UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL SCIENCES

Mathematics

Modelling and Projecting Mortality Rates Using Adaptive P-splines

by

Kai Hon Tang

Thesis submitted for the degree of Doctor of Philosophy

January 2021

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES

Mathematics

Doctor of Philosophy

MODELLING AND PROJECTING MORTALITY RATES USING ADAPTIVE P-SPLINES

by Kai Hon Tang

In this thesis we propose models for estimating and projecting mortality rates using adaptive splines. Mortality modelling has various applications from social planning to insurance. However, raw mortality data often exhibits irregular patterns due to randomness. The data at the oldest ages are also very scarce and unreliable as there are only very little survivors at these ages, adding difficulty to estimation. Graduation refers to the act of smoothing crude mortality rates, during which extrapolation to older ages where data is non-existent is usually also performed. We first propose a flexible and robust model for mortality graduation of static life tables using adaptive splines. Male and female mortality rates are graduated jointly, as opposed to previous English Life Tables (ELTs) where they were smoothed independently. Therefore our model borrows information across sexes, which is especially helpful at the oldest ages. Often when male and female mortality rates are estimated independently, implausible age patterns may occur, such as intersecting male and female mortality schedules. This has been addressed using rather ad hoc procedures in previous ELTs, for example, by calculating the weighted average of the estimated mortality rates starting at the age where they intersect or by discarding data at the oldest ages. By utilising the locality of B-spline basis, constraints can be imposed effectively such that female mortality rates are always lower than or equal to male mortality rates at all ages, even at extrapolation ages, hence does not involve subjective adjustments.

We then extend the model to forecast mortality rates. Building upon models by Dodd et al. (2020) and Hilton et al. (2019), we jointly model and project male and female mortality rates of England Wales and Scotland. The joint sex model produces more reasonable long term male and female mortality projections that are non intersecting. Information is borrowed at the highest ages where exposures are small. By doing so the extrapolation to higher ages beyond data range gives more plausible estimates, especially for the mortality improvement rates for females at the oldest ages where a worsening mortality is otherwise projected. We also jointly model mortality rates of the same sex across the two countries, as they are expected to have similar mortality structures for the same sex. England Wales populations have a wider age range with available data, therefore the joint country model provides a way for the smaller Scottish populations to borrow information and learn from the bigger English Welsh populations. The joint country model is able to produce non-divergent long term projections between the countries for both males and females.

Finally, a joint model for all of the four populations is proposed. The model combines features of the joint sex and joint country models, and borrows strength across sexes and countries.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Kai Hon Tang , declare that the thesis entitled *Modelling and Projecting Mortality Rates Using Adaptive P-splines* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- 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;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission

Signed:.....

Date:.....

Acknowledgements

The completion of this thesis would not have been possible without the support and encouragement of several people who I wish to acknowledge.

First and foremost, I would like to thank my supervisors Dr Erengul Dodd and Professor Jon Forster for their guidance. I have encountered numerous challenges and they have provided invaluable insights along the way to make this thesis possible. Therefore I would like to express my gratitude towards their continuous efforts and devotion. I would also like to thank the Department of Mathematical Sciences at the University of Southampton for enabling me to pursue my doctorate with the help of a studentship.

Next, I would like to show my gratitude towards my friends for their companionship during the hard times. Especially, Ivan Ivanov, Mia Tackney, Meg Wong, Melody Lun and Kensey Lee for their emotional support and for always being there for me. Additionally, I would like to give my most sincere appreciation to Kate Ham, who has been a great mentor that I look up to. She has taught me how to be confident, positive and considerate.

Finally, I would like to express my deepest thanks to my family. They have been encouraging and supportive in every possible way. I would not have been able to accomplish this without their financial, moral and emotional support.

Chapter 1

Introduction

Mortality modelling has seen rapid developments in the past decades, different models have been proposed for the purpose of smoothing and forecasting mortality rates. Mortality modelling is necessary for studying the demography of countries such as population estimation and reconstruction, male-to-female ratios and life expectancies etc. Mortality modelling also has various applications from social planning to insurance and they are especially crucial in the insurance sector as many of the insurance products are related to the longevity risk. Huge amounts are invested in longevity risk and pensions. These often relate to the old age mortality, where uncertainty and variability are the highest. The data at these ages are also usually very scarce, adding difficulty to estimation. In addition, these actuarial products often rely on future mortality rates which, of course, are unknown. Therefore, mortality projection has a significant impact in actuarial applications.

For most countries, it is well-established that the mortality is generally decreasing in the past decades (Wilson, 2011). Typically, it is observed that infants experience exceptionally higher mortality due to neo-natal causes. After that, the child mortality rates decrease steadily until they reach late teenage and young adulthood, where a bump can be observed. This temporary increase in mortality can be attributed to teenage activities and accidents, which is often more prominent in males. For females, there is also excessive maternal mortality at these ages. Then the mortality rates increase steadily into old ages due to senescence. Given this unique shape, there is hardly any simple parametric function that could provide adequate fit to the whole age range. Researchers have proposed several models that involve quite complicated forms in order to fully capture the age pattern in mortality schedules, which will be explored in more details in Chapter 3.

Another feature that is often observed in mortality is the excess of male mortality over females, i.e. male mortality are often higher than female mortality. In fact, the female life expectancies are higher than male life expectancies across almost all the world (Barford et al., 2006). The gender gap has been widening in the period 1950-1970 and declining afterwards (Schünemann et al., 2017; Luy, 2003). Luy (2003) provides a review of the potential explanations of the male

excess mortality in the literature, which can be divided broadly into two basic categories: the biological approach and the non-biological approach. The biological approach states that women are less prone to disease due to physiological and genomic reasons (Smith and Warner, 1989; Waldron, 1985; Lopez, 1983), while the non-biological attempts to explain the sex differences based on behavioral factors such as smoking and alcohol consumption (Oksuzyan et al., 2008; Waldron, 1985). Sex differences in mortality exist even among infants and children, where the higher male mortality rates could not have been caused by behavioral differences, hence confirming the biological contribution to the female mortality advantage. It is concluded that sufficient explanation should be based on both biological and non-biological factors (Schünemann et al., 2017; Oksuzyan et al., 2008; Luy, 2003). Nonetheless, it is evident that there exists a sex differential in mortality, in particular, female mortality is lower than male mortality, which is a fundamental assumption used in the thesis and a key criteria that is used to assess the reasonableness of the mortality estimates.

In addition to the peculiar age patterns in mortality, a challenge often associated with mortality modelling is the lack of data at old ages. As mortality at these ages is the highest, there are much lower or even no exposures. Mortality modelling at old ages have been explored and analysed (Pitacco, 2016; Saikia and Borah, 2014), and there are controversies about whether the human lifespan is limited (Rootzén and Zholud, 2017; Antero-Jacquemin et al., 2015; Couzin-Frankel, 2011; Oeppen and Vaupel, 2002; Olshansky et al., 1990). A decelerating rate of increase in mortality at the oldest ages has been noted in the literature (Rootzén and Zholud, 2017; Pitacco, 2016; Saikia and Borah, 2014; Carriere, 1992). Rootzén and Zholud (2017) focused on the mortality of supercentenarians using Extreme Value Theory and found that for most developed countries the mortality rates after age 110 can be adequately described by a constant. In addition, they found that there are no significant differences in survival at extreme ages between sexes, between lifestyles or genetic backgrounds, between different time periods or between countries.

The concept of 'frailty' has been introduced to account for heterogeneity in mortality as well as to explain the decelerating rate of increase at the oldest ages (Pitacco, 2019; Wienke, 2014; Vaupel et al., 1979; Beard, 1959). The core idea of frailty is that the population consist of mixtures of sub-populations with different frailties, i.e. different levels of mortality. As the cohort advances to higher ages, the more frail individuals/populations would be more susceptible to dying. Therefore as age increases, the frailty of the whole population becomes more concentrated. At the oldest ages where frailty of the surviving population is the lowest and most concentrated, a levelling-off trend or a 'mortality plateau' can then be observed. While frailty is not addressed in this thesis, we acknowledge the potential implications it may have on the estimates. When frailty is taken into account, a more dispersed marginal distribution will be allowed. For example, under the usual Poisson assumption on the number of deaths, this would allow a variance higher than the mean on the marginal distribution, which otherwise maybe too restrictive as over-dispersion is usually observed in mortality data. If frailty is considered, then one should expect the over-dispersion to decrease in age, as frailty is most concentrated at the

oldest ages. For a more detailed discussion on different forms of frailty and heterogeneity, see Pitacco (2019).

Forecasting on the other hand is also challenging due to the many factors involved, such as medical advancements, natural disasters, epidemics, political changes and climate changes etc. Explanatory modelling and forecasting has found little success (Booth and Tickle, 2008). Modelling mortality by exogenous covariates is in fact being recommended *against* by several institutions such as the UK Government Actuary's Department and the USA Social Security Administration (Booth and Tickle, 2008). The main reason for this is that the decomposition of the driving force in mortality is difficult in practice as it is impossible to exhaust all possible covariates and hence leads to high risk of model misspecification. Another reason is that at the oldest ages, the reported causes of death maybe unreliable. McNown and Rogers (1992) concluded that there is no consistent gain in accuracy from cause-of-death decomposition. Thus, models using age and time as covariates are far more popular and accepted, unless it is cause-specific mortality that is of interest.

A key issue when projecting mortality is that the future mortality rates will not always be a continuation of the past trends, and having more historical data does not necessarily improve the quality of the projections. In the literature a somewhat linear trend is usually assumed when projecting mortality rates, with the fitting period chosen such that the linearity fits. In this thesis it is also demonstrated a way to incorporate expert opinion into mortality projection, such that projected trends are moderated by experts that produce more robust and sensible forecasts. Another issue to consider is the plausibility of the projections, or 'biological reasonableness' as described by Cairns et al. (2008). As mentioned above, female mortality rates are expected to be lower than that of males, as a gender gap in mortality is observed across almost all the world. In addition, a global convergence in mortality is also observed (Wilson, 2011), therefore divergences in mortality trends of the same sex across countries considered in this thesis are unlikely. Finally, the mortality age pattern is expected to be monotonically upwards, albeit possible flattening at the oldest ages.

In the following, definitions and notations in mortality modelling are introduced, followed by a discussion of the data used in this thesis.

1.1 Mortality Rate, Probability of Death and Force of Mortality

Let d_x denotes the number of deaths of individuals aged x last birthday and E_x the total exposure of individuals exact age x within some period of time (a year in our case). Then the central exposed to risk at age x is defined as

$$E_x^C = \int_0^1 E_{x+s}\, ds, \tag{1.1}$$

which is an averaged total exposure within the year. E_x is sometimes also called the initial exposed to risk or denoted by E_x^0 .

Let T be the time to death of a newborn (i.e. aged exactly 0). Define $p_x = P(T > x+1|T > x)$ as the probability of an individual aged x surviving for 1 year, then the 'probability of death' of an individual aged x within a year is defined as $q_x = 1 - p_x = P(T < x + 1|T > x)$. Generalisation of the survival and death probability of an individual aged x for t years are ${}_tp_x = P(T > x + t|T > x)$ and ${}_tq_x = P(T < x + t|T > x)$, respectively. Then the 'force of mortality' is defined as

$$\mu_{x+t} = \lim_{\delta t \to 0} \frac{P(T \leq x + t + \delta t|T > x + t)}{\delta t}, \tag{1.2}$$

$$= \lim_{\delta t \to 0} \frac{P(x + t < T < x + t + \delta t|T > x)}{\delta t} \frac{1}{P(T > x + t|T > x)}, \tag{1.3}$$

$$= \frac{1}{{}_tp_x} \lim_{\delta t \to 0} \frac{{}_tp_x - {}_{t+\delta t}p_x}{\delta t}, \tag{1.4}$$

$$= \frac{1}{{}_tp_x} \lim_{\delta t \to 0} \frac{-({}_{t+\delta t}p_x - {}_tp_x)}{\delta t}, \tag{1.5}$$

$$= -\frac{1}{{}_tp_x} \frac{d}{dt} {}_tp_x, \tag{1.6}$$

$$= -\frac{d}{dt} \log {}_tp_x. \tag{1.7}$$

Force of mortality is a transition intensity or a hazard rate. Therefore, the probability of death q_x and force of mortality μ_{x+t} are related through

$$q_x = 1 - p_x = 1 - e^{-\int_0^1 \mu_{x+s}\, ds}. \tag{1.8}$$

The central mortality rate is then defined as

$$m_x = \frac{\int_0^1 E_{x+t}\mu_{x+t}\, dt}{\int_0^1 E_{x+t}\, dt} = \frac{\mathbb{E}(d_x)}{E_x^C}, \tag{1.9}$$

which can be viewed as the average of the force of mortality weighted by the corresponding exposures in $[x, x + 1)$.

Throughout the thesis the term 'mortality rate' refers to the central mortality rate, not to be confused with 'probability of death' or 'force of mortality' (hazard).

The crude mortality rates $\tilde{m}_x$ are obtained by substituting $\mathbb{E}(d_x)$ with the observed numbers of deaths and E_x^C with the exposed-to-risk (or mid-year population) estimates. However the mid-year population estimate sometimes may not be a good approximation to the central exposed to risk when the death distribution is highly asymmetric. Pointed out by Cairns et al. (2014), the cohort born around 1920 displays high unevenness in their death distribution. Cairns et al.

(2014) coined the term 'phantom effect', due to the fact that the unevenness results in an over-estimation of the central exposed to risk by the mid-year population estimate, resembling the existence of 'phantoms' inflating the exposure size.

Often simplifying assumptions are made for the relationships between the mortality rate, the probability of death and the force of mortality. For instance, a constant force of mortality can be assumed such that $\mu_{x+t} = \mu_{x+1/2} = m_x \quad \forall t \in [0,1)$, hence the death probability $q_x = 1-\exp(-\int_0^1 \mu_{x+1/2}\, ds) = 1-\exp(-m_x)$. Another assumption that is often used is the uniform distribution of deaths assumption (UDD), under which we have $q_x = \frac{m_x}{1+\frac{1}{2}m_x}$.

The number of deaths are usually assumed to have a Poisson distribution with mean $E_x^C m_x$,

$$d_x \sim \text{Poisson}(E_x^C m_x). \tag{1.10}$$

In the literature it is acknowledged that the Poisson assumption for the number of deaths might be too restrictive as this assumes equal mean and variance. Often an over-dispersion can be noticed. To allow for over-dispersion, the number of deaths can be assumed to have a Negative Binomial distribution, where there is an extra dispersion parameter which allows variance to be different than the mean. This can also be expressed as a Poisson-Gamma mixture as follows

$$z \sim \text{Gamma}(s, s) \tag{1.11}$$

$$d_x|z \sim \text{Poisson}(z \cdot E_x^C m_x). \tag{1.12}$$

Then the marginal distribution for d_x will be a Negative Binomial with mean $E_x^C m_x$ and variance $E_x^C m_x(1 + E_x^C m_x/s)$. Here s is the dispersion parameter. Comparing to a Poisson where the mean and variance are equal, the Negative Binomial distribution allows a different variance than the mean.

1.2 Data

The data is obtained from the Human Mortality Database (HMD) which contains the number of deaths and the size of the exposures at risk for both males and females at each age in England and Wales (EW) and Scotland (SC) from 1961 to 2016 for ages 1 to 109. However, sometimes at the highest ages there are no exposures, rendering the central mortality rates undefined. Therefore, ages with no exposure in any year within the time period are discarded, resulting in a data set spanning from age 1 to 104 for England and Wales and from age 1 to 99 for Scotland.

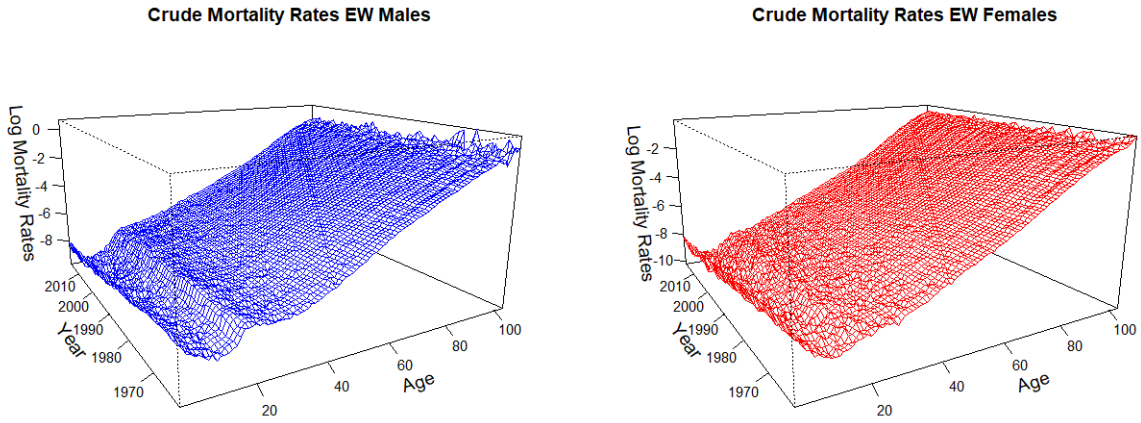


Figure 1.1: Crude mortality rates of EW males and females [1]

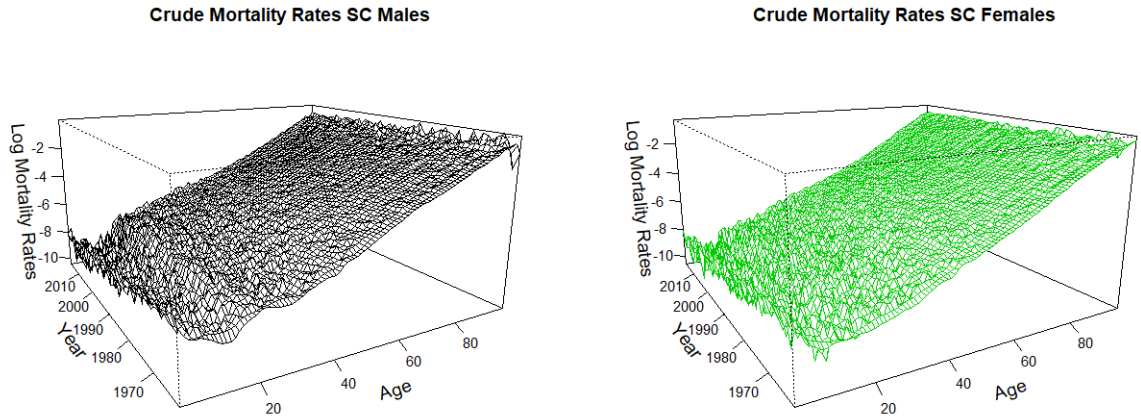


Figure 1.2: Crude mortality rates of SC males and females [1]

Figures 1.1 and 1.2 present the crude mortality rates for males and females in EW and SC respectively. Some features can be immediately noticed. For each year, the early childhood mortality rates are exceptionally high, which then decrease to the lowest level at around age 10, followed by a bump at late teenage and young adulthood called the 'accident hump', and steadily increasing mortality rates into the oldest ages. The accident hump is more prominent in males, regardless of the country. The youngest and the oldest mortality rates display more random variations than adult mortality rates (about ages 40 to 80). In addition, the variations in

[1] Data obtained from the Human Mortality Database. Accessed on 22nd July, 2019.

males are larger than that of females and the variations in Scotland are much larger than that in England and Wales, possibly due to lower exposures.

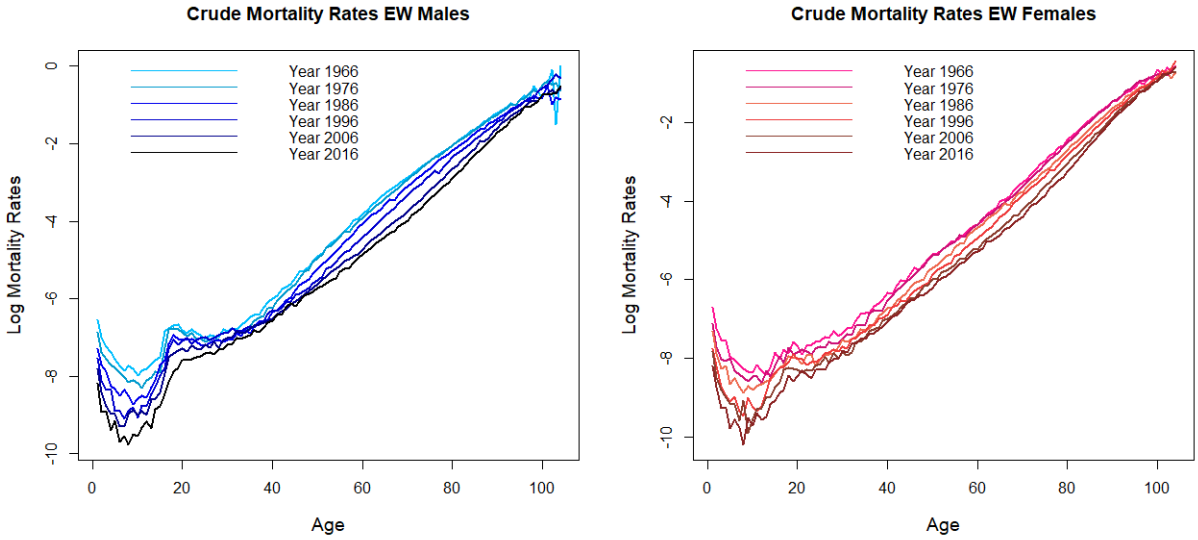


Figure 1.3: Crude mortality rates of EW males and females in some selected years [2]

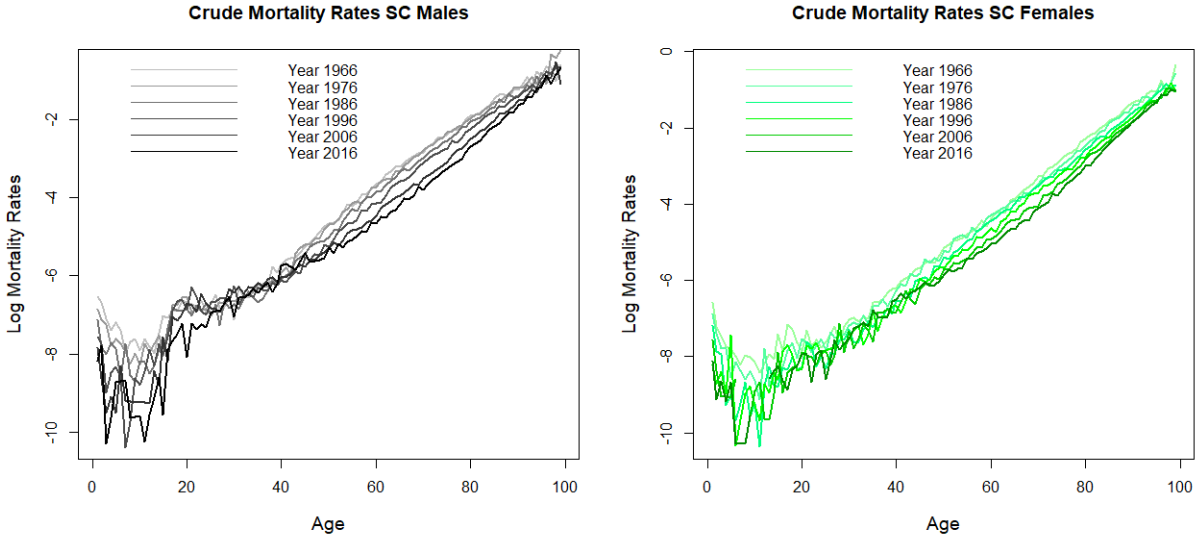


Figure 1.4: Crude mortality rates of SC males and females in some selected years [2]

Figures 1.3 and 1.4 present the mortality profiles for males and females in years 1966, 1976, 1986, 1996, 2006 and 2016, for EW and SC respectively. It is evident that mortality rates at most ages are decreasing in all populations. For each country, mortality improvements in males seem to be larger at about ages 1-20 and 60-80 than remaining ages, while the mortality improvements in females are more even across ages. There is little mortality improvement at the oldest ages. The mortality rates for Scottish males at ages around 40 seem to be stagnated.

[2]Data obtained from the Human Mortality Database. Accessed on 22nd July, 2019.

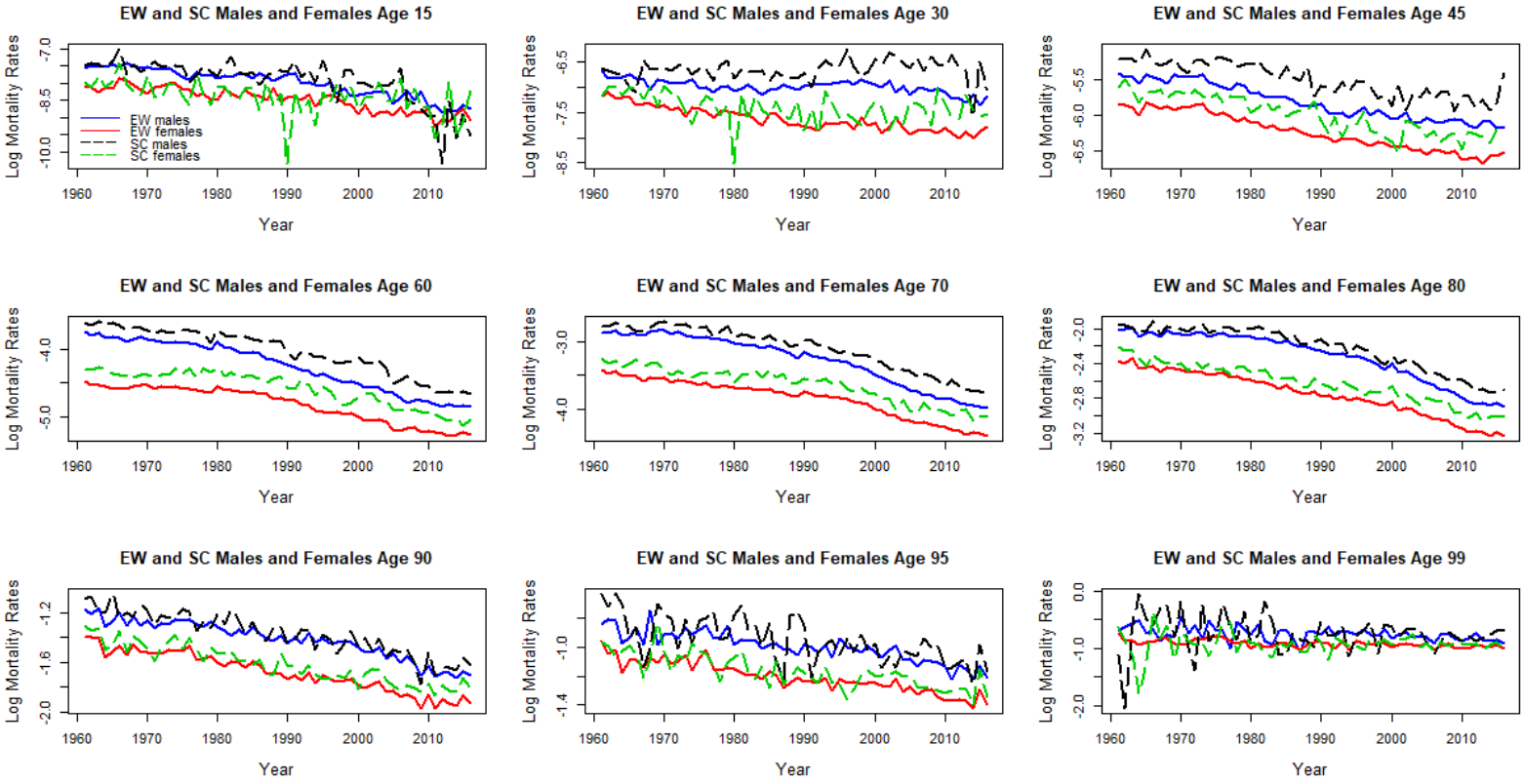


Figure 1.5: Crude mortality rates of EW males (blue), EW females (red), SC males (black) and SC females (green) at some selected ages between 1961 and 2016 [3]

[3] Data obtained from the Human Mortality Database. Accessed on 22nd July, 2019.

Figures 1.5 presents the male and female crude mortality rates at some selected ages between 1961 and 2016, for EW and SC respectively. Some aforementioned features can be noticed again: the difference between male and female mortality rates (blue against red / black against green) is smaller at higher ages and the variations in male mortality rates are higher especially at the oldest ages compared to that of females. This is because female exposures are higher than male exposures. The male and female mortality rates of each country also seem to be correlated. For both males and females, the difference between the mortality rates of EW and SC (blue against black / red against green) gets smaller as age increases. The Scotland mortality seems to be improving at a slightly slower rate for mid ages (about ages 30 to 80). In addition, it can be confirmed that female mortality rates of each country are almost always lower than that of males and mortality trends of the same sex are non-divergent.

The death counts have been pre-processed by HMD, which are minor corrections and would not impact the results substantially. For England and Wales data, age-heaping problem is observed in data during World War I and between the two World Wars (Philipov et al., 2020). However since the analysis is using data since 1961, this does not have any impact on the results. For Scotland, the documentation is still in preparation. However, a cross check on the input raw data and the HMD death counts verifies that two adjustments have been made. Firstly, for year 1961 and 1962, only the total death counts for 5-year age groups are available in the raw data, HMD disaggregated them into corresponding death counts in single year of age. Secondly, for year 1996, 1998, 2002 and 2003, the HMD death counts at some ages differ from the raw data by 0.01-0.07, which could possibly be the result of distributing deaths of unknown age. Nonetheless these minor differences occur only at some ages and would have little to none impact on estimation.

1.3 Agenda

The rest of the thesis is organised as follows. In Chapter 2, the statistical preliminaries are discussed. In particular, a brief introduction is given on Generalised Linear Model (GLM), Generalised Additie Model (GAM) and Spline, with a focus on P-spline and its estimation. In Chapter 3, a review is given on the state-of-the-art mortality models, beginning with models for static mortality age patterns, followed by mortality forecasting models as well as coherent mortality forecasting models. In Chapter 4, the proposed model for mortality graduation is introduced, which is then extended to mortality projection models in Chapters 5 and 6. In Chapter 7, the proposed models are estimated in the Bayesian framework. Chapter 8 concludes and gives possible further work on the thesis.

Chapter 2

Introduction to Generalised Linear Model, Generalised Additive Model and Spline

In this chapter some preliminary background knowledge is briefly reviewed. It serves as an introduction but not a detailed discussion on the topics.

2.1 Generalised Linear Model and Generalised Additive Model

Generalised Linear Models (GLMs) are used widely in statistical analyses. A GLM has three main important ingredients, namely an exponential family distribution EF, a link function g and a linear predictor η. It has the basic structure

$$Y_i \sim EF(\mu_i) \qquad \text{and} \qquad g(\mu_i) = \eta_i = \boldsymbol{X}_i\boldsymbol{\beta}, \tag{2.1}$$

where $\mu_i = \mathbb{E}(Y_i)$, g is a smooth monotonic function, $\boldsymbol{X}_i$ is the i-th row of the design matrix $\boldsymbol{X}$ and $\boldsymbol{\beta}$ is a vector of parameters. Y_i's are assumed to be independent and follow an exponential family distribution.

The link function is usually chosen such that the transformation on the linear predictor is meaningful with respect to the exponential family distribution. For example, for a Poisson distribution a log link function is usually used so that the inverse transformation of the linear predictor is always positive.

Estimation of the parameters is carried out by maximising the (log)likelihood, which can be achieved efficiently by Iteratively Re-weighted Least Squares (IRLS). Given the observed data vector $\boldsymbol{y}$ and the current estimates of the mean $\boldsymbol{\mu}^{[t]}$ at iteration t, estimate $\boldsymbol{\beta}^{[t+1]}$ for the next iteration $t+1$ by minimising the sum of squares $\|\sqrt{\boldsymbol{W}^{[t]}}(\boldsymbol{z}^{[t]} - \boldsymbol{X}\boldsymbol{\beta})\|^2$, where $\boldsymbol{z}^{[t]}$ is the

'pseudo-data' at iteration t and $\boldsymbol{W}^{[t]}$ is a diagonal matrix with entries equal to the 'working weights' at iteration t given by

$$z_i^{[t]} = g'(\mu_i^{[t]})(y_i - \mu_i^{[t]}) + \eta_i^{[t]}, \tag{2.2}$$

$$W_{ii}^{[t]} = \frac{1}{V(\mu_i^{[t]})g'(\mu_i^{[t]})^2}, \tag{2.3}$$

$$V(\mu_i^{[t]}) = VAR(Y_i)/\phi. \tag{2.4}$$

where ϕ is the scale parameter of the exponential family distribution (e.g. for a Poisson distribution, $\phi = 1$). The process is iterated until convergence is reached.

Generalised Additive Models (GAMs) are similar to GLMs in a way that the expected values of the response variables are again described by a linear predictor through a link function, and the response variables $\boldsymbol{Y}$ are assumed to be independently distributed of some exponential family distribution. The main difference between a GAM and a GLM is that while in a GLM a pre-defined parametric structure on the relationship between the response and the covariates is ncecessary, a GAM allows a more flexible framework. Specifically, a GAM assumes that the linear predictor η is a linear sum of smooth (and parametric) functions of the covariates, i.e.

$$g(\mu_i) = \eta_i = \boldsymbol{X}_i\boldsymbol{\beta} + s_1(z_{1i}) + s_2(z_{2i}) + s_3(z_{3i}, z_{4i}) + \ldots, \tag{2.5}$$

where $\boldsymbol{X}_i$ is the i-th row of the model matrix containing covariates of parametric structure, and $s_j(\boldsymbol{z})$ are smooth functions of covariate(s) $\boldsymbol{z}$. The smooth functions are usually estimated using splines.

2.2 Introduction to Spline Smoothing

In this section, splines are briefly introduced. Splines are non-parametric smoothers for curve estimation. A spline is a piece-wise polynomial of degree $p-1$ joined together at n knots $\mathbf{t} : \{t_1, t_2, ..., t_n\}$, in a way such that it is continuous up to the l-th derivative, $l \leq p-2$. In analogue to the notation used by De Boor et al. (1978), we denote the space of splines of order p (the order of a spline is defined as the degree of the constituent polynomials plus one) and knot sequence $\mathbf{t}$ with $\$_{p,\mathbf{t}}$. For example, the most commonly used splines are the cubic splines that are continuous up to the second derivative, and therefore the cubic spline space $\$_{4,\mathbf{t}} = \{f : f, f', f''\text{are continuous}\}$. If we define f_i to be the i-th piece of polynomial that constitutes the cubic spline, then we have

$$\begin{gathered} f_i'(t_{i+1}) = f_{i+1}'(t_{i+1}), \\ \text{and} \\ f_i''(t_{i+1}) = f_{i+1}''(t_{i+1}). \end{gathered} \tag{2.6}$$

2.2.1 B-spline

There are many different ways to set up a basis for a spline space. One of the most popular basis is the B-spline basis, which will be used throughout this paper. Each B-spline basis function is only non-zero over the $p+1$ neighbouring knots. The locality of the B-spline basis functions means that each basis function exerts only little effect on basis functions that are far away, hence improving numerical stability and efficiency. De Boor et al. (1978) provided a recursive algorithm of setting up such a basis easily. Figure 2.1 shows a B-spline basis with 40 basis functions of order 4 with equally spaced knots. Note that the B-spline spans the space $\$_{p,\mathbf{t}:\{t_1,t_2,....t_{k+p}\}}[t_p, t_{k+1}]$ where k is the number of basis, that is, to have a spline spanning the interval $[a, b]$ we need to specify $2p-2$ knots outside the interval, $p-1$ knots at each end. At every point in the domain, the B-spline basis functions sum to unity.

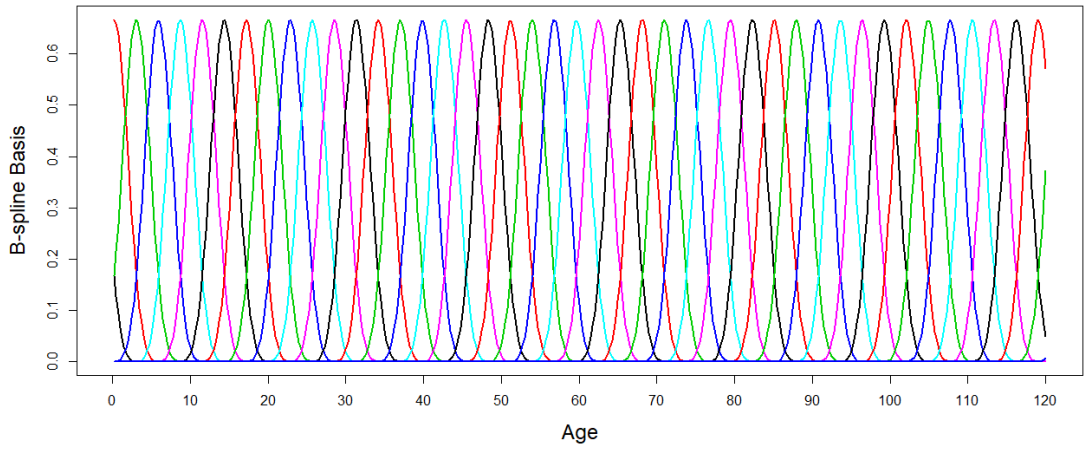


Figure 2.1: cubic B-spline basis

From Figure 2.1 we can see that each basis function is non-zero only over a fixed interval. For cubic splines, at any age that is not a knot, there are exactly 4 basis functions that are non-zero; while at each knot there are exactly 3 non-zero basis functions.

Let $B_i(x)$ denotes the i-th B-spline basis functions, then the spline is the sum of the basis functions, $\sum_{i=1}^{k} \beta_i B_i(x)$ where β_i is the coefficient of the i-th basis function. Given that the B-spline basis functions sum to unity and that they are only local over some knots, the function value at any point can be viewed as a weighted average of neighbouring coefficients, with weights given by the basis function.

The high flexibility makes splines susceptible to over-fitting. For example, in Figure 2.2, the mortality rates are fitted with a B-spline with k=60 (this is chosen to highlight the over-fitting behavior). It is clear that fitted values are erratic and implausible, especially at the youngest and oldest ages where the mortality rates are highly variable.

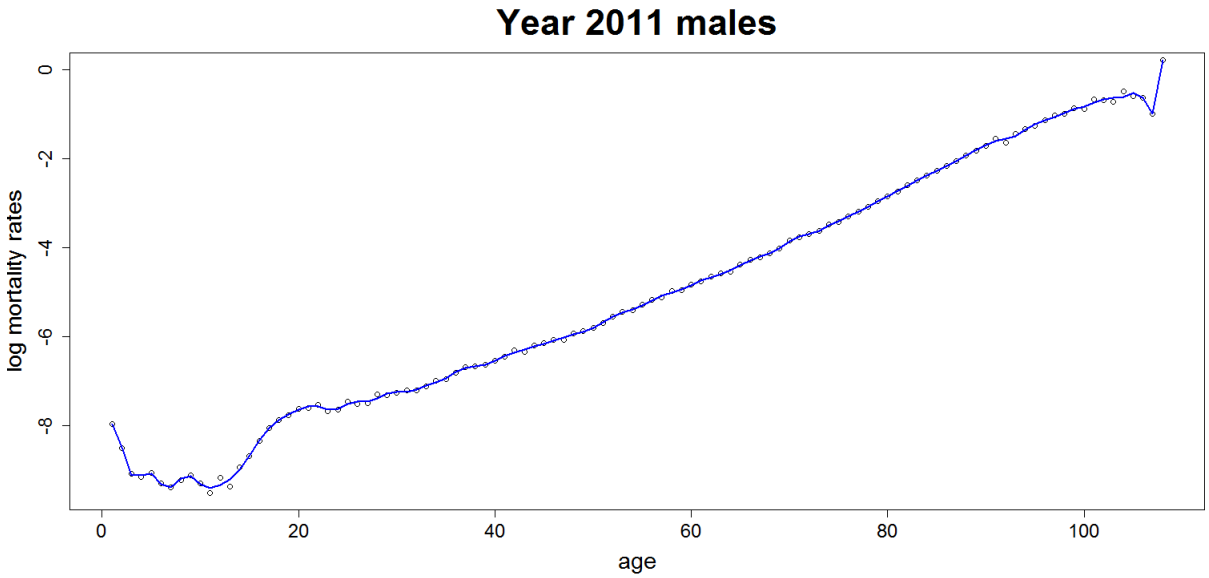


Figure 2.2: Fitted values using a rank 60 B-spline with equally spaced knots

2.2.2 Spline fitting

2.2.2.1 Optimal Knot Selection

It is seen that naive fitting of splines may suffer from over-fitting. One remedy would be to use fewer knots and hence a smaller basis. As a result the basis functions will span a larger interval, thus relying on more data. Conceptually this is equivalent to having a wider bandwidth in kernel estimation. However, a too sparse basis might lead to over-smoothing. Therefore choosing the appropriate number and position of knots is essential. Information criteria such as AIC, BIC or GCV can be used when choosing the optimal knot sequence as they are objective measures that take into account both the goodness of fit and parsimony. This then becomes a variable selection problem where the relevant basis functions from a candidate set that minimises the criterion are chosen. Some adaptive knot searching methods have also been developed to estimate the optimal number and position of knots, such as the evolutionary algorithm, with mutation, insertion and deletion of knots.

2.2.2.2 Penalised Splines

The decision of an optimal knot sequence (both the number and location) can be a complicated and time consuming task. Another approach is to set up a basis with a dimension generous enough to capture all the variations in the function, and then introduce a penalty term to the fitting objective such that over-fitting is penalised. These are called penalised splines. Currie and Durban (2002) and Currie et al. (2004) suggest placing one knot for every four or five data points. The penalty controls the parsimony and smoothness of the fitted curve. A commonly

used smoothness penalty is the integrated squared second derivatives $\int [f''(x)]^2 dx$, which measures the total curvature of the spline (Wood, 2006). In other words, rapid changes in the slope of the function are penalised. Under this approach, the fitting objective is to maximise the penalised (log)likelihood, given by $l(x) - \frac{1}{2}\lambda \int [f''(x)]^2 dx$, where $l(x)$ is the log-likelihood of the data and λ is a positive constant. The λ is called the 'smoothing parameter' and it controls how heavily the fitting is penalised and hence how much weight is given into producing a smooth curve. This new fitting objective represents and balances two often conflicting aims in curve fitting, namely data fidelity and function smoothness.

Estimation of the coefficients of penalised splines with squared penalty (symmetric penalty) is straightforward provided that the smoothing parameters are known or fixed. The squared penalty can be incorporated into a GAM in a natural way. Having set a fixed basis, a GAM can be expressed as a GLM, where the design matrix $\boldsymbol{X}$ are columns of the B-spline basis functions. For normally distributed data, the penalised likelihood can be written as

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda \boldsymbol{\beta}' \boldsymbol{S} \boldsymbol{\beta} \tag{2.7}$$

where $\boldsymbol{S}$ is the penalty matrix. Let $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{S}$, then we can write the penalised likelihood as

$$\left\| \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \boldsymbol{X} \\ \sqrt{\lambda}\boldsymbol{P} \end{pmatrix} \boldsymbol{\beta} \right\|^2. \tag{2.8}$$

Therefore, to incorporate the squared penalty into estimation, one simply needs to augment the data with zeroes and append the square root of the penalty matrix to the design matrix when maximising the penalised likelihood. For non-Gaussian data, this can be done by Penalised Iteratively Re-weighted Least Squares (P-IRLS) and within each working model, estimation is the same as the Gaussian case but with weighted design matrix and weighted pseudo-data instead (Equation 2.4).

2.2.3 P-splines

The integration of the squared second derivative of a B-spline can be complicated. Eilers and Marx (1996) proposed a discretised version of the penalty to be used with the B-spline basis, the squared second differences of the coefficients, $\lambda \sum (\bigtriangledown^2(\beta_i))^2$. This penalised B-spline, hence the name 'P-spline', has been widely used due to its simplicity and ease of implementation. It also has some interesting properties. For example, unlike kernel smoothers, P-splines show no boundary effects, i.e. the fitted curve outside the domain of the data does not bend towards zero. Another attractive property of the P-spline is that the moments in the data are conserved, which is particularly useful in density estimation. According to Eilers and Marx (1996), '*B-splines and difference penalties are the ideal marriage*.'

When using P-splines, the square root of the penalty matrix $\boldsymbol{P}$ is then the $(k-2)\times k$ second order difference matrix $\boldsymbol{D}_2$, that is

$$\boldsymbol{P}=\boldsymbol{D}_2=\begin{bmatrix}1 & -2 & 1 & 0 & 0 & 0 & \cdots\\ 0 & 1 & -2 & 1 & 0 & 0 & \cdots\\ 0 & 0 & 1 & -2 & 1 & 0 & \cdots\\ \vdots & \vdots & \vdots & & \ddots & & \end{bmatrix}, \tag{2.9}$$

and

$$\boldsymbol{S}=\boldsymbol{P'P}=\begin{bmatrix}1 & -2 & 1 & 0 & 0 & 0 & 0 & \cdots\\ -2 & 5 & -4 & 1 & 0 & 0 & 0 & \cdots\\ 1 & -4 & 6 & -4 & 1 & 0 & 0 & \cdots\\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \cdots\\ 0 & 0 & 1 & -4 & 6 & -4 & 1 & \cdots\\ \vdots & \vdots & \vdots & \vdots & & \ddots & & \end{bmatrix}. \tag{2.10}$$

2.2.4 Estimation of Smoothing Parameters

When estimating the smoothing parameters it is preferred to have an automated and objective method. As mentioned before, criteria such as the AIC, BIC or GCV are often used for this purpose. When fitting penalised splines, the coefficients are not entirely 'free' anymore, in the sense that they are constrained or shrunk in a way to produce smooth functions. Intuitively, under penalisation, one parameter does not always enjoy a full degree of freedom, its effective degree of freedom is reduced to somewhere between 0 and 1. Therefore, contrary to the usual cases where the complexity of the model can be measured by the number of parameters, in penalised likelihood estimation, the complexity of the model is measured by the 'effective number of parameters', which is essentially the sum of effective degrees of freedom of all the parameters, defined as the trace of the hat matrix (or influence matrix) $tr(\boldsymbol{A})$, where $\boldsymbol{A}=\boldsymbol{X}(\boldsymbol{X'X}+\lambda\boldsymbol{S})^{-1}\boldsymbol{X'}$. Therefore when calculating scores such as the BIC, the number of parameters has to be replaced by the effective number of parameters. Here we choose to minimise the BIC.

Estimation of smoothing parameters can be based on performance iteration (Gu, 1992) or outer iteration (O'Sullivan et al., 1986). Performance iteration estimates the smoothing parameters by minimising the BIC (or other scores) within each working model of the P-IRLS, while outer iteration estimates the smoothing parameters by minimising the BIC at the final estimates of the coefficients. The advantage of performance iteration is that the gradient and the Hessian can be calculated analytically with less computational cost and hence optimisation of the BIC can be done efficiently, however it does not exactly minimise the BIC of the actual model (see Wood (2006) for a more detailed discussion about these two methods). Our investigations suggest that the estimates from the two methods do not differ a lot, hence we prefer to use performance iteration in this thesis as it is computationally more efficient.

2.2.5 Adaptive Penalty

As discussed in section 2.2.3, a commonly used penalty for P-spline is the sum of squared second differences of the spline coefficients with a smoothing parameter λ controlling the weight. Therefore, we have one smoothing parameter controlling the overall smoothness. The disadvantage of this is that when the underlying function displays varying smoothness, an overall smoothness penalty may not suit well. Therefore, instead of having just one smoothing parameter governing the global smoothness, the penalty can be made to vary along the domain, i.e. $\sum \lambda_i(\bigtriangledown^2(\beta_i))^2$. In other words, different regions are penalised with different weights, allowing varying smoothness throughout the domain. We shall call this the 'local penalty' or 'adaptive penalty', and the opposite the 'global penalty'. Ruppert and Carroll (2000) show that adaptive smoothing is more effective and gives better results than ordinary splines, especially when the underlying function has varying smoothness and performs at least as well as traditional splines even when the underlying function is uniformly smooth.

Different approaches have been suggested to model the varying smoothing parameters λ_i. Pintore et al. (2006) and Liu and Guo (2010) used piece-wise constant functions for the varying penalty. Ruppert and Carroll (2000) and Krivobokova et al. (2008) proposed to use a second layer of linear spline (with less knots) for the penalty. Storlie et al. (2010) suggested to estimate the changing smoothness from an initial fit with ordinary spline while Yang and Hong (2017) recommended to weight the penalty inversely to the volatility of the data in proximity. Bayesian adaptive splines have also been investigated, see Baladandayuthapani et al. (2005), Crainiceanu et al. (2007), Jullion and Lambert (2007), Scheipl and Kneib (2009) and Yue et al. (2012).

In our models we propose to simply use an exponential function for the smoothing function λ_i, because firstly it has only two parameters and secondly we believe that the mortality rates are smoother as age increases, i.e. the λ_i should be increasing in age and penalises roughness at the older ages more than the younger ages. In addition, at the very old ages data are often very scarce and the exposures are extremely low, which are very unreliable and display high random variations. By assigning heavier penalty at the old ages the fitted spline is more parsimonious and more strength can be borrowed from neighbouring ages, hence improves robustness and stability.

2.2.6 Parameter Uncertainty

It is important to incorporate uncertainty in our estimates. Interval estimates reflect uncertainty around the estimates and thus give a better illustration of the fitted values. For Gaussian homoskedastic data, the estimated coefficients are given by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X} + \boldsymbol{S})^{-1}\boldsymbol{X'y}$, therefore $\hat{\boldsymbol{\beta}} \sim N(\mathbb{E}(\hat{\boldsymbol{\beta}}), (\boldsymbol{X'X} + \boldsymbol{S})^{-1}\boldsymbol{X'X}(\boldsymbol{X'X} + \boldsymbol{S})^{-1}\sigma^2)$, where σ^2 is the variance of the response variable. However, since splines are biased smoothers, $E(\hat{\boldsymbol{\beta}}) \neq \boldsymbol{\beta}$ in general, confidence intervals based on this result may not be accurate. An alternative is to use a Bayesian approach to incorporate uncertainty in a natural manner (Wood, 2006). Expressing the penalty as a normal

prior on the coefficients, it can be shown that the posterior distribution of the coefficients is $\boldsymbol{\beta}|\boldsymbol{y} \sim N(\hat{\boldsymbol{\beta}}, (\boldsymbol{X'X} + \boldsymbol{S})^{-1}\sigma^2)$. In this way we can obtain credible intervals of the parameters. For non-Gaussian data, the posterior distribution can be approximated by the final working model at convergence of the P-IRLS algorithm, i.e. $\boldsymbol{\beta}|z \sim N(\hat{\boldsymbol{\beta}}, (\boldsymbol{X'WX} + \boldsymbol{S})^{-1}\phi)$, where $\boldsymbol{z}$ and $\boldsymbol{W}$ are the pseudo-data, iterative weights of the final working model in P-IRLS and ϕ is the scale parameter of the data distribution. Wood (2006) provides a more detailed explanation and proofs of the distributional results.

Chapter 3

Models of Mortality Graduation and Forecasting

In this chapter a review of the state-of-the-art mortality models is given, starting from models for static age patterns of mortality, then models for mortality forecast, followed by models for coherent mortality forecast. The chapter finishes by a section outlining the main contributions of this thesis.

3.1 Mortality Graduation

3.1.1 Adult Mortality Graduation

Crude mortality rates often exhibit natural randomness and irregular patterns, graduation of mortality rates refers to the act of smoothing crude mortality rates. Sometimes extrapolation to higher ages is also performed during the process since the usual assumed maximum lifespan of mankind (e.g. 125) is often beyond the range of available data. Mortality graduation models, sometimes also called 'laws of mortality', have become more and more sophisticated over time. One of the earliest attempts is the Gompertz law (Gompertz, 1825). The Gompertz law of mortality states that the mortality rates is an exponential function of age, i.e. the log of mortality rates is linear in age,

$$\log(\mu_x) = a + bx,$$

which is then modified by Makeham to the Gompertz-Makeham model (Makeham, 1860) where an age invariant constant is added to capture age-independent mortality (e.g. age-independent risk of accidents):

$$\mu_x = A + BC^x.$$

The Gompertz and Gompertz-Makeham laws are simple and straightforward that fit well to adult mortality, however, at the oldest ages the log-linearity assumption may not be suitable. Carriere (1992) noted that the log-linearity is only applicable to adult ages up to 94, thereafter the (log) mortality trend becomes non-linear. In fact, a decelerating rate of increase is often observed at these ages, therefore the Gompertz and the Gompertz-Makeham model may over-estimate the mortality rates at the oldest ages. Coale and Kisker (1990) proposed to model the rate of change as a linear function in age, which is equivalent to adding a quadratic term to the Gompertz law (Pitacco, 2016; Saikia and Borah, 2014),

$$m(x) = Ae^{Bx+Dx^2}.$$

This does provide a decelerating rate of increase in mortality rates, however, the applicable age range is restricted to the oldest ages due to the quadratic term (after age 85 as suggested by the authors).

Perks (1932), Beard (1959) and Thatcher (1999) suggested the use of logistic functions,

$$\mu_x = \frac{cz}{1+z} + \gamma,$$

where $z = e^{a+bx}$. The logistic model produces mortality rates that level-off to an asymptote, c, at the highest ages. In addition to the decelerating increase in mortality rates at the highest ages, at younger adult ages where mortality rates are lower, the logistic curves behave similarly to the Gompertz or Gompertz-Makeham laws, hence is able to capture not only the decreasing rate of increase in mortality rates at the oldest ages but also the steadily increasing adult mortality rates.

Instead of using a logistic function, Lindbergson (2001) suggested a piecewise model such that for ages between a cut-off age w the mortality rates follow the Gompertz-Makeham law, while after the cut-off age they are modelled by a linear function, i.e.

$$m(x) = \begin{cases} Ae^{Bx} + C & \text{for } x \leq w \\ Ae^{Bw} + C + D(x-w) & \text{for } x > w \end{cases}.$$

After the cut-off point, the Linbergson model behaves as the Weibull function with rate 1, therefore the Lindbergson model can be viewed as a compromise between the Gompertz-Makeham and Weibull model. Saikia and Borah (2014) did a comparative study on these models for the oldest ages and showed that the logistic model is the most reasonable choice among the laws mentioned above. Pitacco (2016) provides an excellent and comprehensive review of a range of old age models, and discussed the age pattern of mortality from different perspective.

3.1.2 Full Age Range Mortality Graduation

The aforementioned models are relatively simple and are only applicable to adult ages. Yet, the extension to the whole age range is not so straightforward due to the shape of the infant

and child mortality rates and the presence of the accident hump at teen years. Several more complicated mathematical functions have been proposed to model mortality rates for the whole range. Thiele (1871) proposed the following function:

$$m(x) = \underbrace{\varphi e^{-\psi x}}_{\text{negative exponential}} + \underbrace{\lambda e^{-\delta(x-\epsilon)^2}}_{\text{normal}} + \underbrace{A e^{Bx}}_{\text{Gompertz}} \qquad \varphi, \psi, \lambda, \delta, \epsilon > 0,$$

where the negative exponential, normal and Gompertz functions capture the decreasing child mortality, the accident hump and the senescent mortality, respectively. It decomposes the age pattern of mortality into 3 stages, the declining infant-child mortality, the accident hump and the Gompertzian adult mortality. Another model is the 8-parameters Heligman-Pollard Model proposed by Heligman and Pollard (1980) which shares a similar concept. Instead of modelling the mortality rates, they model the log odds of death probability,

$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + De^{-E(lnx-lnF)^2} + GH^x,$$

where q_x is the one-year death probability at age x. Likewise, the function can be viewed as a decomposition of the mortality schedule into three terms capturing the child mortality, accident hump and adult mortality. Carriere (1992) again decomposes mortality into 3 stages and proposed the Carriere Model,

$$\begin{aligned}
S(x) &= \psi_1 S_1(x) + \psi_2 S_2(x) + \psi_3 S_3(x) \\
S_1(x) &= \exp\{-\big(\frac{x}{m_1}\big)^{\frac{m_1}{\sigma_1}}\} \\
S_2(x) &= 1 - \exp\{-\big(\frac{x}{m_2}\big)^{-\frac{m_2}{\sigma_2}}\} \\
S_3(x) &= \exp\{e^{-\frac{m_3}{\sigma_3}} - e^{\frac{x-m_3}{\sigma_3}}\} \text{and} \\
\psi_3 &= 1 - \psi_1 - \psi_2.
\end{aligned}$$

Here $S(x)$ is a survival function, $S_1(x)$, $S_2(x)$ and $S_3(x)$ are survival functions of the Weibull, Inverse-Weibull and Gompertz distributions, respectively. The Carriere Model is a mixture of survival functions, that correspond to the child ($S_1(x)$), young adult ($S_2(x)$) and adult mortality ($S_3(x)$), respectively. The model has eight interpretable parameters of the scales and locations of each of the survival functions. This parametrisation is also equivalent to a multiple-decrement life table (Booth and Tickle, 2008), as the survival function is a mixture of three survival functions that correspond to mortality at different stages. Despite the ability to model mortality rates of the whole range with interpretable parameters, these models are often difficult to fit in practice, due to the high correlation in the estimated parameters. The high correlation also compromises interpretability.

