} \sim Gamma(r, s)$ independently where r and s are constants. In order to obtain an uninformative prior distribution (i.e. with large variance), Czado et al. choose a small $s = 0.001$. Afterwards, r is chosen equal to $s\exp(\alpha_x)$.

The full conditionals for $\widetilde{\alpha_x}$, ϕ, ρ and σ_κ^2 are standard distributions.

For the parameters β and κ the full conditionals are non-standard distributions and draws were generated using the MH algorithm.

3.7 The Renshaw and Haberman Generalised Linear Model

Like Brouhns et al. (2002), Renshaw and Haberman (2003a) modelled the number of deaths as a Poisson variable but with a different parameter structure:

$$
\begin{aligned}
& d_{x,t} \sim Poi(E_{x,t}\mu_{x,t}) \\
& \log \mu_{x,t} = \alpha_{xt_0} + \beta_x(t - t_0) + \beta'_x(t - t_j)_- + \gamma_t(t - t_0) \\
& \text{for } t \in [t_1, t_n],\ t_1 < t_j < t_0 \leq t_n,\ (t)_- = t,\ t \leq 0;\ (t)_- = 0,\ t > 0\,.
\end{aligned}
\tag{3.7.1}
$$

The value t_0 is a time origin and, while it can be set at any point on the time axis, the authors choose to select it to coincide with a standard mortality table so that the mortality factor α_{xt_0} can be compared with the standard (or published) mortality table. Similarly to the Lee-Carter model, α_{xt_0} is set as the average of the logarithm of the observed mortality, not over the full number of years, but over the three years of data used in the construction of the standard table:

$$\hat{\alpha}_{xt_0} = \log \prod_{t=t_0-1}^{t_0+1} \hat{m}_{x,t}^{\frac{1}{3}}\,.$$

The model includes a break point t_j, to allow for a change in the age differential time trend, and the break point is fixed as part of the initial fitting process.

The model is fitted using a log-link function with the linear predictor $\eta_{x,t}$, where $\eta_{x,t} = \log E_{x,t} + \log \mu_{x,t}$, and so includes an offset of $\log E_{x,t} + \hat{\alpha}_{xt_0}$.

A key difference between the method of Renshaw and Haberman and that of Lee-Carter and Brouhns et al. concerns the treatment of time. Lee-Carter and Brouhns et al. model time as a factor κ that is estimated, whereas Renshaw and Haberman treat time as a known covariate.

Forecasts of future mortality are generated by extrapolating the time t in the $\beta_x(t - t_0)$ factor, so that the mortality rate for age x in year $t_n + s$ is

$$\hat{m}_{x,t_n+s} = \hat{m}_{x,t_n} \exp(\hat{\beta}_x s)\,.$$

3.8 Bayesian State Space Model

Reichmuth and Sarferaz (2008) proposed a generalisation of the Lee-Carter model. Firstly, they ensured smoothness along the age and period variables by assuming

that both follow vector autoregressive processes. Secondly, they introduced additional period and age-related parameters. Thirdly, they incorporated additional observable covariates and modelled their dynamics jointly with the other parameters. The general model is

$$\log m_{x,t} = \overline{\alpha_x} + r_x^{'} z_t + \epsilon_{x,t}^m \,, \tag{3.8.1}$$

where $\overline{\alpha}_x$ is the arithmetic mean of the $\log \hat{m}_{x,t}$ for age x, i.e., $\overline{\alpha_x} = \frac{1}{n_y}\sum_{t=1}^{n_y} \log \hat{m}_{x,t}$, and the explanatory variables are $z_t = [\kappa_t \; v_t]^{'}$, where $\kappa_t = [\kappa_{1t}, \ldots, \kappa_{Nt}]^{'}$ a $N \times 1$ vector and v_t is a $L \times 1$ vector of observed covariates $[v_{1t}, \ldots, v_{Lt}]^{'}$. Therefore z_t is a $R \times 1$ vector with $R = N + L$. The age-related parameters have a similar structure, with $r_x = [r_x^{\kappa} \; r_x^{v}]^{'}$, where $r_x^{\kappa} = [r_{1x}^{\kappa}, \ldots, r_{Nx}^{\kappa}]^{'}$ is a $N \times 1$ vector and r_x^v is a $L \times 1$ vector $[r_{1x}^v, \ldots, r_{Lx}^v]^{'}$, so that r_x is $R \times 1$.

The z_t and r_x follow VAR(p) and VAR(q) processes respectively:

$$z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + \epsilon_t^z \tag{3.8.2}$$

and

$$r_x = \xi_1 r_{x-1} + \xi_2 r_{x-2} + \cdots + \xi_q r_{x-q} + \epsilon_x^r \,, \tag{3.8.3}$$

where c is a $R \times 1$ vector of constants, $\phi_1, \ldots, \phi_p$ are $R \times R$ matrices, $\xi_1, \ldots, \xi_q$ are $R \times R$ diagonal matrices, the $\epsilon_{x,t}^m$ are i.i.d. $N(0, \sigma_m^2)$; the ϵ_t^z are i.i.d. $N_R(0, \Sigma_z)$, and ϵ_x^r are i.i.d. $N_R(0, \Sigma_r)$. Each component of r_x is a priori independent of the other components and so Σ_r is diagonal. The disturbances $\epsilon_{x,t}^m$, ϵ_t^z and ϵ_x^r are independent.

The parameters of the model are therefore $r = (r_1, \ldots, r_{n_a})$, $Z = (z_1, \ldots, z_{n_y})$, $\Phi = (\phi_1, \ldots, \phi_p, c)$, $\Xi = (\xi_1, \ldots, \xi_q)$, Σ_r, Σ_z and σ_m^2.

The following priors were used:

- For the coefficients $\xi_1, \ldots, \xi_q$, flat priors were assumed.
- For $\sigma_m^2 \sim IG\left(\frac{0.01}{2}, \frac{3}{2}\right)$.
- The prior on Φ derived from normal linear regression model $Z^* = X^*\Phi + \epsilon^z$ as

$$\Phi|\Sigma_z \sim N\left(vec(\hat{\Phi}^*), \Sigma_z \otimes (X^{*'}X^*)^{-1}\right)$$

where $\hat{\Phi}^* = (X^{*'}X^*)^{-1}X^{*'}Z^*$ the ordinary least squares estimate;
$Z^* = Diag_{Rp}\left[5A, Diag_{R(p-1)}[0,\ldots,0]\right]$ $\epsilon^z = Diag_{Rp}[\epsilon_1,\ldots,\epsilon_R]$ and
where $\epsilon_t \sim N_R(0,\Sigma_z)$,
Σ_z was given an uninformative prior with the density equal to a constant.
$X^* = Diag_{Rp}[5A, 10A,\ldots,5pA]$ where $A = Diag_p[\hat{\sigma}_1,\ldots,\hat{\sigma}_p]$,
$\hat{\sigma}_1,\hat{\sigma}_2,\ldots,\hat{\sigma}_R$ are the empirical standard deviations obtained from the first p observations.

- The priors on r and z are induced by the other parameters.

The authors used the Gibbs sampler to estimate the model by generating parameter draws from the posterior $\pi(r, Z, \Phi, \Xi, \Sigma_r, \Sigma_z, \sigma_m^2 | \log \hat{m}_{x,t})$ by iteratively drawing from the full conditionals.

Reichmuth and Sarferaz (2008) applied their model to United States data with two period-related parameters (i.e. $N = 2$) with two additional covariates GDP per capita and unemployment rate. The order of the VAR processes were selected with $p = q = 4$.

3.9 Cairns-Blake-Dowd Models

There have been two variants of the Cairns-Blake-Dowd Model, of which the first (CBD-1), introduced in 2006 (Cairns et al. (2006)) and involving two stochastic factors, is a special case of the Perks model described in section 3.1 above. The first factor $\kappa_t^{(1)}$ affects mortality at all ages in an equal manner, whereas the second $\kappa_t^{(2)}$ has an effect on mortality that is proportional to age. If $q_{x,t}$ is the underlying mortality rate in year t at age x, then

$$q_{x,t} = \frac{\exp\{\kappa_t^{(1)} + x\kappa_t^{(2)}\}}{1 + \exp\{\kappa_t^{(1)} + x\kappa_t^{(2)}\}}. \tag{3.9.1}$$

For each t, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ were estimated using least squares by transforming the

observed mortality rates from $q_{x,t}$ to

$$\text{logit } q_{x,t} = \kappa_t^{(1)} + x\kappa_t^{(2)}.$$

In a later paper by the same authors (Cairns et al. (2009)), the age x factor was centred i.e. transformed to $x - \overline{x}$.

For forecasting, Cairns et al. (2006) fitted a 2-dimensional random walk with drift, based on parameters that had to be estimated, i.e.

$$\kappa_{t+1} = \kappa_t + \mu + CZ_t\,, \tag{3.9.2}$$

where $\kappa_t = (\kappa_t^{(1)}, \kappa_t^{(2)})'$, μ is a constant 2×1 vector, C is a constant 2×2 upper triangular matrix and Z_t is a 2-dimensional standard normal random vector.

Distributions of the μ and C parameters were obtained using Bayesian analysis. The process for κ_t is a random walk with mean μ and covariance matrix $V = CC'$.

Non-informative priors were used:

$$p(\mu) \propto constant$$

$$p(V) \propto det(V)^{-\frac{3}{2}}.$$

With this prior and with n_y observations $F = (F(1), \ldots, F(n_y))$, where $F(t) = \kappa_t - \kappa_{t-1}$, the authors derived the posterior distributions as

$$V^{-1}|F \sim Wishart\left(n_y - 1, \frac{1}{n_y}\widehat{V}^{-1}\right)$$

$$\mu|V, F \sim N_{n_y}\left(\widehat{\mu}, \frac{1}{n_y}\widehat{V}^{-1}\right),$$

where

$$\widehat{\mu} = \frac{1}{n_y}\sum_{t=1}^{n_y} F(t)$$

and

$$\widehat{V} = \frac{1}{n_y} \sum_{t=1}^{n_y} \left[F(t) - \widehat{\mu}\right] \left[F(t) - \widehat{\mu}\right]' .$$

This model allows for mortality improvements at different ages to vary by time because of the interaction term $x\kappa_t^{(2)}$. Also a very simple age effect enabled the authors easily to incorporate parameter uncertainty, with all parameters being identifiable. The model's simplistic age effect will result in any non-linearity in the mortality curve due to age being ignored. As with the Lee-Carter model, the (CBD-1) model ignores cohort effects.

The second Cairns-Blake-Dowd Model, (CBD-2), was introduced in 2007 (Cairns et al. (2009)). This model addressed some of the weaknesses of (CBD-1), in particular introducing a cohort parameter and a more flexible age structure that would capture some non-linearity in the mortality curve by age:

$$\text{logit } q_{x,t} = \kappa_t^{(1)} + (x - \overline{x})\kappa_t^{(2)} + ((x - \overline{x})^2 - s_x^2)\kappa_t^{(3)} + \gamma_{t-x} \tag{3.9.3}$$

where γ_{t-x} is the cohort parameter, $s_x^2 = \frac{1}{n_a}\sum_x (x - \overline{x})^2$ and n_a is the number of ages. Unlike with (CBD-1) there is an identifiability problem with the parameters. This was resolved by imposing constraints,

$$\sum_{k=1}^{n_c} \gamma_k = 0$$

$$\sum_{k=1}^{n_c} k\gamma_k = 0,$$

$$\sum_{k=1}^{n_c} k^2\gamma_k = 0,$$

where n_c is the number of cohorts and γ_k is the cohort parameter for cohort k. The authors choose these constraints because, if least squares is used to fit a quadratic function $\phi_1 + \phi_2 k + \phi_3 k^2$ to γ_k, the constraints ensure that $\phi_1 = \phi_2 = \phi_3 = 0$ and that γ_k fluctuates around zero so that improvements with period are captured within the period parameters.

Ideally the investigator should have a valid reason for this restriction, rather than

it being an arbitrary choice to reduce the dimension of a system of equations. The valid restriction must be derived from outside the data, which may not be possible.

3.10 Age-Period-Cohort Models

The concept of cohort has been central in demography, epidemiology and other social sciences. In this context cohort refers to a group of people who share a common year of birth. Age is often the main factor in such models as it accounts for consistent extrinsic factors. The effect of the period accounts for all factors that affect every person at a particular time in history, such as pollution or medical advances. The cohort effect accounts for events which affect generations, e.g. malnutrition of children during or after wars. The use of Age-Period-Cohort (APC) models predates the Lee-Carter model. For example, Holford (1983) and Clayton and Schifflers (1987a,b) used the APC model to analyse data on rates of mortality from cancer.

In the context of a Poisson model for the death count, the APC model has the general form

$$\log m_{x,t} = \mu + \alpha_x + \kappa_t + \gamma_k \,, \tag{3.10.1}$$

where μ is the grand mean, α_x is the age-effect parameter for age x, κ_t is the period-effect parameter for year t, and γ_k is the parameter for the k^{th} cohort where $k = n_a - x + t$ and as a consequence the total number of cohorts is $K = n_a + n_y - 1$.

Parameter identifiability is a key issue with APC models, as is obvious from the following example. If in equation (3.10.1) one adds $(t - \bar{t}) - (x - \bar{x})$ to γ_{t-x} and then adds $(x - \bar{x})$ to the α_x component and deducts $(t - \bar{t})$ from the κ_t component the parameter values are different but result is the same $\log m_{x,t}$ and therefore the same likelihood.

The identifiability problem with APC models arises from the linear dependence of the three temporal factors of interest. By subtraction, one can readily calculate year of birth from age and date of death. Hence, in a real sense there are just two time dimensions that can be used to fully describe the time trends for mortality rates.

To deal with the simultaneous estimation of age, period and cohort effects Fienberg and Mason (1979) and Barrett (1973, 1978) introduced an arbitrary restriction on the parameters, in that two of the effects are equated. As mentioned earlier there should be a valid reason for any restriction. In various papers, Holford (1983, 1991) has explored the problem of parameter identification within APC models. In Holford (1983) he shows that, whilst the individual parameters are not identifiable, certain functions of the age, period and cohort parameters are identifiable. Holford (2006) shows that the non-identifiability only affects linear trends; change points and other non-linear trends can be identified (up to the linear trend). This was demonstrated by firstly removing the linear trend from each parameter, take for example the vector of age-related parameters α to illustrate the process. If we denote α_L as the linear part, that is a linear combination of the components of α, i.e.,

$$\alpha_L = C\sum_{i=i}^{n_a} c_i\alpha_i, \text{ where } c_i = i - \frac{1}{2}n_a - \frac{1}{2} \text{ and } C = (\sum_i c_i^2)^{-1}.$$

The i^{th} component of the non-linear part, $\widetilde{\alpha}$, is:

$$\widetilde{\alpha}_i = \alpha_i - c_i\alpha_L.$$

The non-linear part $\widetilde{\alpha}$ can then be expressed as a sum of orthogonal polynomials of second degree or higher. This process is repeated for the period and cohort parameters. Then a design matrix can be created for the linear parameters i.e. the c_i's and the coefficients of the orthogonal polynomials. If two of the columns of linear components are removed from this design matrix, then the rank of the resultant matrix is full, indicating parameter identifiability.

An important point made by Holford (1991) is that although the individual parameters and their forecasts are non-identifiable, the mortality rate and projected mortality rates are estimable so the non-identifiability problem is irrelevant.

Section 4 of Besag et al. (1995) applied a Bayesian framework to the APC model to analyse the data used by Holford (1983) on mortality from prostate cancer among non-whites in the United States. Besag et al. used logistic regression assuming that the data are binomial rather than Poisson, as done by Holford. A key point made is that the Bayesian formulation avoids the identifiability difficulty through

the adoption of mildly informative priors Besag et al. (1995). Section 4 of the paper included a number of very useful features that have been employed in this thesis.

Besag et al. (1995) showed that additional unobserved covariates $z_{x,t}$ can be included as a random effect, a sample from a Gaussian with zero mean, so that the $m_{x,t}$'s are related via a logistic-normal model:

$$\text{logit } m_{x,t} = \mu + \alpha_x + \kappa_t + \gamma_k + z_{x,t}.$$

The meaning of the parameters μ, α_x, κ_t and γ_k are the same as in the APC model described above. The paper highlights a very useful transformation whereby logit $m_{x,t}$ is treated as a parameter and the $z_{x,t}$'s are expressed in terms of the $m_{x,t}$'s and the μ, α, κ and γ parameters. This enables full conditionals on μ, α, κ and γ to be derived as Gaussians.

Smoothness across age, period and cohort parameters is achieved by Besag et al. using improper priors that correspond to random walks on the first differences. For example, for the period parameters,

$$\pi(\kappa|\lambda) \;\propto\; \lambda^{n_y/2} \exp\left(-\frac{\lambda}{2} \sum_{j=2}^{n_y-1} (\kappa_{j-1} - 2\kappa_j + \kappa_{j+1})^2 \right)$$

and the hyper-parameter λ has a $Gamma(c, d)$ prior, where c and d are constants chosen to be weakly informative.

The authors noted that this form of prior on the period and cohort parameters preserves the trend and is therefore useful for forecasts. Using MCMC, samples from the posterior for each of the parameters were obtained. Using these samples the model can be extended easily to predictions by exploiting the smoothing; i.e. the future period parameters can be generated by a recursive formula so that, for year $n_y + t$,

$$\kappa^{(r)}_{n_y+t} \text{ is a draw from } \; N\left(2\kappa^{(r)}_{n_y+t-1} - \kappa^{(r)}_{n_y+t-2}, 1/\lambda^{(r)}\right),$$

where the superscript $'(r)'$ indicates that the posterior samples are obtained from the r^{th} cycle of the MCMC process. The cohort parameter can be handled in a similar

way.

A number of other authors followed the Besag et al. (1995) model by applying the APC model with random effects to different data; see for example, Knorr-Held and Rainer (2001), Schmid and Held (2007) and Riebler and Held (2010). The last authors Riebler and Held (2010) extended the model to the case when the data are stratified, for example, by different geographical areas, different causes of mortality, and by gender.

3.11 APC Models with Bilinear Age Effects

A generalisation of the age-period-cohort model allows the period and cohort variables to include interactions with age. This model was included in the paper Cairns et al. (2009). From Figure 2.3 it is apparent that the period improvements in mortality vary by age.

The model as presented in the paper by Cairns et al. (2009) is

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x} \,. \tag{3.11.1}$$

To identify the parameters the following constraints were imposed:

$$\sum_{x=1}^{n_a} \beta_x = 1 \,, \; \sum_{x=1}^{n_y} \kappa_t = 0 \,, \; \sum_{t=1}^{n_y} \sum_{x=1}^{n_a} \gamma_{t-x} = 0 \,, \; \sum_{x=1}^{n_a} \delta_x = 1.$$

In this thesis this model is developed further using Bayesian analysis, with a few enhancements. The model as specified lacks smoothness across ages and years and this is introduced by smoothing via the priors. The parameter of interest is the mortality rate and not the individual age, period and cohort parameters. The constraints are therefore dropped. Finally, the inclusion of a random effect to allow for additional variability on the observed mortality rates is included in the model here.

3.12 P-Spline Mortality Model

Currie et al. (2004) used a penalized generalized linear model to smooth and forecast two-dimensional mortality rates. The method used local cubic polynomials, B-splines, as the basis for regression. The B-splines were joined at regularly spaced knots along the age and year dimensions. The parameters were selected so as to maximize the penalised likelihood. The model, for the number of deaths $d_{x,t}$ for age x in year t, is

$$d_{x,t} \sim Poi(E_{x,t}m_{x,t}) .$$

Putting the elements into matrix form, D and E are matrices whose $(x,t)^{th}$ entries are $d_{x,t}$ and $E_{x,t}$, respectively. In the generalized linear model framework with a log link function,

$$\mathbb{E}(d_{x,t}) = \mu_{x,t}$$

and

$$\begin{aligned} vec(\log \mu) &= vec(\log E) + vec(B_a \Theta B_y^{'}) \\ &= vec(\log E) + (B_y \otimes B_a) vec(\Theta), \\ &= vec(\log E) + B\theta. \end{aligned}$$

In the formulae above, Θ is a $c_a \times c_y$ matrix of the regression coefficients for age and year, $\theta = vec(\Theta)$ and c_y and c_a depend on the number of knots and the degree of the B-spline. The matrices B_y and B_a are matrices of B-splines for smoothing by year and age respectively:

$$B_y = \begin{pmatrix} b_1(t_1) & \dots & b_{c_y}(t_1) \\ \vdots & & \vdots \\ b_1(t_{n_y}) & \dots & b_{c_y}(t_{n_y}) \end{pmatrix} \quad B_a = \begin{pmatrix} b_1(x_1) & \dots & b_{c_a}(x_1) \\ \vdots & & \vdots \\ b_1(x_{n_a}) & \dots & b_{c_a}(x_{n_a}) \end{pmatrix} .$$

Further, $B = B_y \otimes B_a$ is an $n_a n_y \times c_a c_y$ matrix of B-splines in both age and year.

The linear predictor corresponding to the j^{th} column of D is therefore

$$\sum_{k=1}^{c_y} b_k(t_j) B_a \Theta_{k\cdot},$$

where $\Theta_{k\cdot}$ denotes the k^{th} row. Also $\Theta_{\cdot k}$ denotes the k^{th} column of Θ. In this structure the maximum likelihood estimate of θ can be obtained by applying the standard GLM scoring algorithm to a Poisson model with log link function.

To introduce smoothing within Θ, a penalty factor is added for both the age (row elements of Θ) and the year (column elements of Θ). There are many different smoothing penalties the choice of which is critical to forecasting, as the authors demonstrate. They selected a quadratic difference penalty for both the age and year dimensions. The penalty along the year dimension for age i is

$$(\Theta_{i1} - 2\Theta_{i2} + \Theta_{i3})^2 + \cdots + (\Theta_{ic_y-2} - 2\Theta_{ic_y-1} + \Theta_{ic_y})^2,$$

which can be written as

$$\sum_{k=1}^{c_y} \Theta_{\cdot k}^{'} G_a^{'} G_a \Theta_{\cdot k} = \theta^{'} (I_{c_y} \otimes G_a^{'} G_a) \theta,$$

where I_{c_y} is an identity matrix of size c_y and G_a is a difference matrix with dimension $(c_a - 2) \times c_a$. Similarly, the row penalty is

$$\theta^{'} (G_y^{'} G_y \otimes I_{c_a}) \theta \ .$$

The penalty on ages and years can therefore be expressed as

$$P = \lambda_a (I_{c_y} \otimes G_a^{'} G_a) + \lambda_y (G_y^{'} G_y \otimes I_{c_a}).$$

The parameters λ_a and λ_y are penalty constants applying, respectively, to the age and year smoothing factors. The sizes of these parameters influence the degree of smoothing: if $\lambda = 0$ then there is no imposed smoothing on the parameters and as $\lambda \to \infty$ the penalty factor dominates and in the limit the parameters take the form

corresponding to minimizing the smoothing penalty. The smoothing parameters can therefore be viewed as balancing fidelity to the data and the level of smoothing.

The penalized log-likelihood for the model is therefore

$$\ell_p(\theta, D) = \ell(\theta, D) - \frac{1}{2}\theta' P\theta \,, \tag{3.12.1}$$

where $\ell(\theta, D)$ is the usual Poisson log-likelihood. The regression coefficients (θ) can be derived by maximizing 3.12.1, which can be achieved with the following scoring algorithm

$$(B' W_c B + P)\theta_{c+1} = (B' W_c B)\theta_c + B'(vec(D) - vec(\mu)) \,, \tag{3.12.2}$$

where θ_c is the value of the parameters during the current iteration and θ_{c+1} is the next iterated value, $vec(\log \mu) = vec(\log E) + B\theta_c$ and $W_c = diag(vec(\mu))$.

The optimal values for the penalty factors λ_a and λ_y were selected by minimizing the Bayesian information criterion (BIC), which has the form

$$deviance(D, \theta, \lambda_a, \lambda_y) - \log(N) tr(H) \,.$$

Here $H = B(B'\hat{W}B + P)^{-1}B'\hat{W}$, which is referred to as the hat matrix, the trace of which gives the effective dimension of the model, where $\hat{W}$ contains the weights of the last iteration after convergence. N is the number of data points and in this case is $n_a n_y$. BIC was chosen rather than the Akaike Information Criterion (AIC) as it introduces much heavier penalties and therefore produces smoother curves. Currie et al. (2004) remark that AIC is likely to result in undersmoothing for over-dispersed data, giving an unsatisfactory result.

3.13 Mortality Forecasts by ONS and CMI

The Office for National Statistics (ONS) and the Continuous Mortality Investigation (CMI) follow a similar methodology for mortality forecasting. The methodologies of the ONS and CMI are explained in Office for National Statistics (2011b) and Contin-

uous Mortality Investigation (2009), respectively. Initial mortality rates are set using current levels of population mortality which have been smoothed using p-spline models similar to that described in section 3.12. Future mortality rates are then estimated from these initial mortality rates by applying deterministic improvement factors that depend upon age, sex and the length of the projection term. Specifically, the CMI model allows for initial rates of mortality improvement, reflecting the current estimate of rates of change in the recent past which converges to assumed long-term rates of mortality improvement. The speed and pattern of convergence from 'initial' to 'long-term' improvement rates are variable and can be set by the users of the model. The ONS mortality improvement figures also vary by age and sex being derived from recent mortality trends of England and Wales population. These mortality improvement figures are fixed and are blended into the mortality improvements for 25 years ahead. Mortality improvements beyond 25 years ahead of the base year are constant at a target rate. The target rate used in the life expectancy figures in the National Population Projections 2010-based report was 1.2% for males and females.

Effectively the approach followed by the CMI and ONS assumes that, in the very short-term, the best guide as to the likely pace of change in mortality rates is the most recently observed experience. In the long-term, the forces driving mortality change are likely to be very different from those currently influencing patterns of improvement. Oeppen and Vaupel (2002) points out that mortality experts have repeatedly asserted that life expectancy is close to an ultimate ceiling and that these experts have repeatedly been proven wrong.

CHAPTER 4

Mortality Models

The previous chapter shows that there has been a large number of different models and modelling techniques used to model mortality. This chapter considers various mortality models, including three new models that are developments and generalisations of the Lee-Carter model.

In all models it is assumed that the number of deaths for age x in calendar year t is a Poisson random variable with a different structure of the mean parameter, i.e.,

$$d_{x,t} \sim Po(E_{x,t} m_{x,t}) \qquad x = x_1, \ldots, x_{n_a}; \;\; t = t_1, \ldots, t_{n_y},$$

where $E_{x,t}$ is the number exposed to risk of death aged x in calendar year t and $m_{x,t}$ follows the structure below:

Model 1 $\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t$

Model 2 $\log m_{x,t} = \mu + \alpha_x + \kappa_t + \gamma_{t-x}$

Model 3 $\log m_{x,t} = \mu + \alpha_x + \kappa_t + \gamma_{t-x} + z_{x,t}$

Model 4 $\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x}$

Model 5 $\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x} + z_{x,t}$

Model 6 $\log \frac{m_{x,t}}{1-m_{x,t}} = \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x} + z_{x,t}$

Model 7 $\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x}$

In Models 1 to 6 $E_{x,t} = E^o_{x,t}$, where $E^o_{x,t}$ is the observed exposure as estimated by HMD, see section 2.2, and is treated as a known constant. In Model 7, each $E_{x,t}$ is treated as an estimated value with the following structure: $\frac{E^o_{x,t}}{1-w_{x,t}}$ where the $w_{x,t}$ parameters represent measurement errors in the exposures and are discussed in section 4.1.1.

The $z_{x,t}$ parameters that appear in Models 3, 5 and 6 are considered as additional unobserved covariates that provide added heterogeneity or extra random variation of mortality.

In Models 1 to 5 and Model 7 the link function is log which can result in $m_{x,t}$ values larger than one, which is theoretically possible but very unlikely in large populations like the data set being used. The HMD data for England and Wales extends from 1841 up to 2009 for ages 0 to 110 and no mortality rate exceeded 1. In section 2.2 it was shown that neighbouring cohorts are of similar sizes and so a mortality rate of 1 means that the expected number of deaths is equal to the average population size, which is unrealistic. Model 6 uses a logit link function, which restricts $m_{x,t}$ to being in the range $(0, 1)$. This restriction to being no greater than 1 is not a material limitation for fitting but has the major benefit that long-range forecasts of $m_{x,t}$ will be better behaved.

In each of the models, μ is a parameter, representing the overall mean logarithm (or mean logit) of mortality; α_x is the age-related parameter for age x; κ_t is the period parameter for year t; γ_{t-x} is the cohort parameter for year of birth $t - x$, there being $n_c = n_a + n_y - 1$ distinct cohorts; β_x and δ_x are age-related parameters used to scale the period and cohort effects.

As noted earlier, in each model some of the parameters are non-identifiable. For example, in Model 1, β_x and κ_t are identifiable up to scale. In the other models the α_x, κ_t and γ_{t-x} are non-identifiable, as are β_x and δ_x in Models 4 to 7. A number of modellers (e.g., Brouhns et al. (2002) and Cairns et al. (2009)) have sought to analyse trends in these underlying parameters by imposing constraints, such as sum-to-zero constraints. The main focus of this thesis is on the mortality rate ($m_{x,t}$), which is identifiable for Models 1 to 6. However, for Model 7, $m_{x,t}$ is identifiable up to scale and age and period because of the presence of measurement errors in the

exposures. This can be shown by noting that if $\exp(\phi_{x,t}) = \frac{1}{1-w_{x,t}}$ then the Poisson mean of Model 7 can be written as $E^o_{x,t} \exp(\phi_{x,t} + \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x})$. Then, for a constant shift to μ say $\mu + c$, a corresponding adjustment to $\phi_{x,t}$ can be made that gives the same overall mean parameter. Also, for an adjustment to the age parameters $\alpha_x + a_x$ then an adjusted $\phi'_{x,t} = \phi_{x,t} - a_x$ preserves the mean. Similarly for period parameter $\kappa_t + b_t$ an adjusted $\phi'_{x,t} = \phi_{x,t} - \beta_x b_t$ preserves the mean. The informative prior on the error parameters $w_{x,t}$'s serves to significantly reduce this issue.

From a biological standpoint one would expect mortality to change gradually and relatively smoothly with age, period and cohort and therefore it is desirable that the respective parameters exhibit smoothness. Importantly, smoothing the period and cohort parameters facilitates forecasting of mortality rates. The form of the smoothing has a significant impact on the forecasts so a number of different smoothing assumptions are applied via the prior distribution. The resulting model is given one of the following additional labels:

(0) No smoothing of the parameters (for Models 1, 2 and 4 only)

(a) Random walk on the levels of the parameters

(b) Autoregressive process of order 1 on the first differences of the parameters

(c) Random walk on the first differences of the parameters.

For example, Model 3(a) denotes Model 3 with the prior specifying that each age, period and cohort parameter follows an independent random walk with its own precision. Model 3(b) denotes Model 3 with each age, period and cohort parameter following an independent autoregressive process of order 1 on the first differences of the parameters, with each parameter having its own precision and autoregressive coefficient.

Models 1 to 4 have been used in the literature. Model 1 where the mortality parameter has a similar form to the Lee-Carter model has been used in many papers; for example see Czado et al. (2005), Pedroza (2006) and Cairns et al. (2009). Model 2, the age-period-cohort model, also occurs frequently in the literature for many different studies; for example see Holford (1983, 1991, 2006) and Clayton and Schifflers

(1987a,b). Model 3 has also appeared in the literature; for example Berzuini and Clayton (1994), Besag et al. (1995), Knorr-Held and Rainer (2001). Model 4 without smoothing of parameters appeared in Cairns et al. (2009). To my knowledge Models 5, 6 and 7 have not appeared in the literature.

The notation follows that of Cairns et al. (2009) which along with Besag et al. (1995) were major influences on this thesis.

4.1 Framework

To implement the Bayesian approach, Markov-Chain-Monte-Carlo (MCMC) methods were employed to sample from the posterior distribution of the parameters in each model. More specifically, draws from the joint posterior distribution were obtained from the full conditional distributions when the full conditionals were easily derived; otherwise the Metropolis-Hastings algorithm (MH) was used. Details of the implementation of each model are described in the following subsections, and exploiting the formulae in Appendix B.

4.1.1 Parameters

If we define $\alpha = (\alpha_1, \ldots, \alpha_{n_a})', \kappa = (\kappa_1, \ldots, \kappa_{n_y})', \gamma = (\gamma_1, \ldots, \gamma_{n_c})', \beta = (\beta_1, \ldots, \beta_{n_a})'$ and $\delta = (\delta_1, \ldots, \delta_{n_a})'$ then discussion of the priors will be in terms of the vectors α, κ, γ, β and δ. Here n_a, n_y and n_c denotes the number of age groups, year groups and cohorts respectively. So n_a is 41 as the ages range from 60 to 100, n_y is 40 as the calendar years covered are from 1960 to 1999 and $n_c = n_a + n_y - 1$ cohorts.

Model 1(0)

The prior distribution used for μ was an improper uninformative prior, uniform on the real line. The priors on α, β and κ are independent multivariate Gaussian distributions centred on the zero vector and with covariance matrices $1000\mathbb{I}_{n_a}$, $1000\mathbb{I}_{n_a}$ and $1000\mathbb{I}_{n_y}$, respectively. Here $\mathbb{I}_n$ denotes the $n \times n$ identity matrix. Therefore, each

parameter is a priori independent of the others and each element of a parameter is a priori independent of the others.

Models 1(a), (b) and (c)

The prior on μ is the same as for Model 1(0). However, the prior assumption on α, β and κ follows the smoothing process applied within the model; the autoregressive process disturbances follow independent identical Gaussian distributions with mean zero. For the three precision parameters of the autoregressive processes associated with α, β and κ, separate independent priors were assumed corresponding to Exponential distributions with mean 2,000.

For Model 1(b) the prior on the autoregressive coefficients for α, β and κ was $Un(-1, 1)$, thereby constraining the model to describe only stationary processes.

Model 2(0)

The prior assumptions on the parameters μ, α, β and κ are the same as for Model 1(0). The prior on the additional parameter γ is a multivariate Gaussian distribution centred on zero with covariance matrix $1000\mathbb{I}_{n_c}$, independently of the other parameters.

Models 2(a), (b) and (c)

The prior assumptions on the parameters μ, α and κ are the same as for Models 1(a), 1(b) and 1(c). The prior on the additional parameter γ also follows the smoothing process applied within the model. The autoregressive process disturbances follow independent identical Gaussian distributions with mean zero and precision parameter assigned to an Exponential prior with mean 2,000.

For Model 2(b) the prior on the autoregressive coefficients for γ was $Un(-1, 1)$ as above.

Models 3(a), (b) and (c)

The prior assumptions on the parameters μ, α, κ and γ are the same as for Models 2(a), 2(b) and 2(c). The random effects $z_{x,t}$ are assumed to come from independent Gaussian distributions with mean zero and precision parameter τ. The prior on τ was the Exponential distribution with mean 2,000.

Model 4(0)

The prior assumptions on the parameters μ, α, κ and γ are the same as for Model 2(0). The priors on the additional parameters β and δ are independent multivariate Gaussians centred on zero with covariance matrices $1000\mathbb{I}_{n_a}$.

Models 4(a), (b) and (c)

The prior assumptions on the parameters μ, α, κ and γ are the same as for Models 3(a), 3(b) and 3(c). The priors on the additional parameters β and δ follow the smoothing process applied within the model; the autoregressive process disturbances correspond to independent identical Gaussian distributions with mean zero. For the precision parameters of the autoregressive processes, independent Exponential priors with mean 2,000 were assigned.

Models 5(a), (b) and (c)

The prior assumptions on the parameters are the same as for Models 4(a), 4(b) and 4(c), respectively, and the prior on the random effects $z_{x,t}$ is the same as for Model 3.

Models 6(a), (b) and (c)

The prior assumptions on the parameters are the same as for Models 5(a), 5(b) and 5(c).

Models 7(a),(b) and (c)

The prior assumptions on the parameters are the same as for Models 5(a), 5(b) and 5(c).

The prior assumption on the exposures reflects the fact that the exposures have been estimated. In England and Wales a population census is taken every 10 years. For years between censuses there is an interpolation between the two neighbouring censuses. In section 2.2 a brief explanation of the process was outlined. This introduces a number of possible errors: not everyone will be captured within the census and there is a small risk that individuals do not accurately record their date of birth; migration flows and the interpolation process can also introduce errors. Migration flows at ages 60 and above are small relative to the exposure and the assumption is that the errors within the population counts at the census dates dominate these other errors. The population counts at the census dates are underestimates of the true population size and, as these values are used for interpolation for population counts between censuses, the population counts at all points will be underestimates of the true values and therefore the exposures will be under-estimates. Although the population counts are under-estimates we would expect the underestimation to be quite small and therefore the exposure estimates will be reasonably good as well. Thus, an informative prior assumption is used to reflect that the observed exposure is a reasonably good estimate. The $w_{x,t}$ represent measurement errors in the exposures so $w_{x,t} \in (0,1)$ and, because of the assumption that the population estimates are likely to be reasonable, one possible approach is to assign independent Beta(1,s) priors for $w_{x,t}$ where $s \sim Un(1,10000)$. Fixing the first parameter of the Beta prior to be 1 simplifies the density and setting the scale parameter s to being larger than or equal to 1 ensures that the distributions of the errors are weighted toward zero. For $s = 1$ the density of the error is constant, for $s > 1$ the slope is downwards and the larger s is the tighter the distributions of errors are to zero. The upper limit of s was set to 10000 as at this point the likely errors are very close to zero. The final results were not sensitive to changes in this upper limit. Figure 4.1 shows a histogram of 2000 simulations from the prior on $w_{x,t}$. The figure illustrates that the samples are close to zero; the upper quartile value is 0.00045.

The counting of deaths is a more robust process and therefore the number of deaths

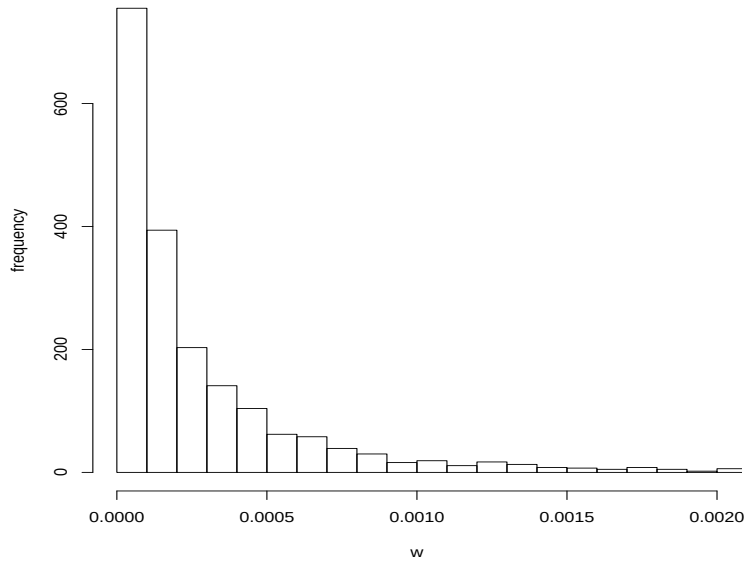


Figure 4.1: Histogram of 2000 simulations from the prior distribution of $w_{x,t}$.

is likely to be much more accurate than the exposures, so measurement error in death counts was ignored.

4.2 Implementation

For Models 1, 2, 4 and 7 parameter draws were obtained using the Metropolis-Hastings algorithm. In each iteration of the MH algorithm, 10 randomly selected elements for each parameter α, κ, γ, β and δ (provided the model included the parameter) in turn were chosen for updating and proposals made to update these parameter elements whilst keeping the remaining parameter elements in the vector unchanged. Proposals were generated by adding a disturbance, that has a 0 mean vector, to the current parameter values. For the models with parameter smoothing, the disturbance followed a highly correlated multivariate Gaussian distribution, with correlations of 0.99 between parameter elements, and a common variance being tailored for each parameter group to give a reasonable mixing rate; an acceptance rate of around 40% to 50% was targeted. For the models with no smoothing of parameters, the disturbance followed a multivariate Gaussian distribution with diagonal covariance matrix with the variance being tailored for each parameter group again to give an acceptance rate of around 40% to 50%.

For the AR(1) coefficients, proposals were generated firstly by transforming the current value, ρ say, to $\log(\frac{1+\rho}{1-\rho})$ and then adding a Gaussian disturbance with mean zero and with variance parameter set so that the acceptance rates of ρ were around 50%.

In all models that include smoothing priors, the choice of conjugate Gamma priors for the precision parameters results in the full conditional distributions of the precision parameters also being Gamma distributions and therefore parameter draws were obtained (by Gibbs sampling) from these conditional distributions.

Models 3, 5 and 6 include random effect parameters $z_{x,t}$. Besag et al. (1995) explained that, by treating the $\log m_{x,t}$'s (or in the case of Model 6, logit $m_{x,t}$'s) as parameters and expressing the random effects in terms of other parameters including $\log m_{x,t}$, one can derive full conditionals for the parameters α, κ, γ, β and δ that are multivariate Gaussian distributions. The full conditional distribution of the $\log m_{x,t}$ parameters is a non-standard distribution and samples are drawn using the MH algorithm. Proposals were generated by adding disturbances to the current parameter values. Each disturbance was Gaussian with mean zero and the variance was set to give an acceptance rate of approximately 50%.

Prior sensitivity analysis was carried out. The key prior sensitivity was that of the smoothing priors for the main parameters. The impact of the smoothing prior was not influential in the fitting but was very influential for the forecasting. These results are discussed in detail in chapter 5. Apart from the errors on the population counts associated with Models 7(a), 7(b) and 7(c), the remaining priors were non-informative or mildly informative. The fixed values of the hyper-parameters were reduced and increased and the model results checked (using the analysis of fit statistics described in Chapter 5) to ensure that the prior had little impact on the results. The prior on the population errors was influential and sensitivity was done on the hyper-parameter s the fixed parameter was 10000 reduced to 50 without any material changes to the results.

Run Times

This section gives details of the average run time of the different models that in-

corporate smoothing priors. Table 4.1 gives the time in hours to complete 2 million iterations, (implementation of the model and prediction of the log-mortality rates), using R version 2.15.1 (R Development Core Team, 2012) using 1 core of a Twin Quad Core Intel(R) Xeon(R) CPUs W5590 @ 3.33GHZ 8MB Cache 24GB (12X2GB)1333MHZ DDR3 ECC-RD RAM.

Model	1	2	3	4	5	6	7
Time (hrs)	2.7	3.1	7.0	3.9	8.8	8.7	8.8

Table 4.1: Average model runtime in hours.

4.2.1 Parameter Transformation

For the Models 1(0), 2(0) and 4(0) where the parameters were independent (no a priori smoothing), draws from the posterior were unconstrained. The μ and α parameters were combined into one age related parameter to reduce the number of constraints required. However, to produce the graphs in section 5.2.1 and section 5.4.1 the unconstrained parameter samples were projected on to the parameter space that satisfies the following constraints: $\sum_x \beta_x = 1$, $\sum_x \delta_x = 1$, $\sum_t \kappa_t = 0$, $\sum_i \gamma_i = 0$ and $\gamma_1 = \gamma_2$. The projection method is illustrated for Model 4(0). Model 1(0) and Model 2(0) follow a similar algorithm. The age, period and cohort parameters are centred on zero, the β_x and δ_x parameters are considered as weights and are constrained to sum to one. The suffices c and n denote the current untransformed and transformed parameters respectively. The process follows the following steps, with $c^{'}$ denoting an intermediate transformation:

$$B^c = \sum_{x=1}^{n_a} \beta_x^c, \quad \Delta^c = \sum_{x=1}^{n_a} \delta_x^c, \quad \overline{\kappa}^c = \frac{1}{n_y} \sum_{t=1}^{n_y} \kappa_t^c, \quad \overline{\gamma}^c = \frac{1}{n_c} \sum_{j=1}^{n_c} \gamma_j^c,$$

$$\beta_x^{c^{'}} = \frac{\beta_x^c}{B^c}, \quad \delta_x^{c^{'}} = \frac{\delta_x^c}{\Delta^c}, \quad \kappa_t^{c^{'}} = \kappa_t^c - \overline{\kappa}^c, \quad \gamma_j^{c^{'}} = \gamma_j^c - \overline{\gamma}^c,$$

$$\kappa_t^n = \kappa_t^{c^{'}} B^c, \quad \gamma_j^n = \gamma_j^{c^{'}} \Delta^c$$

$$\alpha_x^n = \alpha_x^c + \beta_x^c \overline{\kappa}^c + \delta_x^c \overline{\gamma}^c.$$

Model 3 is very similar to the model used by Besag et al. (1995). They suggested

that, for computational stability, at each complete cycle the parameters α, κ and γ are centred. This centring was adopted here for each of the models where there is a priori smoothing of parameters. This centring does not violate the convergence of the Markov chain as the prior probabilities are invariant under this transformation and the likelihood is unchanged. The method is demonstrated below for Model 4; the transformations for other models are obtained similarly. As mentioned above the transformation should enhance computational stability for the age, period and cohort parameters. There is no scaling move on the β and δ parameters as this would require changes to the precision parameters and the process would become very complex. Using the notation above the transformation is described below.