3.1.3 Non-parametric Mortality Graduation

Recent advancement in non-parametric smoothing provides a flexible methodology for smoothing mortality rates. For example, since the 13-th English Life Table (ELT13), subsequent ELTs are all produced using spline-based methods. In ELT14, variable knot cubic spline is adopted and the optimal number and location of knots are estimated. In ELT15 weighted least squares smoothing spline is used, with a twist that the user specifies a set of weights based on their judgment in respect of the regions where the closest fit to the observed data is desired. In addition, some data at the highest ages are discarded to maintain a monotonic (upward) progression in the graduated mortality rates. In ELT16 Geometrically Designed variable knot regression spline (GeDS) (Kaishev et al., 2006) is used. It is a 2-stage method where they first fit a variable knots linear spline, and then estimate the optimal control polygon of higher order splines, resulting in estimates of the optimal knot sequence, the order of the spline and the spline coefficients all together. Detailed information on methods used in the production of previous ELTs can be found in Gallop (2002). In the latest English Life Table, ELT17, Dodd et al. (2018) suggested a hybrid function for mortality graduation. Specifically, the mortality schedule is partially non-parametric (at the younger ages) and partially parametric (at the older ages). Mortality rates before some cut-off age w are modelled using splines while mortality rates after the cut-off age are modelled using a parametric model (log-linear or logistic), i.e.

$$\log(m(x)) \text{ or } \text{logit}_c(m(x)) = \begin{cases} s(x) & \text{for } x \leq w \\ a + bx & \text{for } x > w. \end{cases}$$

At younger ages there is sufficiently dense data, therefore the flexibility of spline makes it a very effective tool in graduating mortality rates at these ages, whereas at the oldest ages a parametric function helps produce more robust estimates and extrapolation. The cut-off point is chosen from a candidate set by minimising the cross-validation error. In other words, the model requires multiple fit depending on the size of the candidate set.

Several mortality graduation models have been introduced, parametric models often have the advantage of having interpretable parameters, however they are usually of complicated non-linear forms and the estimated parameters are usually highly correlated, hence compromising the interpretability. There is no guarantee that a unique minimum exists and the high correlation also means that it is difficult to find the minimum. The parametric form also means that there is a pre-determined structure to the mortality schedule, for instance the previous models all decompose the age pattern into 3 stages: child, young adult (accident hump) and adult mortality. Should there be fundamental changes in the structure of mortality age pattern that these decomposition cannot capture, these models might fail. Non-parametric models offer higher flexibility, with usually more local basis functions. However the high flexibility may introduce robustness problems and over-fit at ages with sparse and unreliable data. This has been dealt with in most of the previous ELTs by rather subjective judgments, such as down-weighting ages with sparse data or discarding data at the oldest ages. In the latest ELT17, a parametric model is used at the

oldest ages to maintain robustness while a non-parametric model is used at younger ages given its flexibility.

3.2 Mortality Forecasting

In section 3.1 some laws of mortality have been discussed, they are useful in smoothing the static age pattern of mortality and therefore producing period life tables. However, more information could be learnt from considering the mortality schedules over time as a whole. It is also necessary to understand the evolution of mortality schedules for the purpose of mortality forecasting. A straightforward and intuitive approach is to repeatedly fit the laws of mortality to each year, then forecast each of the model parameters. For example, McNown and Rogers (1992) fitted the Heligman-Pollard model to the US data and projected each of the eight parameters using time series independently. The use of stochastic processes has the advantage of producing probabilistic intervals for forecasts instead of deterministic ones. However, the neglect of the correlations between parameters when McNown and Rogers (1992) projected the parameters independently might introduce biases and under-estimate the variances of the forecasts.

3.2.1 The Birth of Age-period Factor Models

Instead of repeatedly fitting the annual data and then projecting all the parameters from the mortality laws, Lee and Carter (1992) proposed the Lee-Carter (LC) Model which condenses all time-dependent developments into a single index. The model is non-parametric and employs a very simple yet effective structure,

$$\log(m_{xt}) = \alpha_x + \beta_x \kappa_t.$$

The α_x is the baseline mortality that describes the overall age pattern of mortality, while κ_t is the period effect and β_x measures the age-specific sensitivity to changes in the period effect. The model can be viewed as a principle-component decomposition with only the first component included, and it is shown that this is usually sufficient to explain most of the variations in the data. The LC Model has gained popularity quickly due to its simplicity and ease of parameter interpretation. It has also made forecasting readily available by simply forecasting the only time-dependent parameter κ_t. The LC Model pioneered the use of age and period effects and since then a lot of the models have been built around it. However, several drawbacks have been pointed out. As it can be seen, the model produces perfectly correlated mortality forecasts across ages over time. In addition, the LC model lacks smoothness. For example, the baseline mortality profile α_x can be jagged and hence non-smooth age patterns might be obtained. Delwarde et al. (2007) also noticed that the β_x exhibits irregular patterns. Cairns et al. (2008) showed that the lack of smoothness in the β_x might result in irregular patterns in the residuals. It is also criticised that the LC Model is not robust to the fitting period, meaning that using different fitting periods

might result in quite drastically different estimates. Smoothing techniques have been suggested in order to increase the robustness of the model and reduce the risk of over-fitting. Delwarde et al. (2007) fitted the LC Model with the β_x smoothed out using P-splines, while Richards and Currie (2009) smoothed out the β_x as well as the κ_t. Other several variants and modifications to the LC Model have also been proposed, such as adjusting the κ_t to match the life expectancy or calibrating the parameters against the mortality rates in the jump-off year for better forecasts (Lee and Miller, 2001), while Booth et al. (2002) investigated the optimal estimation period as well as the inclusion of higher order terms to the model in the form,

$$\log(m_{xt}) = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \dots.$$

In the original LC Model, Lee and Carter (1992) estimated the parameters by minimising the squared differences of log mortality rates. Statistically this is equivalent to assuming a normal distribution on the log of mortality rates, therefore errors are assumed to be homoskedastic, which is quite unrealistic as the observed mortality rates are often more variable at older ages due to the much smaller number of deaths at older ages. Brouhns et al. (2002) assumed a Poisson distribution for the number of deaths and proposed a Poisson parallel approach to the LC Model, hence introducing a more natural framework accounting for the heteroskedasicity. Specifically, mortality rates are modelled through a Poisson distribution on the number of deaths d_{xt} with mean $E_{xt}^C m_{xt}$, where E_{xt}^C and m_{xt} are the central exposure and mortality rate at age x in year t, respectively. Alternatively, a Binomial distribution on number of deaths d_{xt} with counts E_{xt}^0 and probability q_{xt} are also sometimes used, where E_{xt}^0 and q_{xt} are the initial exposure and death probability at age x in year t, respectively. Moreover, it is often observed that the Poisson variance (which is equal to the mean) is too restrictive, therefore a dispersion parameter can be introduced. A particular choice of parametrisation leads to a Negative Binomial distribution. The LC model and its variants have been applied widely, for example to the G7 countries (Tuljapurkar et al., 2000), Australia (Tickle and Booth, 2014), Netherlands and Belgium (Antonio et al., 2017) and Italy (D'Amato et al., 2011).

Using age and period effects, Cairns et al. (2006) focused only on higher ages (60 to 89) and proposed the CBD Model,

$$\operatorname{logit} q_{xt} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}).$$

It can be thought as collection of Gompertz models, i.e. repeatedly fitting a linear model to each annual data. The intercept and slope parameters $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are then modelled as a bivariate random walk time series. The model offers advantages over the LC Model such as having non-trivial correlation of future mortality rates and a smooth pattern of the age effect (the linear term). The model can also be fitted within a linear model framework as there is not any bi-linear terms as in the LC models. However, the applicable range is limited to older ages due to the linearity assumption. Several variants of the CBD model have also been proposed (Cairns et al., 2006, 2009; Cairns, 2013), such as adding a quadratic age term for a better fit to the very old ages and including a cohort effect, which is discussed below. Instead of a parametric linear (and quadratic) age term, Currie (2011) proposed to use a non-parametric smooth function, a smooth

CBD model, i.e.

$$\text{logit}\, q_{xt} = \kappa_t^{(1)} + \kappa_t^{(2)} s(x - \bar{x}),$$

where $s(\cdot)$ is a smooth function that is estimated using P-splines by the authors. This method avoids the rigid parametric structure while allowing extrapolation in the age direction by means of penalty.

3.2.2 Inclusion of Cohort Effects

In addition to the age and period effects, Willets (2004) noticed patterns in the data that relates to the year of birth which cannot be easily explained by age-period terms. Researchers have found that the inclusion of cohort effects is necessary in some populations, such as the United Kingdom. Cairns et al. (2009) demonstrated that adding a cohort effect provides a statistically significantly better fit to these populations and the standardised residuals display more randomness. Mortality models generally incorporate any of these three factors, namely age, period and cohort effects. These models are sometimes categorised into one, two or three-factor model. For instance, The LC Model and the CBD Model is a two-factor model with age and period effect.

However, Age-Period-Cohort (APC) models suffer a common drawback of non-identifiability, such that the age, period and cohort effects are inestimable. The root cause for the non-identifibaility lies in the fact that the age, period and cohort effects are exchangeable, one cannot move along any of the two dimensions without moving in the third one. This is best understood if one imagines a Lexis diagram, with age being the horizontal direction and period the vertical direction, then the cohort travels in the diagonal direction. It is then apparent that these three effects are always confounded, rendering any three-factor (age, period and cohort) models not identifiable. The non-identifiability means that there exist infinitely many possible solutions to the problem, and interpretation to the estimated age/period/cohort effects could be corrupted and compromised.

In order to circumvent this issue, constraints are usually placed on the coefficients to regain identifiability. Several estimation techniques have been proposed, such as the Constrained Generalized Linear Model (CGLM) (Mason et al., 1973) which imposes constraints on the age, period and cohort effects or more recently the Intrinsic Estimator (IE) (Yang et al., 2004) that utilises principal component approach to separate the null-space of the age-period-cohort space and assigning it with zero mass when estimating the effects, which is essentially equivalent to applying certain implicit constraints on the coefficients (Luo, 2013). Random effect approach (Luo and Hodges, 2020; Yang and Land, 2006, 2008) and Bayesian framework (Schmid and Held, 2007) have also been used to estimate APC models, where they both handle the non-identifiability problem by assigning distributions (priors) over the age/period/cohort effects, which can be viewed as adding regularisations to the original problem. In Fienberg's comment (2013) on Luo's paper (2013), it is argued that there is no technical way to solve the APC problem, since the issue lies in the heart of the APC model formulation (i.e. the linear

dependency between the three effects) instead of the methods of estimation. Fienberg (2013) as well as Smith (2008, 2004) and Mason and Smith (1985) argued that substantive judgment and knowledge are required when resolving the APC issue. Therefore, one has to be cautious whether all three effects together would give significantly better fit than either one or two of the effects alone, which do not suffer from any non-identifiability problems. It should also be noted that the age, period and cohort effects are only exactly linear dependent, meaning that non-linear trends in these effects are still estimable.

A classic three-factor model is the simple APC Model, which is simply

$$\log(\mu_{xt}) = \beta_x + \kappa_t + \gamma_{t-x}.$$

Renshaw and Haberman (2006) also extended the LC Model by adding a age-cohort interaction term to account for the cohort effects:

$$\log(\mu_{xt}) = \alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(2)}\gamma_{t-x},$$

where α_x, $\beta_x^{(1)}$ and κ_t are again the baseline mortality schedule, age-specific sensitivity to period changes and period effect, respectively. $\beta_x^{(2)}$ is the age-specific response to the cohort effect, γ_{t-x}, of the cohort born in year $t-x$. However, Cairns et al. (2009) and CMI (2007) found that this model lacks robustness. Cairns et al. (2011a) also noted that the age-cohort term appears to be compensating for the lack of a second age-period term, hence suggested replacing the age-cohort term with a second age-period term and simply a cohort effect (instead of an age-cohort interaction) for better robustness . If one think of the mortality surface with age as the horizontal direction and time the vertical, then the year-of-birth (cohort effect) would affect the surface diagonally. It is obvious how the age-period-cohort effects may compromise each other hence introducing robustness problems and optimisation difficulties. Some authors have experienced difficulties in estimating and projecting the cohort effects in a robust way (Plat, 2009a; Renshaw and Haberman, 2006). Antonio et al. (2017) have even opted out not to include cohort effects into their mortality model for Netherlands and Belgium as they are deemed not as significant and that they would complicate their other objectives.

3.2.3 Non-parametric Mortality Models

Currie et al. (2004) took a fully non-parametric approach and modelled the mortality surface using a two-dimensional (2-D) P-spline. Being totally structure-free means that the model is able to capture age-time interactions more flexibly. The model also produces smooth mortality trends in both the age and time direction. Forecasting is done by extrapolating the mortality surface in the time direction. However the forecasts are highly dependent on the penalty used for the P-spline, for example, a second order difference penalty would produce linear-trend like mortality forecasts while a first order difference penalty would produce flat mortality forecasts. Another possible drawback is that the forecasts produced this way are deterministic, instead of

stochastic. CMI (2006) has shown that the model produces good in-sample fit, but Cairns et al. (2009) revealed that there is a genuine random period effect that is over-smoothed in the model. Richards et al. (2006) later proposed an alternative 2-D P-spline model that models mortality in the age-cohort surface instead of the age-period surface. Camarda (2019) has also employed the 2-D spline approach to mortality modelling, with constraints imposed on the coefficients such that the forecast age profiles and improvement rates do not deviate from past historic trends significantly.

Hilton et al. (2019) extended the hybrid model by Dodd et al. (2018) in section 3.1 for mortality forecast, which takes the form:

$$m_{xt} = \begin{cases} \exp(s_\alpha(x) + s_\beta(x)t + \kappa_t + s_\gamma(t-x)) & \text{for } x < w \\ \frac{\exp(\beta_0+\beta_1 x+\beta_2 t+\beta_3 xt)}{1+\exp(\beta_0-\log(\psi)+\beta_1 x+\beta_2 t+\beta_3 xt)} \exp(\kappa_t + s_\gamma(t-x)) & \text{for } x \geq w, \end{cases}$$

where $s_\alpha(x)$, $s_\beta(x)$ and $s_\gamma(x)$ are smooth functions estimated by P-splines representing the baseline, the age-specific improvement rates and the cohort effect. Therefore, similarly to the Dodd et al. (2018) semi-parametric mortality graduation model, at younger ages splines are used in explaining variations while after a certain cut-off point w, mortality pattern takes the shape of a logistic form tending towards the asymptote ψ. Contrary to the usual assumption of the LC Model that κ_t follows a random walk with drift, Hilton et al. (2019) decomposed it into a deterministic linear trend t plus the random walk κ_t. This avoids estimating the bi-linear age-period term, making the fitting process easier.

3.2.4 Structural Breaks

Many of the stochastic mortality models assume linear trends in mortality rates (the main period effect) and project them forward to obtain forecasts. However, recent literature have started to examine the propriety of this assumption. O'Hare and Li (2014) investigated 30 countries and found that structural breaks (departure from the linearity assumption) are present in some countries. O'Hare and Li (2015) as well as Coelho and Nunes (2011) proposed measures and strategies for identifying structural breaks and modelled the period effects as broken linear trends in the event of structural breaks.

A number of mortality forecasting models have been discussed, along with some of their variants. In particular, the LC Model notably serves as a 'base' from which various other models are derived. It has a simple form that pioneered the use of age and period effects. Mortality projection is done by forecasting the period effects with the assumption that future period effects are expected to be a continuation of the present trend. More recent literature has advocated the inclusion of cohort effects, which is found to be significant in some countries. Some authors have encountered a lack of robustness in their estimation and have found it difficult to project and forecast cohort effects in a reliable way. Forecasting mortality rates mainly rely

on projecting current trends forward, yet the assumption that the same trend holds throughout the whole fitting period seems less likely the longer the fitting period. One way to circumvent this is to select an optimal fitting period such that the linear trend holds, as mentioned above. Scholars have also tried to find ways to identify significant changes in mortality trends, sometimes called structural breaks, such that the fitting period can be split up into sub-periods with different trends. When forecasts of different populations are compared, more problems arise. For example, divergences in mortality forecasts among populations are usually unappealing as a global convergence has been observed in the past (Wilson, 2011). This is particularly true when the populations considered are a main population and its sub-population (e.g. an insured population and the national population), as characteristics are shared between them to some extent. Moreover, female mortality rates are expected to be always less than or equal to the male mortality rates at each age, based on empirical data. When mortality of populations are modelled independently, they may fail to satisfy the aforementioned notions. This give rise to numerous joint population models, sometimes called 'coherent models', stemming from many of the single population models introduced in this section. Not only does jointly modelling different populations avoid said problems, but it also enjoys the advantage of pooling data sets, hence allows sharing of information among populations.

3.3 Coherent Models

Benefits can often be gained when different populations are modelled jointly, these are sometimes called coherent models. Coherent modelling of mortality is of more recent endeavours. When populations are modelled independently, implausible results are sometimes obtained, such as diverging mortality trends. For example, Jarner and Kryger (2011) observed that small populations such as Denmark are generally more irregular and that they are very sensitive to the fitting period, hence naive extrapolation of historic trends is likely to produce unrealistic forecast, such as a lower old age mortality than that of the younger ages. Booth et al. (2006) has also pointed out the lack of fit of the LC Model for small populations. Modelling mortality of different populations jointly has the advantage of borrowing strength, this is particularly useful in age range with scarce or even missing data. Coherent modelling is also very logical and sensible especially when the mortality of a sub-population is of interest, as the sub-population and the main population are expected to share some similar characteristics.

3.3.1 Common and Specific Factor Models

In order to coherently model and forecast mortality rates of multiple populations, Li and Lee (2005) extended the single population LC Model and proposed the Common Factor LC Model, which assumes a common $\beta_x\kappa_t$ term among different populations, i.e. $\log(m^j_{xt}) = \alpha^j_x + \beta_x\kappa_t$ where the superscript j indicates the population. This means that, starting from their respective mortality baselines, mortality rates of different populations have the same evolution over

time hence effectively avoids divergent mortality trends. In fact, mortality forecasts of different populations will be parallel to each other at each age. Clearly this common factor could be too restrictive and might lead to lack of fit to historical data, therefore Li and Lee (2005) also proposed a variant, the Augmented Common Factor Model, which includes population-specific age and period effects:

$$\log(m_{xt}^j) = \alpha_x^j + \beta_x \kappa_t + \beta_x^j \kappa_t^j.$$

The population-specific period effects κ_t^j's are assumed to be mean-reverting in the long run to maintain the non-divergence, while allowing for short term variations. Other variants of this type include the Common Age-effect Model (Kleinow, 2015) which assumes a common β_x but different κ_t^j; and the joint-κ Model (Lee and Carter, 1992; Delwarde et al., 2006) which on the other hand assumes a common κ_t but different β_x^j for different populations. Zhou et al. (2013) adopted the Common Age-effect Model and further assumed that the difference between the period effects of the two populations follow a mean-reverting AR(1) process, hence the projected mortality rates are non-divergent in the long run. Instead of having additive population-specific age-period terms as in the Common Factor Model and its variants, Russolillo et al. (2011) constructed a multi-population model with a multiplicative population-specific factor to the age-period effect, the three-way Lee-Carter Model,

$$\log(m_{xt}^j) = \alpha_x^j + \beta_x \kappa_t \varphi_j.$$

3.3.2 Associated Mortality Indices Model

Alternative to jointly modelling multiple populations using common effects, Cairns et al. (2011b) focused on forecasting period effects (and cohort effects) of multiple populations coherently. Using a simple age-period-cohort model as example, Cairns et al. (2011b) modelled the period and cohort effects jointly as correlated time series such that they produce non-divergent mortality trends. They focused on modelling mortality rates between a large population and a sub-population, and assumed that the difference of the period and cohort effects between the large population and sub-population are mean-reverting, so that the projected rates at each age are non-divergent over time. Specifically, let $j = 1$ be the (larger) main population and $j = 2$ be the (smaller) sub-population and let $c = t - x$ be the year of birth, then

$$\begin{aligned}
\log(m_{xt}^j) &= \alpha_x^j + n_a^{-1}\,\kappa_t^j + n_a^{-1}\,\gamma_c^j, && \text{for} \quad j = 1, 2 \\
\kappa_t^1 &= \mu_\kappa^1 + \kappa_{t-1}^1 + \epsilon_t^1, \\
\kappa_t^1 - \kappa_t^2 &= (1 - \psi_\kappa^2)\mu_\kappa^2 + \psi_\kappa^2(\kappa_{t-1}^1 - \kappa_{t-1}^2) + \epsilon_t^2, \\
\tilde{\gamma}_c^1 &= \gamma_c^1 - \mu_\gamma^1 - \rho_\gamma^1(c - \bar{c}), \\
\tilde{\gamma}_{c+1}^1 &= \psi_{\gamma,1}^1 \tilde{\gamma}_c^1 + \psi_{\gamma,2}^1 \tilde{\gamma}_{c-1}^1 + \delta_c^1, \\
\gamma_{c+1}^1 - \gamma_{c+1}^2 &= (1 - \psi_{\gamma,1}^2 - \psi_{\gamma,2}^2)\mu_\gamma^2 + \psi_{\gamma,1}^2(\tilde{\gamma}_c^1 - \tilde{\gamma}_c^2) + \psi_{\gamma,2}^2(\tilde{\gamma}_{c-1}^1 - \tilde{\gamma}_{c-1}^2) + \delta_c^2.
\end{aligned}$$

Here n_a is the number of ages and $\bar{c}$ is the average year of birth, μ_γ^1 and ρ_γ^1 are the parameters for the linear trend of the cohort effect of the large population, $\mu_\kappa^1, \mu_\kappa^2, \mu_\gamma^2, \psi_\kappa^2, \psi_{\gamma,1}^1, \psi_{\gamma,2}^1, \psi_{\gamma,1}^2$, $\psi_{\gamma,2}^2$ are parameters of the time series processes, and $(\epsilon_t^1, \epsilon_t^2)$ and (δ_c^1, δ_c^2) are bivariate normal error terms. The period effect of the larger population κ_t^1 follows a random walk with drift and the spread of the period effect $\kappa_t^1 - \kappa_t^2$ is a mean-reverting AR(1) process. The cohort effect of the larger population γ_c^1 is an AR(2) process around a linear trend and the spread of the cohort effect $\gamma_c^1 - \gamma_c^2$ is a mean-reverting AR(2) process. Therefore, in the long run the projected mortality rates of the sub-population would not diverge from that of the main population. In addition to modelling a main and sub population, they have also considered the case for two equal populations in the paper.

A similar approach is proposed by Dowd et al. (2011), which in addition to producing non-divergent trends, the period effect (and cohort effect) of the smaller population, labelled as 2 is actually being 'pulled' towards the period effect (and cohort effect) of the larger population, labelled as 1,

$$\begin{aligned}
\kappa_t^1 &= \mu_\kappa^1 + \kappa_{t-1}^1 + \epsilon_t^1, \\
\kappa_t^2 &= \mu_\kappa^2 + \kappa_{t-1}^2 + \phi_\kappa^2(\kappa_{t-1}^1 - \kappa_{t-1}^2) + \epsilon_t^2, \\
\tilde{\gamma}_c^1 - \tilde{\gamma}_{c-1}^1 &= (1-\psi_\gamma^1)\mu_\gamma^1 - \psi_\gamma^1(\tilde{\gamma}_{c-1}^1 - \tilde{\gamma}_{c-2}^1) + \delta_c^1, \\
\tilde{\gamma}_c^2 - \tilde{\gamma}_{c-1}^2 &= (1-\psi_\gamma^2)\mu_\gamma^2 - \psi_\gamma^2(\tilde{\gamma}_{c-1}^2 - \tilde{\gamma}_{c-2}^2) + \phi_\gamma^2(\tilde{\gamma}_{c-1}^1 - \tilde{\gamma}_{c-1}^2) + \delta_c^2.
\end{aligned}$$

Similarly, $\mu_\kappa^1, \mu_\kappa^2, \mu_\gamma^1, \mu_\gamma^2, \psi_\gamma^1, \psi_\gamma^2, \phi_\kappa^2, \psi_\gamma^2$ are parameters of the time series processes and $(\epsilon_t^1, \epsilon_t^2)$ and (δ_c^1, δ_c^2) are bivariate normal error terms. The ϕ_κ^2 and ϕ_γ^2 are parameters that dictates how heavy the pull is between the period and cohort effects of the two populations. The larger the gap is between the period (cohort) effect of the two populations, the stronger the pull is, hence in the long run the projected mortality rates of the two populations will converge.

3.3.3 Relational Models

Instead of introducing common or correlated effects, Jarner and Kryger (2011) proposed the SAINT Model, where they first model the mortality surface of a larger population as the reference schedule and then model the ratio of the sub-population mortality rates to the reference,

$$m_{xt}^{ref} = H_\theta(x,t), \tag{3.1}$$

$$m_{xt}^{sub} = H_\theta(x,t)\exp(\sum_{i=1}^{n}\beta_x^{(i)}\kappa_t^{(i)}). \tag{3.2}$$

In their paper they chose the β_x^i's to be some fixed functions of age as regressors and regard κ_t^i as parameters to be estimated. Specifically, they modelled the ratio by three orthogonal regressors,

$$\begin{aligned}\beta_x^{(1)} &= 1,\\ \beta_x^{(2)} &= (x-60)/40,\\ \beta_x^{(3)} &= (x^2-120x+9160/3)/1000,\end{aligned}$$

and modelled the $\{\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}\}$ as a VAR(1) process. Plat (2009b) also considered a relational model in the form of 3.2 where he focused on modelling an insured sub-population relative to the country population. The author suggested various possible ways of specifying the regressors $\beta_x^{(i)}$, and further imposed a constraint on the spread such that at the closing age (say 120) the ratio of the insured population mortality rates to that of the country population is exactly 1. This is because differences between the portfolio mortality and main population mortality is expected to wear off at higher ages. In other words, the modelled mortality rates of the main and sub-population converge in the age direction. Villegas and Haberman (2014) also proposed a relational model for the mortality of different socio-economic sub-populations. Similarly, a larger population is treated as the reference schedule and the ratio of each socio-economic group to the reference is then modelled. The Villegas and Haberman (2014) model is similar to the Augmented Common Factor LC Model but the additional age effect is assumed to be the same among different groups, i.e

$$\begin{aligned}\log(m_{xt}^{ref}) &= \alpha_x^{ref} + \beta_x^{ref}\kappa_t^{ref} + \gamma_c^{ref},\\ \log(m_{xt}^{j}) &= \log(m_{xt}^{ref}) + \alpha_x^{j} + \beta_x\kappa_t^{j},\end{aligned}$$

where j indicates the population. The period effects κ_t^{ref} and κ_t^j are assumed to be random walk with drifts.

The Product-ratio Model by Hyndman et al. (2013) can also be viewed as a relational model, instead of treating a single population as the reference schedule, the geometric mean of all the populations considered is taken as the reference, and the ratios of each population to the mean are then modelled. One of the strength of the Product-ratio Model is that if the populations have approximately equal variances, then the product (geometric mean) and each of the ratios will have approximately zero correlation, making forecasting more efficient. The product and ratios are each modelled using functional analysis methods and then forecast using time series processes. In addition, the ratio model forecasts are assumed to follow stationary time series processes so that the projected mortality rates are non-divergent at each age.

3.3.4 Non-parametric Models for Multi-population

Biatat and Currie (2010) extended the 2-D spline model by Currie et al. (2004) to a joint model

by adding another 2-D spline for the spread between a population and a reference population. They introduced the concept of 'similarity' among populations and specified conditions in which a more parsimonious surface should be used in describing the gap between populations. More specifically, when the two populations are 'similar' in some sense, then a uniform (across age and/or period) spline surface is used for the spread, resulting in a more parsimonious structure. This can also be used to classify whether the populations are in anyway 'similar' to each other.

Shang et al. (2016) proposed a functional approach to jointly model mortality of different populations. The model has a similar structure to the augmented common factor model (Li and Lee, 2005), with more principal components included. Shang and Hyndman (2017) also adopted the functional approaches but focused on mortality rates of disaggregated populations. A main goal is to produce forecasts that are consistent at all aggregation levels, also called 'forecasts reconciliation'. They analysed the mortality time series of different aggregation levels using two grouped time series forecasting methods, namely the 'bottom-up' approach and the 'optimal combination method (Hyndman et al., 2011)' and found that the 'bottom-up' approach gives more accurate forecast in their example. The bottom-up approach first forecasts mortality rates at each age at the lowest aggregation level independently, then forecasts higher level mortality from these base level forecasts, such that the aggregated estimates are ensured to be consistent with the lower level estimates. The optimal combination method on the other hand forecast mortality at all aggregation levels independently, then revise the estimates by fitting a regression model.

The aforementioned hybrid model by Hilton et al. (2019) also modelled males and females jointly by assuming a common asymptote of the logistic function for both sexes, hence the mortality rates of males and females tend to the same level at the highest ages.

De Jong et al. (2016) provided an entirely different and interesting angle to coherent mortality modelling: a complex number framework is introduced to jointly model male and female mortality. More specifically, the male and female mortality rates at each age in each year are first condensed complex numbers, with one representing the real part and the other the imaginary part. The mean of them are then subtracted and complex number SVD is applied to the complex number data matrix as in the LC model. The age and period effects can then be extracted from the components of the SVD. The procedure is very similar to the usual LC model, the main difference being that in the complex LC model, optimisation is done in the age and time direction as well as across sex.

Coherent modelling of mortality has gained increasing attention in the literature. Coherent models are able to produce more reasonable forecasts, as opposed to fitting single population models to different populations separately. For example, divergent mortality trends are often resulted when populations are forecast independently, which contradicts to the global demographic convergence observed in the past. Jointly modelling mortality of multiple populations also has the advantage of pooling data, and borrowing strength across populations. Researchers have suggested various way to achieve coherent modelling, such as using common factors among

populations or modelling the mortality rates using correlated time series. Relational models have also been proposed and are especially effective when the populations of interest are a large population an its sub-population. The spread between the sub-population and a large population can then be modelled and quantified. Splines models have proven to give satisfactory fit to the data due to the high flexibility. When multiple populations are considered, a parsimonious structure should be employed whenever it is sufficient as it is crucial to avoid over-fitting. A complex number framework has also been adopted for male and female mortality modelling which performs optimisation across age, time and sex at the same time.

3.4 Main Contributions of the Thesis

Computational advancements have made more sophisticated models feasible. Researchers have constructed advanced models for mortality graduation attempting to fit the whole age range with interpretable parameters. However these models often involve complicated functions such that estimation of the parameters are difficult. In addition, the estimated parameters are often highly correlated which compromises their interpretability. Splines have also been used in mortality graduation which offers higher flexibility and computation convenience, yet it usually suffers lack of robustness at the highest ages where data is sparse. Several approaches have been adopted to circumvent the said problem such as discarding or down-weighting data at the oldest ages. Dodd et al. (2018) on the other hand proposed the semi-parametric model, which exploits the high flexibility of splines for younger ages while maintaining a parsimonious and robust parametric structure for the oldest ages.

Building upon the Dodd et al. (2018) model, instead of switching from a spline to a parametric function, we suggest using adaptive P-spline for mortality graduation in Chapter 4. In other words, a spline is used for the whole age range with varying smoothness in age. This avoids the need to estimate or specify a cut-off point at which the spline transitions into a parametric function. At higher ages the adaptive spline should be smoother, and hence more parsimonious and robust. This method is suitable for the whole age range that can be estimated efficiently under the generalised linear models framework, as opposed to the more difficult non-linear optimisation of complicated parametric functions which often involve highly correlated estimated parameters. Moreover, most of the English Life Tables are produced by smoothing male and female mortality rates separately, this might create unreasonable trends such as divergence at the highest ages or male mortality rates being higher than that of females. Previously this is eliminated with ex post ad-hoc fixes. We propose a way to borrow information from each sex to further increase the robustness of the estimation at the oldest ages, and avoid divergence and intersection of male and female mortality rates, thus producing more plausible smooth schedules. We utilise properties of the B-spline basis and impose hard constraints on the parameters such that cross-over of male and female mortality rates is strictly prohibited. The proposed method performs better than traditional splines.

We then extend the model for mortality projection in Chapters 5 and 6. The model has a similar structure as the Hilton et al. (2019) model, however we use adaptive splines instead of a semi-parametric model. Therefore the model does not require us to choose an 'optimal age' to switch between models and therefore we obtain smooth estimates for the whole age range.

In this thesis we jointly model English-and-Welsh and Scottish males and females moratlity rates. Following Li et al. (2014), models discussed in Section 3.3 can be broadly categorised into three broad classes: Common and Specific Factor Model, Associated Mortality Indices Model and Ratio Model. Common and Specific Factor Models assume common trends for multiple populations and model the population-specific deviations from the common trend. Associated Mortality Indices Models focus on correlation of the projected mortality rates of different populations and aim at producing non-diverging trends in the long run. Ratio Models usually start with modelling the mortality rates of a large reference population and then model the ratio or spread of a smaller or sub-population to the reference population. We model mortality rates of different populations by means of penalty. This can be viewed as a halfway between jointly modelling them through a common factor models and modelling them independently. The proposed model is not as restrictive as the common factor models in the sense that multiple populations are allowed to have their own respective age/period/cohort effects, but is stronger than the independent models in the sense that the shapes of the effects of different populations depend on each other.

There has been little focus on mortality forecasts at the oldest ages (extrapolation in the age direction) where data is non-existent, however sensible and plausible forecasts of mortality rates at these ages are crucial. Extrapolation to higher ages where data is sparse or non-existent often results in implausible trends. One main challenge is that since data is non-existent at these ages, extrapolation relies heavily on the model structure. Assessments on these projections are mainly based on their plausibility and reasonableness. For example, Cairns et al. (2008) introduced and explained the concept of 'biological reasonableness'. Since period mortality tables have historically exhibited increasing rates of mortality with age at higher ages, decreasing mortality age patterns after the young adulthood may be deemed implausible. In addition, as discussed in Chapter 1, evidence is shown supporting the male excess mortality from both a biological and behavioral points of view, hence female mortality rates should be expected to be lower than that of males. Divergences in estimated mortality trends are also often deemed to be unappealing, as historical data has shown otherwise. As in mortality graduation, information can be borrowed from each sex especially at the oldest ages. We expect the mortality improvement rates of males and females are similar at the oldest ages. This is supported by empirical data that similar mortality levels tend to have similar improvement rates. This also avoids divergences in male and female mortality rates in the mid to long run, particularly at the oldest ages where the forecasts are most susceptible to cross-over. We also model mortality rates across countries, specifically England-and-Wales' and Scotland's populations are considered. These two populations are expected to have similar mortality structures for the same sex. England-and-Wales has a wider age

range of available data, therefore by borrowing information from this population, the extrapolation for Scottish population can be made more reliably. This is especially important at the highest ages where there are significant social and financial implications such as pensions and health care.

We demonstrate a way to incorporate expert opinion into the projections. Following Dodd et al. (2020), weights are assigned between the current mortality improvement rates and the expert-determined target rates such that future mortality improvement rates will approach the expert-determined target rates. The incorporation of expert opinion avoids the unlikely assumption of perpetual linear improvement trends at current rates and improves the plausibility of projection mortality rates. It also ensures the projections are non-divergent as the target rates are the same for both sexes and countries.

Chapter 4

Mortality Graduation for Static Life Tables

In this chapter a method for robust mortality graduation using penalised adaptive splines is proposed. In Section 3.1 various mortality graduation models have been introduced. In order to model the mortality rates over the entire age range, these models incorporate functions of quite complicated forms, such as the Heligman-Pollard 8-parameter model (Heligman and Pollard, 1980). They often require non-linear optimisation and the estimated parameters are often highly correlated, adding difficulties to the estimation. On the other hand, splines are well known for their flexibility and have been used widely as a smoothing tool. The local support of B-spline basis also offers computational efficiency. Nonetheless, when splines are used to smooth mortality rates, a lack of robustness is observed at the oldest ages where data is scarce. Figures 4.1 and 4.2 show the smoothed mortality rates of England and Wales males and females in 2010, 2011 and 2012 using rank 40 uniform P-splines (i.e. 40 basis functions with equally spaced knots), extrapolated to age 120. The lack of robustness at the oldest ages is more apparent in males, as we do not expect the yearly mortality schedule to vary a lot over consecutive years. The splines are then re-fitted discarding the last 5 data points for each year, from age 1 to 100, again revealing the lack of robustness at the oldest ages. The exclusion of the last 5 data points changes the mortality profiles quite drastically. Extrapolations based on these mortality schedules are very sensitive to the unreliable data at the oldest ages. In addition, sometimes an unreasonable mortality schedule is produced. For example, Figure 4.3 plots the estimated mortality rates for England and Wales males in 1980. The fitted mortality rates are decreasing at the highest ages, which is not 'biologically reasonable' also pointed out by Cairns et al. (2008).

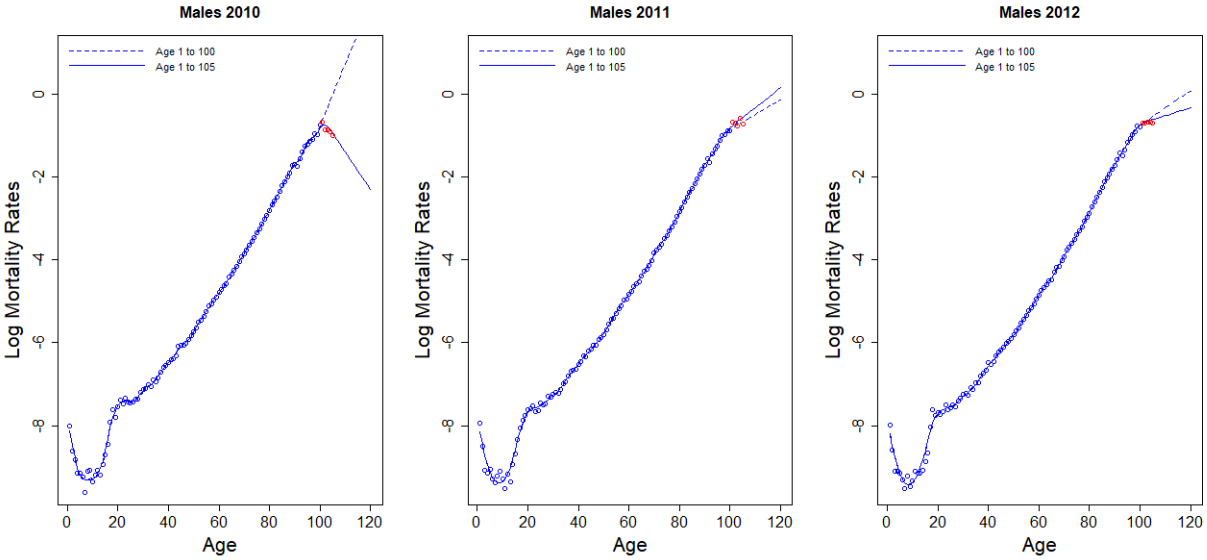


Figure 4.1: P-spline fit of the England and Wales males in 2010, 2011 and 2012, extrapolated to age 120. The solid lines are the estimated mortality rates using data from age 1 to 105, while the dotted lines are the estimated mortality rates using data from age 1 to 100.

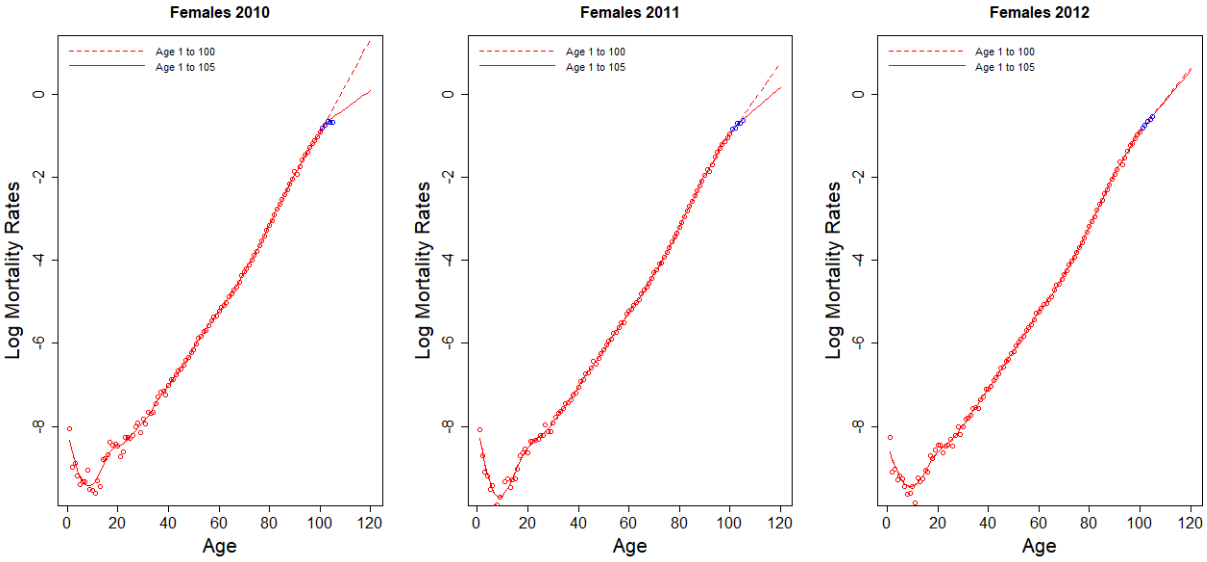


Figure 4.2: P-spline fit of the England and Wales females in 2010, 2011 and 2012, extrapolated to age 120. The solid lines are the estimated mortality rates using data from age 1 to 105, while the dotted lines are the estimated mortality rates using data from age 1 to 100.

Dodd et al. (2018) acknowledged this lack of robustness at the oldest ages and constructed the semi-parametric model (as discussed in Section 3.1). They proposed switching from a spline to a parametric model at some cut-off age, therefore a more robust and parsimonious model is used for the oldest ages. The cut-off point has to also be estimated or specified. They have fitted the model with different cut-off ages from a candidate set and performed model averaging. This requires multiple fittings of the model and smoothness at the transition from the spline to the parametric model is not guaranteed.

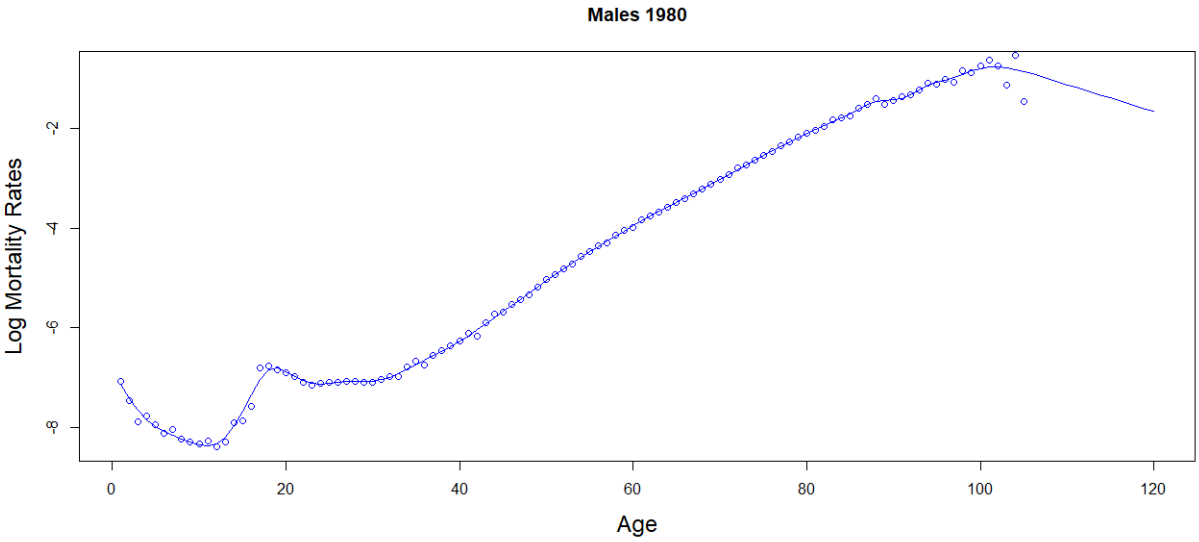


Figure 4.3: P-spline fit of the England and Wales males in year 1980, extrapolated to age 120.

The main reason of the lack of robustness at the oldest ages is that conventional penalised splines has only one smoothing parameter governing the overall smoothness, resulting in either over-smoothed mortality rates at the young ages (less likely due to much bigger exposures compared to older ages) or a lack of robustness at the oldest ages, as shown in Figures 4.1 - 4.3. This motivates the use of adaptive P-splines.

Under an adaptive splines, the degree of smoothness can vary over the domain. It has been shown that adaptive splines perform at least as good as ordinary splines when the true underlying function has uniform smoothness and perform better when the true underlying function has changing smoothness over the domain (Ruppert and Carroll, 2000). Using adaptive P-splines, instead of having a single parameter governing the global smoothness penalty, we assumed that the penalty function is an exponential function.

$$\log(\boldsymbol{m}(\boldsymbol{x})) = \boldsymbol{B}(\boldsymbol{x})\boldsymbol{\beta} \tag{4.1}$$

with penalty

$$P_s = \sum_{i=3}^{k} \zeta(i)(\bigtriangledown^2(\beta_i))^2, \tag{4.2}$$

where

$$\zeta(i) \;=\; \lambda_1 \,\exp(\,\lambda_2\, i), \qquad\qquad \lambda_1 > 0. \tag{4.3}$$

Here $\boldsymbol{B}(\boldsymbol{x})$ is the design matrix of B-spline basis functions, λ_1 and λ_2 are parameters controlling the smoothness and k is the number of basis functions. To ensure the penalty is positive λ_1 has to be positive. The penalty can be written more compactly as $\boldsymbol{\beta}'\boldsymbol{P}'\boldsymbol{\Lambda}\boldsymbol{P}\boldsymbol{\beta}$ where $\boldsymbol{P}$ is the second

order difference matrix with appropriate dimension and $\mathbf{\Lambda}$ is a diagonal matrix with entries $\mathbf{\Lambda}_{ii} = \lambda_1 e^{\lambda_2 i}$.

4.1 Basis Dimension

When using P-splines, the number of basis has to be chosen prior to model fitting. There are no fixed rules on how to select the basis dimension, as the essence of penalised splines is that the user construct a basis that is generous enough to capture variations of the underlying function and then penalise the roughness. One could have as many basis functions as the number of distinct data points and the smoothness penalty would then control the optimal smoothness and hence prevent over-fitting. Currie and Durban (2002) and Currie et al. (2004) suggested placing one knot for every four or five data points. Here the basis dimension is chosen according to preliminary analysis. The data is fitted with ordinary P-splines of dimension 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 for each sex and the total effective degrees of freedom for males and females is examined. Males and females have the same knots sequence and hence the same B-spline basis. Figure 4.4 plots the total effective degrees of freedom (males + females) against the total basis dimension. It can be seen that the effective degrees of freedom increases steadily from 20 basis functions for each sex (40 in total) to 40 basis functions for each sex (80 in total), indicating that a basis dimension of at least 40 is needed. After that, the total effective degrees of freedom starts to level-off at around 40 basis functions for each sex (80 total dimension), indicating that further increase in the dimension has limited benefits and might not be necessary. Therefore we choose the basis dimension to be 40 for each sex.

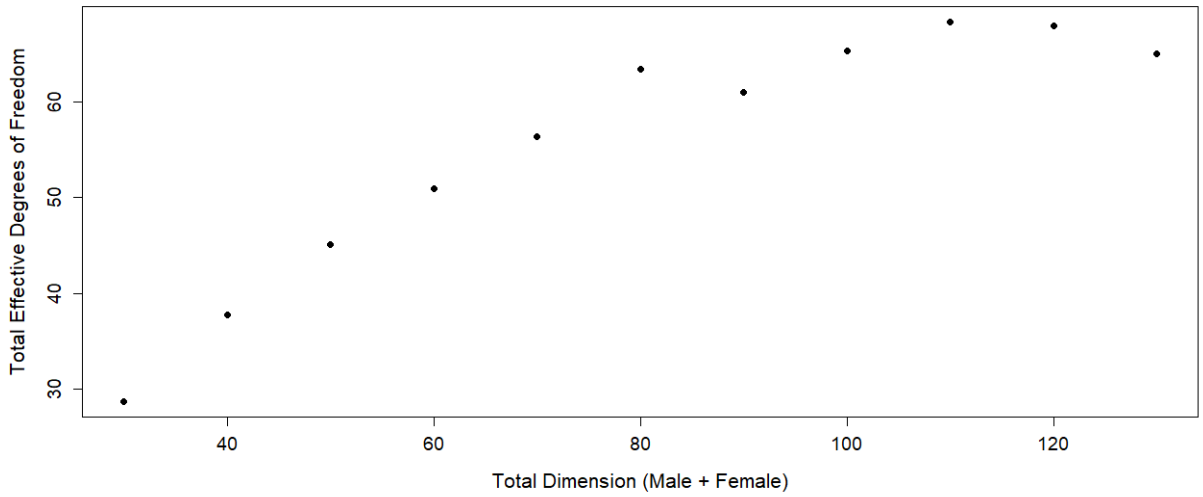


Figure 4.4: The total effective degrees of freedom against the total number of basis dimension for males and females.

4.2 Results

Figure 4.5 plots the graduated mortality rates using the proposed adaptive P-splines. It is evident that the proposed adaptive P-splines are more robust at the oldest ages compared to ordinary P-splines with global penalty. The yearly mortality schedules display less irregular variations. Figure 4.6 plots the smoothed male mortality rates in 1980. With the global penalty, the estimated mortality rates seem to be under-smoothed and are decreasing at the oldest ages, while with the adaptive penalty the mortality schedule is much more reasonable and smooth.

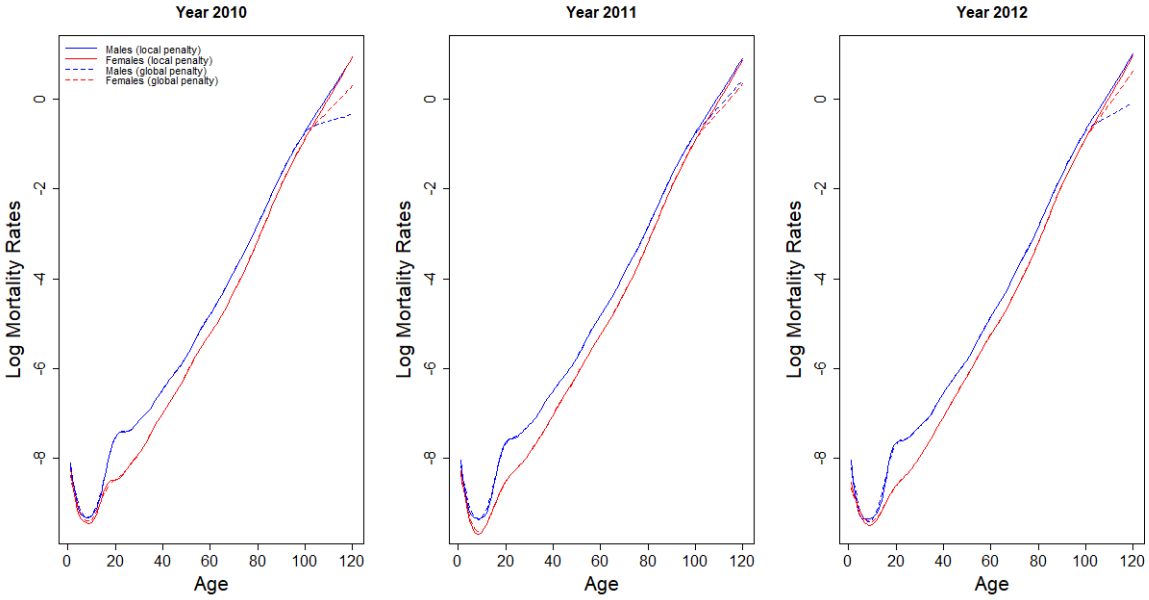


Figure 4.5: The smoothed mortality rates of England and Wales males and females in year 2010, 2011 and 2012. The solid lines are the estimated rates using adaptive P-splines with exponential penalty function, while the dotted lines are the estimated rates using P-splines with global penalty.

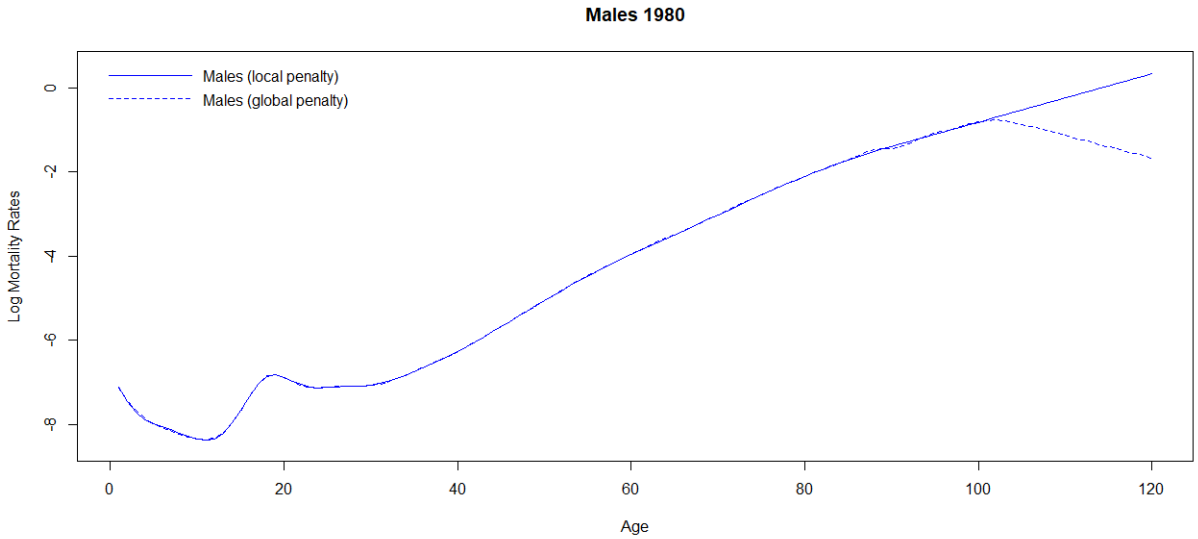


Figure 4.6: The smoothed mortality rates of England and Wales males in 1980. The solid line is the estimated rates using adaptive P-splines with exponential penalty function, while the dotted line is the estimated rates using P-splines with global penalty.

Figure 4.7 shows the adaptive P-spline fit for males in 2011 and the corresponding estimate of the exponential smoothness penalty $(\lambda_1 \exp(\lambda_2 i))$. The smoothness penalty is relatively low at younger ages and it rapidly increases at about age 90. This is desirable as the exposures and the reliability start to decrease at ages beyond 90.

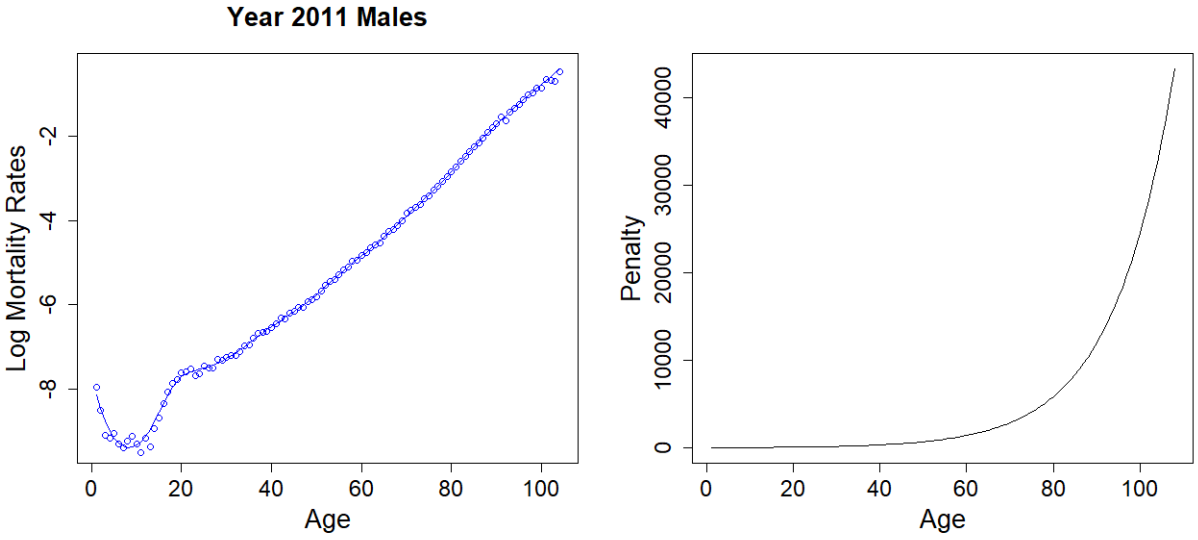


Figure 4.7: P-spline with exponential penalty fit for 2011 England and Wales males and the corresponding adaptive smoothness penalty

4.3 Joint Mortality Graduation of Males and Females

It is shown in section 4.2 how a locally adaptive penalty for splines can largely eliminate implausible trends and improve parsimony and robustness. Nonetheless, it is still possible to obtain unrealistic mortality schedules when male and female mortality rates are modelled independently. For example, the estimated mortality rates for males and females might sometimes diverge or intersect at the oldest ages. Figure 4.8 plots the estimated mortality rates using adaptive P-splines in 2007 and 2017. It is generally expected that female mortality rates are lower than that of males at all ages and that the differences between male and female mortality rates decrease as age increases. Although the possibility of divergent or intersecting mortality rates can not be ruled out, we would not expect this to happen at ages where we have very sparse or no data.