Firstly, centre the period and cohort parameters by deducting the means:

$$\kappa_t^n = \kappa_t^c - \overline{\kappa}^c, \quad \gamma_j^n = \gamma_j^c - \overline{\gamma}^c.$$

A corresponding adjustment needs to be made to compensate for the centring of the κ and γ parameters. The adjustment depends upon the model and is illustrated for Model 4. The adjustment to the current value of the age-related α parameters is as follows:

$$\alpha_x^{c'} = \alpha_x^c + \beta_x^c \overline{\kappa}^c + \delta_x^c \overline{\gamma}^c, \quad \overline{\alpha}^{c'} = \frac{1}{n_a} \sum_{x=1}^{n_a} \alpha_x^{c'}.$$

Then the adjusted α parameters are centred and μ correspondingly adjusted:

$$\alpha_x^n = \alpha_x^{c'} - \overline{\alpha}^{c'}, \quad \mu^n = \mu^c + \overline{\alpha}^{c'}.$$

If the model does not contain β or δ parameters then adjustment to compensate for the centring of the κ and γ parameters is made to μ rather than to α. These transformations keep $\log m_{x,t}$ or $\text{logit}\, m_{x,t}$ and the likelihood unchanged.

CHAPTER 5

Mortality Models - Results

This section reviews how well the various models fit the observed data, firstly for male data and then for female data. For the models that best fit the data, out-of-sample testing is done to assess how well these models predict mortality in the near future.

Computations are carried out using Markov-Chain-Monte-Carlo methods (MCMC) by setting up a Markov chain whose stationary distribution is the posterior distribution of the model parameters. The Monte-Carlo approach allows for inference based on sampling the posterior distribution of the parameters $m_{x,t}$ and its hyper-parameters.

5.1 Model Convergence

To assess convergence of the MCMC sampler, for each model several simulations were run, with different starting values and seed values, and the results were compared. To allow for the initial transient phase, a number of samples were discarded - often referred to as the burn-in - before posterior draws of the parameters were stored. The length of the burn-in period depended upon the model complexity; for Model 1 the burn-in was 200,000 and for the other models the burn-in was the first 1,000,000 parameter draws. The amount of burn-in was assessed by considering the trace output of the parameter of interest, $\log m_{x,t}$ or in the case of Model 6 logit $m_{x,t}$. This is illustrated in Figure 5.1 for Model 5(b) using male data for a selection of the 1640 $\log m_{x,t}$ parameters using 4 different starting values for the chain. In addition

plots of the log-likelihood at the sampled values were considered as another ad-hoc assessment of convergence as illustrated in Figure 5.2, also for Model 5(b).

A more formal assessment is that of Gelman and Rubin's $\hat{R}$ (Gelman et al. (2003)). Suppose we collect S samples of $m_{x,t}$ from each of C chains and denote this as $m^{sc}_{x,t}$. Define the within-sequence mean ($\bar{m}^{c}_{x,t}$) and overall mean ($\bar{m}_{x,t}$) as

$$\bar{m}^{c}_{x,t} = \frac{1}{S}\sum_{s=1}^{S} m^{sc}_{x,t},$$

$$\bar{m}_{x,t} = \frac{1}{C}\sum_{c=1}^{C} \bar{m}^{c}_{x,t}.$$

Define the between-sequence variance ($B_{x,t}$) and within-sequence variance ($W_{x,t}$) as

$$B_{x,t} = \frac{S}{C-1}\sum_{c=1}^{C}(\bar{m}^{c}_{x,t} - \bar{m}_{x,t})^2$$

$$W_{x,t} = \frac{1}{C}\sum_{c=1}^{C}\left[\frac{1}{S-1}\sum_{s=1}^{S}(m^{sc}_{x,t} - \bar{m}^{c}_{x,t})^2\right].$$

We can now construct two estimates of the variance of $m_{x,t}$. The first estimate is $W_{x,t}$ which should underestimate $var(m_{x,t})$ if the chains have not ranged over the full posterior. The second estimate

$$\hat{V}_{x,t} = \frac{S-1}{S}W_{x,t} + \frac{1}{S}B_{x,t},$$

is an estimate of $var(m_{x,t})$ that is unbiased when stationarity conditions are reached but is an overestimate when the starting points were over-dispersed. The convergence diagnostic statistic developed by Gelman and Rubin is defined as $\hat{R}$, where

$$\hat{R}_{x,t} = \sqrt{\frac{\hat{V}_{x,t}}{W_{x,t}}} = \sqrt{\frac{S-1}{S} + \frac{B_{x,t}}{SW_{x,t}}}.$$

$\hat{R}$ measures the degree to which the posterior variance would decrease if we were to continue sampling by increasing S. If $\hat{R} \approx 1$ for any $m_{x,t}$, then that estimate is reliable. Essentially it means the variance between the chains is similar to the variance within each chain.

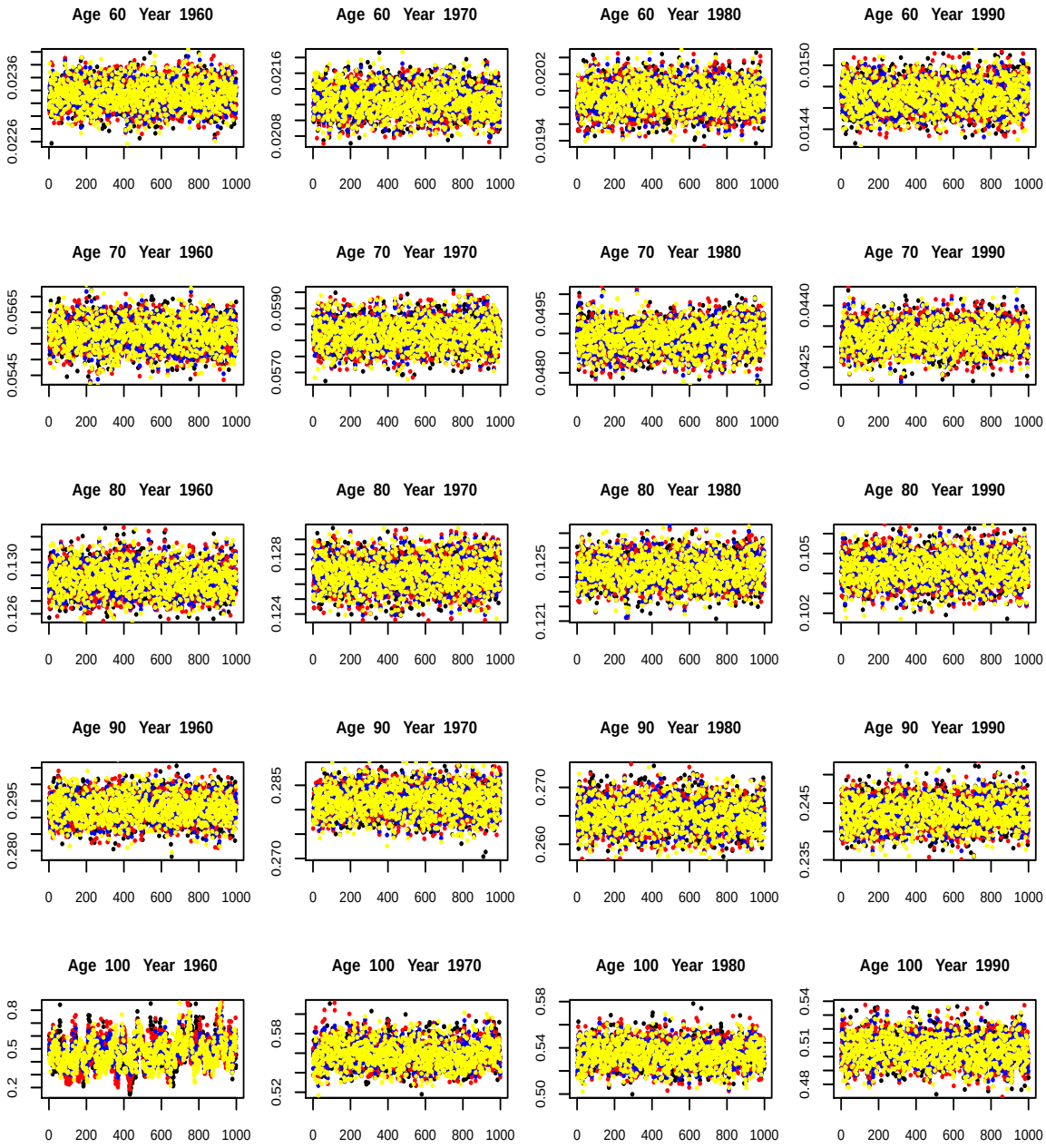


Figure 5.1: Panel of trace plots of $m_{x,t}$ for a selection of male ages in different calendar years from Model 5(b). Each plot shows 4 trace plots generated from chains with different starting values and after a burn-in of 1,000,000 iterations.

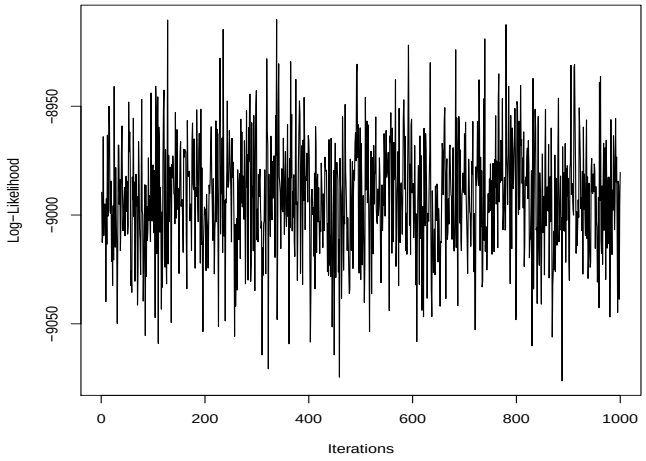


Figure 5.2: Plot of the Log-likelihood of Model 5(b) at the sampled values using male data. The samples were thinned by storing every 400th sample.

For Model 5(b) across all $m_{x,t}$ parameters the maximum $\hat{R}_{x,t}$ value was 1.02 for males and 1.03 for females, indicating that the chains have converged. Similar results were obtained for the other models.

The draws from the posterior are correlated and, to reduce this correlation, the draws were thinned with every 400th draw retained as a nearly independent sample. The level of thinning was chosen based on plots of the autocorrelation function of the posterior draws of the parameters. Apart from the very oldest of ages, at the bottom of Figure 5.3, thinning of every 400th draw results in very little or no autocorrelation between the draws. This is illustrated in Figure 5.3 for Model 5(b) using male data for a selection of the 1640 $m_{x,t}$ parameters.

For each model there are three different forms of smoothing priors; these do not have a significant differential impact on the fit of the model as demonstrated in the case of Model 6 in Table 5.6. Therefore, in the following sections that discuss model fit, the smoothing priors correspond to autoregressive processes of order one on the first differences of the parameters, i.e., version (b) of each model.

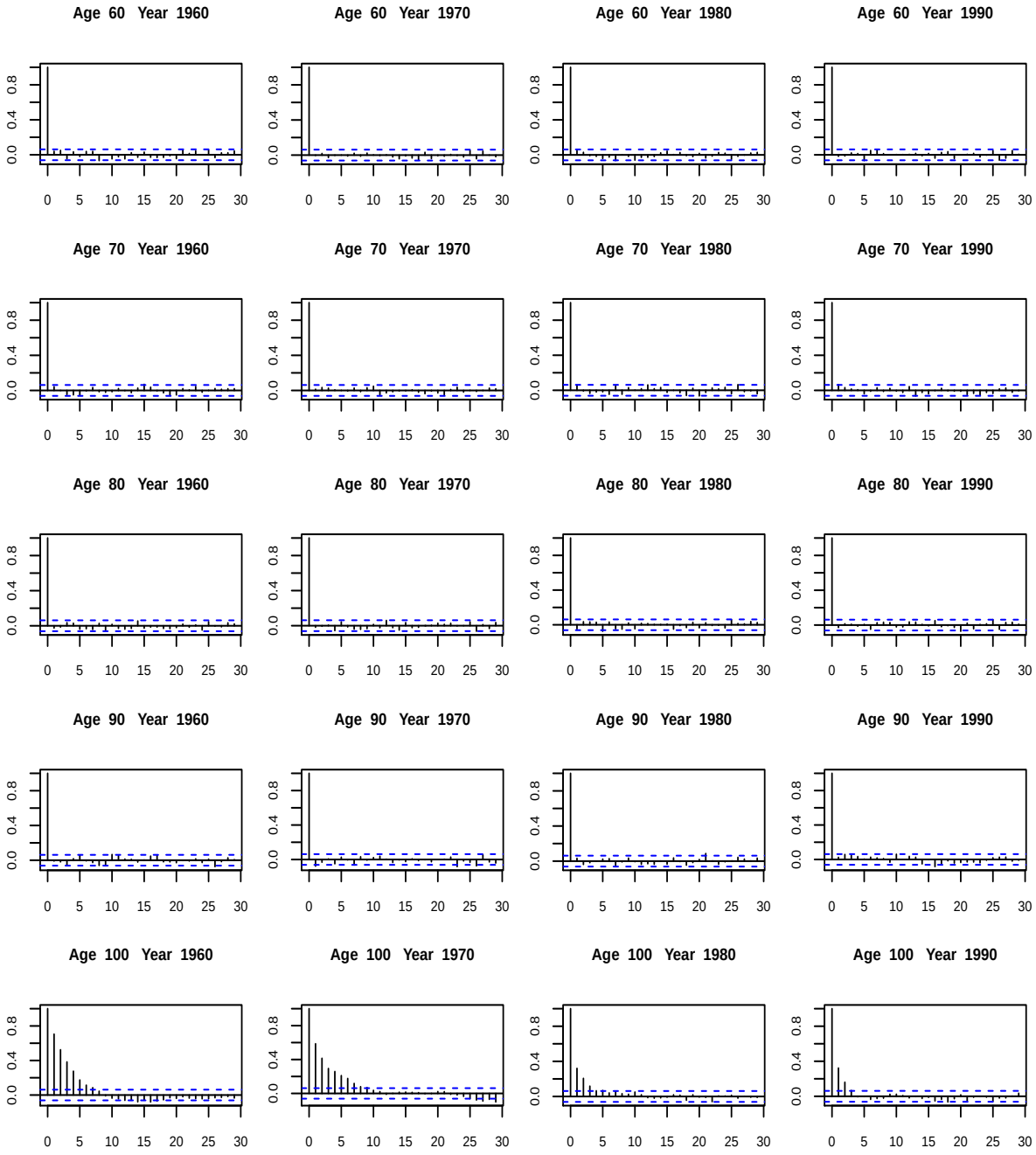


Figure 5.3: Model 5(b) - estimated autocorrelation function of parameter $m_{x,t}$ for a selection of male ages in different calendar years for lags up to 30, corresponding to equivalent panel in Figure 5.1.

5.2 England and Wales male mortality

5.2.1 Unsmoothed Parameters

Although the primary focus of the modelling is the set of mortality rates, it is instructive to consider the parameters α, β, κ, δ and γ without any smoothing. To render the parameters identifiable the following constraints were applied following the algorithm outlined in section 4.2.1 : $\sum_x \beta_x = 1$, $\sum_x \delta_x = 1$, $\sum_t \kappa_t = 0$, $\sum_i \gamma_i = 0$ and $\gamma_2 = \gamma_2$. Here i indexes the distinct cohorts. Box-plots of the posterior parameter draws obtained from Models 1(0), 2(0) and 4(0), are shown in Figures 5.4, 5.5 and 5.6 respectively.

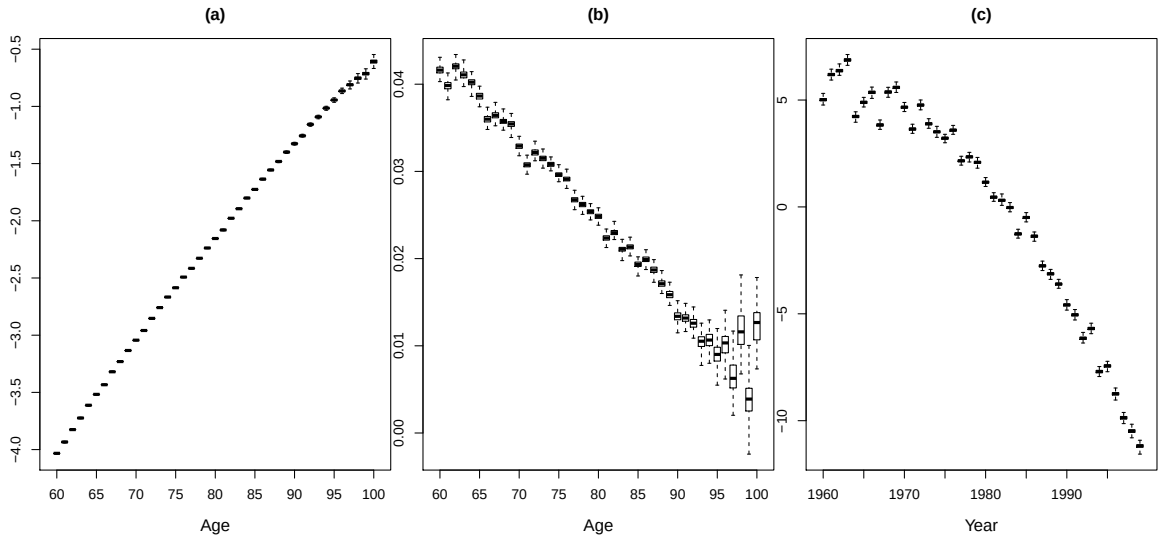


Figure 5.4: Boxplots of parameter samples from Model 1 using male data with $\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t$. (a) is $\mu + \alpha$ for each age, (b) is β for each age and (c) is κ for each calendar year

What is clear from Figures 5.4, 5.5 and 5.6 is that, as expected, even without incorporating a smoothing process in the model the parameters show a relatively smooth pattern. In all three models the age-effect parameter α increases with age reflecting that mortality rates generally increase with age. The trend in the period parameter κ generally reduces with time indicating a decreasing trend in mortality rates with time for each age. The cohort parameter γ has much wider boxplots for the oldest and youngest cohorts because these number of data points included are very few. The width of the boxplots for both β and δ at the older ages are much wider than for

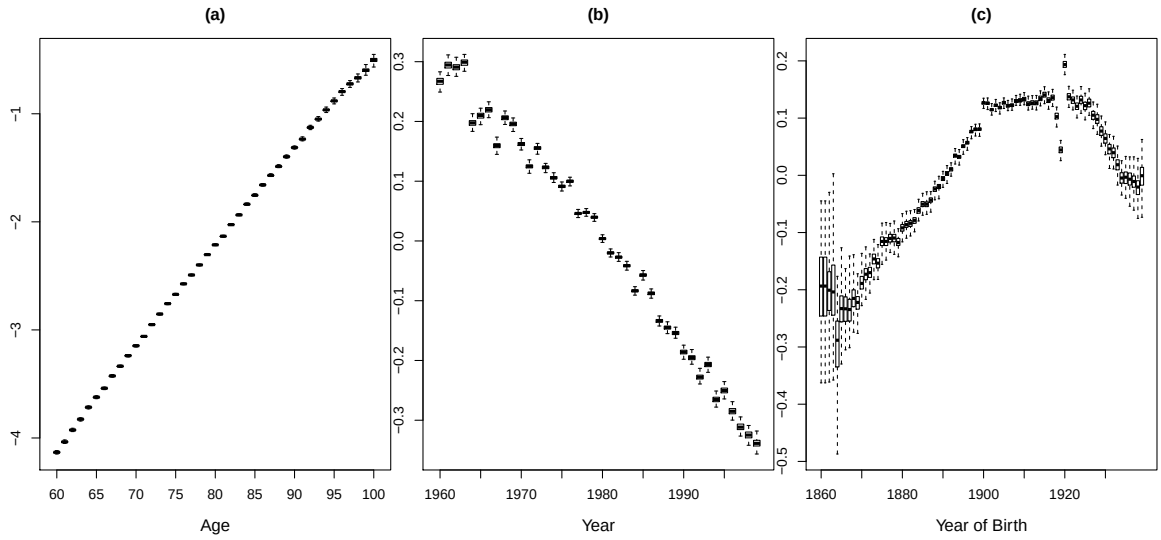


Figure 5.5: Boxplots of parameter estimates from Model 2 using male data with $\log m_{x,t} = \mu + \alpha_x + \kappa_t + \gamma_{t-x}$. Fig. (a) is $\mu + \alpha$ for each age, (b) is κ for each calendar year and (c) is γ for each cohort

the younger ages; this reflects the increased variability of empirical mortality rates over time at these older ages.

5.2.2 Random Effects

The random effects are Gaussian with mean zero and the precision is a parameter in the model. Figure 5.7 shows the posterior distribution of the random effects precision parameter for Models 3(b), 5(b) and 6(b) using male data. The distribution of the precision parameter is similar for the other versions of the model with different smoothing priors. The age-period-cohort model (Model 3(b)) has a much greater spread of random effects than the other models and therefore may include some allowance for model misspecification. For Models 5(b) and 6(b) the precision parameters of the random effects are of a similar order but much larger than that of Model 3(b). Therefore, Model 3(b) incorporates much greater random effects than Models 5(b) and 6(b).

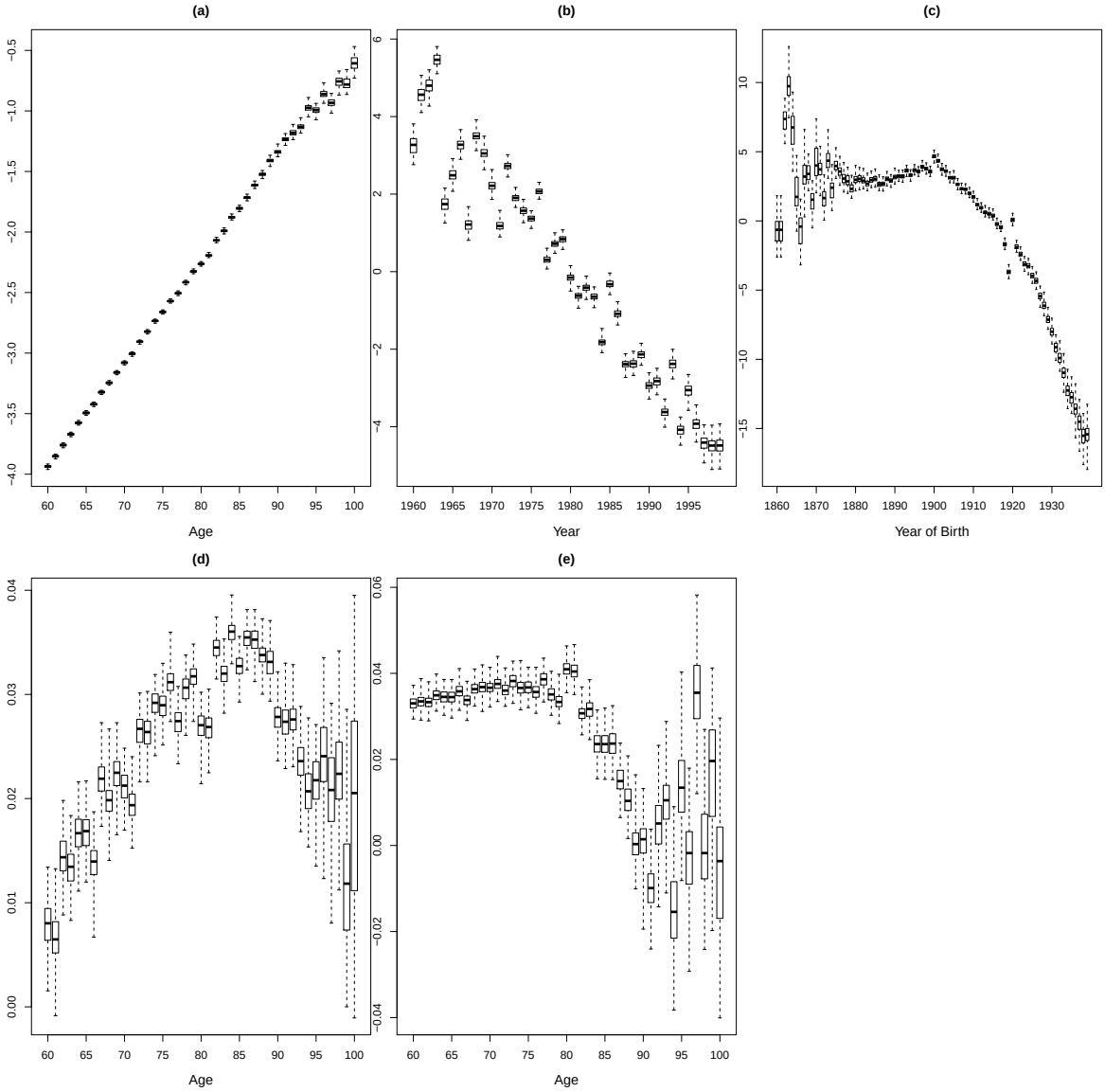


Figure 5.6: Boxplots of parameter estimates from Model 4 using male data with $\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x}$. (a) is $\mu + \alpha$ for each age, (b) is κ for each calendar year, (c) is γ for each cohort, (d) is β for each age and (e) is δ for each age

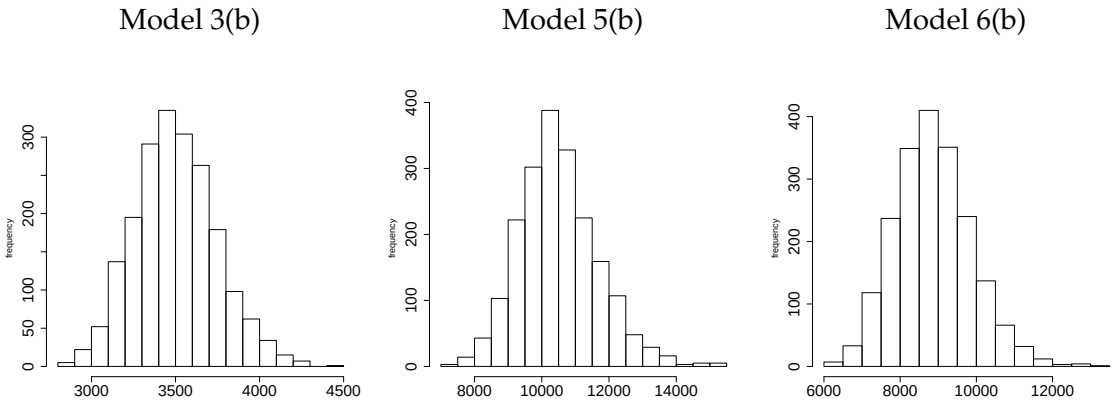


Figure 5.7: Histograms of 2000 samples from the posterior distributions of the random effect precision parameter of Models 3(b), 5(b) and 6(b) for male data.

5.2.3 Autoregressive smoothing coefficient

Figures 5.8 and 5.9 show the posterior distributions of the autoregressive coefficient for the age, period and cohort parameters for each of the models that include the autoregressive smoothing priors based on the male data. Table 5.1 shows 95% credible intervals for these autoregressive coefficients. The age autoregressive coefficients are very close to 1 for all models (i.e. very similar to a random walk on first differences). For the period autoregressive coefficients, the 95% central posterior intervals for Models 3(b), 5(b) and 6(b) indicate a relatively small negative coefficient. For the other models the 95% central posterior intervals straddle zero. Therefore, the difference in the neighbouring period parameters will quickly reduce to around zero. For the cohort autoregressive coefficients for all models except Model 3, the 95% central posterior intervals indicate that the coefficient is strongly positive, the median values being within the range 0.5 to 0.8. For Models 3 the 95% central posterior interval indicates a negative coefficient.

5.2.4 Fitted Models - Male Data

Analysis of Standardised Residuals

One method for assessing the fit of the models is by analysing the residuals, that is the difference between the observed number of deaths and the estimated expected number of deaths produced by the model for each age and year. Equation 5.2.1 for

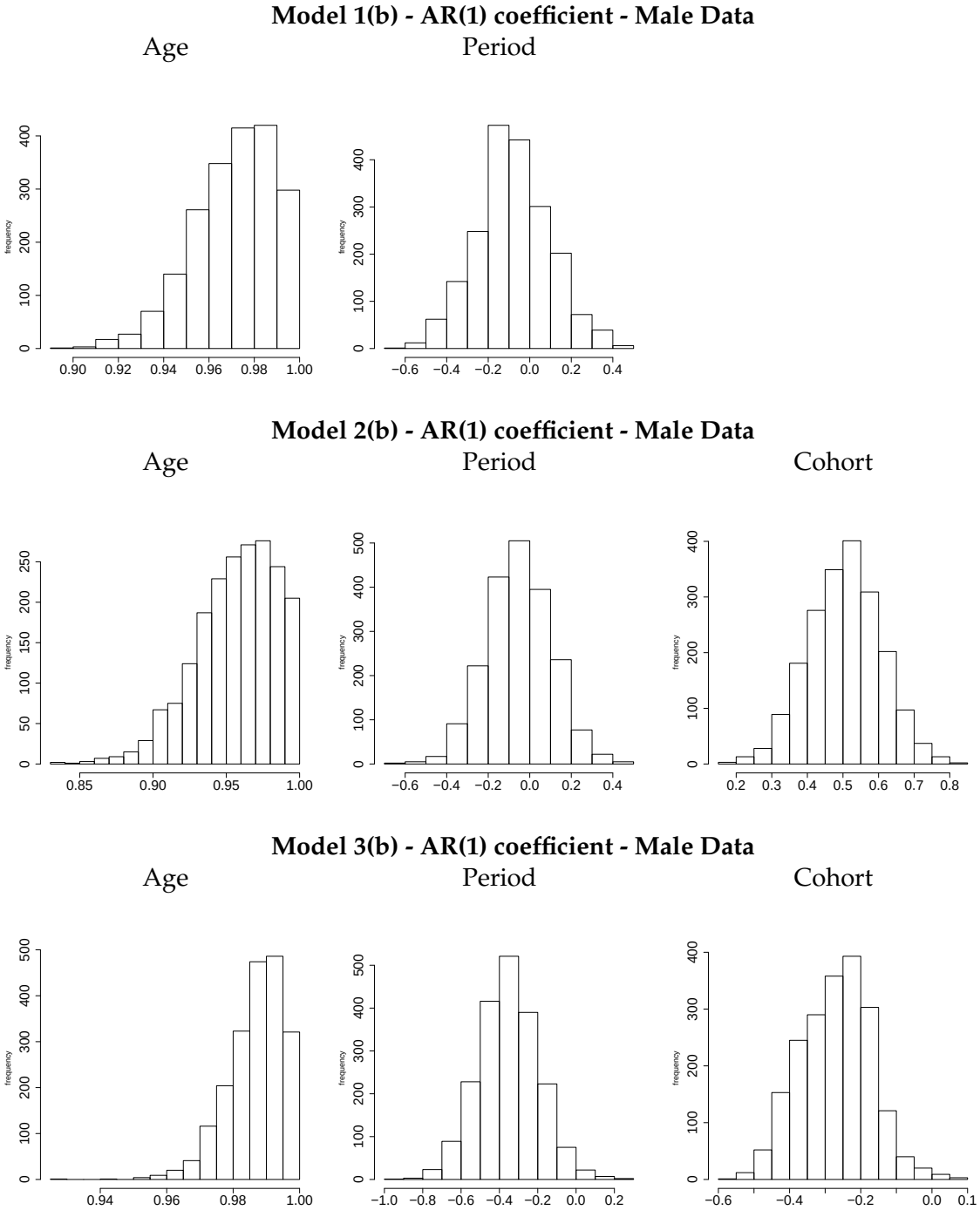


Figure 5.8: Histograms of 2000 samples from the posterior distributions of the coefficient of the autoregressive series for the age, period and cohort parameters for Models 1(b), 2(b) and 3(b) for male data.

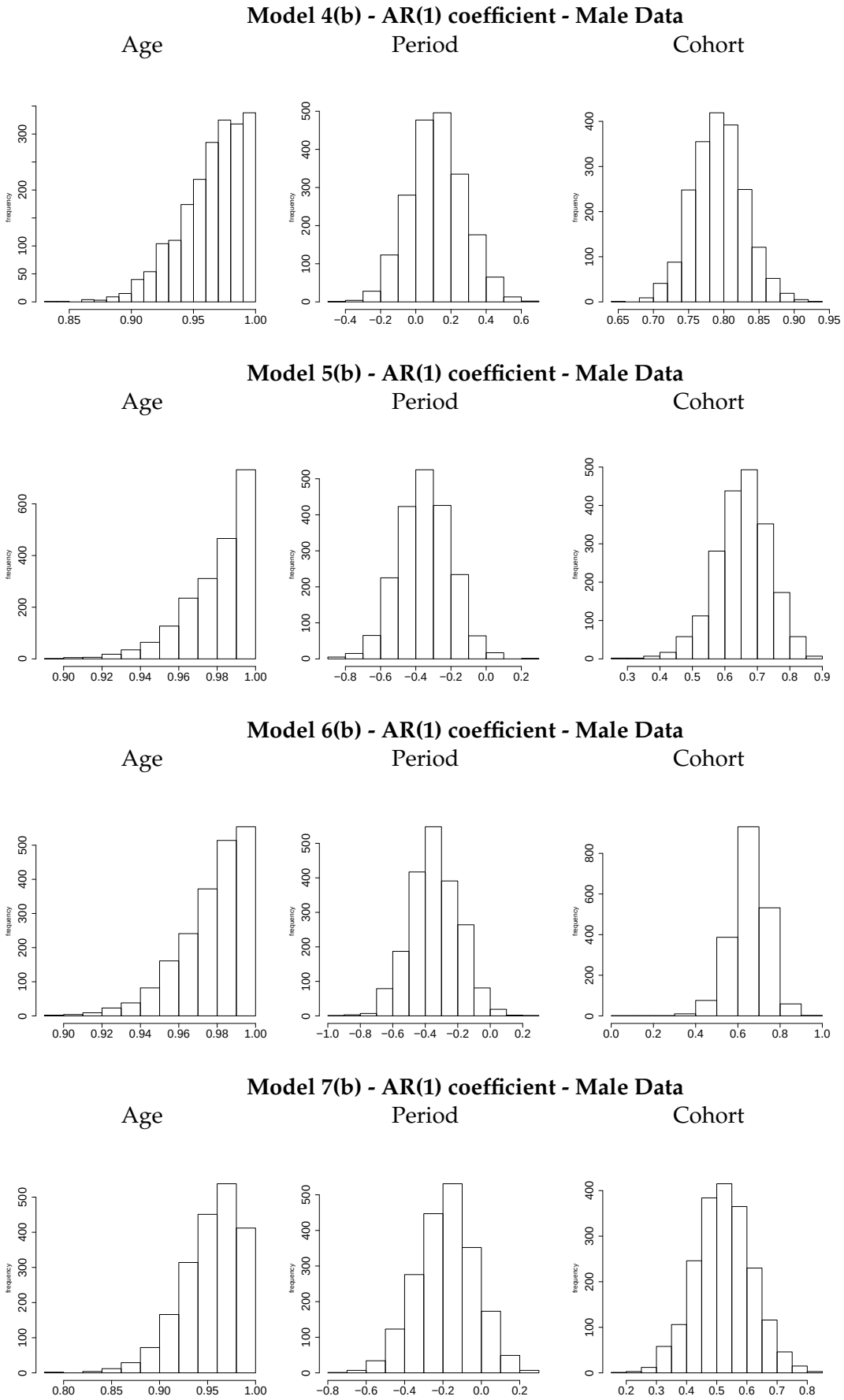


Figure 5.9: Histograms of 2000 samples from the posterior distributions of the coefficient of the autoregressive series for the age, period and cohort parameters for Models 4(b), 5(b), 6(b) and 7(b) for male data.

	coeff. of α		coeff. of κ		coeff. of γ	
Model	lower	upper	lower	upper	lower	upper
1(b)	0.938	1.000	-0.435	0.278		
2(b)	0.906	1.000	-0.357	0.252	0.303	0.701
3(b)	0.972	1.000	-0.667	-0.041	-0.459	-0.077
4(b)	0.915	1.000	-0.177	0.429	0.720	0.866
5(b)	0.947	1.000	-0.642	-0.067	0.471	0.806
6(b)	0.944	1.000	-0.638	-0.047	0.470	0.808
7(b)	0.896	1.000	-0.492	0.105	0.338	0.708

Table 5.1: Table shows 95% posterior intervals for the autoregressive coefficient of the smoothing prior for α, κ and γ for each model using male data. For alpha, given the asymmetric shape of the posterior sample, it seemed more appropriate to leave all the 5% to the left of the interval, rather than using a 95% central posterior intervals.

the standardised residual $r_{x,t}$ for age x in calendar year t is

$$r_{x,t} = \frac{d_{x,t} - E_{x,t}\widetilde{m}_{x,t}}{\sqrt{E_{x,t}\widetilde{m}_{x,t}}}, \tag{5.2.1}$$

where $\widetilde{m}_{x,t}$ is the mean of the $m_{x,t}$ parameters calculated from the sampled model parameters. The variance of the standardised residuals is then the variance of $r_{x,t}$ over all ages x and years t. Table 5.2 shows the variance of the standardised residuals from each model.

Model	1	2	3	4	5	6	7
	4.26	2.22	0.51	1.33	0.62	0.64	0.73

Table 5.2: Variance of standardised residuals for male data

A good model would have a variance of standardised residuals close to 1. Model 1, based on the Lee-Carter parameter structure, with a variance of standardised residuals equal to 4.26 gives a poor fit to the male data. Adding the cohort parameters improves the fit but it is still unsatisfactory. Comparing the variances of the standardised residuals for Models 2 and 3 as well as Models 4 and 5 indicates that the addition of the random effects moves the residuals from being over-dispersed to being under-dispersed. The variances of the standardised residuals for Models 3, 5, 6 and 7 are similar. Model 3 has the lowest figure, largely due to the much greater random effects allowed for within this model. There are other methods that can be employed where the overdispersion parameter can be calculated directly, for example, by using a chi-squared statistic divided by the number of data points less

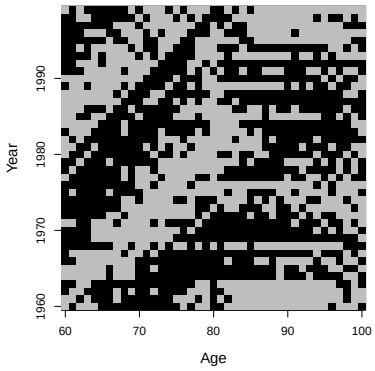
the effective number of parameters. The effective number of parameters for models including random effect models can be calculated by following the approaches detailed in Burnham and Anderson (1998). Underpinning the overdispersion parameter calculation is the assumption that the systematic part of the Poisson regression is correct.

Join Count Statistic

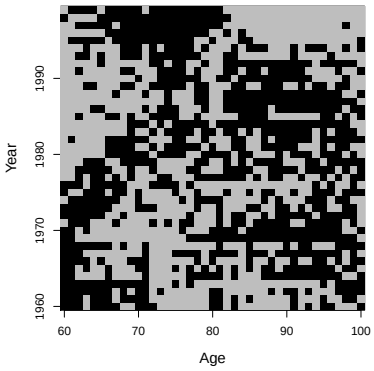
Cairns et al. (2009) employ a graphical presentation of the standardised residuals to consider the fit of models. This presentation is repeated for the Models 1(b) to 7(b) and is shown in panel 1 of Figure 5.10. Each plot shows the sign of the standardised residuals $r_{x,t}$ at each age x from 60 to 100 and calendar year t, from 1960 to 1999. For a given age and year, black pixels show the points where the mean of the modelled deaths exceeds the observed number of deaths, and the grey pixels show where the mean of the modelled deaths is below the observed number of deaths. For a good model we would look to see a random pattern of black and grey pixels if the assumptions of the model are satisfied and that the $d_{x,t}$'s, conditional on the $m_{x,t}$'s, are independent random variables.

The plot of residuals for Model 1, the first panel plot in Figure 5.10, shows diagonal blocks of the same colour, indicating the presence of cohort effects. This is not surprising as the model does not include any cohort parameters. The plots of the residuals for the other models do not display any obvious cohort effect. However, the patterns of residuals do not appear to be random. Certainly the plots for Models 4, 5 and 6 look to have less structure than those of Models 2 and 3. The residuals of Model 7 show a pattern similar to those of Models 4, 5 and 6. This is not surprising as these models have similar parameter structures.

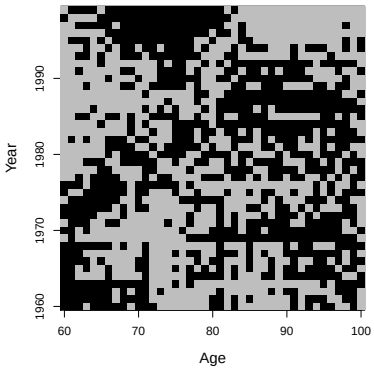
In Cairns et al. (2009) the authors assessed the fit by visual inspection. However, a more objective approach to assessing the randomness of the residual patterns can be carried out by analysing the colours of neighbouring cells; this method was set out in Cliff and Ord (1981). Their method focuses on whether or not there is spatial autocorrelation in the map of coloured cells, using the number of joins of grey and black cells. Table C.1 in Appendix C illustrates a join. A test statistic is constructed based on the distribution of the number of J joins of black and grey cells in the map, where



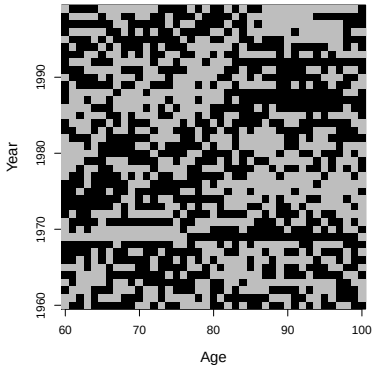
(1) Model 1(b)



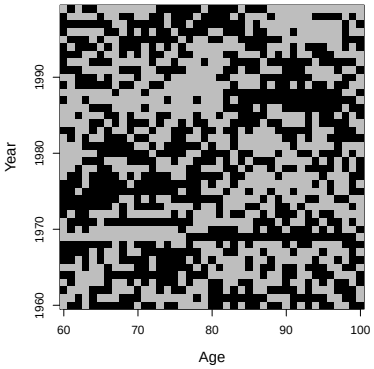
(2) Model 2(b)



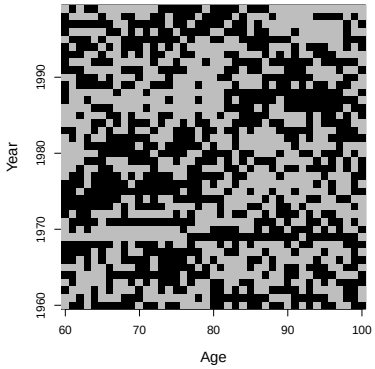
(3) Model 3(b)



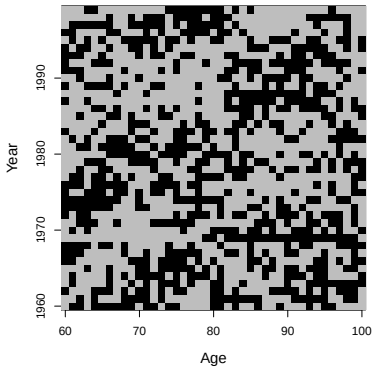
(4) Model 4(b)



(5) Model 5(b)



(6) Model 6(b)



(7) Model 7(b)

Figure 5.10: Panel of plots for each model based on male data showing the sign of the standardised residuals $r_{x,t}$ over the years 1960 to 1999 and ages 60 to 100. The black pixels indicate the points where the mean of the modelled deaths exceeds the observed number of deaths, and the grey pixels indicate where the mean of the modelled deaths is below the observed number of deaths.

$J \in \{0,1,2,3,4\}$, i.e., the number of joins between neighbouring cells of different colours. 'Neighbourhood' in this context refers to a first-order neighbourhood. The null hypothesis H_0 is that there is no autocorrelation in the residual pattern. Under H_0 the expected frequency distribution of J joins of neighbouring black and grey cells can be computed, say $N(J)$. From the map of residuals the number of J joins is counted for each value of J, giving $O(J)$; then, using a modified chi-squared test based on $N(J)$ and $O(J)$, one can assess whether or not the H_0 can be rejected. The number of degrees of freedom for the modified chi-squared test is 4 as J can take 5 possible values and the probability of whether a cell is black or white is known, in this case 0.5. A standard chi-squared test cannot be used because the observed counts are not independent; each cell appears as a neighbour for several other cells. Moreover, each cell appears both as a reference cell and as a neighbour, thus inducing an element of double counting. As a result of both theoretical and Monte-Carlo studies, Cliff et al. (1975) suggested a modification to the standard chi-squared test. The detailed calculations as applied to a 2-dimensional 2-colour problem are shown in Appendix C.

Table 5.3 shows the modified chi-squared p-value assuming that the pattern of residuals displayed in Fig. 5.10 shows no autocorrelation. The number of degrees of freedom is 4.

Model	1	2	3	4	5	6	7
Chi-squared	<0.0001	<0.0001	<0.0001	.0008	<0.0001	<0.0001	<0.0001

Table 5.3: Table of p-values from modified chi-squared distribution assuming that the pattern of residuals displayed in Fig. 5.10 shows no spatial autocorrelation.

Given the very small p-values, the residuals of each model show clear evidence of autocorrelation, indicating that there is some remaining structure in the residuals that is not included within the models. However, as the model increases in complexity visually there is a substantial reduction in the structure of the residuals.

5.2.5 Deviance Information Criterion

The Deviance Information Criterion (DIC) is a Bayesian method for model comparison, based on a trade-off between the fit of the data to the model and the corre-

sponding complexity of the model. Full details of DIC can be found in Spiegelhalter et al. (2002). If we denote by π the likelihood of the model in question then the Deviance Information Criterion is defined for the parameters in focus, Φ, as

$$DIC = \mathbb{E}_{\Phi|y}\left(D(\Phi)\right) + P_D,$$

where

$$D(\Phi) = -2\log \pi(y|\Phi)$$

and P_D is the effective number of parameters:

$$P_D = \mathbb{E}_{\Phi|y}\left(D(\Phi)\right) - D\left(\mathbb{E}_{\Phi|y}(\Phi)\right).$$

To calculate DIC, one estimates $\mathbb{E}_{\Phi|y}\left(D(\Phi)\right)$ as the sample mean of $D(\Phi^{(i)})$ over the N MCMC samples, and $D(\mathbb{E}_{\Phi|y}(\Phi))$ as the value of the deviance evaluated at the average of the N samples of Φ. Then

$$P_D \approx \frac{1}{N}\sum_{i=1}^{N} D(\Phi^{(i)}) + 2\log \pi(y|\widetilde{\Phi}) \tag{5.2.2}$$

and

$$DIC \approx \frac{2}{N}\sum_{i=1}^{N} D(\Phi^{(i)}) + 2\log \pi(y|\widetilde{\Phi}), \tag{5.2.3}$$

where $\widetilde{\Phi} = \frac{1}{N}\sum_{i=1}^{N}\Phi^{(i)}$.

DIC provides a penalised fit criterion that is applicable to comparing non-nested models and models including random effects Spiegelhalter et al. (2002). The model with the smallest DIC is regarded as the model of choice.

The parameters in focus, Φ, are $\log m$ or $\text{logit}\, m$. The resulting DIC for the models is shown in Table 5.4, which indicates Model 5(b) and Model 6(b) followed by Model 7(b) as the preferred models of those considered. The effective number of parameters is the highest for Model 3 due to the much greater random effects allowed for within this model.

In addition to DIC there are a number of other statistics that can be used for model selection. For example, Bayesian analysis often uses Bayes factors but these cannot

Model	1	2	3	4	5	6	7
DIC	23,504	20,267	19,014	18,868	18,609	18,607	18,754
P_D	120	140	920	174	625	607	627

Table 5.4: Deviance Information Criteria for Models 1(b) to 7(b) using male data

be used in this instance as most of the fitted models incorporate improper priors. There are non-Bayesian statistics for model comparison like Akaike information criterion (AIC) or Bayesian information criterion BIC. The calculation of either AIC or BIC involves the calculation of the maximum value of the log likelihood which is a significant challenge because most of the models include well over 100 parameters as well as hyper-parameters and random effects.

5.2.6 Posterior predictive intervals - Male Data

Model checks can be done by creating posterior predictive intervals using samples from the posterior density. Replicate data d^{rep} are generated from

$$\begin{aligned} p(d^{rep}|\mathcal{D}) &= \int p(d^{rep}|\mathcal{D},\Phi)p(\Phi|\mathcal{D})d\Phi \\ &= \int p(d^{rep}|\Phi)p(\Phi|\mathcal{D})d\Phi, \end{aligned} \tag{5.2.4}$$

where $\mathcal{D}$ denotes the observed number of deaths and Φ denotes all the parameters in the model. The last equation holds because of the assumption that d^{rep} and $\mathcal{D}$ are independent given the parameters Φ. The posterior predictive distribution $p(d^{rep}|\mathcal{D})$ is obtained by iteratively sampling from $p(\Phi|\mathcal{D})$ and $p(d^{rep}|\Phi)$ to form the distribution. Using the samples from $p(d^{rep}|\mathcal{D})$, samples of $\hat{m}^{rep}$ can be generated by dividing the replicate number of deaths by the corresponding exposure. Then from these $\hat{m}^{rep}$ values posterior predictive intervals can be constructed. In a satisfactory model, namely, one that adequately reproduces the data being modelled, 95% of the empirical mortality rates are within, 95% of the posterior predictive intervals of $\hat{m}^{rep}$.

Figures 5.11, 5.12 and 5.13 show for each age, from 60 to 100, the trend in observed male mortality rates over the years 1960 to 1999 indicated by the solid blue line and

the dashed red lines are the boundary of 95% posterior predictive intervals according to Model 6(b). The posterior predictive interval for a mortality rate at given age and calendar year was derived by first generating 1000 replicated death counts from the model using the known exposure and each of the simulated values of $\hat{m}_{x,t}$. These replicated deaths where then divided by the known exposure to give replicated mortality rates that were then used to construct the posterior predictive intervals. For the younger ages up to around age 85 the posterior predictive intervals at the 95% level are quite narrow. For the very highest ages, 95 and over, the posterior predictive interval are much wider reflecting the higher variability of observed mortality rates at these high ages. These figures show that the empirical coverage of the posterior predictive interval at the 95% level is conservative covering almost all the observed rates. A summary of the empirical coverage by the posterior predictive intervals at different levels for each model is shown in Table 5.5. The results show that the coverage of the empirical rates by predictive posterior intervals from Model 1(b) and Model 2(b) are too low. This is consistent with the overdispersion of these models shown in Table 5.2. The posterior predictive intervals of Model 4(b) give the best coverages although slightly narrow followed by Models 5(b), 6(b) and 7(b) which provide conservative intervals.

PPI level	80%	95%	99%
Model 1(b)	0.56	0.75	0.86
Model 2(b)	0.65	0.84	0.93
Model 3(b)	0.95	0.99	1.00
Model 4(b)	0.77	0.93	0.98
Model 5(b)	0.93	0.99	1.00
Model 6(b)	0.92	0.99	1.00
Model 7(b)	0.92	0.98	1.00

Table 5.5: Coverage of empirical male mortality rates for ages 60 to 100 and years 1960 to 1999 by posterior predictive intervals (PPI) at different probability levels for each model.

5.2.7 Impact of Smoothing Priors on Model Fit

As mentioned at the beginning of this chapter the choice of smoothing prior is not very important for the fit of the model, as this section aims to demonstrate. The results shown are for Model 6 with the three different forms of smoothing priors

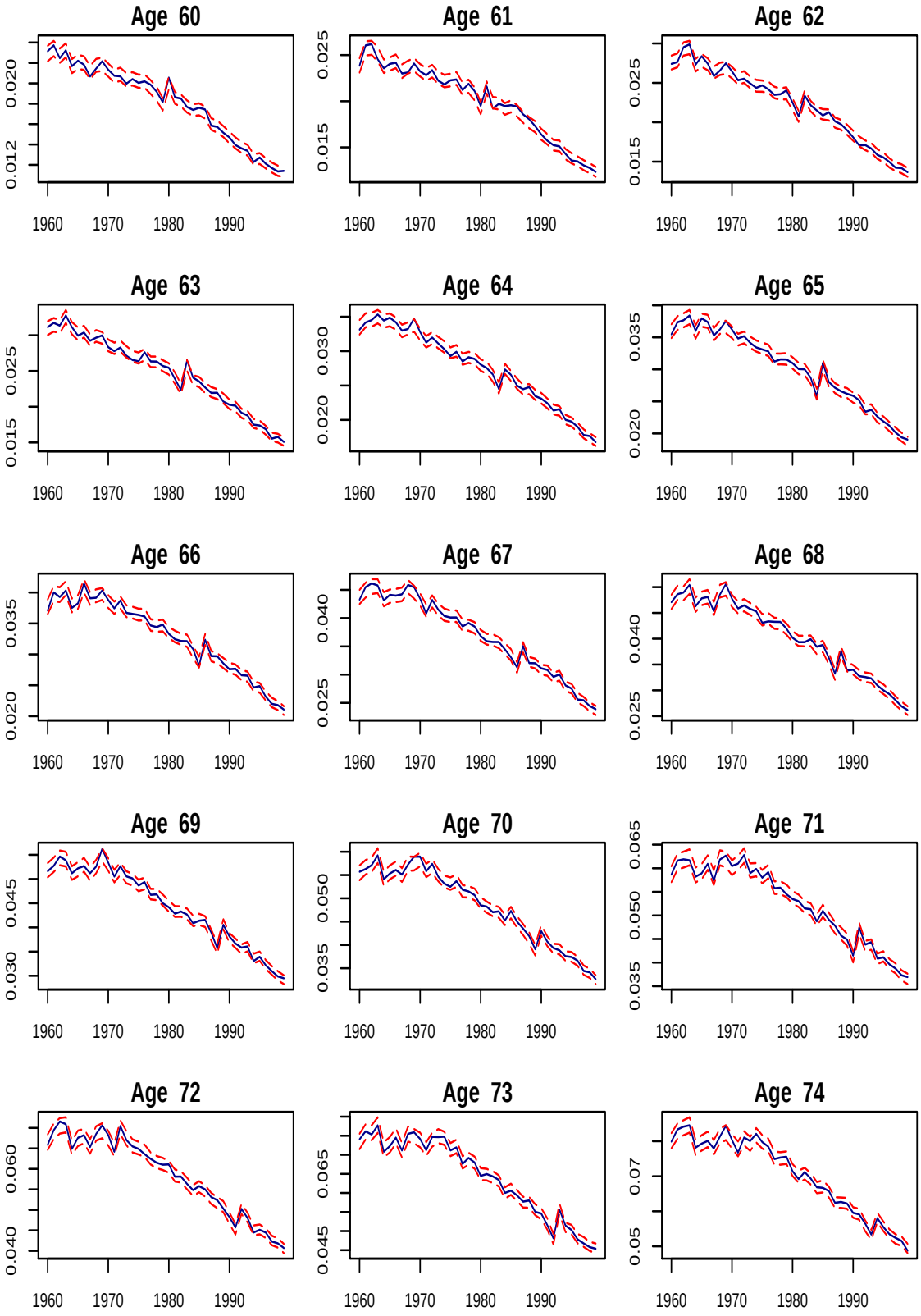


Figure 5.11: Panel of plots for male ages between 60 and 74, showing the trend in the observed age-specific male mortality rates over the years 1960 to 1999 indicated by the blue line and the corresponding red dashed lines providing 95% posterior predictive intervals from Model 6(b). The y-axis in each plot has a different scale of mortality rate.

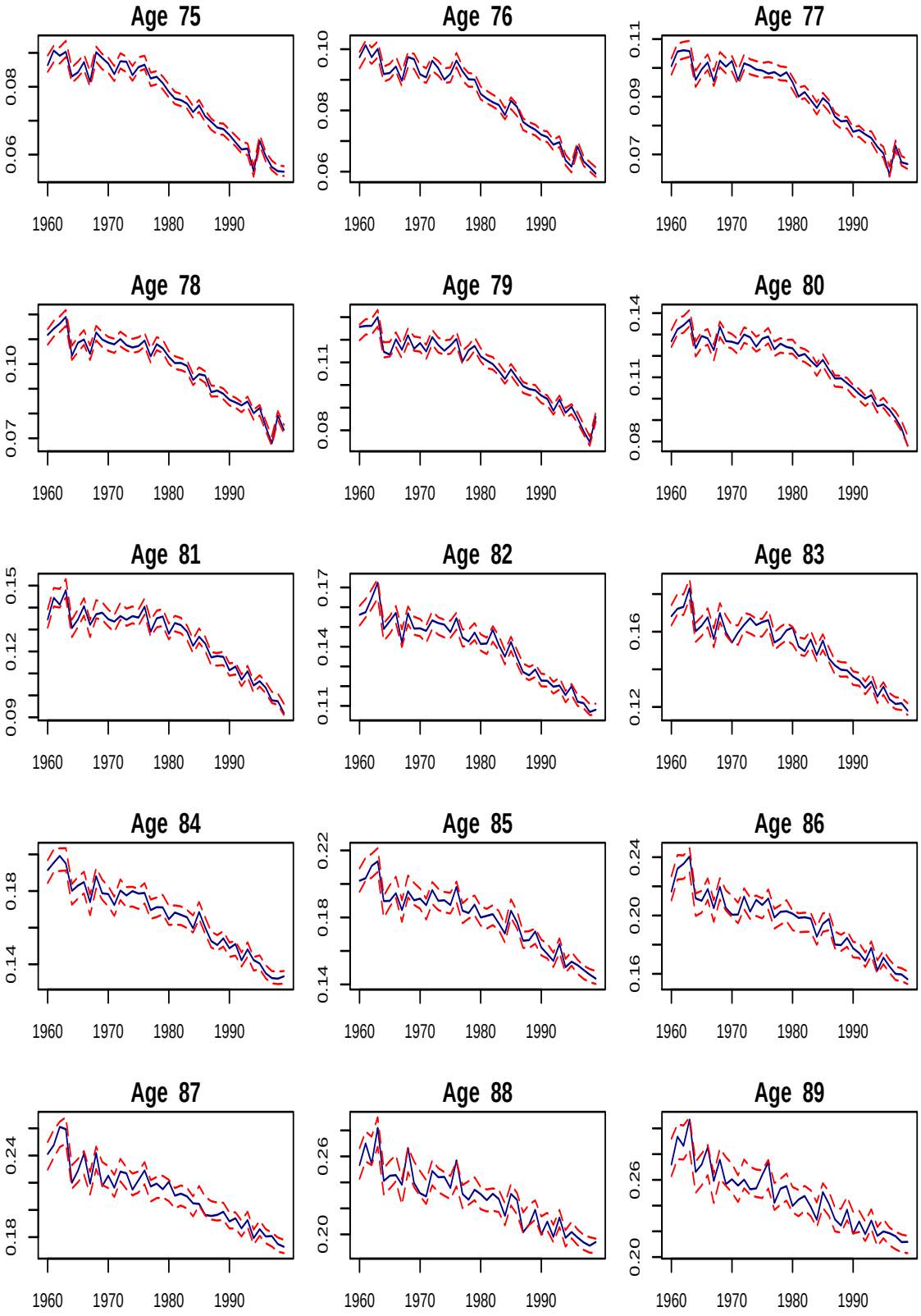


Figure 5.12: Panel of plots for male ages between 75 and 89, showing the trend in the observed age-specific male mortality rates over the years 1960 to 1999 and the corresponding 95% posterior predictive intervals from Model 6(b). See legend to Fig. 5.11 for explanation.

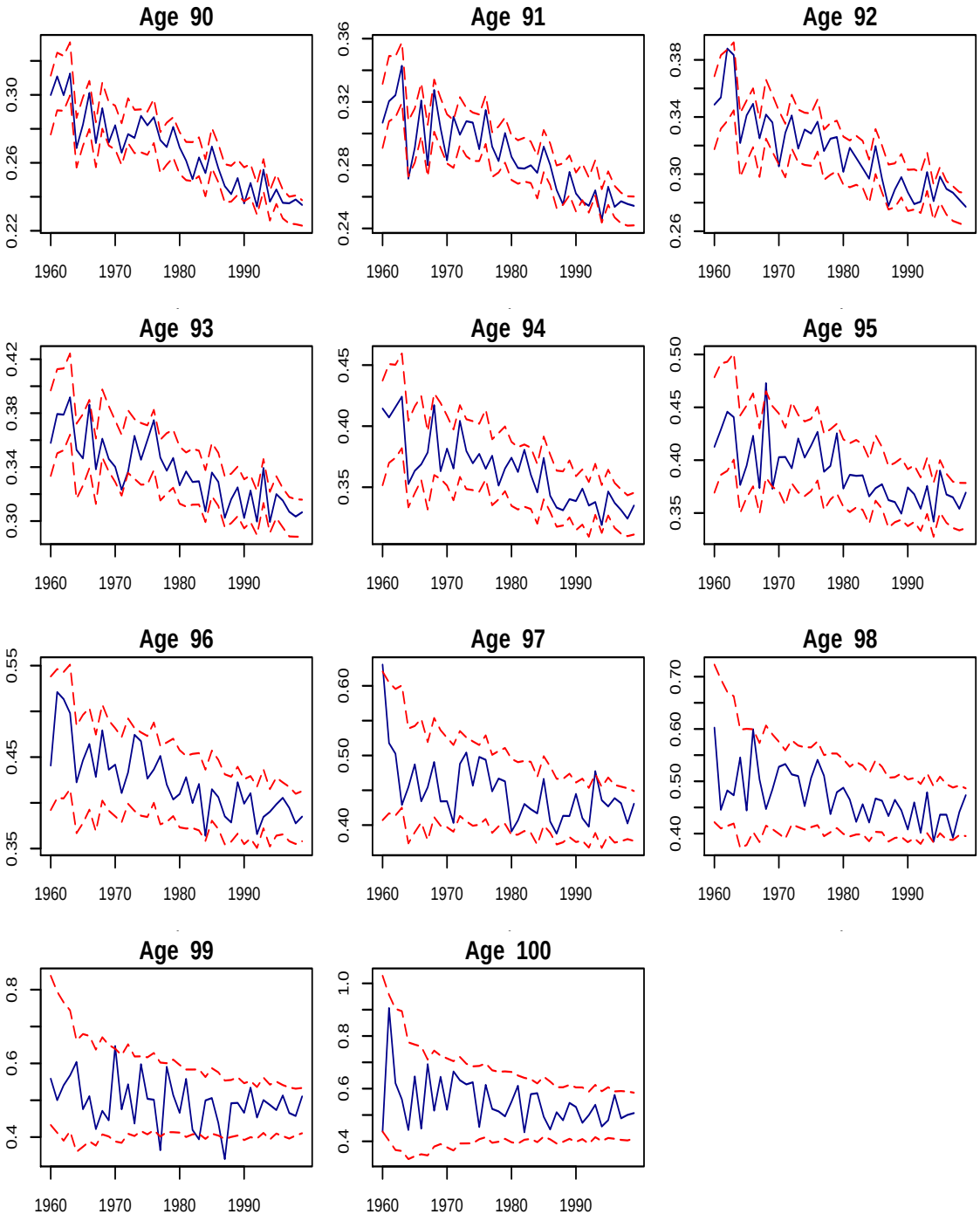


Figure 5.13: Panel of plots for male ages between 90 and 100, showing the trend in the observed age-specific male mortality rates over the years 1960 to 1999 and the corresponding 95% posterior predictive intervals from Model 6(b). See legend to Fig. 5.11 for explanation.

applied to the parameters α, β, δ, κ and γ, as described in chapter 4. For Model 6(a) each parameter follows a random walk on the levels of the parameter in question; for Model 6(b) each parameter follows an autoregressive process of order 1 on the first differences of the parameter in question and for Model 6(c) each parameter follows a random walk on the first differences of the parameter in question. Table 5.6 shows some key results for each of these models; the figures in the column headed posterior predictive interval show the proportion of empirical age-specific male mortality rates covered by the respective posterior predictive interval for each model. The proportion of empirical mortality rates covered by each interval is similar across the different models. The variance of the standardised residuals, as described earlier, is similar for all three smoothing priors. The p-value assumes that the pattern of residuals from the model shows no autocorrelation. The DIC figures are of similar magnitudes but show that Model 6(b) is the best. Similar conclusion is obtained if Model 6 is replaced by one of the other models.

	Posterior Predictive Interval			Variance of Standardised Residuals	Join Count p-value	DIC
	80% level Coverage	95% level Coverage	99% level Coverage			
Model 6(a)	0.93	0.99	1.00	0.64	<0.0001	18,645
Model 6(b)	0.89	1.00	1.00	0.64	<0.0001	18,607
Model 6(c)	0.93	0.99	1.00	0.63	<0.0001	18,609

Table 5.6: Table of various statistics for the fit of Models 6(a), 6(b) and 6(c).

5.3 Model Forecasting - Male Data

From the models considered, Models 5, 6 and 7 based on DIC fit the data best. Although forecasts were done using the better-fitting models based on DIC, it is important to point out that that a better fitting model does not automatically imply that the forecasts will be better than a slightly poorer fitting model. This section considers how well these models forecast age-specific male mortality using out-of-sample for the years 2000 to 2006 and ages 60 to 100. The forecasting of future mortality rates is achieved by forecasting each time-related parameter into the future. Below is a description of the method using Model 6(b) as an example. A complete description of Model 6(b) and all the parameters in the model are given in Appendix B. Forecasting period and cohort effects independently is not without some contro-

versy; see Clayton and Schifflers (1987b), Goldstein (1984) and Currie (2012) but the method described here follows Besag et al. (1995), Knorr-Held and Rainer (2001), Schmid and Held (2007) and supported by Holford (1991).

We denote the i^{th} draws from the posterior for the model parameters by $\mu^{(i)}$, $\alpha_x^{(i)}$, $\beta_x^{(i)}$, $\delta_x^{(i)}$, $\kappa_t^{(i)}$, $\gamma_{t-x}^{(i)}$, for the precision parameters by $\lambda^{(i)}$, $\nu^{(i)}$, $\zeta^{(i)}$ and for the autoregressive coefficients by $\rho^{(i)}$ and $\phi^{(i)}$.

The forecast $m_{x,t+n_y}^{(i)}$ of future mortality for age x and year $t+n_y$, i.e. t years out of sample, is calculated by transforming the logit $m_{x,n_y+t}^{(i)}$ which for future year t is

$$\begin{aligned} \xi_{x,n_y+t}^{(i)} &= \mu^{(i)} + \alpha_x^{(i)} + \beta_x^{(i)}\kappa_{n_y+t}^{(i)} + \delta_x^{(i)}\gamma_{n_a+n_y+t-x}^{(i)} + z_{x,n_y+t}^{(i)}, \\ m_{x,n_y+t}^{(i)} &= \exp(\xi_{x,n_y+t}^{(i)})/(1+\exp(\xi_{x,n_y+t}^{(i)})). \end{aligned}$$

where

$$\kappa_{n_y+t}^{(i)} \text{ is a draw from } N((1+\rho^{(i)})\kappa_{n_y+t-1}^{(i)} - \rho^{(i)}\kappa_{n_y+t-2}^{(i)}, 1/\lambda^{(i)}),$$

$$\gamma_{n_c+t}^{(i)} \text{ is a draw from } N((1+\phi^{(i)})\gamma_{n_c+t-1}^{(i)} - \phi^{(i)}\gamma_{n_c+t-2}^{(i)}, 1/\nu^{(i)}) \text{ and}$$

$$z_{x,n_y+t}^{(i)} \text{ is a draw from } N(0, 1/\zeta^{(i)}).$$

Next, generate the posterior predictive number of deaths in year n_y+t using the known exposure E_{x,n_y+t}, from which predictive mortality rates $\hat{m}_{x,n_y+t}^{(i)}$ can be obtained as described in section 5.2.6. Finally, posterior predictive intervals can be created using the $\hat{m}_{x,n_y+t}^{(i)}$ values as follows,

$$d_{x,n_y+t}^{(i)} \text{ is a draw from } Poi(E_{x,n_y+t}m_{x,n_y+t}^{(i)}),$$

$$\hat{m}_{x,n_y+t}^{(i)} = d_{x,n_y+t}^{(i)}/E_{x,n_y+t}.$$

For predictions beyond the time when the exposures are known, as we are interested only in the predicted mortality risks, a dummy exposure of 10,000 is used. The use of a fixed value simplifies the calculations but affects the variance of the forecasts. An improved method is to forecast the future exposures. For example, if we start

with a known population age x in year T we can forecast the number of deaths from this population age in the following year and so obtain a forecast of the population aged $x+1$ in year $T+1$. This process can be followed recursively for population counts in future years, see Richards et al. (2013).

Figures 5.14, 5.15 and 5.16 show for each age, 60 to 100, a comparison of the forecast of the age-specific male mortality rates from Models 6(a), 6(b) and 6(c) with out-of-sample data for male mortality between 2000 and 2006. Also plotted are 95% posterior predictive intervals associated with the 3 different forms of smoothing priors. These figures demonstrate that the choice of smoothing prior has a very material impact on forecasts as seen on the width of the posterior predictive intervals.

Figures 5.14, 5.15 and 5.16 show that prior smoothing corresponding to a random walk on the first differences results in posterior predictive intervals that are very wide. The reason why these intervals are so wide is that the parameter forecasts retain the curvature within the posterior parameter draws. So when one simulation produces a path where the parameters are reducing the forecasts of this path continue this trend into the future so producing increasingly lower mortality rates in the future. There will be other simulations where the parameters are increasing and the forecasts will continue this trend into the future so producing very high mortality rates in the future." The other two forms of prior smoothing give narrower 95% posterior predictive intervals that are conservative and provide modelled rates that seem reasonable as short and medium term estimates to the observed rates at most ages.

Appendix D contains forecasts of the age-specific male mortality for Models 5(a)-5(c) and 7(a)-7(c) similar to the plots for Models 6(a)-6(c) in Figures 5.14, 5.15 and 5.16. Forecast age-specific male mortality rates from Models 7(a), 7(b) and 7(c) follow the procedure above but do not include the random effect $z_{x,t}$, but instead $w_{x,t}$, the parameter estimating the error in measured exposures. For the comparison with the observed age-specific male mortality rates, 95% posterior predictive intervals have been calculated using $\hat{m}^{(i)}_{n_y+t}$ as follows;

$$w^{(i)}_{x,n_y+t} \text{ is a draw from } Beta(1, s^{(i)}),$$

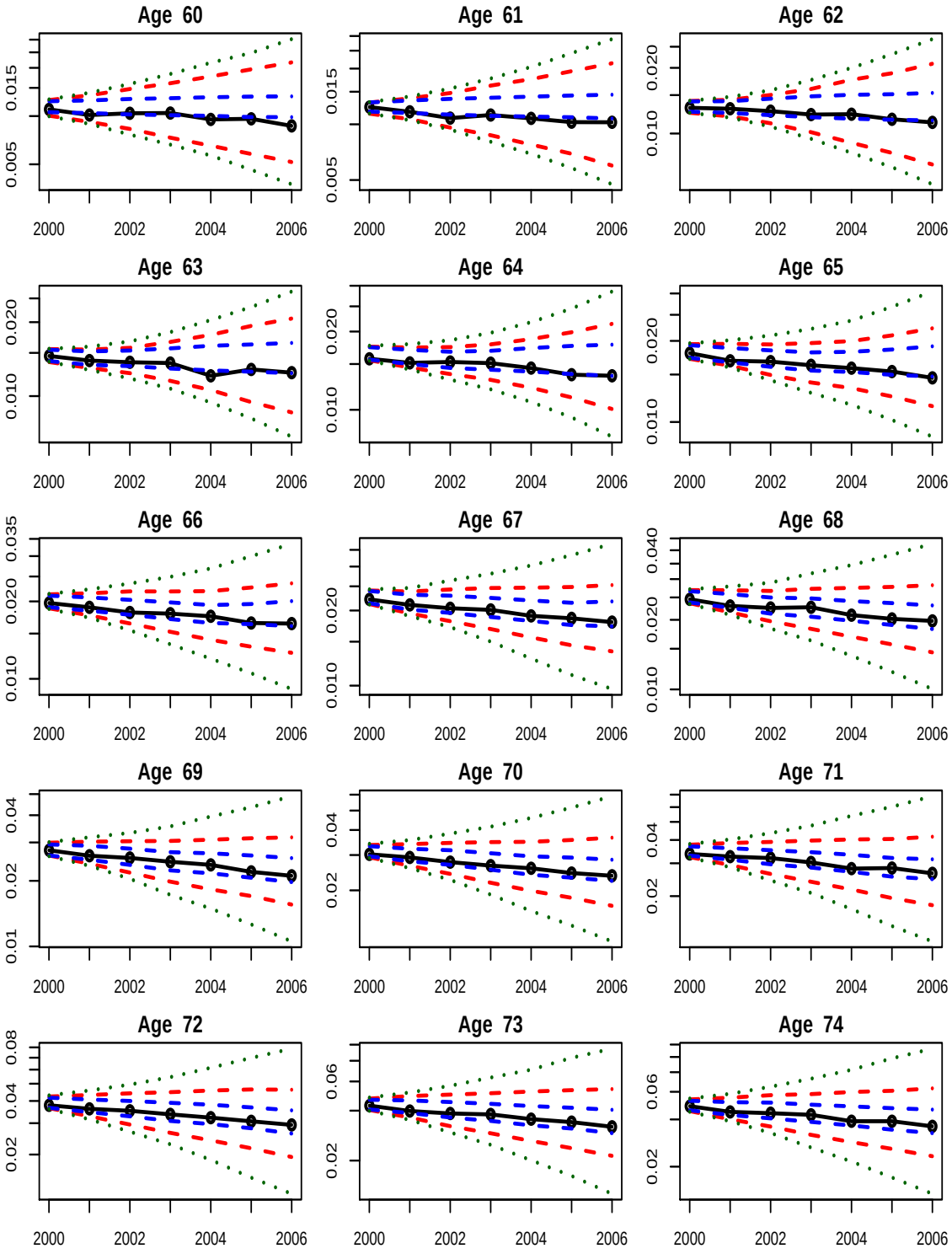


Figure 5.14: Plots of out-of-sample empirical male mortality rates for individual ages 60 to 74, for the years 2000 to 2006 and the corresponding forecasts 95% posterior predictive intervals from Models 6(a), 6(b) and 6(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithm scale. Each plots has its own y-axis scale.

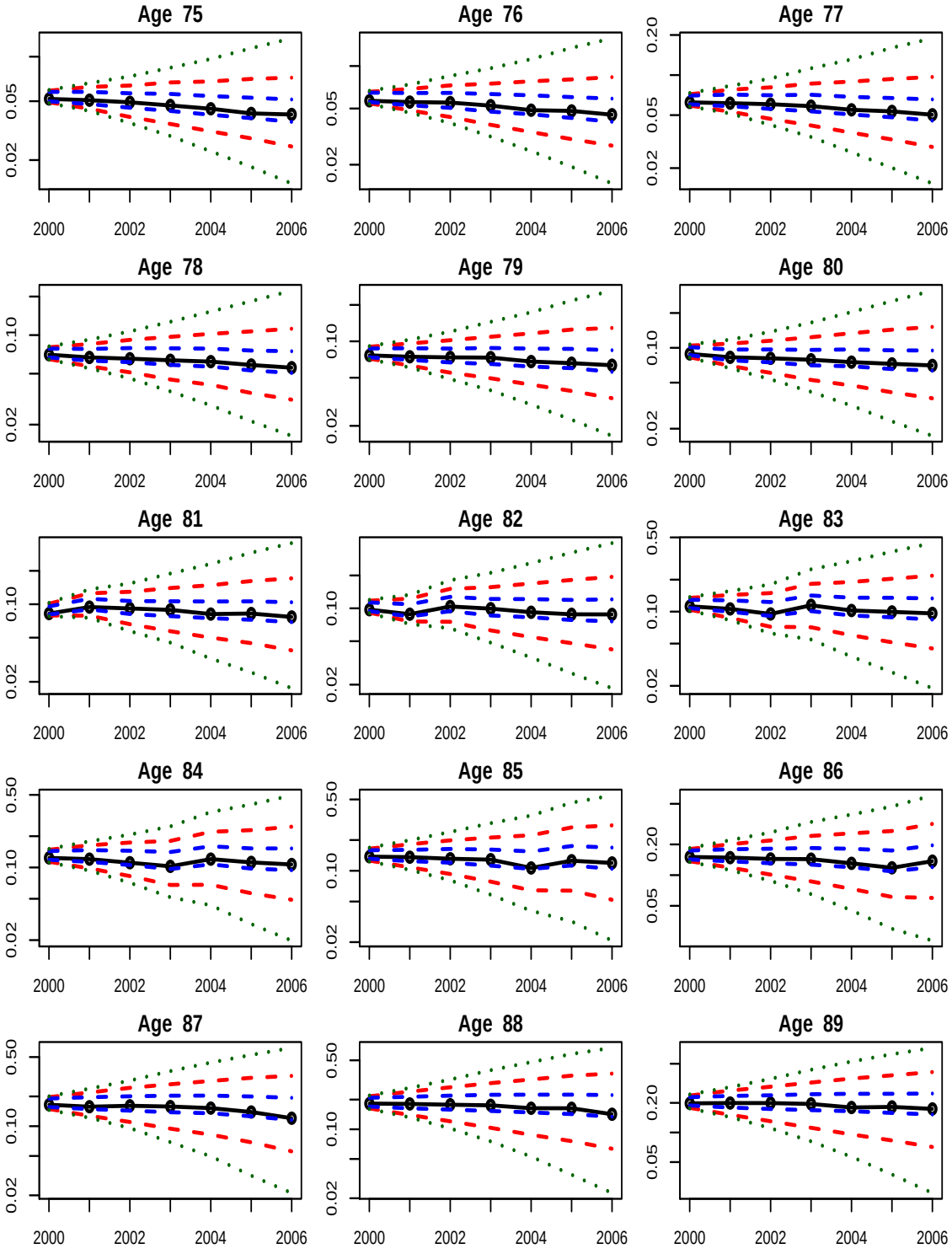


Figure 5.15: Plots of out-of-sample empirical male mortality rates for individual ages 75 to 89, for the years 2000 to 2006 and the corresponding forecasts 95% posterior predictive intervals from Models 6(a), 6(b) and 6(c). See legend to Fig. 5.14 for explanation.

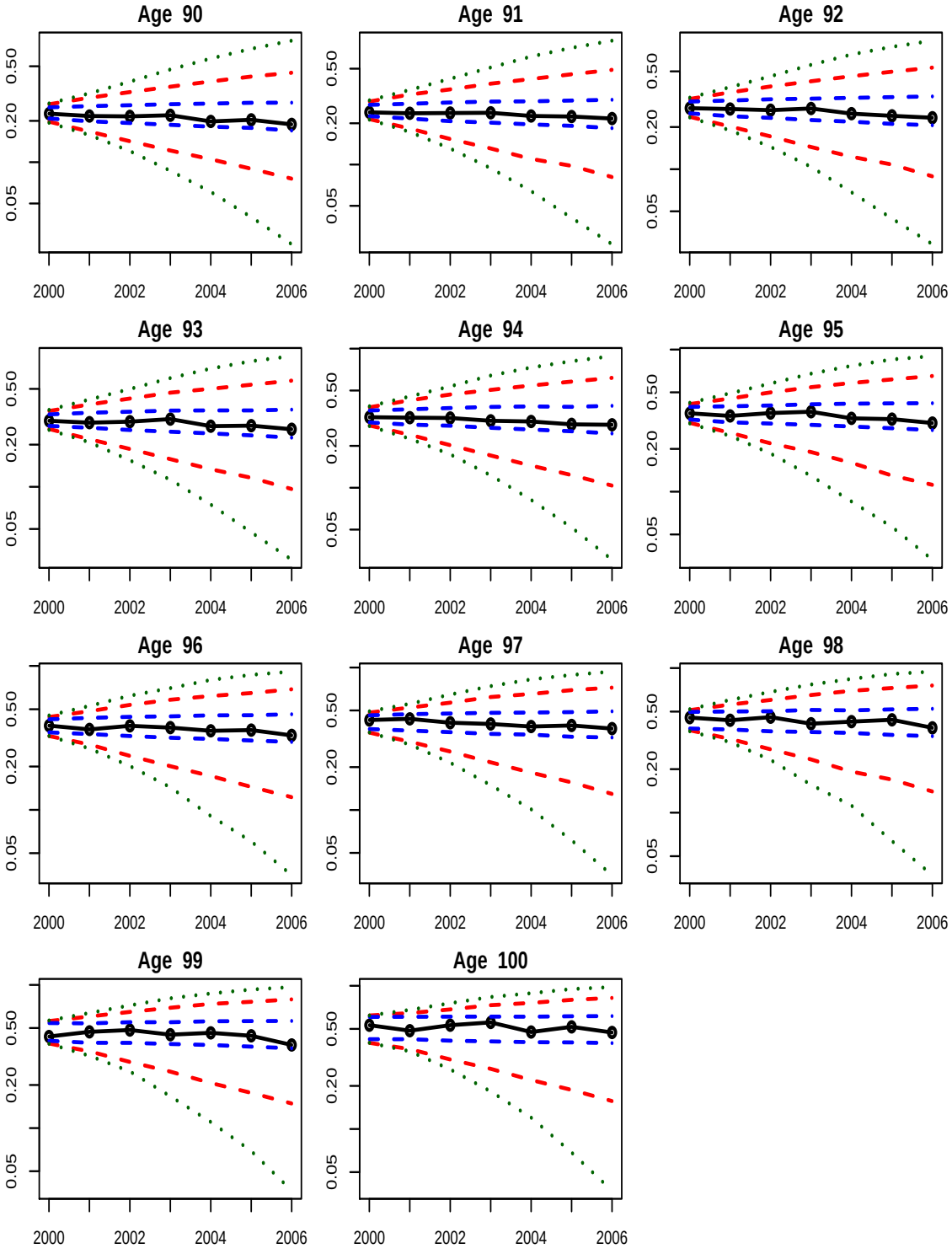


Figure 5.16: Plots of out-of-sample empirical male mortality rates for individual ages 90 to 100, for the years 2000 to 2006 and the corresponding forecasts 95% posterior predictive intervals from Models 6(a), 6(b) and 6(c). See legend to Fig. 5.14 for explanation.

$$d^{(i)}_{n_y+t} \text{ is a draw from } Poi(E_{x,n_y+t}/(1-w^{(i)}_{x,n_y+t})m^{(i)}_{x,n_y+t}),$$

$$\hat{m}^{(i)}_{n_y+t} = d^{(i)}_{n_y+t}/(E_{x,n_y+t}/(1-w^{(i)}_{x,n_y+t})).$$

A summary of the proportion of empirical mortality rates covered by posterior predictive intervals at selected levels for each model is shown in Table 5.7. Models 5(a) and 6(a) give the appropriate coverage of the empirical mortality rates. Models 5(b) and 6(b) give slightly conservative coverage levels. Model 7(a) and 7(b) give more conservative intervals than the corresponding models 5 and 6. Finally, for Models 5(c), 6(c) and 7(c) the posterior predictive intervals at the different levels shown are very conservative.

	Model 5			Model 6			Model 7		
PPI level	80%	95%	99%	80%	95%	99%	80%	95%	99%
Model (a)	0.78	0.95	0.99	0.83	0.96	0.99	0.95	1.00	1.00
Model (b)	0.85	0.98	1.00	0.89	0.99	1.00	0.94	1.00	1.00
Model (c)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 5.7: The proportion of the out-of-sample observed age-specific male mortality rates $\hat{m}_{x,t}$ falling within the posterior predictive intervals at selected levels.

5.4 England and Wales female mortality

5.4.1 Unsmoothed Parameters

Parallel to section 5.4.1, it is instructive to consider the parameters α, β, κ, δ and γ for females without any smoothing. Box plots of the posterior parameter draws obtained from Models 1, 2 and 4 are shown in Figures 5.17, 5.18 and 5.19, respectively:

As with the male data, Figures 5.17, 5.18 and 5.19 show that even without smoothing the parameters display a relatively smooth pattern. In all three models the age effect parameter α increases with age, a declining trend in the period parameter κ indicating that like male mortality female mortality is also declining. Similar comments made regarding males for the cohort parameter γ and age parameters β and κ apply also to the female parameters.

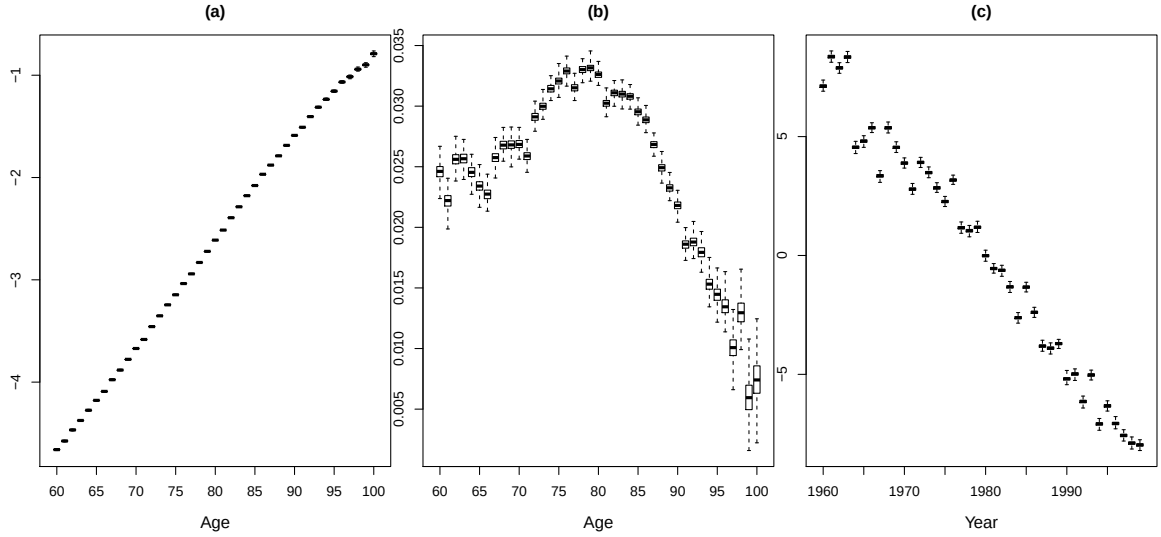


Figure 5.17: Boxplots of parameter estimates from Model 1 using female data with $\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t$: (a) is $\mu + \alpha$ for each age, Fig, (b) is β for each age and Fig. (c) is κ for each calendar year.

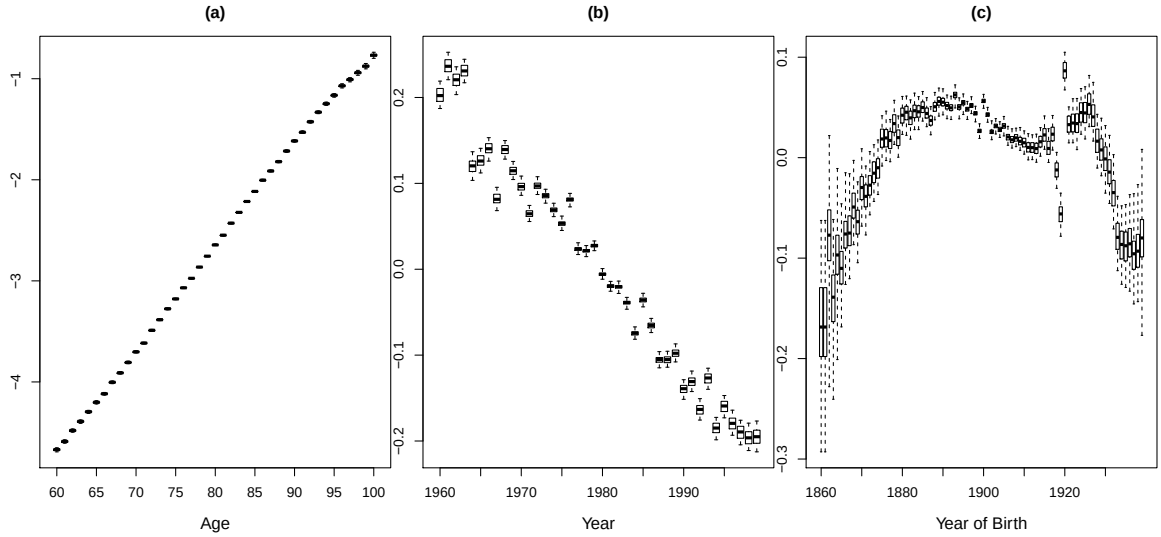


Figure 5.18: Boxplots of parameter estimates from Model 2 using female data with $\log m_{x,t} = \mu + \alpha_x + \kappa_t + \gamma_{t-x}$. (a) is $\mu + \alpha$ for each age, (b) is κ for each calendar year and (c) is γ for each cohort.