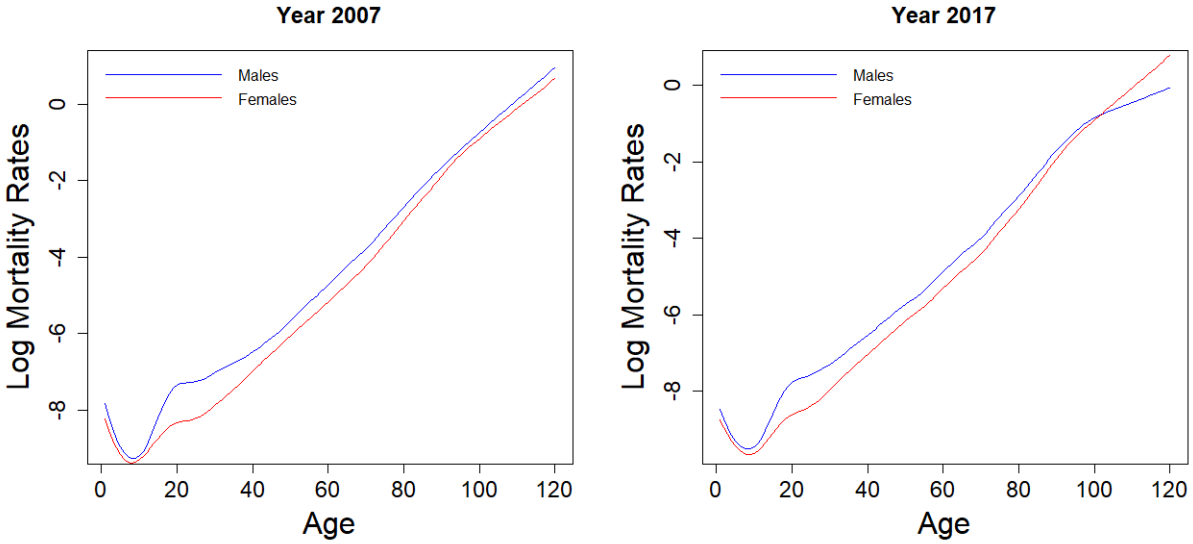


Figure 4.8: The estimated mortality rates of England and Wales males and females in 2007 and 2017.

4.3.1 Preventing Divergences of Male and Female Mortality Rates

Since there are more female data at the oldest ages, by jointly modelling male and female mortality rates we can borrow information and avoid divergence. To do this we introduce a cross-sex penalty in addition to the adaptive smoothness penalty for males and females. Assuming that the splines for males and females have the same knot sequence (i.e. the same B-spline basis), we penalise the squared differences of the coefficients of males and females with an exponential penalty, i.e.

$$P_d = \sum_{i=1}^{k} \zeta^D(i) \, (\beta^M i - \beta_i^F)^2, \tag{4.4}$$

where

$$\zeta^D(i) = \lambda_1^D \exp(\lambda_2^D\, i). \tag{4.5}$$

The superscripts M, F and D indicate that the corresponding quantities relate to males, females and their differences, respectively. At younger ages, the splines shall enjoy more freedom as data at this region is more reliable. It is also believed that there is genuine difference in the levels and patterns of mortality between male and female mortality (especially at the accident hump). At adult ages the differences between male and female mortality rates seem to diminish gradually as age increases. Therefore an exponential penalty function is used for the cross-sex difference penalty. Thus, the joint graduation model has three penalties, an exponential smoothness penalty for males, an exponential smoothness penalty for females and an exponential penalty for the differences between male and female mortality rates. We found that when the differences between the first few male and female coefficients are left un-penalised, the estimation of smoothing parameters is easier and more stable. One of the reasons for this might be the apparent converging trends in infant mortality, which adds difficulty to the estimation of the difference penalty. Hence we are only penalising the differences between males and females after the 8-th coefficient (around age 24, which is just after the accident hump). Results suggest that the selection of this coefficient is immaterial as long as the childhood mortality is excluded. Our proposed model then becomes,

$$\log\Big(\boldsymbol{m^*(x)}\Big) = \log\begin{pmatrix}\boldsymbol{m^M(x)}\\ \boldsymbol{m^F(x)}\end{pmatrix} = \begin{pmatrix}\boldsymbol{B(x)} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{B(x)}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta^M}\\ \boldsymbol{\beta^F}\end{pmatrix}, \tag{4.6}$$

with the following three penalties

$$P_{s,M} = \sum_{i=3}^{k} \zeta^M(i)(\bigtriangledown^2(\beta_i^M))^2, \tag{4.7}$$

$$P_{s,F} = \sum_{i=3}^{k} \zeta^F(i)(\bigtriangledown^2(\beta_i^F))^2, \tag{4.8}$$

$$\text{and} \quad P_d = \sum_{i=9}^{k} \zeta^D(i)(\beta_i^M - \beta_i^F)^2, \tag{4.9}$$

where

$$\zeta^j(\cdot) \;=\; \lambda_1^j \exp(\lambda_2^j \cdot) \qquad \text{and} \qquad \lambda_1^j > 0. \tag{4.10}$$

Here the first two penalties relate to the smoothness of male and female mortality rates while the third penalty corresponds to the difference between male and female mortality rates. The penalties can be written more compactly as $\boldsymbol{\beta' P' P \beta}$ where the overall coefficient vector is $\boldsymbol{\beta} = (\,\boldsymbol{\beta^M}\ \boldsymbol{\beta^F}\,)$ and the overall penalty matrix is $\boldsymbol{P} = \begin{pmatrix}\sqrt{\boldsymbol{\Lambda^M}}\boldsymbol{P^M}\\ \sqrt{\boldsymbol{\Lambda^F}}\boldsymbol{P^F}\\ \sqrt{\boldsymbol{\Lambda^D}}\boldsymbol{P^D}\end{pmatrix}$, $\boldsymbol{P^M} = (\,\boldsymbol{\nabla_2}\ \mathbf{0}\,)$, $\boldsymbol{P^F} = (\,\mathbf{0}\ \boldsymbol{\nabla_2}\,)$.

Here $\boldsymbol{\nabla_2}$ is the second order difference matrix, $\boldsymbol{P^D} = \begin{pmatrix} 1 & -1 \end{pmatrix} \otimes \boldsymbol{I}$ and $\mathbf{0}$ is simply a null matrix of appropriate dimension.

Figure 4.9 plots the estimated mortality rates of the joint model in 2007 and 2017. Comparing the graduated rates of the joint and separate models, we can see that the difference penalty is effective in preventing divergences between male and female mortality rates. Figure 4.10 plots the estimated mortality rates of the joint model in 2007 alongside with the estimated cross-sex penalty function, $\zeta^D(i)$. The cross-sex difference penalty $\zeta^D(i)$ is increasing in age, resulting in a diminishing gap between male and female estimated mortality rates. Even though the cross-sex difference penalty is able to prevent divergences, there is no guarantee that female mortality rates will always be lower than male mortality rates (for example, the graduated mortality rates at the oldest ages in 2017 in Figure 4.9).

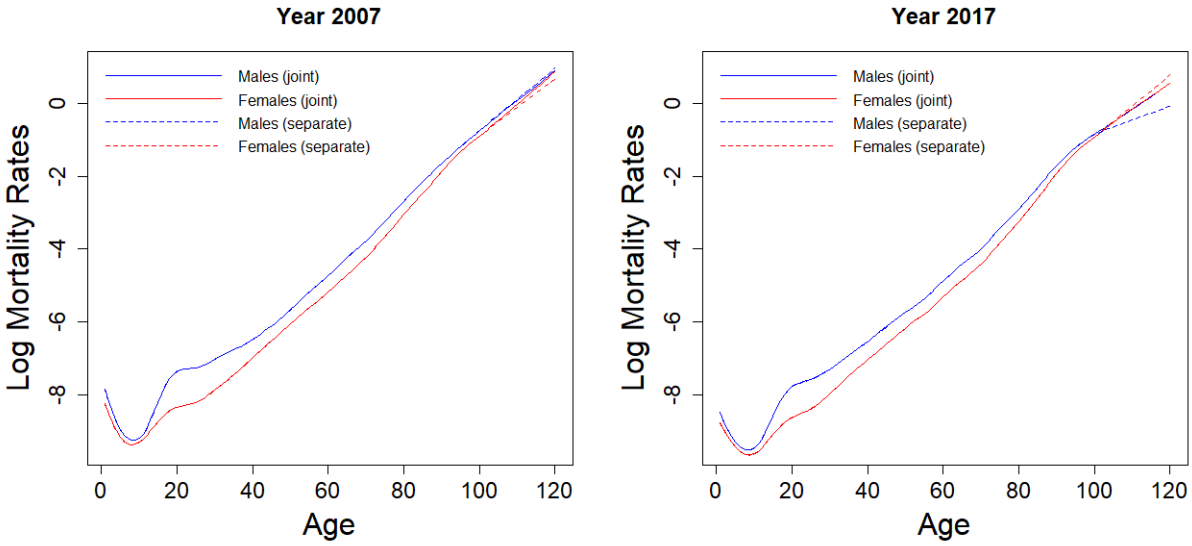


Figure 4.9: The estimated mortality rates of England and Wales males and females in 2007 and 2017. The solid lines are the estimated rates from the joint model while the dotted lines are the estimated rates from smoothing male and female mortality rates separately.

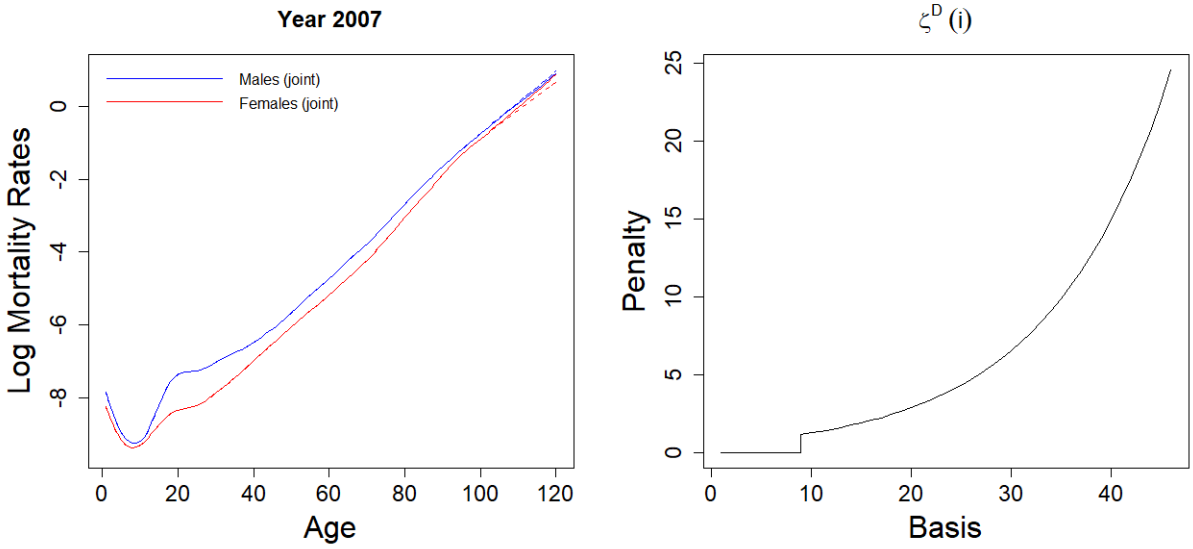


Figure 4.10: The estimated mortality rates of England and Wales males and females in 2007 and the estimated penalty on the difference between male and female spline coefficients.

4.3.2 Preventing Cross-overs of Male and Female Mortality Rates

By introducing the difference penalty in section 4.3.1, we solved the problem of divergent mortality trends at the oldest ages. However, this does not necessarily guarantee that the female mortality rates will always stay lower than the male mortality rates. One further improvement is to prevent male and female mortality rates from crossing each other. Figure 4.11 plots the graduated mortality rates from the joint model in 2010, 2011 and 2012. Sometimes the estimated male mortality rates are lower than that of females, for example, the child mortality in 2010 and 2012.

As mentioned before, we do not expect male and female mortality rates to cross over at ages where we have very little or no data. In order to avoid this, a hard constraint is imposed on the spline coefficients such that the each of the spline coefficients of females are smaller than or equal to the corresponding spline coefficients of males, i.e. $\beta_i^F \leq \beta_i^M \quad \forall i$. Since it is assumed that males and females have the same knots sequence, this is a sufficient condition such that $m^F(x) \leq m^M(x) \quad \forall x$. This constraint is implemented by using Non-Negative Least Squares (NNLS) instead of least squares within each P-IRLS iteration. The NNLS is done using the *lsei* package in R.

To proceed, the parameter vector $\boldsymbol{\beta^*} = \begin{pmatrix} \boldsymbol{\beta^M} \\ \boldsymbol{\beta^F} \end{pmatrix}$ is first transformed to $\boldsymbol{\beta^{**}} = \begin{pmatrix} \boldsymbol{\beta^M} - \boldsymbol{\beta^F} \\ \boldsymbol{\beta^F} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta^D} \\ \boldsymbol{\beta^F} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta^M} \\ \boldsymbol{\beta^F} \end{pmatrix} = \boldsymbol{D\beta^*}$. The constraint $\beta_i^F \leq \beta_i^M \quad \forall i$ is equivalent to $\beta_i^M - \beta_i^F \geq 0 \quad \forall i$, hence the non-negative constraint is only applied to $\boldsymbol{\beta^D}$, i.e. $\beta_i^D \geq 0 \quad \forall i$. The design matrix and penalty matrix have to be transformed accordingly. Since this is not a

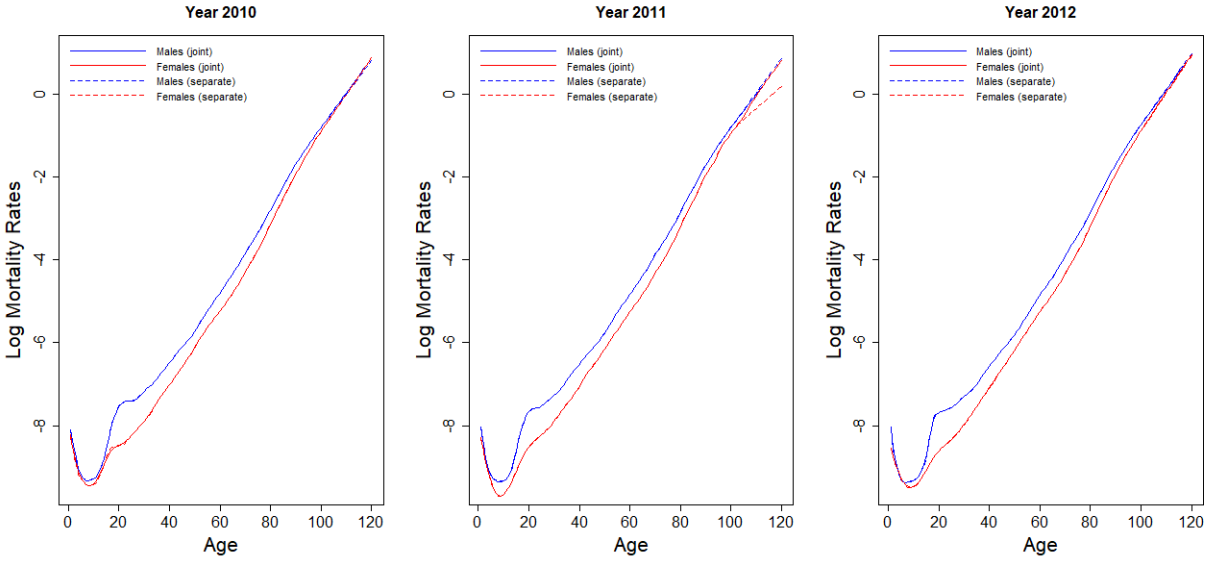


Figure 4.11: The estimated mortality rates of England and Wales males and females in 2010, 2011 and 2012 from the joint model.

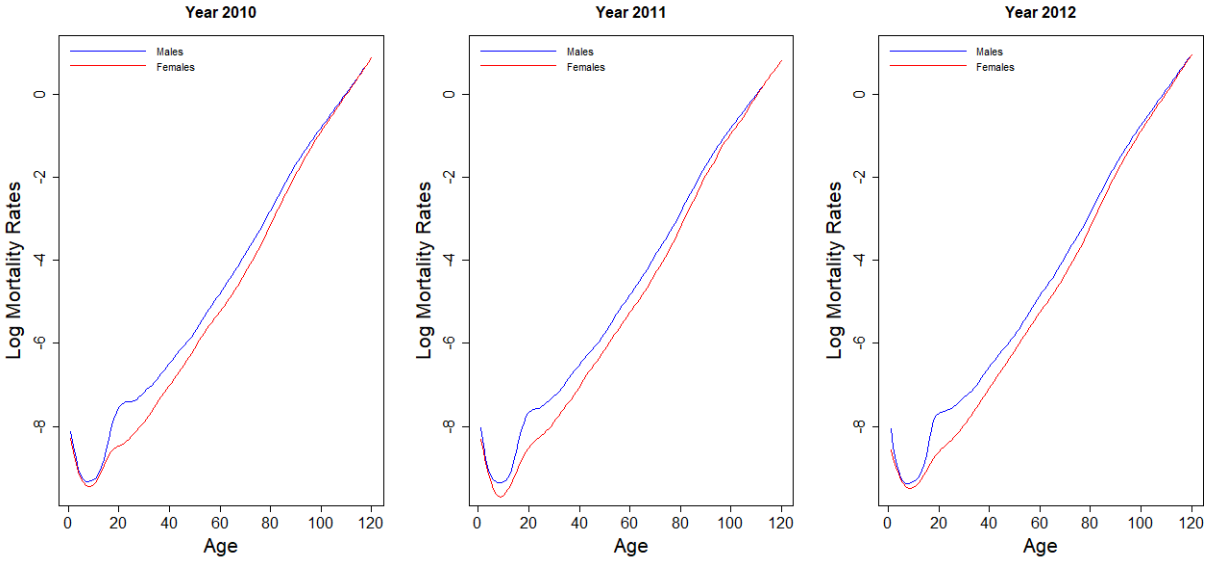


Figure 4.12: The estimated mortality rates of England and Wales males and females in 2010, 2011 and 2012 from the joint model with non-negative constraints.

linear smoother, i.e. it cannot be written as the general form $\hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{y}$, the gradient and Hessian for the minimisation of the BIC cannot be found analytically. In addition, to the best of our knowledge, the effective degrees of freedom of a non-negative least square model is not known. We believe that adding the non-negative constraints shall not change the overall optimal smoothness considerably, therefore the smoothing parameters in the constrained model are fixed at the estimated smoothing parameters from the unconstrained joint model. Figure 4.12 plots the graduated mortality rates under the non-negative constraints. As expected, the male mortality curve now always lies above female mortality curve.

4.4 Conclusion

Crude mortality rates often exhibit irregular and wiggly patterns, due to natural randomness. Therefore the crude rates have to be smoothed before they are used, this process is sometimes called mortality graduation. The core objective of graduating mortality rates is to produce a smooth mortality schedule, as well as eliminating any trends that are deemed unreasonable, such as decreasing mortality rates at high ages, or male mortality rates being lower than female mortality rates. The latter issue is addressed using rather ad-hoc methods in previous ELTs. For example, In ELT15 the data at the oldest ages are discarded such that a monotonic mortality scheulde would be obtained at these ages. In this chapter we proposed a robust method for mortality graduation for the entire age range using adaptive P-splines. We assumed that the smoothness penalty is an exponential function, hence at younger ages where more reliable data is available, the model enjoys higher freedom, whereas at the oldest ages where data is sparse, the heavier penalty improves robustness. Under this penalty, the model benefits from the flexibility offered by P-splines and is robust at ages with low exposures, making extrapolation to higher ages more stable.

Joint graduation of male and female mortality rates was also performed. Borrowing information in this way is useful especially for males at the highest ages since there are usually more female data at these ages compared to males. In addition, by introducing an additional penalty for the difference between male and female spline coefficients and constraining the spline coefficients, we could avoid divergences and cross-overs between male and female mortality rates. As a result, the approach described in this thesis provides a coherent way of modelling mortality across the whole age range even in regions where data is sparse or non-existent.

One potential drawback of our model is that mortality rates at ages beyond available data is extrapolated almost linearly, while in the literature there is evidence showing a decreasing increase in the log mortality rates at the oldest ages (Pitacco, 2016). A remedy would be to use a logistic model instead of a log-linear model, so that an asymptotic limit is introduced.

An alternative way of modelling mortality would be to use a fully Bayesian approach. That way the prior beliefs of the increasing smoothness and increasing similarity between male and female mortality rates can be incorporated into the model in a more natural framework.

Chapter 5

Mortality Projection - Joint Sex and Joint Country Model

Forecasts of mortality rates are necessary in many situations, for example it is needed for the estimation of future life expectancies and sizes of the old age population. Many insurance products such as life annuities also depend on these forecasts. Existing models for mortality forecast are generally based on extrapolations of historic trends (see Section 3.2). The LC model (Lee and Carter, 1992) pioneered the form $\log(m_{xt}) = \sum_i \alpha_x^i \kappa_t^i$ for mortality forecasting. Several models have built upon the LC model and proposed amendments focusing on different aspects. It is then realised that for some populations such as England and Wales, the year of birth has a significant effect on mortality rates, leading to a more general form of the age-period-cohort effect models $\log(m_{xt}) = \sum_i \alpha_x^i \kappa_t^i \gamma_{t-x}^i$. In these models, future mortality rates are produced by forecasting the time-relevant effects, i.e the period and cohort effects.

Renshaw and Haberman (2003) proposed a parallel GLM approach to the LC model. In this model, they model mortality improvements as deterministic linear trends; avoiding the bi-linear age-period term. Hilton et al. (2019) also modelled mortality improvements as deterministic linear trends under a GAM framework, and further added a simple period effect as random walk without drift to capture the stochastic behaviour and allow for probabilistic forecast intervals. The cohort effect has found to be significant in the United Kingdom (Cairns et al., 2009) and therefore it will also be considered in our model. In this chapter, we extend our mortality graduation model presented in Chapter 4 to mortality projection using adaptive P-splines. Our proposed model is

$$\log(m_{xt}^j) = s_\alpha^j(x) + s_\beta^j(x)t + \kappa_t^j + s_\gamma^j(t-x), \tag{5.1}$$

where $s_\alpha^j(\cdot), s_\beta^j(\cdot)$ and $s_\gamma^j(\cdot)$ are smooth functions representing the baseline mortality, age-specific mortality improvement rates and the cohort effects for population j. The period effect, κ_t^j, is assumed to be either an AR(1) or ARIMA(1,1,0). The model structure is similar to the spline component of the model in Hilton et al. (2019), except that in our case the splines are

fitted to the whole age range with specially constructed penalties, avoiding the necessity of estimating a 'threshold age' to switch from splines for the majority of ages to a parametric model for the oldest ages. All the smooth functions $s^j_\alpha(\cdot), s^j_\beta(\cdot)$ and $s^j_\gamma(\cdot)$ are estimated by P-splines, therefore we can write

$$\log(m^j_{xt}) = \boldsymbol{B}(x)\boldsymbol{\alpha}^j + \boldsymbol{B}(x)\boldsymbol{\beta}^j\, t + \kappa^j_t + \boldsymbol{B}_{\boldsymbol{\gamma}}(t-x)\boldsymbol{\gamma}^j, \tag{5.2}$$

where $\boldsymbol{B}(x)$ is the B-spline basis at age x and $\boldsymbol{B}_{\boldsymbol{\gamma}}(t-x)$ is the B-spline basis for year-of-birth $t-x$. All the B-spline bases are of dimension 40 with equally spaced knots. $\boldsymbol{\alpha}^j, \boldsymbol{\beta}^j$ and $\boldsymbol{\gamma}^j$ are spline coefficients of the baseline mortality schedule, improvement rates and cohort effects, respectively. For each smooth term, $s^j_\alpha(\cdot), s^j_\beta(\cdot)$, and $s^j_\gamma(\cdot)$, the same knot sequence is used for different populations, hence the same basis.

Since there is a linear relationship between age, period and cohort terms, the model is unidentifiable under some affine transformations. For example, for any constant c, we have the following relationships:

$$\begin{aligned}
\{s_\alpha(x), \kappa_t\} &\longmapsto \{s_\alpha(x) + c, \kappa_t - c\},\\
\{s_\alpha(x), s_\gamma(t-x)\} &\longmapsto \{s_\alpha(x) + c, s_\gamma(t-x) - c\},\\
\{\kappa_t, s_\gamma(t-x)\} &\longmapsto \{\kappa_t + c, s_\gamma(t-x) - c\},\\
\{s_\beta(x), \kappa_t\} &\longmapsto \{s_\beta(x) + c, \kappa_t - ct\},\\
\{s_\alpha(x), s_\beta(x), s_\gamma(t-x)\} &\longmapsto \{s_\alpha(x) + cx, s_\beta(x) - c, s_\gamma(t-x) + ct - cx\},\\
\{s_\alpha(x), \kappa_t, s_\gamma(t-x)\} &\longmapsto \{s_\alpha(x) + cx, \kappa_t - ct, s_\gamma(t-x) + ct - cx\},\\
\{s_\alpha(x), s_\beta(x), \kappa_t, s_\gamma(t-x)\} &\longmapsto \{s_\alpha(x) + cx^2, s_\beta(x) - 2cx, \kappa_t + ct^2, s_\gamma(t-x) - c(t-x)^2\}.
\end{aligned}$$

In each line, the set of coefficients on the left hand side and the set on the right after the transformations will give the same estimates. In order to overcome this unidentifiability problem, we impose the following constraints:

$$\begin{aligned}
\sum \kappa_t &= 0,\\
\sum t\kappa_t &= 0,\\
s_\gamma(1) &= 0,\\
s_\gamma(n_c) &= 0,\\
\sum s_\gamma(t-x) &= 0,
\end{aligned}$$

where n_c is the total number of cohort years.

With these constraints, we ensure that information about mortality improvements are contained in $s_\beta(x)$ with κ_t capturing yearly variations around it, $s_\alpha(x)$ is the mortality schedule on average

and the cohort effect $s_\gamma(t-x)$ is zero on average. Without loss of generality, the age and year-of-birth are rescaled to $[0,1]$, and the year is rescaled and centered to $[-0.5, 0.5]$.

5.1 Penalties

As discussed in Section 2.2.2, the essence of penalised splines is that the user supply a generous enough basis to capture variations and then introduce a penalty to prevent over-fitting. In this section the penalties used in our model are discussed.

5.1.1 Smoothness penalties

The baseline $s^j_\alpha(\cdot)$ captures the average mortality profile in age, therefore it is expected to behave similarly to the period mortality schedules (see Chapter 4), i.e. smoother at the higher ages. The smoothness penalty function is assumed to be an exponential function as given in (5.3) such that the penalty varies across ages.

On the other hand, for the age-specific improvement rates, $s^j_\beta(\cdot)$, a global smoothness penalty is assumed as given in (5.4). This is because the empirical data does not show a considerable change in smoothness over the age range. Similarly, the smoothness penalty for the cohort effect $s^j_\gamma(\cdot)$ is also global as given in (5.5).

Therefore, for each population j, there are three smoothness penalties,

$$P_{j,\alpha} = \sum_{i=3}^{k} \lambda_1^{j,\alpha} \exp(\lambda_2^{j,\alpha}\ i)(\bigtriangledown^2(\alpha_i^j))^2, \tag{5.3}$$

$$P_{j,\beta} = \lambda^{j,\beta} \sum_{i=3}^{k} (\bigtriangledown^2(\beta_i^j))^2 \tag{5.4}$$

$$\text{and} \quad P_{j,\gamma} = \lambda^{j,\gamma} \sum_{i=3}^{k} (\bigtriangledown^2(\gamma_i^j))^2, \tag{5.5}$$

where $P_{j,\alpha}$, $P_{j,\beta}$ and $P_{j,\gamma}$ are the smoothness penalties applied on the baseline mortality schedule $s^j_\alpha(\cdot)$, age-specific-improvement rates $s^j_\beta(\cdot)$ and cohort effect $s^j_\gamma(\cdot)$ for population j respectively.

5.1.2 Cross-sex penalties

As with mortality graduations of period life tables, a divergence in male and female mortality rates in the oldest ages where data is sparse or non-existent is undesirable. If the populations of interest are the males and females of the same country, additional cross-sex penalties are

introduced as follows. Let the superscripts m, f denote males and females respectively, then the cross-sex penalties are

$$P_{d,\alpha} = \sum_{i=9}^{k} \lambda_1^{d,\alpha} \exp(\lambda_2^{d,\alpha}\ i)(\alpha_i^m - \alpha_i^f)^2 \tag{5.6}$$

$$\text{and} \quad P_{d,\beta} = \sum_{i=9}^{k} \lambda_1^{d,\beta} \exp(\lambda_2^{d,\beta}\ i)(\beta_i^m - \beta_i^f)^2, \tag{5.7}$$

where $P_{d,\alpha}$ and $P_{d,\beta}$ are the cross-sex penalties applied on the baseline mortality schedules and the age-specific improvement rates respectively, and the superscript d indicates that the smoothing parameters are related to the cross-sex difference penalties. The cross-sex penalty is again only penalised from the 9-th basis function onwards (which is approximately right after the accident hump) for easier estimation of the smoothing parameters.

To avoid divergences in the baseline mortality rates, we employ the penalty given in (5.6), i.e. the squared differences between the male and female P-spline coefficients are penalised. Here the penalty function is exponential hence allows a lighter penalty at younger ages, where data is abundant and the levels and patterns of male and female mortality rates are expected to be less similar.

The same has to be applied to the improvement rates to maintain a non-divergent trends over time. Generally, if the mortality rates are already low, the improvement will be at a slower rate. Therefore if the male and female mortality improvement rates were to be modelled and extrapolated to higher ages independently, the projection of male mortality rates will eventually take over the projection of female mortality rates even in a short period of time. By penalising the differences between the male and female spline coefficients of the improvement rates as given in (5.7), this can be avoided. This also serves as a means to borrow information at ages with unreliable and scarce data. At the highest ages where the male and female mortality levels are very similar, we expect the improvement rates also to be similar.

5.1.3 Cross-country penalties

When populations of different countries are considered, in addition to the smoothness penalties for each term (5.3), (5.4) and (5.5), we introduce a cross-country penalty. For each sex, let the superscripts EW and SC denotes the England and Wales and Scotland populations respectively. Then we have the cross-country penalties

$$P_{s,\alpha} = \sum_{i=2}^{k} \lambda_1^{s,\alpha} \exp(\lambda_2^{s,\alpha}\ i)(\bigtriangledown\alpha_i^{EW} - \bigtriangledown\alpha_i^{SC})^2 \tag{5.8}$$

$$\text{and} \quad P_{s,\beta} = \sum_{i=2}^{k} \lambda_1^{s,\beta} \exp(\lambda_2^{s,\beta}\ i)(\bigtriangledown\beta_i^{EW} - \bigtriangledown\beta_i^{SC})^2, \tag{5.9}$$

where $P_{s,\alpha}$ and $P_{s,\beta}$ are the cross-country penalties applied on the baseline mortality schedules and the age-specific improvement rates respectively, and the superscript s indicates that the smoothing parameters are related to the cross-country shape difference penalties. For each sex, the squared differences between the first differences of the England and Wales and Scotland P-spline coefficients are penalised, both in the baseline mortality (5.8) and the age-specific improvement rates (5.9). In other words, any discrepancy in the shapes of the baseline and improvement rates is penalised. Therefore even in ages where Scottish data is non-existent, information can be learnt from the larger England and Wales population. A global cross-country penalty is assumed hence we can learn how 'similar' the two populations are for the whole age range.

To simplify the estimation process, common smoothing parameters for the smoothness of $s_\alpha(x)$, $s_\beta(x)$ and $s_\gamma(t-x)$ are assumed for EW and SC males and females. In our application with Scottish male data when they are modelled independently, the estimated smoothness penalty of the baseline mortality schedule is very large at the oldest ages (5.3). This is possibly due to the fact that we only have data up to age 99 with very low exposures, where the log mortality rates are still quite linear. Therefore the smoothness penalty would dominate and render the cross-country shape penalty ineffective. Hence, by assuming common smoothing parameters for the smoothness penalties between the two countries, we reduce the number of parameters to be estimated, and also avoid the possibility of a dominating Scottish males smoothness penalty.

5.2 Period and Cohort Effects

In our proposed model (5.1), period and cohort effects can be interpreted as the accumulated period shocks and cohort development to the annual mortality improvement. To see this, for a given population, consider

$$\begin{aligned}\log(m_{x,t}) - \log(m_{x,t-1}) &= s_\alpha(x) - s_\alpha(x) + s_\beta(x)t - s_\beta(x)(t-1) + \kappa_t - \kappa_{t-1} \\ &\quad + s_\gamma(t-x) - s_\gamma(t-1-x) \\ &= s_\beta(x) + (\kappa_t - \kappa_{t-1}) + (s_\gamma(t-x) - s_\gamma(t-x-1)) \\ &= s_\beta(x) + \kappa_t^* + s_\gamma^*(t-x) \qquad (5.10)\end{aligned}$$

where κ_t^* and $s_\gamma^*(t-x)$ are the first differences of the period and cohort effects respectively. Once written out in the form of (5.10), it can be seen that κ_t^* and $s_\gamma^*(t-x)$ are the period and generational impact on mortality improvement respectively.

In the literature, to forecast the period and cohort effects, relatively simple time series models are used. For example, in most of the LC type models, the main period effect is forecast using a random walk with drift. However, since there is already a linear term for mortality improvements in our model, the period effect should capture annual variations around the deterministic trend. We consider projecting the period effect using relatively simple time series,

either an AR(1) process or an ARIMA(1,1,0) process without intercept. The advantage of using an AR(1) process is that it produces stationary forecasts. On the other hand, the ARIMA(1,1,0) has the meaning that the expected future period shocks to the annual mortality improvement are zero, which maybe a desirable feature.

In the literature the cohort effect is also forecast using time series methods similar to the period effect. Here we model the cohort effect as a smooth term. In other words, mortality changes for cohorts born in different years is a smooth process. We assume the generational impact of future cohorts to be zero (i.e. $s^*_\gamma(\cdot)$ in (5.10) for future cohorts is zero), therefore the we let the first difference of future cohort effects tail off to zero. This can be achieved easily by setting the coefficients of the B-spline basis functions of the future cohorts to be the same as the last coefficient of the estimated $s_\gamma(t-x)$.

5.3 Forecast Uncertainty

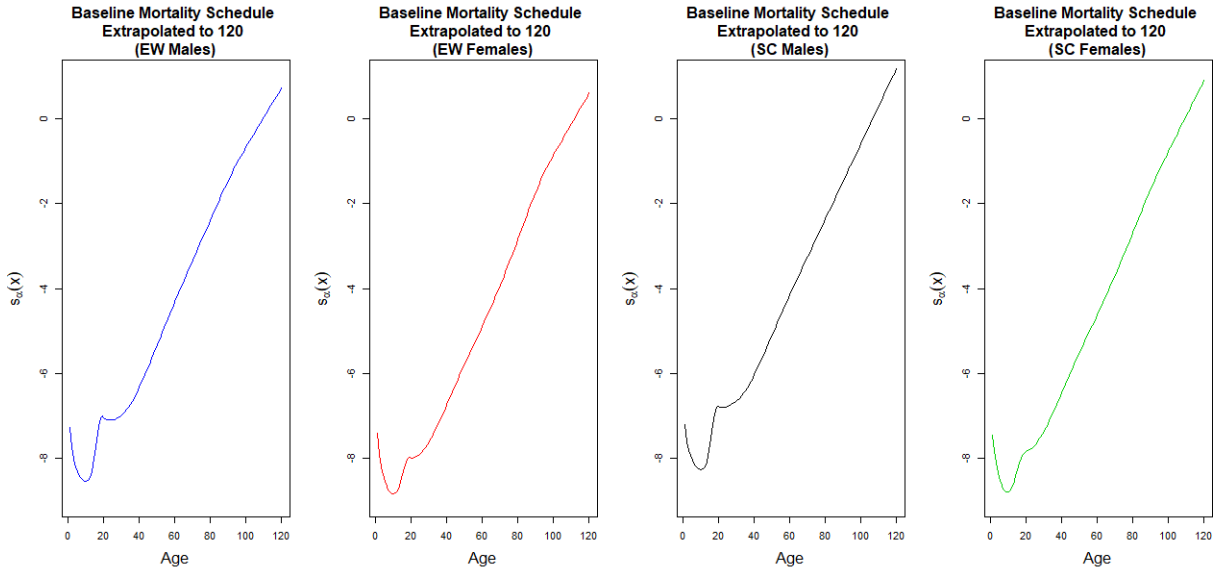
In Section 2.2.6, the parameter uncertainty and approximations to the distributions of the estimated parameters have been discussed. However, when mortality forecasts are considered, the forecast uncertainty also has to be taken into account. More specifically, recall that in our model, mortality projection reduces to projecting the period effect, therefore the uncertainty about the projected period effect has to be quantified. When the historic period effect is known or given, the forecast uncertainty can be calculated in a relatively straightforward way, however, when the historic period effect is also being estimated with uncertainty, as well as being correlated with other estimated parameters such as the splines coefficients and the cohort effects, quantifying the uncertainty of the projected period effect becomes a complicated task. Therefore we resort to bootstrapping techniques in order to measure the uncertainty of the projected mortality rates. As the joint distribution of the estimated parameters is available, we first draw samples from this distribution. After that a time series model (AR(1) or ARIMA(1,1,0)) is fitted to the sampled historic period effect to obtain an estimate of the variance of the innovations and estimates of the mean and variance of the estimated time series coefficients. The time series coefficients are then sampled as well as the projected period effects. Finally the quantiles of these forecast samples are obtained.

5.4 Results

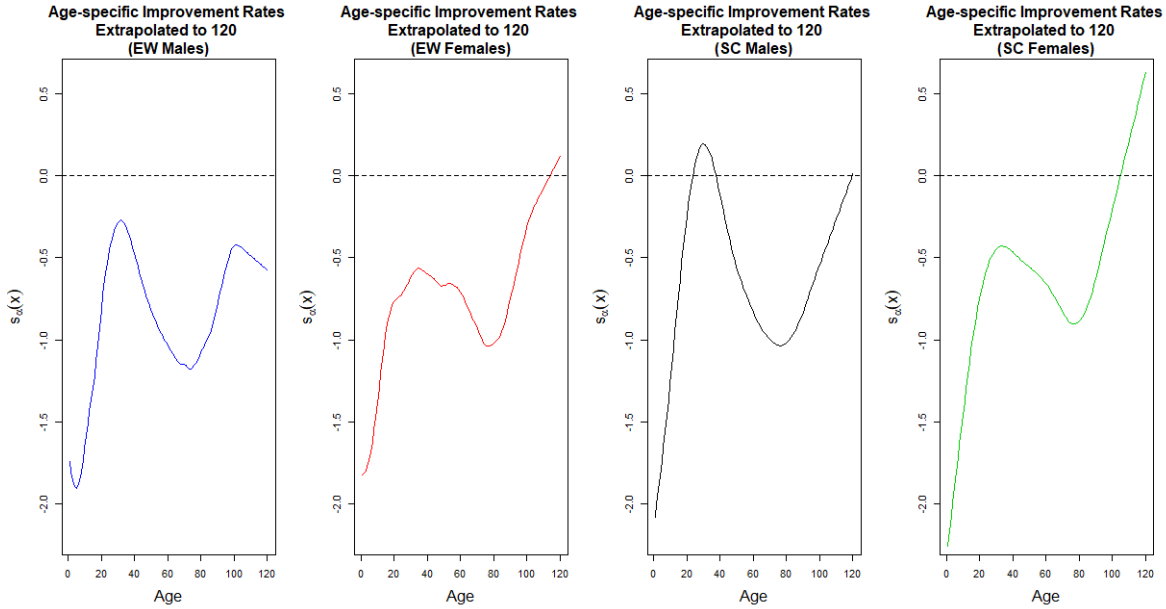
5.4.1 Independent Models

In this subsection we first present the results of individual models for each population. When we model each population independently we only need the smoothness penalties (i.e. (5.3), (5.4) and (5.5)). Figure 5.1 shows the parameter estimates for each population when they are

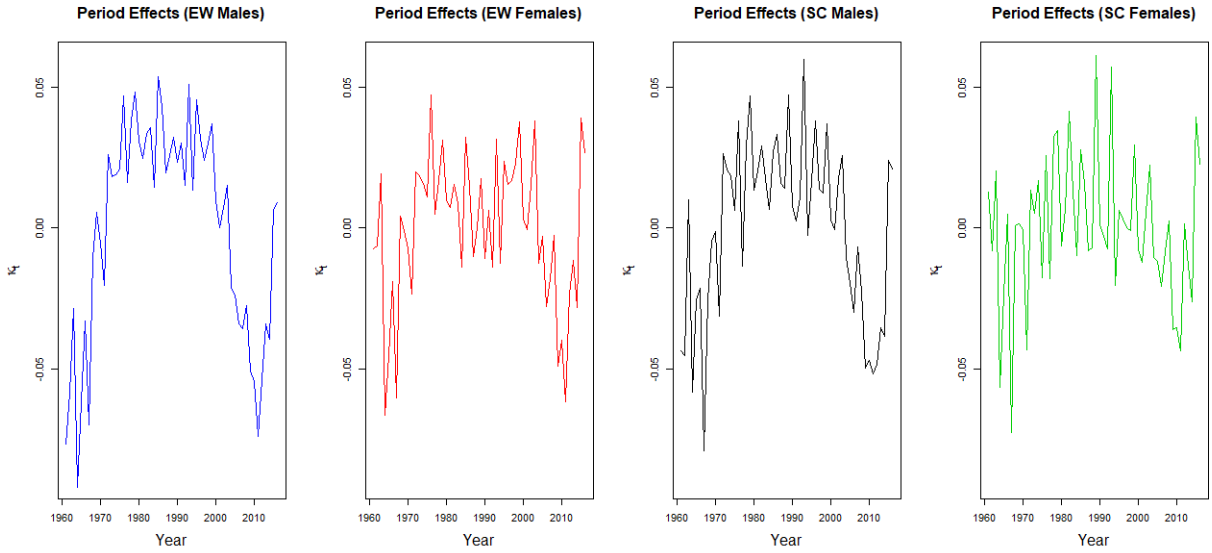
modelled individually. A slight curvature (decelerating rate of increase) can be observed at the oldest ages in the estimated baseline mortality schedules for EW males and females, while the same cannot be noticed for SC males and females. As discussed in Section 3.1, studies in the literature have pointed out that a decreasing rate of increase in the log mortality rates is often observed at the highest ages (approximately after age 94 according to Carriere (1992)). The linearity in the highest ages of the SC populations is possibly due to the limited data range the SC population has. Recall that for SC males and females only data up to age 99 is available, where within this range the log mortality rates are still relatively linear, therefore the decelerating rate of increase may not show. Figure 5.1b shows the estimated age-specific improvement rates. Note that the more negative the estimates, the higher the mortality improvement. Some common characteristics seem to be shared among the populations. Mortality improvement is the fastest at the youngest ages, then slowing down until it reaches young adulthood, especially for males. From late 30s to late 70s there is a huge mortality improvement for both males and females in EW and SC. Finally the mortality improvement slows down up to the oldest ages. The age-specific improvement rates of the male and female have similar shapes between countries. However, note that the estimated mortality improvement is above zero for SC young adult males and at the extrapolated ages for EW and SC females, meaning that mortality is worsening. From Figure 5.1c we see that the period effects for males and females for both countries are very similar to each other, and that they are highly correlated. Figure 5.1d presents the estimated cohort effects for these populations. The cohort effects for EW males and females have a similar structure.



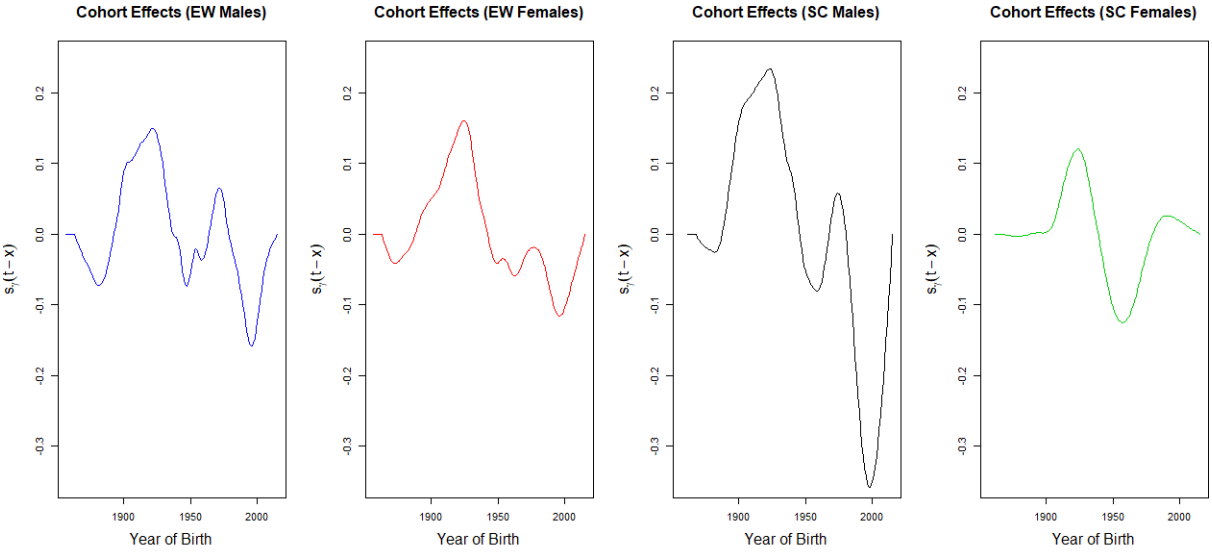
(a) estimated baseline mortality schedules extrapolated to age 120



(b) estimated age-specific improvement rates extrapolated to age 120



(c) estimated period effects

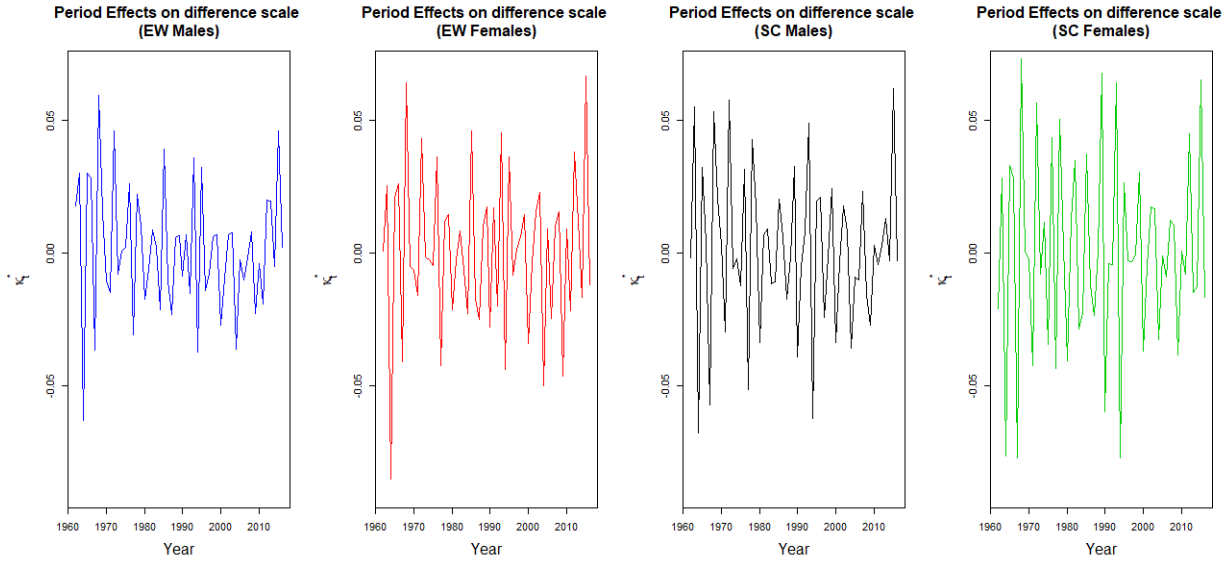


(d) estimated cohort effects

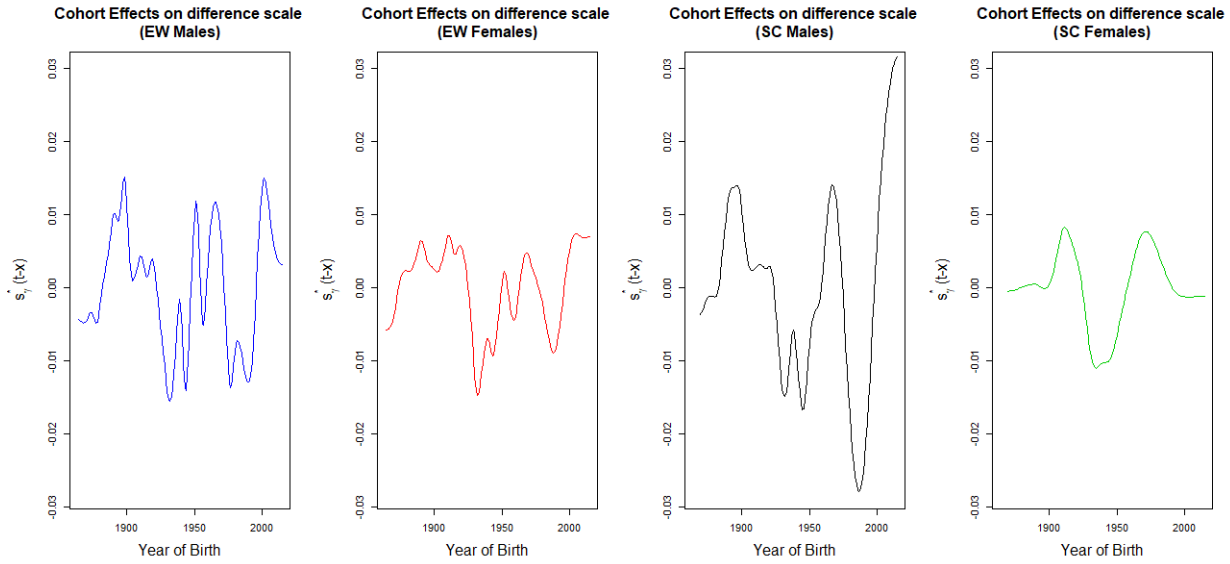
Figure 5.1: The parameter estimates for EW males (blue), EW females (red), SC males (black) and SC females (green).

More insights can be gained by looking at the period and cohort effects on the difference scale (5.10), as they resemble their impacts on the annual mortality improvement. Figure 5.2 plots the estimated period and cohort effects on a difference scale. The estimated κ_t^*'s are oscillating around zero, indicating that the linear mortality improvement trend is a suitable assumption, although in the most recent years there seems to be a deviation from the linear trend. The estimated $s_\gamma^*(t-x)$'s show significantly better mortality improvement for cohorts born around years 1931 and 1944. The first trough is a well documented phenomenon in the literature (Willets, 2004; Richards et al., 2006; Murphy, 2009) and the second one is also pointed out by Willets

(2004). A better mortality improvement is also observed for cohorts born around 1988, except for SC females.



(a) estimated period effects on a difference scale



(b) estimated cohort effects on a difference scale

Figure 5.2: The estimated period and cohort effects for each population on a difference scale.

5.4.2 Joint Sex Models

In this subsection we present the results of the joint sex model for EW and SC. Male and female mortality rates are expected to converge at the oldest ages. In addition, at ages where data is non-existent, it is essential to borrow information from the two sexes about the extrapolation trends. Therefore to model males and females jointly, we employ the smoothness penalties

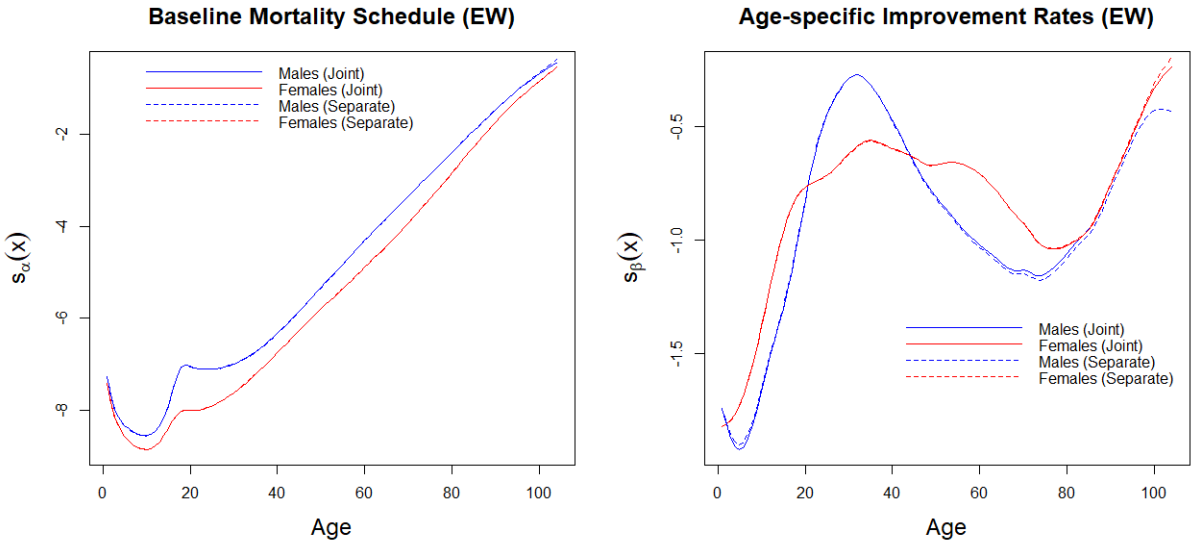
given in Section 5.1 as well as the cross-sex penalties given in Section 5.1.2. The penalties that are used in the joint model of males and females are then given in Table 5.1.

Function	Smoothness Penalty	Cross-sex Penalty
$s_\alpha(x)$	$\sum e^{\lambda_1^{m,\alpha}+\lambda_2^{m,\alpha}i}(\bigtriangledown^2\alpha_i^m)^2$ $+$ $\sum e^{\lambda_1^{f,\alpha}+\lambda_2^{f,\alpha}i}(\bigtriangledown^2\alpha_i^f)^2$	$\sum e^{\lambda_1^{d,\alpha}+\lambda_2^{d,\alpha}i}(\alpha_i^m-\alpha_i^f)^2$
$s_\beta(x)$	$\sum \lambda^{m,\beta}(\bigtriangledown^2\beta_i^m)^2$ $+$ $\sum \lambda^{f,\beta}(\bigtriangledown^2\beta_i^f)^2$	$\sum e^{\lambda_1^{d,\beta}+\lambda_2^{d,\beta}i}(\beta_i^m-\beta_i^f)^2$
$s_\gamma(t-x)$	$\sum \lambda^{m,\gamma}(\bigtriangledown^2\gamma_i^m)^2$ $+$ $\sum \lambda^{f,\gamma}(\bigtriangledown^2\gamma_i^f)^2$	

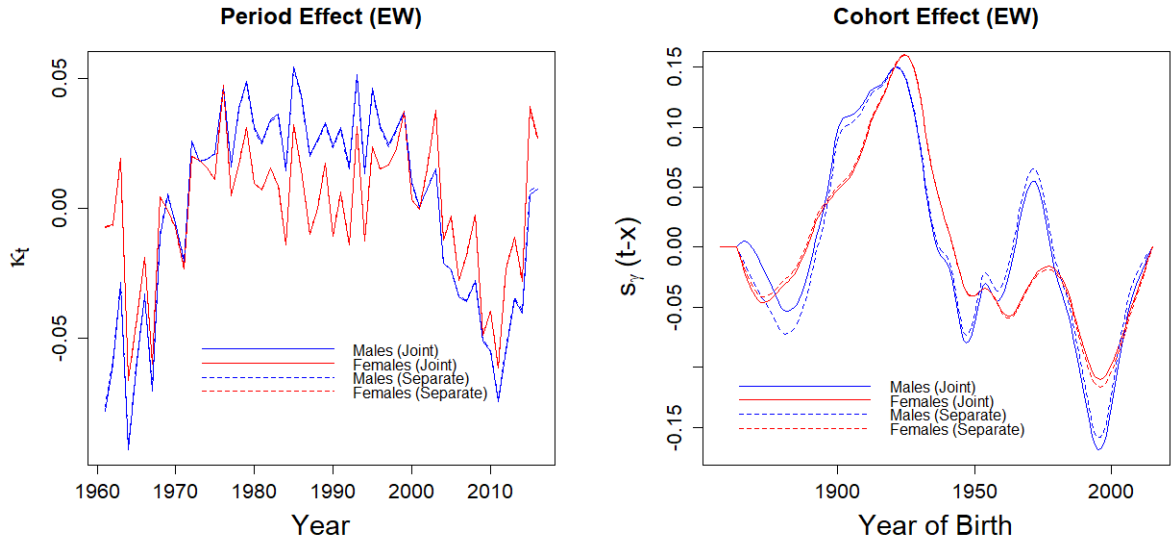
Table 5.1: The penalties for the Joint Sex Model

In the estimation of the smoothing parameters for EW, some computational challenges are experienced. Specifically, the estimated penalty function of the differences between male and female improvement rates was very high at old ages, approaching infinity (i.e. the $\exp(\lambda_1^{d,\beta}+\lambda_2^{d,\beta}\ i)$), causing numerical issues in the optimisation. Comparing the age-specific improvement rates of EW males and females when they were fitted independently (Figure 5.1b), it can be seen that the improvement rates are very similar to each other after around age 85. This is the reason why the estimated cross-sex difference penalty function is approaching infinity after a certain basis function when they are jointly modelled. In order to avoid the computational issues, instead of estimating these smoothing parameters, we fix $\lambda_1^{d,\beta}$ and $\lambda_2^{d,\beta}$ at -55 and 85. Once these parameters are fixed at these levels, the remaining smoothing parameters are optimised successfully.

Figure 5.3 plots the estimated baseline mortality schedules $s_\alpha(x)$, improvement rates $s_\beta(x)$, period effects κ_t and cohort effects $s_\gamma(t-x)$ of the joint model, for EW males and females while Figure 5.4 plots the same for SC males and females. The corresponding estimates from the single sex model (i.e. without the cross-sex penalties) are also shown for comparison. The corresponding estimates from the independent models (i.e. without the cross-sex penalties) are also plotted for comparison.