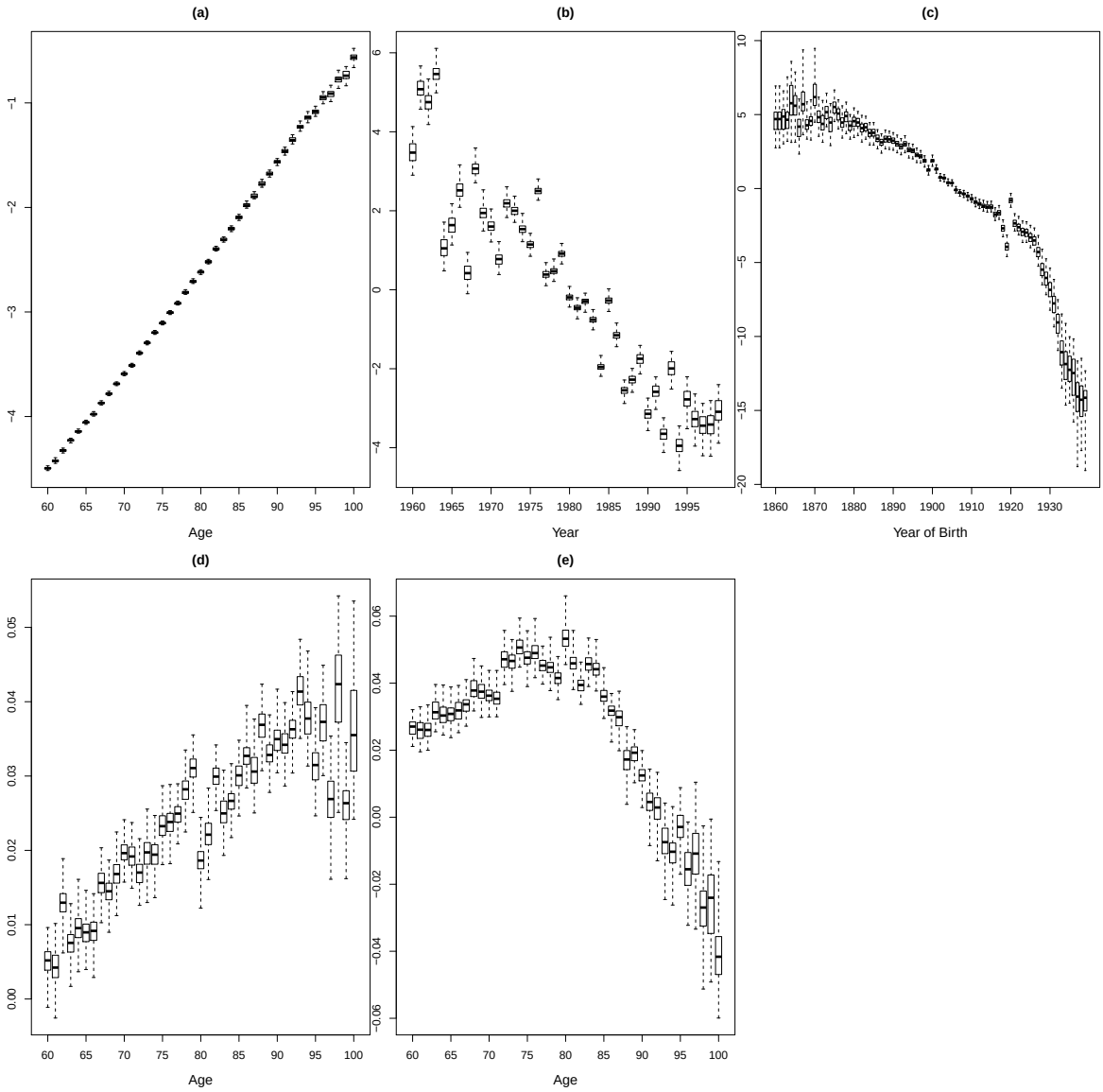


Figure 5.19: Boxplots of parameter estimates from Model 4 using female data with $\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x}$. (a) is $\mu + \alpha$ for each age, (b) is κ for each calendar year, (c) is γ for each cohort, (d) is β for each age and (e) is δ for each age.

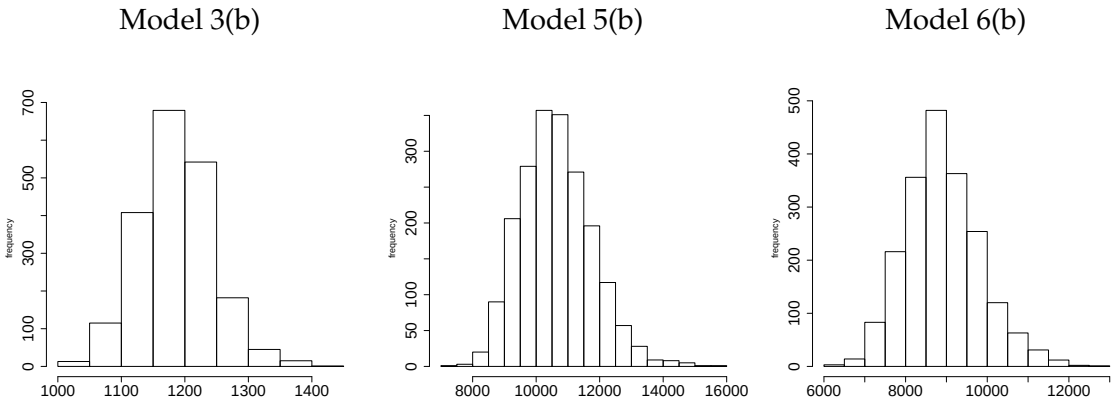


Figure 5.20: Histograms of 2000 samples from the posterior distributions of the random effect precision parameter of Models 3(b), 5(b) and 6(b) for female data.

5.4.2 Random Effects

Figure 5.20 shows the distribution of the random effects precision parameter for Models 3(b), 5(b) and 6(b) using female data. As with the male data, the distribution of the precision parameter is similar for the other versions of the model with different smoothing priors. The results are similar to those for male data, with Model 3(b) having a much smaller precision than Model 5(b) and Model 6(b). However, the precision parameter for Model 3(b) is much smaller than for the corresponding model with male data. This indicates that the random effects incorporated into Model 3(b) are much greater than Models 5(b) and 6(b).

5.4.3 Autoregressive smoothing coefficient

Figures 5.21 and 5.22 show the posterior distributions of the autoregressive coefficient for the age, period and cohort parameters for each of the models that include the autoregressive smoothing priors. Table 5.8 shows 95% credible intervals for these autoregressive coefficients. As for the male data, the autoregressive coefficient for the age parameter is very close to 1 for all models. For the period autoregressive coefficients, the 95% central posterior intervals for all models, except Models 2(b), indicate a negative coefficient. Therefore, the difference in neighbouring period parameters will reduce to around zero with the difference in parameters alternating between positive and negative differences. For Model 2(b), the 95% central posterior interval straddles zero. For the cohort autoregressive coefficients, Models 2(b),

	coeff. of α		coeff. of κ		coeff. of γ	
Model	lower	upper	lower	upper	lower	upper
1(b)	0.946	1.000	-0.623	-0.009		
2(b)	0.919	1.000	-0.382	0.215	0.343	0.727
3(b)	0.980	1.000	-0.724	-0.121	-0.551	0.026
4(b)	0.946	1.000	-0.748	-0.205	0.159	0.603
5(b)	0.970	1.000	-0.783	-0.215	-0.278	0.247
6(b)	0.967	1.000	-0.774	-0.210	-0.270	0.278
7(b)	0.922	1.000	-0.672	-0.126	0.305	0.709

Table 5.8: Table shows 95% posterior intervals for the autoregressive coefficient of the smoothing prior for α, κ and γ for each model using female data. For alpha, given the asymmetric shape of the posterior sample, it seemed more appropriate to leave all the 5% to the left of the interval, rather than using a 95% central posterior intervals.

4(b) and 7(b), the 95% central posterior intervals indicate that the coefficient is positive. For the other models the 95% central posterior interval straddles zero.

5.4.4 Fitted Models - Female Data

Analysis of Standardised Residuals

Based on female data, Table 5.9 shows the variance of the standardised residuals from each model. Similarly to the male data, Model 1 gives a poor fit to the female data. However, adding the cohort parameters and dropping the age-related β parameters, i.e. comparing from 1(b) to 2(b), results in an increase in variance, suggesting that the age-related interaction parameters β are a very important model feature for females more so than the cohort parameters. Models 3(b), 5(b), 6(b) and 7(b) include random effect type parameters that result in the standardised residuals being under-dispersed indicating a degree of over-fitting. For the other models that do not include random effect parameters, 1(b), 2(b) and 4(b), the standardised residuals are over-dispersed. As with the male data, Model 3 has the lowest variance of standardised residuals, even lower than the corresponding male figure. This is largely due to the very much greater random effects allowed for within this model with the female data.

Model	1(b)	2(b)	3(b)	4(b)	5(b)	6(b)	7(b)
	4.17	4.73	0.25	1.41	0.61	0.63	0.73

Table 5.9: Variance of standardised residuals for female data

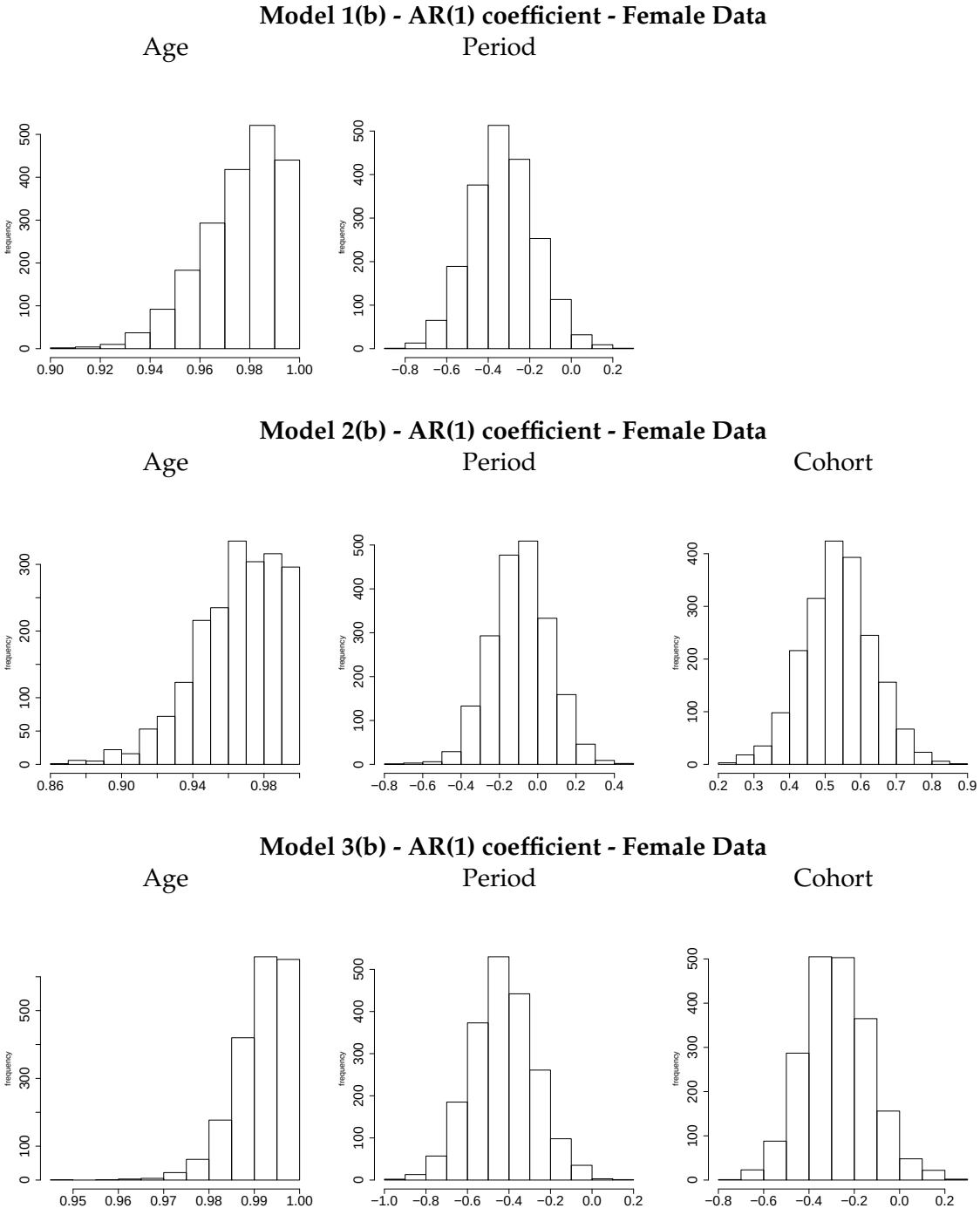


Figure 5.21: Histograms of 2000 samples from the posterior distributions of the coefficient of the autoregressive series for the age, period and cohort parameters for Models 1(b), 2(b) and 3(b) for female data.

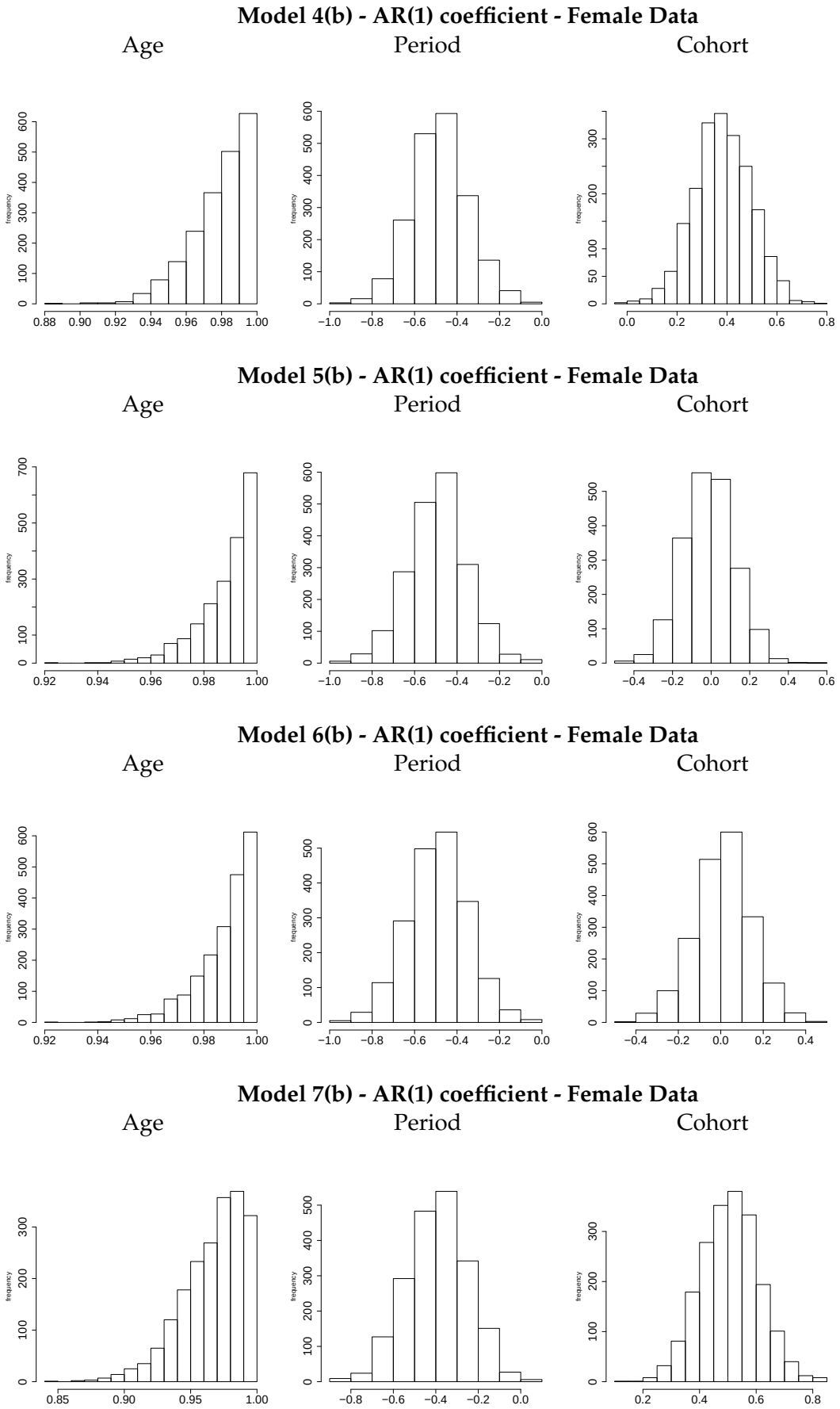


Figure 5.22: Histograms of 2000 samples from the posterior distributions of the coefficient of the autoregressive series for the age, period and cohort parameters for Models 4(b), 5(b), 6(b) and 7(b) for female data.

Join Count Statistic

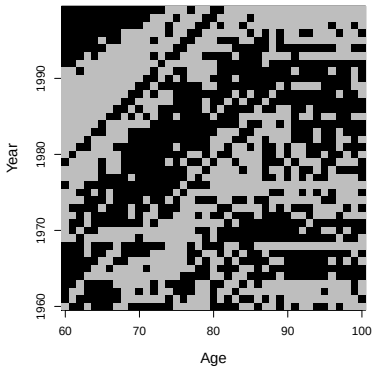
Figure 5.23 is a panel of plots of the sign of the median of the standardised residuals $r_{x,t}$ at each age and year data point for each model with prior smoothing following an autoregressive process of order 1 on the first differences of the parameters. Model 1 (panel 1 of Figure 5.23) shows diagonal blocks of the same colour, indicating the presence of cohort effects. This is not surprising as the model does not include any cohort parameters. The plots of the residuals for the other models do not display any obvious cohort effect. The patterns of residuals for Models 2(b) and 3(b) look to have large regions where the medians of modelled mortality consistently overestimate or underestimate the observed female age-specific rates. Hence, the inclusion of age-related period mortality improvements is an important feature. As with the male data the patterns of residuals do not appear to be random but again the plots for Models 4(b), 5(b) and 6(b) look to have less structure than those of the other models. The residuals for Model 7(b) show a pattern similar to those of Models 4(b), 5(b) and 6(b) as these models have a similar parameter structure.

Applying the method described in Cliff and Ord (1981) to the pattern of residuals displayed in Fig. 5.23 a p-value assuming that there is no autocorrelation can be calculated. Table 5.10 shows the modified Chi-Squared p-value. The number of degrees of freedom is 4.

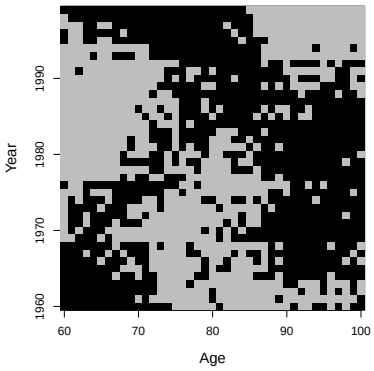
Model	1(b)	2(b)	3(b)	4(b)	5(b)	6(b)	7(b)
Chi-squared	<0.0001	<0.0001	<0.0001	<0.0001	0.0008	0.0008	<0.0001

Table 5.10: Table of p-values from modified chi-squared distribution assuming that the pattern of residuals displayed in Fig. 5.23 shows no spatial autocorrelation.

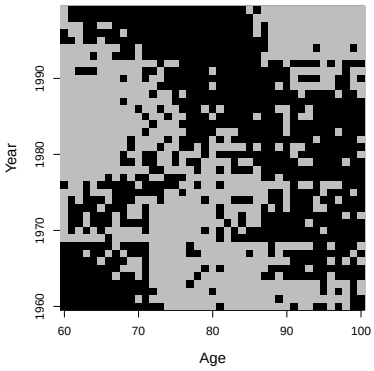
Given the very small p-values, the residuals of each model show clear evidence of autocorrelation, indicating that there is some remaining structure in the residuals that is not included within the models. However, as the model increases in complexity visually there is a substantial reduction in the structure of the residuals.



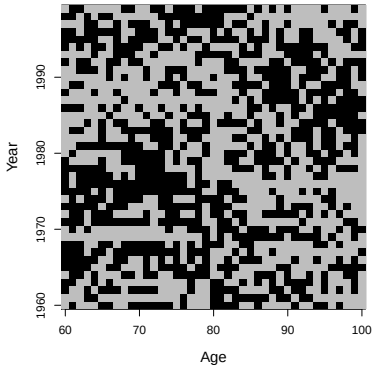
(1) Model 1(b)



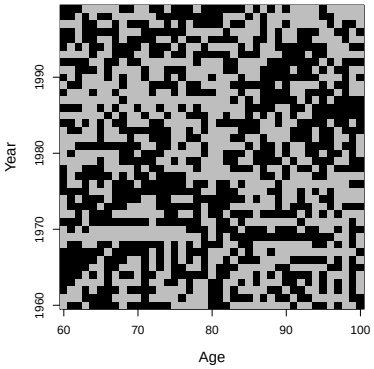
(2) Model 2(b)



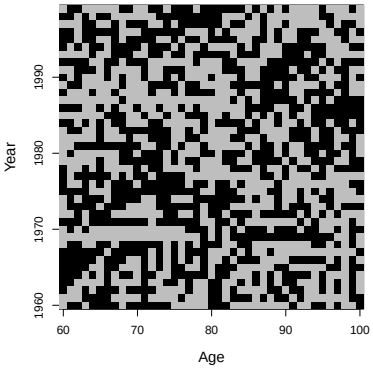
(3) Model 3(b)



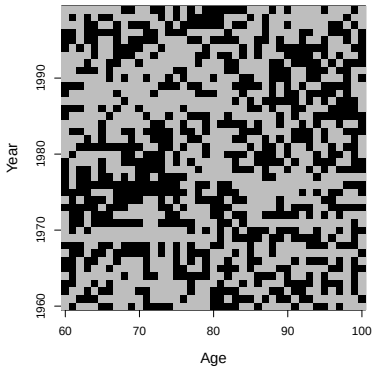
(4) Model 4(b)



(5) Model 5(b)



(6) Model 6(b)



(7) Model 7(b)

Figure 5.23: Panel of plots for each model based on female data showing the sign of the standardised residuals $r_{x,t}$ over the years 1960 to 1999 and ages 60 to 100. The black pixels indicate the points where the mean of the modelled deaths exceeds the observed number of deaths, and the grey pixels indicate where the mean of the modelled deaths is below the observed number of deaths.

5.4.5 Deviance Information Criterion

The Deviance Information Criterion (DIC) is calculated exactly as for the males (section 5.2.5). The resulting DIC for the models is shown in Table 5.11, which indicates Model 6(b) very closely followed by Model 5(b) as the best models of those considered. Of all the models analysed, Model 6(a) gave the lowest DIC of 19,209 with pD of 650. The effective number of parameters for Model 3 is very large. This is mainly due to the much greater random effects allowed for within this model.

Model	1(b)	2(b)	3(b)	4(b)	5(b)	6(b)	7(b)
DIC	23,904	24,932	19,886	19,576	19,217	19,213	19,324
pD	120	144	1293	186	665	650	623

Table 5.11: Deviance Information Criteria for Models 1(b) to 7(b) for females.

5.4.6 Posterior predictive intervals - Female Data

Figures 5.24, 5.25 and 5.26 show, for each age, from 60 to 100, the trend in observed female mortality rates over the years 1960 to 1999 indicated by the solid blue line and the dashed red lines are the boundary of 95% posterior predictive intervals according to Model 6(b). For the younger ages up to around age 85 the posterior predictive intervals at the 95% level are quite narrow. For the very highest ages, 95 and over, the posterior predictive intervals are much wider reflecting the higher variability of observed mortality rates at these high ages. These figures show that the empirical coverage of the posterior predictive interval at the 95% level is conservative covering almost all the observed rates. A summary of the empirical coverage achieved by the posterior predictive intervals at different probability levels for each model is shown in Table 5.12. The results show that the coverage of the empirical rates by predictive posterior intervals from Model 1(b) and Model 2(b) are too low. The posterior predictive intervals of Model 4(b) give the best coverages although slightly narrow followed by Models 5(b), 6(b) and 7(b) which provide conservative intervals.

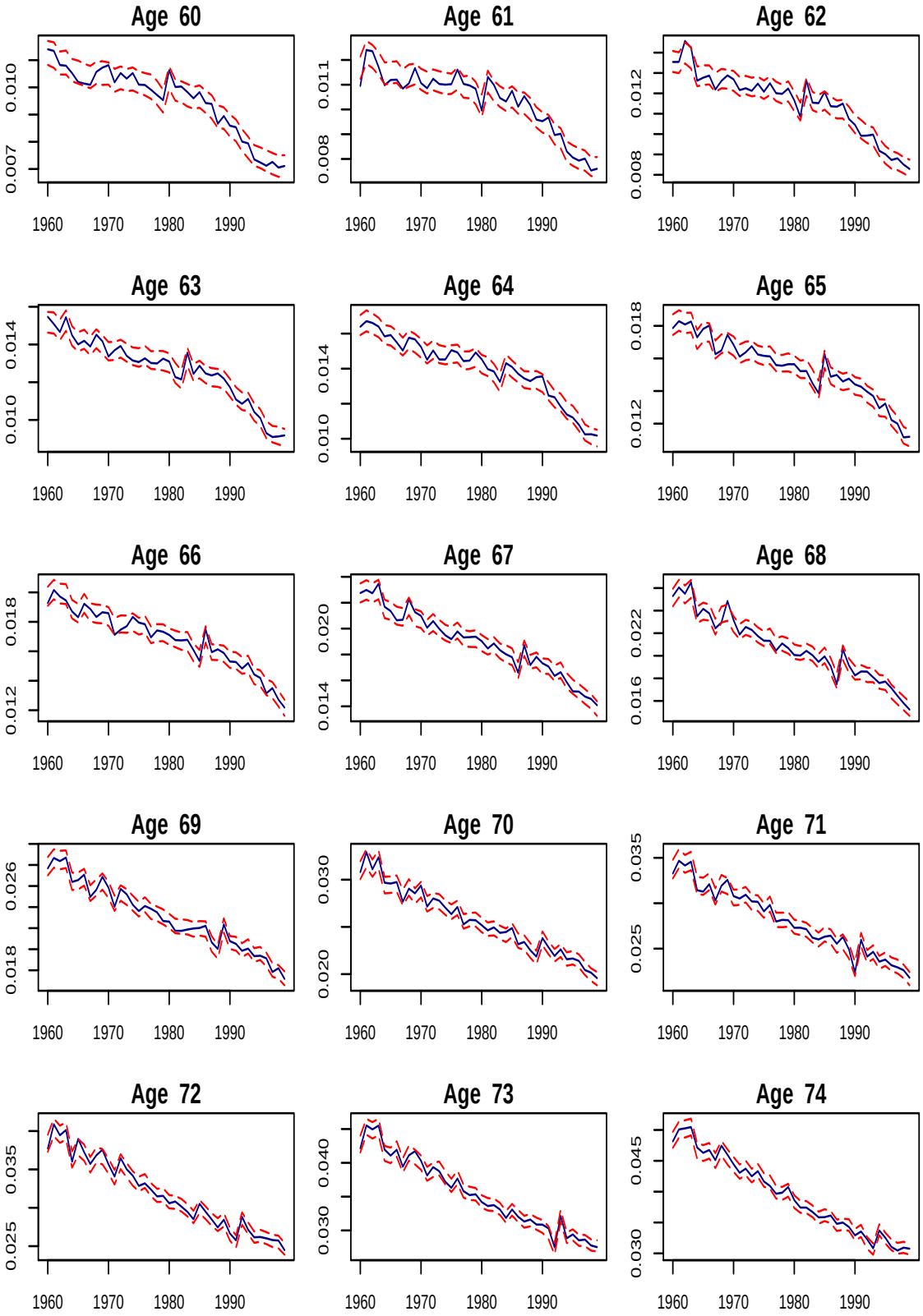


Figure 5.24: Panel of plots for female ages between 60 and 74, showing the trend in the observed age-specific female mortality rates over the years 1960 to 1999 indicated by the blue line and the corresponding red dashed lines providing 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Model 6(b). The y-axis in each plot has a different scale of mortality rate.

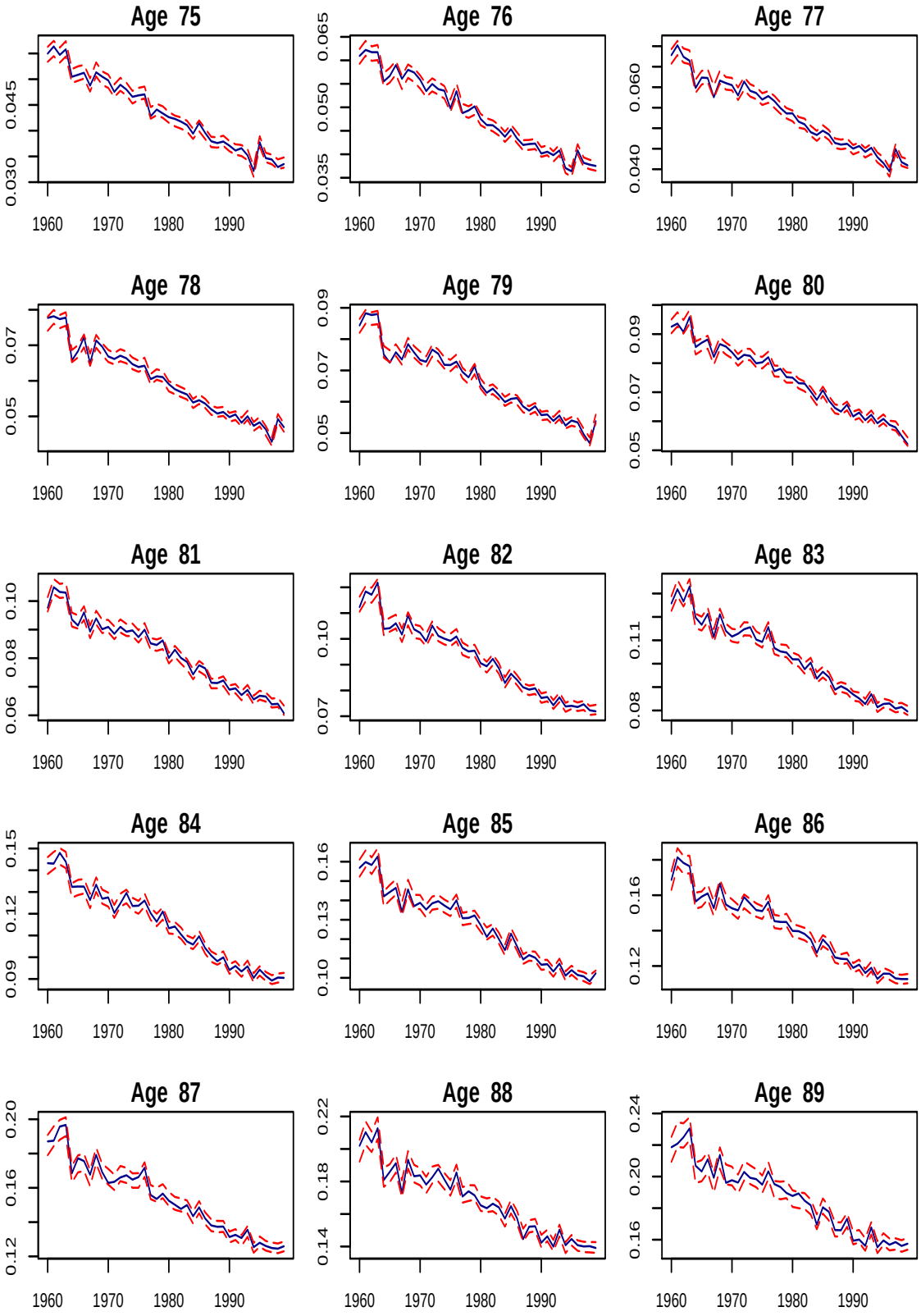


Figure 5.25: Panel of plots for female ages between 75 and 89, showing the trend in the observed age-specific male mortality rates over the years 1960 to 1999 and the corresponding 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Model 6(b). See legend to Fig. 5.24 for explanation.

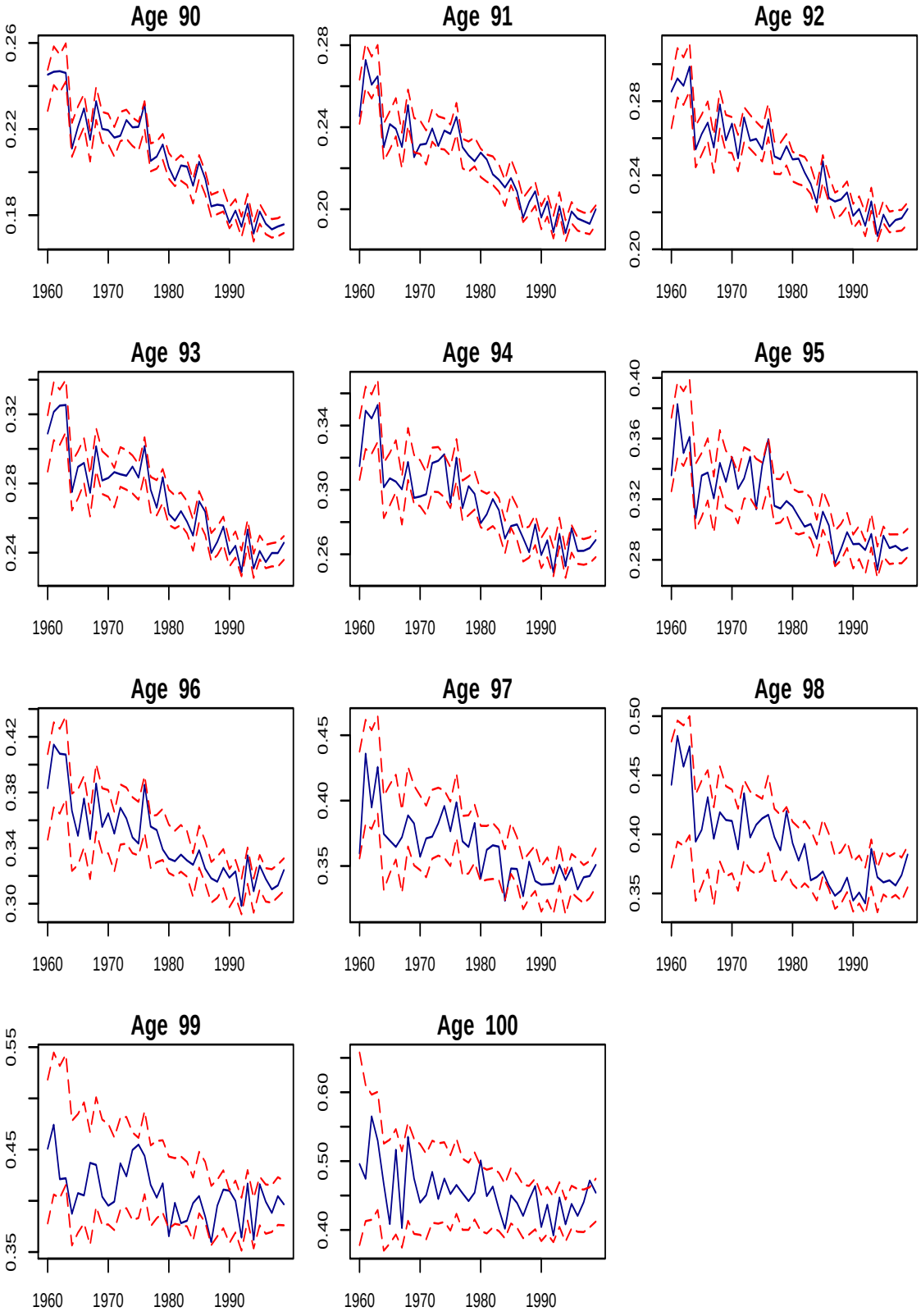


Figure 5.26: Panel of plots for female ages between 90 and 100, showing the trend in the observed age-specific male mortality rates over the years 1960 to 1999 and the corresponding 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Model 6(b). See legend to Fig. 5.24 for explanation.

CI level	80%	95%	99%
Model 1(b)	0.55	0.75	0.86
Model 2(b)	0.49	0.68	0.82
Model 3(b)	0.99	1.00	1.00
Model 4(b)	0.76	0.92	0.97
Model 5(b)	0.94	0.99	1.00
Model 6(b)	0.93	0.99	1.00
Model 7(b)	0.93	0.99	1.00

Table 5.12: Coverage of empirical female mortality rates for ages 60 to 100 and years 1960 to 1999 by posterior predictive intervals at different probability levels for each model.

5.5 Model Forecasting

From the models considered for female data, Models 5, 6 and 7 based on DIC fit the data best. The forecast method is outlined in section 5.3. Figures 5.27, 5.28 and 5.29 show for each age, 60 to 100, a comparison of the forecast of the age-specific female mortality rates from Models 6(a), 6(b) and 6(c) with out-of-sample data empirical female mortality between 2000 and 2006. Also plotted are 95% posterior predictive intervals associated with the three different forms of smoothing priors. As with the male data, these figures show that prior smoothing corresponding to a random walk on the first differences results in posterior predictive intervals that are very wide. Therefore this smoothing would provide poor forecasts over the near to longer term. The other two forms of prior smoothing give narrower 95% posterior predictive intervals that are conservative and provide modelled rates that seem reasonable as short and medium term estimates to the observed rates at most ages.

Appendix D contains the forecasts of the age-specific female mortality rates for Models 5(a)-5(c) and Models 7(a)-7(c) similar to the plots for Models 6(a)-6(c) and the conclusions are essentially the same for each of the models.

Table 5.13 shows a summary of the proportion of empirical mortality rates covered by posterior predictive intervals at selected levels for each model. For a given model and probability level, the figures in the table are the numbers of out-of-sample observed age-specific female mortality rates that fall within the credible region. Models 5(a), 6(a), 7(a) and 7(b) appear to give the appropriate coverage of the empirical mortality rates. Models 5(b) and 6(b) appear to give excessively narrow intervals and for Models 5(c), 6(c) and 7(c) the posterior predictive intervals at the different

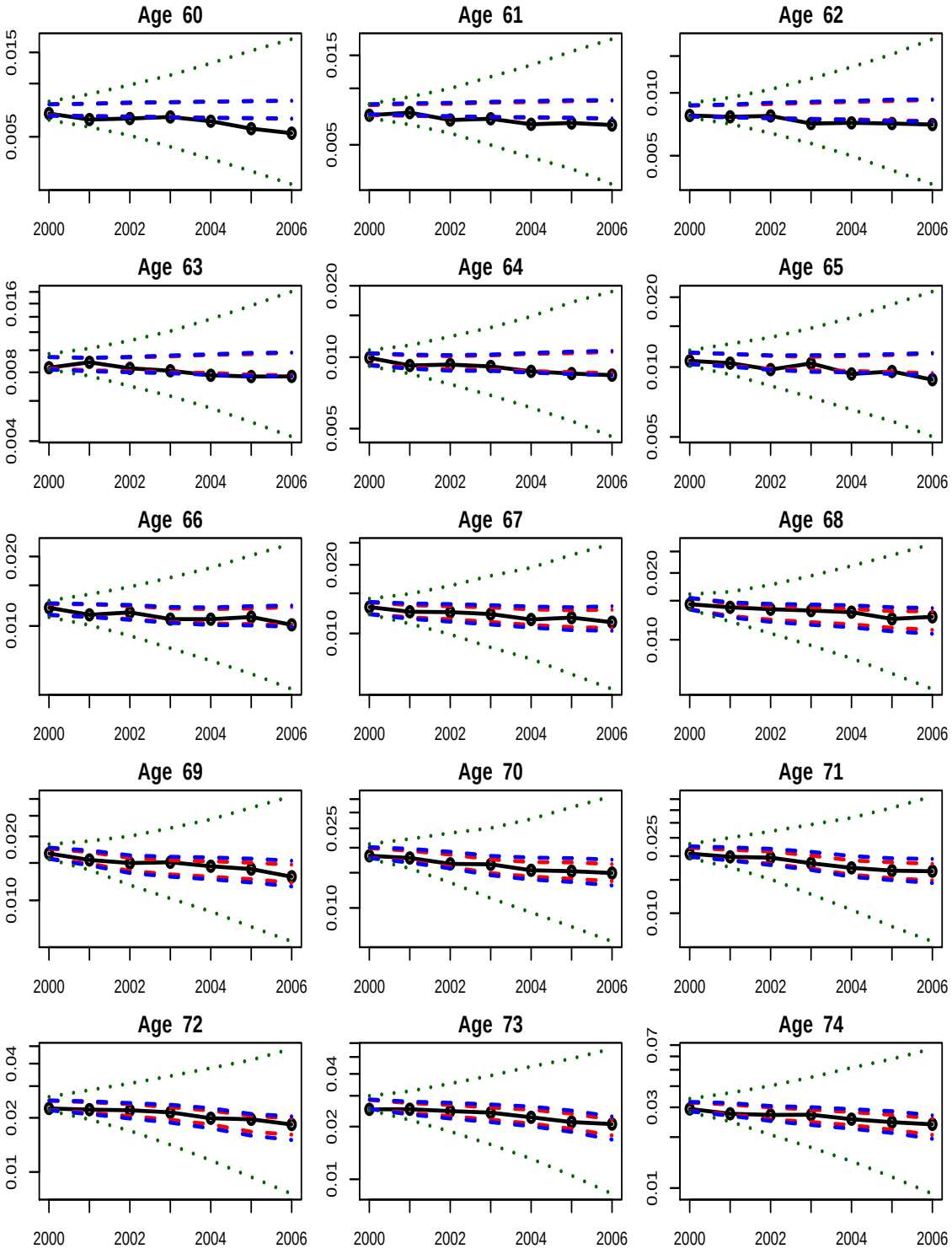


Figure 5.27: Plots of out-of-sample empirical female mortality rates for individual ages 60 to 74, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 6(a), 6(b) and 6(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

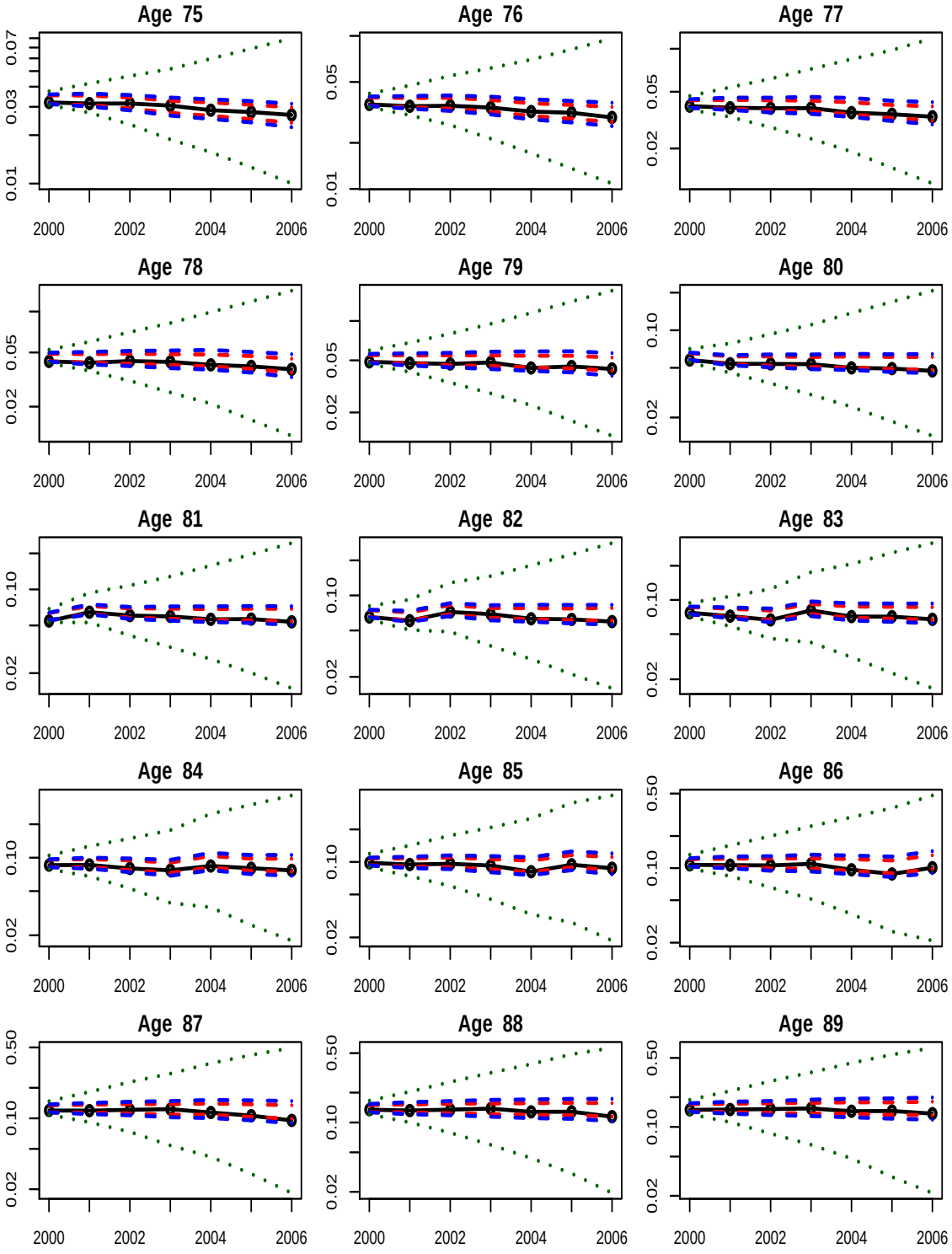


Figure 5.28: Plots of out-of-sample empirical female mortality rates for individual ages 75 to 89, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 6(a), 6(b) and 6(c). See legend to Fig. 5.27 for explanation.

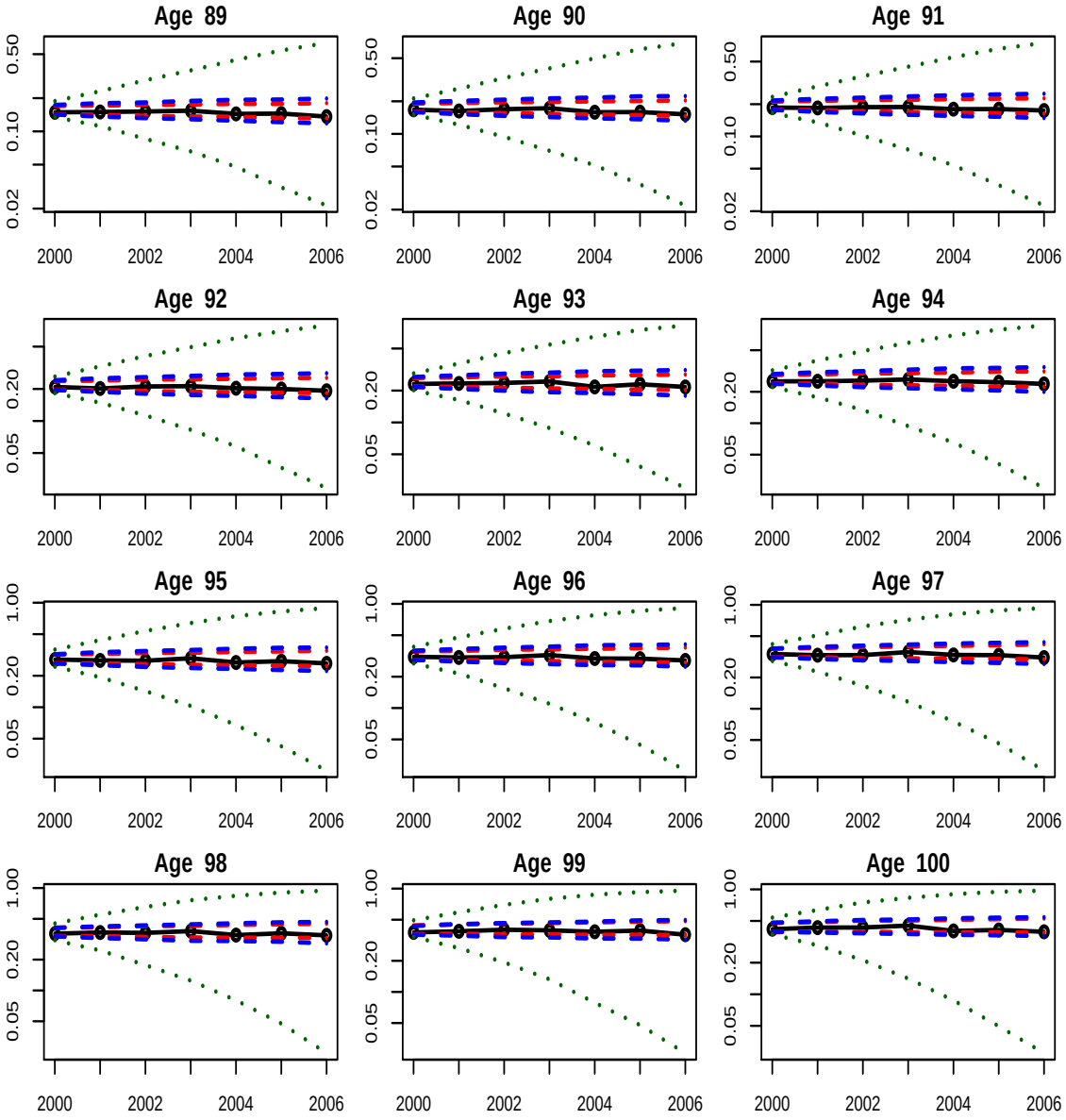


Figure 5.29: Plots of out-of-sample empirical female mortality rates for individual ages 90 to 100, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 6(a), 6(b) and 6(c). See legend to Fig. 5.27 for explanation.

levels shown are very conservative.

	Model 5			Model 6			Model 7		
Prob. level	80%	95%	99%	80%	95%	99%	80%	95%	99%
Model (a)	0.78	0.92	0.97	0.81	0.92	0.96	0.83	0.96	1.00
Model (b)	0.63	0.85	0.96	0.67	0.88	0.96	0.77	0.93	0.98
Model (c)	0.98	1.00	1.00	0.98	1.00	1.00	0.98	1.00	1.00

Table 5.13: Table shows the proportion of the out-of-sample observed age-specific female mortality rates $\hat{m}_{x,t}$ falling within the posterior predictive at selected levels.

5.6 Comparison of Male and Female Mortality

Model 1 is the only model in which the period parameter is identifiable up to scale and so comments regarding mortality improvements can be made in a more straightforward manner without the need to impose what could be considered relatively arbitrary constraints. Both males and females show improvements in mortality with time (Fig. 5.4(b) and Fig. 5.17(b), respectively). For males, Fig. 5.4(c) also indicates that improvements have been greatest for the youngest ages particularly 60 to 65 and, for females, Fig. 5.17(c) indicates that the improvements have been greatest for ages 72-86.

Comparing the model fit results, moving from Model 1 to Model 2 provided a dramatic improvement for males but not for females, suggesting that the cohort factor is more pronounced for males, while age-related interactions are a more important feature than are cohort features for females.

Model 6(b) fits best for the males, both for in-sample (based on the DIC results, table 5.4), and for out-of-sample data (based on table 5.7). For the females, Model 6(a) fits best for the in-sample data (based on the DIC results) and the out-of-sample data (based on table 5.13).

Tables 5.14 and table 5.15 show how the trend in age-specific male mortality rates varies relative to age-specific female mortality. The forecast figures (years 2000 to 2070) for table 5.14 were obtained from Model 6(b) for both males and females. The figures in table 5.15 were taken from the models with the lowest DIC score, i.e., 6(b) for males and 6(a) for females. The first column shows the ages of the age-specific

	1960	1970	1980	1990	2000	2010	2020	2030
60	2.03	1.97	1.93	1.71	1.58	1.49	1.52	1.54
65	1.99	2.15	1.98	1.80	1.67	1.50	1.46	1.45
70	1.81	2.01	1.93	1.80	1.67	1.55	1.46	1.48
75	1.57	1.75	1.84	1.77	1.62	1.67	1.53	1.48
80	1.37	1.51	1.65	1.70	1.52	1.37	1.37	1.28
85	1.29	1.38	1.41	1.52	1.43	1.23	1.22	1.11
90	1.22	1.29	1.33	1.34	1.30	1.13	0.93	0.81
95	1.23	1.16	1.18	1.29	1.22	1.05	0.89	0.68
100	0.88	1.18	1.10	1.31	1.16	0.98	0.89	0.69

Table 5.14: Table shows the ratio of male mortality over female mortality for the years detailed along the horizontal and ages down the vertical. For the years 1960 to 1990 the ratio is that of the observed mortality rates and for the years 2000 to 2030 the ratio is that of the median forecast mortality rates for males and females for Model 6(b).

	1960	1970	1980	1990	2000	2010	2020	2030
60	2.03	1.97	1.93	1.71	1.58	1.50	1.52	1.54
65	1.99	2.15	1.98	1.80	1.67	1.50	1.47	1.46
70	1.81	2.01	1.93	1.80	1.66	1.55	1.48	1.48
75	1.57	1.75	1.84	1.77	1.61	1.67	1.56	1.49
80	1.37	1.51	1.65	1.70	1.51	1.39	1.43	1.34
85	1.29	1.38	1.41	1.52	1.43	1.23	1.23	1.13
90	1.22	1.29	1.33	1.34	1.30	1.12	0.93	0.84
95	1.23	1.16	1.18	1.29	1.22	1.05	0.90	0.72
100	0.88	1.18	1.10	1.31	1.15	0.99	0.92	0.74

Table 5.15: Table shows the ratio of male mortality over female mortality for the years detailed along the horizontal and ages down the vertical. For the years 1960 to 1990 the ratio is that of the observed mortality rates and for the years 2000 to 2030 the ratio is that of the median forecast mortality rates from Model 6(b) for males and from Model 6(a) for females.

mortality rates being compared and each column heading shows the calendar year of comparison. The ratios in the first 4 columns of each table, 1960 to 1990, are based on the observed age-specific mortality rates and the last 4 columns, 2000 to 2030, are the ratios of the median forecast male and female rates for the particular age and calendar year. For ages 60 to 90 there is a general move towards 1 of the ratio of the median male and female age-specific mortality rates. For ages 90 and above the trend of male and female age-specific mortality rates continues and eventually the male age-specific mortality falls below the female age-specific mortality for longer term forecasts. A male age-specific mortality rate that is very much lower than the corresponding female age-specific mortality rate whilst not impossible is very unlikely and indicates that perhaps the models require some modification to

ameliorate this feature. A suggestion to produce coherent male-female forecasts is explained in section 7 "Future Research". A non-Bayesian approach is set out by Biatat and Currie (2010).

CHAPTER 6

Applications to Insurance

In previous chapters, a number of different stochastic mortality models have been constructed and calibrated using male and female population data from England and Wales. The models are very flexible and can be easily calibrated to other population groups. The posterior predictive distribution of the parameters enables forecasts of the number of deaths and the central mortality rate. Using the relationship of $m_{x,t}$ and $q_{x,t}$ as detailed in equation (2.1.5), we can derive the estimated posterior probability of death rates $q_{x,t}$. The $q_{x,t}$'s can then be used in many different insurance applications. This chapter applies the model results to three different insurance applications:

- future expectation of life for males and females;
- solvency capital requirement for longevity risk of insurers under EU solvency regulations;
- future number of centenarians.

6.1 Forecast Expectation of Life

As described in Chapter 2, there are two different methods for calculating life expectancy: *period life expectancy*, where the future life expectancy is calculated using the current age-specific mortality rates without any allowance for future improvement in longevity as specified by equation (2.1.6), and *cohort life expectancy*, which

is calculated using age-specific mortality rates, allowing for future improvement in longevity, as specified by equation (2.1.7).

In the graphs that follow, 2000 simulated life expectancy figures are calculated from 2000 posterior draws of the mortality model parameters, each set of parameters giving rise to one life expectancy. Forecasts of period and cohort parameters are obtained following the method of section 5.3.

If we assume that we have a model that fits the data reasonably well, then the most important factor affecting forecasts is the forecast method. For the models under consideration this is driven by the smoothing prior on the period and cohort parameters. Models 5(a) to 5(c), 6(a) to 6(c) and 7(a) to 7(c) produce reasonably similar fits to data. However, according to DIC, Models 6(a) to 6(c) are the best. As mentioned in section 4, the restriction of $m_{x,t}$ to the interval $(0, 1)$ reduces materially the probability of obtaining more deaths than the exposure and so should improve the long-term forecasts of mortality. For this reason and to reduce the number of results I have chosen Models 6(a) to 6(c) to demonstrate the influence of smoothing priors on the forecast results.

Figures 6.1 and 6.2 show the distribution of the expected additional years of life for males and females aged 60 in the particular calendar year. First consider Model 6(c), where the smoothing priors follow a random walk on the first differences. This process preserves the past trend in parameters and so, for simulations where the overall parameter trend is decreasing (increasing), this will result in very low (high) mortality rates in the medium to longer term forecasts. The outcome is a very wide range of life expectancy in the future, as shown in Fig. 6.1c and Fig. 6.2c. Thus the value of such a smoothing algorithm for long-term forecasts is limited. Furthermore, the distribution of life expectancy is actually bimodal with modes at very low and very high life expectancies.

Model 6(a) incorporates smoothing priors that follow a random walk on the levels, whilst Model 6(b) incorporates an autoregressive process of order 1 on the first differences. The box-plots of life expectancy for both these models (Fig. 6.1a, b and Fig. 6.2a, b) show much narrower ranges of future life expectancies than Model 6(c). The median values of life expectancy by sex are quite consistent between these two models. Interestingly, both Models 6(a) and 6(b) result in a wider range of forecast

life expectancy for males than for females. For example, the forecast lower and upper quartiles of period life expectancy of males in 2030 are (20.7, 23.4) (Fig. 6.1a) and (20.5, 24.5) (Fig. 6.1b). For females the corresponding numbers are (24.5, 26.6) (Fig. 6.2a) and (24.6, 26.1) (Fig. 6.2b).

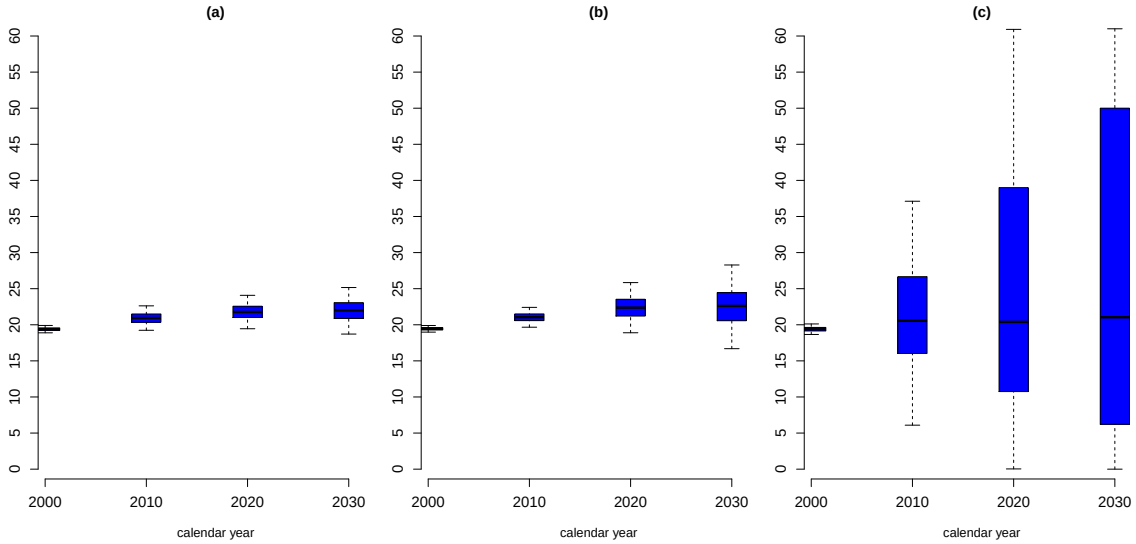


Figure 6.1: Panel of box plots of forecast period life expectancy as additional years of life for males aged 60 in future calendar years 2000 to 2030. (a) is calculated from Model 6(a), where the period and cohort parameters follow a random walk on the levels. (b) is calculated from Model 6(b), where the period and cohort parameters follow an autoregressive process on the first differences. (c) is calculated from Model 6(c), where the period and cohort parameters follow a random walk on the first differences.

To forecast period life expectancy for males and females, the model with the lowest DIC is selected (sections 5.2.5 and 5.4.5). For males this is Model 6(b) and for females Model 6(a). Figures 6.3 and 6.4 show how the distribution of period life expectancy progresses with time for ages 60, 65, 75 and 85. These ages and calendar years were selected because other organisations, namely HMD and ONS, have provided official figures that put the results of the models in some context. In the figures, the black line denotes the median life expectancy, the blue lines denote the upper and lower quartiles and the gaps between the red lines are 95% posterior predictive intervals. The figures show that the median male and female life expectancies increase until the early 2020's and then level off. For ages up to 85 and future years to 2030, females in general have a higher median life expectancy than the corresponding males. However, the differential in life expectancy reduces with time. For very old ages, age 90 and more, by 2020 male median life expectancy exceeds

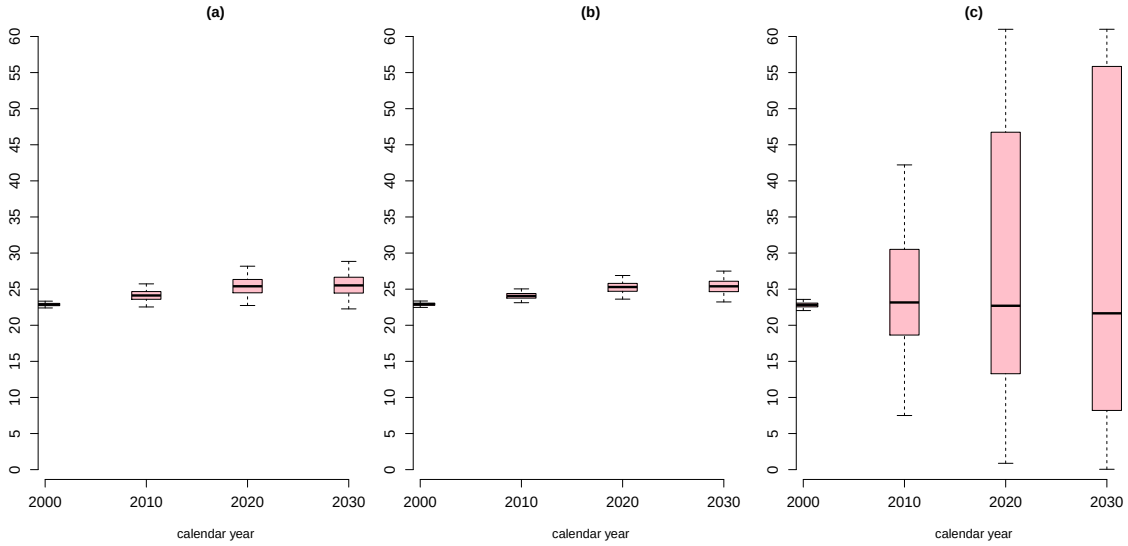


Figure 6.2: Panel of box plots of forecast period life expectancy as additional years of life for females aged 60 in future calendar years 2000 to 2030. See legend to Fig. 6.1

female median life expectancy.

For the calendar years 2000 and 2009 the HMD have published life expectancies based on observed mortality. These are denoted by the black triangles in Figures 6.3 and 6.4. The HMD life expectancies are just above the upper quartile from the models but within the 95% posterior predictive intervals. The specific estimates are shown in tables 6.1 and 6.2.

The ONS also produce regular forecasts of period life expectancy in their National Population Projections. The ONS report (Office for National Statistics (2011c)) includes forecasts of period life expectancy for selected ages (60, 65, 75 and 85) and future calendar years (2021, 2031 and 2033). These figures (excluding those relating to 2033, as they are similar to 2031) are included in figures 6.3 and 6.4 as solid black dots. It is important to note that the ONS figures relate to the UK as a whole. One would expect estimates for England and Wales to be slightly higher (by about 0.25 years) because Scotland's period expectation figures have been generally slightly lower than for the rest of the UK. Clearly the ONS view is that mortality will continue to improve at a greater rate than implied by our various mortality models.

Tables 6.1 and 6.2 show the HMD and ONS period life expectation figures together with the lower and upper limits of 95% posterior predictive intervals from the mortality model 6(b) for males and 6(a) for females. The ONS methodology for fore-

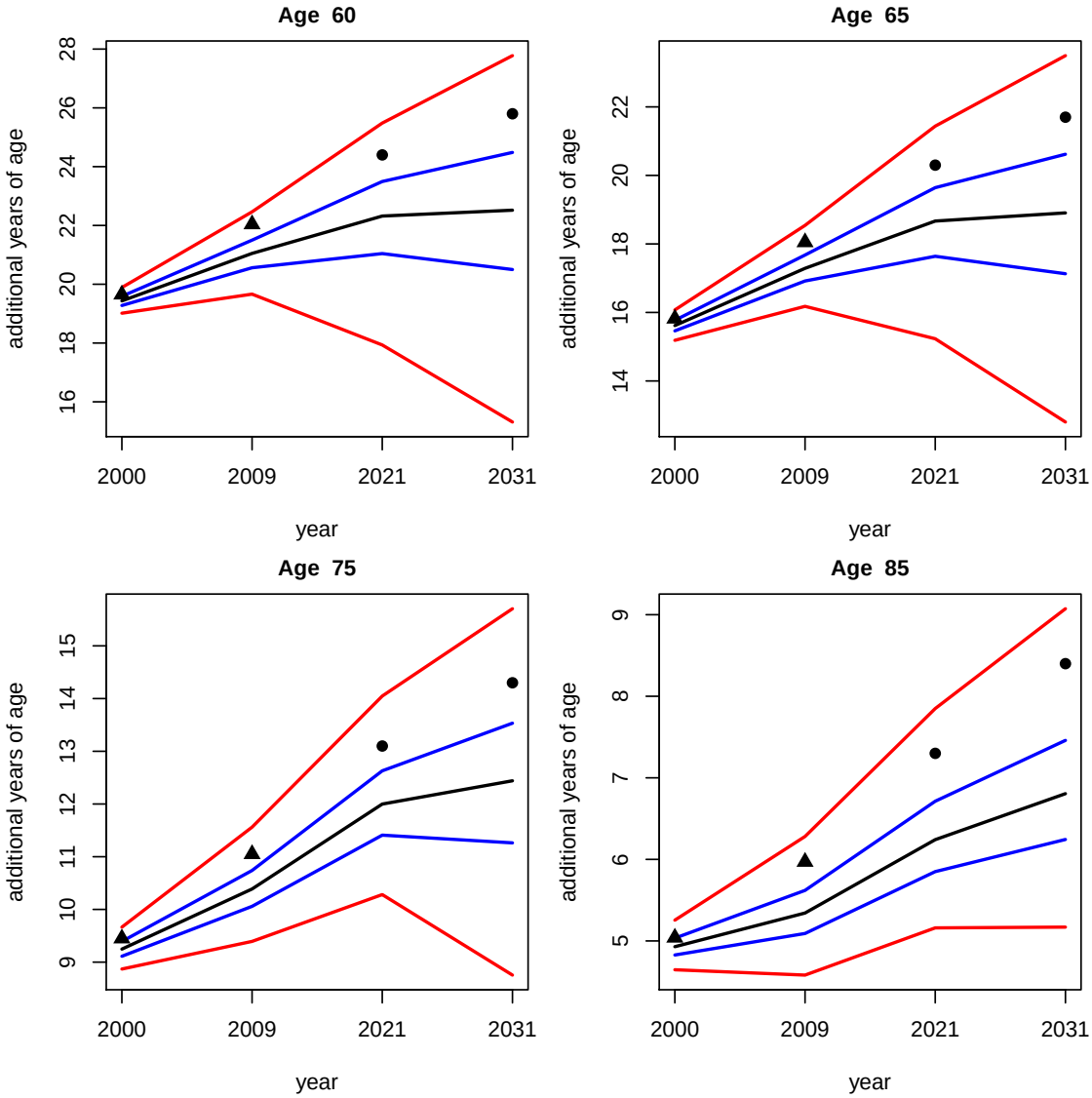


Figure 6.3: The progression of period life expectancy for various male ages over the period 2000 to 2031. The black line denotes the median period life expectancy, the blue lines denote the upper and lower quartiles and the red lines denote the 95^{th} and 5^{th} percentiles. The black triangles denote the observed period expectancy in 2000 and 2009 by HMD. The black dots are values forecast by the ONS for the years 2021 and 2031.

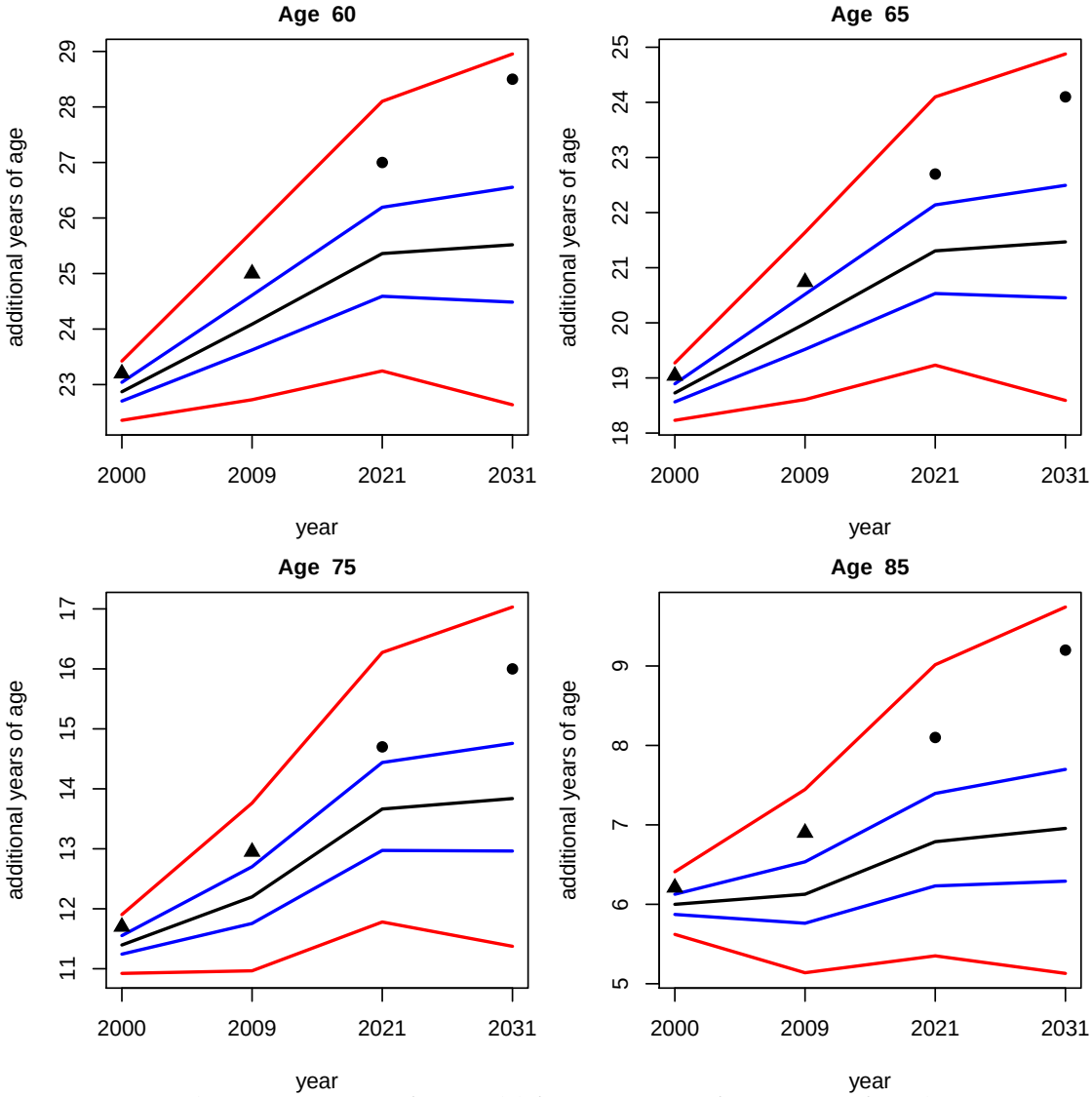


Figure 6.4: The progression of period life expectancy for various female ages over the period 2000 to 2031. See legend to Fig. 6.3

Table 6.1: Table shows the estimated period life expectancy for males by the HMD for the years 2000 and 2009; and forecast period life expectancy for males by the ONS for the years 2021 and 2031. The 95% posterior predictive intervals from Model 6(b) are also shown by the $2\frac{1}{2}$% and $97\frac{1}{2}$% percentile values.

Age	60			65			75			85		
		HMD			HMD			HMD			HMD	
percentile	$2\frac{1}{2}$%	ONS	$97\frac{1}{2}$%	$2\frac{1}{2}$%	ONS	$97\frac{1}{2}$%	$2\frac{1}{2}$%	ONS	$97\frac{1}{2}$%	$2\frac{1}{2}$%	ONS	$97\frac{1}{2}$%
2000	19.0	19.6	19.9	15.2	15.8	16.0	8.9	9.4	9.6	4.6	5.0	5.2
2009	19.7	22.0	22.5	16.2	18.0	18.5	9.4	11.0	11.5	4.6	6.0	6.2
2021	18.0	26.1	25.6	15.3	20.7	21.5	10.2	13.4	14.0	5.1	7.4	7.7
2031	15.5	26.3	28.1	13.0	22.0	23.7	8.8	14.5	15.8	5.1	8.4	9.0

Table 6.2: The estimated period life expectancy for females by the HMD for the years 2000 and 2009; and forecast period life expectancy for females by the ONS for the years 2021 and 2031. The 95% posterior predictive intervals from Model 6(a) are also shown by the $2\frac{1}{2}$% and $97\frac{1}{2}$% percentiles.

Age	60			65			75			85		
		HMD			HMD			HMD			HMD	
percentile	$2\frac{1}{2}$%	ONS	$97\frac{1}{2}$%	$2\frac{1}{2}$%	ONS	$97\frac{1}{2}$%	$2\frac{1}{2}$%	ONS	$97\frac{1}{2}$%	$2\frac{1}{2}$%	ONS	$97\frac{1}{2}$%
2000	22.4	23.2	23.4	18.3	19.0	19.2	11.0	11.7	11.9	5.7	6.2	6.4
2009	22.7	25.0	25.6	18.6	20.7	21.5	10.9	12.9	13.7	5.3	6.9	7.4
2021	23.1	27.2	27.9	19.1	22.9	23.9	11.6	14.9	16.1	5.3	8.1	8.8
2031	22.5	28.7	28.6	18.5	24.3	24.5	11.3	16.2	16.7	5.1	9.3	9.4

casting mortality rates is detailed in Office for National Statistics (2012a) and Office for National Statistics (2011b). In summary the methodology derives crude mortality rates from the actual experience for the period 1961 to 2009. These are then smoothed using a p-spline model to produce a fitted, smoothed mortality surface to the historical data for each gender. Age-specific rates of mortality improvement were then calculated for the year 2007 using the smoothed mortality rates calculated for 2006 and 2007. These mortality improvement rates for 2007 were then projected forward to 2010 by assuming that the same rates of improvement applied in 2008 to 2010. The smoothed base mortality rates at 2010 are then reduced by applying mortality improvement rates that vary by age and sex, that were derived from recent mortality trends. Mortality improvements for 25 years ahead of the base year and beyond are fixed at a target rate. The target rate used in the life expectancy figures in the National Population Projections 2010-based report was 1.2% for males and females.

Putting the median values into historical context, figure 6.5 shows the progression of period life expectation for males and females at age 60. The first four periods

of each bar chart, years 1960 to 1990, show the period life expectancy as calculated by the HMD. The last four periods of each chart show the median forecast period life expectancy for the calendar years 2000 to 2031. For the year 2000 the median forecast period life expectancies for males and females aged 60 are 19.44 and 22.90, respectively which can be compared with the HMD values of 19.65 for males aged 60 and 23.20 for females aged 60. Both the male and female HMD figures are within the 95% posterior predictive intervals of the estimated period expectancy values. Figure 6.5 displays a continued improvement in longevity for males and females up to the calendar year 2030, when the improvements level off at 22.7 years for males and 25.5 for females. This also indicates a narrowing of the differential between male and female period life expectation. In 1970 the difference was around 4.5 years, whereas by 2030 the difference is predicted to reduce to around 2.8 years.

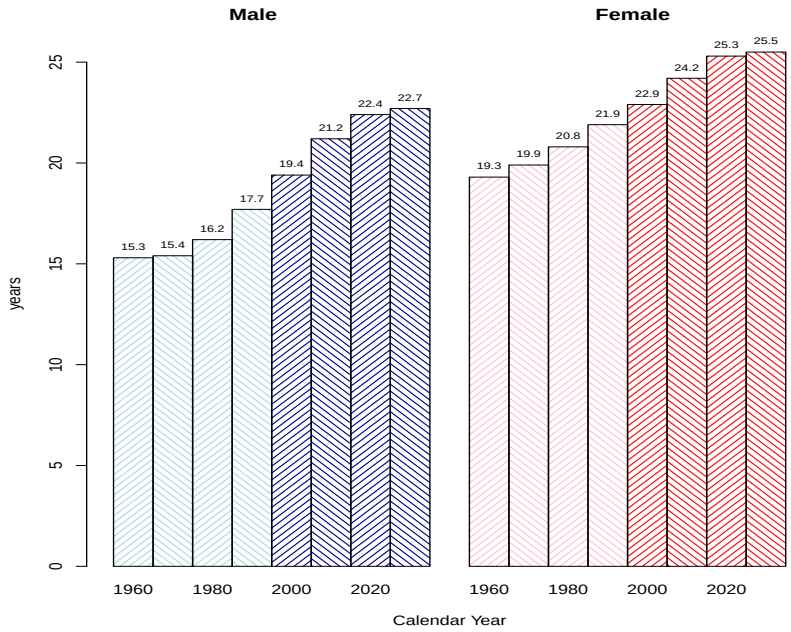


Figure 6.5: Period expectation of life, for males and females aged 60. The first four columns of each chart give expectation of life based on observed mortality whereas the last four columns give forecast period life expectancy from Model 6(b) for males and Model 6(a) for females.

We turn now to cohort life expectancy. Model 6(b) was again used for male mortality and Model 6(a) for female mortality as above. Figures 6.6 and 6.7 show how the distribution of cohort life expectancy progresses with time for ages 60, 65, 75 and 85. The progression of the median cohort life expectancy is flatter and above the corresponding median period life expectancy because the proportion of simulations

where mortality reduces exceeds the proportion where mortality increases and so median mortality reduces with time.

The solid black dots denote forecasts of cohort life expectancy by the ONS extracted from their report "National Population Projections Base 2008" (Office for National Statistics (2011b)). Figure 6.7 also shows that three of the ONS predicted female cohort life expectancies lie outside the 95% posterior predicted intervals whilst all of the female period life expectancies forecasts were inside the 95% intervals. This departure can be attributed to a difference in the allowance for future improvements in mortality between the ONS and our models. Forecast period life expectancy incorporates improvements in mortality up to the point the forecast is made. Thereafter, no further improvement is assumed, whereas forecast cohort life expectancy makes allowance for assumed future mortality improvements, not only up to the point of the forecast, but also into the full prospective lifetime of the individual. The fact that the ONS life expectancies are larger than our models' predictions is a clear indication that the ONS assume that mortality improvements will occur at a higher rate than the models. As with period life expectancy, the ONS believe that cohort life expectancy will be several years (3-5 years) greater than the median cohort life expectancy obtained from our mortality models.

6.2 Solvency Capital Requirement under Solvency II

At some date in the future, all but the smallest companies engaging in the business of insurance and reinsurance within Europe must follow the European Solvency II directive. Solvency II is a revised set of EU-wide capital requirements and risk management standards. The Solvency II framework has three main areas (pillars):

- Pillar 1 consists of the quantitative requirements (for example, the amount of capital an insurer should hold);
- Pillar 2 sets out requirements for the governance and risk management of insurers, as well as for the effective supervision of insurers;
- Pillar 3 focuses on disclosure and transparency requirements.

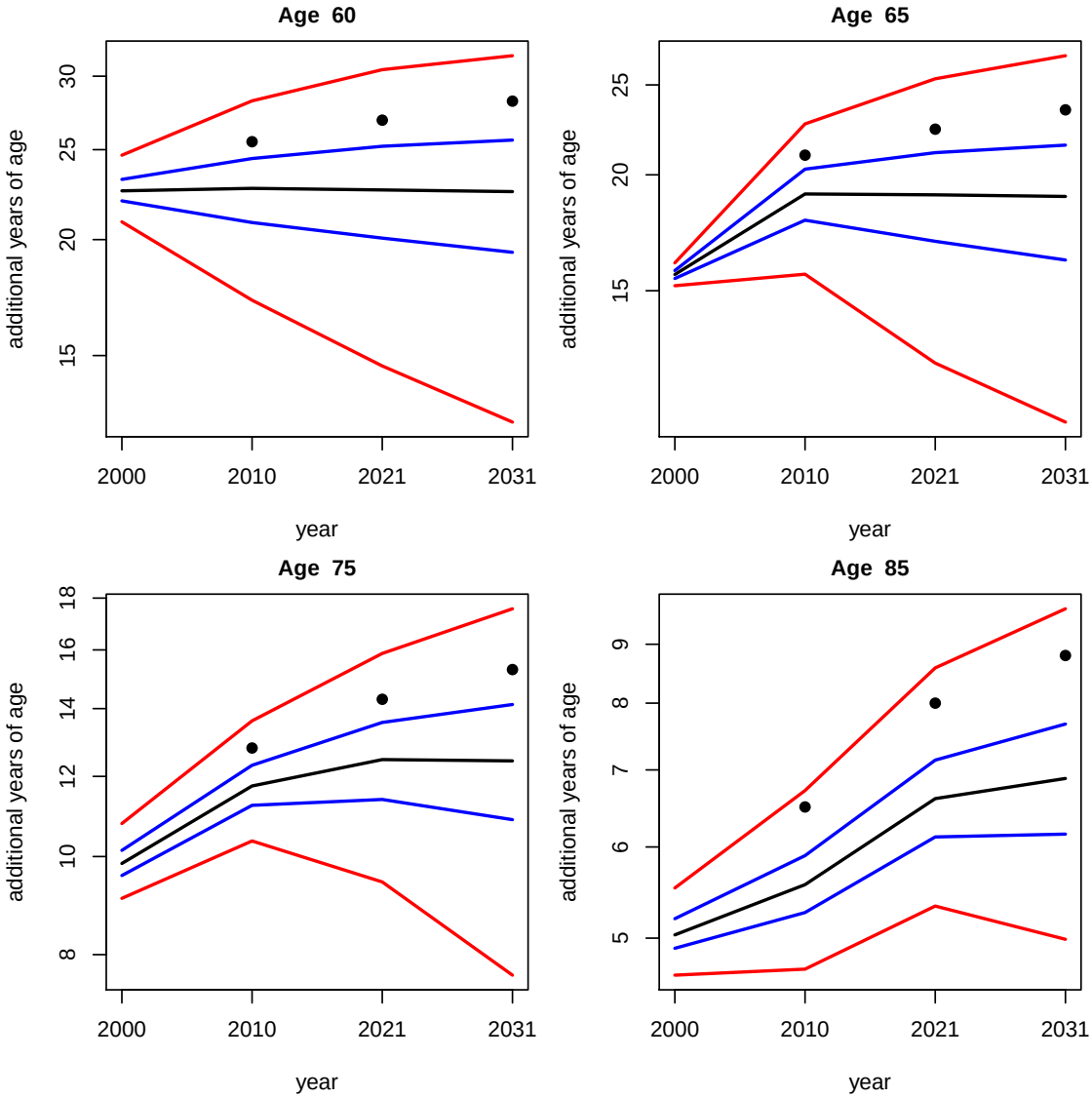


Figure 6.6: The progression of cohort life expectancy for males of various ages over the period 2000 to 2031. The black line denotes the median cohort life expectancy, the blue lines provide the upper and lower quartiles and the red lines denote the 95^{th} and 5^{th} percentiles. The black dots indicate the values forecast by the ONS for the years 2021 and 2031.

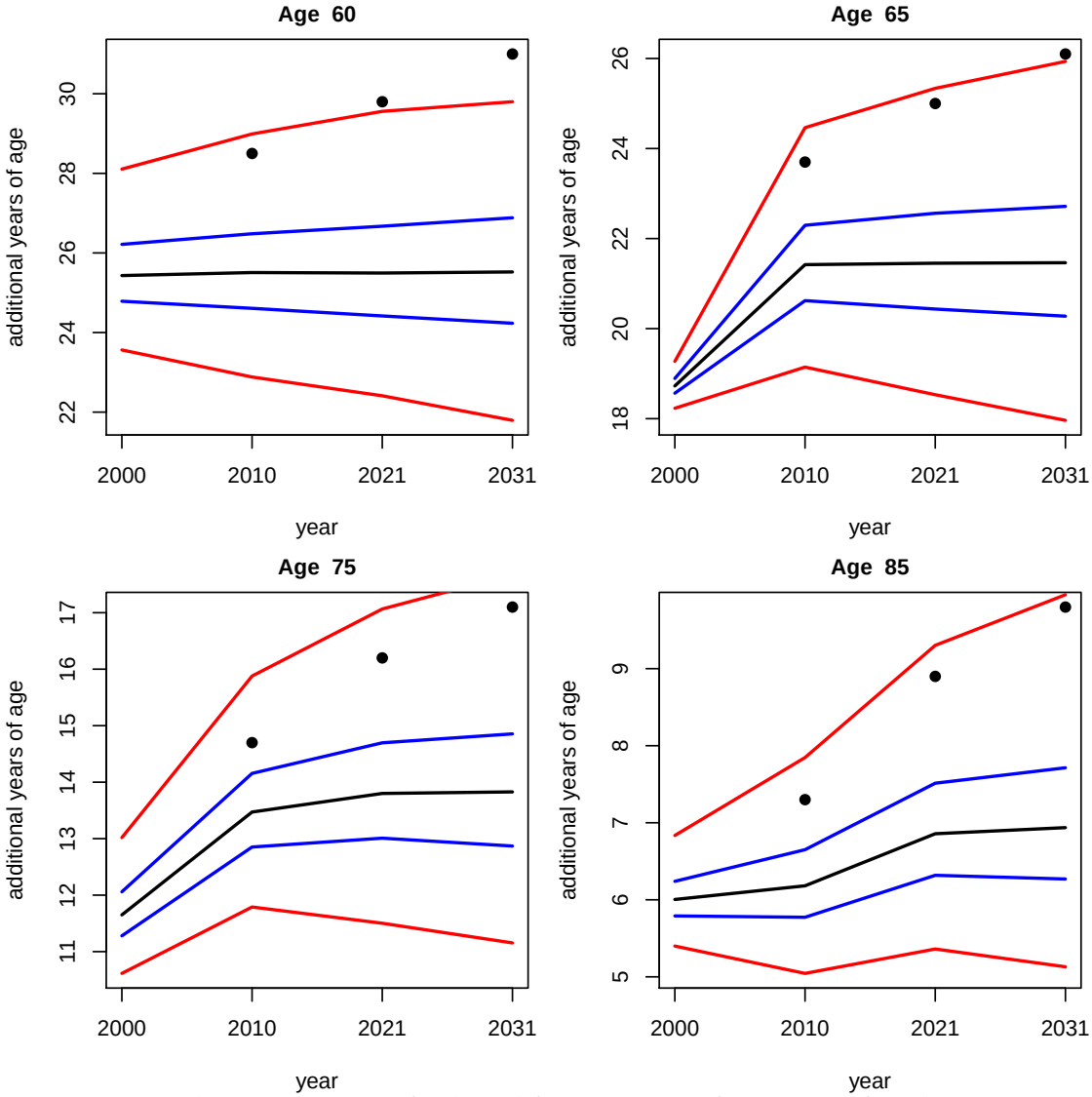


Figure 6.7: The progression of cohort life expectancy for various female ages over the period 2000 to 2031. See legend to Figure 6.6

This section focuses on Pillar 1, and in particular on the *solvency capital requirement* for longevity risk for annuity business and how Bayesian models can assist in examining and quantifying the key elements of longevity risk.

The quantifying of longevity risk needs to account for many sources of variation in mortality forecasts; firstly errors in the base mortality enter in as part of the fitting process. Secondly, future improvements in mortality may be different from those estimated and, thirdly, potentially extreme increases in longevity beyond those captured within the data used for fitting. The Bayesian Statistics provides an integrated approach to the first two sources of error. The third needs to be estimated by experts with information outside the model. The risks connected to the third factor are much more subjective and have been ignored here.

The *solvency capital requirement* (SCR) under Solvency II is defined as the amount of capital necessary at the valuation date (say the current date) to cover, with a probability of at least 99.5% all losses which may occur in the following 12 month period. This can be re-expressed as a 1-year 99.5% Value at Risk (VaR). So for a given portfolio of liabilities, probability and time horizon, VaR is defined as a threshold value such that the probability that the loss on the portfolio over the given time horizon exceeds this value is the given probability level. Figure 6.8 illustrates this in graphical form. The chart shows the distribution of losses from a hypothetical portfolio after 12 months from the valuation date. The line dividing the white and red areas would be the 99.5% VaR. So the SCR would be the loss corresponding to $\alpha = 99.5\%$, shown as 4 in the illustration. The company would therefore require to hold sufficient capital at the valuation date to cover a potential loss of 4.

Glossing over exactly what is meant by the value of assets and liabilities, so as to elucidate the calculation of the SCR and how this is relatively straightforward application of Bayesian models, we denote by AC_t the difference between the value of assets and the value of the liabilities at time t. Then the SCR required at the current valuation date $t = 0$ can be written more formally as the smallest positive x that satisfies the following condition

$$Pr(AC_1 > 0 | AC_0 = x) \geq 0.995. \tag{6.2.1}$$

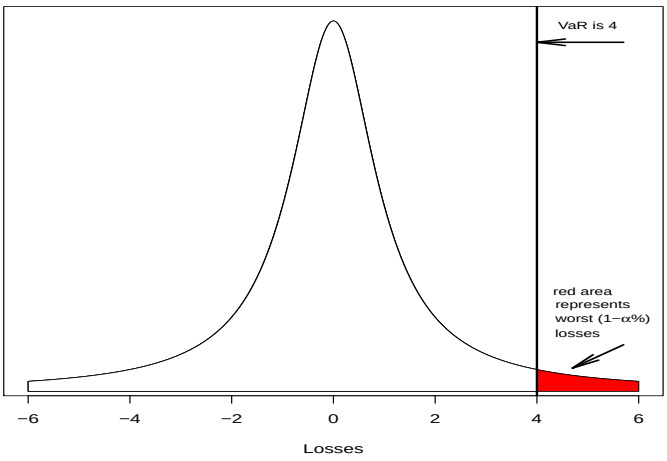


Figure 6.8: An illustration of a hypothetical distribution of profit and loss generated by a central t-distribution with 5 degrees of freedom. The VaR at probability level α is represented by the vertical black line bordering the red area.

That is, having x additional assets over the liabilities at the start of the year will result in assets being equal to or greater than the liabilities at the end of the year with a probability of 99.5%. The value of x is the SCR. If we denote by i as the 1 year annual return on the assets, Bauer et al. (2009) shows that the condition in equation 6.2.1 is approximately equivalent to

$$Pr(AC_0 - \frac{AC_1}{(1+i)} > x) \leq 0.005, \tag{6.2.2}$$

that is, the probability that the loss over one year ($(AC_0 - \frac{AC_1}{(1+i)}$) exceeds x (i.e. the SCR) is less than or equal to 0.005.

The problem with the SCR risk measure for longevity risk is that longevity risk lies in the long-term path taken by mortality rates rather than over a short 1-year period. It is therefore difficult to identify for longevity risk a transparent method that equates to the 1-year 99.5% Value at Risk measure. Platt (2011) and Richards et al. (2013) suggest possible approaches. Clearly the risk measure needs to include more than 1-year 99.5% VaR on current mortality but should be less than 99.5% VaR over the term until the last annuity risk ends, denoted by ω. The regulations allow companies to use a permanent 25% decrease in mortality rates to calculate the SCR rather than calculations based on their own specific longevity risks. The implication is that a permanent 25% decrease in mortality rates is equivalent to a 1-year 99.5% VaR, or perhaps slightly more penal since the regulators want to encourage

companies to do their own risk calculations.