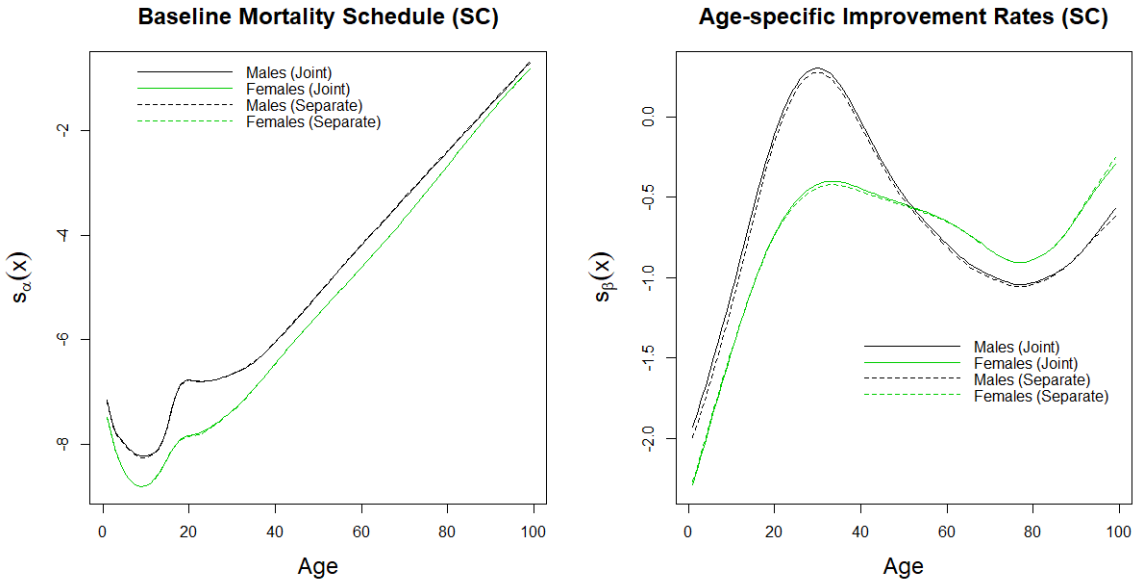


(a) estimated baseline mortality schedules and age-specific improvement rates

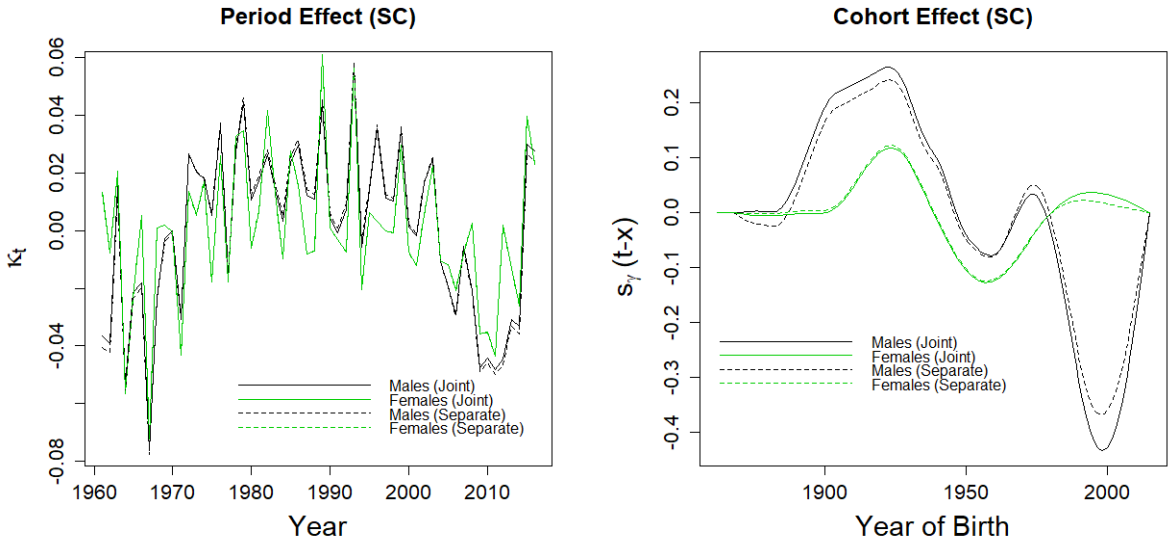


(b) estimated period and cohort effects

Figure 5.3: The parameter estimates of the joint sex model for EW males (blue) and females (red). The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.



(a) estimated baseline mortality schedules and age-specific improvement rates

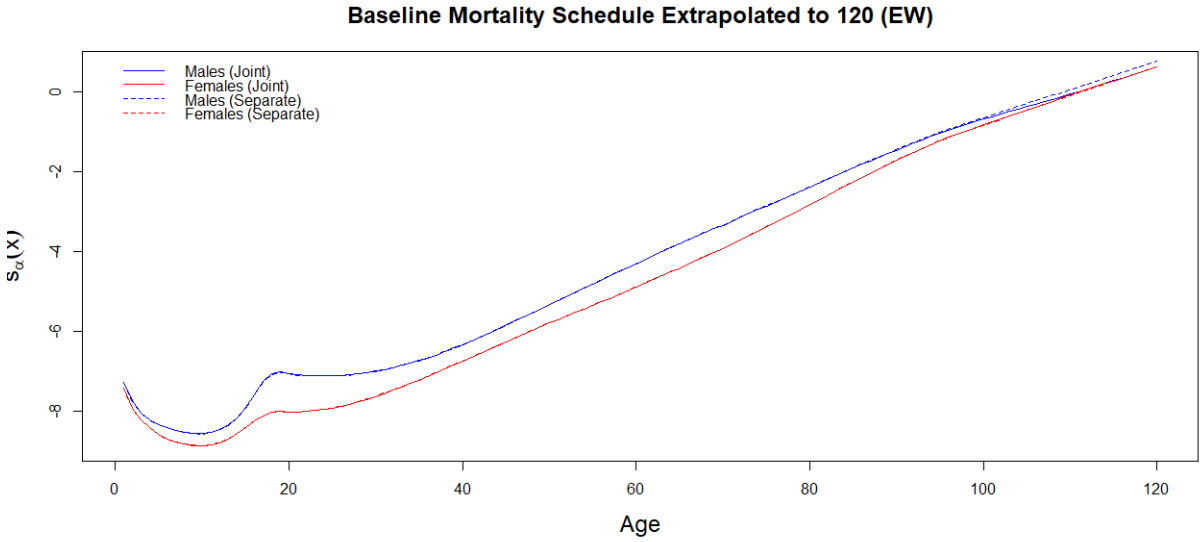


(b) estimated period and cohort effects

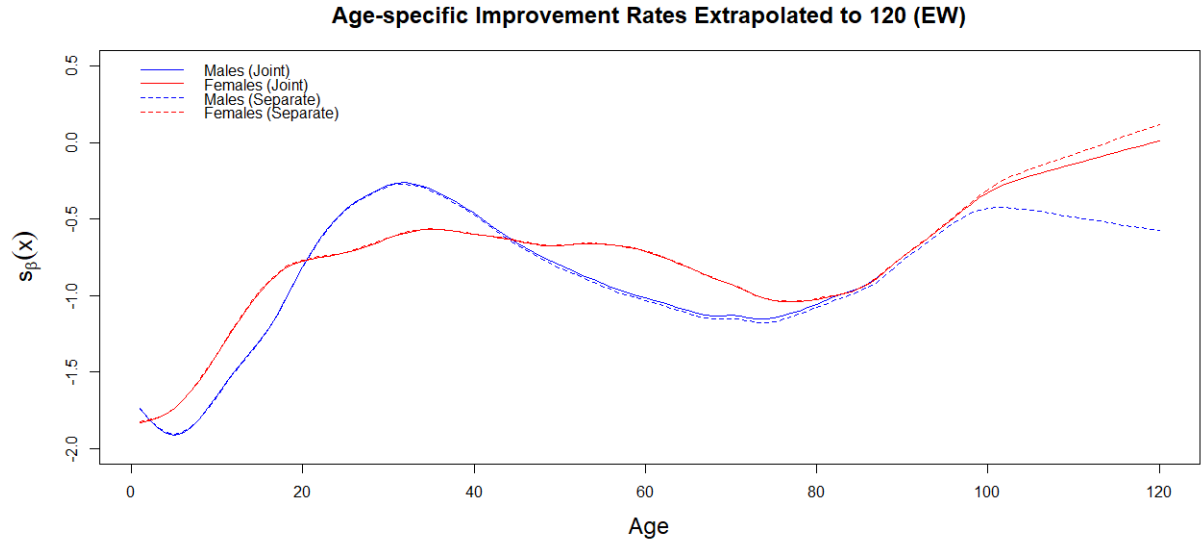
Figure 5.4: The parameter estimates of the joint sex model for SC males (black) and females (green). The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.

Comparing the joint sex and single sex (independent) models (solid and dotted lines respectively) for EW males and females, for the baseline mortality schedules the difference is almost indistinguishable, except at the oldest ages where the estimated curves start to converge in the joint model. For the age-specific improvement rates, the effect of the cross-sex penalty is more apparent, the male and female improvement rates tend to a common level at high ages. Under the independent model (dotted lines), at old ages where male and female baseline mortality rates are similar, the faster male improvement rate compared to that of females means that the projection of male mortality rates will drop below that of female mortality rates even in a short period

of time. On the other hand, this problem is not experienced in the joint sex model, providing more reasonable future mortality rates for both males and females. For Scotland, the data shows less convergence in both the baseline mortality and improvement rates compared to the England and Wales. The estimated penalties on the cross-sex differences are also relatively smaller. The estimated period effects from the joint sex and single sex models are very similar and the male and female period effects seem to be correlated. This is sensible as they are of the same country, hence exposed to similar shocks in mortality rates. The cohort effect changes slightly while maintaining the overall shape.

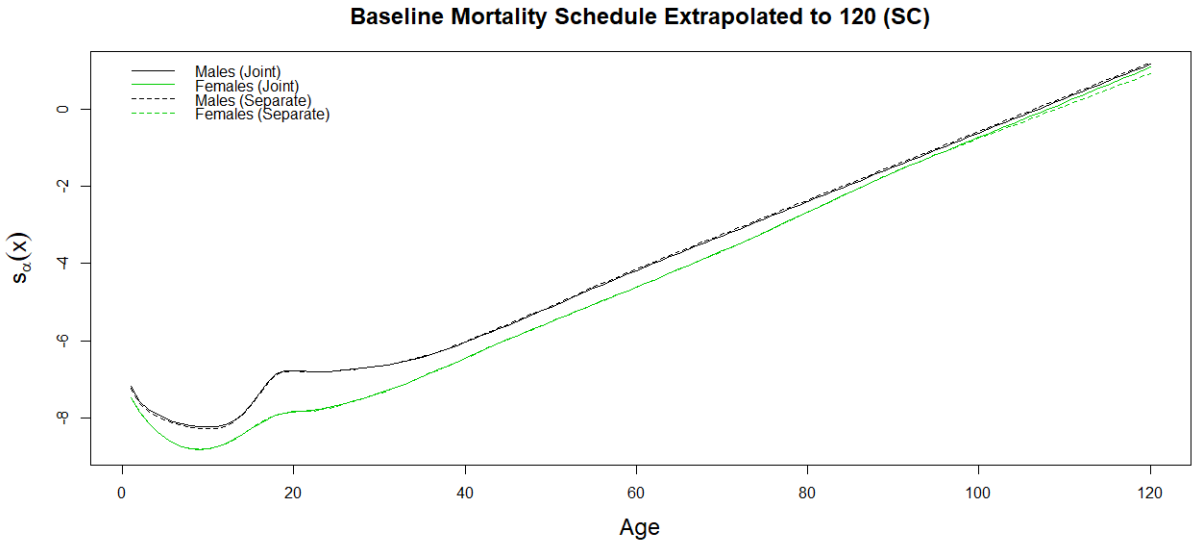


(a) estimated baseline mortality schedules extrapolated to age 120

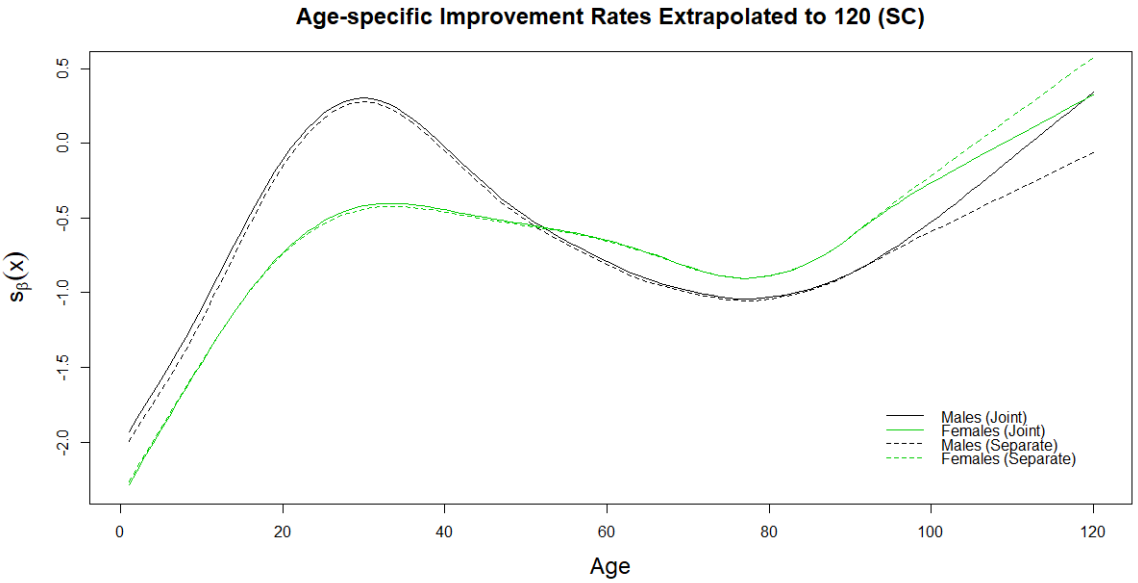


(b) estimated age-specific improvement rates extrapolated to age 120

Figure 5.5: The extrapolated baseline mortality schedule and age-specific improvement estimates of the joint sex model for EW males (blue) and females (red). The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.



(a) estimated baseline mortality schedules extrapolated to age 120



(b) estimated age-specific improvement rates extrapolated to age 120

Figure 5.6: The extrapolated baseline mortality schedules and age-specific improvement rates for SC males (black) and females (green). The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.

Figures 5.5 and 5.6 plot the baseline mortality schedule and the improvement rates extrapolated to age 120 under both the joint model and the independent model for EW and SC respectively. For EW, the female improvement rates are almost unchanged in the two models, except at the oldest ages where the exposure of males is comparable to that of females. The extrapolated EW male improvement rates are faster than that of EW females under the independent model. As a result, projections based on these trends will be implausible whereas under the joint model, strength is borrowed and the projected mortality rates are much more reasonable at these ages (Figure 5.15b). For SC males and females, under the joint model the male and female baseline

mortality profiles still tend to a common level, although less significant than the EW males and females. The SC male mortality improvement rates are also way lower than that of SC females at adult ages even with the presence of the cross-sex difference penalty, but eventually converge.

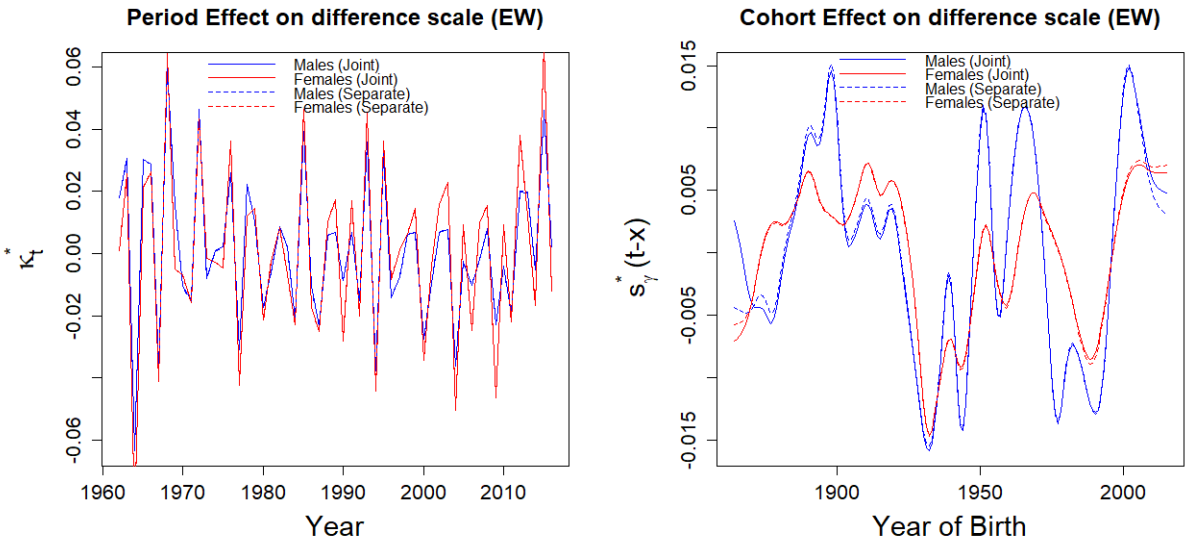


Figure 5.7: The estimated period and cohort effects for EW males (blue) and females (red) on a difference scale. The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.

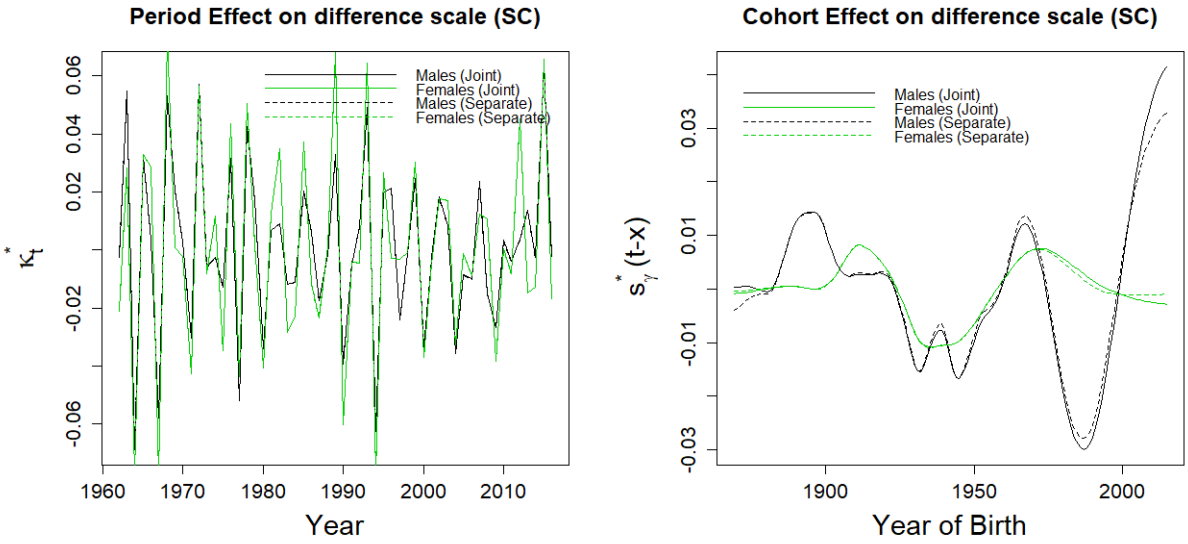


Figure 5.8: The estimated period and cohort effects for SC males (black) and females (green) on a difference scale. The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.

Figures 5.7 and 5.8 plot the estimated period and cohort effects on a difference scale for EW and SC respectively. For both EW and SC, the κ_t^*'s oscillate around 0 for both males and females, indicating that the linearity assumption for mortality improvement is appropriate (i.e. $s_\beta(x)t$

in (5.1)). It can also be noted that the κ_t^* for males and females are correlated, which is not surprising as males and females of the same country are likely to experience common mortality shocks.

The $s_\gamma^*(t-x)$'s oscillate around zero. There is a pronounced trough for cohorts born about 1932 for EW males, EW females and SC males, which agrees with the well documented cohort effect (Willets, 2004; Richards et al., 2006; Murphy, 2009), where they experience higher mortality improvements than neighbouring cohorts. Another trough could be observed for cohorts born about 1944, which is also pointed out by Willets (2004). For SC females, there is only one big trough spanning around 1925 to 1950. In addition, cohorts born around 1988 also seem to have better mortality improvement rates, except for SC females. The relatively larger discrepancy between the joint and independent model in the earliest and latest cohorts is possibly due to lack of data, as we have only less than few years of data for those cohorts, hence the variability.

5.4.2.1 Backtesting

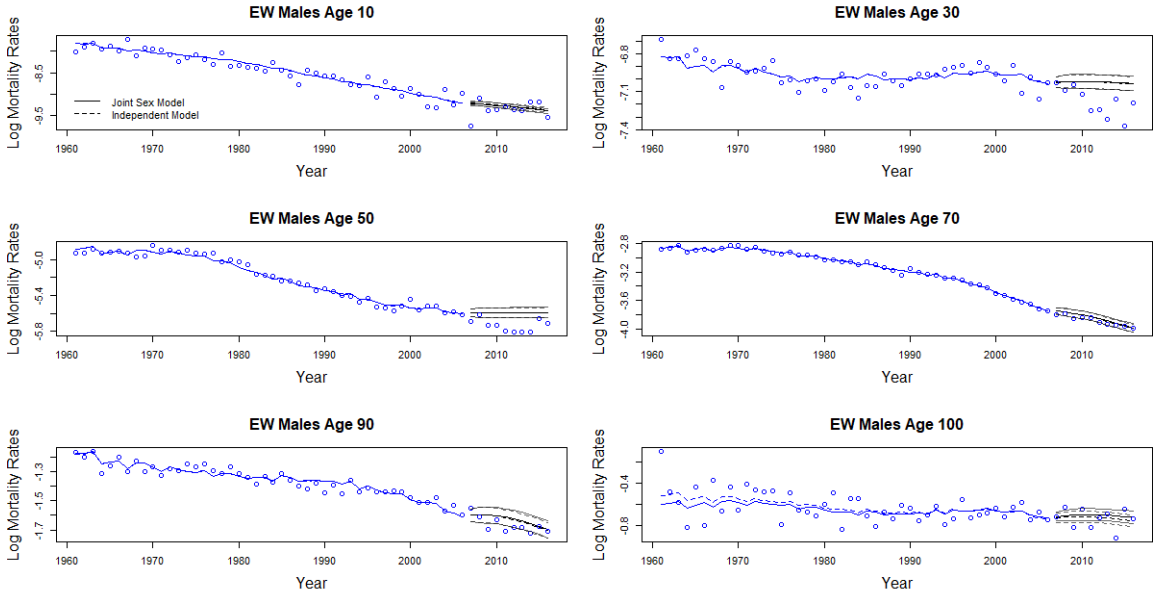
We conduct backtests on the models excluding data from the last 10 years, i.e. data from 1961 to 2006 are used to train the model while data from 2007 to 2016 are used as validation. Following Booth et al. (2006), accuracy is measured by the mean absolute error (MAE) of the log mortality rates as given in (5.11).

$$MAE = \sum_x \sum_t |\log \tilde{m}_{xt} - \log \hat{m}_{xt}|. \tag{5.11}$$

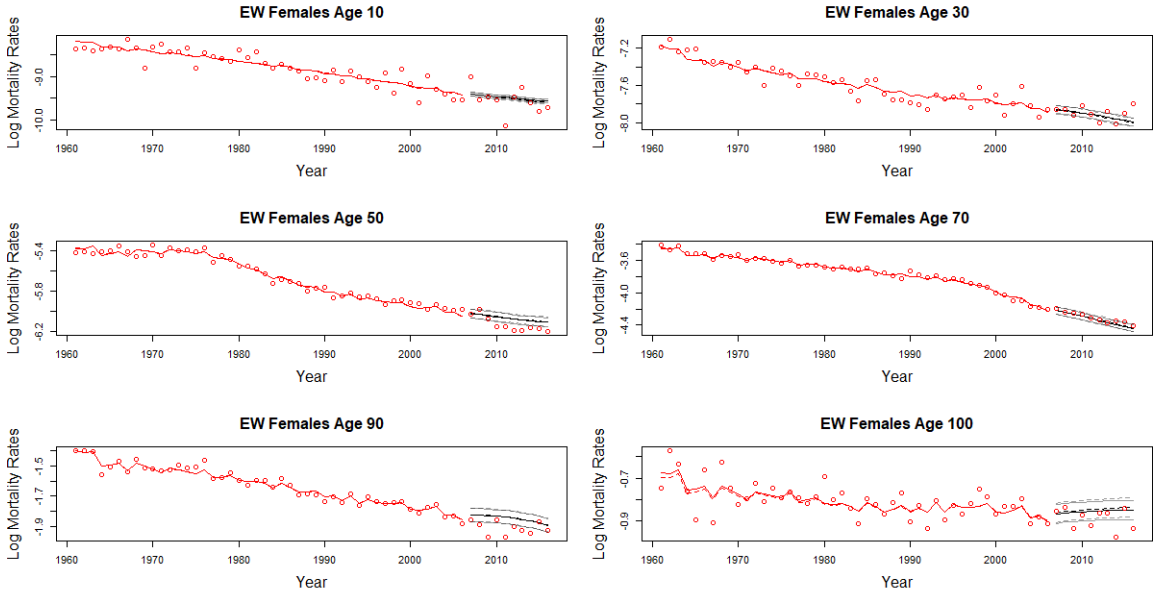
Here $\tilde{m}_{xt}$ is the crude (observed) mortality rates at age x in year t and $\hat{m}_{xt}$ the corresponding estimate.

Figures 5.9 and 5.10 plot the fitted (1961 to 2006) and projected mortality rates (2007 to 2016) at some selected ages for EW and SC respectively using an AR(1) process on the period effects together with the 95% CI. The black solid and dotted lines correspond to the projected mortality rates produced by the joint and independent models, respectively. There are little differences between the joint model and the independent model except at the oldest ages for EW males and females where the contrasts become more noticeable. Tables 5.2 and 5.3 show the MAE of the joint model and the independent model for EW and SC respectively. The joint models have slightly higher MAE except at the youngest ages. An ARIMA(1,1,0) is also fitted to the period effects instead of an AR(1), and the projections are shown in Figures 5.11 and 5.12. This is equivalent to fitting an AR(1) process to the first differences of the period effects. As shown above, the first differences of the period effects resemble the shocks to the annual mortality improvement rates, therefore the ARIMA(1,1,0) means that the expected future shocks to annual mortality improvements will gradually revert to zero. Contrary to the results for EW when using an AR(1), the MAE of the joint model using an ARIMA(1,1,0) is lower than the MAE of the independent model in all but the 31-60 age group. The MAE of the models with ARIMA(1,1,0) is also lower in all age groups compared to the models with AR(1). For SC, the results are

similar when an ARIMA(1,1,0) is used compared to the AR(1) process, with ARIMA(1,1,0) having a slightly higher MAE than the AR(1). Table 5.4 shows the coverage of the estimated 95% intervals of the joint sex model. On average, the intervals produced using AR(1) are too narrow. On the other hand, ARIMA(1,1,0) produces wider intervals and hence higher coverage. The overall coverage of the intervals using ARIMA(1,1,0) are still below 95%, however, for older age groups, the coverage are much closer to 95%, especially for EW males aged 61-90. The EW populations also have better coverage than the SC populations.

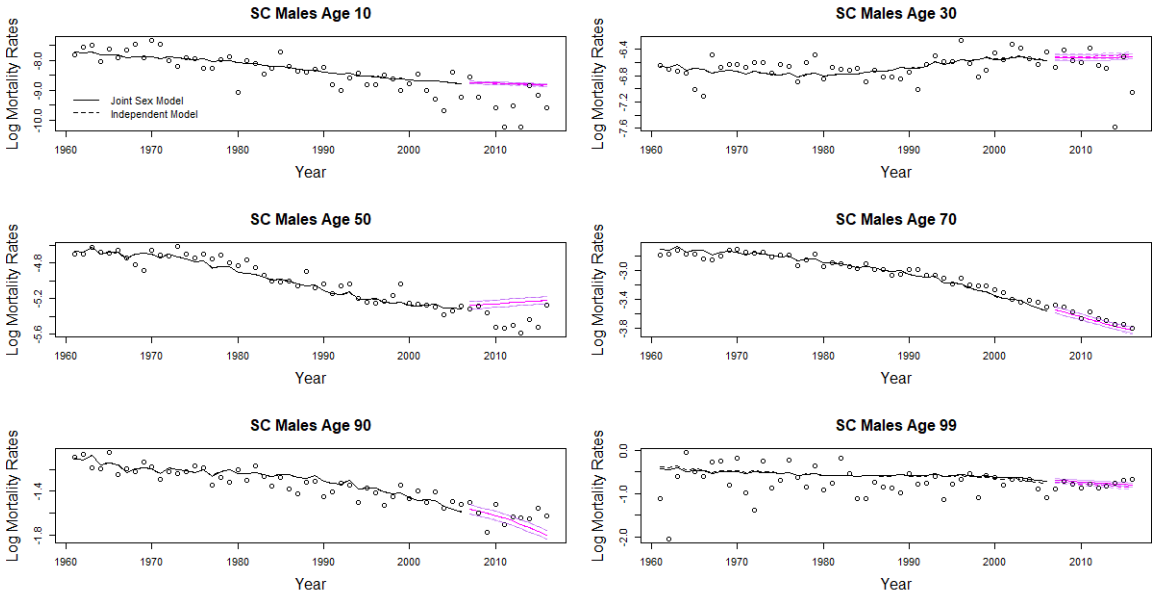


(a) EW males

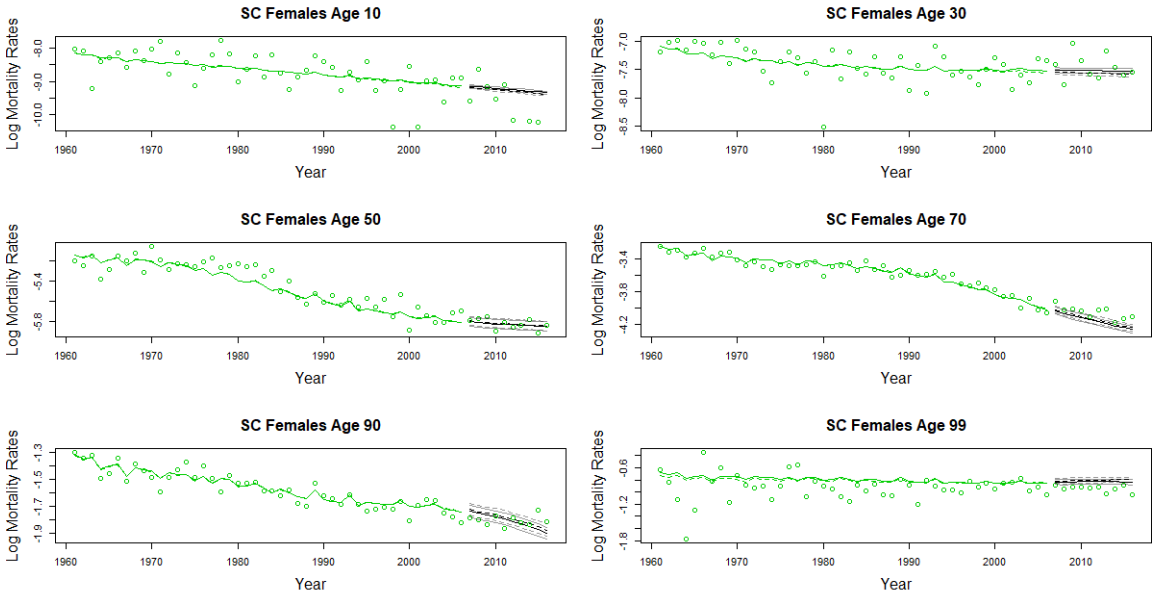


(b) EW females

Figure 5.9: Estimated mortality rates and the projected 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an AR(1). The points are the observed mortality rates.

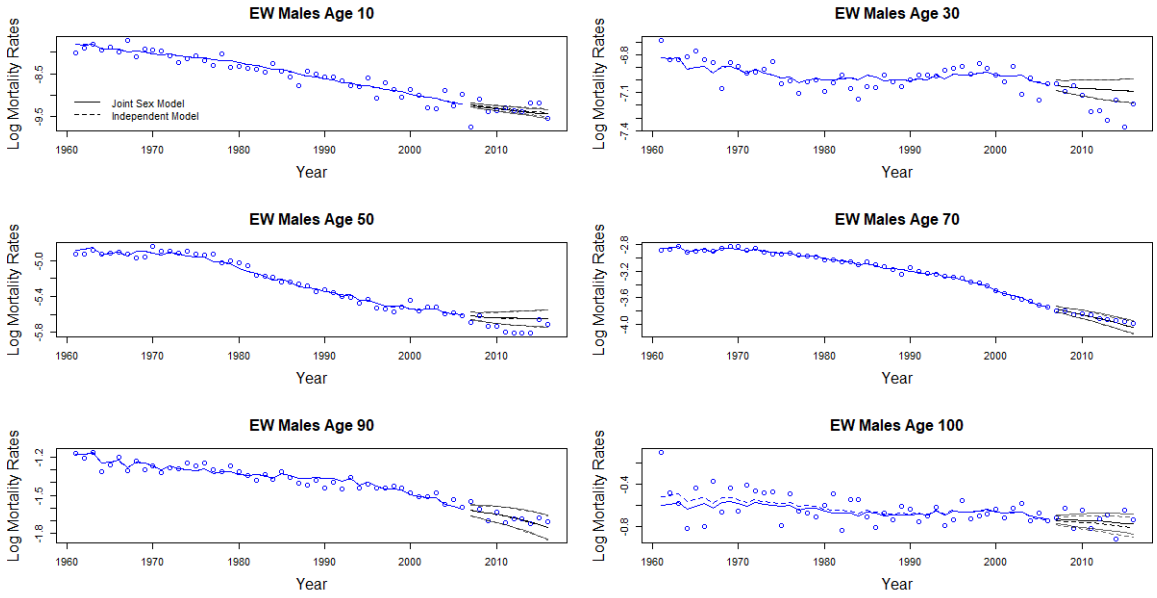


(a) SC males

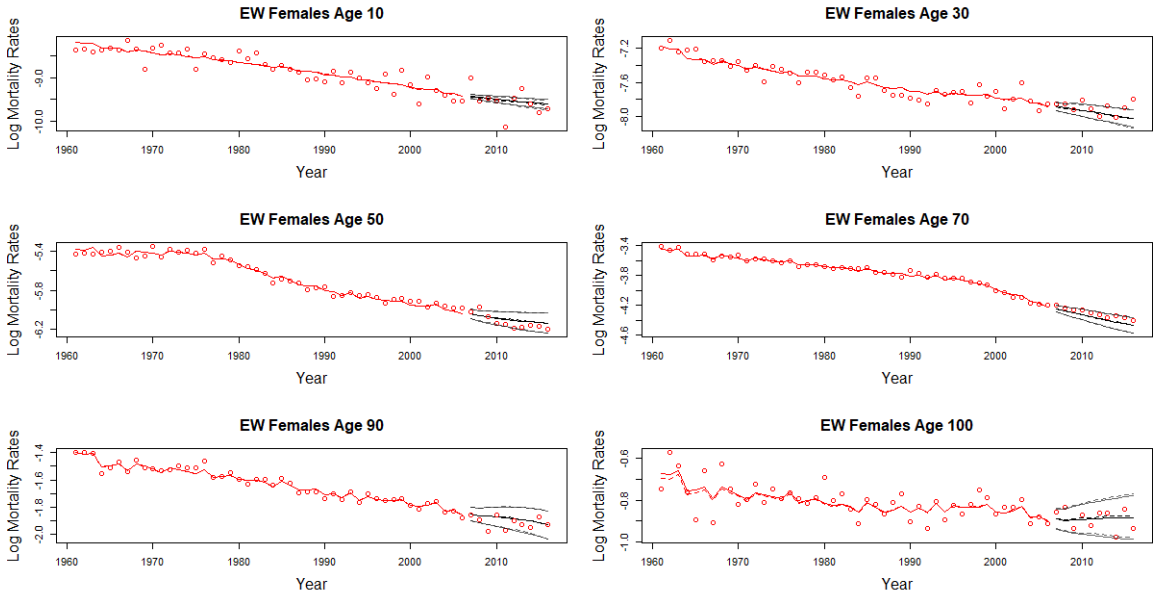


(b) SC females

Figure 5.10: Estimated mortality rates and the projected 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an AR(1). The points are the observed mortality rates.

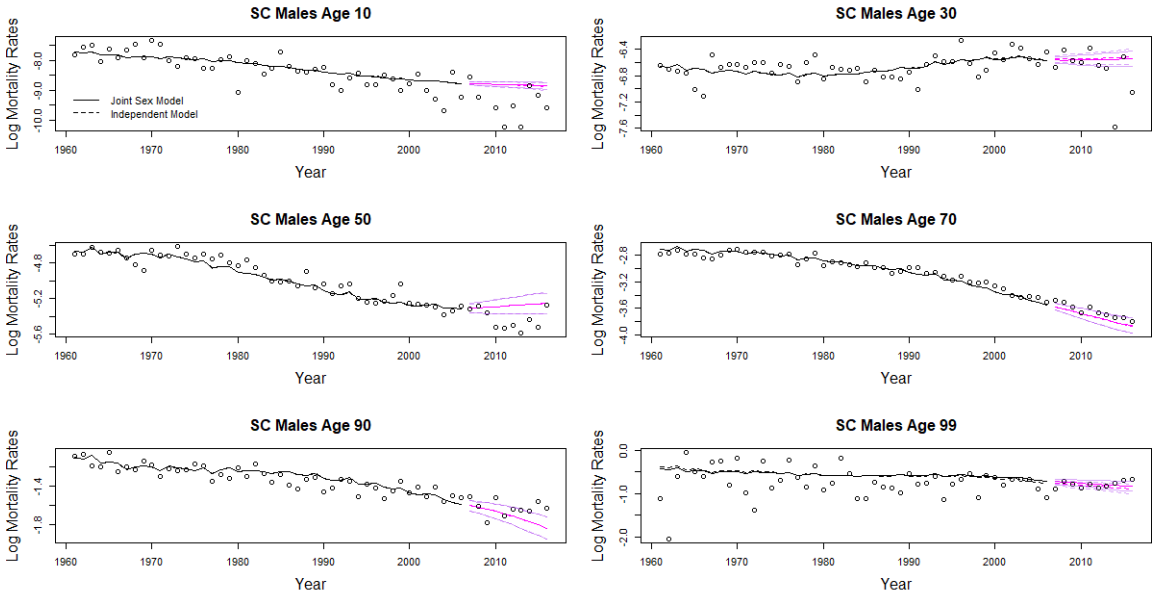


(a) EW males

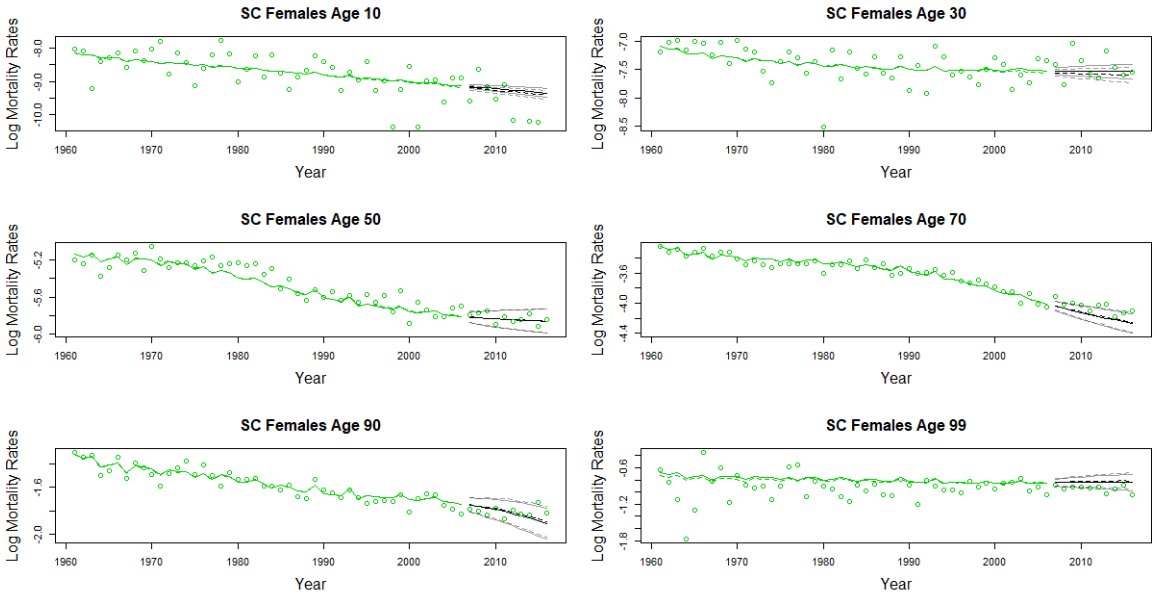


(b) EW females

Figure 5.11: Estimated mortality rates and the projected 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an ARIMA(1,1,0). The points are the observed mortality rates.



(a) SC males



(b) SC females

Figure 5.12: Estimated mortality rates and the projected 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an ARIMA(1,1,0). The points are the observed mortality rates.

Age	EW males and females			
	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	0.1537691	0.1546165	0.1376645	0.1391921
31 to 60	0.08711875	0.086363	0.06133899	0.06077347
61 to 90	0.04220626	0.04166751	0.03567749	0.03597661
91+	0.06147689	0.06051831	0.0479942	0.04903926
all ages	0.08993749	0.08967949	0.07449548	0.07499984

Table 5.2: The mean absolute error of log projected mortality rates for EW males and females

Age	SC males and females			
	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	2.824223	2.844317	2.84601	2.86908
31 to 60	1.574693	1.574558	1.59648	1.599322
61 to 90	1.014488	1.006697	1.036275	1.03146
91+	0.3062166	0.302636	0.3212329	0.3211467
all ages	1.668263	1.671625	1.689435	1.69582

Table 5.3: The mean absolute error of log projected mortality rates for SC males and females

Age	95% intervals coverage							
	EW males		EW females		SC males		SC females	
	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)
all ages	0.4576923	0.6163462	0.4865385	0.7644231	0.302020191	0.434343403	0.331313127	0.536363633
61-90	0.7666667	0.9633333	0.63	0.89	0.4833333	0.5833333	0.4133333	0.6466667
91+	0.6285714	0.7428571	0.4571429	0.8857143	0.3111111	0.5111111	0.4444444	0.8333333

Table 5.4: The coverage of the 95% intervals of the joint sex model

5.4.2.2 Projection

We present the projections (from 2016) of the joint sex model in the age and time direction, alongside with the projections of the single sex models. The male and female mortality rates should be non-intersecting such that the female mortality rates are always below male mortality rates at each age in any time. Figures 5.13 and 5.14 plot the 1, 25 and 50-year ahead projections of the models for EW and SC respectively. The joint sex model largely eliminates the problem of female mortality rates overtaking male mortality rates at older ages. At younger ages, the differences between the two models are smaller and the female childhood mortality eventually takes over the male childhood mortality in the 50-year ahead projections for EW males and females.

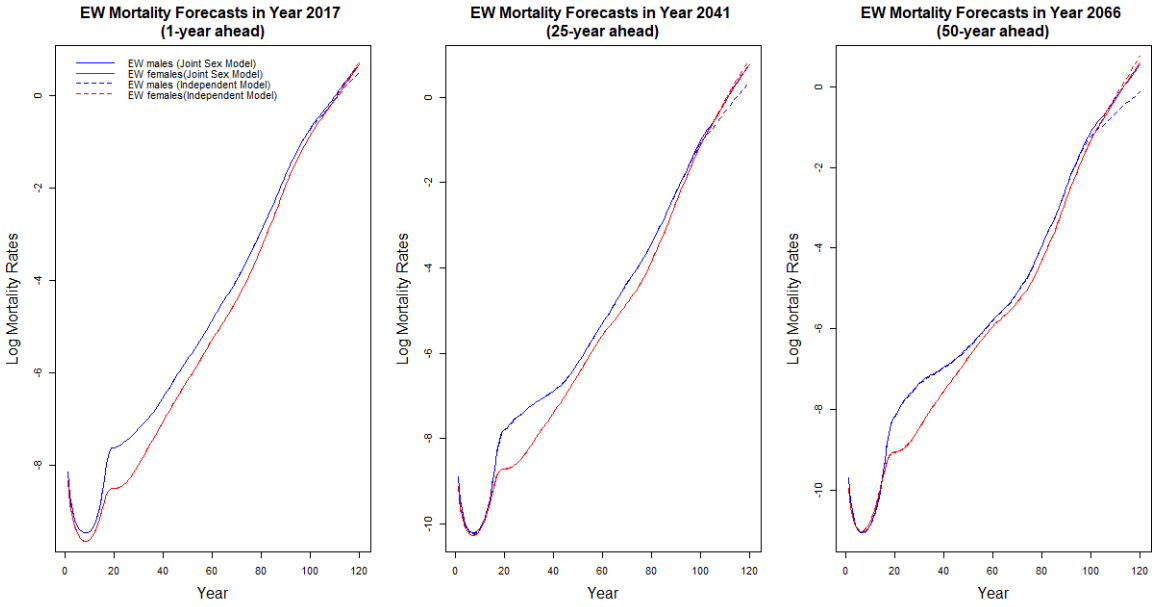


Figure 5.13: The projected mortality schedules in 1, 25 and 50 years ahead. The blue and red lines are the projections for male and female mortality rates respectively while the solid and dotted lines correspond to the projections of the joint and independent models respectively.

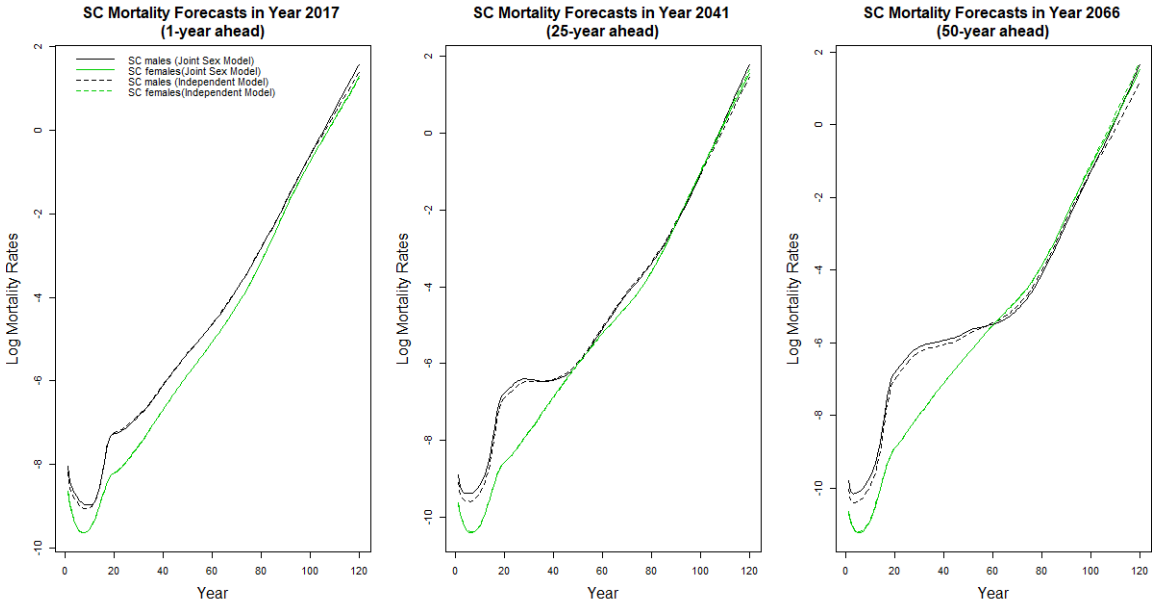
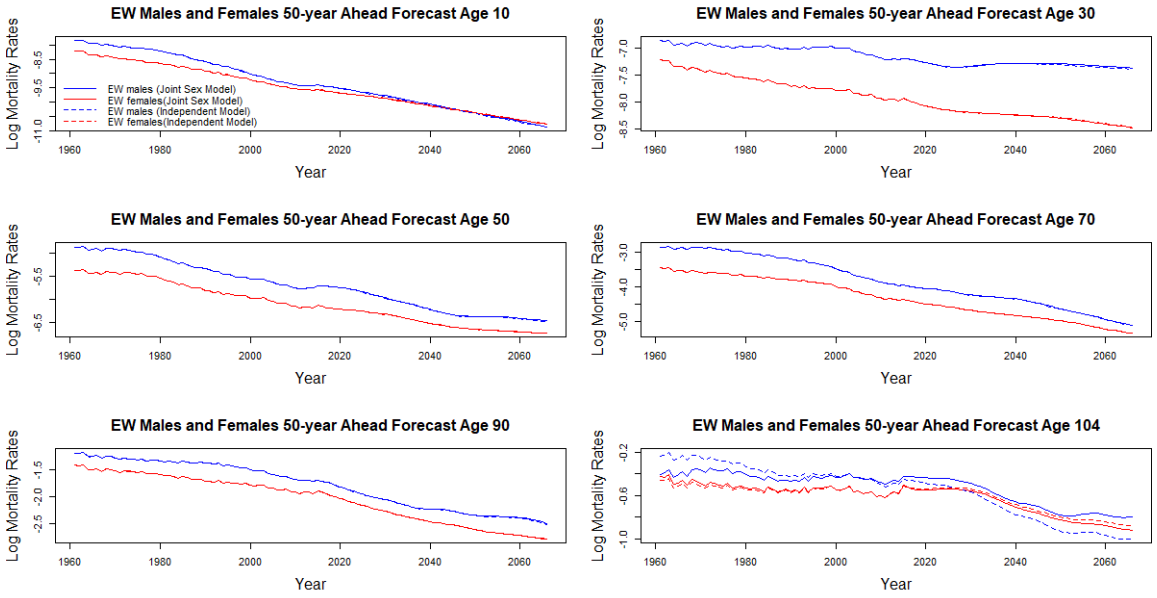


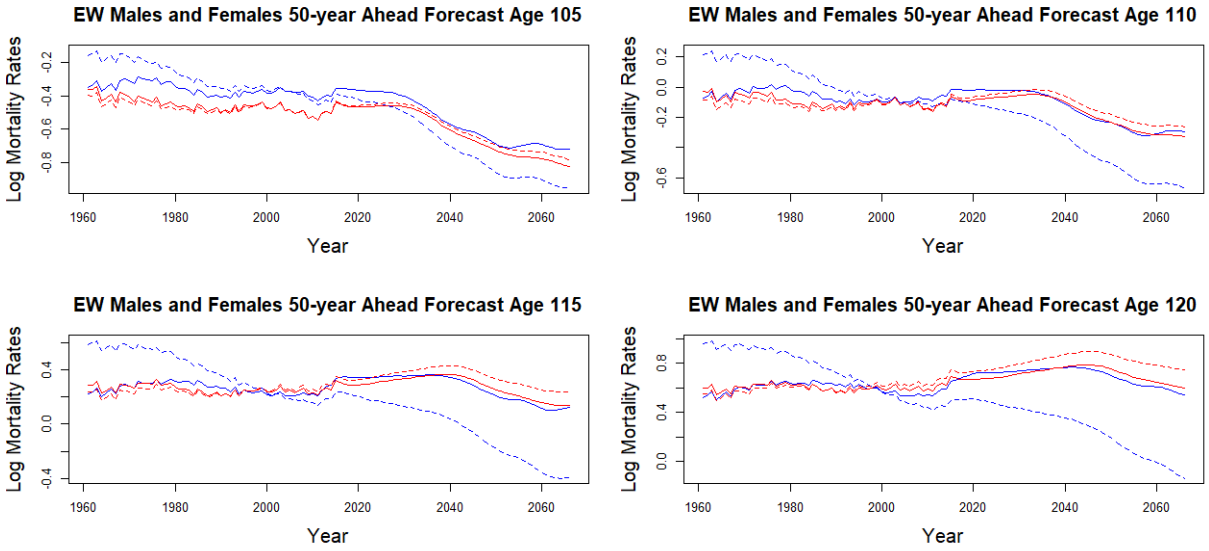
Figure 5.14: The project mortality schedules in 1, 25 and 50 years ahead. The black and green lines are the projections for male and female mortality rates respectively while the solid and dotted lines correspond to the projections of the joint and independent models respectively.

Figures 5.15a and 5.16a plot the 50-year ahead projected mortality rates at some selected ages for EW and SC respectively. At younger ages the differences between the two models are relatively small. At the oldest ages, the joint sex model produces much more reasonable long term projections compared to the independent model for EW males and females. For example, the EW male and female mortality projections at age 104 of the independent models intersect at

around 2030; whereas in the joint sex model the female mortality projections are always lower than that of males. By modelling male and female mortality rates jointly we avoided divergences and cross-overs of the mortality trends, especially at high ages where male and female mortality rates are very similar. For SC, the differences between the two models are relatively small due to the small estimated cross-sex penalty. It should also be noted that there are relatively less changes in the female projections compared to the male projections between the joint model and the independent model, reflecting the fact that the female population gives more information due to having bigger exposures. Figures 5.15b and 5.16b plot the estimated and projected mortality rates at the extrapolated ages. The joint model largely improves the plausibility of the projections. Even though the male and female projections in these extrapolated ages still cross-over in some years, the gap between them are much smaller compared to the independent models and they are non-divergent. For example, at age 110 the projected EW female mortality rate becomes higher than that of males in around year 2040 in the joint model, while in the independent model it happens in about year 2010. Note that the extrapolated mortality improvement rates after about age 115 are positive for SC males and females, meaning that mortality is worsening over time. Although we can not rule out this possibility, we do not have a strong reason to believe this, especially in the area where we have extrapolated. Therefore expert opinions play a more important role here for the smaller SC population.

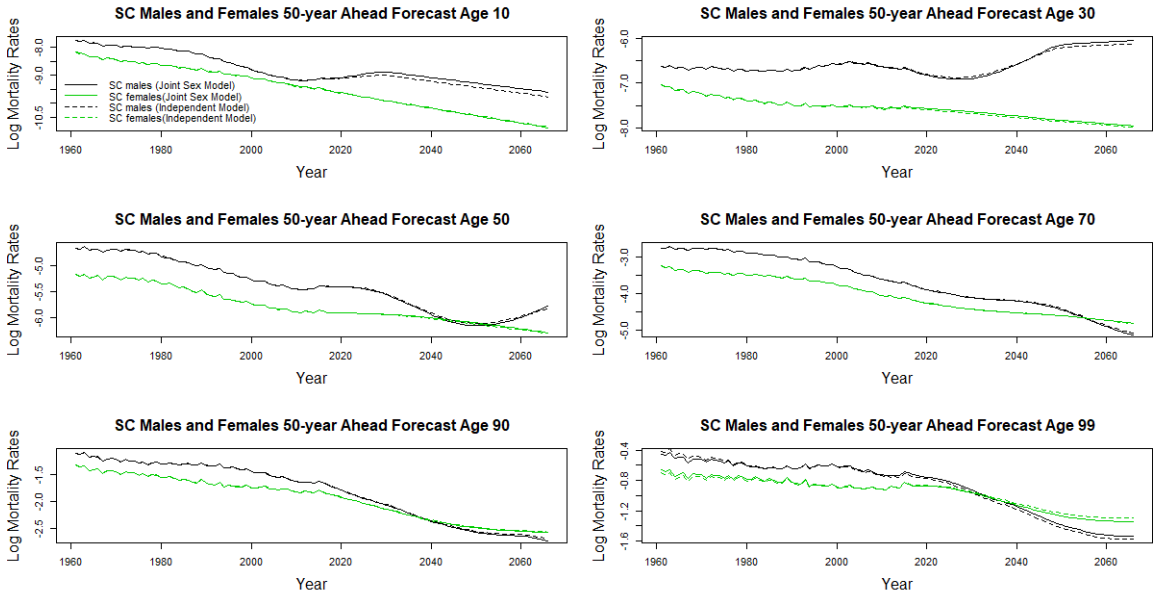


(a) age 10, 30, 50, 70, 90 and 104

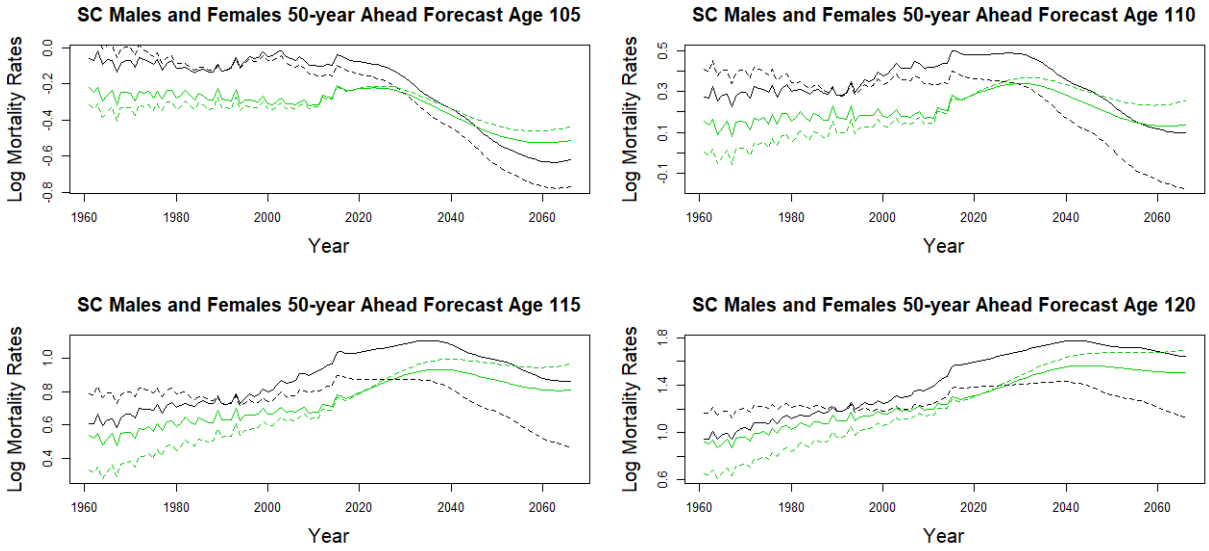


(b) age 105, 110, 115 and 120

Figure 5.15: 50 years ahead projections of EW male (blue) and female (red) mortality rates at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.