In the remainder of this section we will consider what a permanent 25% decrease in mortality rates equates to in terms of probability level over the term of the risk and also what permanent reduction in mortality equates to a 99.5% VaR over the term of the risk. The following paragraphs set out a number of assumptions to simplify the calculations of the SCR.

At the valuation date ($t = 0$) we will assume for the time being that $AC_0 = 0$, that is, the value of assets is equal to the value of liabilities, and look at the distribution of the differences in the value of assets and liabilities after one year at $t = 1$. A portfolio of life annuities (the liabilities) would typically comprise males and females of different ages with different annuity payments. We will consider a set of hypothetical annuity portfolios each one consisting of N individuals all of the same age and a single sex. The assumption of individuals in the portfolio having the same age and sex are not realistic but have been chosen to simplify the calculations and to make the pattern in SCR requirements by age and sex clearer. To simplify the calculations further we assume that the life annuities pay a fixed annuity amount of A yearly at the end of each year provided the life is still alive. The formula for the liability of a portfolio of lives aged x in year T is

$$Liab_{x,T} = AN \sum_{k=1}^{\omega - x} \frac{\prod_{j=1}^{k}(1 - q_{x+j-1,T+j-1})}{(1 + i_k)^k}, \tag{6.2.3}$$

where $q_{x,t}$ is the mortality rate for age x in year t and i_k is the interest rate for term k years; i.e, the current value of AN payable in 10 years time is $AN/(1 + i_{10})^{10}$. The expression $\prod_{j=1}^{k}(1 - q_{x+j-1,T+j-1})$ is an estimate of the probability that an individual age x at the beginning of year T survives to age $x + k$ at the beginning of year $T + k$.

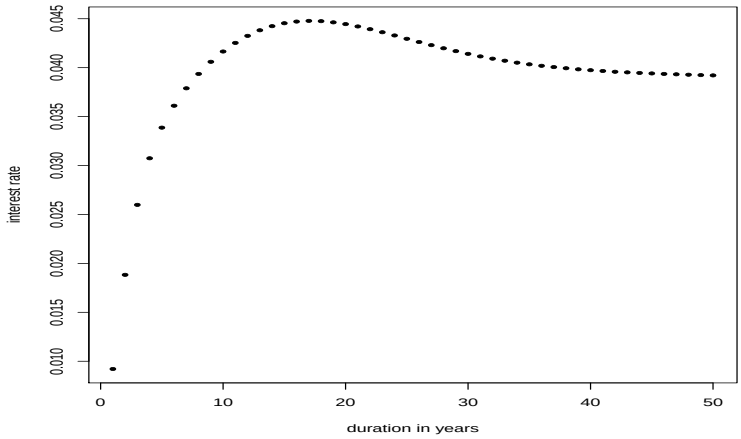
The assets provide an interest return equivalent to i_k for $k = 1, \ldots, \omega$.

As we are considering only longevity risk then we assume that the interest rates and all other factors remain unchanged (otherwise we would have to consider the dependency between these other risks and longevity risk). If the current calendar year is T_0, then from our mortality model we can use the simulations of future mortality rates to produce simulations of liabilities at the beginning and end of calendar year

T_0 (i.e., $Liab^{(r)}_{x,T_0}$ and $Liab^{(r)}_{x+1,T_0+1}$) by substituting the appropriate mortality and interest rates in equation 6.2.3. For $Liab^{(r)}_{x,T_0}$, replace q_{x+j-1,T_0+j-1} by $q^{(r)}_{x+j-1,T_0+j-1}$, the mortality rates derived from the parameters of the r^{th} MCMC cycle for age $x+j-1$ and year T_0+j-1. The i_k are unchanged. For $Liab^{(r)}_{x+1,T_0+1}$ replace q_{x+j-1,T_0+j-1} by $q^{(r)}_{x+j,T_0+j}$. The interest rates are an unwelcome complication where the i_k is replaced by i'_k where $i'_k = (\frac{(1+i_{k+1})^{k+1}}{1+i_1})^{1/k}$. The formula for i'_k assumes that interest rates in one year's time are consistent with current interest rates i_k.

For those familiar with Solvency II the actual interest rates used in the calculation of liabilities are those from Quantitative Impact Assessment No. 5 (European Commission (2010)). A plot of these interest rates is shown in Figure 6.9.

Figure 6.9: Term-dependent spot interest rates (i_k) for discounting future values. The present value of X payable in k years is $\frac{X}{(1+i_k)^k}$.



Let us assume that the liability at $t=0$ is the median of $\{Liab^{(r)}_{x,T_0},\ r=\{1,\ldots,S\}\}$ where S represents the number of samples. As we are assuming that value of the assets equals the value of the liability at $t=0$, then we have the value of assets. As we have assumed that mortality is the only factor that is changing, we have sufficient information to calculate equation 6.2.2 for each simplified portfolio.

In tables 6.3 and 6.4 the columns labelled $Liab_{x,T_0}$ shows the median current annuity liability for a male or female aged x receiving an income of £1 per annum (i.e., $AN=1$ in equation 6.2.3). The columns labelled SCR shows the approximate SCR obtained if we assume that the 99.5% VaR is equivalent to a permanent reduction in mortality by 25%. The columns labelled $\frac{SCR}{Liab_0}$% show the SCR as a proportion of median current annuity liability. The figures relate to different annuity portfolios, each

portfolio consisting solely of either males or females aged either 60, 65, 75 or 85 as indicated. The figures in table 6.3 assume that mortality follows Model 6(a) and the figures in table 6.3 assume that mortality follows Model 6(b). The results indicate that for both models the *SCR* as a proportion of liability increases with age for both males and females. This is not surprising as the standard deviation of mortality increases with age. Furthermore the *SCR* for a given age is greater for males than for females; again this is expected as the models produce male mortality rates that have a greater standard deviation than for females and the future improvements at most ages are on average greater for males than for females.

As mentioned above, arriving at a transparent method that equates to a 1-year 99.5% Value at Risk measure for a liability that has a much longer term than one year presents a challenge. Tables 6.5 and 6.6 apply a lower probability level over the full term of the liability as one possible approach. The underlying assumptions and description for tables 6.5 and 6.6 are identical to tables 6.3 and 6.4 respectively. The figures tables 6.5 and 6.6 are calculated in a two stage process. Firstly, for each of the ages and percentile levels, the corresponding liability (referred to here as the liability in the stressed condition) is obtained. The second stage is to determine the percentage permanent reduction in mortality that would gives the same annuity liability as the liability in the stressed condition. It is these percentage permanent reductions in mortality that are shown in the tables. Table 6.5 indicates that, under Model 6(a), for a given probability level the permanent reduction in mortality is reasonably stable over age and by sex. However, Table 6.6 shows that, under Model 6(b), this is not the case and therefore under different mortality models different age-related factors would arise.

My conclusion is that rather than a fixed permanent reduction in mortality an age-related (perhaps also sex-related) function should be used for the calculation of the *SCR*. This agrees with Figure 6 in Richards et al. (2013).

6.3 Future Numbers of centenarians in England and Wales

This section applies the results of Models 6(b) and 6(a) to forecasting how the numbers of male and female centenarians in England and Wales may develop over the

Table 6.3: For individuals aged 60, 65, 75 and 85, the current annuity liability of £1 p.a., the additional risk capital SCR corresponding to a permanent 25% reduction in mortality and the ratio of SCR to the current annuity liability. Figures assume that mortality follows Model 6(a) separately for males and females.

	Males			Females		
Age	$Liab_0$	SCR	$\frac{SCR}{Liab_0}$%	$Liab_0$	SCR	$\frac{SCR}{Liab_0}$%
60	13.15	1.17	8.9%	14.53	0.94	6.5%
65	11.65	1.48	12.7%	13.06	1.01	7.8%
75	8.10	1.33	16.4%	9.37	1.12	12.0%
85	4.10	1.36	33.1%	4.77	1.30	27.3%

Table 6.4: For individuals aged 60, 65, 75 and 85, the current annuity liability of £1 p.a., the additional risk capital SCR corresponding to a permanent 25% reduction in mortality and the ratio of SCR to the current annuity liability. Figures assume that mortality follows Model 6(b) separately for males and females.

	Males			Females		
Age	$Liab_0$	SCR	$\frac{SCR}{Liab_0}$%	$Liab_0$	SCR	$\frac{SCR}{Liab_0}$%
60	13.37	1.12	8.4%	14.50	0.95	6.5%
65	11.92	1.21	10.2%	13.01	1.01	7.7%
75	8.35	1.26	15.1%	9.34	1.10	11.7%
85	4.34	1.31	30.1%	4.79	1.29	26.9%

Table 6.5: The permanent reduction in future mortality assumed in calculating the annuity liability equating to a 75%, 90%, 95% and 99.5% VaR over the term of the liability for both males and females. The results assume that mortality follows Model 6(a).

	Males				Females			
Age	75%	90%	95%	99.5%	75%	90%	95%	99.5%
60	10%	24%	29%	38%	10%	23%	28%	37%
65	12%	20%	23%	39%	11%	23%	27%	37%
75	11%	23%	27%	35%	11%	23%	27%	38%
85	9%	19%	22%	29%	10%	23%	27%	35%

Table 6.6: Table shows what permanent reduction in future mortality assumed in calculating the annuity liability equating to a 75%, 90%, 95% and 99.5% VaR over the term of the liability for both males and females. The results assume that mortality follows Model 6(b).

	Males				Females			
Age	75%	90%	95%	99.5%	75%	90%	95%	99.5%
60	18%	38%	44%	60%	8%	18%	21%	27%
65	12%	28%	33%	44%	8%	18%	21%	27%
75	9%	18%	21%	31%	7%	16%	18%	24%
85	8%	18%	21%	29%	7%	14%	17%	23%

next 20 years to 2030 if future mortality follows Model 6(b) for males and Model 6(a) for females and compares the results with official figures from the ONS.

The forecast starts with the population of males and females at the beginning of 2010. The population data by age and sex was obtained from the HMD (www.mortality.org). The data was the size of the population aged x on 1st January of calendar year t and so refers to all persons of the same sex in the age range $[x, x+1)$ on January 1st of calendar year t. Let these be denoted by $Pop_{x,t}$. For each of the 2000 MCMC simulations of parameters, forecast values of $m_{x,t}$ were obtained following the method detailed in section 5.3. The number of deaths ($d_{x,t}$) aged x in year t were then generated as a variate from a Poisson distribution with mean $Pop_{x,t}m_{x,t}$. The forecast population size aged $x+1$ at the beginning of calendar year $t+1$ is calculated as $Pop_{x+1,t+1} = Pop_{x,t} - d_{x,t}$. This process is repeated for the years 2010 to 2030 using each simulation to generate a new future scenario. It is then a straightforward matter to count the number of individuals aged 100 and over in each future year for each scenario. The calculation includes the simplifying assumption that the mortality rate of those above 100 is the same as those who are 100. This will be immaterial to the overall conclusion.

Figure 6.10 illustrates for both males and females the trend in the number of centenarians for the two forms of smoothing priors up to 2030. The vertical line at 2010 indicates the divide between actual figures to the left and forecast figures to the right. For each sex, there is a similar pattern and population size between the two smoothing priors. For males there is a general increasing trend over time. Initially the median forecast number of male centenarians is considerably below the median number of female centenarians in 2010 but increases more rapidly than females. For females the increase is much more modest and for first-order autoregressive smoothing priors on the differences the median forecast is almost constant; although there is considerable uncertainty, a sizeable proportion of the scenarios include a noticeable increase in female centenarians. The female result is a little at odds with official forecasts.

The U.K. Department for Work and Pensions (DWP) has produced official forecasts of centenarians for males and females combined (Price (2010)), (see also Office for National Statistics (2012c)). In order to compare this, the simulations of the number

of deaths of males and females from Model 6(a) were added together following the order of the simulations, i.e. the i^{th} simulation of the number of deaths for males and females were added together to give an i^{th} simulation of the total number of centenarians. This is reasonable because the male and female mortality models are considered independent.

This process was repeated for the number of deaths of males and females from Model 6(b). The results are shown in Figure 6.11 together with the DWP official forecast. The figures show that the official forecasts of number of centenarians are much greater than those from the model; only about 10% of the Model 6(a) and Model 6(b) simulations gave forecasts in excess of the official forecast at 2020 and about 2% of simulations gave a higher figure than the official forecast at 2030. By 2020 the official forecast of the number of centenarians is about 1.8 times the median value from the two models and by 2030 over 3.5 times the median. These results are consistent with the conclusion mentioned in section 6.1 that the ONS appear to assume that mortality improvements will occur at a higher rate than our models.

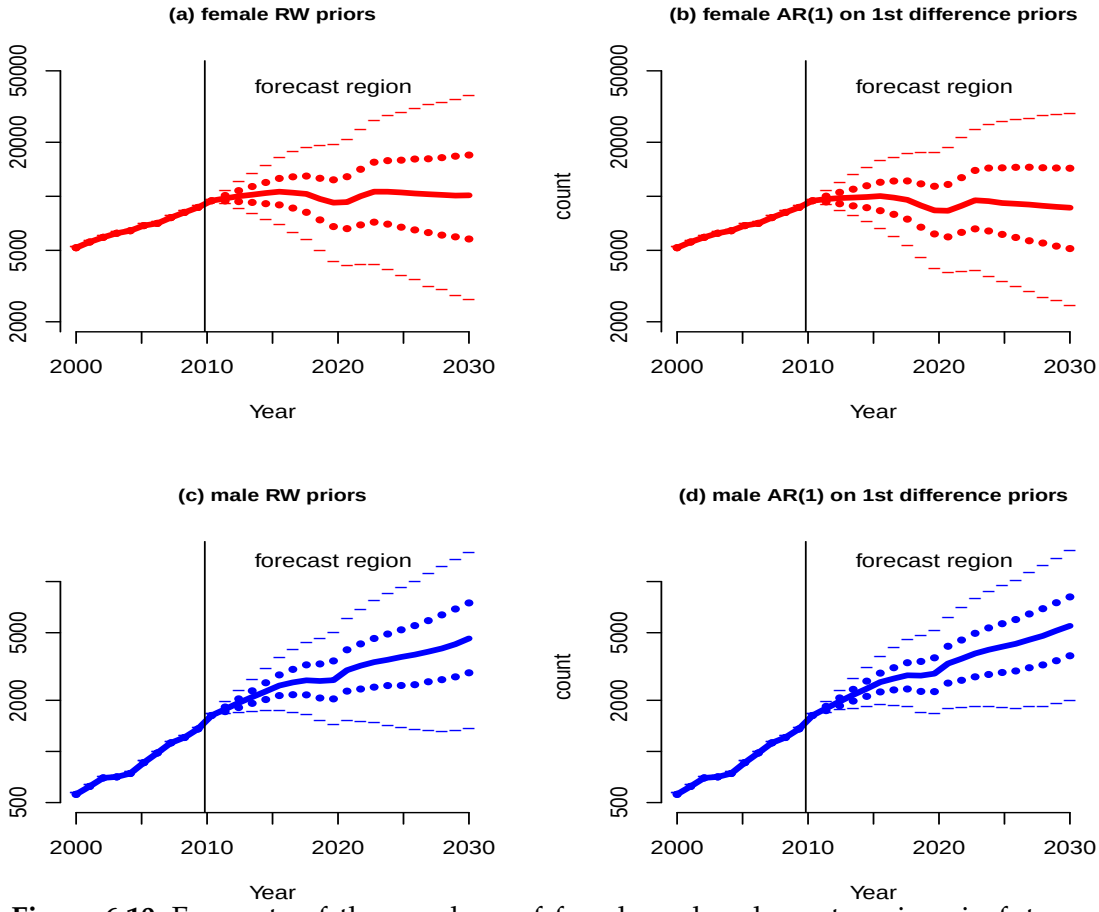


Figure 6.10: Forecasts of the numbers of female and male centenarians in future calendar years using different smoothing priors. The vertical line at 2010 indicates the change from actual numbers to forecast numbers. (a) and (b) are for females, (c) and (d) are for males. (a) and (c) are from Model 6(a), (b) and (d) are from Model 6(b). The heavier solid dots provide the upper and lower quartile figures and the solid line is the median forecast. The dashes indicate the 95^{th} and 5^{th} percentiles. The 2010 and earlier population counts were obtained from the HMD. The numbers on the y-axis in each plot are the population count but the plot is on a logarithm scale.

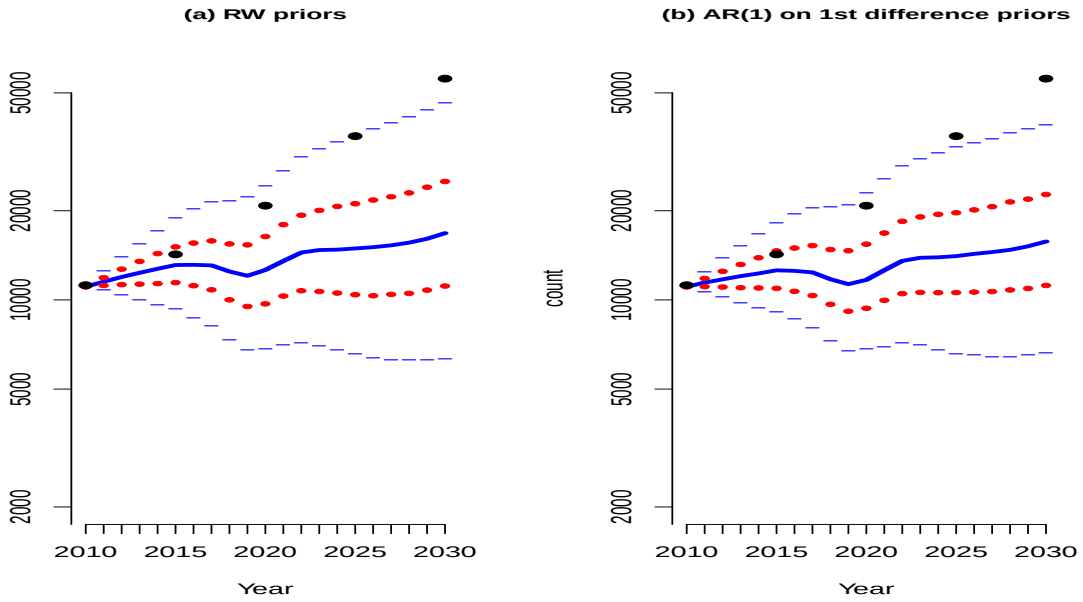


Figure 6.11: Forecasts of the numbers of all centenarians in future calendar years. (a) is model 6(a), (b) is for Model 6(b). The upper and lower red dashes provide the upper and lower quartile figures and the blue line indicates the median forecast. The blue dashes are the 95^{th} and 5^{th} percentiles. The black dots indicate the DWP official forecasts. The numbers on the y-axis in each plot are population count but the plot is on a logarithm scale.

CHAPTER 7

Discussion and Conclusions

The aim of this thesis was to develop flexible mortality models for the analysis and forecasting of mortality using a Bayesian approach. Several new models were introduced, mainly as generalisations of existing models, but in addition a model was developed that treats the exposures as estimated values rather than as known fixed values.

The introduction presented the problem, which is one that faces the U.K. government as well as the life assurance industry and sponsors of pension schemes: over the last 90 years period life expectancy at birth for both males and females has increased by 20 years, based on figures from HMD (www.mortality.org). This raises the question of how life expectancy will change over the coming decades. This dramatic increase in life expectancy has had and continues to have far-reaching effects on pension provision and future social care in the United Kingdom. The use of Bayesian mortality models can assist policymakers to make better informed decisions in these areas.

Chapter 2 analysed empirical mortality using the population data of England and Wales since the 1960s and demonstrated that the mortality rates of both males and females have consistently reduced over the second half of the 20th century; it is this fall in mortality that has caused the increase in life expectancy. Interestingly, the increase in life expectancy has not been a predictable process; certain decades have contributed to the increase in life expectancy more than others and within decades

the increase in life expectancy is a result of mortality improvements in different age groups. In particular, in the earlier decades of the 20th century, increases in life expectancy were as a result mainly of reductions in mortality at the younger ages, below 40. However, in the later part of the 20th century, the increases in life expectancy have been brought about by reductions in mortality at the older age groups of 60 and over. These older age groups have been the focus of interest for this thesis. Because of this pattern of changes in life expectancy, the mortality models labelled 4, 5, 6 and 7 include mortality improvements that contain age/period and age/cohort interactions. Evidence of the cohort effect in England and Wales population mortality has been well documented by other researchers for example, Willets (2004), Macdonald (2009) and was evident in plots of raw mortality rates and residual plots in chapter 5 associated with models that did not include cohort parameters.

In chapter 4 and appendix B, each of the models used was documented. Models labelled 1, 2 and 3 are models that have been extensively analysed in the literature. Model 1 is a close derivative of the Lee-Carter model, Model 2 is the age-period-cohort model and Model 3 is the age-period-cohort model with random effects. Models labelled 4, which form a generalisation of the age-period-cohort model with age interactions on the period and cohort parameters, has been included in research papers but this thesis is the first to fit this model using a Bayesian framework as well as introducing smoothing of parameters. Models 5 and 6 are new models that generalise Model 4 by introducing random effects on mortality. These random effects have been assumed to follow Gaussian distributions with mean zero and the precision a hyper-parameter within the models. Models labelled 6 involve a logit link function whereas models labelled 5 employ a log link function. The use of the logit function constrains the mortality rates to being in the interval (0,1). Models labelled 7 are new models based on Model 6 with the fundamental difference from all the other models that the exposures are noisy estimates rather than known fixed values. This reflects a genuine feature of reality as population exposures are estimated. Within chapter 4 the structure and prior distributions for these measurement errors is explained.

To consider how different types of smoothing affect the mortality forecasts, for each of the models 1 to 7, three different smoothing priors were assumed giving rise to three versions of each model denoted by an additional label, (a), (b) or (c). These smoothing priors not only enforce smoothing of the parameters but also enable forecasts of future mortality rates to be made by forecasting the individual smoothing processes into the future. Version (a) denotes a random walk on the levels of the parameters, (b) denotes an autoregressive process of order 1 on the first differences of the parameters and (c) denotes a random walk on the first differences of the parameters.

With all models, not all parameters are identifiable; this is particularly true of the age, period and cohort parameters in models labelled 2 to 7. Many researchers for example Brouhns et al. (2002), Czado et al. (2005), Cairns et al. (2009), Fienberg and Mason (1979) and Holford (1983, 1991) have chosen to deal with this problem by imposing constraints or by considering particular functions of these parameters that render the parameters identifiable. As the mortality parameter $m_{x,t}$ is identifiable in all models (except models labelled 7) I chose only to interpret the $m_{x,t}$ parameters. For Model 7, a very informative prior on the error parameters greatly reduced concerns regarding identifiability of the $m_{x,t}$ parameter.

Chapter 5 discussed the fit of the models by analysing the residuals. On the basis of these tests models labelled 4 to 7 tended to perform better than the well-researched models labelled 1 to 3. However, there is not a level playing field as models labelled 4 to 7 include many more parameters than models labelled 1 to 3. The Deviance Information Criterion was used to compare the models. Models labelled 6 and 5 gave the best DIC results followed by Model 7 and then Model 4. The model with the lowest DIC for males was Model 6(b), with smoothing priors following an autoregressive process of order 1 on the first differences of the parameters. For females the lowest DIC was obtained with Model 6(a), with smoothing priors following a random walk on the levels of the parameters. Models were also compared by considering the coverage of the empirical mortality rates for both males and females, by posterior predictive intervals at 80%, 95% and 99% probability levels for each of

the models. For the in-sample data, the posterior predictive intervals were slightly too narrow for models 1, 2 and 4 and slightly too conservative for models 3, 5, 6 and 7.

Chapter 5 also described the methodology for producing forecasts of mortality rates. For the out-of-sample empirical mortality rates, the calculation of the posterior predictive intervals involved forecasts of future mortality. The form of the smoothing prior was an important factor influencing the mortality forecasts as seen in the coverage of the empirical rates with the posterior predictive intervals. For the better-fitting models, models labelled 5, 6 and 7, posterior predictive intervals at probability levels 80%, 95% and 99% were obtained for both males and females. For males, Models 5 and 6 with smoothing following a random walk on the levels gave good agreement and smoothing following an AR(1) on first differences gave intervals that were slightly conservative. For Model 7 the intervals were too wide with both of these forms of smoothing. For females, there was a similar story but the intervals produced by Model 7 were not out of line with the coverage of the intervals produced from Models 5 and 6, as was the case with the males. In all models where smoothing followed a random walk on the first differences of the parameters the posterior predictive intervals were very wide and so their value for providing forecasts is very limited.

Chapter 6 demonstrated the value and the straightforward use of Bayesian models in analysing risks and uncertainties. Based on the DIC results, models labelled 6 were selected to produce forecasts. Forecasts of mortality rates were obtained using the three versions of Model 6, (i.e., models labelled 6(a), 6(b) and 6(c)) in order to investigate some of the important issues involving longevity raised in the introduction. The first issue was how life expectancy might develop for males and females in future years. The results of the models indicated that period life expectancy would increase for both males and females until the 2020 decade for ages 60 up to 75 and until the 2030 decade for older ages; at these times life expectancy levels off. Interestingly, for age 60, the median forecasts for life expectancy are similar for each of the smoothing priors but the range of future values does vary noticeably; this is

particularly so where smoothing follows a random walk on the first differences of parameters. The median forecasts of period life expectancy produced by our models are lower than official ONS forecasts, but the ONS forecasts are within our 95% posterior predictive intervals. Figures for median forecast cohort life expectancy for males and females do not change by much from 2020; at age 60 cohort life expectancies for males and females are 83 and 86, respectively. Corresponding ONS forecasts are noticeably higher, at around 87 for males and 91 for females. The ONS forecast methodology is described in chapter 3 and these figures show that the ONS are assuming that mortality improvements will occur at a higher rate than in our models.

Chapter 6 also discussed the proposed EU solvency II regulations that apply to longevity risk. The solvency rules require companies to hold solvency capital equivalent to a 1-year 99.5% VaR. As an approximation to this solvency rule companies are permitted to use a permanent 25% decrease in mortality rates to determine the solvency capital. Based on the results from Model 6 (the other models labelled 5 and 7 would give a similar result) a 25% decrease in mortality results in solvency capital being an increasing proportion of the liability as the age increases. Therefore, perhaps an age-related reduction rather than the flat 25% would give a more consistent capital requirement over ages, this is consistent with Richards et al. (2013). If the flat 25% reduction is not used then it is not obvious how to apply the 1-year 99.5% VaR capital requirement in the context of long-term contracts. This results in possible issues of consistency between different companies for capital measurement and this thesis suggests that applying a different VaR measure, one with a lower probability, say 90% to 95%, over the term of the liability rather than one year may give better consistency of measurement between companies.

Finally, chapter 6 concluded by forecasting the number of centenarians in England and Wales up until 2030, as an illustration of how these models can assist in the quantification of long-term care costs. The median projection from Models 6(a) and 6(b) indicated that the numbers of male centenarians will continue to increase with time from about 2000 in 2010 to around 5000 by 2030. For females the increase is proportionately less marked, the number of centenarians increasing from around

10,000 in 2010 to around 15,000 in 2030. The posterior predictive intervals at 2030 are very wide, reflecting the uncertainty in the forecasts.

In conclusion, the new models introduced in this thesis are flexible models that can be used to analyse and forecast human mortality. Also, as illustrated in chapter 6, which applied the Bayesian mortality models to important longevity issues, the Bayesian approach is seen to be a very powerful tool for analysing uncertainty and risk.

Future Research

All of the models had weaknesses and future research could help address them.

- Consider modelling males and females jointly so that the mortality improvements processes are related; for example, one could assume that the male and female period parameters are cointegrated and male and female cohort processes are cointegrated. So, in the case of the period parameters, suppose that κ_t^M, κ_t^F denote the male and female parameters for year t, and are non-stationary, integrated of order 1. Then a linear combination of κ_t^M and κ_t^F is stationary, i.e., $\kappa_t^F - c\kappa_t^M = u_t$ where u_t is stationary. Then the parameters of the model would include c and, if we assume that u_t is Gaussian, the precision of u_t would be an additional parameter. The female period parameters would then be derived from the male period parameters. This prior structure should ensure that the male and female mortality improvements are prevented from drifting arbitrarily far from each other. This should remove the problem with the current models that, at the very oldest ages 90 and over, for very long-term forecasts, male and female mortality drift away from each other in an implausible way.

- Consider more general smoothing algorithms for the age, period and cohort parameters where each follows an independent ARIMA(p, d, q) process where p, d, q and the AR and MA coefficients as hyper-parameters in the model.

- Consider the inclusion of additional covariates. For example, include the rates of death from the major causes of death as covariates. The causes of death will

be according to International Classification of Diseases e.g., circulatory diseases, respiratory diseases etc. There will be some practical issues to be dealt with; for example, the classification and definitions of diseases have changed over time but using broad groupings should reduce this issue. Another point requiring attention is that the rates by cause of death are not by specific age but by age groupings.

- Consider a more complex model for the errors in population estimates, for example by increasing the variance of the error depending on how far the date of the population estimate is from a census date. Another possibility would be to assume that the errors in population size vary by age on the basis that at the oldest ages individuals may not remember ages accurately.

APPENDIX A

Period & Cohort Life Expectancy

Based on the notation in chapter 2, the formulae 2.1.6 for *period life expectancy* and 2.1.7 for *cohort life expectancy* are derived as follows.

The general definition of period life expectancy is shown in equation A.0.1 below. In practice it is common to set an upper age limit, indicated by ω, as 110 or 120 or some other age; in this thesis 120 was chosen. Equation A.0.2 is equation A.0.1 divided into individual ages with the age limit ω imposed:

$$e^{p}_{x,T} = \int_0^{\infty} S_x(r,T)dr \tag{A.0.1}$$

$$e^{p}_{x,T} = \frac{(1-\exp(-m_{x,T}))}{m_{x,T}} + \sum_{u=1}^{\omega-1}\prod_{r=0}^{u-1}(1-q_{x+r,T})\frac{(1-\exp(-m_{x+u,T}))}{m_{x+u,T}}. \tag{A.0.2}$$

Adopting the assumption that the force of mortality is constant over each age and calendar year as in chapter 2, we can then derive the following results, A.0.3, A.0.4 and A.0.5, that are useful in deriving equation 2.1.6:

$$\begin{aligned}
\int_0^t \mu_{x+r,T}dr &= -[\log S_x(r,T)]_0^t \quad t \leqslant 1 \\
\mu_{x,T}\int_0^t dr &= \log S_x(0,T) - \log S_x(t,T) \\
\mu_{x,T}t &= \log 1 - \log S_x(t,T) \\
S_x(t,T) &= \exp(-t\mu_{x,T}),
\end{aligned} \tag{A.0.3}$$

$$\begin{aligned}\int_0^1 S_x(r,T)dr &= \int_0^1 \exp(-r\mu_{x,T})dr \\ &= [-\frac{1}{\mu_{x,T}}\exp(-r\mu_{x,T})]_0^1 \\ &= \frac{1}{\mu_{x,T}}(1-\exp(-\mu_{x,T})) \\ &= \frac{1}{m_{x,T}}(1-\exp(-m_{x,T})), \end{aligned} \tag{A.0.4}$$

$$\begin{aligned}S_x(r,T) &= \prod_{u=0}^{r-1} S_{x+u}(1,T) \\ &= \prod_{u=0}^{r-1}(1-q_{x+u,T}). \end{aligned} \tag{A.0.5}$$

We now derive equation 2.1.6:

$$\begin{aligned}e^p_{x,T} &= \int_0^{\omega} S_x(r,T)dr \\ &= \int_0^1 S_x(r,T)dr + \sum_{u=1}^{\omega-1}\{S_x(u,T)\int_0^1 S_x(u+r,T)dr\} \\ &= \frac{(1-\exp(-m_{x,T}))}{m_{x,T}} + \sum_{u=1}^{\omega-1}\prod_{r=0}^{u-1}(1-q_{x+r,T})\frac{(1-\exp(-m_{x+u,T}))}{m_{x+u,T}}. \end{aligned}$$

The derivation of equation 2.1.7 for cohort life expectancy follows the derivation of 2.1.6 with the adjustment that the mortality rates do not relate to a single calendar year but to the mortality rate in the particular year during which each age is attained.

APPENDIX B

Model Specifications

General Model:

$$d_{x,t} \sim Poi(E_{x,t}m_{x,t}) \qquad x = x_1, \ldots, x_{n_a};\ \ t = t_1, \ldots, t_{n_y}. \tag{B.0.1}$$

B.1 Model 1

In this model

$$\log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t. \tag{B.1.1}$$

The likelihood is given by

$$L(D|\mu, \alpha, \beta, \kappa) = \prod_{x,t} \frac{[E_{x,t}m_{x,t}]^{d_{x,t}} e^{-E_{x,t}m_{x,t}}}{d_{x,t}!}. \tag{B.1.2}$$

where D denotes the matrix of death counts $[d_{x,t}]$.

The Metropolis-Hastings algorithm is used to make draws of each parameter, μ, α_x, β_x and κ_t from the posterior. Proposal values were generated from the current value by adding a disturbance from a Gaussian distribution with mean zero and variance chosen for each parameter so as to ensure good mixing.

B.1.1 Model 1(0)

For Model 1(0), weakly informative priors are used for the parameters. Defining $\mu_\alpha = \mu_\beta = (0, \ldots, 0)'$ of length n_a, $\mu_\kappa = (0, \ldots, 0)'$ of length n_y and $\sigma_\alpha^2 = \sigma_\beta^2 = \sigma_\kappa^2 = 1000$, we have the following priors:

$$p(\mu) \propto constant\,, \tag{B.1.3}$$

$$p(\alpha) \propto \exp(-\frac{1}{2\sigma_\alpha^2}(\alpha - \mu_\alpha)'\mathbb{I}_{n_a}(\alpha - \mu_\alpha)), \tag{B.1.4}$$

$$p(\beta) \propto \exp(-\frac{1}{2\sigma_\beta^2}(\beta - \mu_\beta)'\mathbb{I}_{n_a}(\beta - \mu_\beta)), \tag{B.1.5}$$

$$p(\kappa) \propto \exp(-\frac{1}{2\sigma_\kappa^2}(\kappa - \mu_\kappa)'\mathbb{I}_{n_y}(\kappa - \mu_\kappa)). \tag{B.1.6}$$

Sections B.1.2 to B.1.4 illustrate the various smoothing priors for α. Similar formulae apply to κ and β and their respective variance hyper-parameters λ and τ.

B.1.2 Model 1(a)

The smoothing priors follow a random walk on the parameters and for α we have the following prior:

$$\alpha_x - \alpha_{x-1} \sim N(0, \omega^{-1}) \text{ for } x = 1, \ldots, n_a. \tag{B.1.7}$$

The priors on α and ω are as expressed in equations (B.1.8) and (B.1.9):

$$p(\alpha|\omega) \propto \exp(-\frac{\omega}{2}\alpha'\Sigma_\alpha\alpha) \tag{B.1.8}$$

$$p(\omega) \propto \exp(-a\omega),\ a = 2000^{-1}, \tag{B.1.9}$$

where

$$\Sigma_\alpha = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & -1 & 2 & -1 & \\ & & & & -1 & 1 \end{bmatrix}_{n_a \times n_a}.$$

The full conditional of $\omega|...$ can be expressed as a gamma distribution:

$$\begin{aligned} \pi(\omega|\cdots) &\propto \omega^{n_a/2} \cdot \exp\left(-\frac{\omega}{2}\alpha'\Sigma_\alpha\alpha\right) \cdot \exp(-a\omega), \\ \text{so that} \quad \omega|\cdots &\sim Gamma\left(1 + n_a/2, a + \frac{1}{2}\alpha'\Sigma_\alpha\alpha\right). \end{aligned} \tag{B.1.10}$$

B.1.3 Model 1(b)

The smoothing priors follow an autoregressive process of order 1 on the first differences of the parameters and for α this can be expressed as

$$\begin{aligned} \Delta\alpha_x - \psi\Delta\alpha_{x-1} &\sim N(0, \omega^{-1}), \\ \alpha_x - (1+\psi)\alpha_{x-1} + \psi\alpha_{x-2} &\sim N(0, \omega^{-1}). \end{aligned} \tag{B.1.11}$$

The prior on ω is the same as for Model 1(a) (equation (B.1.9)). However, the priors on α and ψ are

$$p(\alpha|\omega, \psi) \propto \exp\left(-\frac{\omega}{2}\alpha'\Sigma_{\alpha(\psi)}\alpha\right), \tag{B.1.12}$$

$$\psi \sim Un(-1, 1). \tag{B.1.13}$$

Here,

$$\Sigma_{\alpha(\psi)} = \begin{bmatrix} \psi^2 & -\psi(1+\psi) & \psi & & & & \\ -\psi(1+\psi) & \psi^2+(1+\psi)^2 & -(1+\psi)^2 & \psi & & & \\ \psi & -(1+\psi)^2 & 1+\psi^2+(1+\psi)^2 & -(1+\psi)^2 & \psi & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & \psi & -(1+\psi)^2 & 1+\psi^2+(1+\psi)^2 & -(1+\psi)^2 & \psi \\ & & & \psi & -(1+\psi)^2 & 1+(1+\psi)^2 & -(1+\psi) \\ & & & & \psi & -(1+\psi) & 1 \end{bmatrix}_{n_a \times n_a}.$$

The Metropolis-Hastings algorithm is used to make draws of ψ. The proposal value ($\tilde{\psi}$) was generated from the current value (ψ) using the following transformation: $q = \log\frac{1+\psi}{1-\psi}$ (i.e. $q \in (-\infty, \infty)$) so that $\psi = \frac{\exp(q)-1}{1+\exp(q)}$. A new $\tilde{q}$ was generated by

adding a disturbance, i.e. $\tilde{q} = q + \epsilon$, where ϵ is a draw from $N(0, ss)$, in which ss is chosen to ensure good mixing. Thus the proposal density is $\pi(\tilde{\psi}|\psi) = \pi(\tilde{q}|q)\frac{2}{(1+\tilde{\psi})(1-\tilde{\psi})}$. Hence, the acceptance ratio is

$$\frac{\exp\{-\frac{\omega}{2}\alpha'\Sigma_{\alpha(\tilde{\psi})}\alpha\}(1+\tilde{\psi})(1-\tilde{\psi})}{\exp\{-\frac{\omega}{2}\alpha'\Sigma_{\alpha(\psi)}\alpha\}(1+\psi)(1-\psi)}. \tag{B.1.14}$$

The full conditional of ω is the same as in equation (B.1.10) but with $\Sigma_{\alpha(\psi)}$ replacing Σ_{α}.

B.1.4 Model 1(c)

The smoothing priors follow a random walk on the first differences of the parameters and for α we have the following prior:

$$\begin{aligned} \Delta\alpha_x - \Delta\alpha_{x-1} &\sim N(0, \omega^{-1}), \\ \alpha_x - 2\alpha_{x-1} + \alpha_{x-2} &\sim N(0, \omega^{-1}). \end{aligned} \tag{B.1.15}$$

The priors on α and ω can be expressed as

$$p(\alpha|\omega) \propto \exp(-\frac{\omega}{2}\alpha'\Sigma_{\alpha}\alpha), \tag{B.1.16}$$

$$p(\omega) \propto \exp(-a\omega), \; a = 2000^{-1}, \tag{B.1.17}$$

where

$$\Sigma_{\alpha} = \begin{bmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & & \ddots & \ddots & & & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{bmatrix}_{n_a \times n_a}.$$

The full conditional of ω can be expressed as a gamma distribution:

$$\begin{aligned} \pi(\omega|\cdots) &\propto \omega^{n_a/2} \cdot \exp\{-\frac{\omega}{2}\alpha'\Sigma_\alpha\alpha\} \cdot \exp(-a\omega), \\ \text{so that} \quad \omega|\cdots &\sim Gamma(1+n_a/2, a+\frac{1}{2}\alpha'\Sigma_\alpha\alpha). \end{aligned} \tag{B.1.18}$$

B.2 Model 2

In this model

$$\log m_{x,t} = \mu + \alpha_x + \kappa_t + \gamma_{t-x}. \tag{B.2.1}$$

The likelihood for Model 2 is the same as in equation (B.1.2) but with $\log m_{x,t}$ defined as in equation (B.2.1).

The Metropolis-Hastings algorithm is used to make draws of each parameter, μ, α_x, κ_t and γ_{t-x}, from the posterior. Proposal values were generated from the current value by adding a disturbance from a Gaussian distribution with mean zero and variance chosen for each parameter so as to ensure good mixing.

B.2.1 Model 2(0)

For Model 2(0), the same weakly informative priors are used for the parameters μ, α, κ as in Model 1(0); see equations (B.1.3, B.1.4, B.1.6), respectively. If we define $\mu_\gamma = (0,\ldots,0)'$ of length n_c and $\sigma_\gamma^2 = 1000$, the prior on γ is

$$p(\gamma) \propto \exp\left(-\frac{1}{2\sigma_\gamma^2}(\gamma-\mu_\gamma)'\mathbb{I}_{n_c}(\gamma-\mu_\gamma)\right). \tag{B.2.2}$$

independently of μ, α, κ.

B.2.2 Models 2(a), 2(b) and 2(c)

The priors for μ, α and κ are the same as in the corresponding Model 1(a), 1(b) and 1(c). The prior on the new parameter γ has the same structure as that for α in the corresponding Model 1 (e.g. Model 2(a) has equation B.1.8) but with matrix

dimension $n_c \times n_c$. Smoothing of γ is independent of the other parameters and has precision hyper-parameter ν. In Model 2(b), where γ is assumed to follow an autoregressive process of order 1, the coefficient is ϕ. The prior distributions of ν and ϕ are similar to those of ω and ψ in Model 1(b); see equations (B.1.9) and (B.1.13).

B.3 Model 3

In this model

$$\xi_{x,t} = \log m_{x,t} = \mu + \alpha_x + \kappa_t + \gamma_{t-x} + z_{x,t}. \tag{B.3.1}$$

B.3.1 Model 3(0)

For Model 3(0), the same weakly informative priors are used for the parameters μ, α, κ and γ as in Model 2(0); see equations (B.1.3), (B.1.4), (B.1.6) and (B.2.2), respectively. The prior on the random effects $z_{x,t}$ is $N(0, \zeta^{-1})$. If Z denotes the collection of all the random effects, and $\pi(\zeta)$ denotes the prior on the precision ζ, then

$$\pi(Z|\zeta) \propto \zeta^{n_a n_y/2} \exp\left(-\frac{\zeta}{2} \sum_{x=1}^{n_a} \sum_{t=1}^{n_y} z_{x,t}^2\right), \tag{B.3.2}$$

$$\pi(\zeta) \propto \exp(-d\zeta), \quad d = 2000^{-1}. \tag{B.3.3}$$

B.4 Models 3(a), 3(b) and 3(c)

The priors for μ, α, κ and γ are the same as in the corresponding Model 2(a), 2(b) and 2(c). The prior on the set of random effects $z_{x,t}$ and precision ζ are as in equations (B.3.2) and (B.3.3).

If we use the notation Σ_α, Σ_κ and Σ_γ to represent the respective matrices of the smoothing algorithm for α, κ and γ, as detailed in Models 1(a), 1(b) and 1(c), the following derivation of full conditional distributions can then be applied to each of the different smoothing models.

The likelihood for Model 3 is the same as in equation (B.1.2) but with $\log m_{x,t}$ defined as in equation (B.3.1).

The joint posterior is therefore

$$\begin{aligned}\pi(\mu,\alpha,\kappa,\gamma,z,\omega,\lambda,\nu,\zeta,\psi,\rho,\phi|D) \;\propto\; & \prod_{x,t}\exp\{-E_{x,t}\exp(\xi_{x,t})\}\exp(d_{x,t}\xi_{x,t}) \\ & \cdot\; \omega^{n_a/2}\cdot\exp\{-\frac{\omega}{2}\alpha^{'}\Sigma_\alpha\alpha\} \\ & \cdot\; \lambda^{n_y/2}\exp\{-\frac{\lambda}{2}\kappa^{'}\Sigma_\kappa\kappa\} \\ & \cdot\; \nu^{n_c/2}\exp\{-\frac{\nu}{2}\gamma^{'}\Sigma_\gamma\gamma\} \\ & \cdot\; \zeta^{n_a n_y/2}\exp\{-\frac{\zeta}{2}\sum_{x,t}z_{x,t}^2\} \\ & \cdot\; \exp(-a\omega)\;\cdot\;\exp(-b\lambda) \\ & \cdot\; \exp(-c\nu)\;\cdot\;\exp(-d\zeta). \end{aligned} \tag{B.4.1}$$

By re-expressing equation (B.3.1) in terms of $z_{x,t}$ and treating $\xi_{x,t}$ as a parameter, we have:

$$z_{x,t} = \xi_{x,t} - \mu - \alpha_x - \kappa_t - \gamma_{t-x}. \tag{B.4.2}$$

The full conditionals for the parameters can be expressed as standard distributions by first considering the term $\sum_{x,t} z_{x,t}^2 = \sum_{x,t}(\xi_{x,t} - \mu - \alpha_x - \kappa_t - \gamma_{t-x})^2$, for each parameter α, κ, γ and μ in turn.

In terms of the age-related parameter α, $\exp\{\sum_{x,t} z_{x,t}^2\}$ can be written as

$$\begin{aligned}\exp\{\sum_{x,t} z_{x,t}^2\} &= \exp\{\sum_{x,t}(\alpha_x^2 - 2\alpha_x w_{x,t} + w_{x,t}^2)\} \\ &= \exp\{n_y(\sum_x(\alpha_x^2 - 2\alpha_x\sum_t w_{x,t}/n_y + \sum_t w_{x,t}^2/n_y))\} \\ &\propto \exp\{n_y(\alpha^{'}\alpha - 2\alpha^{'}W)\},\end{aligned}$$

where $w_{x,t} = \xi_{x,t} - \mu - \kappa_t - \gamma_{t-x}$ and $W = (W_1,\ldots,W_{n_a})^{'}$, $W_x = \sum_t w_{x,t}/n_y$.

Then the full conditional for α can then be expressed as

$$\begin{aligned} \pi(\alpha|\cdots) &\propto \exp\{-\frac{\omega}{2}\alpha'\Sigma_\alpha\alpha - \frac{n_y\zeta}{2}(\alpha'\alpha - 2\alpha'W)\} \\ &\propto \exp\{-\frac{1}{2}(\alpha - m_\alpha)'(\omega\Sigma_\alpha + n_y\zeta\mathbb{I}_{n_a})(\alpha - m_\alpha)\}, \\ \text{so that} \quad \alpha|\cdots &\sim N_{n_a}(m_\alpha, (\omega\Sigma_\alpha + n_y\zeta\mathbb{I}_{n_a})^{-1}), \end{aligned} \tag{B.4.3}$$

where $m_\alpha = n_y\zeta(\omega\Sigma_\alpha + n_y\zeta\mathbb{I}_{n_a})^{-1}W$. Thus $\pi(\alpha|\cdots)$ is a multivariate Gaussian distribution.

In terms of the period-related parameter κ, $\exp\{\sum_{x,t} z_{x,t}^2\}$ can be written as

$$\begin{aligned} \exp\{\sum_{x,t} z_{x,t}^2\} &= \exp\{\sum_{x,t}(\kappa_t^2 - 2\kappa_t u_{x,t} + u_{x,t}^2)\} \\ &= \exp\{n_a(\sum_t(\kappa_t^2 - 2\kappa_t\sum_x u_{x,t}/n_a - \sum_x u_{x,t}^2/n_a))\} \\ &\propto \exp\{n_a(\kappa'\kappa - 2\kappa'U)\}, \end{aligned}$$

where $u_{x,t} = (\xi_{x,t} - \mu - \alpha_x - \gamma_{t-x})$ and $U = (U_1, \ldots, U_{n_y})'$, $U_t = \sum_x u_{x,t}/n_a$.

The full conditional on κ can then be expressed as

$$\begin{aligned} \pi(\kappa|...) &\propto \exp\{-\frac{\lambda}{2}\kappa'\Sigma_\kappa\kappa - \frac{n_a\zeta}{2}(\kappa'\kappa - 2\kappa'U)\} \\ &\propto \exp\{-\frac{1}{2}(\kappa - m_\kappa)'(\lambda\Sigma_\kappa + n_a\zeta\mathbb{I}_{n_y})(\kappa - m_\kappa)\} \\ \text{so that} \quad \kappa|... &\sim N_{n_y}(m_\kappa, (\lambda\Sigma_\kappa + n_a\zeta\mathbb{I}_{n_y})^{-1}), \end{aligned} \tag{B.4.4}$$

where $m_\kappa = n_a\zeta(\lambda\Sigma_\kappa + n_a\zeta\mathbf{I}_{n_y})^{-1}U$.

Thus $\pi(\kappa|\cdots)$ is a multivariate Gaussian distribution.

In formulating the full conditional for the cohort-related parameter γ, the random effect parameters $z_{x,t}$ can be written in matrix form as Z, an $n_a \times n_y$ matrix whose $(x,t)^{th}$ entry is $z_{x,t}$. If we call V an $n_a \times n_y$ matrix whose $(x,t)^{th}$ entry is $v_{x,t}$ where $v_{x,t} = (\xi_{x,t} - \mu - \alpha_x - \kappa_t)$, the matrix Z can be expressed as

$$\begin{bmatrix} v_{1,1}-\gamma_{n_a} & & & v_{1,n_y}-\gamma_{n_a+t-1} & v_{1,n_y}-\gamma_{n_a+n_y-1} \\ & \ddots & & \ddots & \\ & & v_{x,x}-\gamma_{n_a} & & \vdots \\ v_{n_a,1}-\gamma_1 & & & & v_{n_a,n_y}-\gamma_{n_y} \end{bmatrix}_{n_a \times n_y}.$$

If we define the k^{th} diagonal of a matrix as the k^{th} diagonal from the bottom left corner (cell $(n_a, 1)$ to the top right corner (cell $(1, n_y)$) then k can take the values $1, \ldots, (n_a + n_y - 1)$. Then $\exp\{\sum_{x,t} z_{x,t}^2\}$ can be written as

$$\begin{aligned} \exp\{\sum_{x,t} z_{x,t}^2\} &= \exp\{\sum_{x,t}(\gamma_{n_a+t-x}^2 - 2\gamma_{n_a+t-x} v_{x,t} + v_{x,t}^2)\} \\ &\propto \exp\{\gamma^{'} T\gamma - 2\gamma^{'}\tilde{V}\}, \end{aligned}$$

where $\tilde{V}$ is an $(n_a + n_y - 1) \times 1$ matrix whose i^{th} entry is the sum of the i^{th} diagonal of V. Also, T is a $(n_a + n_y - 1)$ diagonal matrix,
$T = diagonal(1, 2, 3, \ldots, \min(n_a, n_y), \ldots, \min(n_a, n_y), (\min(n_a, n_y) - 1), \ldots, 3, 2, 1)$.

The full conditional on γ can then be expressed as

$$\begin{aligned} \pi(\gamma|\cdots) &\propto \exp\{-\frac{\nu}{2}\gamma^{'}\Sigma_\gamma\gamma - \frac{\zeta}{2}(\gamma^{'}T\gamma - 2\gamma^{'}\tilde{V})\} \\ &\propto \exp\{-\frac{1}{2}(\gamma - m_\gamma)^{'}(\nu\Sigma_\gamma + \zeta T)(\gamma - m_\gamma)\}, \\ \gamma|\cdots &\sim N_K(m_\gamma, (\nu\Sigma_\gamma + \zeta T)^{-1}), \end{aligned} \tag{B.4.5}$$

where $m_\gamma = \zeta(\nu\Sigma_\gamma + \zeta T)^{-1}\tilde{V}$.

Thus $\pi(\gamma|...)$ is a multivariate Gaussian distribution.

The full conditional of μ is

$$\begin{aligned} \pi(\mu|...) &\propto \exp\{-\frac{\zeta}{2}\sum_{x,t} z_{x,t}^2\} \\ &\propto \exp\{-\frac{\zeta}{2}\sum_{x,t}(\mu - (\xi_{x,t} - \alpha_x - \kappa_t - \gamma_{t-x}))^2\} \\ &\propto \exp\{-\frac{\zeta n_a n_y}{2}(\mu - Q)^2\}, \\ \text{so that} \quad \mu|... &\sim N(Q, \eta^2), \end{aligned} \tag{B.4.6}$$

where $Q = \sum_{x,t}(\xi_{x,t} - \alpha_x - \kappa_t - \gamma_{t-x})/(n_a n_y)$ and $\eta^2 = 1/(n_a n_y \zeta)$.

For ρ, ϕ and ψ draws are generated using a Metropolis-Hastings algorithm based on the same prior and moves as for ψ in Model 1.

The full conditionals of ω, λ and ν are as for Models 2(a), 2(b) and 2(c). The full conditional of ζ is derived in a similar way and is also a gamma distribution:

$$\zeta|\cdots \sim Gamma\left(1 + n_a n_y/2, d + \frac{1}{2}\sum_{x,t} z_{x,t}^2\right). \tag{B.4.7}$$

The full conditional for each $\xi_{x,t}$ is a non-standard distribution and is sampled using a MH move with a proposal from a random walk centred on the current parameter value with variance set so as to give good mixing:

$$\pi(\xi_{x,t}|\cdots) \propto \exp\{-E_{x,t}\exp(\xi_{x,t}) - \frac{\zeta}{2}z_{x,t}^2)\}\exp(d_{x,t}\xi_{x,t}). \tag{B.4.8}$$

B.5 Model 4

In this model

$$\log m_{x,t} = \mu + \alpha_x + \beta_x\kappa_t + \delta_x\gamma_{t-x}. \tag{B.5.1}$$

The MH algorithm is used to make draws of each parameter, μ, α_x, β_x, κ_t, δ_x and γ_{t-x}, from the posterior. Proposal values were generated from the current value by adding a disturbance from a Gaussian distribution with mean zero and variance chosen for each parameter so as to ensure good mixing.

B.5.1 Model 4(0)

For Model 4(0), the same weakly informative priors are used for the parameters μ, α, κ and γ as in Model 2(0), see equations (B.1.3),(B.1.4),(B.1.6) and (B.2.2), respectively, and for β as in Model 1(0) equation (B.1.5). If we define $\mu_\delta = (0,\ldots,0)'$ of length n_a and σ_δ^2 = 1000 , the prior on δ is

$$p(\delta) \propto \exp\left(-\frac{1}{2\sigma_\delta^2}(\delta - \mu_\delta)'\mathbb{I}_{n_a}(\delta - \mu_\delta)\right). \tag{B.5.2}$$

B.5.2 Models 4(a), 4(b) and 4(c)

The priors for μ, α, κ and γ are the same as those in the corresponding Models 2(a), 2(b) and 2(c). The priors on the new parameter δ and also for β have the same structure as that for α in the corresponding Model 1's (equations (B.1.8), (B.1.12) and (B.1.16)). Smoothing of δ is independent of the other parameters with the precision hyper-parameter ι. In Model 4(b), where δ is assumed to follow an autoregressive process of order 1, the coefficient is ζ. The prior distributions of ι and ζ are similar to those for ω and ψ respectively in Model 1(b); see equations (B.1.9) and (B.1.13)).

B.6 Model 5

In this model

$$\xi_{x,t} = \log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x} + z_{x,t}. \tag{B.6.1}$$

B.6.1 Models 5(a), 5(b) and 5(c)

This model is similar to Models 4(a), 4(b) and 4(c) and the priors on the common parameters and hyper-parameters are the same. The priors on $z_{x,t}$ and ζ are the same as for Model 3.

If we use the notation Σ_β, Σ_κ, δ and Σ_γ to represent the appropriate matrices of the smoothing algorithm, then, just as in Models 3(a), 3(b) and 3(c), full conditional distributions for the parameters can be derived using the same trick as used with Models 3.

We treat $z_{x,t}$ as a function of the other parameters, writing $\sum_{x,t} z_{x,t}^2 = \sum_{x,t} (\xi_{x,t} - \mu - \alpha_x - \beta_x \kappa_t - \delta_x \gamma_{t-x})^2$.

The full conditional on α is a multivariate Gaussian distribution:

$$\alpha | \cdots \sim N_{n_a}(m_\alpha, (\omega \Sigma_\alpha + n_y \zeta \mathbb{I}_{n_a})^{-1}), \tag{B.6.2}$$

where $m_\alpha = n_y\zeta(\omega\Sigma_\alpha + n_y\zeta\mathbb{I}_{n_a})^{-1}W$, $W = (W_1, \ldots, W_{n_a})'$, $W_x = \sum_t w_{x,t}/n_y$ and $w_{x,t} = (\xi_{x,t} - \mu - \beta_x\kappa_t - \delta_x\gamma_{t-x})$.

The full conditional on κ is a multivariate Gaussian:

$$\kappa|\cdots \sim N_{n_y}(\zeta(\sum_x \beta_x^2)(\lambda\Sigma_\kappa + \zeta(\sum_x \beta_x^2)\mathbb{I}_{n_y})^{-1}U, (\lambda\Sigma_\kappa + \zeta(\sum_x \beta_x^2)\mathbb{I}_{n_y})^{-1}), \quad \text{(B.6.3)}$$

where $U = (U_1, \ldots, U_{n_y})'$, $U_t = \frac{1}{\sum_x \beta_x^2}\sum_x \beta_x u_{x,t}$ and $u_{x,t} = (\xi_{x,t} - \mu - \alpha_x - \delta_x\gamma_{t-x})$.

The full conditional on β is a multivariate Gaussian:

$$\beta|\cdots \sim N_{n_a}(\zeta(\sum_t \kappa_t^2)(\tau\Sigma_\beta + \zeta(\sum_t \kappa_t^2)\mathbb{I}_{n_a})^{-1}\tilde{U}, (\tau\Sigma_\beta + \zeta(\sum_t \kappa_t^2)\mathbb{I}_{n_a})^{-1}), \quad \text{(B.6.4)}$$

where $u_{x,t} = (\xi_{x,t} - \mu - \alpha_x - \delta_x\gamma_{t-x})$, $\tilde{U}_x = \frac{1}{\sum_t \kappa_t^2}\sum_t U_{x,t}$ and $\tilde{U} = (\tilde{U}_1, \ldots, \tilde{U}_{n_a})'$.

In formulating the full conditional for the cohort-related parameter γ, if we call V an $n_a \times n_y$ matrix whose $(x,t)^{th}$ entry is $v_{x,t}$ where $v_{x,t} = (\xi_{x,t} - \mu - \alpha_x - \beta_x\kappa_t)$, the matrix $[z_{x,t}]$ can be written as

$$\begin{bmatrix} v_{1,1} - \delta_1\gamma_{n_a} & & v_{1,n_y} - \delta_1\gamma_{n_a+t-1} & v_{1,n_y} - \delta_1\gamma_{n_a+n_y-1} \\ \ddots & & \ddots & \\ & v_{x,x} - \delta_x\gamma_{n_a} & & \vdots \\ & & & \\ v_{n_a,1} - \delta_{n_a}\gamma_1 & & & v_{n_a,n_y} - \delta_{n_a}\gamma_{n_y} \end{bmatrix}_{n_a \times n_y}.$$

If we again define the k^{th} diagonal of a matrix as the k^{th} diagonal from the bottom left corner (cell $(n_a, 1)$ to the top right corner (cell $(1, n_y)$) then k can take the values $1, \ldots, (n_a + n_y - 1)$. Then $\exp\{\sum_{x,t} z_{x,t}^2\}$ can be written as

$$\begin{aligned} \exp\{\sum_{x,t} z_{x,t}^2\} &= \exp\{\sum_{x,t}(\delta_x^2\gamma_{n_a+t-x}^2 - 2\delta_x\gamma_{n_a+t-x}v_{x,t} + v_{x,t}^2)\} \\ &\propto \exp\{\gamma' T\gamma - 2\gamma'\tilde{V}\}, \end{aligned}$$

where $\tilde{V}$ is an $(n_a + n_y - 1) \times 1$ matrix whose i^{th} entry is the sum of the i^{th} diagonal of $A * V$. Where A is an $n_a \times n_y$ matrix whose $(x,t)^{th}$ element is δ_x and $*$ is element-

wise multiplication of A and V. Here T is a $(n_a + n_y - 1)$ diagonal matrix, $T = diag(\delta^2_{n_a}, \delta^2_{na} + \delta^2_{n_a-1}, \delta^2_{n_a} + \delta^2_{n_a-1} + \delta^2_{n_a-2}, \ldots, \delta^2_2 + \delta^2_1, \delta^2_1)$.

The full conditional on γ can therefore be expressed as

$$\begin{aligned} \pi(\gamma|\cdots) &\propto \exp\left(-\frac{\nu}{2}\gamma'\Sigma_\gamma\gamma - \frac{\zeta}{2}(\gamma' T\gamma - 2\gamma'\tilde{V})\right) \\ &\propto \exp\left(-\frac{1}{2}(\gamma - m_\gamma)'(\nu\Sigma_\gamma + \zeta T)(\gamma - m_\gamma)\right), \\ \gamma|\cdots &\sim N_{n_c}\left(m_\gamma, (\nu\Sigma_\gamma + \zeta T)^{-1}\right), \end{aligned} \tag{B.6.5}$$

where $m_\gamma = \zeta(\nu\Sigma_\gamma + \zeta T)^{-1}\tilde{V}$.

Thus $\pi(\gamma|...)$ is a multivariate normal.

In terms of the parameter δ, $\exp\{\sum_{x,t} z^2_{x,t}\}$ can be written as

$$\begin{aligned} \exp\{\sum_{x,t} z^2_{x,t}\} &= \exp\{\sum_{x,t}(\delta^2_x\gamma^2_{n_a+t-x} - 2\delta_x\gamma_{n_a+t-x}v_{x,t} + v^2_{x,t})\} \\ &\propto \exp\{\sum_x(\delta^2_x\sum_t\gamma^2_{n_a+t-x} - 2\delta_x\sum_t\gamma_{n_a+t-x}v_{x,t})\} \\ &\propto \exp\{\delta' S\delta - 2\delta' R\}, \end{aligned}$$

where $v_{x,t} = \xi_{x,t} - \mu - \alpha_x - \beta_x\kappa_t$, $r_{x,t} = \gamma_{n_a+t-x}v_{x,t}$ and $R = (\sum_t r_{1,t}, \ldots, \sum_t r_{n_a,t})'$. $S_x = \sum_t \gamma^2_{n_a+t-x}$, and $S = diag(S_1, \ldots, S_{n_a})$.

Hence, the full conditional of δ can be expressed as

$$\begin{aligned} \pi(\delta|\cdots) &\propto \exp\{-\frac{\iota}{2}\delta'\Sigma_\delta\delta - \frac{\zeta}{2}(\delta' S\delta - 2\delta' R)\} \\ &\propto \exp\{-\frac{1}{2}(\delta - m_\delta)'(\iota\Sigma_\delta + \zeta S)(\delta - m_\delta)\}, \\ \text{so that} \quad \delta|... &\sim N_{n_a}(m_\delta, (\iota\Sigma_\delta + \zeta S)^{-1}), \end{aligned} \tag{B.6.6}$$

where $m_\delta = \zeta(\iota\Sigma_\delta + \zeta S)^{-1}R$.

Thus $\pi(\delta|...)$ is a multivariate normal.

The full conditional of μ is

$$\mu|... \sim N(Q, \eta^2), \tag{B.6.7}$$

where $Q = \sum_{x,t}(\xi_{x,t} - \alpha_x - \beta_x\kappa_t - \delta_x\gamma_{t-x})/(n_a n_y)$ and $\eta^2 = 1/(n_a n_y \zeta)$.

For ζ and ϱ draws are generated using the MH algorithm with the same prior and moves as for ψ in Model 1.

The full conditional of ι is a standard gamma distribution:

$$\iota|... \sim Gamma(1 + n_a/2, f + \frac{1}{2}\delta' \Sigma_\delta \delta). \tag{B.6.8}$$

The full conditional for each $\xi_{x,t}$ is a non-standard distribution and is sampled using a MH move:

$$\pi(\xi_{x,t}|...) \propto \exp\left(-E_{x,t}\exp(\xi_{x,t}) - \frac{\zeta}{2}z_{x,t}^2\right)\} \exp(d_{x,t}\xi_{x,t}). \tag{B.6.9}$$

B.7 Model 6

In this model

$$\xi_{x,t} = \log\left(\frac{m_{x,t}}{1-m_{x,t}}\right) = \mu + \alpha_x + \beta_x\kappa_t + \delta_x\gamma_{t-x} + z_{x,t}. \tag{B.7.1}$$

B.7.1 Models 6(a), 6(b) and 6(c)

Models 6 are similar to Models 5 apart from the use of the logit link function. If we define $\xi_{x,t} = \log(\frac{m_{x,t}}{1-m_{x,t}})$ then the priors and full conditionals on the parameters and hyper-parameters are the same as the corresponding parameters and hyper-parameters of Models 5(a), 5(b) and 5(c).

The full conditional for each $\xi_{x,t}$ is a non-standard distribution and is sampled using a MH move with proposal constrained to be in the interval (0,1) by a logit transform of a random walk centred on the logit transform of the current parameter value and with variance set to give good mixing:

$$\pi(\xi_{x,t}|...) \propto \exp\left(-E_{x,t}\exp(\xi_{x,t}) - \frac{\zeta}{2}z_{x,t}^2\right) \exp(d_{x,t}\xi_{x,t}). \tag{B.7.2}$$

B.8 Model 7

In this model

$$\xi_{x,t} = \log m_{x,t} = \mu + \alpha_x + \beta_x \kappa_t + \delta_x \gamma_{t-x}, \tag{B.8.1}$$

$$E_{x,t} = \frac{E^o_{x,t}}{1 - w_{x,t}}. \tag{B.8.2}$$

The MH algorithm is used to make draws of each parameter, μ, α_x, β_x, κ_t, δ_x and γ_{t-x}, from the posterior. Proposal values were generated from the current value by adding a disturbance from a Gaussian distribution with mean zero and variance chosen for each parameter so as to ensure good mixing.

B.8.1 Models 7(a), 7(b) and 7(c)

Models 7(a), 7(b) and 7(c) are similar to Models 4(a), 4(b) and 4(c). The priors on the parameters and hyper-parameters are the same as for the corresponding parameters and hyper-parameters of Model 6(a), 6(b) and 6(c). Section 4.1.1 discusses the rationale for the prior on $w_{x,t}$. The prior on each $w_{x,t}$ is a Beta(1,s) where $s \sim Un(1, 10000)$.

The full conditional for each $\xi_{x,t}$ is a non-standard distribution and is sampled using a MH move with proposal generated from a random walk centred on the current parameter value and variance set to give good mixing:

$$\pi(\xi_{x,t}|...) \propto \exp\{-E_{x,t} \exp(\xi_{x,t})\}(\exp(d_{x,t}\xi_{x,t})). \tag{B.8.3}$$

APPENDIX C

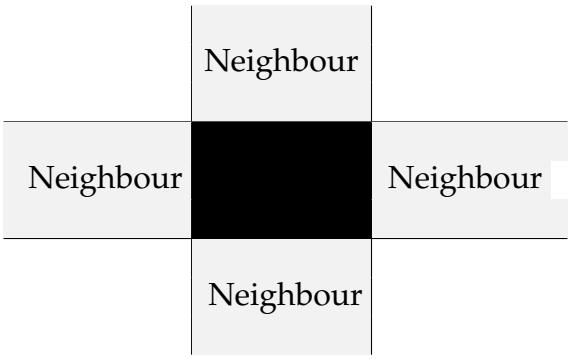
Residual Autocorrelation Test - 2 Colour Map

The question is whether or not the patterns of residuals shown in Figures 5.10 and 5.23 show no autocorrelation in the coloured map of cells. This appendix follows the derivation of the Residual Autocorrelation Test as set out by Cliff and Ord (1981).

The null hypothesis H_0 is that there is no autocorrelation in the residual pattern.

The test statistic is constructed based on the distribution of the number of joins between connecting cells of different colours cells in a 2-colour map. For a 2-dimensional map the range of the number of joins J, is $J \in \{0, 1, 2, 3, 4\}$; this is illustrated for an interior cell coloured black in Fig. C.1. The grey cells are the neighbours of the black cell, so in this case the number of joins J is 4. If a cell is on the edge of the map then it has 3 neighbours and a corner cell has only 2 neighbours.

Table C.1: Diagram of interior cell of a map coloured black, the neighbouring cells are coloured grey. So in this example J the number of joins for the black cell is 4.



Expected frequency distribution for J under H_0 i.e. no spatial autocorrelation

The expectation is calculated under the assumption that each cell is independently coded black with probability p, and grey with probability $q = 1 - p$. In our case p is assumed to be $\frac{1}{2}$.

In the $n_a \times n_y$ map, c_j denotes the number of cells that have j joins with other cells. Then the total number of joins, A, in the map is given by

$$A = \frac{1}{2} \sum_{j=1}^{R} j c_j$$

where $R = max\{j | c_j > 0\}$.

In the specific case in hand, the 2-colour map had dimension $n_a \times n_y$ map, $R = 4$ and $c_1 = 0$, $c_2 = 4$, $c_3 = 2(n_a - 2) + 2(n_y - 2)$ and $c_4 = (n_a - 2)(n_y - 2)$. So, the total number of joins A is $n_a n_y$.

For a given cell s, with j joins in all, the probability of J joins between neighbouring black and grey cells is

$$\begin{aligned} P(J|j) &= P(s^{th} \text{ cell is black})\ P(J \text{ neighbours are grey} | s \text{ is black}) \\ &+ P(s^{th} \text{ cell is grey}) P(J \text{ neighbours are black} | s \text{ is grey}). \end{aligned}$$

If we denote c_i as the number of cells which have i joins then the expected number of J joins between black and grey cells is E_J say, where

$$E_J = \sum_{i=1}^{R} c_i P(J|i), \qquad J = 0, 1, 2, \ldots, 4.$$

This expression equals

$$E_J = \sum_{i=1}^{R} c_i \binom{i}{J} \{p^{i+1-J} q^J + p^J q^{i+1-J}\}.$$

Observed frequency distribution for J.

For a residual map in Fig. 5.10, a frequency table for J joins, $J \in \{0, 1, 2, 3, 4\}$, is

constructed by counting the number of joins between neighbouring cells of different colours. Let O_J be the observed number of J joins in the map.

The modified Chi-squared statistic see Cliff and Ord (1981), is

$$\tilde{X}^2 = \beta \sum_J \frac{(O_J - E_J)^2}{E_J},$$

where

$$\beta = \frac{1}{2}\{1 - \frac{\frac{1}{2}B(1-4pq)}{A(1-2pq) + B(1-4pq)}\},$$

$A = \frac{1}{2}\sum jc_j$ and $B = \frac{1}{2}\sum_{i=1}^{n} w_{i.}^2 - A$.

Here we define $w_{i.}$ as the sum of the i^{th} row of the matrix w where w is a matrix of weights for neighbouring cells; in this case it is a $n_a n_y \times n_a n_y$ matrix. The (x, y) cell of w has value 1 if the x^{th} and y^{th} cells are neighbours; otherwise the (x, y) cell has value 0.

$\tilde{X}^2$ may be treated as having a chi-squared distribution under H_0 provided that the number of cells is not too small. Numerical studies in Cliff et al. (1975) indicate that the appropriate degrees of freedom is R when p is known.

APPENDIX D

Out of Sample Forecast Results for Models 5 and 7

Figure D.1: Plots of out-of-sample empirical male mortality rates for individual ages 60 to 74, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 5(a), 5(b) and 5(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

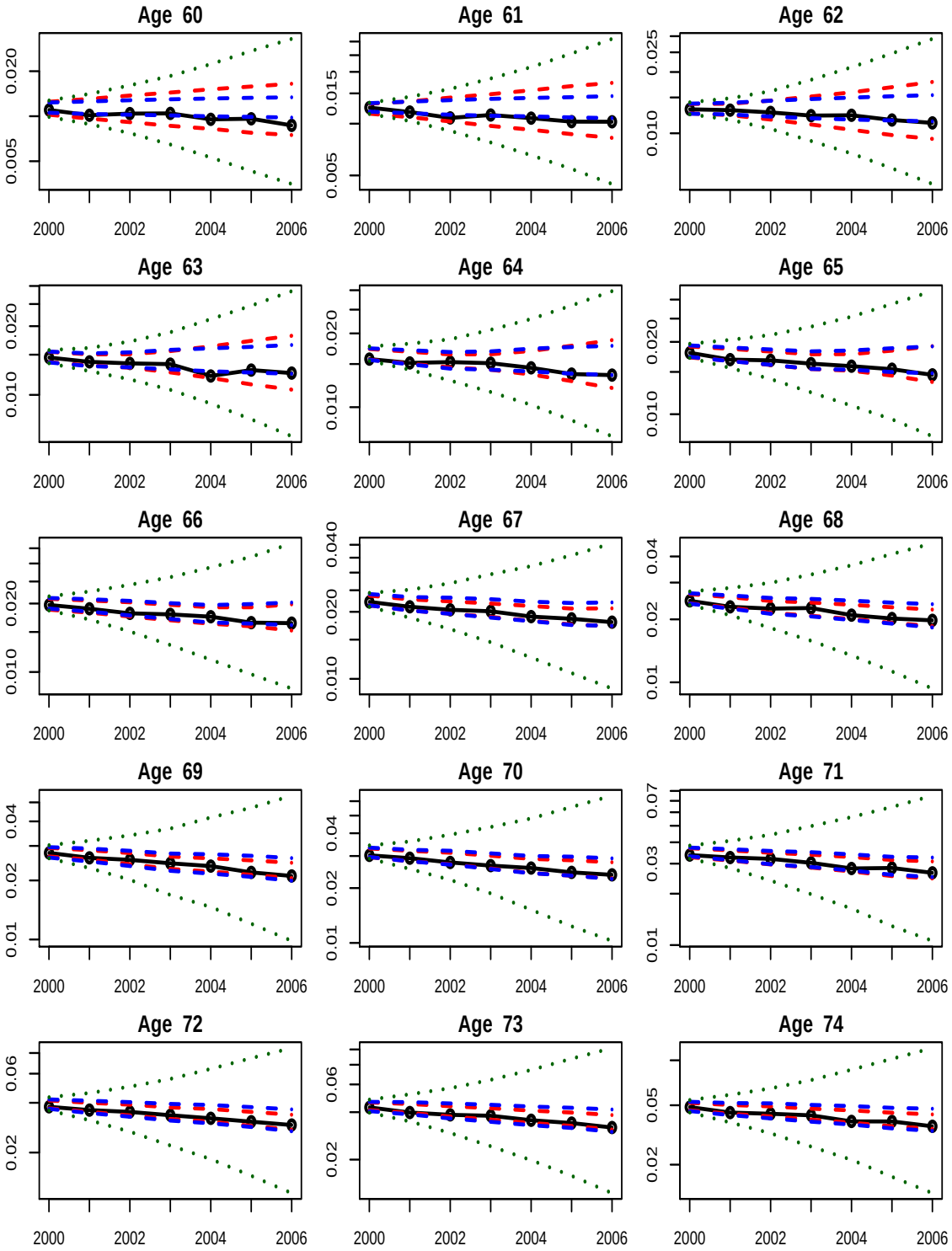


Figure D.2: Plots of out-of-sample empirical male mortality rates for individual ages 75 to 89, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 5(a), 5(b) and 5(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

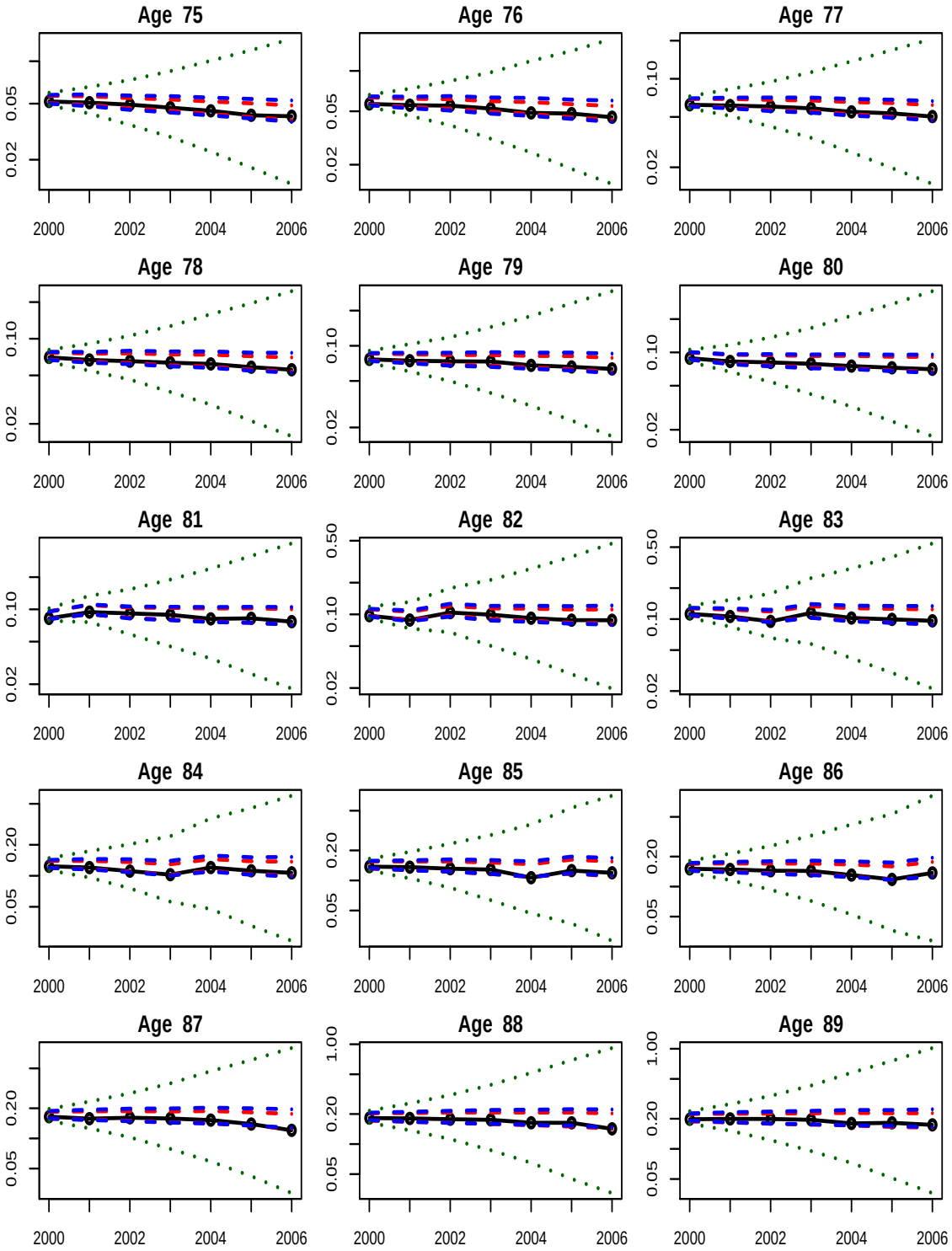


Figure D.3: Plots of out-of-sample empirical male mortality rates for individual ages 90 to 100, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 5(a), 5(b) and 5(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

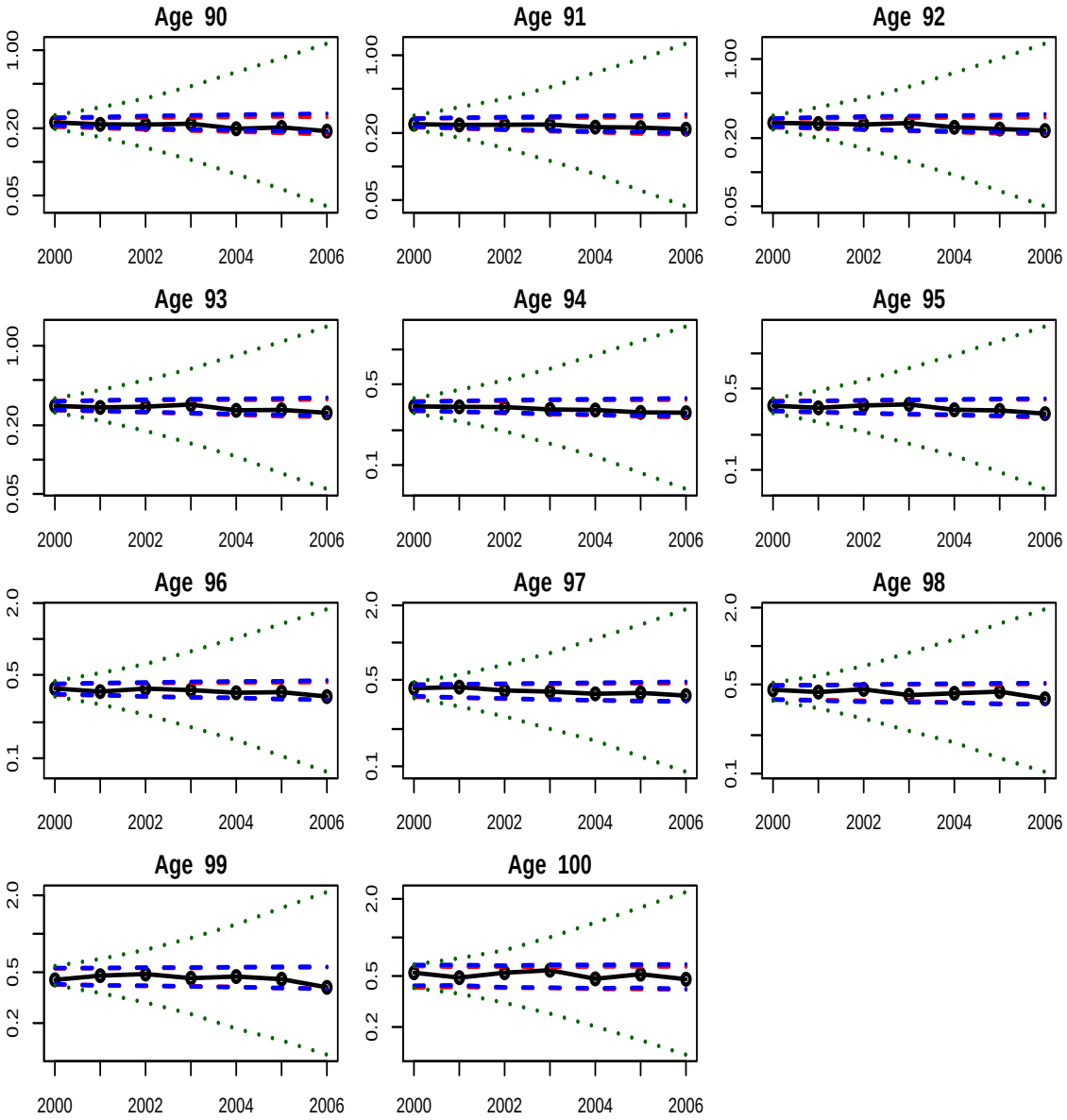


Figure D.4: Plots of out-of-sample empirical male mortality rates for individual ages 60 to 74, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 7(a), 7(b) and 7(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

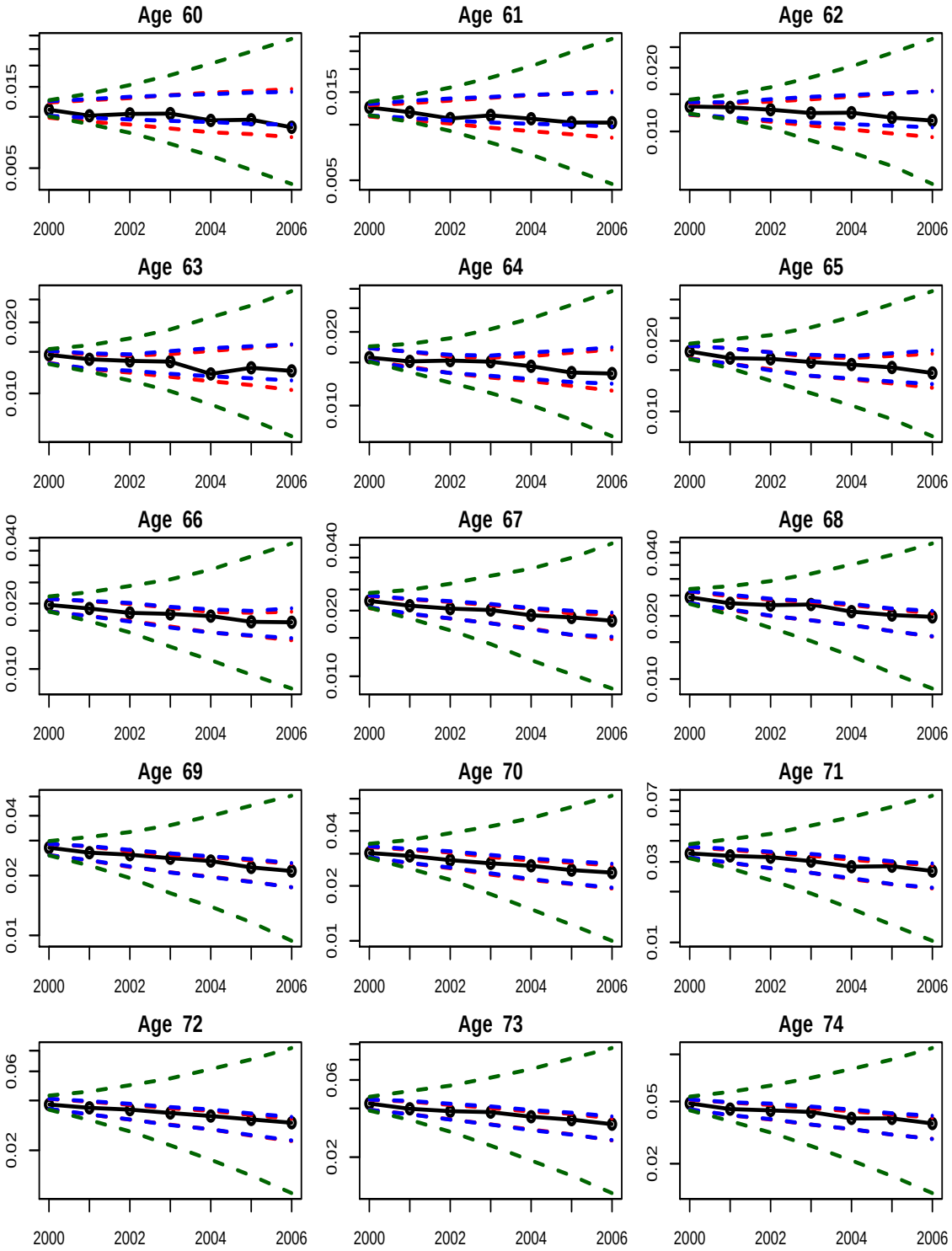


Figure D.5: Plots of out-of-sample empirical male mortality rates for individual ages 75 to 89, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 7(a), 7(b) and 7(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

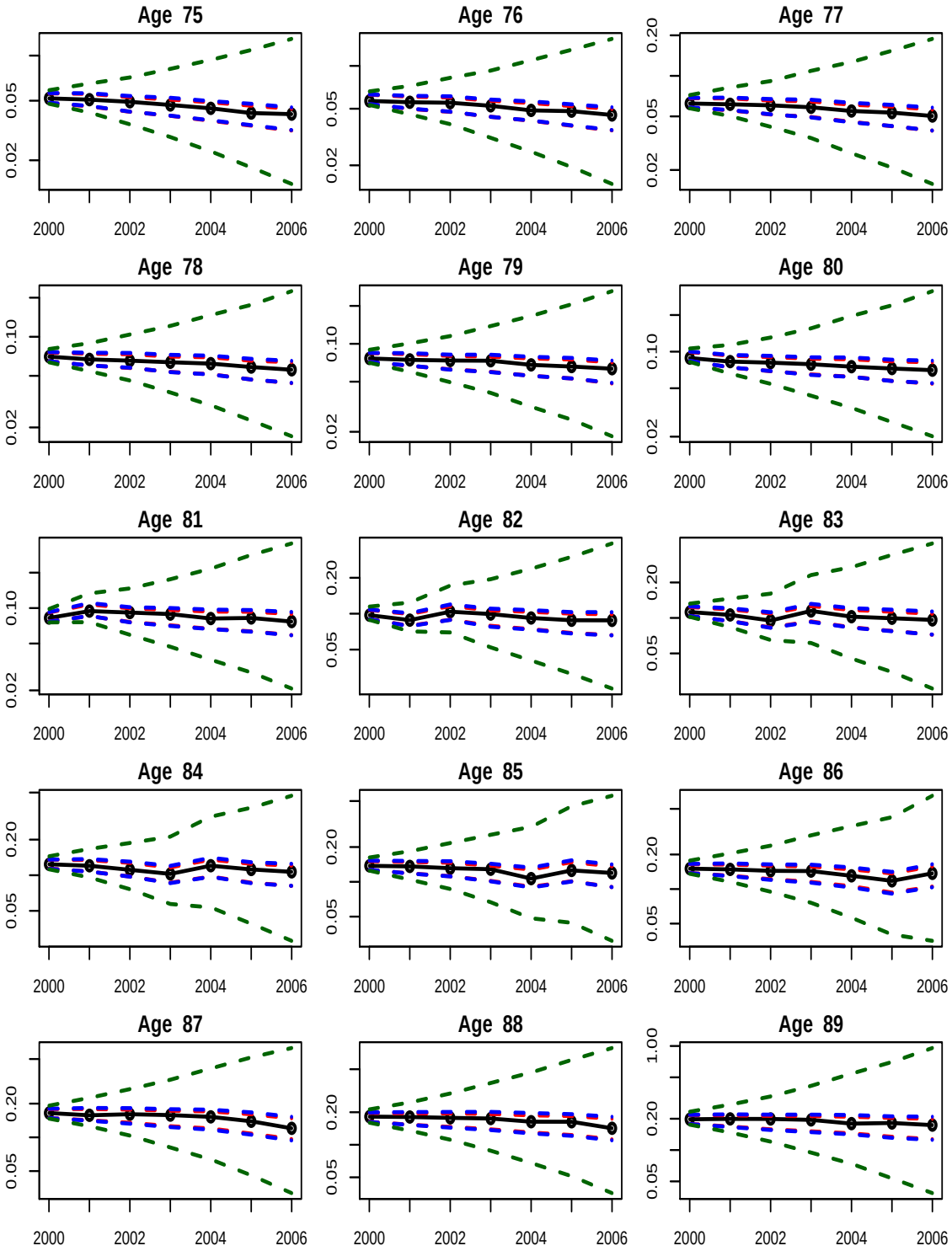


Figure D.6: Plots of out-of-sample empirical male mortality rates for individual ages 90 to 100, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 7(a), 7(b) and 7(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