(a) age 10, 30, 50, 70, 90 and 99



(b) age 105, 110, 115 and 120

Figure 5.16: 50 years ahead projections of SC male (black) and female (green) mortality rates at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.

5.4.2.3 Expert Opinions

It is unrealistic to have perpetual constant mortality improvement rates based on the historical data. In our example, the age-specific mortality improvement rates are estimated using around 50 years of data, long term projections using these estimated improvement rates maybe unreasonable. To avoid risks of implausible projections, expert opinion can be incorporated into

mortality projections. The expert opinion moderates the effect of the model and model misspecification on long term forecasts and produces more robust and sensible estimates. The principal mortality assumptions used in the Office for National Statistics (ONS) for the 2016-based UK national population projections are that annual mortality improvement rates will converge to 1.2% for most ages in 2041, which is the 25-th year of the projections, and remains at 1.2% thereafter for both males and females. For the oldest ages there is little past evidence of the rates of improvement, therefore in order to avoid implausible numbers surviving to extreme ages, the target rates of improvement for those born before 1924 are assumed to decline from 1.1% to 0.1% for those born in 1906 and earlier. The transition from the current mortality improvement rates to the target rates is not assumed to happen linearly over the forecast horizon, but the speed is assumed to be the same for males and females. Following Dodd et al. (2020), the convergence to the target rates is incorporated using a weight function w that assigns more weight to the target rates over the forecast horizon,

$$w(t) = \begin{cases} 1 - 3(\frac{t}{25})^2 + 2(\frac{t}{25})^3 & \text{for } 0 \leq t \leq 25 \\ 0 & \text{for } t > 25, \end{cases} \tag{5.12}$$

and as a result the rate of improvement at age x in forecast year t equals to

$$\beta_{x,t} = \begin{cases} (1 - w(t))\beta_x^e + w(t)\hat{s}_\beta(x) & \text{for } 0 \leq t \leq 25 \\ \beta_x^e & \text{for } t > 25, \end{cases} \tag{5.13}$$

where β_x^e is the experts advised target rate of improvements at age x and $\hat{s}_\beta(x)$ is the estimated current mortality improvement rates at age x. Therefore in each forecast year the improvement rates are getting closer to the target rates and eventually reach the target improvement rates in the 25-th forecast year.

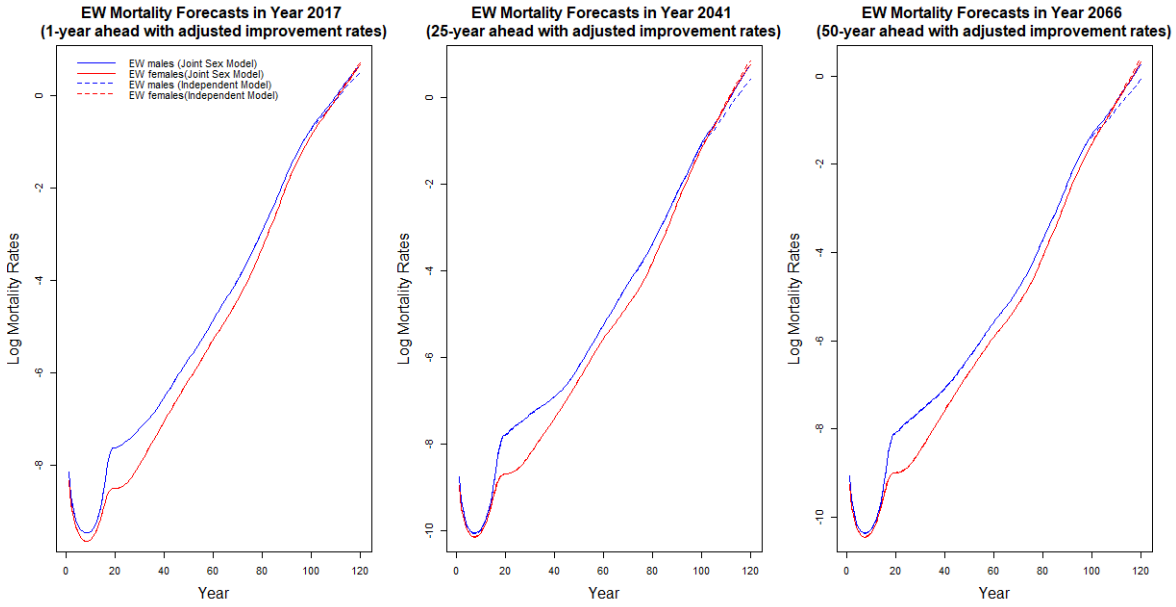
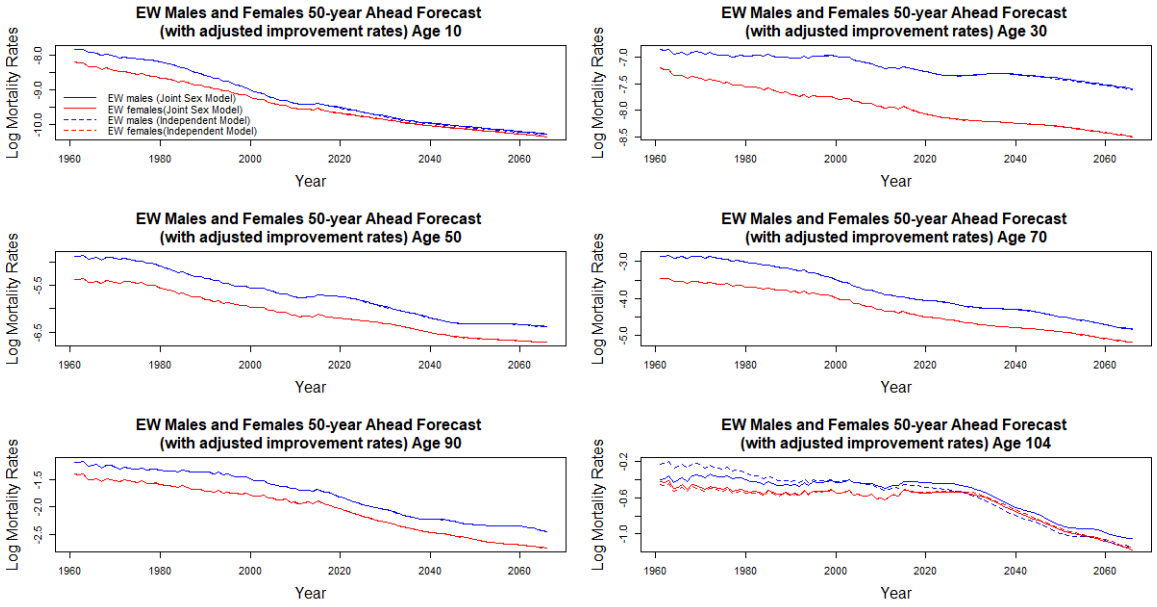
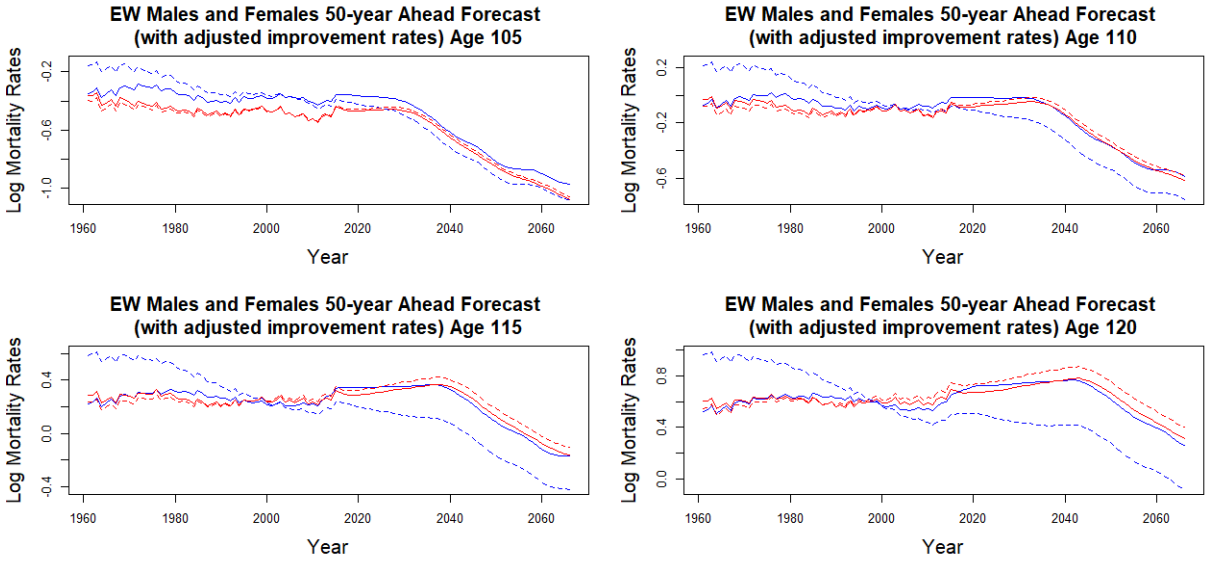


Figure 5.17: The project mortality schedules in 1, 25 and 50 years ahead with the incorporation of expert opinion. The blue and red lines are the projections for male and female mortality rates respectively while the solid and dotted lines correspond to the projections of the joint and independent models respectively.



(a) age 10, 30, 50, 70, 90 and 104



(b) age 105, 110, 115 and 120

Figure 5.18: 50 years ahead projections of EW male (blue) and female (red) mortality rates with the incorporation of expert opinion at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.

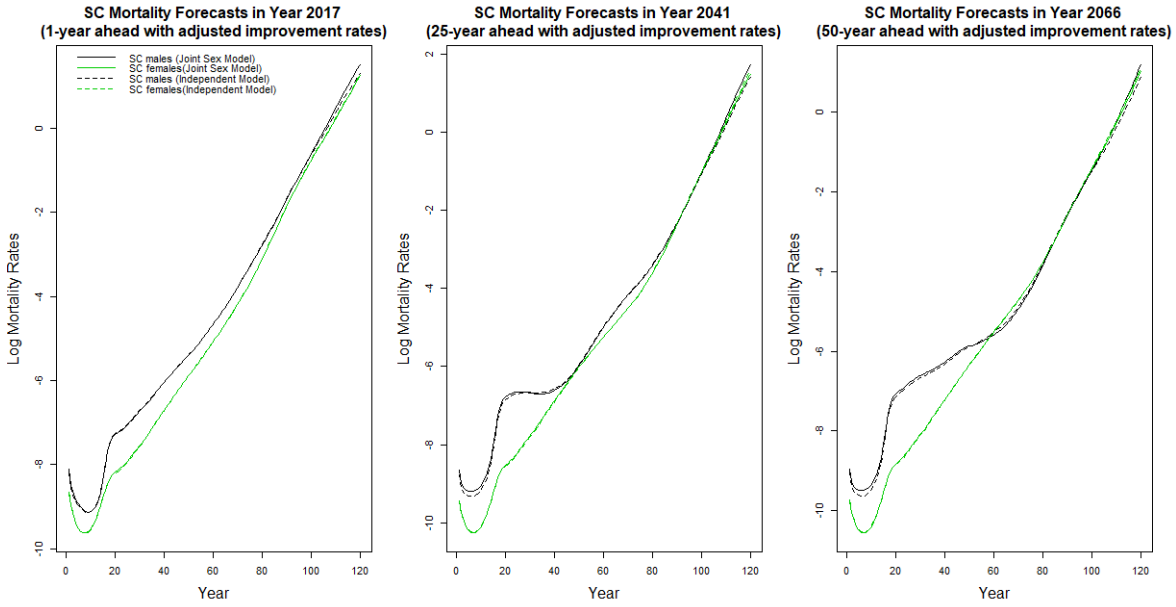
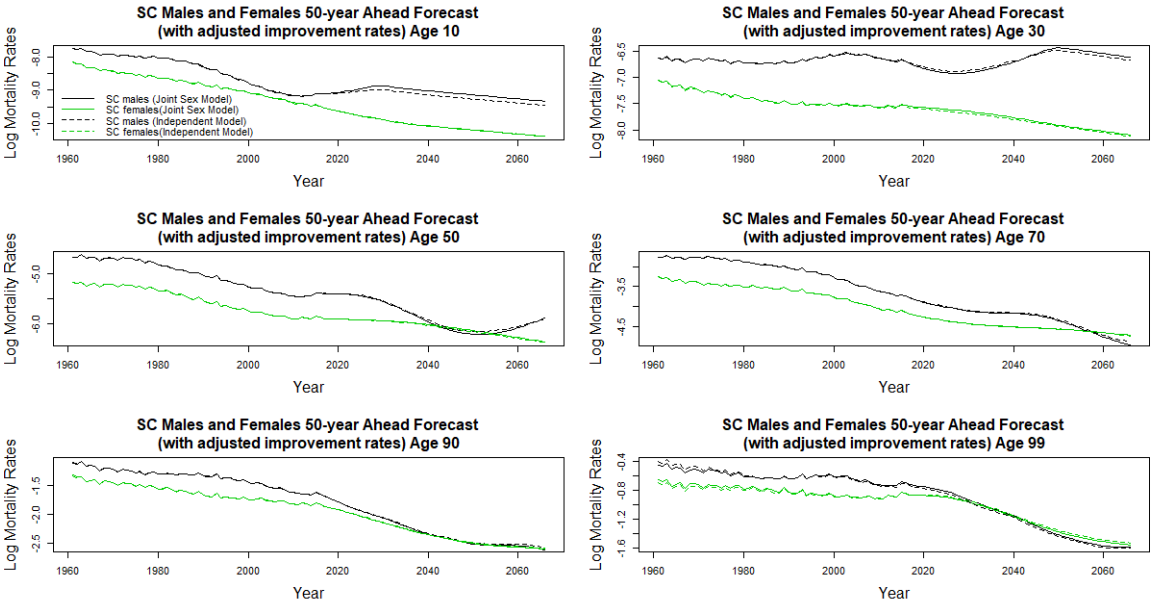
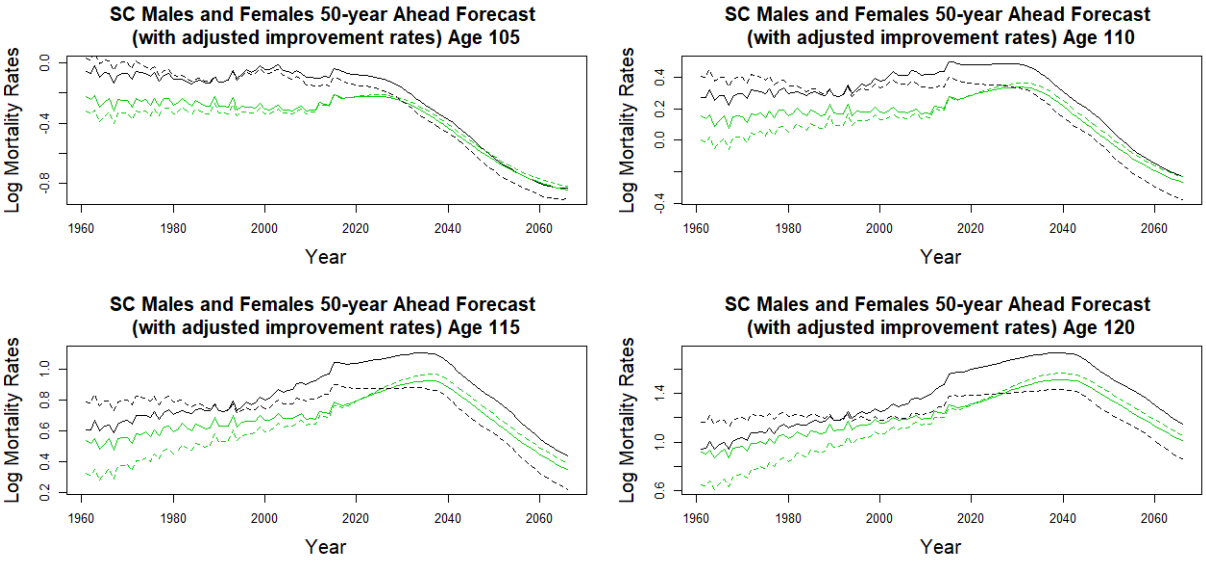


Figure 5.19: The projected mortality schedules in 1, 25 and 50 years ahead with the incorporation of expert opinions. The black and green lines are the projections for male and female mortality rates respectively while the solid and dotted lines correspond to the projections of the joint and independent models respectively.

Figures 5.17 and 5.19 plot the projected mortality profiles in the 1-st, 25-th and 50-th forecast year for EW and SC respectively while Figures 5.18 and 5.20 plot the 50-year ahead projections at some selected ages and the extrapolated ages with expert opinions adjustments. As expected, the mortality projections are now non-divergent in the long run as the improvement rates eventually tend to a common value. Incorporating expert opinion largely improves the plausibility of the long term mortality forecasts. Specifically, when the current mortality improvement rates are continued indefinitely, the male and female childhood mortality rates eventually cross-over as depicted in Figure 5.13, while after the incorporation of expert opinion, this is effectively avoided. The SC male mortality rates still drop below SC female mortality rates at some ages in the long term, but the gaps are much smaller and with fewer ages experiencing this problem after incorporating expert opinion.



(a) age 10, 30, 50, 70, 90 and 99



(b) age 105, 110, 115 and 120

Figure 5.20: 50 years ahead projections of SC male (black) and female (green) mortality rates with the incorporation of expert opinion at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint sex model and the independent models respectively.

5.4.3 Joint Country Models

In this subsection we present the result of the joint country model for males and females. It is evident and reasonable that the EW and SC males (females) have similar mortality structures. Having a wider age range of available data in England and Wales, it would be beneficial to jointly model EW males (females) with SC males (females), which has data only up to age 99.

The penalties employed for the joint country model of England and Wales' males (females) and Scotland's males (females) are given in Table 5.5.

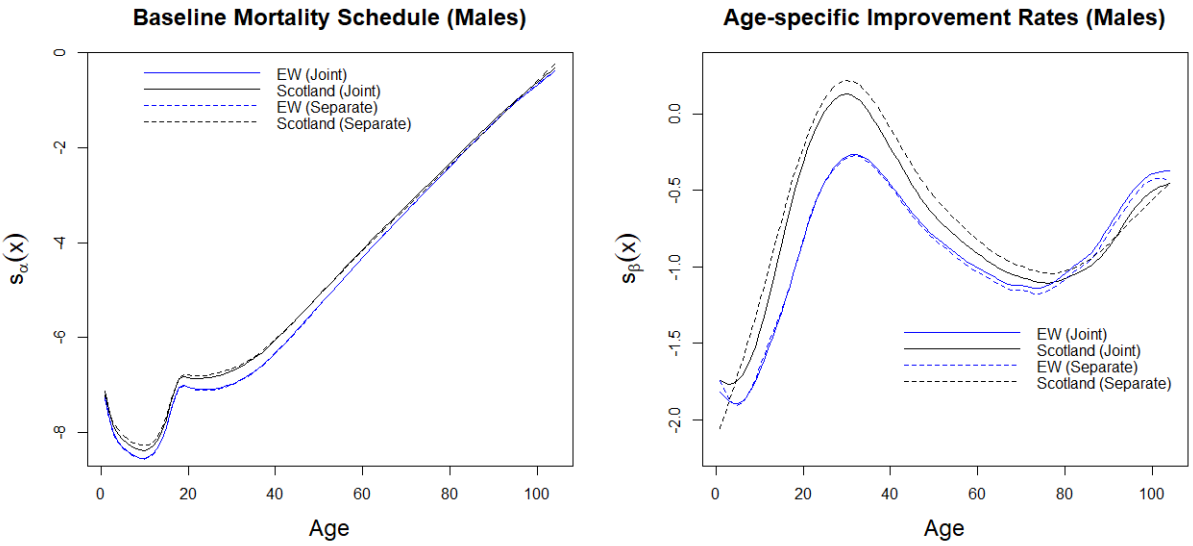
Function	Smoothness Penalty	Cross-country Penalty
$s_\alpha(x)$	$\sum e^{\lambda_1^{c,\alpha}+\lambda_2^{c,\alpha}i}(\bigtriangledown^2\alpha_i^{EW})^2$ + $\sum e^{\lambda_1^{c,\alpha}+\lambda_2^{c,\alpha}i}(\bigtriangledown^2\alpha_i^{SC})^2$	$\sum \lambda^{s,\alpha} \bigtriangledown (\alpha_i^{EW} - \alpha_i^{SC})^2$
$s_\beta(x)$	$\sum \lambda^{c,\beta}(\bigtriangledown^2\beta_i^{EW})^2$ + $\sum \lambda^{c,\beta}(\bigtriangledown^2\beta_i^{SC})^2$	$\sum \lambda^{s,\beta} \bigtriangledown (\beta_i^{EW} - \beta_i^{SC})^2$
$s_\gamma(t-x)$	$\sum \lambda^{c,\gamma}(\bigtriangledown^2\gamma_i^{EW})^2$ + $\sum \lambda^{c,\gamma}(\bigtriangledown^2\gamma_i^{SC})^2$	

Table 5.5: The penalties for the Joint Country Model

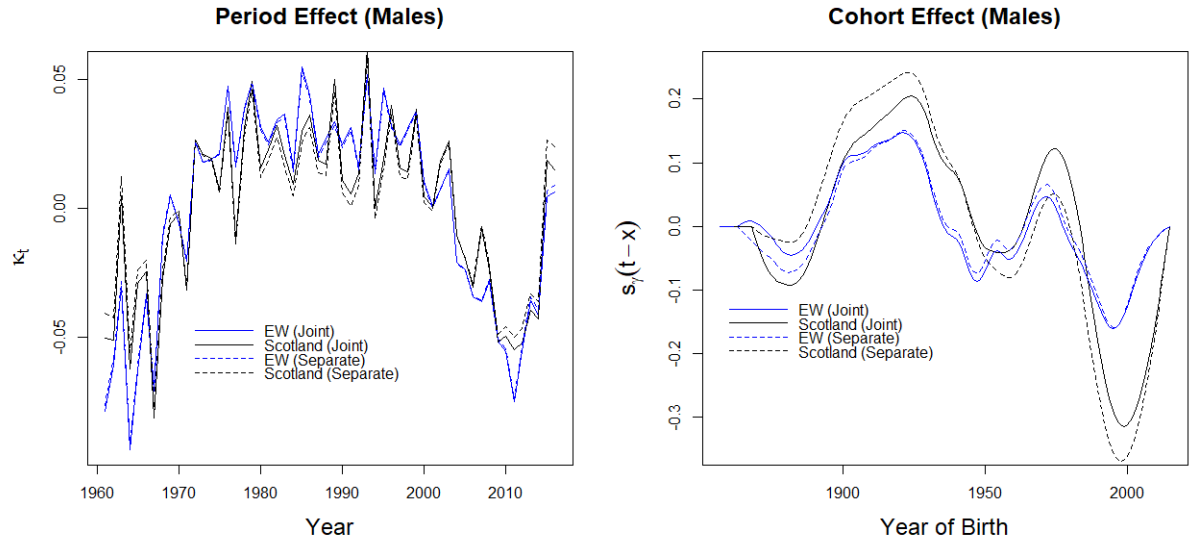
Here the superscript c indicates that the smoothing parameters are common between EW and SC.

Figures 5.21 and 5.22 plot the estimated terms of the joint country model alongside that of the independent models. Comparing the joint country model and independent models, we can see that information is shared between the England and Wales and Scotland's populations. Being the larger population, the estimates of the EW populations are less affected by the cross-country penalties. The shapes of the baseline mortality profile and improvement rates of SC males and females are now more similar to that of the EW populations. Despite the mortality improvement rates for SC adult males around age 30 are still above zero (worsening mortality), it is less extreme than those obtained in the independent model. The estimated cross-country shape penalty in the improvement rates (5.9) for females is much stronger than that for males, hence the shape of the mortality improvement rates of the smaller SC female population is very close to that of the larger EW female population. The period effects do not show significant differences between the two models but the cohort effect for SC females display more variability.

Figures 5.23 and 5.24 plot the extrapolated baseline mortality profiles and improvement rates up to age 120. Instead of linearly projecting SC baseline mortality after age 99 as in the single population model, information is learnt from the EW data about the shape of the mortality schedule at higher ages and the extrapolated SC baseline mortality schedules now show a deceleration in mortality increase in age, a feature that is well established in the literature (Carriere, 1992; Thatcher, 1999; Saikia and Borah, 2014; Pitacco, 2016). The shape of the improvement rates are now more in line with each other even in the extrapolation range.

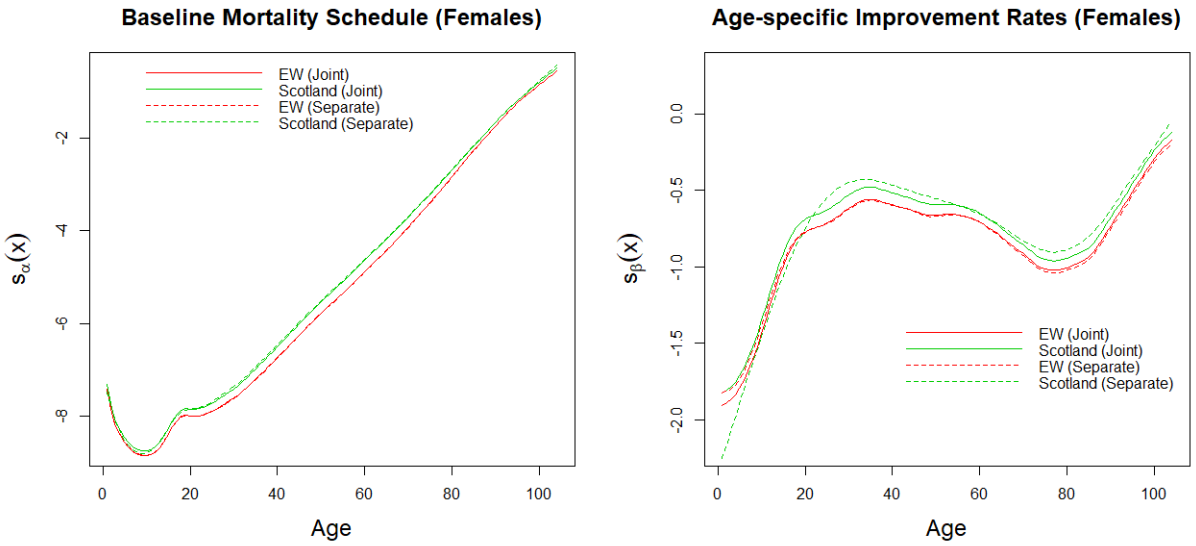


(a) estimated baseline mortality schedules and age-specific improvement rates

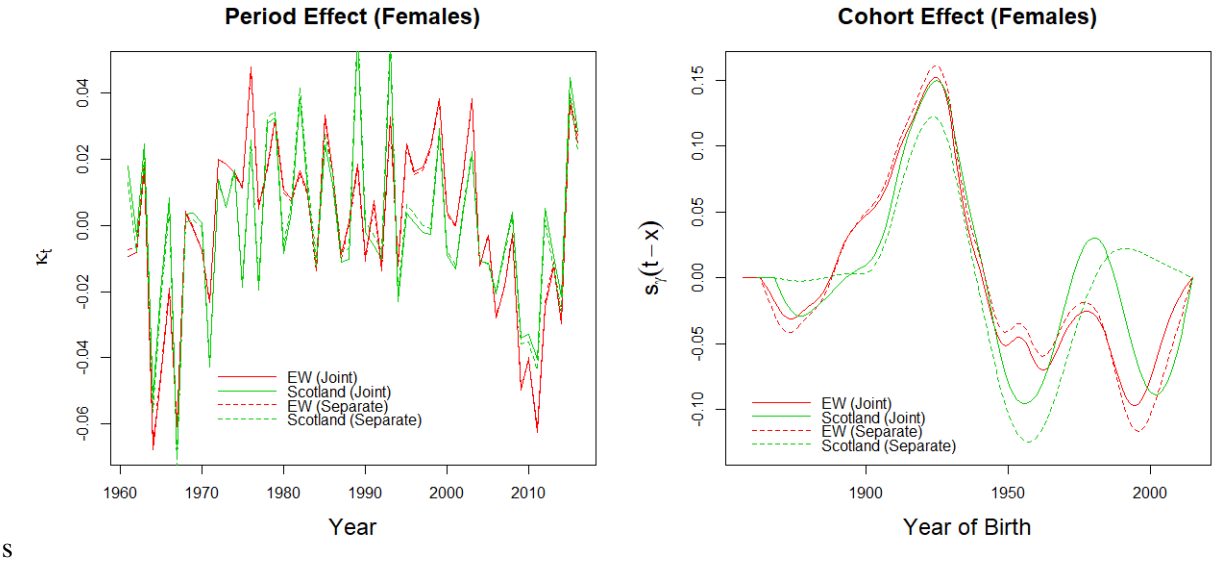


(b) estimated period and cohort effects

Figure 5.21: The parameter estimates of the joint country model for EW (blue) and SC (black) males. The solid and dotted lines correspond to estimates of the joint country model and the independent models respectively.



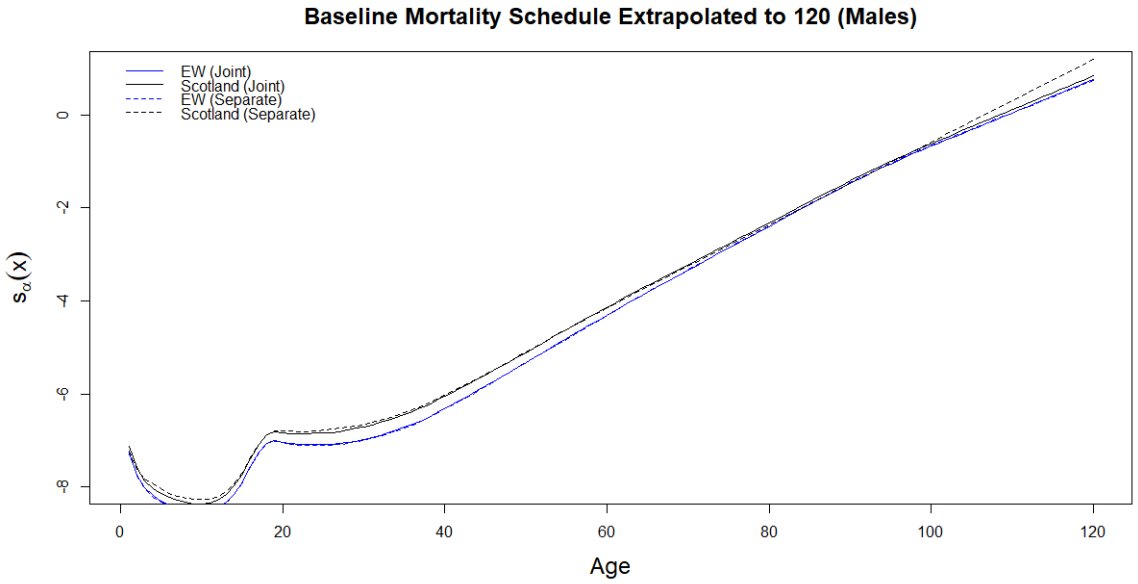
(a) estimated baseline mortality schedules and age-specific improvement rates



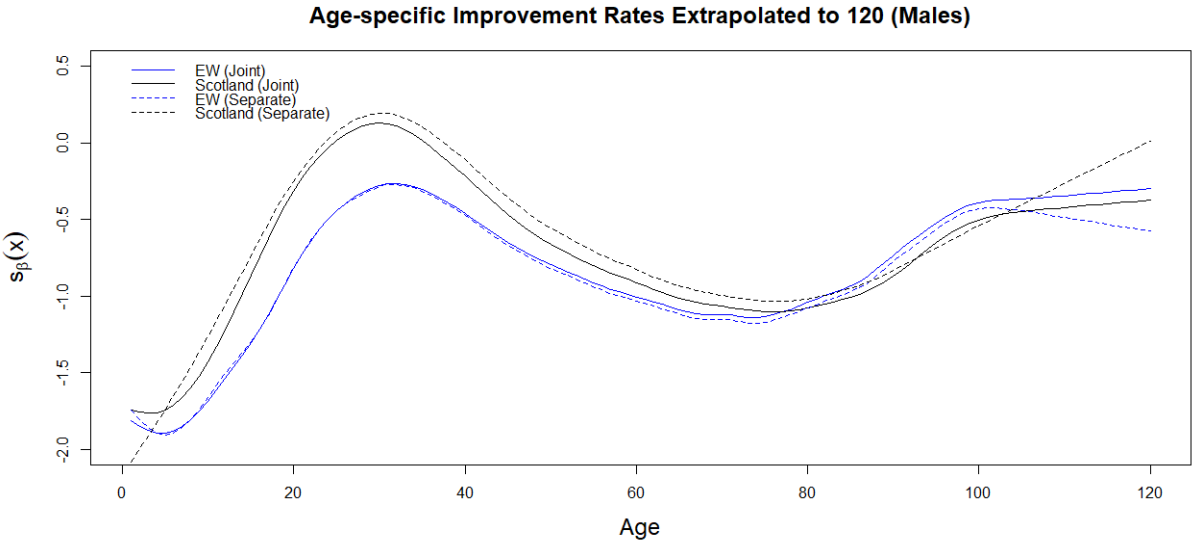
s

(b) estimated period and cohort effects

Figure 5.22: The parameter estimates of the joint country model for EW (red) and SC (green) females. The solid and dotted lines correspond to the estimates of the joint country model and independent models respectively.

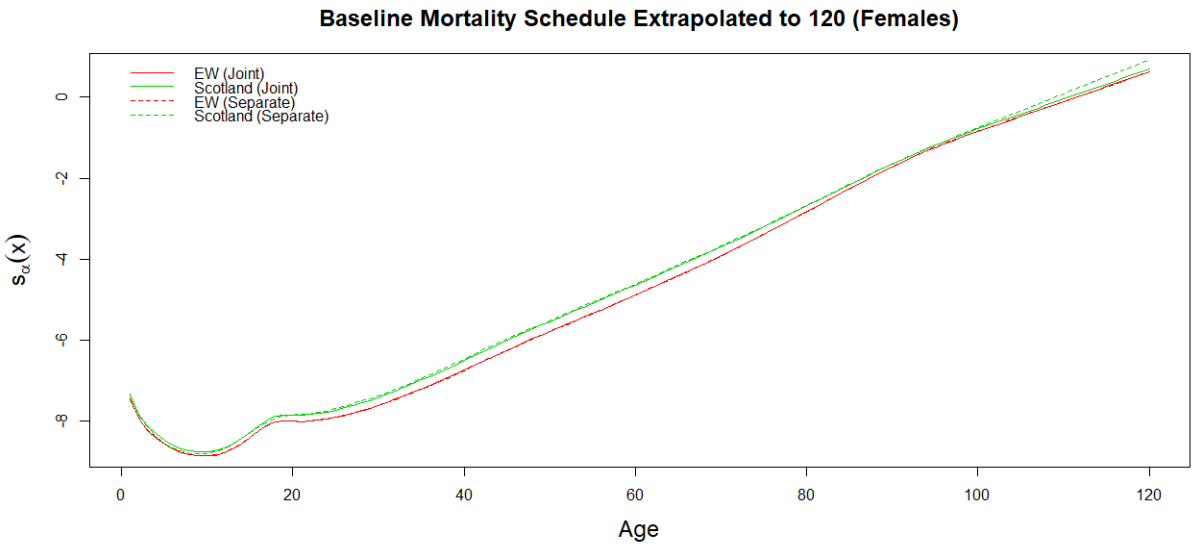


(a) estimated baseline mortality schedules extrapolated to age 120

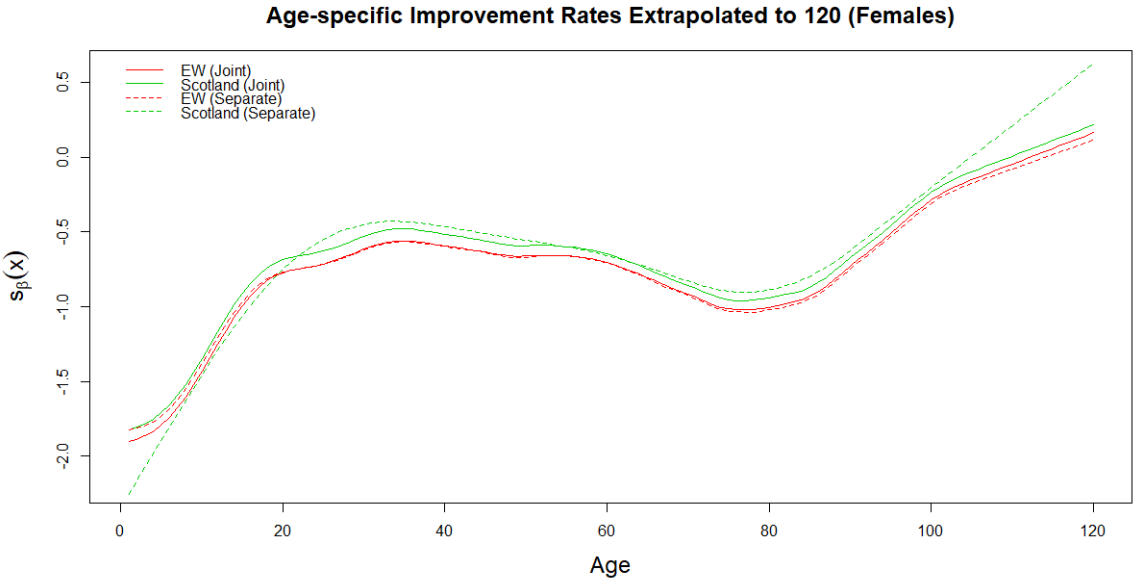


(b) estimated age-specific improvement rates extrapolated to age 120

Figure 5.23: The extrapolated baseline mortality schedules and age-specific improvement rates of the joint country model for EW (blue) and SC (black) males. The solid and dotted lines correspond to estimates of the joint country model and the independent models respectively.



(a) estimated baseline mortality schedules extrapolated to age 120



(b) estimated age-specific improvement rates extrapolated to age 120

Figure 5.24: The extrapolated baseline mortality schedule and age-specific improvement rates for EW (red) and SC (green) females. The solid and dotted lines correspond to estimates of the joint country model and the independent models respectively.

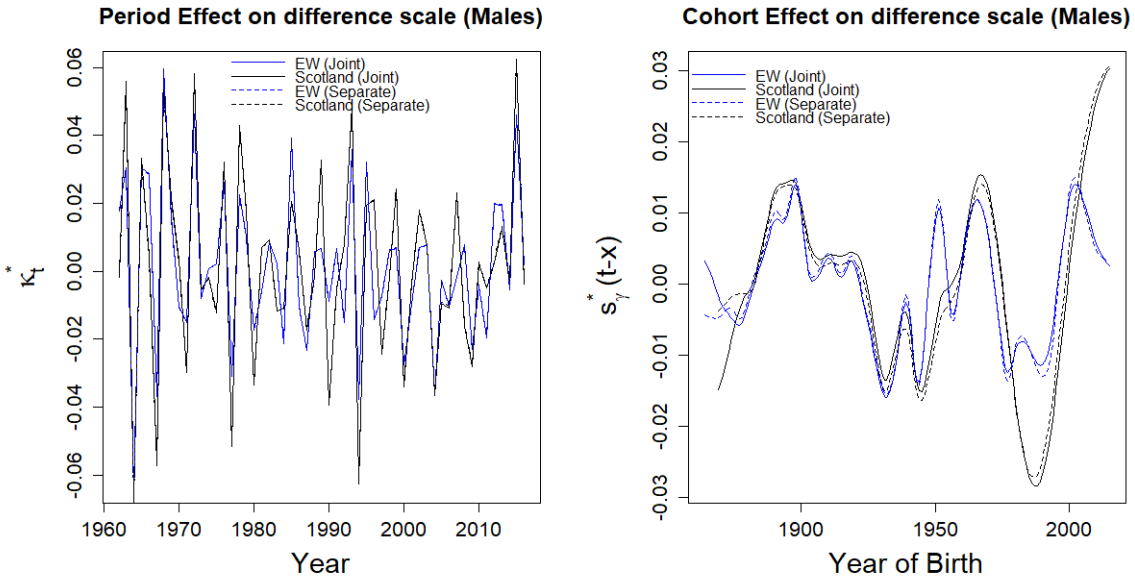


Figure 5.25: The estimated period and cohort effects for EW (blue) and SC (black) males on a difference scale. The solid and dotted lines correspond to estimates of the joint country model and the independent models respectively

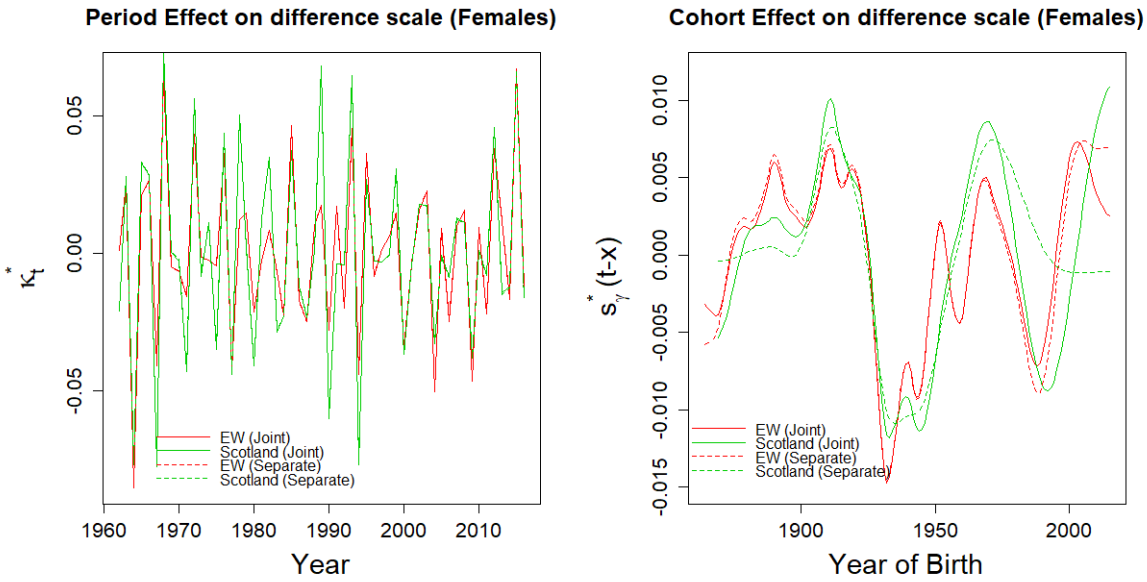


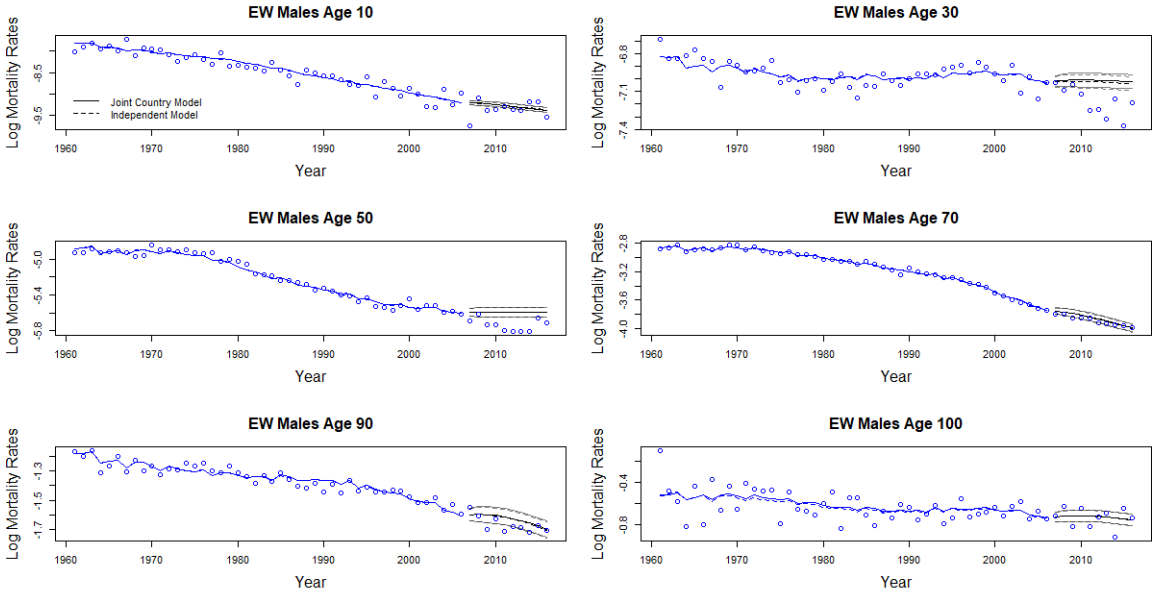
Figure 5.26: The estimated period and cohort effects for EW (red) and SC (green) females on a difference scale. The solid and dotted lines correspond to estimates of the joint country model and the independent models respectively.

Figures 5.25 and 5.26 plot the period and cohort effects on the difference scale. The κ_t^*'s oscillate around 0 evenly. It can also be noted that the κ_t^*'s between EW and SC populations seem to be correlated. The κ_t^*'s between the joint and independent model are indistinguishable.

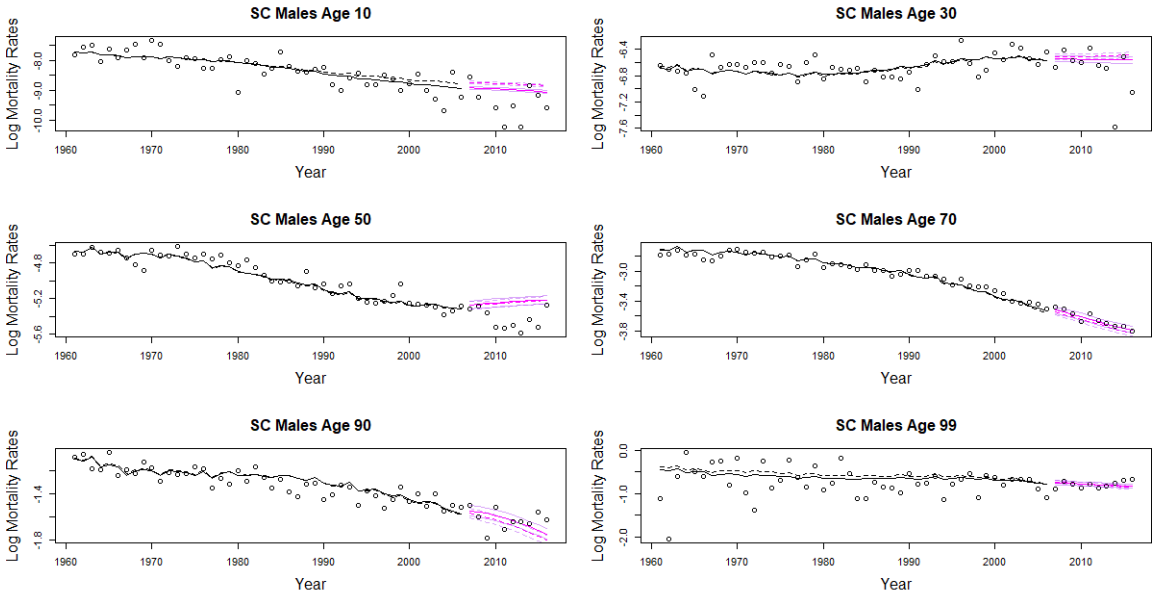
The differences between the $s_\gamma^*(t-x)$'s of the joint and independent models are relatively small, except for SC females. Three pronounced troughs can be observed for cohorts born about 1931, 1944 and 1987 as in the joint sex model, meaning that these cohorts enjoy a faster mortality improvement rates. The generational effect for cohorts born about 1931 is well acknowledged in the literature (Willets, 2004; Richards et al., 2006; Murphy, 2009), and generational effect for the cohorts born about 1944 is also noted in Willets (2004). In the independent model for SC females and the joint sex model for SC males and females, only a single broad trough is observed for cohorts born around 1925 to 1950, whereas in the joint country model two troughs for cohorts born around 1932 and 1945 are observed. In addition, SC female born around 1987 is also identified as having faster improvement rates. This result of faster improvement for SC females born around 1987 is consistent with our results for EW males, EW females and SC males. Note that there is no cross-country penalty for the cohort effects, this is merely a consequence of the cross-country penalty on the baseline mortality schedules and the age-specific improvement rates.

5.4.3.1 Backtesting

Figures 5.27 and Figure 5.28 plot the fitted and projected mortality rates for males and females respectively at some selected ages from 1961 to 2016. The black solid and dotted lines correspond to projections produced by the joint and independent models respectively and the MAE are shown in Table 5.6 and 5.7. The MAE of the joint model in older ages groups (61+) are lower than that of the independent model for both males and females, indicating that the cross-country penalty does indeed produces more accurate predictions at older ages, as information is now shared between populations at these ages where the exposures are relatively small. However the penalty may be too heavy and restrictive on the freedom of the model at younger ages, where data is more abundant. One possible modification would be to use an exponentially varying cross-country penalty instead of a global one. Similar conclusions can be drawn when an ARIMA(1,1,0) is used (Figures 5.29 and 5.30). Table 5.8 shows the coverage of the estimated 95% intervals of the joint country model. On average, the intervals produced using AR(1) are too narrow. On the other hand, ARIMA(1,1,0) produces wider intervals and hence higher coverage. The overall coverage of the intervals using ARIMA(1,1,0) are still below 95%. However, for older age groups, the coverage are much closer to 95%, especially for EW males aged 61-90. The coverage of the intervals produced by the joint country model for SC males have improved compared to the independent model, especially at the older age groups (61+), while for EW males, EW females and SC females, the coverage are quite similar to that of the independent models.

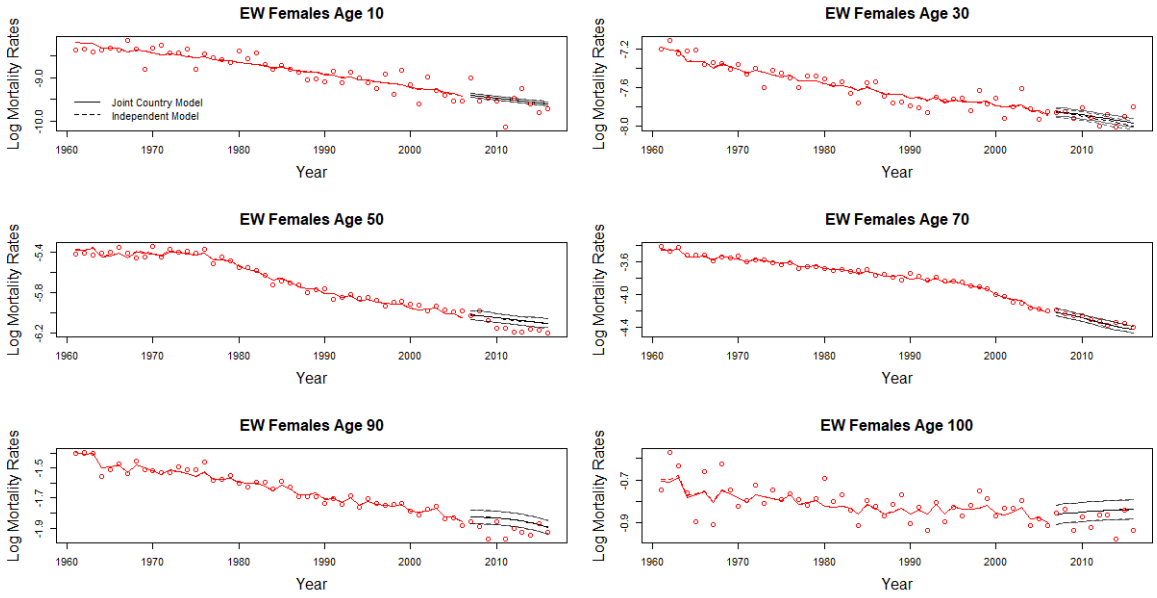


(a) EW males

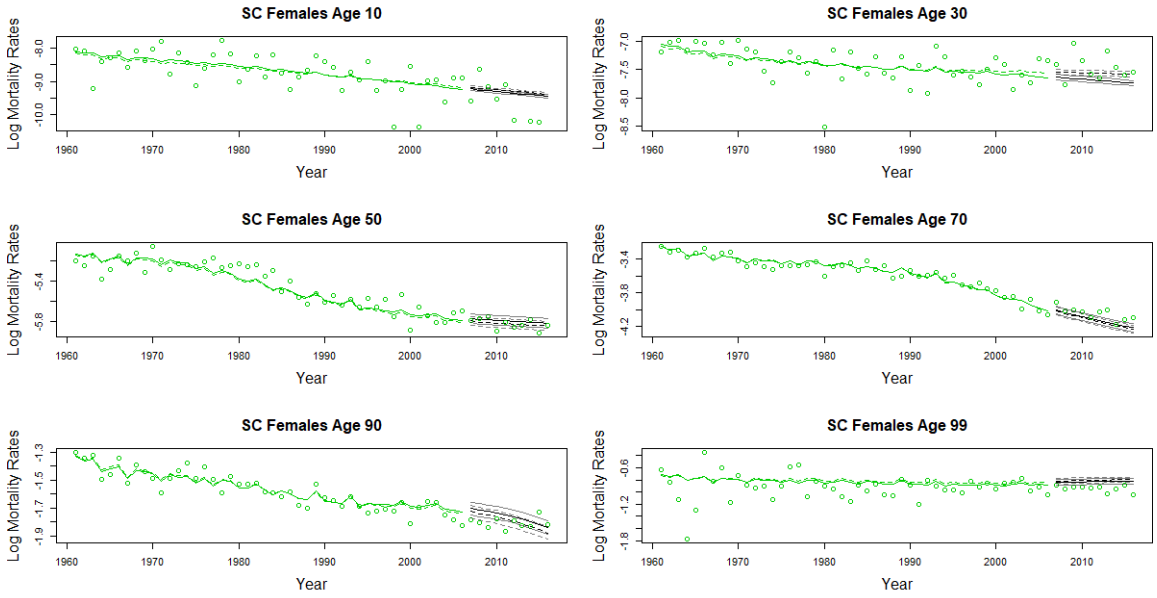


(b) SC males

Figure 5.27: Estimated mortality rates and the project 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an AR(1). The points are the observed mortality rates.

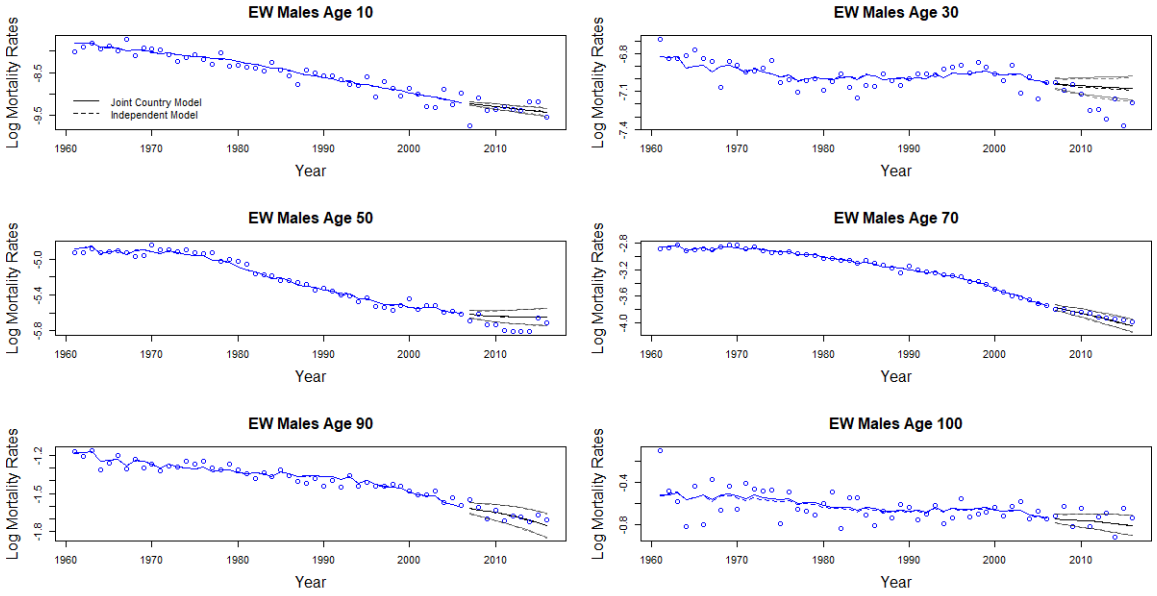


(a) EW females

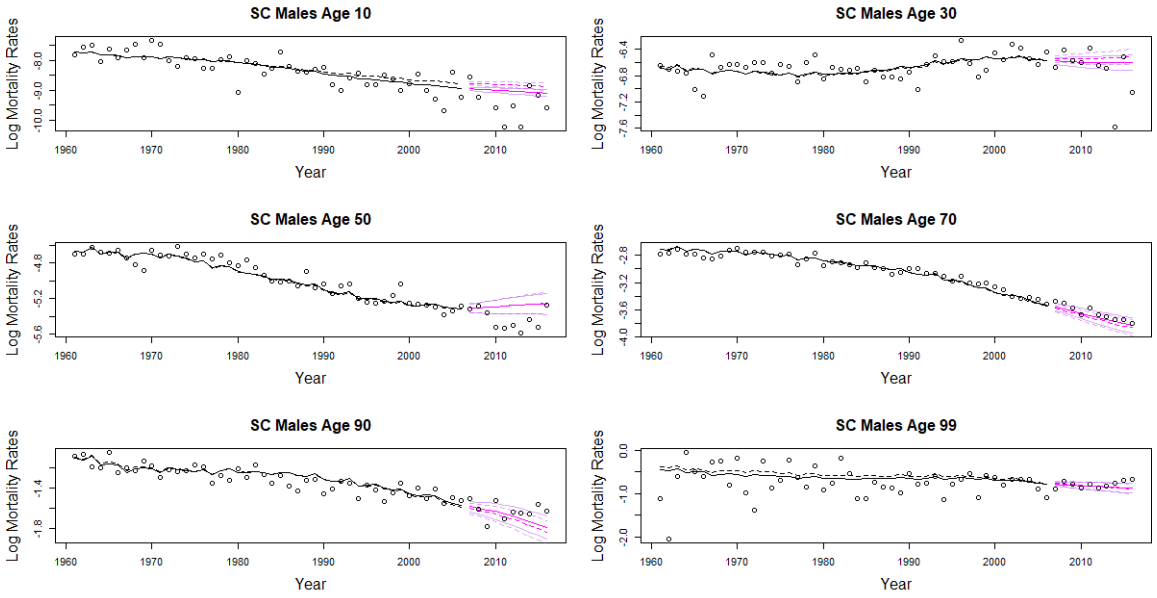


(b) SC females

Figure 5.28: Estimated mortality rates and the project 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an AR(1). The points are the observed mortality rates.

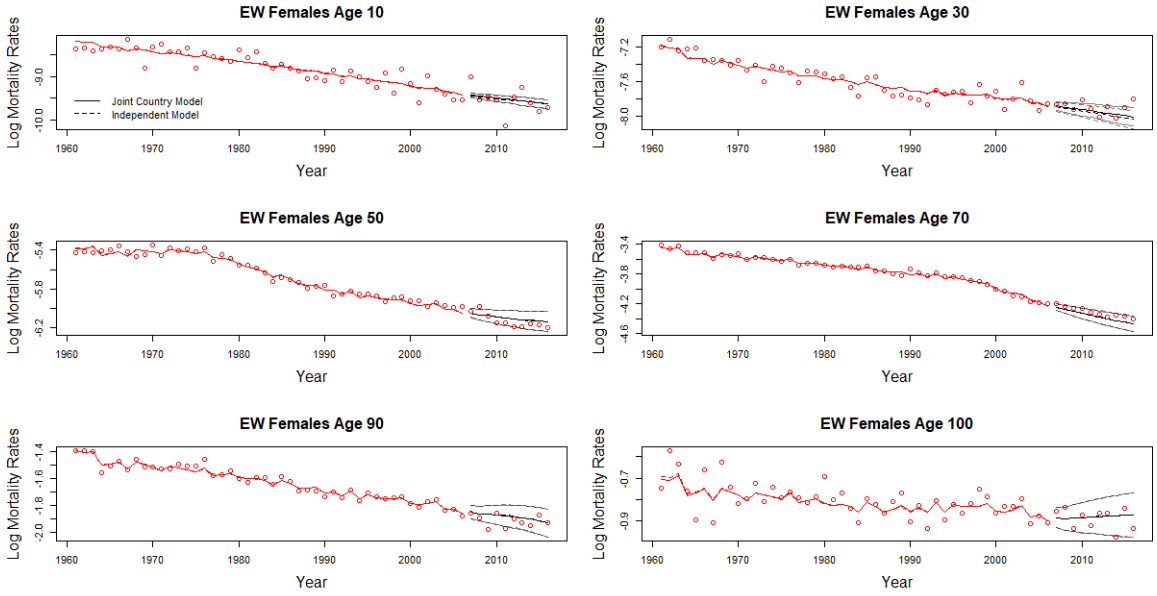


(a) EW males

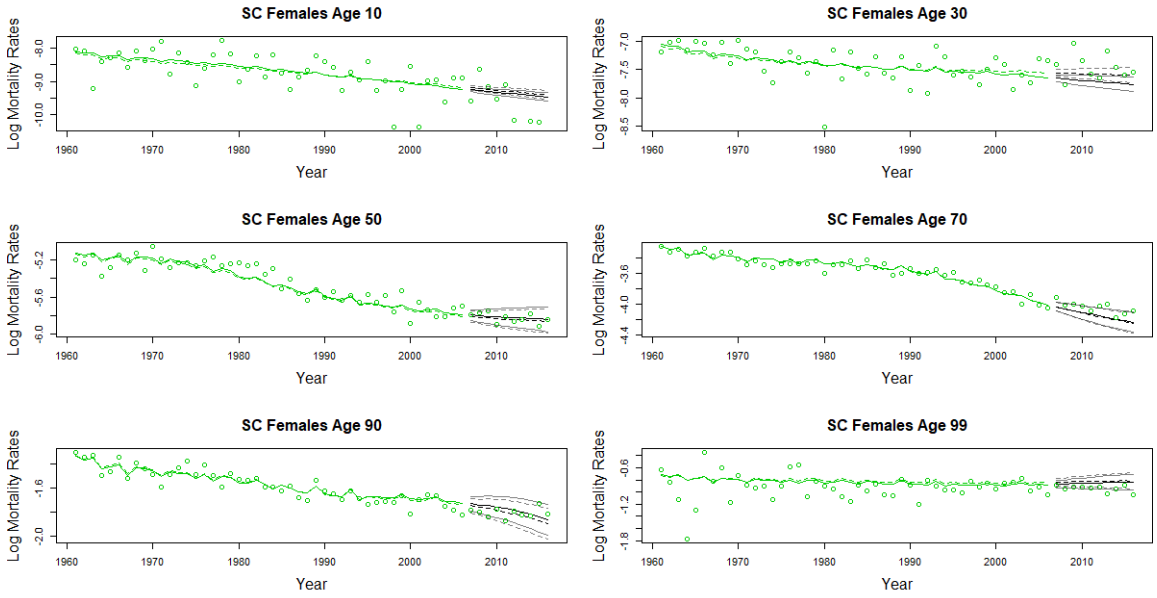


(b) SC males

Figure 5.29: Estimated mortality rates and the project 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an ARIMA(1,1,0). The points are the observed mortality rates.



(a) EW females



(b) SC females

Figure 5.30: Estimated mortality rates and the project 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an ARIMA(1,1,0). The points are the observed mortality rates.

	EW and SC males			
Age	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	1.357501	1.311118	1.365819	1.314342
31 to 60	0.7154661	0.7108316	0.7178916	0.7090071
61 to 90	0.4636356	0.4815504	0.4790211	0.4931601
91+	0.1747451	0.1835653	0.1840951	0.1913881
all ages	0.7695336	0.7607488	0.7783156	0.7654801

Table 5.6: The mean absolute error of log projected mortality rates for EW and SC males

	EW and SC females			
Age	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	1.720643	1.687815	1.729588	1.695103
31 to 60	0.9545456	0.9500895	0.9583566	0.9510878
61 to 90	0.5587298	0.566814	0.5691109	0.5742768
91+	0.1201981	0.1269547	0.1147334	0.1196441
all ages	0.9694566	0.9615916	0.9756758	0.9654181

Table 5.7: The mean absolute error of log projected mortality rates for EW and SC females AR(1)

	95% intervals coverage							
Age	EW males		EW females		SC males		SC females	
	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)
all ages	0.4567308	0.6048077	0.4634615	0.7721154	0.369696982	0.497979782	0.280808073	0.516161609
61-90	0.7633333	0.9566667	0.6166667	0.9033333	0.6	0.7333333	0.4333333	0.71
91+	0.65	0.7357143	0.45	0.9071429	0.4777778	0.5555556	0.1777778	0.7666667

Table 5.8: The coverage of the 95% intervals of the joint country model

5.4.3.2 Projections

Figures 5.31 and 5.32 plot the projected mortality schedules of males and females in the 1-st, 25-th and 50-th forecast year. At the oldest ages the shape of the SC mortality profiles ARE now more similar to that of Ew. The young-adulthood humps for SC females are also more pronounced under the joint model. Looking at the observed Scottish female mortality rates in the most recent years in Figure 5.33, the crude mortality rates of the teenage and young adulthood display a big uncertainty and it is difficult to conclude This is possibly due to the sparse data at these ages are so dispersed and it is difficult to conclude which model has a more appropriate illustration.

Figures 5.34 and 5.35 plot the 50-year ahead mortality projections of EW and SC male and female mortality rates. The joint models produce non-divergent projections within the 50-year forecast window, the main difference being that at high ages the two mortality trends are closer to each other. This is reasonable as divergences in mortality forecasts are usually criticised based on historical data and the general understanding of a global mortality convergence.

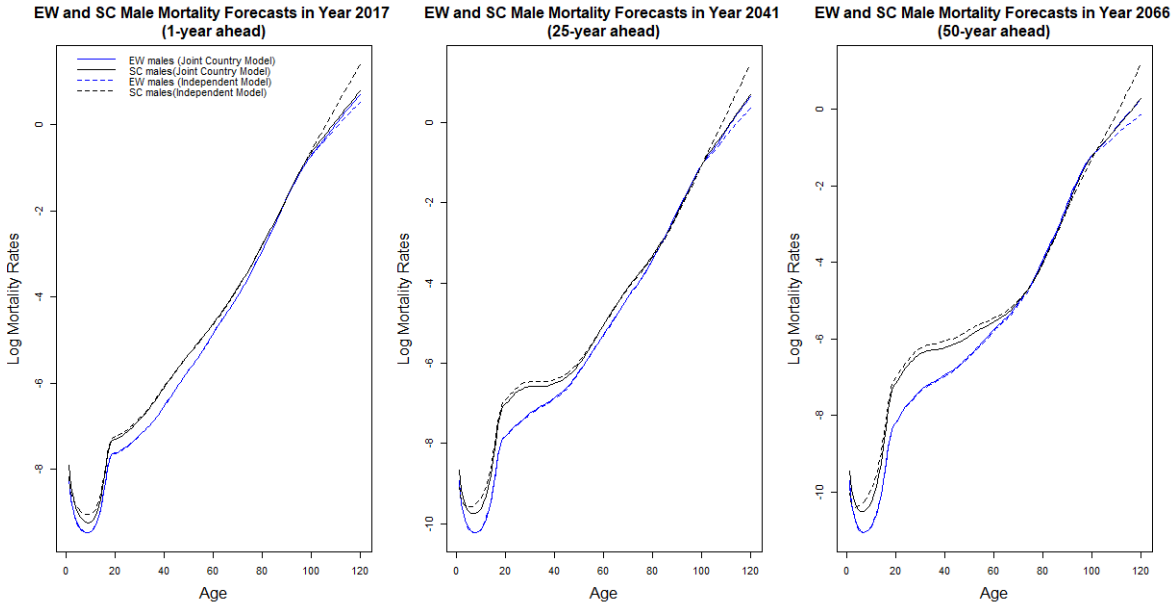


Figure 5.31: The projected mortality schedules in 1, 25 and 50 years ahead. The blue and black lines are the projections for EW and SC male mortality rates respectively while the solid and dotted lines correspond to the projections based on the joint and independent models respectively.

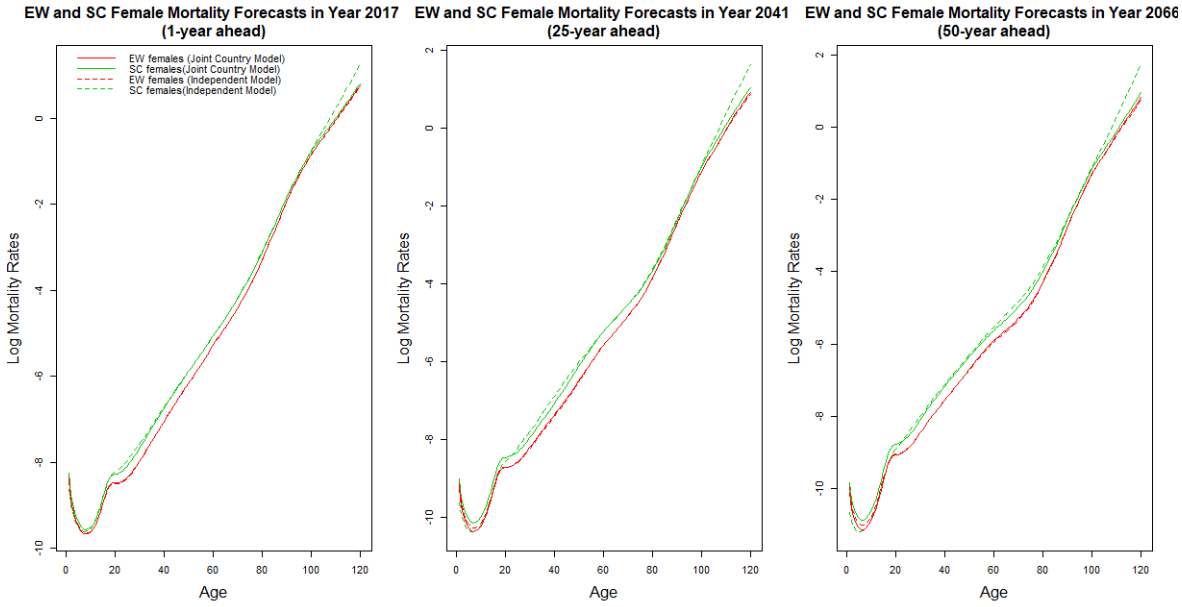
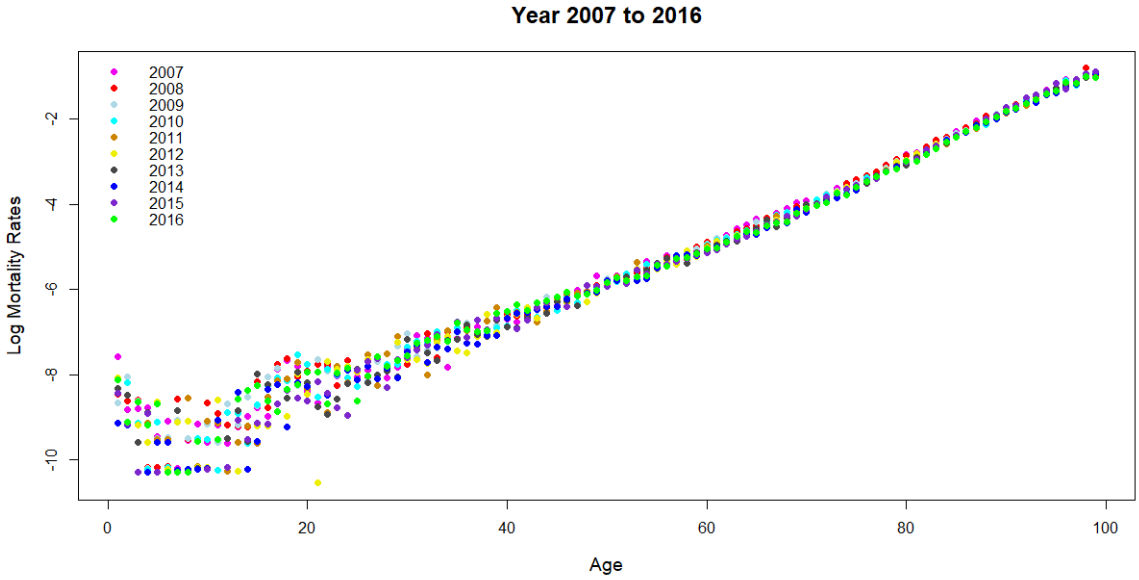
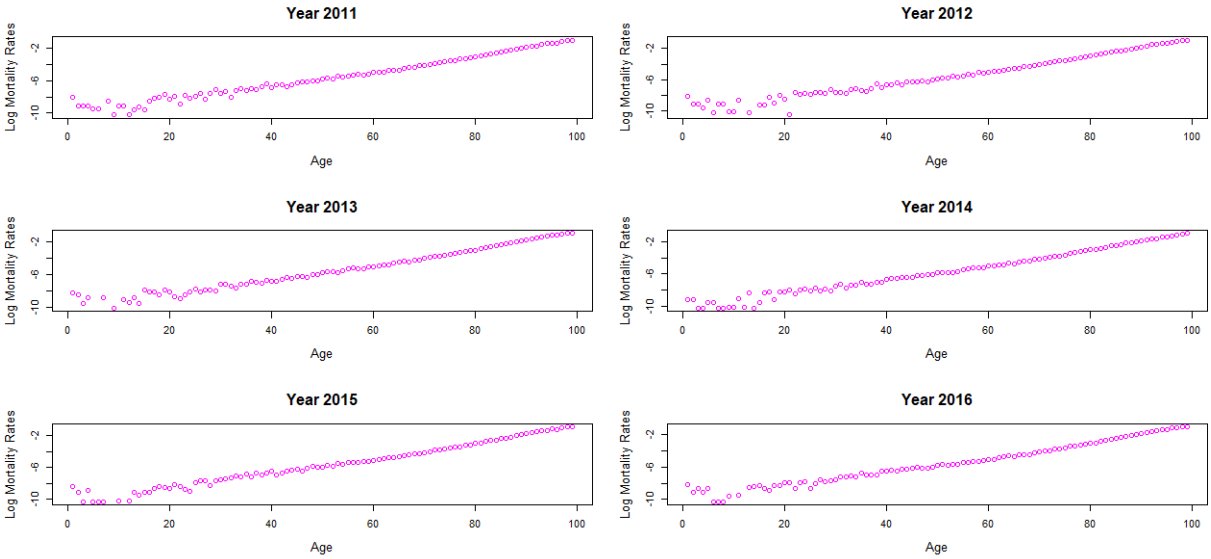


Figure 5.32: The projected mortality schedules in 1, 25 and 50 years ahead. The red and green lines are the projections for EW and SC female mortality rates respectively while the solid and dotted lines correspond to the projections of the joint country model and the independent models respectively.

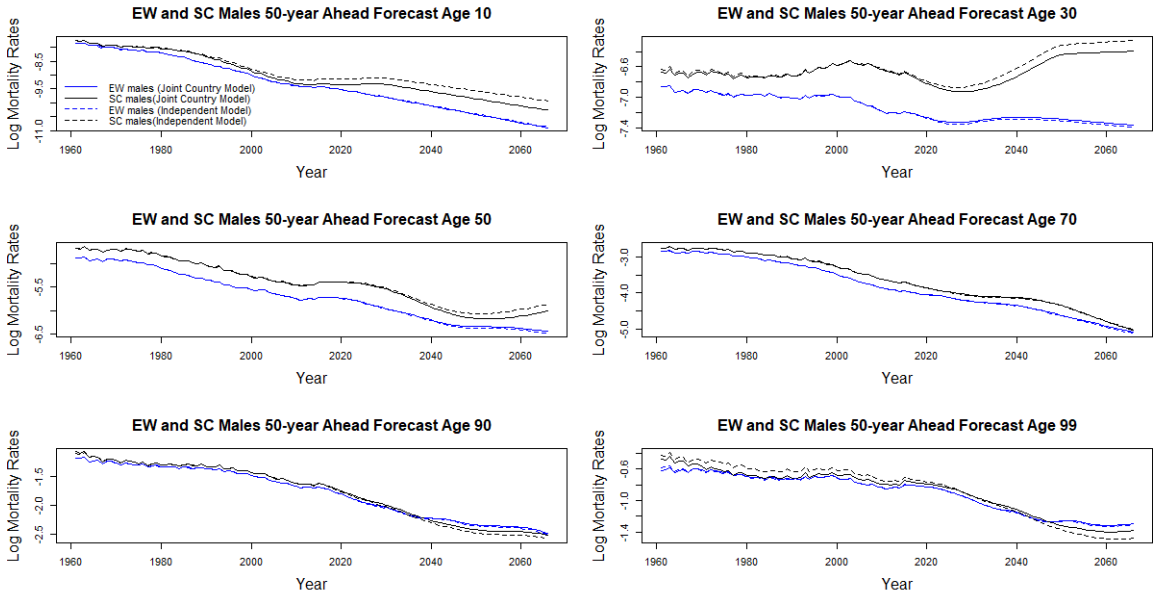


(a) crude SC female mortality rates from year 2006 to 2016

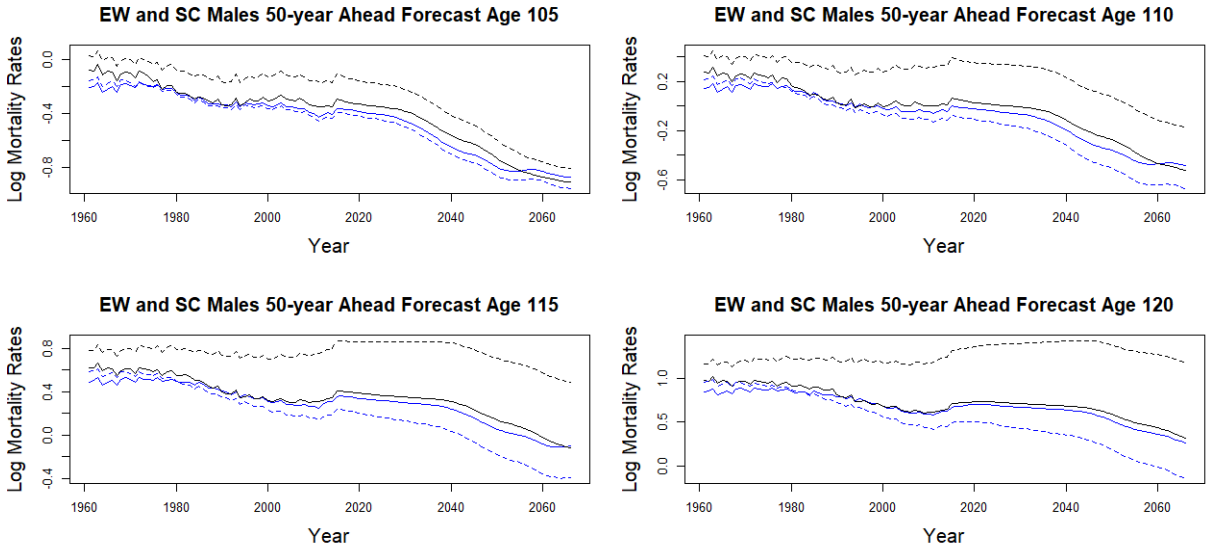


(b) yearly crude SC female mortality profile from year 2011 to 2016

Figure 5.33: The curde mortality rates of Scottish females in the most recent years

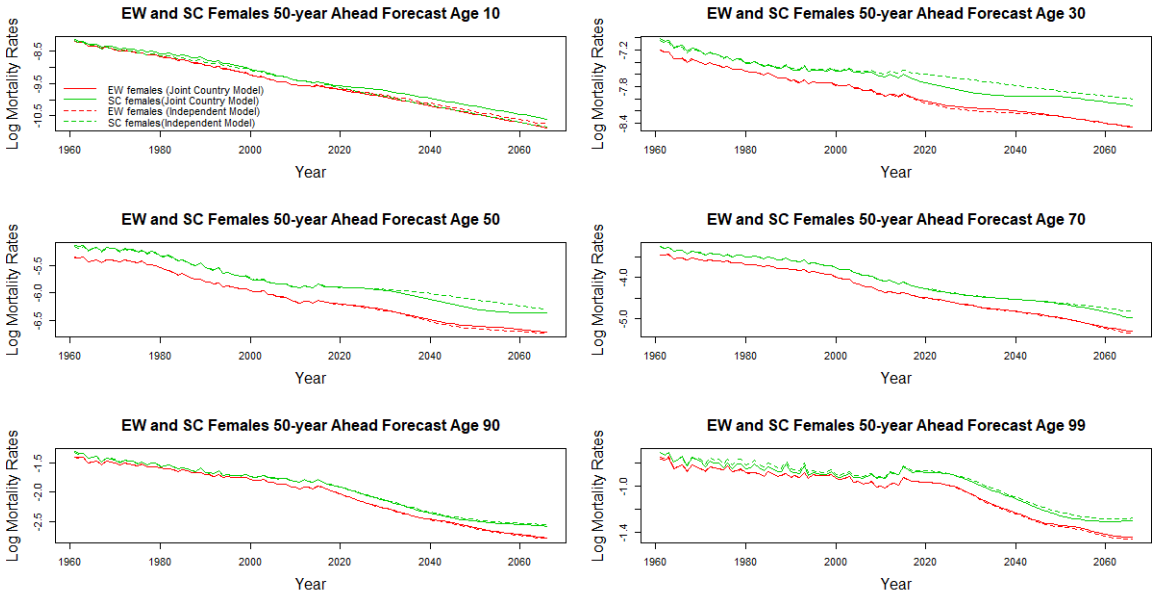


(a) age 10, 30, 50, 70, 90 and 99

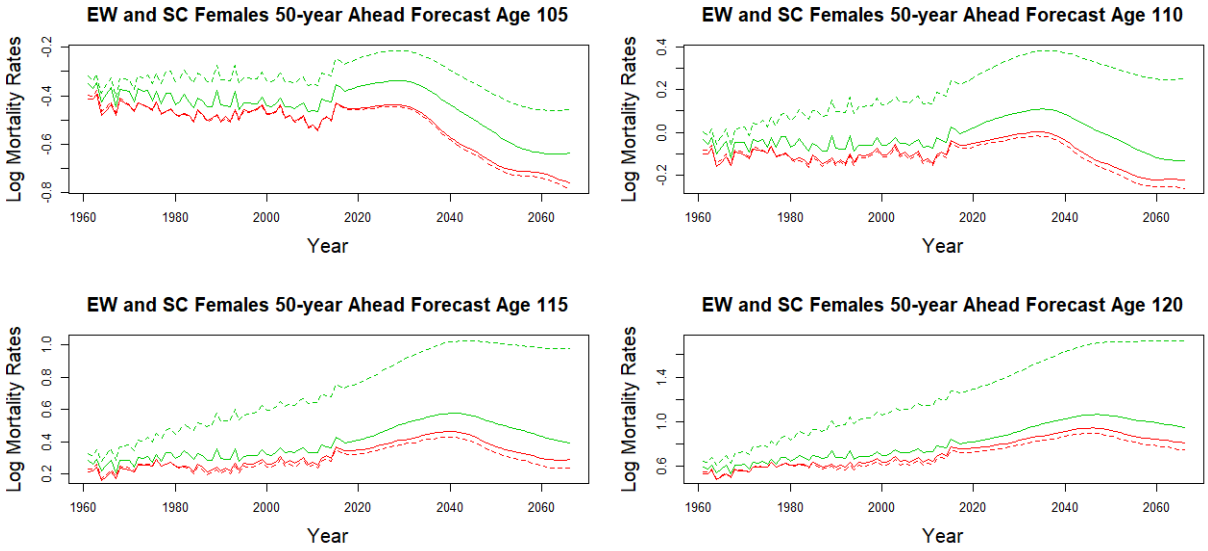


(b) age 105, 110, 115 and 120

Figure 5.34: 50 years ahead projections of EW (blue) and SC (black) male mortality rates at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint country model and the independent models respectively.



(a) age 10, 30, 50, 70, 90 and 99



(b) age 105, 110, 115 and 120

Figure 5.35: 50 years ahead projections of EW (red) and SC (green) female mortality rates at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to esimates of the joint country model and the independent models respectively.

5.4.3.3 Expert Opinions

We incorporate expert opinion as explained in Section 5.4.2.3. The resulting adjusted projections are shown in Figures 5.36, 5.37, 5.38 and 5.39. After the incorporation of expert opinion, the mortality projections are now non-divergent at all ages in the long run and the mortality schedules in each year now maintain a more regular structure over time. Note that under the

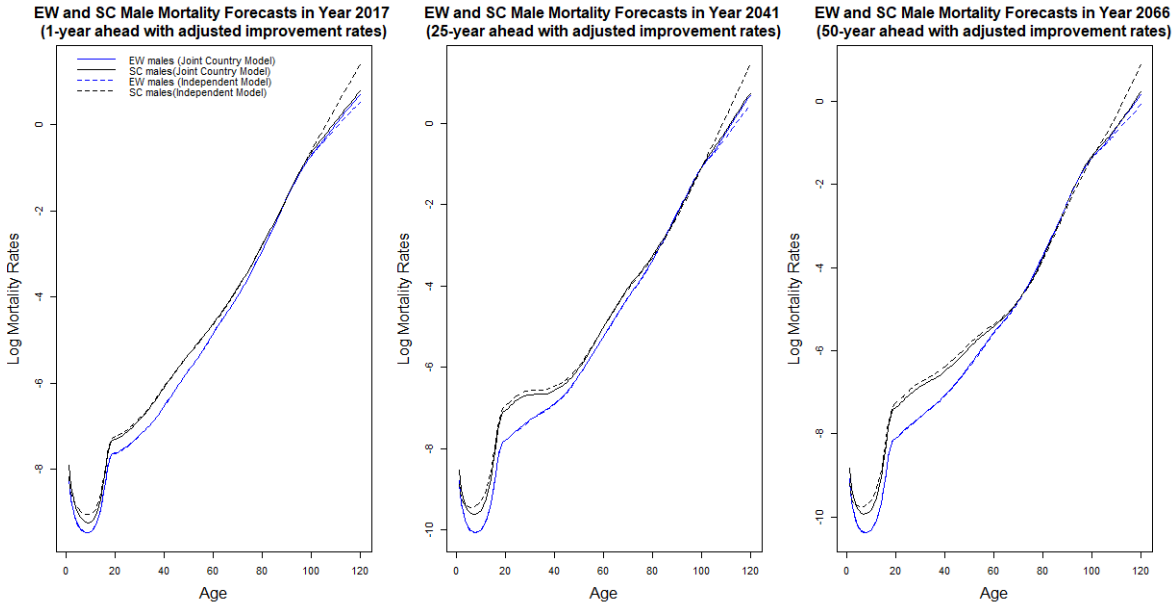


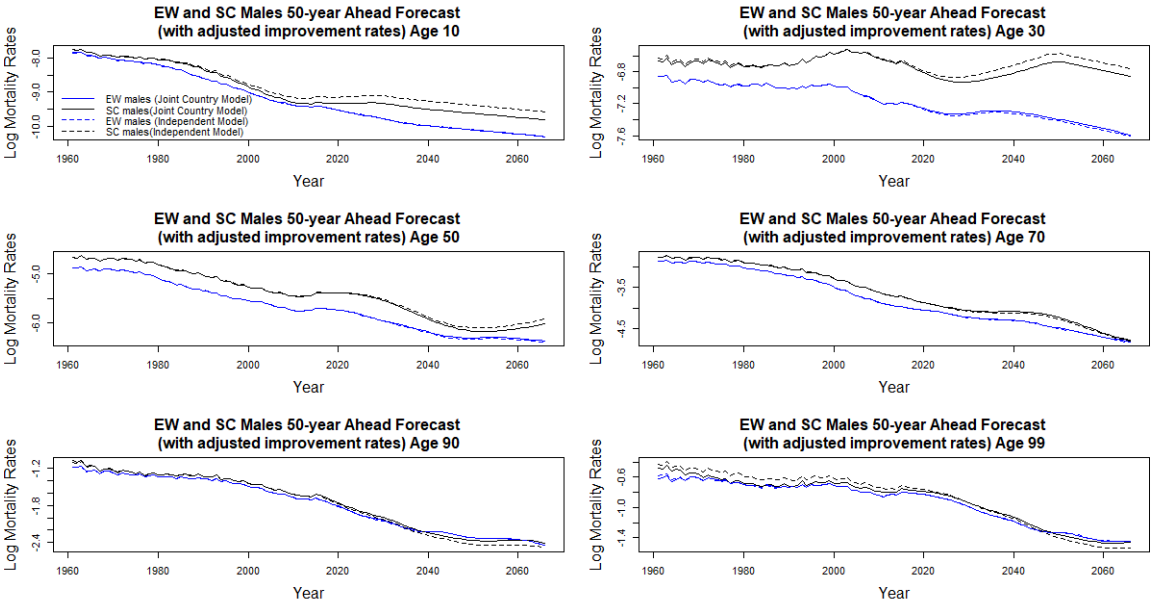
Figure 5.36: The projected mortality schedules in 1, 25and 50 years ahead with the incorporation of expert opinion. The blue and black lines are the projections for EW and SC male mortality rates respectively while the solid and dotted lines correspond to the projections based on the joint and independent models respectively.

independent model, even after incorporating the expert opinion, the projected mortality trends of EW and SC females still diverge in the extrapolated age range in short-to-mid term.

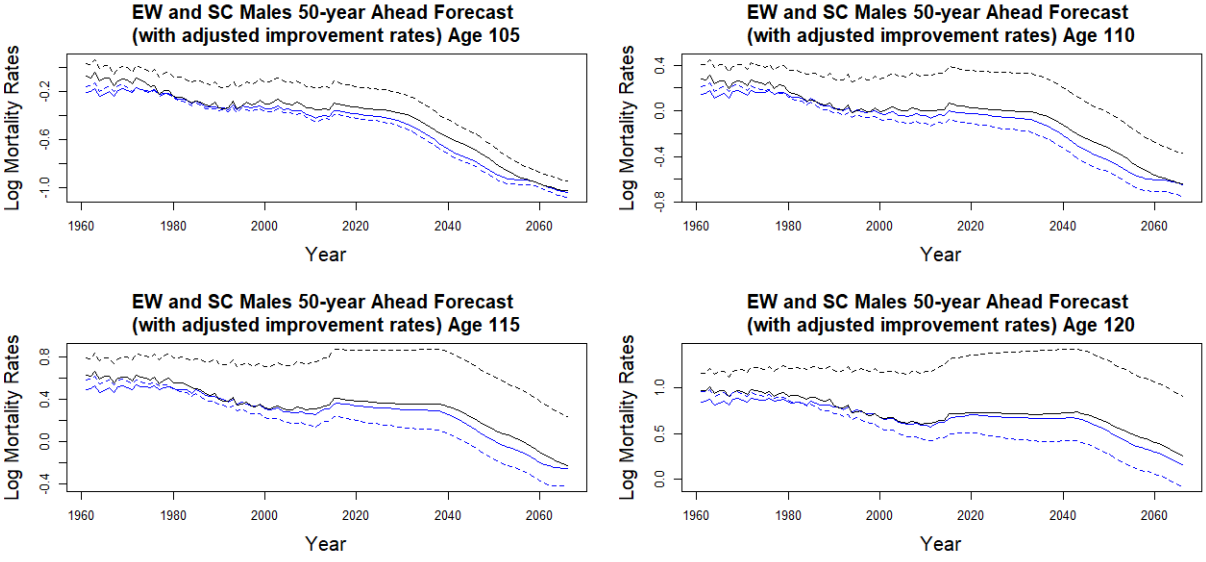
5.5 Conclusion

In this chapter we first introduced a mortality model using adaptive P-splines for single population, then we introduced a joint sex model for male and female mortality rates, followed by a joint country model. We also presented a method to incorporate expert opinion as we believe that the long term projections would benefit from moderation by experts.

The joint sex models have shown to be able to produce more reasonable long term male and female mortality projections that are non intersecting, a quality that single sex models often fail to achieve and requires ad-hoc adjustments. Information is borrowed at the highest ages where exposures are small. By doing so the extrapolation to higher ages beyond data range gives more plausible estimates, especially for the mortality improvement rates for females at the highest ages where a worsening mortality is otherwise projected. The linear mortality improvement assumption is suitable except for the most recent years in which there seems to be a deviation from the linearity. An AR(1) and an ARIMA(1,1,0) processes have been used for projecting the period effects. The evidence suggests that the ARIMA(1,1,0) is more suitable for the EW population, while we do not find strong evidence that they differ for the SC population. While the AR(1) is a stationary process, it gives prediction intervals that are seemingly too narrow. The ARIMA(1,1,0) on the other hand is non-stationary and gives wider prediction intervals,



(a) age 10, 30, 50, 70, 90 and 99



(b) age 105, 110, 115 and 120

Figure 5.37: 50 years ahead projections of EW (blue) and SC (black) male mortality rates with the incorporation of expert opinion at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint country model and the independent models respectively.

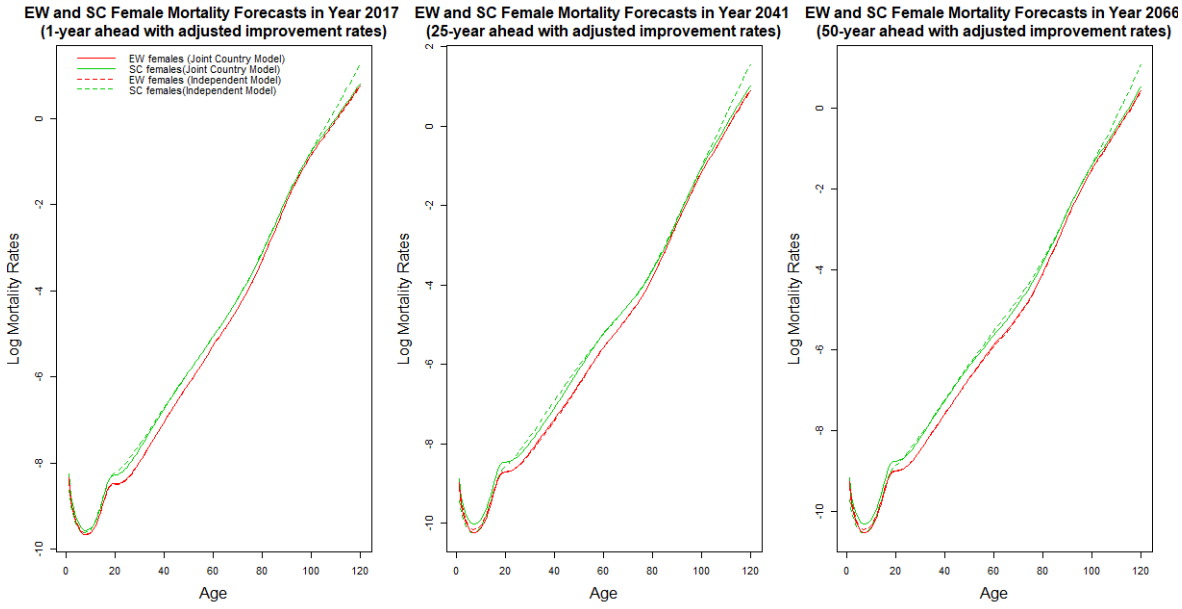
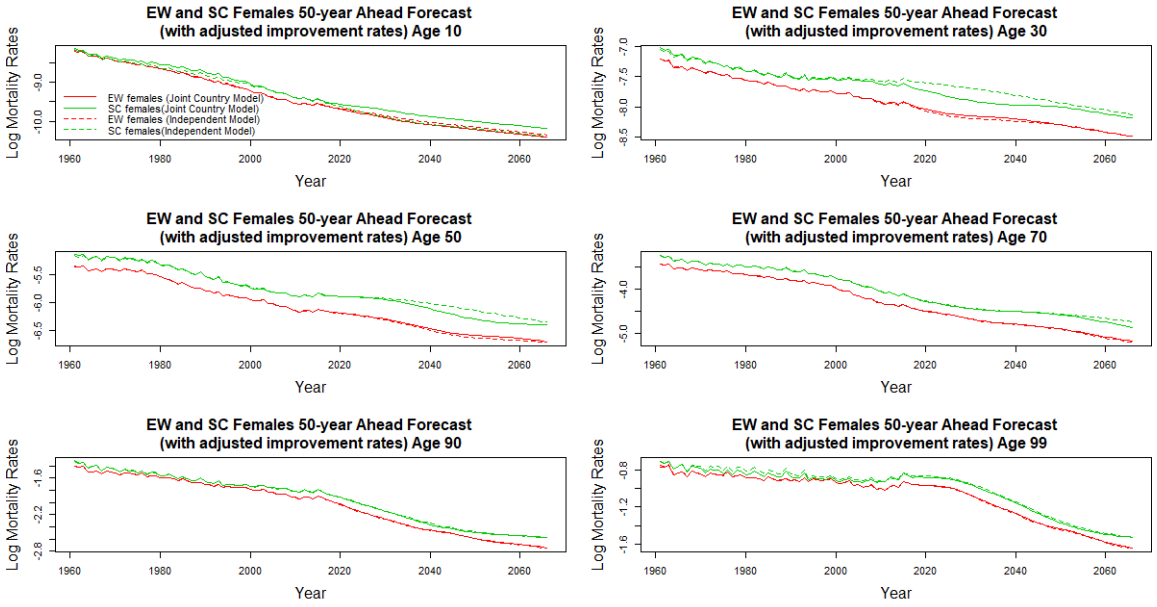


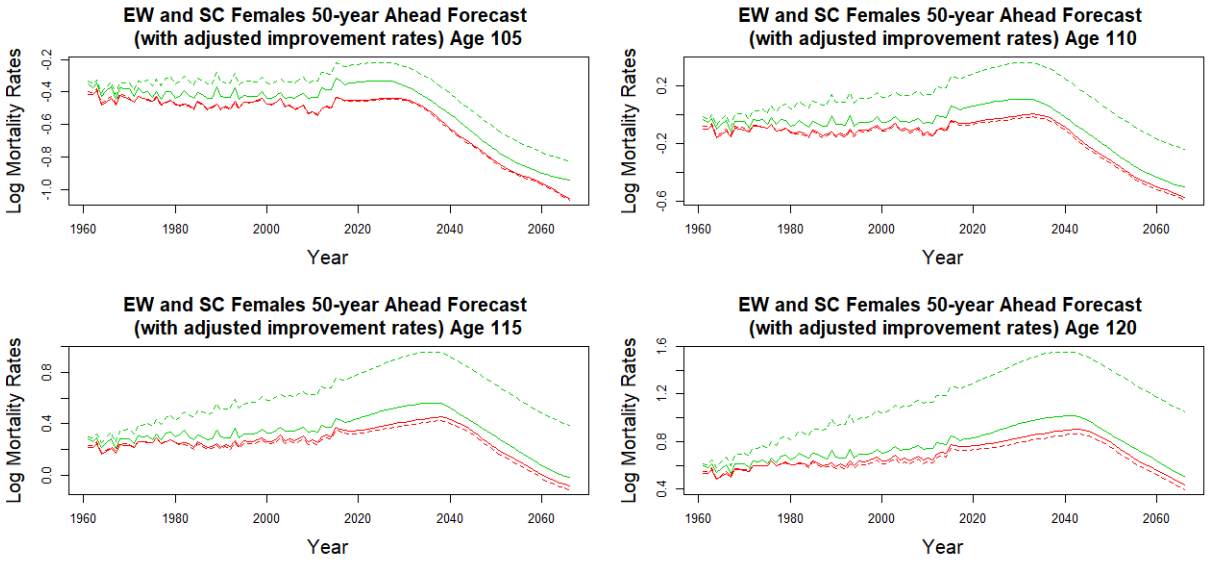
Figure 5.38: The projected mortality schedules in 1, 25 and 50 years ahead with the incorporation of expert opinion. The red and green lines are the projections for EW and SC female mortality rates respectively while the solid and dotted lines correspond to the projections of the joint and the independent models respectively.

and has a meaning that the expected future period shocks to annual mortality improvements are zero in the long run. The models have identified the well-known generational effect for cohorts born about 1932 (Willets, 2004; Richards et al., 2006; Murphy, 2009) as well as cohorts born about 1944 (Willets, 2004), these cohorts have better mortality improvement than neighbouring cohorts. The cohort born around 1988 has also experienced better mortality improvement as indicated by our models. In terms of the forecast accuracy, a backtest have been conducted leaving out the last 10 years of data as the validation set. For EW males and females, when an ARIMA(1,1,0) is in place, the forecast accuracy has slightly improved for most of the ages in the joint sex model, however when an AR(1) is used the accuracy has slightly worsened except for the youngest age group, according to the MAE. For SC males and females, the forecast accuracy has slightly improved only in the younger age groups but the magnitude of the differences in the MAE are relatively small.

The joint country model provides a way for the smaller SC population to borrow strength and learn from the bigger EW population. The model is constructed based on the belief that EW and SC male (female) mortality have similar structures. By jointly modelling the two countries, estimation and extrapolation of the SC baseline mortality schedule and age-specific improvement rates make use of information provided by the EW population which has a wider data range. Instead of linearly extrapolating the baseline mortality schedule, the estimated SC baseline can then follow the baseline mortality schedule of EW, with a decelerating rate of increase at the highest ages. The decelerating rate of increase in mortality rates at the highest ages is a feature has been noted by researchers (Carriere, 1992; Perks, 1932; Beard, 1959; Thatcher, 1999; Saikia



(a) age 10, 30, 50, 70, 90 and 99



(b) age 105, 110, 115 and 120

Figure 5.39: 50 years ahead projections of EW (red) and SC (green) female mortality rates with the incorporation of expert opinion at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint country model and the independent models respectively.

and Borah, 2014; Pitacco, 2016). When Scottish mortality is modelled independently, we observe positive improvement rates (worsening mortality) at some ages, which becomes a problem in long-term projections. Under the joint model this issue is largely avoided or even eliminated. Similarly the linear improvement assumption seems to be suitable except for the most recent years. The generational effects for cohorts born about 1931 and 1944 are identified in the joint country model. An additional cohort with better mortality improvement is identified in the joint country model for SC females born around year 1988, a pattern that is not captured in the single population model. Considering the EW males, EW females and SC males born around 1988 all have better mortality improvement, it is highly likely that SC females born around that time also enjoys better mortality improvement, meaning that the single population model has possibly failed to identify this feature. The joint country model is able to produce non-divergent long term projections between the countries for both males and females. The forecast accuracy has improved for both males and females except for the younger age groups, regardless of the underlying time-series structures (AR(1) or ARIMA(1,1,0)). This indicates that a lighter cross-country penalty at younger ages may perform better as data at these ages are more abundant and reliable. In other words, instead of a global cross-country penalty penalising all coefficient at the same degree, a varying local cross-country penalty may potentially increase the accuracy of predictions.

Chapter 6

Mortality Projection - England-and-Wales and Scotland males and females Joint Model

In this chapter we jointly model the four populations we considered so far, i.e. EW males, EW females, SC males and SC females. To further allow the sharing of information among populations and to simplify estimation, common smoothing parameters are assumed for the cross-sex penalties for each male-and-female pair, regardless of the country; and common smoothing parameters are assumed for the cross-country penalties for each EW-and-SC pair, regardless of the sex. As in Section 5.4.3, common smoothing parameters for the smoothness penalties of the splines between countries are also kept. In other words, the penalty functions that are shared across sexes are the smoothness penalties of the splines and the cross-country shape penalties; and the penalty functions that are shared across countries are the cross-sex difference penalties. Therefore, the penalties that are used in the joint model of all of the 4 populations are as follow, given in Table 6.1.

Function	Smoothness Penalty	Cross-sex Penalty	Cross-country Penalty
$s_\alpha(x)$	$\sum e^{\lambda_1^{m,\alpha}+\lambda_2^{m,\alpha}i}(\bigtriangledown^2\alpha_i^{EW,m})^2$ $\sum e^{\lambda_1^{m,\alpha}+\lambda_2^{m,\alpha}i}(\bigtriangledown^2\alpha_i^{SC,m})^2$ $\sum e^{\lambda_1^{f,\alpha}+\lambda_2^{f,\alpha}i}(\bigtriangledown^2\alpha_i^{EW,f})^2$ $\sum e^{\lambda_1^{f,\alpha}+\lambda_2^{f,\alpha}i}(\bigtriangledown^2\alpha_i^{SC,f})^2$	$\sum e^{\lambda_1^{d,\alpha}+\lambda_2^{d,\alpha}i}(\alpha_i^{EW,m}-\alpha_i^{EW,f})^2$ $\sum e^{\lambda_1^{d,\alpha}+\lambda_2^{d,\alpha}i}(\alpha_i^{SC,m}-\alpha_i^{SC,f})^2$	$\sum \lambda^{s,\alpha} \bigtriangledown (\alpha_i^{EW,m}-\alpha_i^{SC,m})^2$ $\sum \lambda^{s,\alpha} \bigtriangledown (\alpha_i^{EW,f}-\alpha_i^{SC,f})^2$
$s_\beta(x)$	$\sum \lambda^{m,\beta}(\bigtriangledown^2\beta_i^{EW,m})^2$ $\sum \lambda^{m,\beta}(\bigtriangledown^2\beta_i^{SC,m})^2$ $\sum \lambda^{f,\beta}(\bigtriangledown^2\beta_i^{EW,f})^2$ $\sum \lambda^{f,\beta}(\bigtriangledown^2\beta_i^{SC,f})^2$	$\sum e^{\lambda_1^{d,\beta}+\lambda_2^{d,\beta}i}(\beta_i^{EW,m}-\beta_i^{EW,f})^2$ $\sum e^{\lambda_1^{d,\beta}+\lambda_2^{d,\beta}i}(\beta_i^{SC,m}-\beta_i^{SC,f})^2$	$\sum \lambda^{s,\beta} \bigtriangledown (\beta_i^{EW,m}-\beta_i^{SC,m})^2$ $\sum \lambda^{s,\beta} \bigtriangledown (\beta_i^{EW,f}-\beta_i^{SC,f})^2$
$s_\gamma(t-x)$	$\sum \lambda^{m,\gamma}(\bigtriangledown^2\gamma_i^{EW,m})^2$ $\sum \lambda^{m,\gamma}(\bigtriangledown^2\gamma_i^{SC,m})^2$ $\sum \lambda^{f,\gamma}(\bigtriangledown^2\gamma_i^{EW,f})^2$ $\sum \lambda^{f,\gamma}(\bigtriangledown^2\gamma_i^{SC,f})^2$		

Table 6.1: The penalties for the joint model of 4 populations

The $\{\lambda_1^{m,\alpha}, \lambda_2^{m,\alpha}\}$ and $\{\lambda_1^{f,\alpha}, \lambda_2^{f,\alpha}\}$ control the smoothness of male and female baseline mortality schedules respectively, regardless of the country. $\lambda^{m,\beta}$ and $\lambda^{f,\beta}$ control the smoothness of male and female improvement rates respectively, regardless of the country. $\lambda^{m,\gamma}$ and $\lambda^{f,\gamma}$ control the smoothness of male and female cohort effects respectively, again regardless of the country.

The $\{\lambda_1^{d,\alpha}, \lambda_2^{d,\alpha}\}$ and $\{\lambda_1^{d,\beta}, \lambda_2^{d,\beta}\}$ control the penalty of the differences between male and female baseline mortality schedules and improvement rates respectively, regardless of the country.

The $\lambda^{s,\alpha}$ and $\lambda^{s,\beta}$ control the penalty of the differences between the shapes of EW and SC baseline mortality schedules and improvement rates respectively, regardless of the sex.

The same problem discussed in Section 5.4.2 persists, the $\{\lambda_1^{d,\beta}, \lambda_2^{d,\beta}\}$ estimates tend to extreme values during the optimisation. The estimated cross-sex penalty function for the age-specific improvement rates tend to infinity at older ages. Therefore we fix them at -55 and 85 as in Section 5.4.2, and proceed to estimate the remaining smoothing parameters.

6.1 Results

6.1.1 Overview

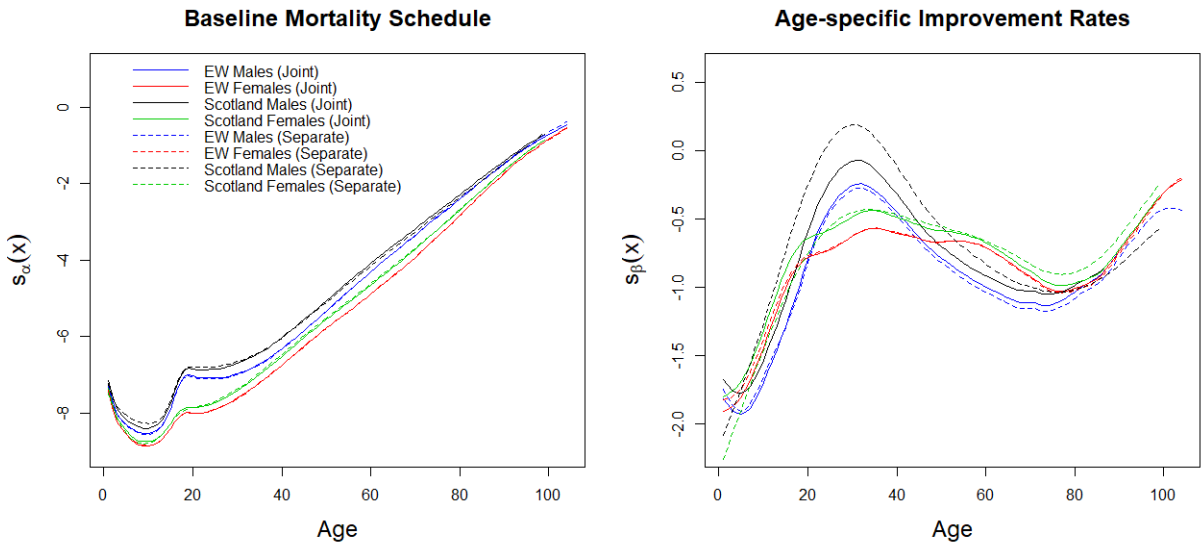
Figure 6.1 plots the parameter estimates of the joint model alongside the estimated terms of the independent models. The baseline schedules and the age-specific improvement rates tend to a common value at the oldest ages in the joint model. Note that the baseline schedules and improvement rates of SC males and females are not penalised towards that of EW males and

females. It is interesting that by penalising the cross-country shape differences and cross-sex differences, we obtain converging trends in all four populations (see Figure 6.1a). The estimated period effects from the joint model remain similar as those from the independent models. The estimated EW male and female cohort effects from the joint model also maintain similar structures to that from the independent models. The estimated SC male and female cohort effects show more considerable variability before and after jointly modelling the populations, due to having smaller exposures (see Figure 6.1b). Figures 6.2 and 6.3 plot the baseline schedules and the improvement rates extrapolated up to age 120. The effect of the penalties are apparent and the estimates are more plausible. Instead of having the SC baseline schedules extrapolating upwards linearly, information is borrowed from the bigger EW population which has data of a wider age range, capturing the deceleration in mortality increase in age. The age-specific improvement rates of different populations are now more consistent with each other especially at older ages, as opposed to them going in different directions under the independent models. The joint model has also largely eliminated unreasonable trends such as worsening mortality (positive $s_\beta(x)$) in the oldest ages for SC females (dotted green line) and around age 30 for SC males (dotted black line) when they are modelled independently. As we will see in later Sections 6.1.2 and 6.1.3, the joint model also effectively avoids implausible projections such as intersecting male and female mortality rates and divergent trends.

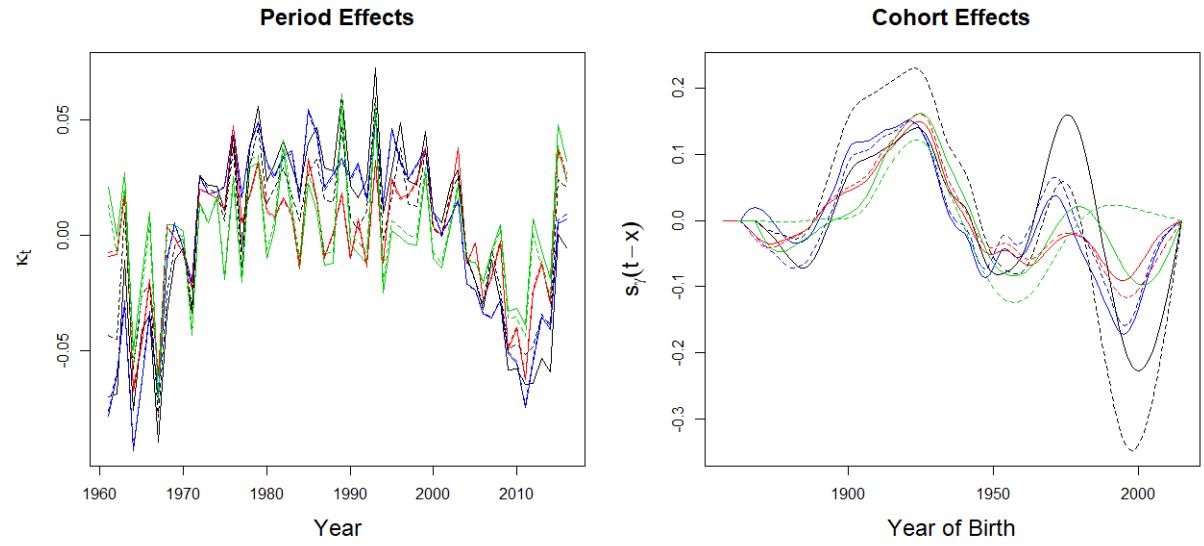
Figure 6.4 plots the first differences of the estimated period effects of each population individually. The estimated period shocks to annual mortality improvements (κ_t^*'s) under the joint model and the independent models are extremely similar to each other. The dotted lines and the solid lines are almost superposed. The first differences oscillate around zero, indicating the linearity mortality improvement assumption is appropriate.

Figure 6.5 plots the first differences of the estimated cohort effects of each population individually. Similar conclusion as in Chapters 5 can be drawn. That is, several troughs can be identified in years around 1932, 1945 and 1988. These cohorts are all identified under both the joint and independent models, except of SC females where the joint model located a better mortality improvement for cohorts born around 1988 that is not observed under the independent model. Considering the EW males, EW females and SC males born around 1988 all have faster mortality improvement, it is highly likely that SC females born around that time also enjoys better mortality improvement, meaning that the single population model for SC females has failed to capture this pattern. The first and the last few cohorts are more variable due to the lack of data.