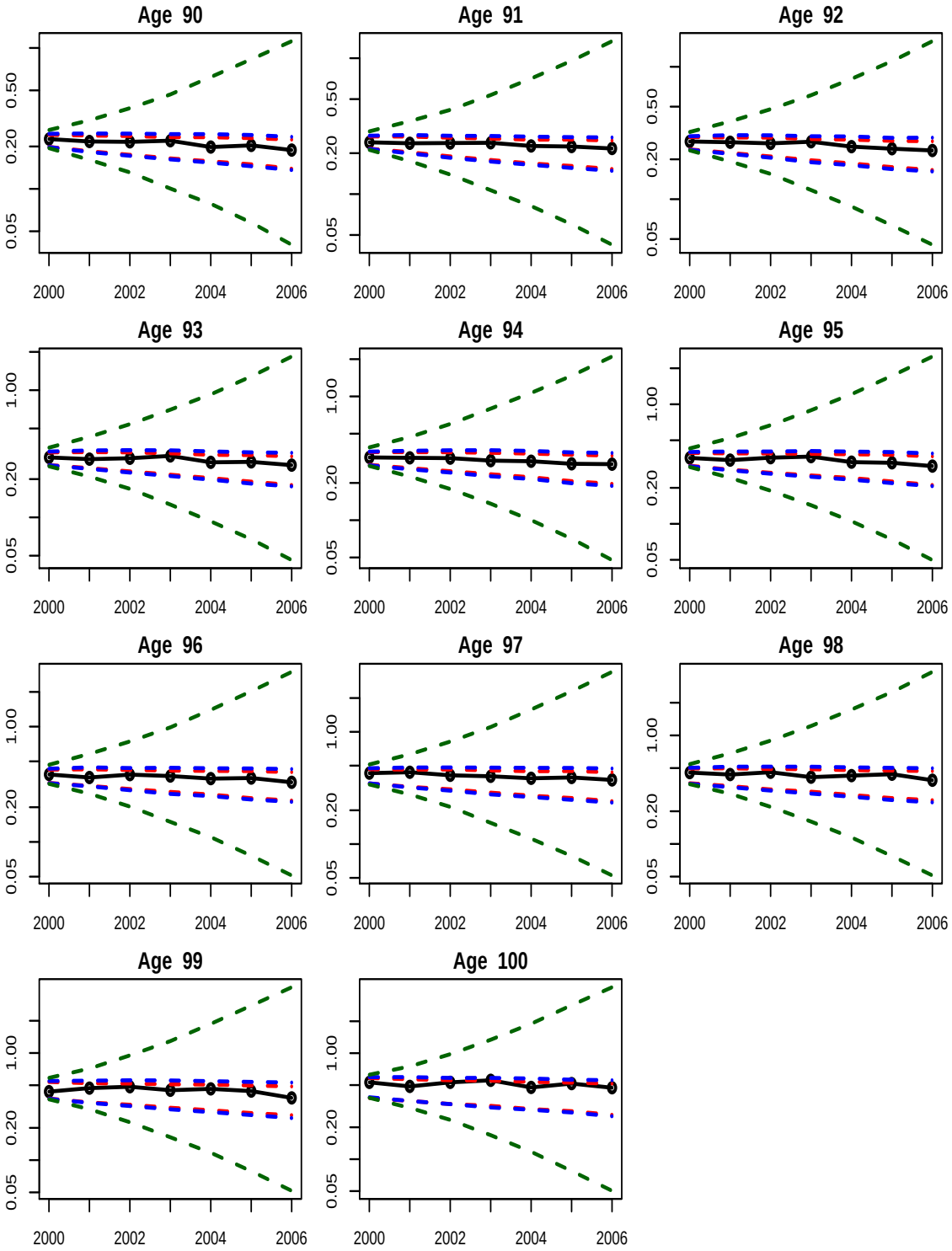


Figure D.7: Plots of out-of-sample empirical female mortality rates for individual ages 60 to 74, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 5(a), 5(b) and 5(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

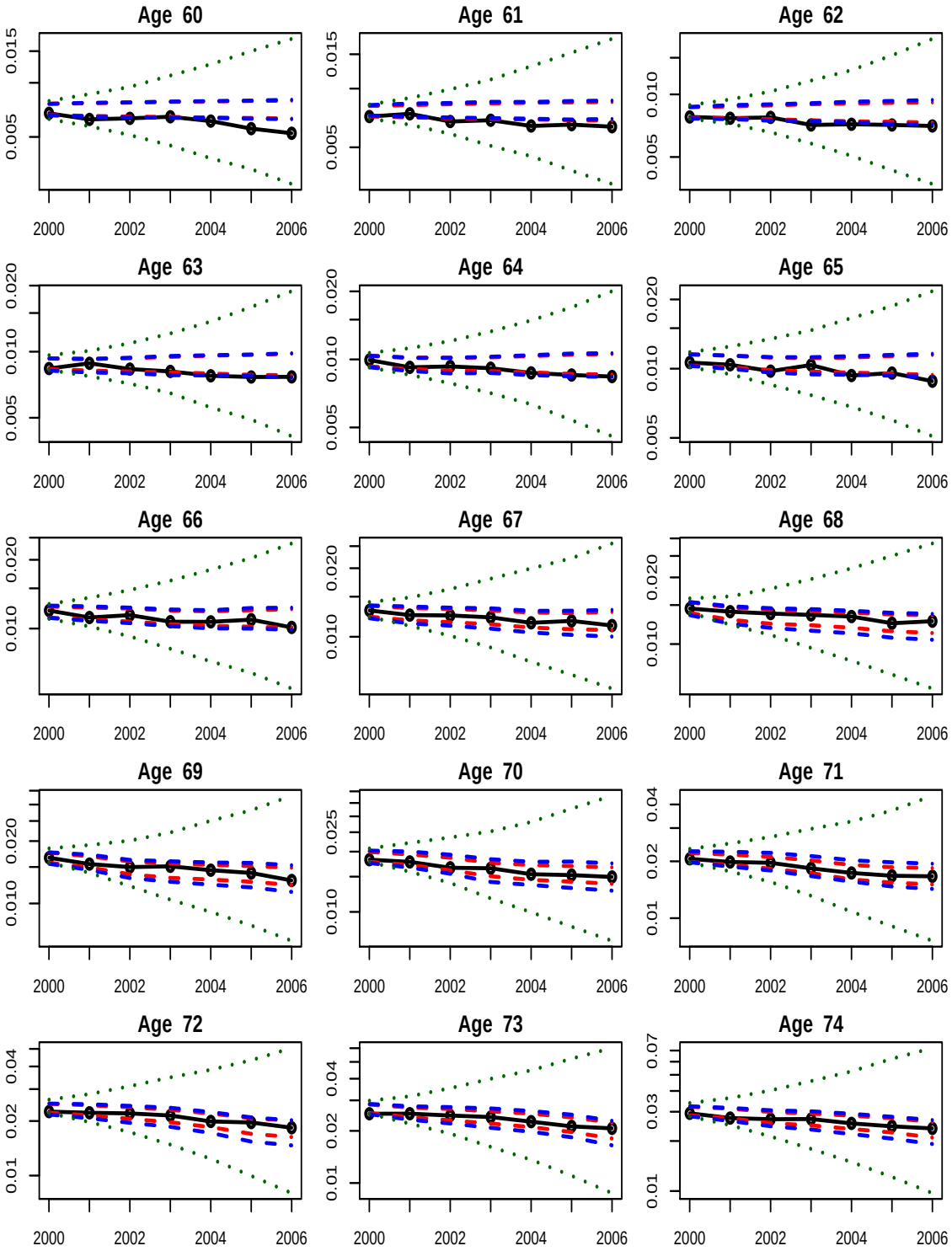


Figure D.8: Plots of out-of-sample empirical female mortality rates for individual ages 75 to 89, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 5(a), 5(b) and 5(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

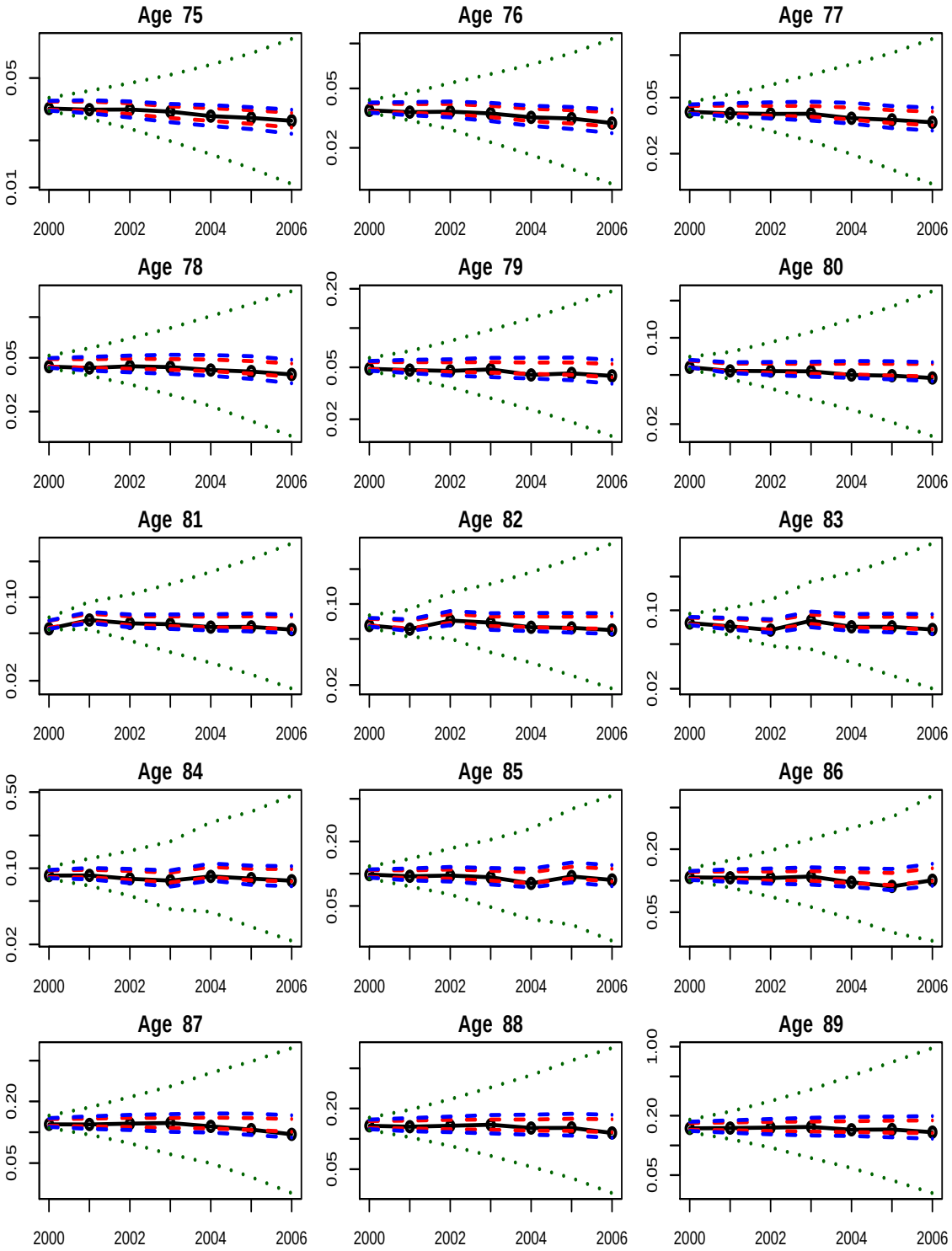


Figure D.9: Plots of out-of-sample empirical female mortality rates for individual ages 90 to 100, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 5(a), 5(b) and 5(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

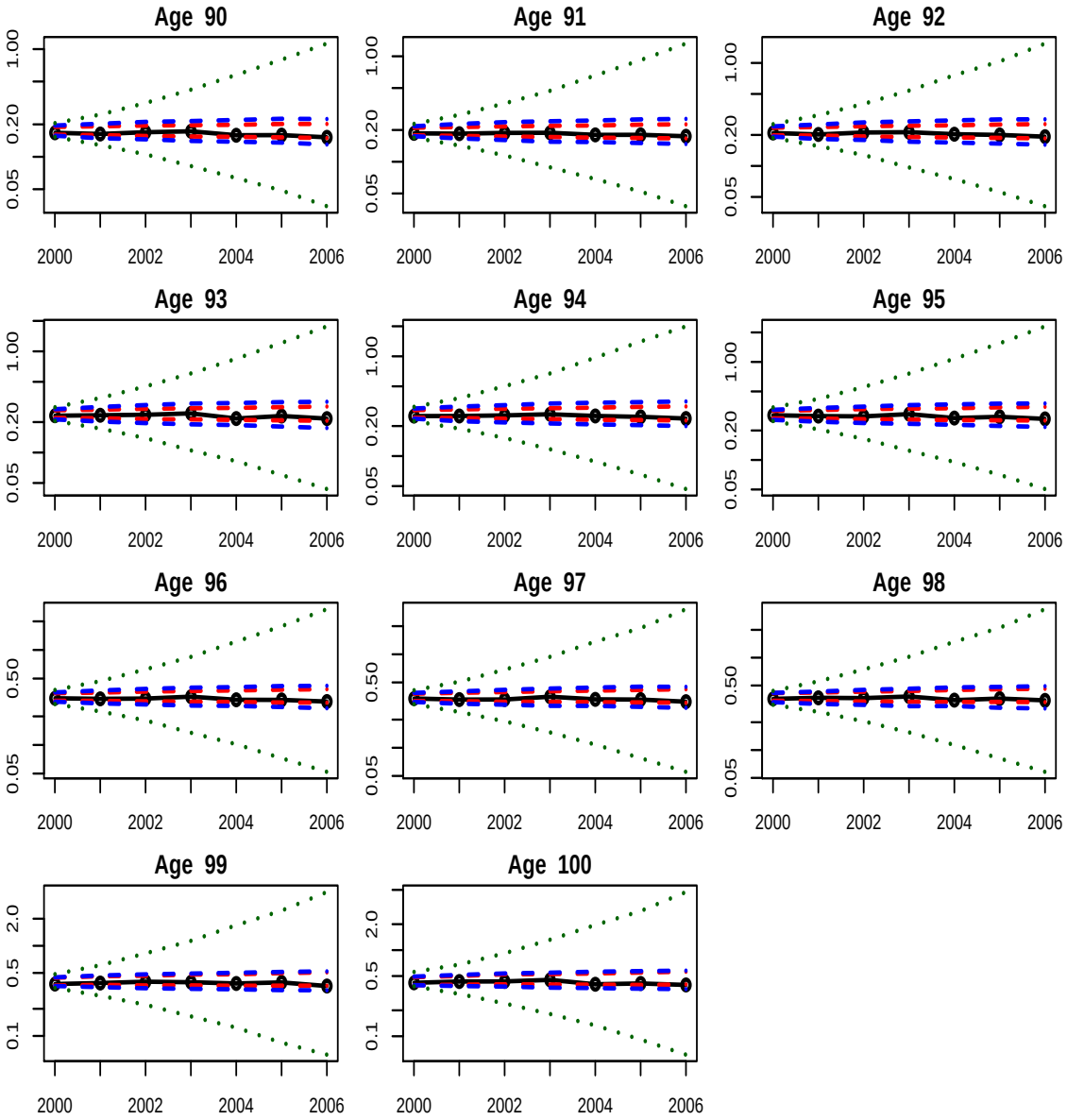


Figure D.10: Plots of out-of-sample empirical female mortality rates for individual ages 60 to 74, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 7(a), 7(b) and 7(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

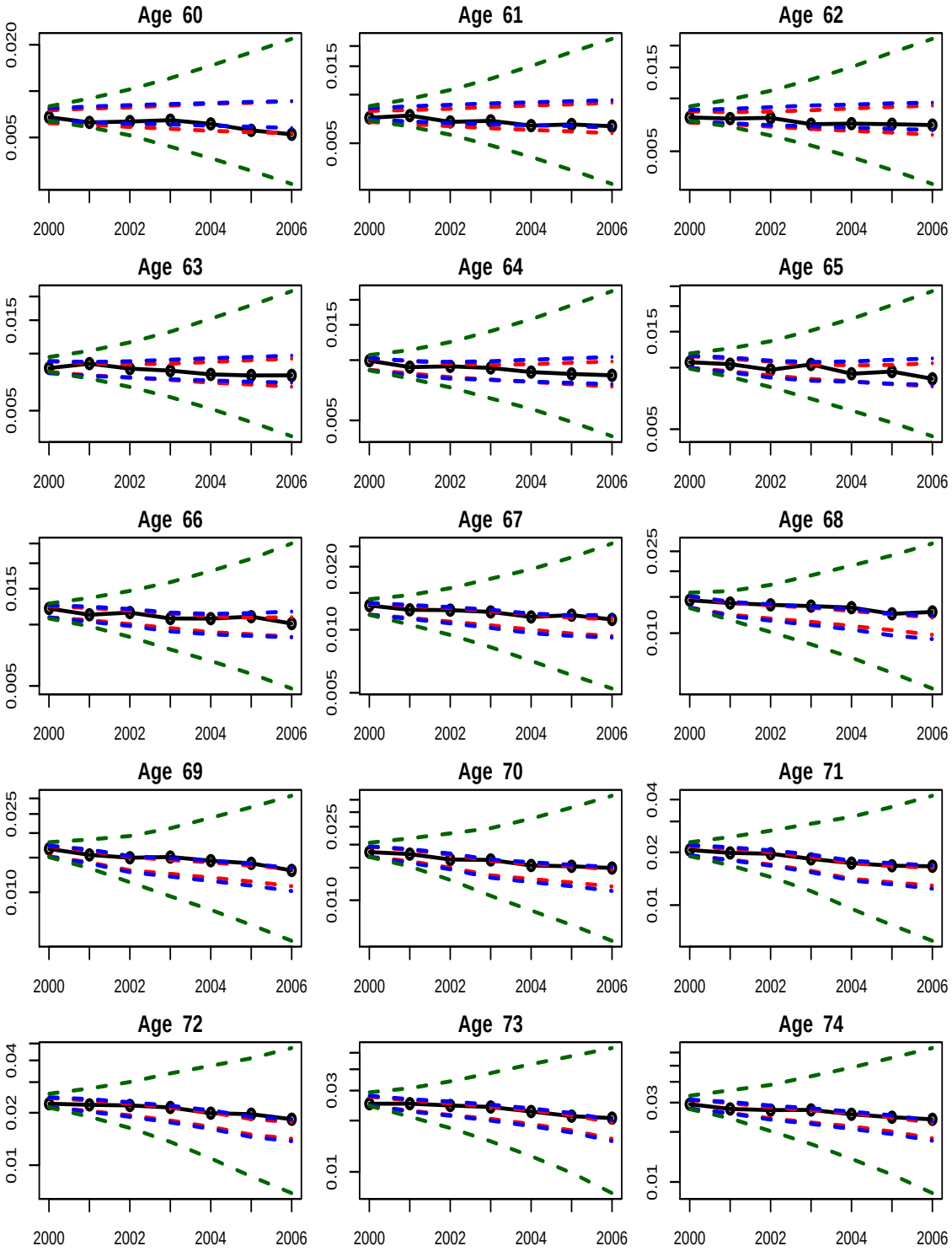


Figure D.11: Plots of out-of-sample empirical female mortality rates for individual ages 75 to 89, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 7(a), 7(b) and 7(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.

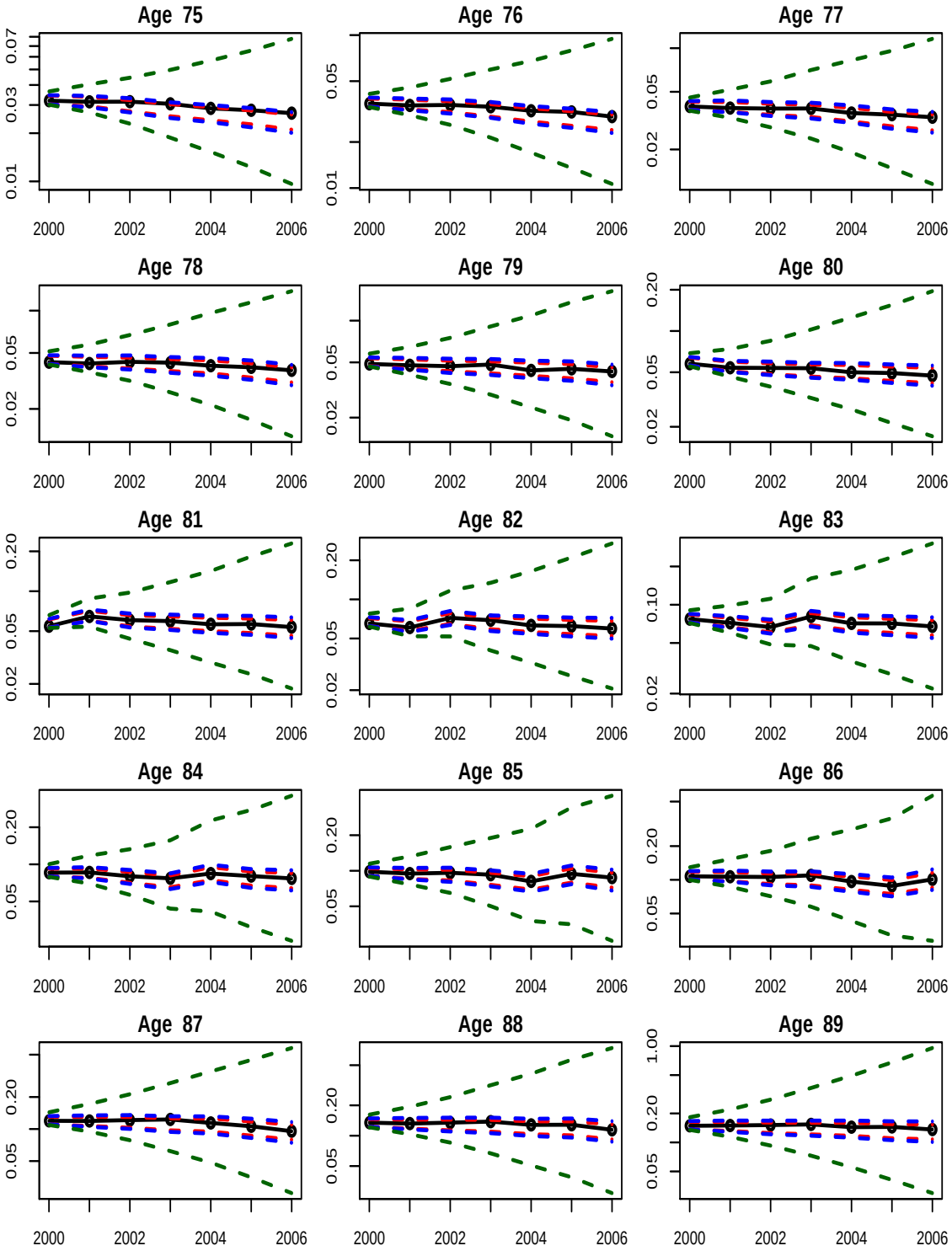
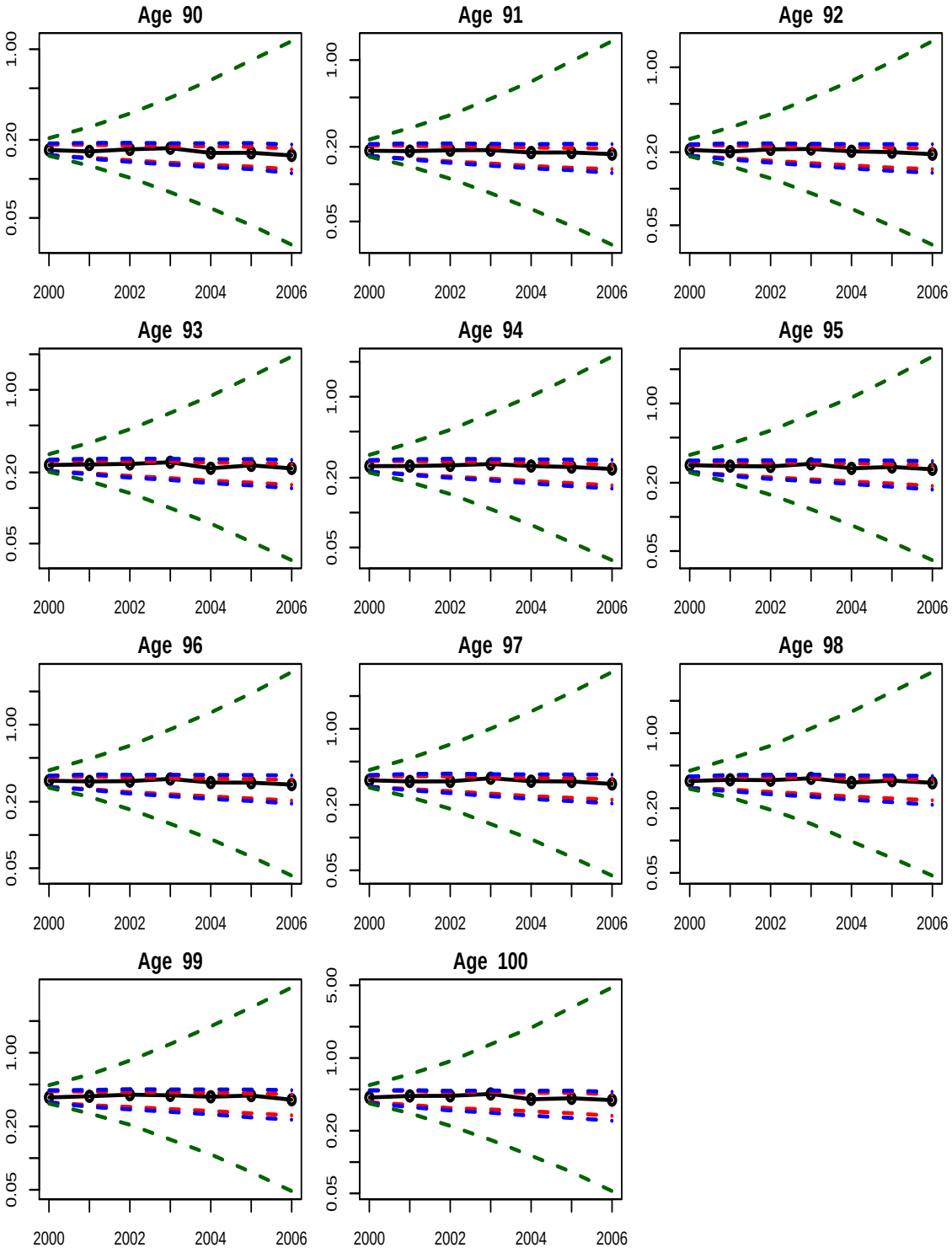


Figure D.12: Plots of out-of-sample empirical female mortality rates for individual ages 90 to 100, for the years 2000 to 2006 and the corresponding forecast 95% posterior predictive intervals for $\hat{m}_{x,t}$ from Models 7(a), 7(b) and 7(c). The blue, red and green lines show 95% posterior intervals with prior smoothing of (i) random walk on the levels, (ii) autoregressive process of order 1 on the first differences and (iii) random walk on first differences respectively. The numbers on the y-axis in each plot are rates but the plot is on a logarithmic scale.



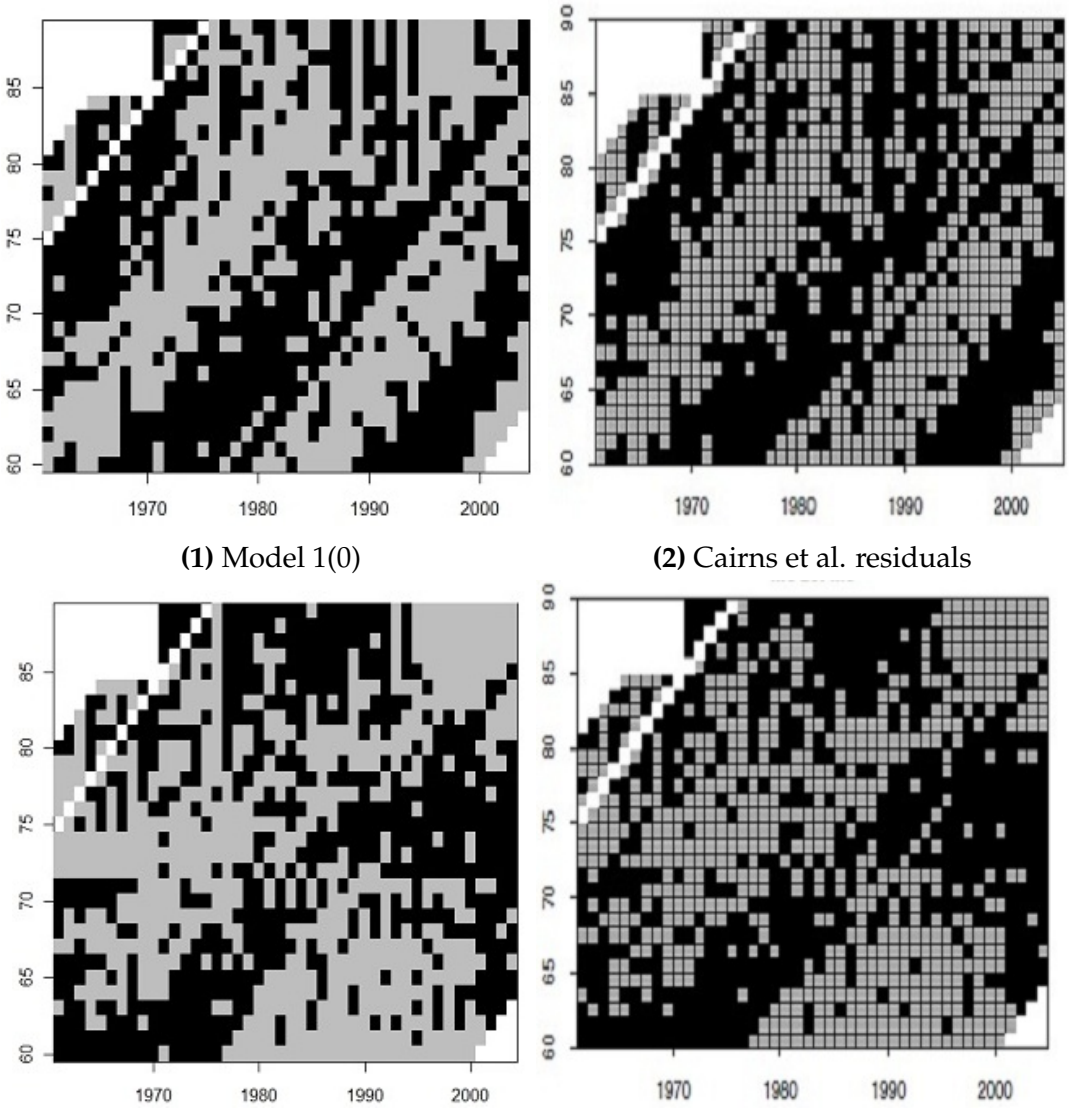
APPENDIX E

Program Testing

A number of tests were done to check that the calculations carried out by various algorithms within the programs were working correctly. In addition, the following models: Model 1(0), Model 2(0) and Model 3(c), data and results have been presented in reports by other researchers. Therefore as an additional check on these models I attempted to reproduce these results.

Model 1(0) and Model 2(0) were used in the paper Cairns et al. (2009). The data used was population data for England and Wales for males aged 60 to 89, for the period 1961 to 2004. The source of the data was not HMD so I contacted the authors for their data but because it was supplied by a 3rd party the data was considered proprietary. There would be differences but probably not that material. One of the known differences was that Cairns at al. excluded certain cohorts, where the year of birth was on or before 1880 and 1886, as well as cohorts that had fewer than 5 separate years, i.e. year of birth after 1940. For comparison purposes, I also excluded these cohorts from the HMD data. Table E.1 compares the standardised residuals from Model 1(0) and Model 2(0) with the corresponding results from Cairns et al. (2009). The calculation of the variance of the standardised residuals is detailed in section 5.2.4. Figure E.1 compares the pattern of residuals from Model 1(0) and Model 2(0) with the corresponding models from Cairns et al. (2009). The explanation of these plots is detailed in section 5.2.4.

A slight modification of Model 3(c) was used by Besag et al. (1995) to analyse cancer deaths. Besag et al. used a Binomial model for the number of deaths rather than a Poisson model. So for testing purposes, Model 3(c) was altered very easily to



(1) Model 1(0) **(2)** Cairns et al. residuals

(3) Model 2(0) **(4)** Cairns et al. residuals

Figure E.1: Comparison of residual plots for Model 1(0) and Model 2(0) with corresponding plots from Cairns et al. (2009). Calendar year on x-axis and age on y-axis.

Model	Cairns et al.	Our Figures
Model 1(0)	4.1	4.2
Model 2(0)	2.2	2.2

Table E.1: Comparison of the variance of standardised residuals using Cairns et al. (2009) England & Wales male data.

become a Binomial model and this model was used together with the exact data set used by Besag et al. Table 2 page 17 of Besag et al. (1995) contained the mean and the standard deviation of the negative of the log-odds of modelled mortality parameter $p_{x,t}$. This table has been reproduced in Tables E.2 and E.3 but with figures from our Model 3(c). There were only 7 small differences indicated by an underline in the wrong digit.

Table E.2: Mean Negative log-odds of Modelled Mortality.

Age	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980
50-54	7.39	7.11	7.04	6.97	7.16	7.30	7.38	7.50	7.62	7.74
55-59	6.71	6.54	6.25	6.19	6.22	6.42	6.41	6.57	6.69	6.81
60-64	6.12	5.99	5.82	5.48	5.59	5.47	5.56	5.73	5.85	5.97
65-69	5.77	5.56	5.33	5.19	4.93	4.90	4.92	5.02	5.13	5.25
70-74	5.44	5.20	4.96	4.73	4.54	4.47	4.38	4.37	4.47	4.58
75-79	5.15	4.90	4.63	4.43	4.32	4.17	4.02	3.91	3.91	4.01
80-84	5.12	4.83	4.65	4.37	4.09	4.01	3.85	3.66	3.59	3.59

Table E.3: Standard Deviation of log-odds of Modelled Mortality.

Age	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980
50-54	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.16	0.27	0.41
55-59	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.11	0.20	0.32
60-64	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.10	0.16	0.26
65-69	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.09	0.15	0.23
70-74	0.05	0.04	0.03	0.03	0.02	0.02	0.02	0.10	0.15	0.23
75-79	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.10	0.15	0.23
80-84	0.08	0.06	0.05	0.04	0.03	0.03	0.03	0.10	0.15	0.23

Besag et al. also produced a table of 80% posterior predictive intervals for the number of deaths per 100,000 of lives. Table 3 on page 18 of Besag et al. (1995) has been reproduced in table E.4 but with figures from our Model 3(c). The final column of table E.4 shows the average difference between the figure produced by our model and that shown in the Besag et al. paper. For each age group, this difference is almost constant over the periods. The figures indicate that our model produces only minor differences and that our posterior predictive intervals are very slightly wider.

Table E.4: 80% posterior predictive intervals for the number of deaths per 100,000 of lives

Age		1935	1940	1945	1950	1955	1960	1965	Difference
	10% lower	55	74	80	86	71	61	56	-3
50-55	observed	59	85	88	97	75	64	62	
	90% Upper	69	91	96	102	85	74	68	3
	10% lower	111	132	179	191	187	153	154	-4
55-59	observed	124	141	198	206	199	158	167	
	90% Upper	134	155	205	217	212	174	174	3
	10% lower	204	235	280	396	354	402	367	-4
60-64	observed	226	247	290	427	361	425	389	
	90% Upper	237	268	315	438	392	439	401	5
	10% lower	290	362	457	530	694	713	700	-6
65-69	observed	310	378	480	548	725	742	725	
	90% Upper	332	408	505	579	748	765	750	5
	10% lower	405	518	666	837	1019	1100	1199	-6
70-74	observed	432	541	694	878	1069	1133	1235	
	90% Upper	464	580	732	911	1094	1173	1273	5
	10% lower	534	695	913	1128	1265	1468	1713	-5
75-79	observed	600	764	990	1190	1310	1517	1753	
	90% Upper	620	785	1012	1232	1368	1571	1818	6
	10% lower	534	736	889	1177	1574	1704	2012	-4
80-84	observed	568	809	891	1218	1686	1771	2082	
	90% Upper	659	858	1009	1312	1723	1842	2155	4

References

Andreev, K. (2001). Kannisto-Thatcher database on population and death counts at older ages. Technical report, Max Planck Institute for Demographic Research.

Barrett, J. (1973). Age, Time and Cohort Factors in Mortality from Cancer of the Cervix. *Journal of Hygene 71*, 253–259.

Barrett, J. (1978). The Redundant Factor Method and Bladder Cancer Mortality. *Journal of Epidemiology and Community Health 32*, 314–316.

Bauer, D., D. Bergmann, and A. Reuss (2009). Solvency II and Nested Simulations: a Least-Squares Monte Carlo Approach . www.researchgate.net/228962781/publicationsSolvencyII-and-nested-simulations-A-Least-Squares-Monte-Carlo-approach. Downloaded Aug 2012.

Beard, R. (1959). Some Aspects of Theories of Mortality, Cause of Death Analysis, Forecasting and Stochastic Processes. In *The Lifespan of Animals*, pp. 302–311. Little, Brown, Boston.

Beard, R. (1971). Some Aspects of Theories of Mortality, Cause of Death Analysis, Forecasting and Stochastic Processes. *Biological Aspects of Demography 999*, 57–68.

Berzuini, C. and D. Clayton (1994). Bayesian analysis of survival on multiple time scales. *Statistics in Medicine 13*, 823–838.

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995). Bayesian Computation and Stochastic Systems. *Statistical Science 10*, 3–41.

Biatat, V. D. and I. Currie (2010). Joint models for classification and comparison of mortality in different countries. In *25th IWSM*, pp. 89–94.

Brouhns, N., M. Denuit, and J. Vermunt (2002). A Poisson Log-bilinear Regression Approach to The Construction of Projected Lifetables. *Insurance: Mathematics and Economics 31*, 373–393.

Burnham, K. P. and D. R. Anderson (1998, Nov). *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer Verlag.

Cairns, A., D. Blake, and K. Dowd (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance 73*, 687–718.

Cairns, A., D. Blake, K. Dowd, G. Coughlan, and D. Epstein (2009). A Quantitative Comparison of Stochastic Mortality Models Using Data from England & Wales and the United States. *North American Actuarial Journal 13*, 1–35.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction* (6th edition ed.). CRC Press.

Clayton, D. and E. Schifflers (1987a). Models for Temporal Variation in Cancer Rates. I: Age-period-cohort models. *Statistics in Medicine 6*, 449–467.

Clayton, D. and E. Schifflers (1987b). Models for Temporal Variation in Cancer Rates. II: Age-period-cohort models. *Statistics in Medicine 6*, 469–481.

Cliff, A., R. Martin, and J. Ord (1975). A test for spatial autocorrelation in choropleth maps based upon a modified χ^2 statistic. *Transactions and Papers, Institute of British Geographers 65*, 109–129.

Cliff, A. and J. Ord (1981). *Spatial Processes Models & Applications*. Pion.

Continuous Mortality Investigation (2009). Working Paper 38 - A Prototype Mortality Projections Model: Part One - An Outline of the Proposed Approach. Technical report, Institute of Actuaries and Faculty of Actuaries.

Currie, I. (2012). Forecasting with the Age-Period-Cohort Model? In *27th IWSM*, pp. 87–92.

Currie, I., M. Durban, and P. Eilers (2004). Smoothing and Forecasting Mortality Rates. *Statistical Modelling 4*, 279–298.

Czado, C., A. Delwarde, and M. Denuit (2005). Bayesian Poisson Log-bilinear Mortality Projections. *Insurance: Mathematics and Economics 36*, 260–284.

Delwarde, A., M. Denuit, and P. Eilers (2007). Smoothing The Lee-Carter and Poisson Log-bilinear Models for Mortality Forecasting. *Statistical Modelling 7*, 29–48.

Department of Work and Pensions (2011). PENSIONS ACT 2011 - Impact Assessments Summary. Technical report, DWP.

European Commission (2010). QIS5 Technical Specifications. Technical report, Brussels.

European Parliament (2009). Directive 2009/138/EC. Technical report, European Parliament.

Fienberg, S. and W. Mason (1979). Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data. *Sociological Methodology 10*, 1–67.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). *Bayesian Data Analysis, Second Edition*. Chapman & Hall/CRC.

Goldstein, H. (1984). Age, Period And Cohort Effects: A Confounded Confusion. *Journal Applied Statistics 16*, 19–24.

Heligman, L. and J. Pollard (1980). The Age Pattern of Mortality. *JIA 107*, 49–80.

Holford, T. (1983). The Estimation of Age, Period and Cohort Effects for Vital Rates. *Biometrics 39*, 311–324.

Holford, T. (1991). Understanding The Effects of Age, Period, and Cohort on Incidence and Mortality Rates. *Annu. Rev. Publ. Health 12*, 425–457.

Holford, T. (2006). Approaches to Fitting Age-Period-Cohort Models with Unequal Intervals. *Statistics in Medicine 25*, 977–993.

Kannisto, V. (1994). *Development of Oldest-Old Mortality, 1950-1990 - Evidence from 28 Developed Countries*. Odense University Press.

Knorr-Held, L. and E. Rainer (2001). Projections of Lung Cancer Mortality in West Germany: A Case Study in Bayesian Prediction. *Biostatistics 2*, 109–129.

Konstantopoulos, T. (2006). Notes on Survival Models.

Lee, R. and L. Carter (1992). Modelling and Forecasting U.S. Mortality. *Journal of the American Statistical Association 87*, 659–675.

Macdonald, C. (2009). Scoping mortality research: (report of the mortality research steering group). *British Actuarial Journal 15*, 33–83.

National Association of Pension Funds (2010). The 2010 NAPF Annual Survey. Technical report, National Association of Pension Funds.

Oeppen, J. and J. Vaupel (2002). Broken Limits to life Expectancy. *Science 296*(5570), 1029–1031.

Office for National Statistics (2011a). Life expectancy at birth and at age 65 by local areas in the United Kingdom, 2004-06 to 2008-10. Technical report, Office for National Statistics.

Office for National Statistics (2011b). National Population Projections 2008-based. Technical Report 27, Office for National Statistics.

Office for National Statistics (2011c). Statistical Bulletin-Life expectancy at birth and at age 65 by local areas in the United Kingdom, 2004-06 to 2008-10. Technical report, Office for National Statistics.

Office for National Statistics (2012a). Mortality Assumptions, 2010-Based National Population Projections. Technical report, Office for National Statistics.

Office for National Statistics (2012b). Occupational Pension Schemes Survey, 2011. Technical report, Office for National Statistics.

Office for National Statistics (2012c). What are the Chances of Surviving to Age 100? Technical report, Office for National Statistics.

Pedroza, C. (2006). A Bayesian Forecasting Model: Predicting U.S. Male Mortality. *Biostatistics 7*, 530–550.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal Institute of Actuaries 63*, 12–57.

Platt, R. (2011). One-year Value-at-Risk for longevity and mortality. *Insurance: Mathematics and Economics 49*(3), 462–470.

Price, E. (2010). Number of Future Centenarians. Technical report, Department for Work and Pensions.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Reichmuth, W. and S. Sarferaz (2008). Bayesian Demographic Modeling and Forecasting: An Application to U.S. Mortality. Technical Report 052, Humboldt-Universität zu Berlin, Germany.

Renshaw, A. and S. Haberman (2003a). Lee Carter Mortality Forecasting-A Parallel Generalized Linear Modelling Approach for England and Wales Mortality Projections. *Journal of the Royal Statistical Society 52*, 119–137.

Renshaw, A. and S. Haberman (2003b). Lee-Carter Mortality Forecasting With Age-Specific Enhancement. *Insurance: Mathematics and Economics 33*, 255–272.

Renshaw, A. and S. Haberman (2006). A Cohort Based Extension to the Lee Carter Model for Mortality Reduction Factors. *Insurance: Mathematics and Economics 38*, 556–570.

Richards, S. J., I. D. Currie, and G. P. Ritchie (2013). A value-at-risk framework for longevity trend risk. *British Actuarial Journal*.

Riebler, A. and L. Held (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models. *Biostatistics 11*, 57–69.

Schmid, V. J. and L. Held (2007). Bayesian Age-Period-Cohort Modeling and Prediction. *Journal of Statistical Software 21*, 1–15.

Spiegelhalter, D., N. Best, B. Carlin, and A. Van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 583–640.

Spurgeon, E. (2011). *Life Contingencies* (3rd edition ed.). Cambridge University Press.

Vojak, F. (2011). Ageing, longevity and demographic change: A factpack of statistics. Technical report, International Longevity Centre-UK.

Willets, R. (2004). The Cohort Effect Insights and Explanations. Technical report, Institute and Faculty of Actuaries.

Wilmoth, J. (1993). Computational Methods For Fitting and Extrapolating The Lee-Carter Model of Mortality Change. Technical report, University of California, Berkeley, Department of Demography.

Wilmoth, J. (1995). Are Mortality-Rates Falling at Exremely High Ages: An Investigation Based on a Model Proposed by Coale and Kisker. *Population Studies-A Journal of Demography 49*, 281–295.

Wilmoth, J., K. Andreev, D. Jdanov, and D. Glei (2007). Methods Protocol for the Human Mortality Database. Technical Report version 5, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).