A backtest is conducted as in Chapter 5 and data from the most recent 10 years is excluded as a validation set, i.e. data from 1961 to 2006 are used to train the model while data from 2007 to 2016 are used as validation. Figure 6.6 plots the projections of each population using an AR(1) process on the period effects while Figure 6.7 plots the projections using an ARIMA(1,1,0). The differences in the projections between the joint model and independent models for EW males and females are less noticeable compared to SC males and females. Both the joint and independent models over-estimate the male mortality rates at middle ages. The MAE figures which give a quantitative measure of the forecast accuracy, are presented in Table 6.2. According

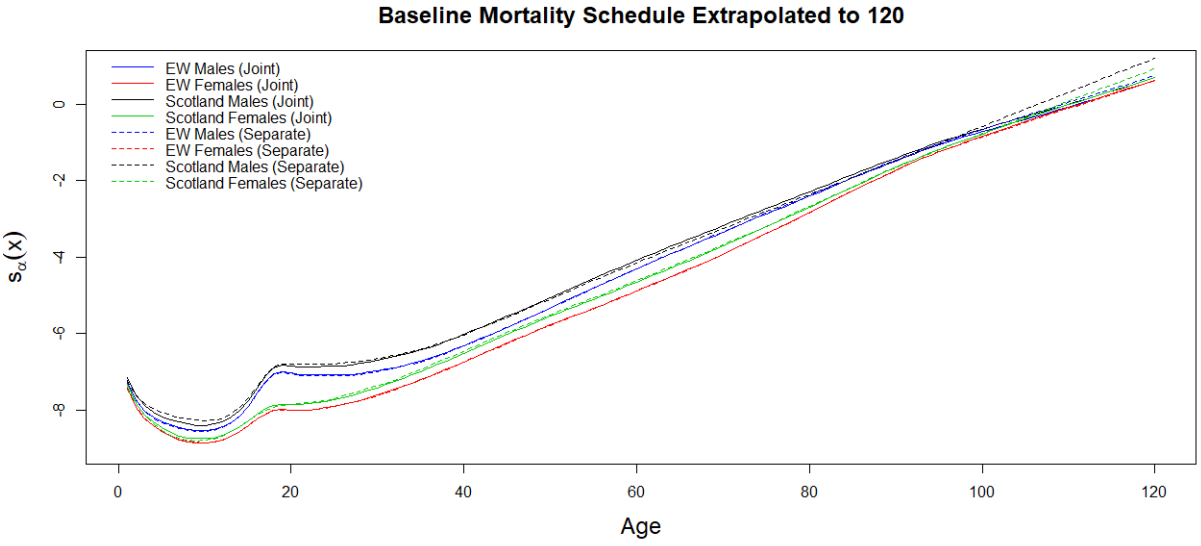


(a) estimated baseline mortality schedules and age-specific improvement rates

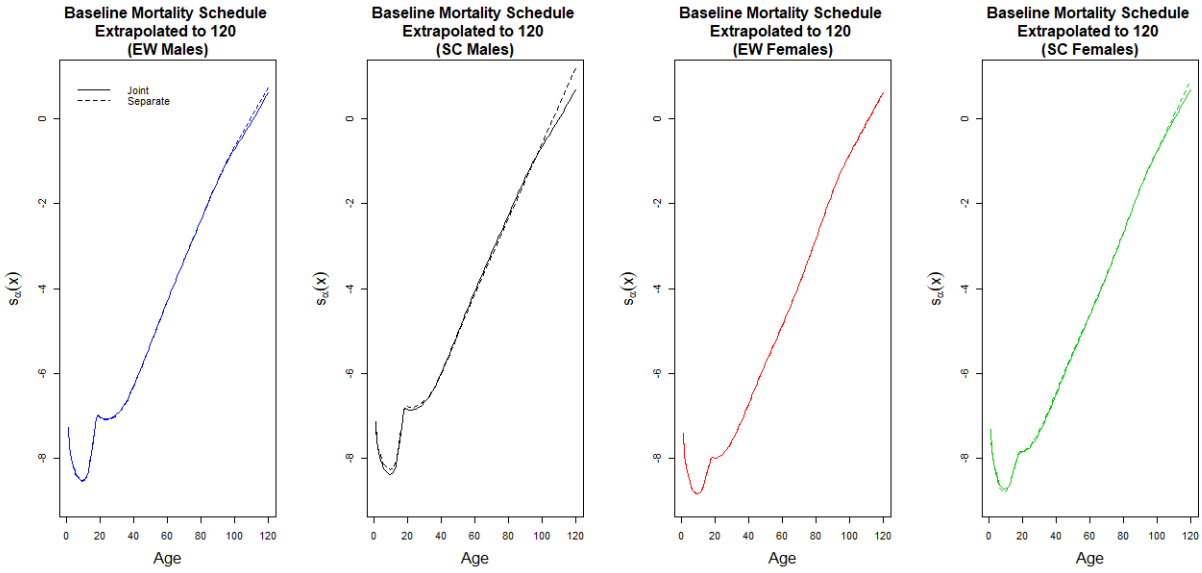


(b) estimated period and cohort effects

Figure 6.1: The parameter estimates of the joint model for EW males (blue), EW females (red), SC males (black) and SC females (green). The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.

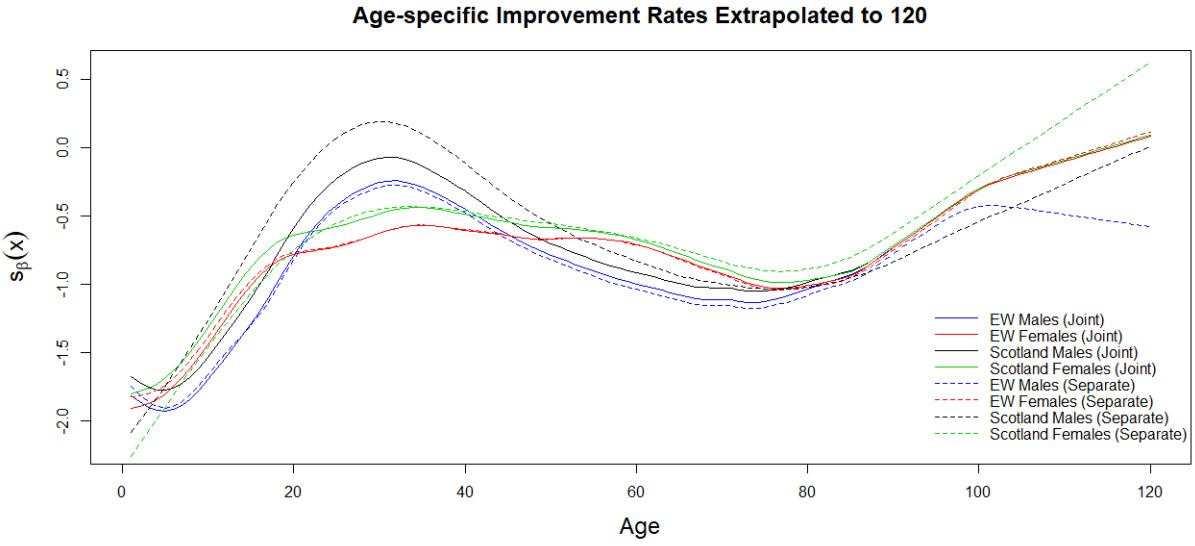


(a) estimated baseline mortality schedules extrapolated to age 120

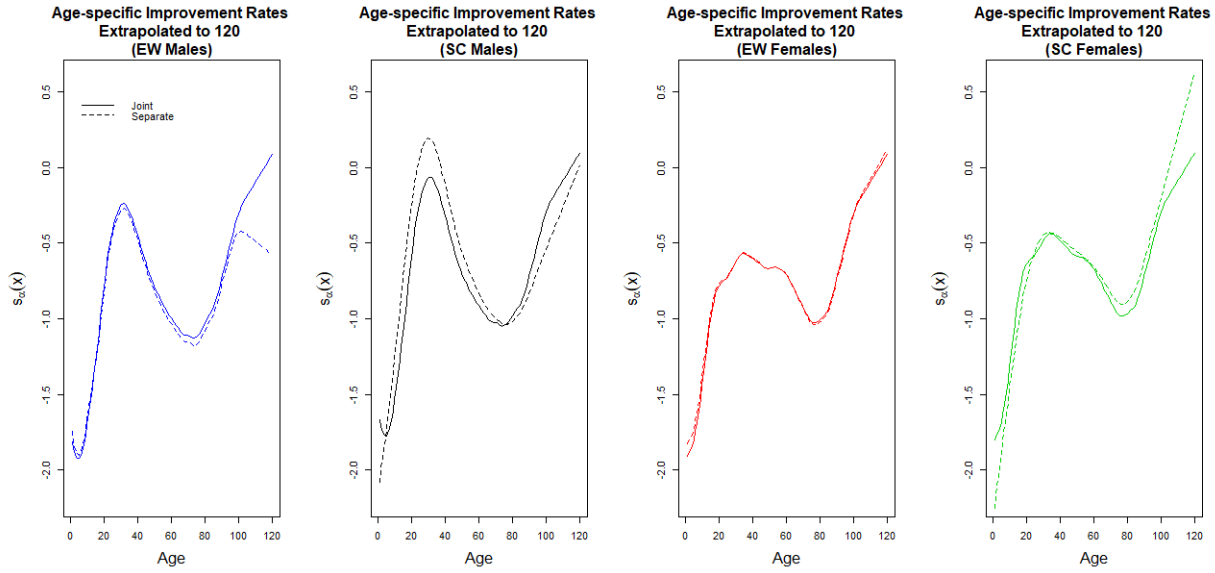


(b) individual estimated baseline mortality schedules

Figure 6.2: The estimated baseline mortality schedules extrapolated to age 120. The baselines for each population are also plotted individually for a clearer illustration. The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.



(a) estimated age-specific mortality improvement rates extrapolated to age 120



(b) individual estimated age-specific mortality improvement rates

Figure 6.3: The estimated mortality improvement rates extrapolated to age 120. The trends for each population are also plotted individually for a clearer illustration. The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.

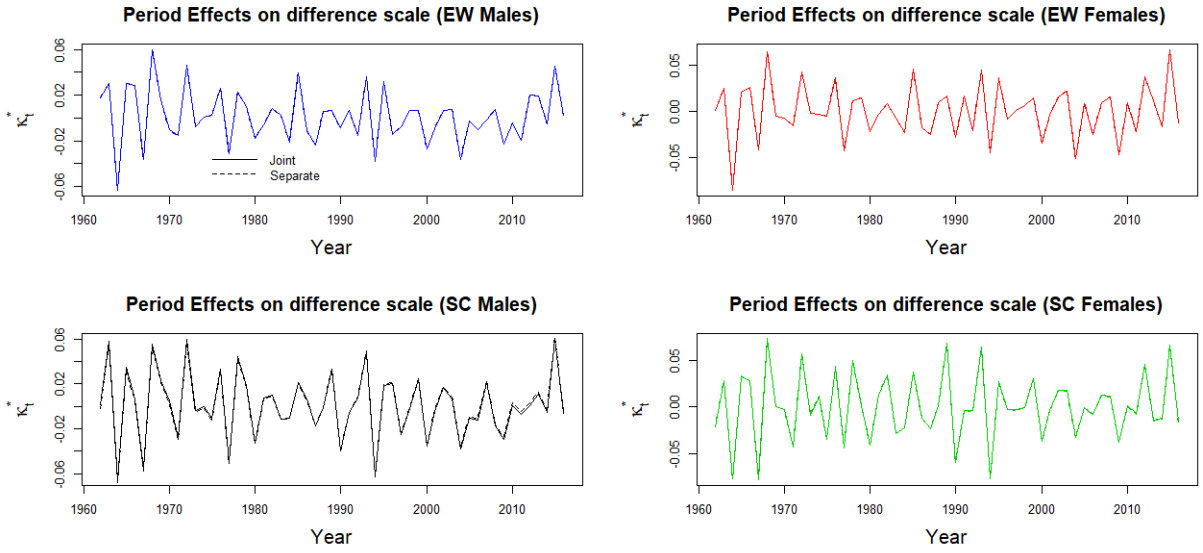


Figure 6.4: The estimated period effects of each population plotted individually on a difference scale.

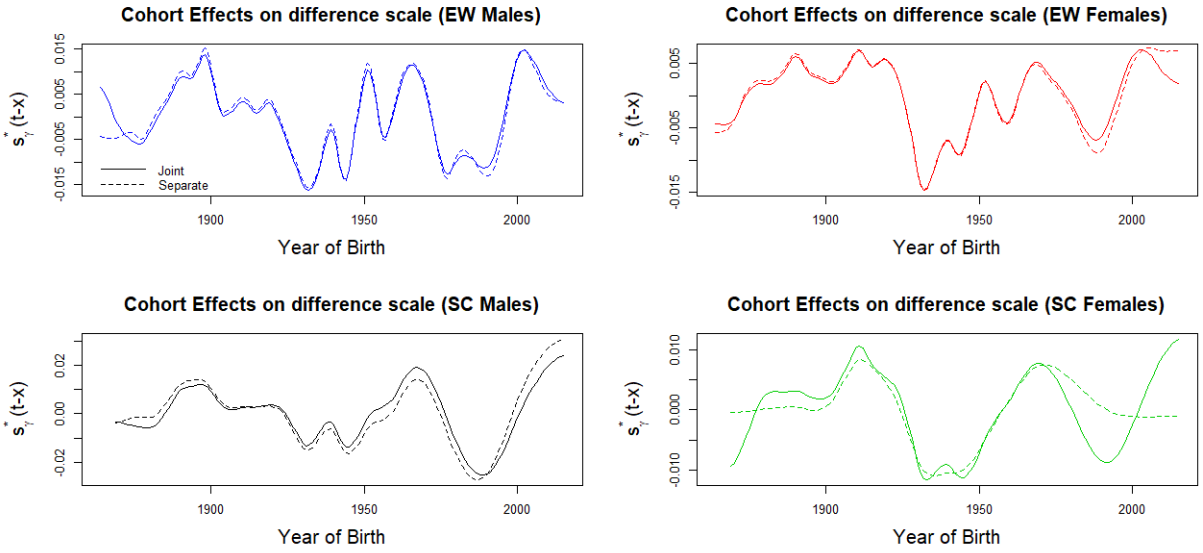
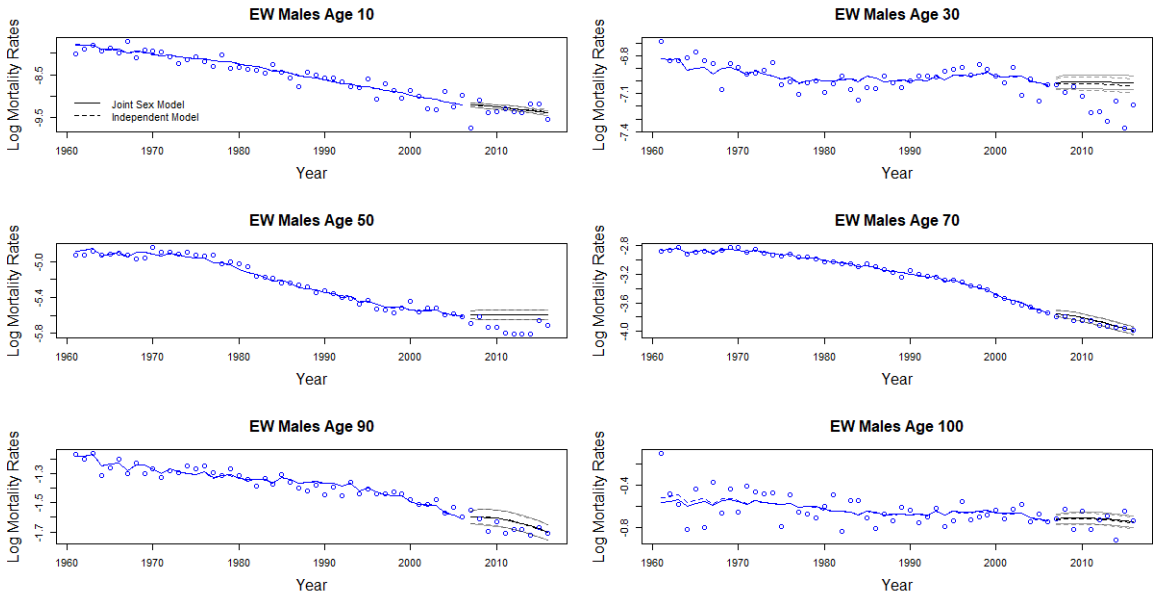


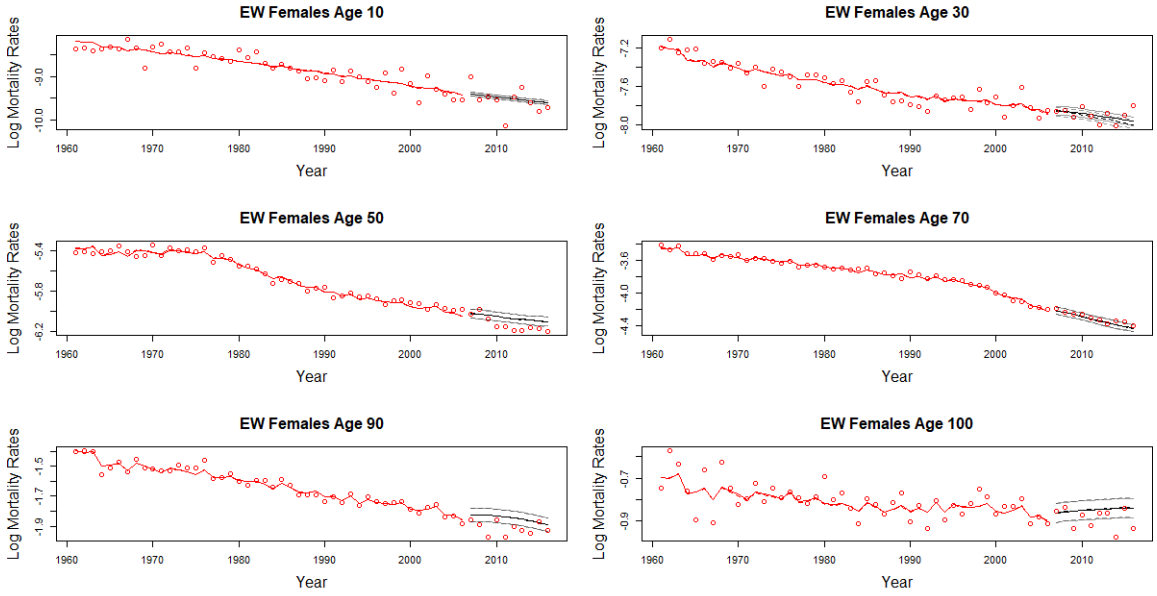
Figure 6.5: The estimated cohort effects of each population plotted individually on a difference scale.

to this, the forecast accuracy has improved under the joint model at most ages except for the youngest age group (1-30) when an AR(1) is used; and for ages 61+ when an ARIMA(1,1,0) is used. Table 6.3 shows the coverage of the estimated 95% intervals of the joint model. On average, the intervals produced using AR(1) are too narrow. On the other hand, ARIMA(1,1,0) produces wider intervals and hence more satisfactory coverage. The overall coverage of the intervals using ARIMA(1,1,0) are still below 95%, however, for older age groups, the coverage are much closer to 95%, especially for EW males aged 61-90. Focusing on older age groups (61+), the coverage of the intervals produced by the joint model for SC males has improved a lot, compared to the independent model. For SC males aged 61-90, the intervals produced

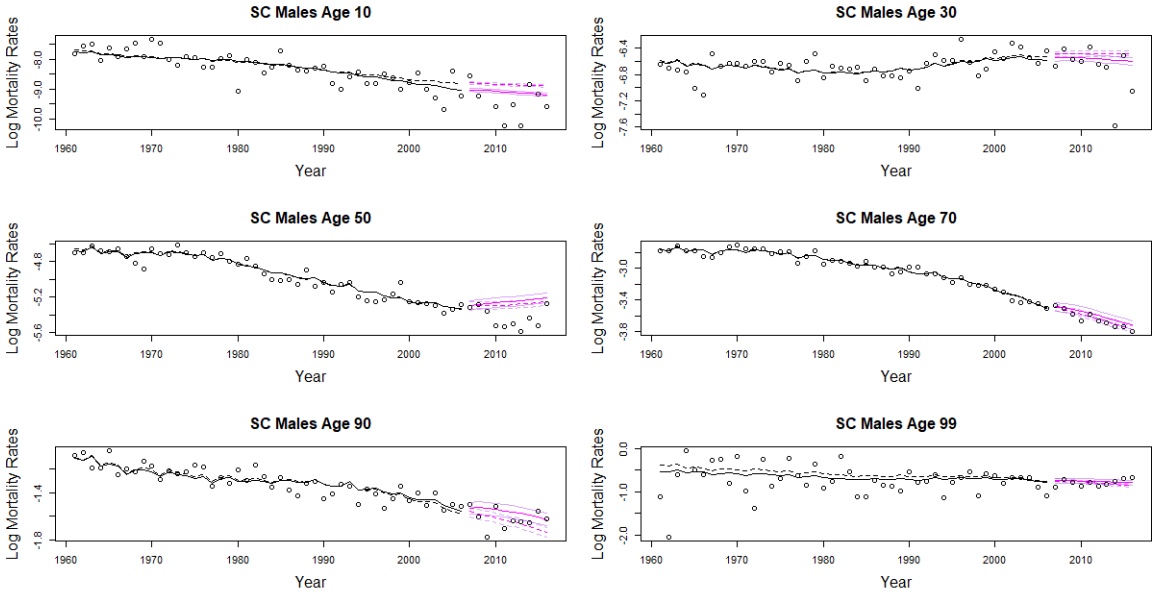
using ARIMA(1,1,0) is about 83% in the joint model, while in the single population model it is only 57%. For SC males aged 91+ using ARIMA(1,1,0), the joint model has a coverage of 67% against a coverage of only 46% in the single population model.



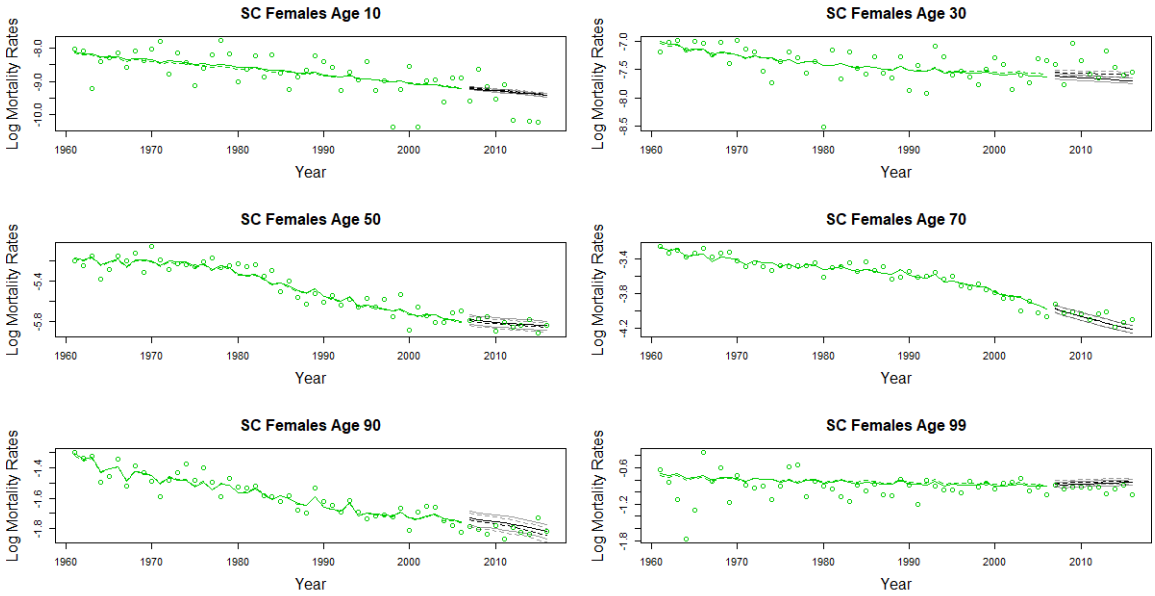
(a) EW males



(b) EW females

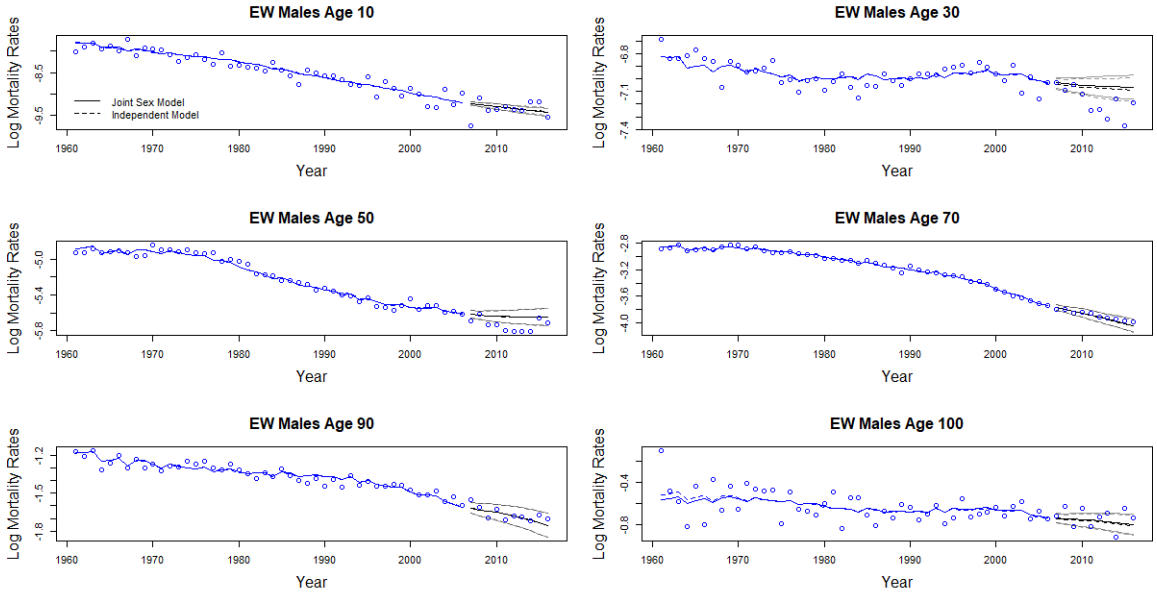


(c) SC males

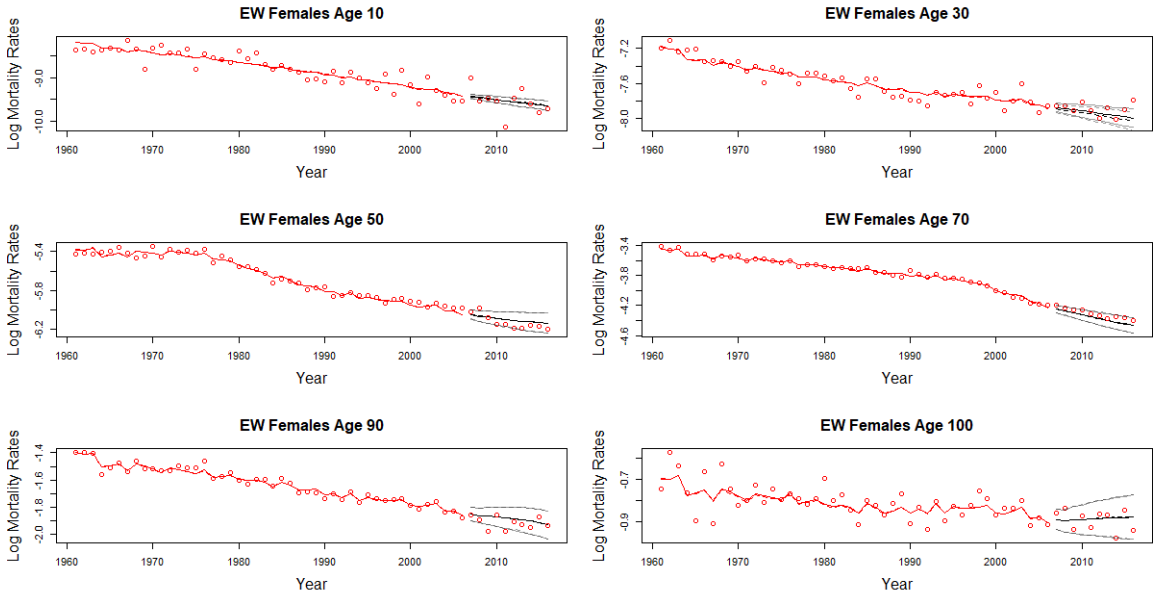


(d) SC females

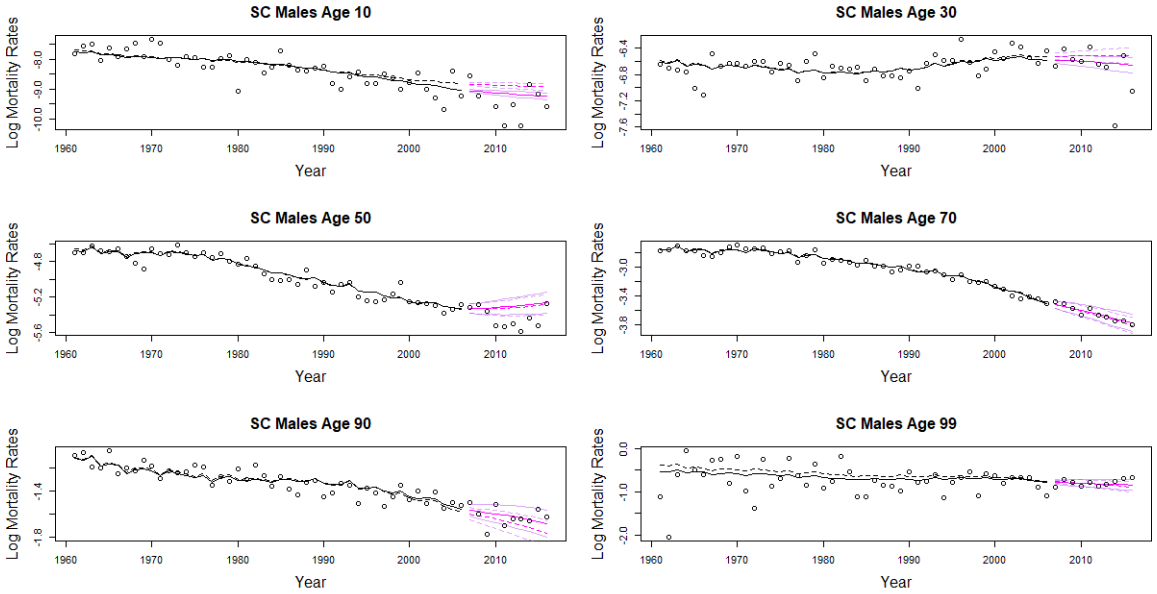
Figure 6.6: Estimated mortality rates and the projected 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an AR(1). The points are the observed mortality rates.



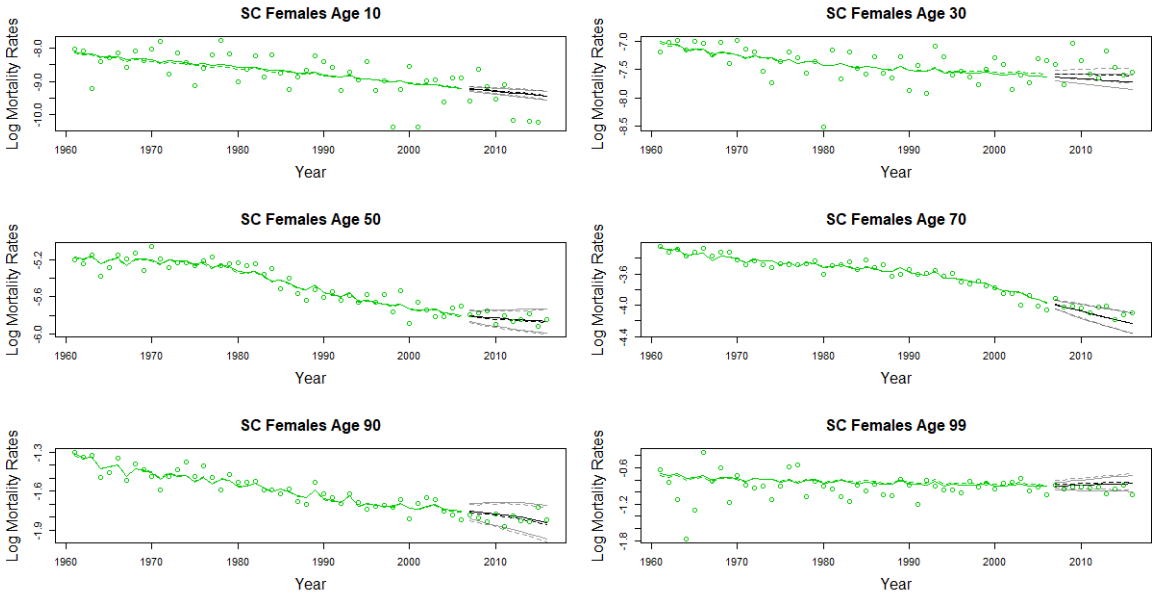
(a) EW males



(b) EW females



(c) SC males



(d) SC females

Figure 6.7: Estimated mortality rates and the projected 95% intervals at some selected ages from 1961 to 2006 (training) and from 2007 to 2016 (validation) using an ARIMA(1,1,0). The points are the observed mortality rates.

Age	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	1.553095	1.498843	1.563237	1.504025
31 to 60	0.8275105	0.8310234	0.8323145	0.8305121
61 to 90	0.5067144	0.5242525	0.5211037	0.5337053
91+	0.1417191	0.1549504	0.1441549	0.1558011
all ages	0.869452	0.8611379	0.8783984	0.8654089

Table 6.2: The mean absolute error of log projected mortality rates for all populations

Age	95% intervals coverage							
	EW males		EW females		SC males		SC females	
	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)	AR(1)	ARIMA(1,1,0)
all ages	0.4480769	0.5961538	0.4663462	0.7634615	0.351515139	0.498989912	0.289898979	0.519191942
61-90	0.77	0.9466667	0.6233333	0.89	0.64	0.8266667	0.4233333	0.7066667
91+	0.65	0.75	0.45	0.8785714	0.5222222	0.6666667	0.2111111	0.7666667

Table 6.3: The coverage of the 95% intervals of the joint model

6.1.2 Pairwise Comparison - Males and females

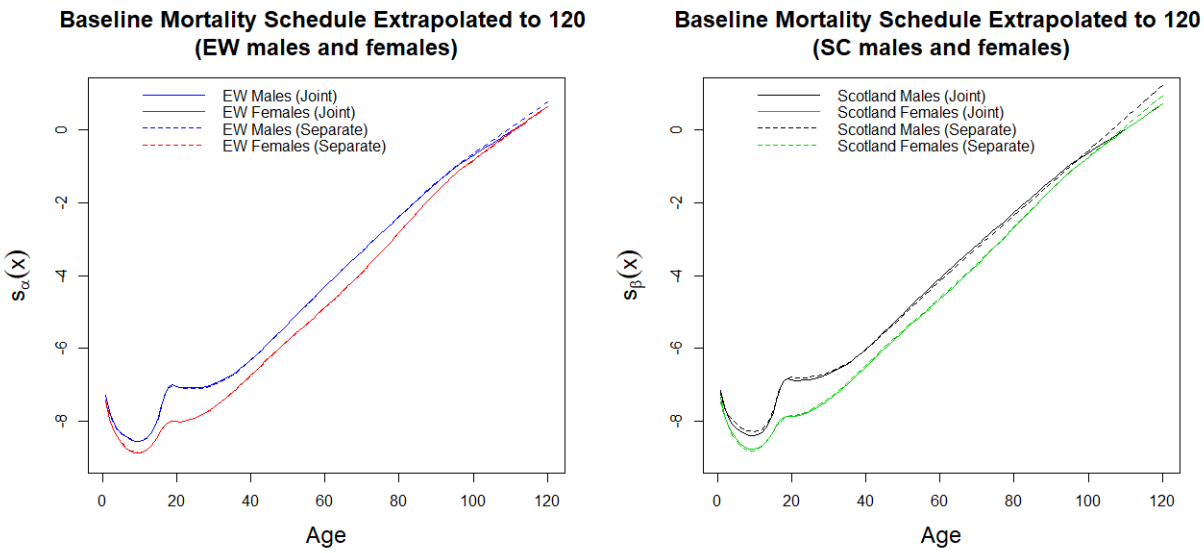
It may be more illustrative and informative to examine the estimated terms in pairs. In this section, the male-and-female pairs are considered. Figure 6.8 shows the estimated baseline mortality schedules and the improvement rates of the EW and SC male-and-female pairs. Contrary to only jointly modelling SC males and females (Figure 5.4), the estimated cross-sex penalty for the age-specific improvement rates is much stronger due to the presence of EW males and females. The cross-country penalty also indicates that the shape of the EW and SC baselines and improvement rates are similar. Hence the estimated trends for SC males and females resemble the EW trends, especially outside the SC data range.

6.1.2.1 Backtesting

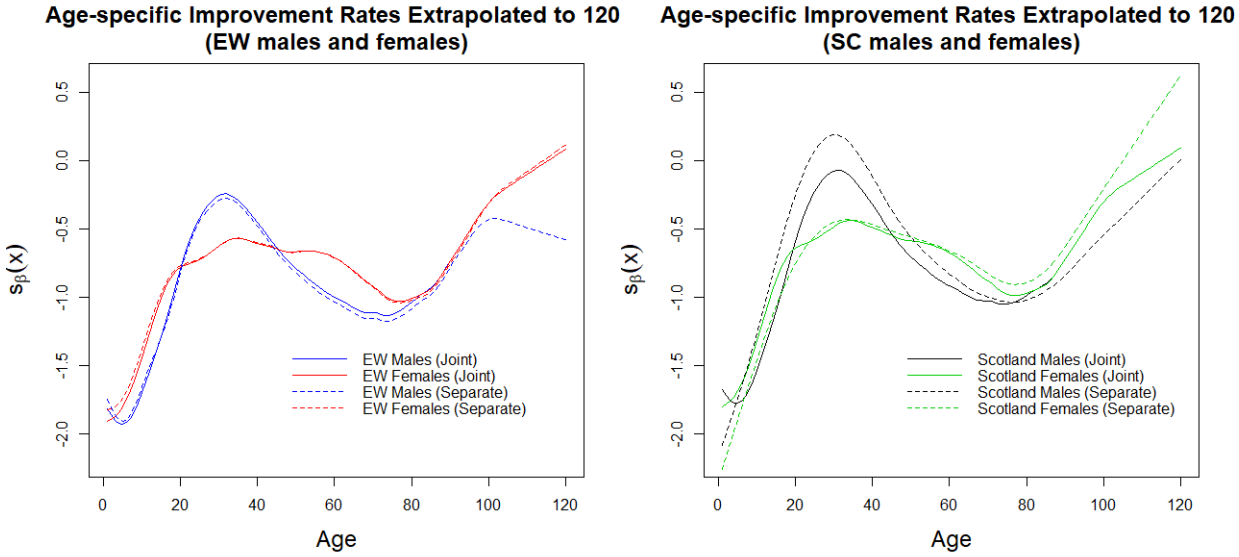
Tables 6.4 and 6.5 show the MAE of EW and SC populations respectively. For EW populations, the fit of the projections has slightly worsened under the joint model except for the age group 61-90 when using an ARIMA(1,1,0) process. However, the magnitudes of the differences in the MAE are very small except for the youngest age group. On the other hand, forecast accuracy has improved for all but the youngest age group for SC males and females. This indicates that the EW information is indeed beneficial to SC. One possible improvement to our model would then be to first model EW males and females jointly, and then model SC males and females jointly using this information.

Age	EW males and females			
	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	0.1639079	0.1534538	0.1484186	0.1392333
31 to 60	0.08808014	0.08648498	0.06191506	0.0608769
61 to 90	0.04222944	0.04168126	0.03523491	0.03600139
91+	0.0608055	0.05952919	0.04911866	0.04907227
all ages	0.0930558	0.08925009	0.07744922	0.074715

Table 6.4: The mean absolute error of log projected mortality rates for EW males and females



(a) estimated baseline mortality schedules for EW and SC male-and-female



(b) estimated age-specific improvement rates for EW and SC male-and-female

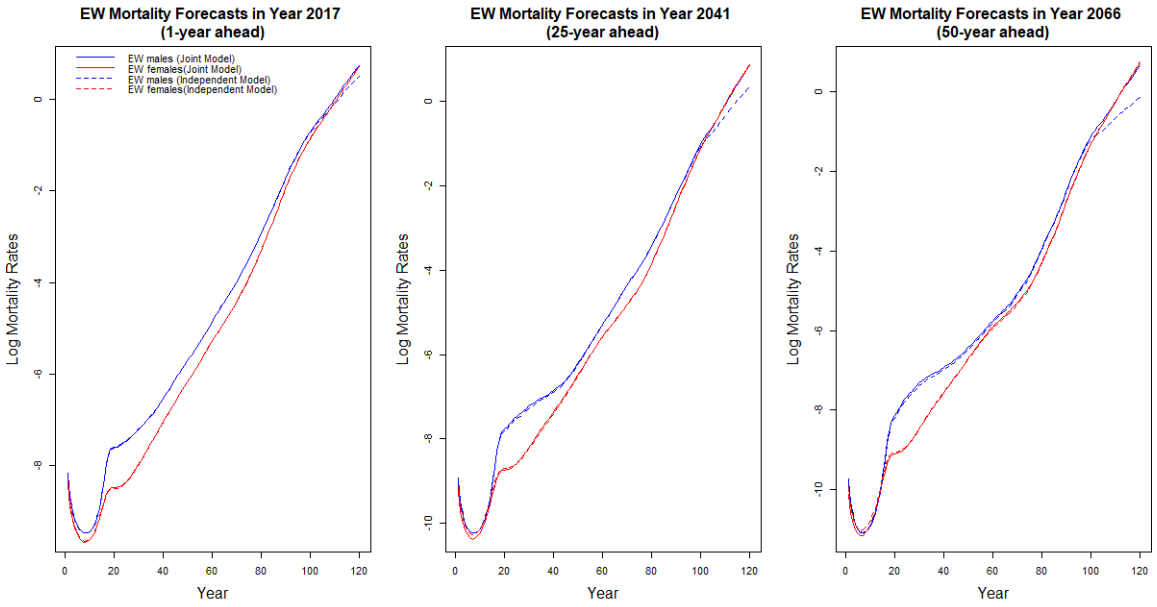
Figure 6.8: The estimated baseline mortality schedules and improvement rates extrapolated to age 120 for each country. The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.

Age	SC males and females			
	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	2.942282	2.844232	2.978055	2.868818
31 to 60	1.566941	1.575562	1.602714	1.600147
61 to 90	0.9711995	1.006824	1.006973	1.031409
91+	0.2675848	0.3033834	0.2919891	0.3218237
all ages	1.68506	1.67201	1.7198	1.696037

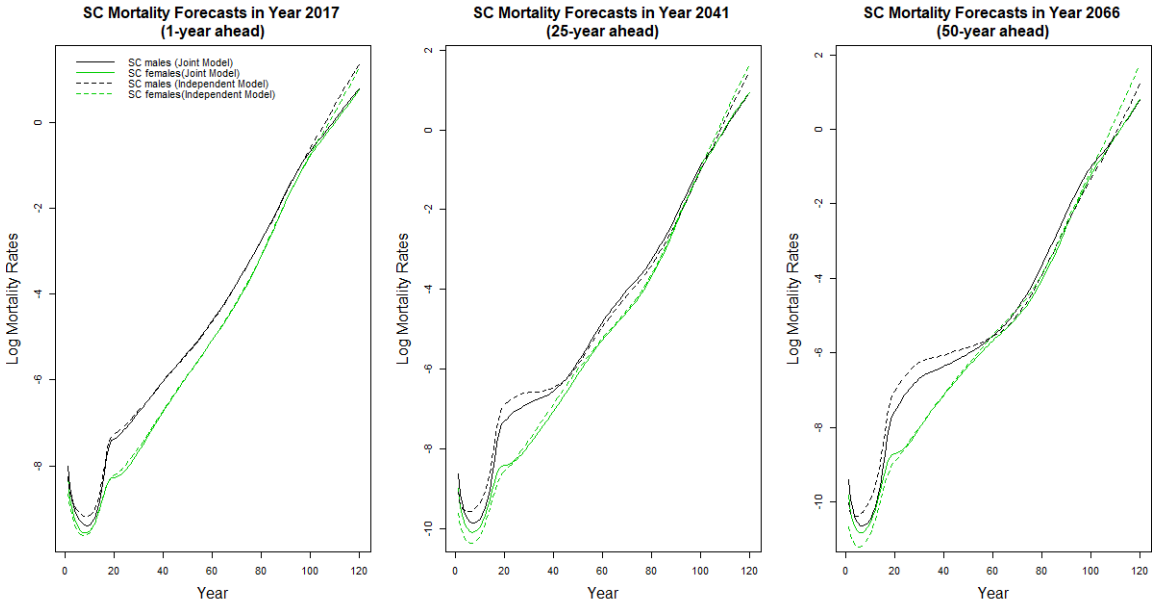
Table 6.5: The mean absolute error of log projected mortality rates for SC males and females

6.1.2.2 Projections

Figures 6.9, 6.10 and 6.11 plot the mortality projections of the joint model for male-and-female pairs of each country. The joint model produces much more reasonable long term projections, especially for SC males and females, when compared to projections obtained from either the independent models or the SC male-and-female joint sex model. Recall in Section 5.4.2 when only the SC males and females are jointly modelled (joint sex model), even with the incorporation of expert opinion, the male mortality projections still fall below that of females at some ages within the data range (Figure 5.16a). Whereas when all four of them are modelled jointly, this does not happen. The projections of EW males and females are very similar to that obtained from the EW male-and-female joint sex model (Section 5.4.2) due to the fact that the cross-country penalties exerts less influence on EW populations, as they have higher exposures. The projected SC mortality schedules are also more reasonable especially at older ages when compared to either the single population models (dotted lines) or the SC male-and-female joint sex model (Figure 5.14) where male and female intersections are still obtained.



(a) EW males and females

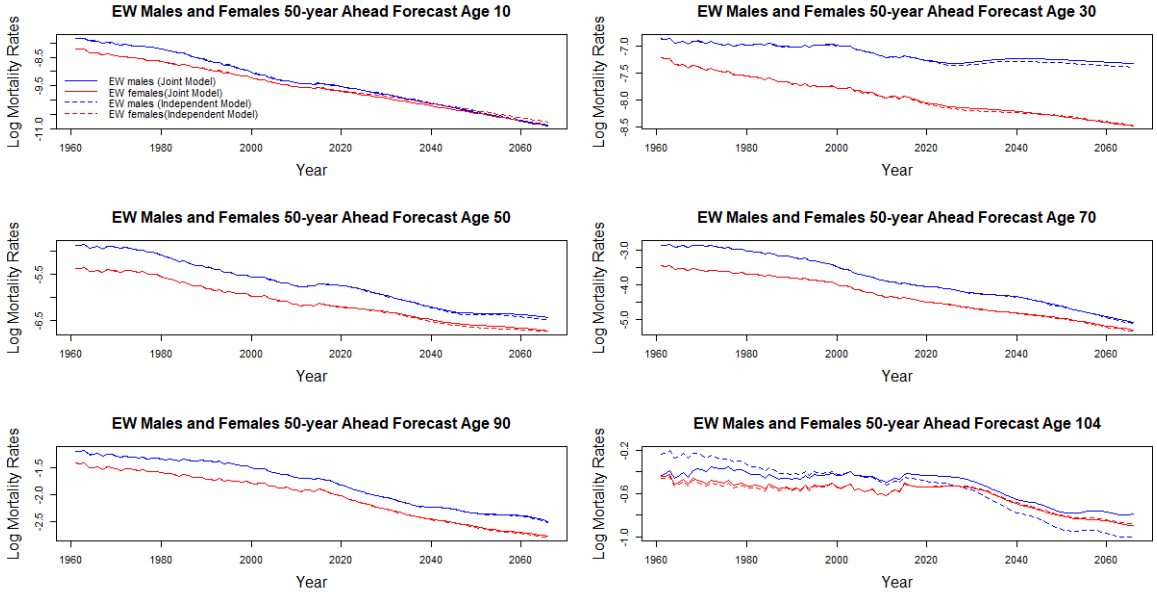


(b) SC males and females

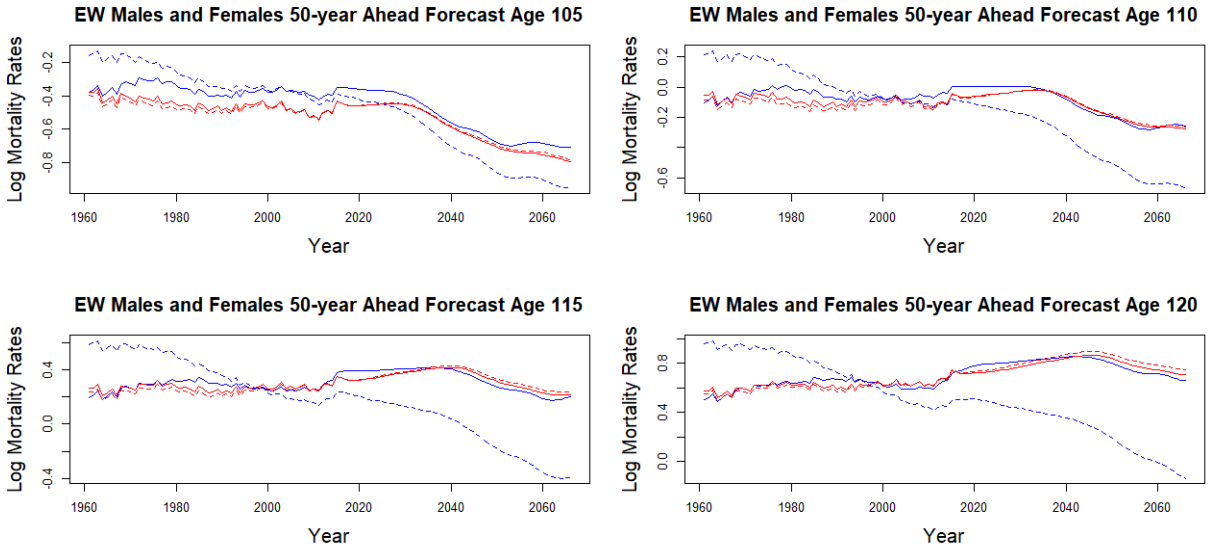
Figure 6.9: The projected mortality schedules in 1, 25 and 50 years ahead. The blue, red, black and green lines are the projections of EW males, EW females, SC males and SC females, respectively while the solid and dotted lines correspond to the projections of the joint model and the independent models respectively.

The 50 years ahead mortality projections for EW males and females (Figure 6.10) are similar to those obtained from the EW male-and-female joint sex model (Figure 5.15). The 50 years ahead mortality projections for SC males and females (Figure 6.11) are more plausible compared to those obtained from either the single population models (dotted lines) or the SC male-and-female joint sex model (Figure 5.16). Both the single population models and the SC male-and-female joint sex model produce male and female mortality forecasts that intersect and diverge

at most ages. However, in the four-population joint model the projected male mortality rates always stay above the projected female mortality rates at all ages within data range within the forecast horizon. At the extrapolation age range, the four-population joint model produces SC male and female forecasts that are much closer to each other, this is due to the fact that the SC male and female baseline mortality schedules are now learning from the diminishing gap in age between EW male and female mortality rates.

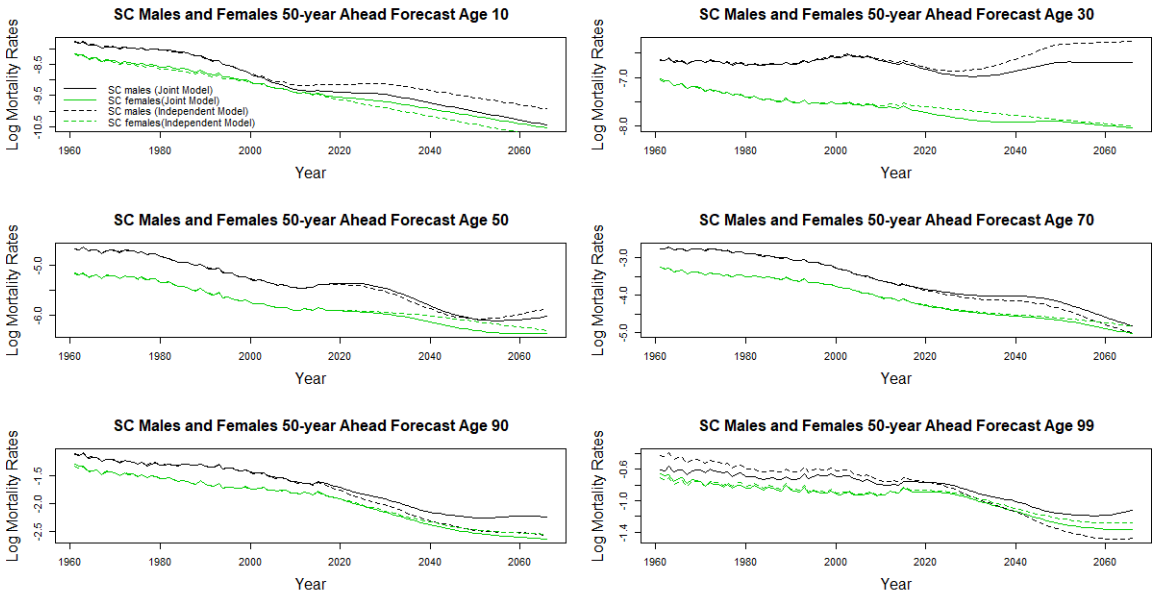


(a) age 10, 30, 50, 70, 90 and 104

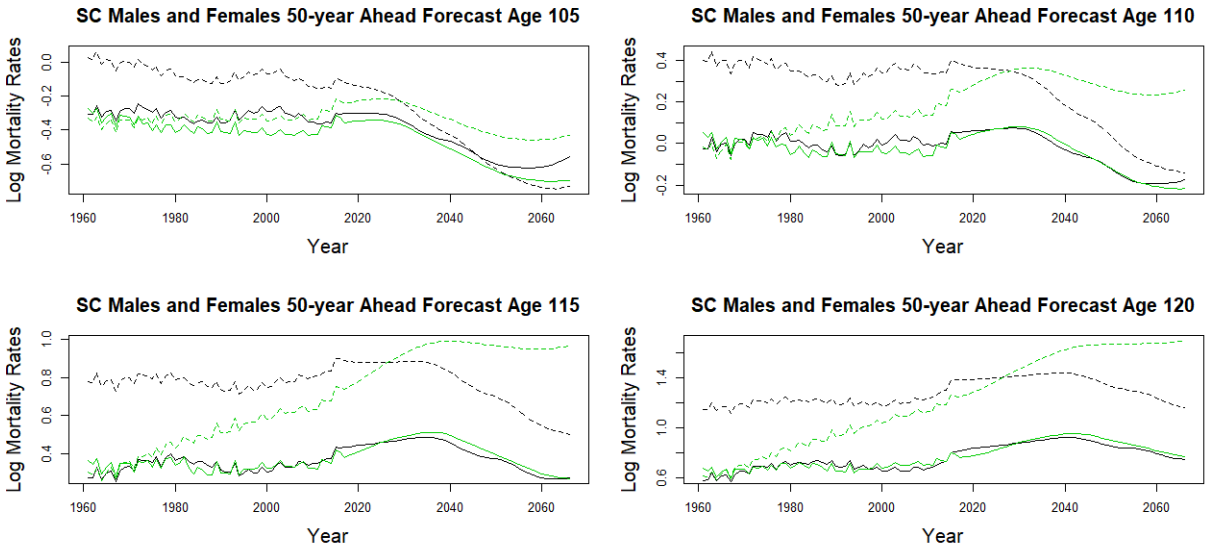


(b) age 105, 110 115 and 120

Figure 6.10: 50 years ahead projections of EW male (blue) and female (red) mortality rates at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.



(a) age 10, 30, 50, 70, 90 and 99

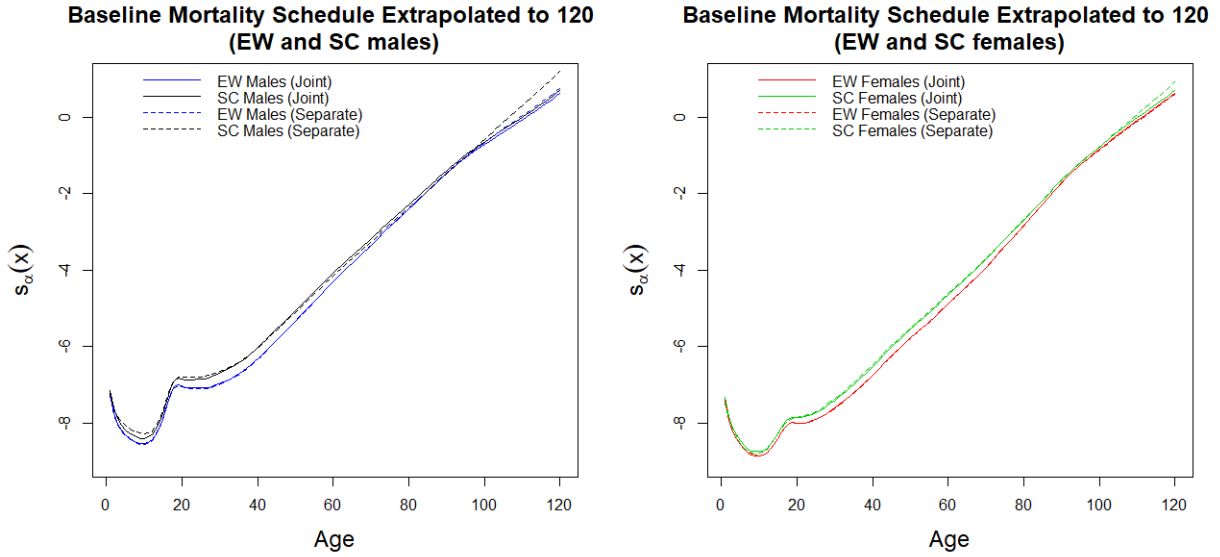


(b) age 105, 110 115 and 120

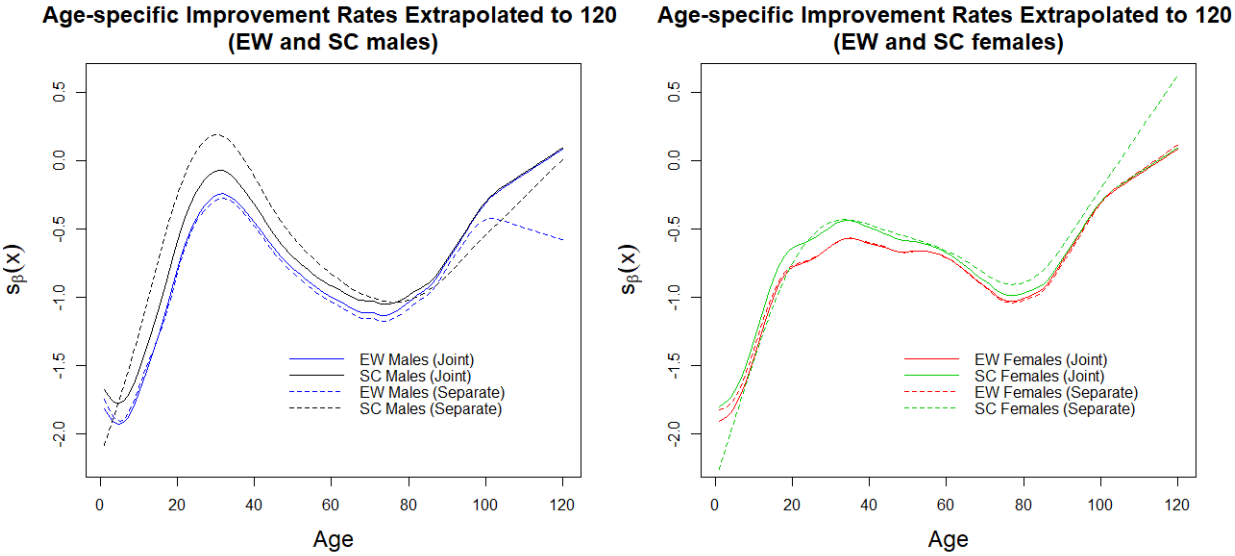
Figure 6.11: 50 years ahead projections of SC male (black) and female (green) mortality rates at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.

6.1.2.3 Expert Opinions

The expert opinion is incorporated in the way described before in Sections 5.4.2.3 and 5.4.3.3. After the incorporation of expert opinion, the long term forecasts are strictly non-divergent as the improvement rates eventually tend to a common value. However, intersections of male and female mortality rates still occurred at the oldest extrapolated ages in the long run.



(a) estimated baseline mortality schedules for EW-and-SC



(b) estimated age-specific improvement rates for EW-and-SC

Figure 6.12: The estimated baseline mortality schedules and improvement rates extrapolated to age 120 for each sex. The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.

6.1.3 Pairwise Comparison - England Wales and Scotland

In this section the EW-and-SC pairs are inspected. Figure 6.12 plots the pairwise estimated baseline mortality schedules and improvement rates for males and females. The estimated SC terms are more consistent with the EW estimates under the joint model. The cross-country penalties have less impact on the EW, due to it being a larger population.

6.1.3.1 Backtesting

Tables 6.6 and 6.7 show the MAE of the male and female populations respectively. The projections for males have improved all but the youngest age group for both the AR(1) and ARIMA(1,1,0), while the female projections have seen only a modest improvement in age groups over 60.

	EW and SC males			
Age	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	1.402152	1.304688	1.414957	1.307595
31 to 60	0.7016268	0.7143507	0.7087438	0.7117318
61 to 90	0.4509867	0.4829384	0.4708807	0.4936692
91+	0.1624962	0.183911	0.1730395	0.192056
all ages	0.773514	0.7603379	0.7864768	0.7645174

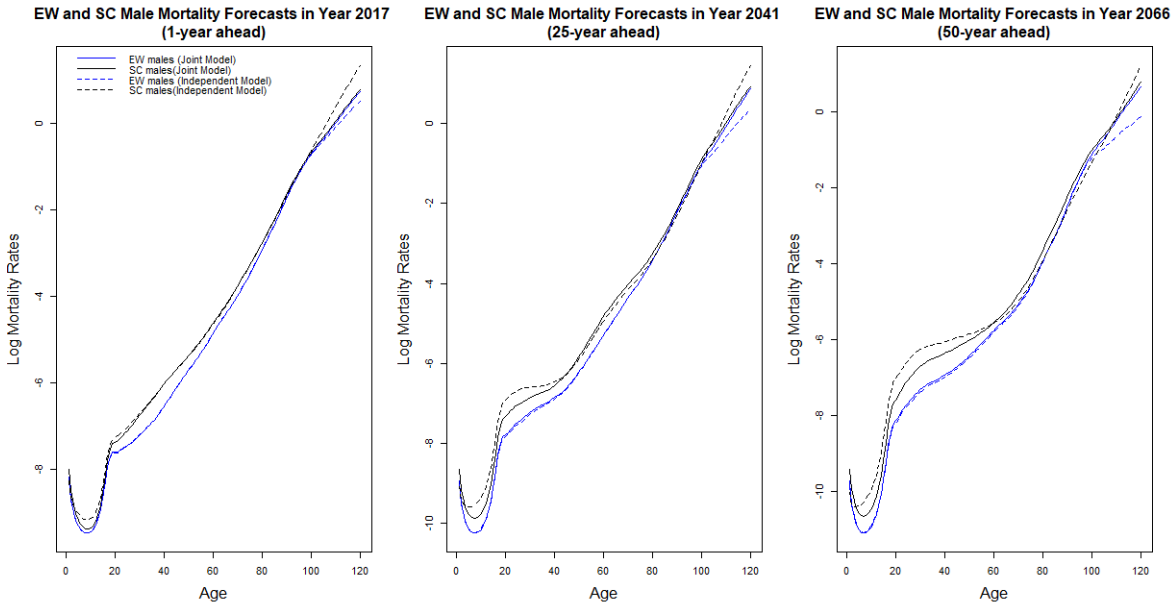
Table 6.6: The mean absolute error of log projected mortality rates for EW and SC males

	EW and SC females			
Age	MAE of Joint Model AR(1)	MAE of Independent Model AR(1)	MAE of Joint Model ARIMA(1,1,0)	MAE of Independent Model ARIMA(1,1,0)
1 to 30	1.704038	1.692998	1.711517	1.700456
31 to 60	0.9533941	0.9476961	0.9558851	0.9492924
61 to 90	0.5624422	0.5655666	0.5713268	0.5737415
91+	0.120942	0.1259898	0.1152703	0.1195462
all ages	0.9653899	0.9619378	0.97032	0.9663003

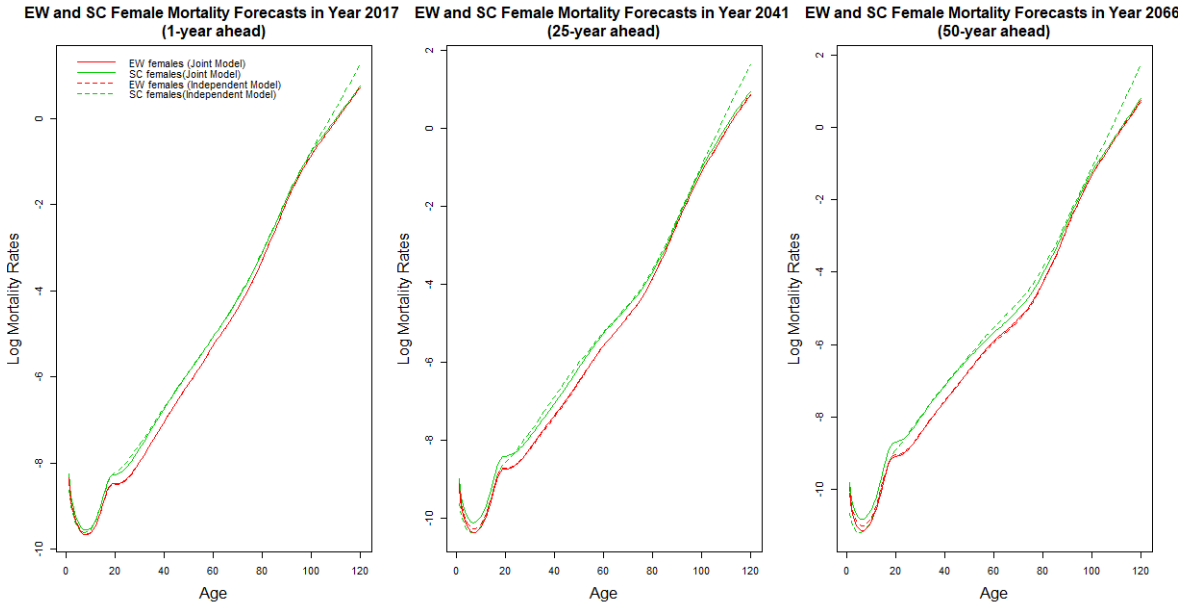
Table 6.7: The mean absolute error of log projected mortality rates for EW and SC females

6.1.3.2 Projections

Figure 6.13 plots the projected mortality schedules for each sex. The 50 years ahead projections of the EW-and-SC pairs are presented in Figures 6.14 and 6.15. The joint model is able to produce non-divergent projections at extrapolation range. The estimates for EW females between the joint and independent model are very similar, possibly due to the fact that it has the largest exposures among the four populations. The estimates for EW males between the two models are also quite similar within the data range (age 1 to 104), whereas estimates for SC males and females are more variable.



(a) EW males and females

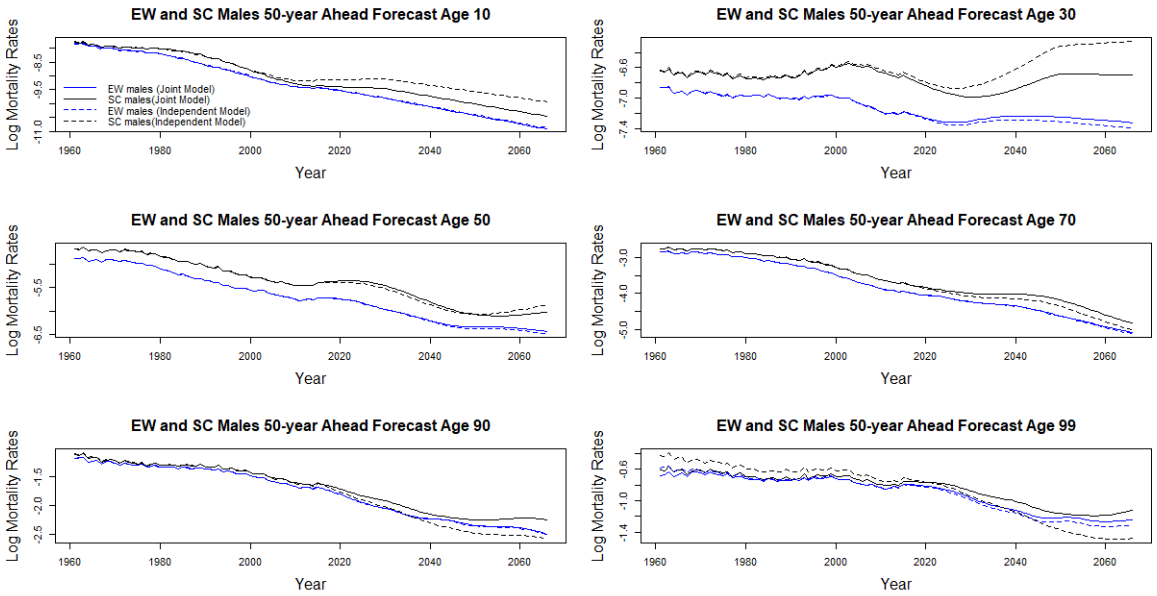


(b) SC males and females

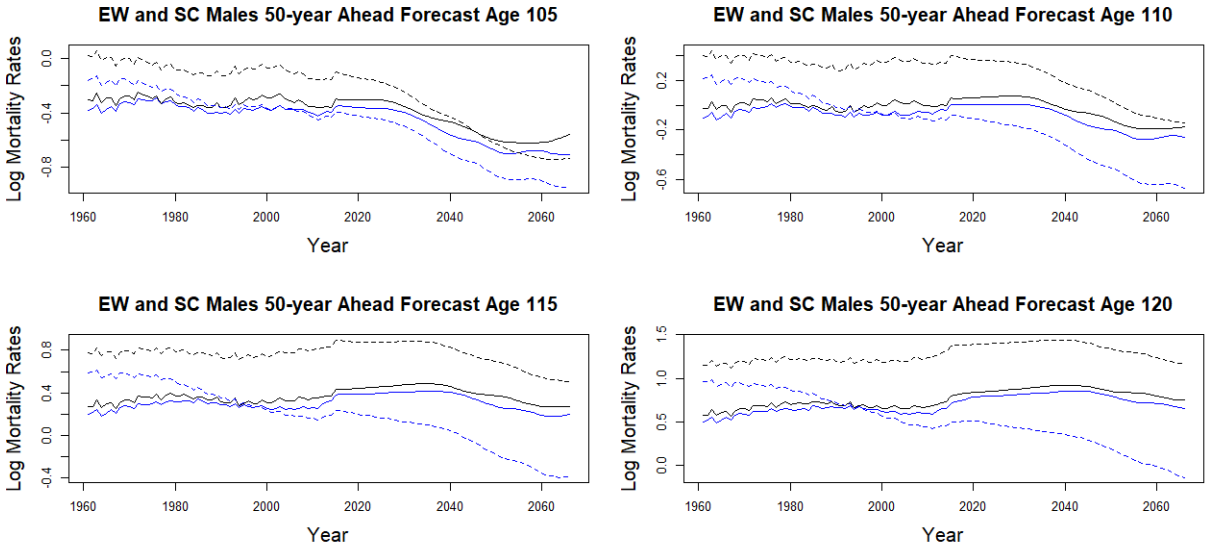
Figure 6.13: The projected mortality schedules in 1, 25 and 50 years ahead. The blue, red, black and green lines are the projections of EW males, EW females, SC males and SC females, respectively while the solid and dotted lines correspond to the projections of the joint model and the independent models respectively.

6.1.3.3 Expert Opinions

The expert opinion is incorporated in the way described before in Sections 5.4.2.3 and 5.4.3.3, similar conclusion can be drawn. After the incorporation of expert opinion, the long term forecasts are strictly non-divergent as the improvement rates eventually tend to a common value.

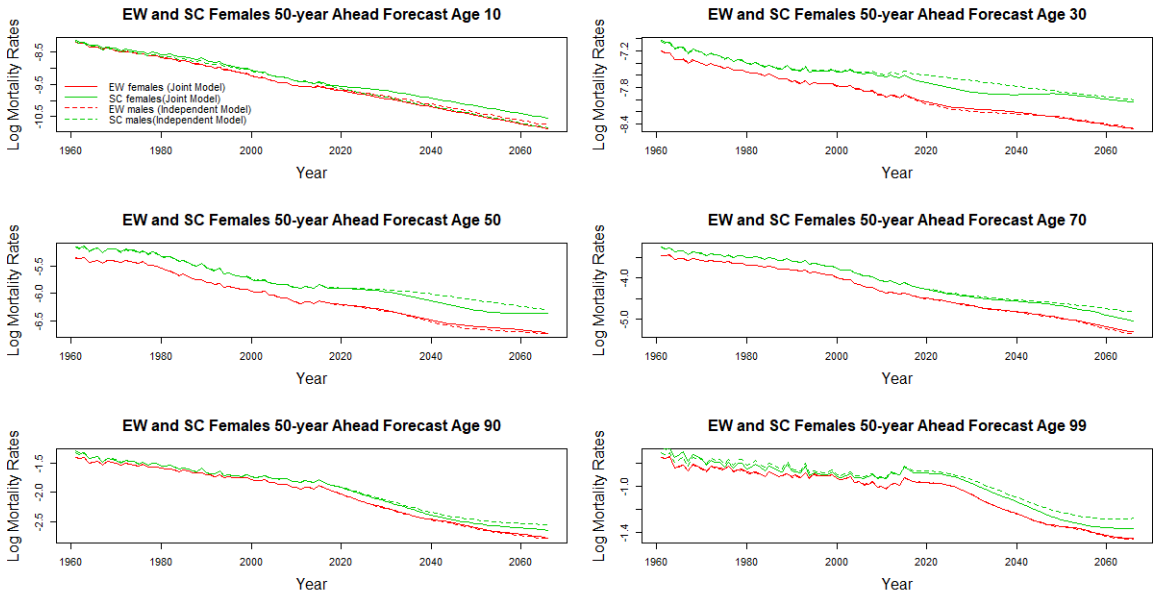


(a) age 10, 30, 50, 70, 90 and 99

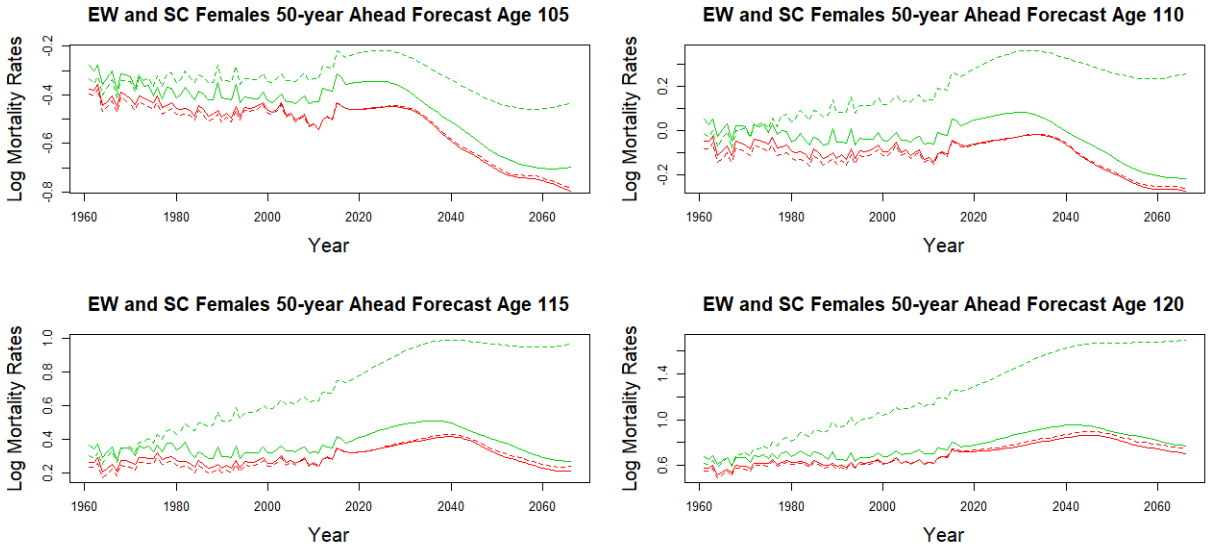


(b) age 105, 110 115 and 120

Figure 6.14: 50 years ahead projections of EW (blue) and SC (black) male mortality rates at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.



(a) age 10, 30, 50, 70, 90 and 99



(b) age 105, 110 115 and 120

Figure 6.15: 50 years ahead projections of EW (red) and SC (green) female mortality rates at selected and extrapolated ages along the time horizon. The solid and dotted lines correspond to estimates of the joint model and the independent models respectively.

6.2 Conclusion

In Chapter 5, a joint sex model and a joint country model are introduced. It is shown that the joint sex model is able to produce much more reasonable long term male and female mortality projections that are non-intersecting, a property that single population models often fail to achieve. Information is borrowed at the highest ages where exposures are small. The EW and SC joint country models provide a way for the smaller SC population to borrow strength and learn from the bigger EW population. By jointly modelling the two countries, estimation and extrapolation of the SC baseline mortality schedules and age-specific improvement rates make use of information provided by the EW population, which has a wider data range, resulting in more plausible estimates. In addition, the joint country model also largely reduces or even eliminates the problem of producing worsening mortality of the SC populations when they are modelled independently. The joint country model is able to produce non-divergent long term projections between the countries for both males and females. The forecast accuracy has improved for both males and females except for the younger age groups.

In this chapter, a joint model for all four populations, i.e. EW males, EW females, SC males and SC females, is proposed, which combine features of both the joint sex and joint country models and borrow strength across sexes and across countries. To simplify estimation process and further allow exchange of information among populations, a common cross-sex penalty is assumed for the male-and-female pairs, regardless of the country; and a common cross-country penalty is assumed for the EW-and-SC pairs, regardless of the sex. Common smoothness penalties are also assumed for the same sex of each country. For each male-and-female pair, the estimated baseline schedules and the mortality improvement rates tend to a common value. The SC male-and-female pair is able to utilize information offered by the EW populations especially when extending to ages beyond the data range. The joint model is able to produce projections that are non-divergent and without intersections of males and females trends. The overall MAE has decreased in all but the youngest age group, meaning that the forecast accuracy has improved, regardless of the time series process used to forecast the period effects (AR(1) or ARIMA(1,1,0)). A possible amendment maybe to have a locally varying cross-country penalty as discussed before. A finer breakdown of the MAE reveals that the improvement in the forecast accuracy mainly comes from the SC populations, as the MAE of the EW male-and-female pair has slightly worsened. When the same sex pairs are considered, the EW and SC males pair has seen larger improvement in the forecast accuracy than the EW-and-SC females pair. In light of the MAE results, a possible improvement maybe to first jointly model EW males and females, and then jointly model the SC males and females together with the results obtained from the EW joint model.

Chapter 7

Bayesian Mortality Projection

The classical approach we used so far has some common aspects with the Bayesian approach. Specifically, the distributional result of the estimated spline coefficients stems from a Bayesian point of view of the smoothness penalty (Section 2.2.6). In addition, as mentioned in Section 5.3, quantifying the forecast uncertainty is not so straightforward. We relied on bootstrapping in order to obtain estimates of the quantiles. A more natural way to incorporate all uncertainty into the forecasts would be to use a fully Bayesian approach.

7.1 Prior Specifications

The smoothness penalty can be viewed as a particular prior on the spline coefficients as shown in Section 2.2.6, a prior that has higher density for smooth functions. To see this, consider the penalised log-likelihood $\ell(\boldsymbol{y}) - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{D}_\lambda\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the vector of coefficients and $\boldsymbol{D}_\lambda$ is the penalty matrix that depends on the smoothing parameter(s) λ. This is equivalent to having a multivariate normal prior on the spline coefficients, $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}_\lambda^{-1})$. Thus the penalty matrix can be viewed as a precision matrix, the higher the penalty, the more concentrated is the prior on smooth functions. This is an improper prior due to the fact that the penalty matrix $\boldsymbol{D}_\lambda$ is often rank deficient. For example, with a squared second difference penalty, the penalty matrix is lacking two ranks that correspond to the null space of constants and linear functions, as these are deemed 'completely smooth' under the squared second difference penalty (their roughness is zero measured by the penalty). Different approaches have been taken to circumvent the rank deficiency. For instance, for simple splines with squared second difference penalty, the improper multivariate normal prior can be expressed in terms of a second order random walk, with the first two coefficients having an improper uniform prior (Lang and Brezger, 2004). The impropriety means that the normalising constant calculated will always be zero, since the determinant of the rank deficient penalty matrix is always zero. Yue et al. (2012) calculated the normalising constant by only multiplying the non-zero eigenvalues of the penalty matrix, $|\boldsymbol{D}_\lambda|_+$, which is also equivalent to having improper uniform priors on the null space. Intuitively, another solution

is to complete the rank, by giving some mass to the null space. This can be done by introducing a small penalty on the null space. Following Wood (2016), let $\boldsymbol{U}$ be columns of eigenvectors of the penalty matrix with zero eigenvalues, then $\boldsymbol{UU'}$ is the un-penalised null space. Therefore, to complete the rank, a small penalty is added to this null space, i.e. adding $c\boldsymbol{UU'}$ to the penalty with some small constant c. In other words, the eigenvalues for these vectors are now no longer zero, but c. We have taken this approach as it is the most straightforward to implement in Stan software. Here we choose c to be 0.001. Bayesian adaptive splines have also been investigated. Baladandayuthapani et al. (2005) demonstrated the adaptive penalty using a polynomial spline basis (albeit in the paper the author used the misleading and confusing term of 'P-spline' as a synonym for penalised spline) for homoskedastic data, where they assumed normal priors on the coefficients each with their own variances, which are themselves realisations of another spline. Crainiceanu et al. (2007) then extended the approach to heteroskedastic data. Scheipl and Kneib (2009) did not assume any underlying smoothness on the prior variances across the domain and proposed piece-wise constant variances with multiple levels of hyperprior. Jullion and Lambert (2007) on the other hand studied the robustness of the hyperprior on non-adaptive peanlised splines and showed that the usual assumption of a Gamma hyperprior on the smoothness parameter (i.e. precision of the prior on spline coefficients) induces potential impact on the posteriors. The authors then suggested placing certain hyperpriors on the Gamma parameters as well as using a mixture prior on the smoothness parameter. In the following we use bounded proper uniform priors on all smoothness-related parameters.

To help with the identifiability problem, the same constraints are applied as in the classical approach, only now they are expressed in terms of conditional distributions on the priors of the period and cohort effects. Specifically, let $\boldsymbol{T}$ and $\boldsymbol{C}$ be matrices such that

$$\boldsymbol{\kappa}^\star = \boldsymbol{T\kappa} = \boldsymbol{T}\begin{pmatrix}\kappa_1\\ \kappa_2\\ \kappa_3\\ \vdots\\ \kappa_T\end{pmatrix} = \begin{pmatrix}\sum \kappa_i\\ \sum i\kappa_i\\ \kappa_3\\ \vdots\\ \kappa_T\end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\gamma}^\star = \boldsymbol{C\gamma} = \boldsymbol{C}\begin{pmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\\ \vdots\\ \gamma_k\end{pmatrix} = \begin{pmatrix}s_\gamma(1)\\ \sum s_\gamma(i)\\ \kappa_3\\ \vdots\\ s_\gamma(n_c)\end{pmatrix},$$

where $s_\gamma(1)$ and $s_\gamma(n_c)$ are the first and last cohort effect. Then the constraints are equivalent to conditioning the prior distributions receptively, i.e. $\boldsymbol{\kappa}|_{\sum \kappa_i, \sum i\kappa_i = 0}$ and $\boldsymbol{\gamma}|_{s_\gamma(1), s_\gamma(n_c), \sum s_\gamma(i) = 0}$.

Therefore for the single population models we have the following priors,

$$\begin{aligned}
\boldsymbol{\alpha} &\sim \mathcal{N}(\mathbf{0}, (\boldsymbol{P}'\boldsymbol{W}^{\alpha}\boldsymbol{P} + \boldsymbol{P}_{null})^{-1}), \\
\boldsymbol{\beta} &\sim \mathcal{N}(\mathbf{0}, (\lambda^{\beta}\boldsymbol{P}'\boldsymbol{P} + \boldsymbol{P}_{null})^{-1}), \\
\boldsymbol{\kappa}|_{\sum \kappa_i, \sum i\kappa_i = 0} &\sim \mathcal{N}(\mathbf{0}, \sigma_{\kappa}^2 \boldsymbol{\Sigma}_{\kappa}), \\
\boldsymbol{\gamma}|_{s_\gamma(1), s_\gamma(n_c), \sum s_\gamma(i) = 0} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma}), \\
\sigma_{\kappa}^2 &\sim \mathcal{U}(0, 100) \\
\rho &\sim \mathcal{U}(-1, 1) \\
\lambda_1^{\alpha}, \lambda^{\beta}, \lambda^{\gamma} &\sim \mathcal{U}(-50, 30) \\
\lambda_1^{\alpha} + \lambda_2^{\alpha} = \lambda_{sum}^{\alpha} &\sim \mathcal{U}(-50, 30)
\end{aligned}$$

where $\boldsymbol{P}$ is the second order difference matrix of appropriate dimension (i.e. the penalty matrix), $\boldsymbol{P}_{null}$ is the small penalty added to the null space and ρ is the time series coefficient, $\boldsymbol{W}^{\alpha}$ is a diagonal matrix with weights $\boldsymbol{W}_{ii}^{\alpha} = e^{\lambda_1^{\alpha} + \lambda_2^{\alpha} i} = e^{\lambda_1^{\alpha} + (\lambda_{sum}^{\alpha} - \lambda_1^{\alpha})i}$, $\boldsymbol{\Sigma}_{\kappa}$ and $\boldsymbol{\Sigma}_{\gamma}$ are the conditional variance matrices of the constrained period effect $\boldsymbol{\kappa}$ and cohort effect spline coefficients $\boldsymbol{\gamma}$. The re-parameterisation λ_{sum}^{α} is to prevent overflow of the smoothness penalty at the highest ages (the last entry of the diagonal of $\boldsymbol{W}^{\alpha}$ is $\lambda_{sum}^{\alpha} = \lambda_1^{\alpha} + \lambda_2^{\alpha}$ since we scale i [-1, 1]). The remaining smoothing parameters $\lambda_1^{\alpha}, \lambda^{\beta}$ and λ^{γ} are also given uniform prior between -50 and 30. Let $\boldsymbol{V}$ be the variance matrix of the unconstrained period effect, the AR(1) variance, then $\boldsymbol{\Sigma}_{\kappa}^{-1} = (\boldsymbol{T'^{-1}}\boldsymbol{V}_{\boldsymbol{\rho}}^{\boldsymbol{-1}}\boldsymbol{T^{-1}})_{[-(1,2),-(1,2)]}$ where the subscript means that the first and second rows and columns are excluded. Similarly, $\boldsymbol{\Sigma}_{\gamma}^{-1} = (\boldsymbol{C'}^{-1}(\lambda^{\gamma}\boldsymbol{P'P} + \boldsymbol{P}_{null})\boldsymbol{C}^{-1})_{[-(1,2,k),-(1,2,k)]}$ where the first, second and the last rows and columns are excluded.

7.2 Posterior Sampling

Posterior sampling is done using 'Stan' (Stan Development Team, 2020) with the R interface, the *rstan* package (Stan Development Team, 2018). Stan is a modelling language that allows full Bayesian inference using Markov Chain Monte Carlo (MCMC) sampling techniques. In particular, it uses Hamiltonian Monte Carlo (HMC) to sample from the posterior distribution. HMC is a powerful tool in the sense that it can handle correlated parameters and usually mixes relatively quickly. Each proposal step is generated from the No U-Turn Sampler (NUTS) exploring the Hamiltonian system formed by the negative log posterior and approximating the trajectory using discrete time steps (leapfrog algorithm). The NUTS is an algorithm that generates proposals by tracing out trajectory in the Hamiltonian system that doubles in length in each step, creating a binary tree. The iteration is halted when the trajectory starts to make a 'U-Turn', and the proposal is accepted with the standard Metropolis acceptance probability. As the NUTS algorithm doubles the trajectory length in each step, a large number of log-probability and gradient evaluations maybe potentially needed if the region is difficult to sample from, therefore a maximum number of steps (treedepth) is set for the NUTS algorithm, which is 10 by default in Stan.

In our practice, however, even with such a powerful tool we still experience difficulties in sampling from the posterior. The samples do not mix well with seemingly poor adaption. There were warnings about the estimated Rhat and Effective Sample Size (ESS) in Stan. In addition, most of the iterations have either resulted in divergent transitions (when the simulated trajectory departs from the true trajectory measured by Stan) or exceeded the maximum treedepth of 10 of the No U-Turn Sampler (NUTS) in Stan, which maybe an indication that the posterior contains regions with high curvatures or is very difficult to sample from. One could decrease the step size for a closer approximation to the trajectory of the Hamiltonian system, or increase the maximum treedepth allowed and see if this problem persists, however this is extremely inefficient as hitting even just one more treedepth means doubling the computation time. Instead, we try to make use of the distributional results from the classical approach (P-IRLS) and re-parameterise the model. More specifically, in Section 2.2.6, it is shown that the distribution of the estimated parameters from the classical approach is approximated by $\boldsymbol{\beta}|z \sim N((\boldsymbol{X'WX}+\boldsymbol{S})^{-1}\boldsymbol{X'}\sqrt{\boldsymbol{W}}\boldsymbol{z}, (\boldsymbol{X'WX}+\boldsymbol{S})^{-1}\phi)$, where $\boldsymbol{z}$ and $\boldsymbol{W}$ are the pseudo-data, iterative weights of the final working model in P-IRLS and ϕ is the scale parameter of the data distribution. For Poisson responses with log-link, the scale parameter ϕ is simply one and the iterative weights w are the estimated numbers of death. We set the weights to be the observed deaths plus one. During each iteration of the HMC, we let $\boldsymbol{M} = (\boldsymbol{X'WX}+\boldsymbol{S_\lambda})^{-1}$ and $\boldsymbol{L_\lambda L_\lambda'} = \boldsymbol{M}$ be the Cholesky decomposition. Here the subscript λ indicates that the matrix varies with the sampled smoothing parameters at each iteration. Then we re-parameterise $\boldsymbol{\eta}^* = \boldsymbol{L}^{-1}(\boldsymbol{\eta} - (\boldsymbol{X'WX}+\boldsymbol{S})^{-1}\boldsymbol{X'}\sqrt{\boldsymbol{W}}\boldsymbol{z})$, where $\boldsymbol{\eta}$ is the concatenated vector of parameters $\boldsymbol{\eta'} = (\boldsymbol{\alpha'}, \boldsymbol{\beta'}, \boldsymbol{\kappa'}, \boldsymbol{\gamma'})'$. Note that the Jacobian adjustment of the re-parameterisation has to be included accordingly. After the re-parameterisation, the HMC is more efficient.

7.3 Incorporation of Expert Opinion

Expert opinion is incorporated into long-term projections in a similar fashion to the frequentist approach in Chapters 5 and 6. The same function 5.13 is used to weigh between the estimated current improvement rates and the target rates over the projection horizon. In a Bayesian setting, however, the target rates are given a normal prior to reflect the uncertainty around these expert-based mortality improvement rates, i.e.

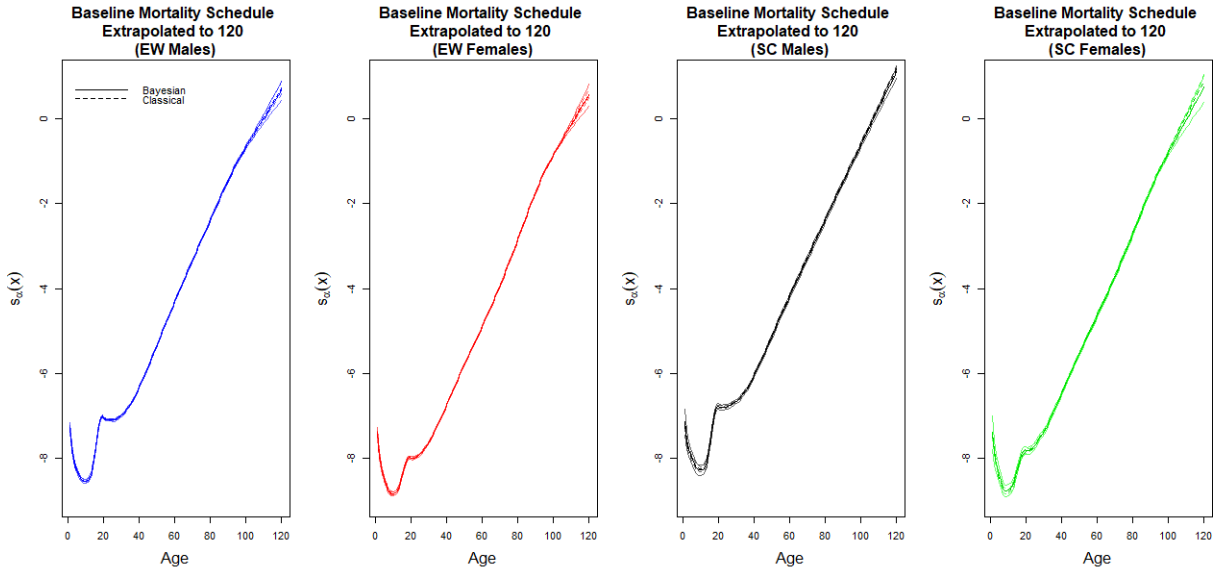
$$\beta_x^e \sim N(c_x^e, \sigma_x^e). \tag{7.1}$$

The principal mortality assumptions used in the Office for National Statistics for the 2016-based UK national population projections are that annual mortality improvement rates will converge to 1.2%. Following Dodd et al. (2020), we assume the uncertainty around these target rates $c_x^e = -0.012$ to be expressed by $\sigma_x^e = 0.006$.

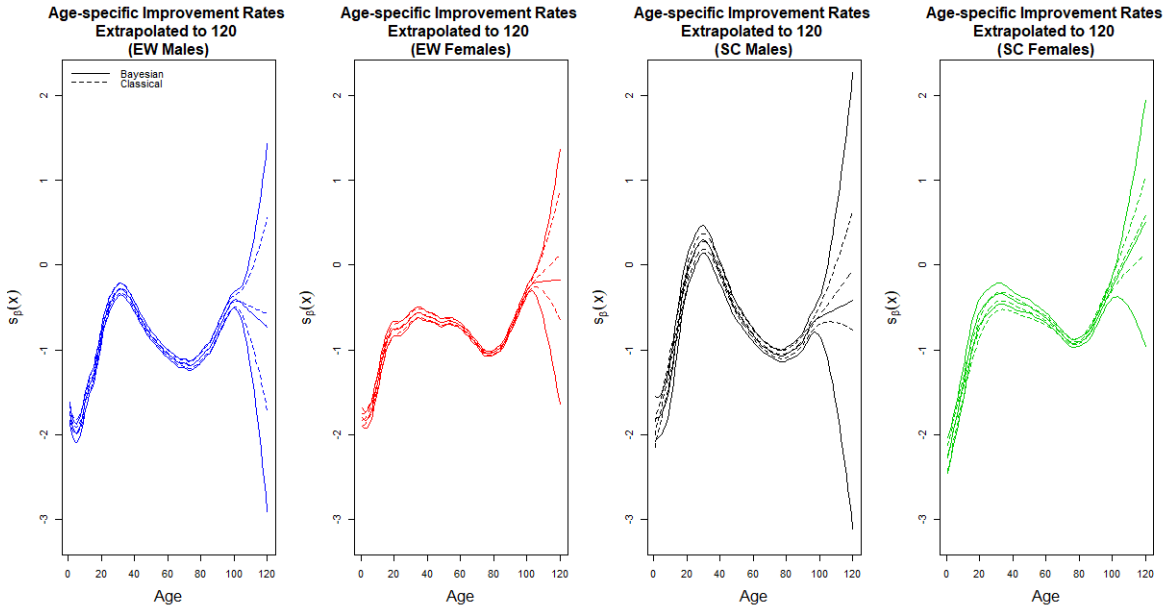
7.4 Results

7.4.1 Independent Model

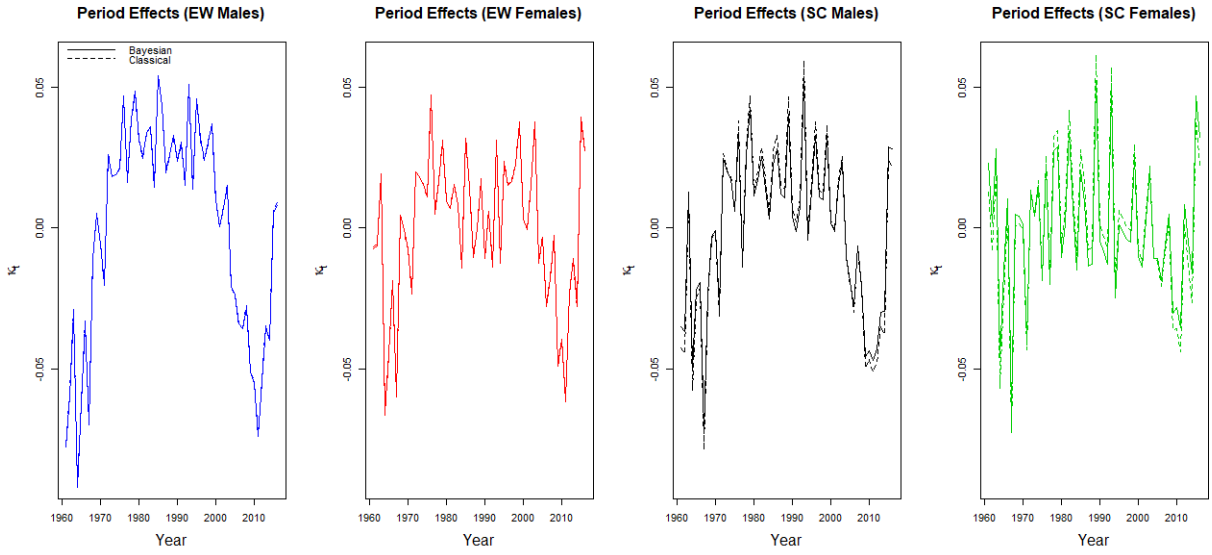
The posterior means of the parameters calculated from the Monte Carlo are similar to the frequentist point estimates. The posterior means of the smoothing parameters are also very close to that estimated from the classical approach except for SC males, where the posterior means of the smoothing parameters for the baseline in the Bayesian setting is less extreme than the frequentist estimates. Note that this specific group has the fewest data among all groups. Figure 7.1 plots the posterior means of the baseline schedules, age-specific improvement rates, period effects (interval estimates omitted for clarity) and cohort effects of EW males, EW females, SC males and SC females individually, while Figure 7.2 plots the first differences of their period and cohort effects. The solid lines plot the posterior means of the Bayesian models while the dotted lines plot the frequentist maximum likelihood estimates. The 95% credible intervals are also plotted. Clearly, the 95% intervals in the fully Bayesian model are wider due to the fact that the smoothing parameters are now themselves random variables, instead of constants. For EW males and EW females, the posterior means and the frequentist maximum likelihood estimates are very similar, except for the baseline schedules and improvement rates at the extrapolation range. For SC males and SC females, even within the data range there seems to be larger discrepancies between the posterior means from the Bayesian approach and the frequentist maximum likelihood estimates. Of course, the posterior means and intervals from the Bayesian approach and the maximum likelihood estimates from the frequentist approach are not directly comparable. The most notable change is the cohort effects of SC females, where a better mortality improvement is observed for cohorts born around 1988 in the Bayesian model. This pattern is also observed in the Bayesian model of the remaining three populations and all of the joint models in Chapters 5 and 6. Therefore we conclude that the classical single population model for SC females seems to be failing to capture this generational effect.



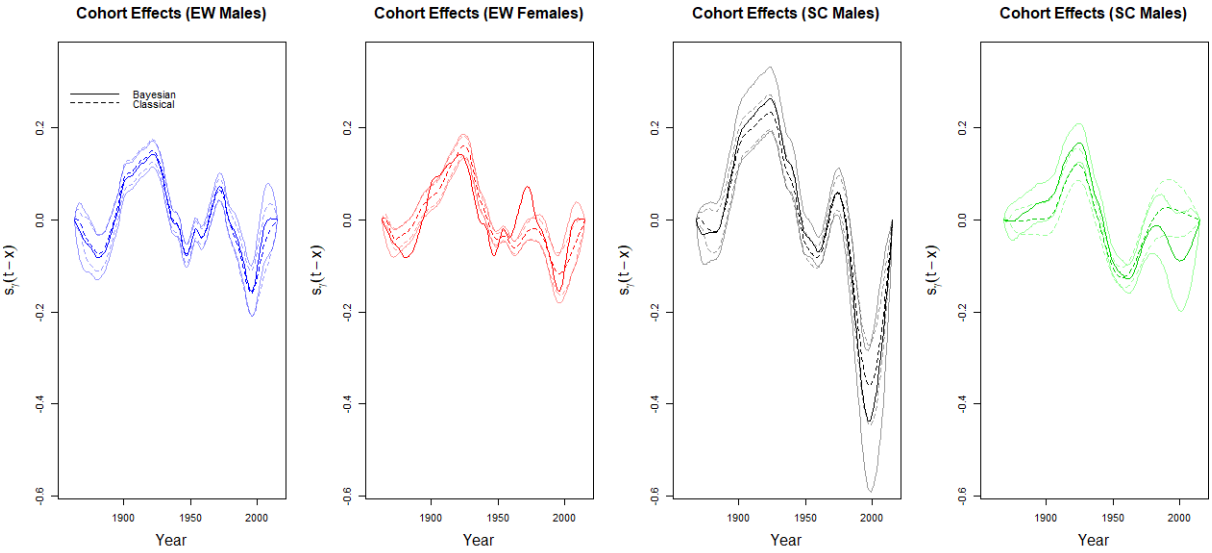
(a) baseline mortality schedules



(b) age-specific improvement rates

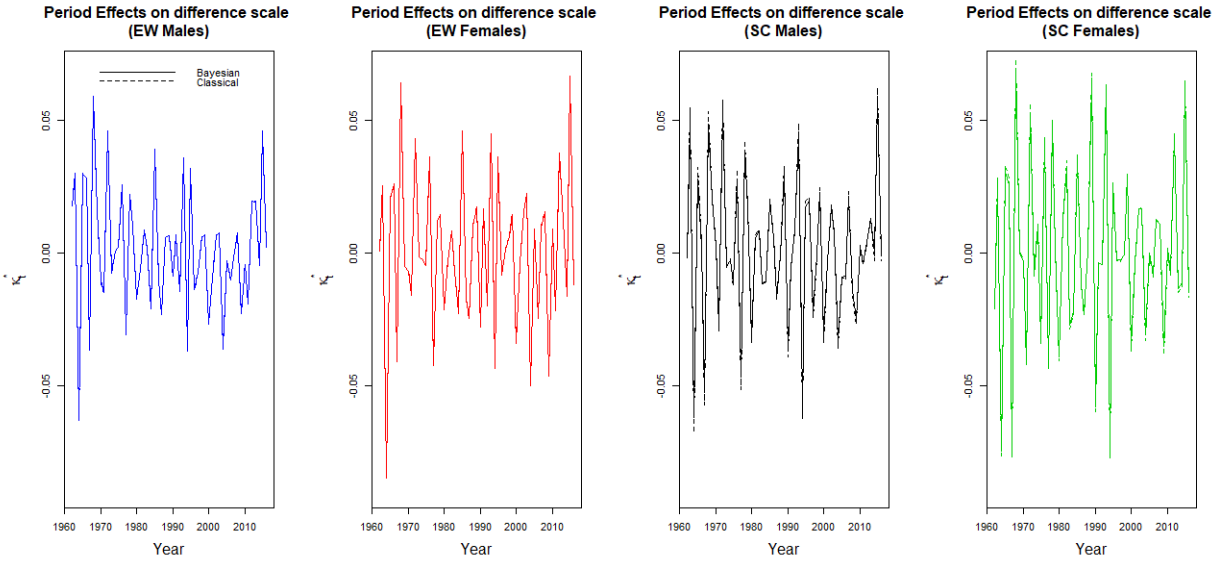


(c) period effects

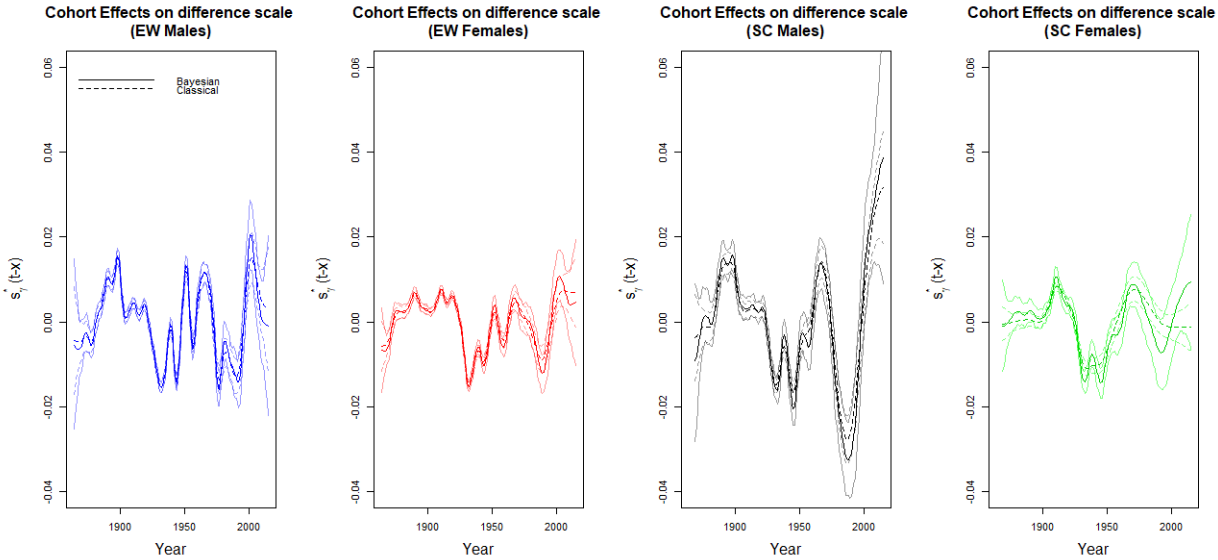


(d) cohort effects

Figure 7.1: Posterior means and 95% credible intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects (solid lines) of the Bayesian single population model for EW males, EW females, SC males and SC females. The intervals of the period effects are omitted for clarity. The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).



(a) first differences of the period effects



(b) first differences of the cohort effects

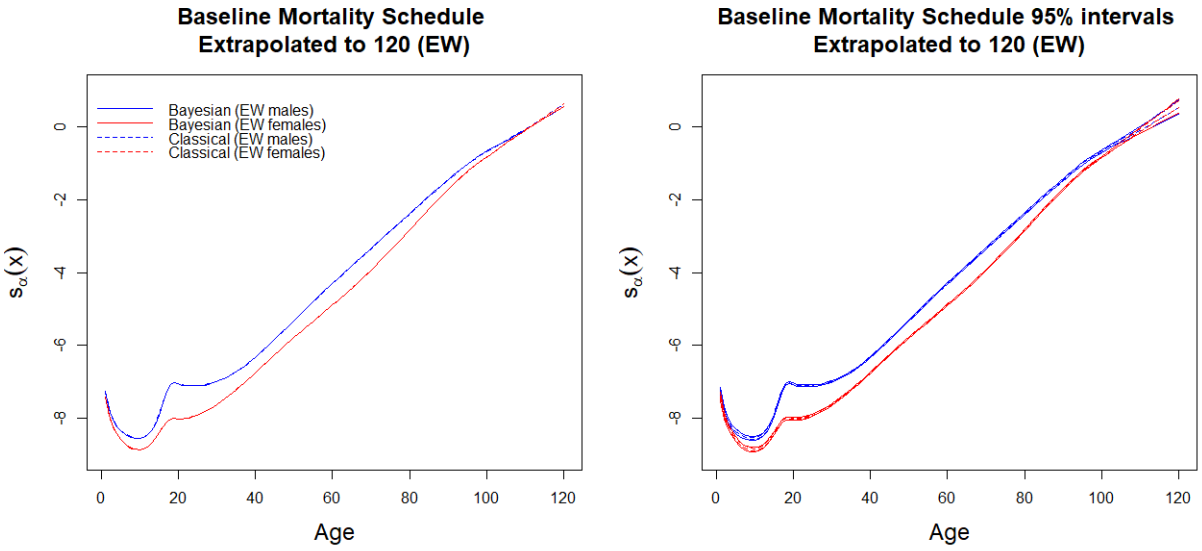
Figure 7.2: Posterior means and 95% credible intervals of the first differences of the period effects and cohort effects (solid lines) of the Bayesian single population model for EW males, EW females, SC males and SC females. The intervals of the first differences of the period effects are omitted for clarity. The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

7.4.2 Male and Female Joint Sex Model

7.4.2.1 England and Wales

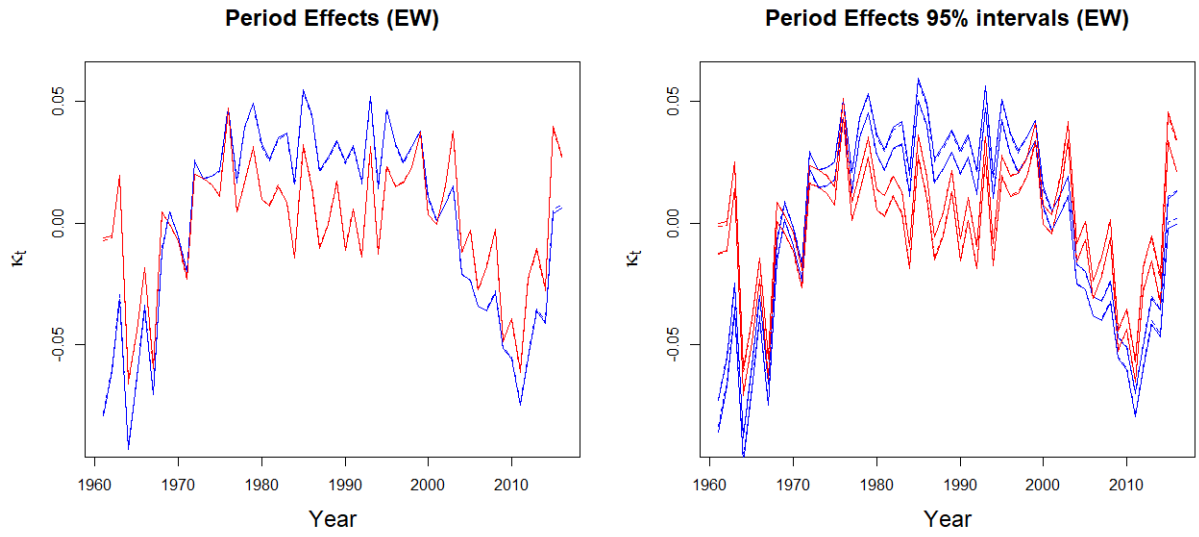
The posterior means and the 95% intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects of the EW male-and-female joint Bayesian model are shown in Figure 7.3. Within the data range, the posterior means of the parameter

estimates are very similar to the corresponding maximum likelihood estimates under the classical approach, except for the earliest and latest cohort effects, where there is only little data. The intervals of the splines are wider due to the fact that the smoothing parameters are now variables. At the extrapolated ages, the posterior means of the age-specific improvement rates from the Bayesian model are flatter, i.e. bigger mortality improvement compared to the classical approach.

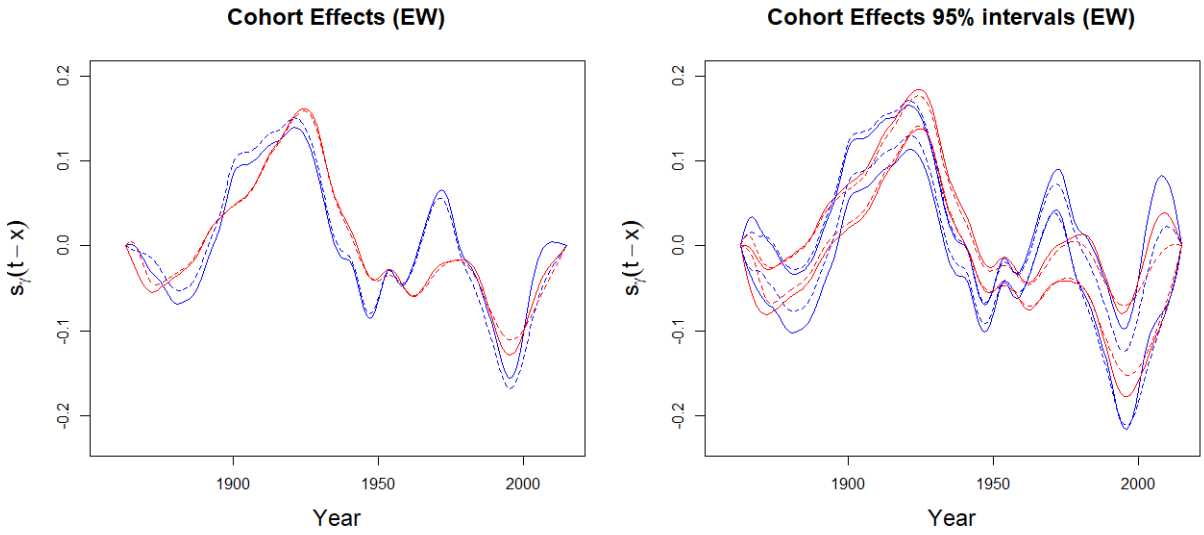


(a) baseline mortality schedules

(b) age-specific improvement rates



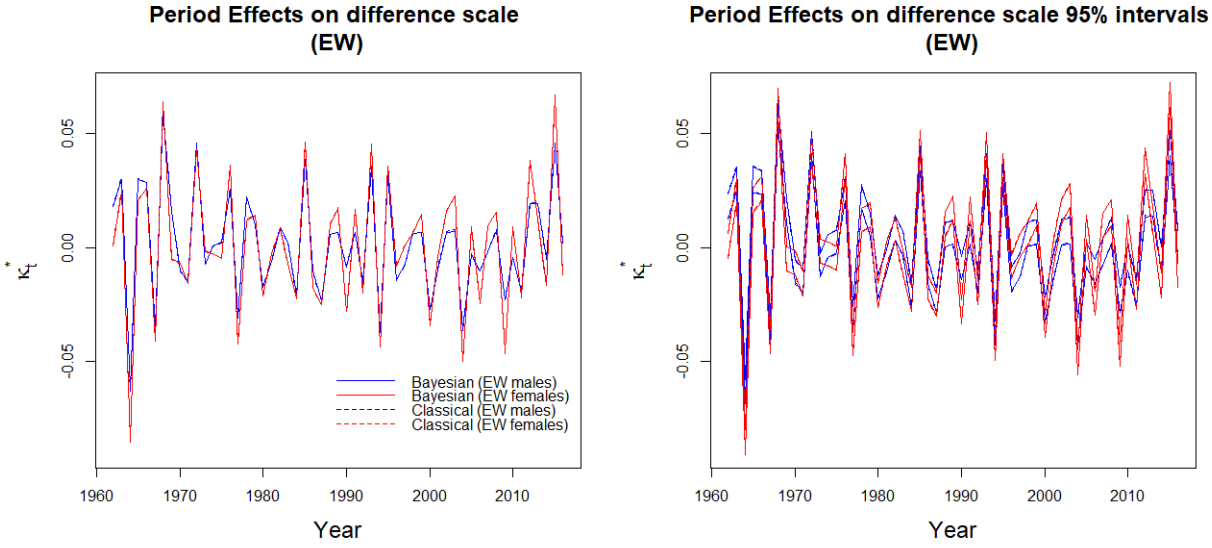
(c) period effects



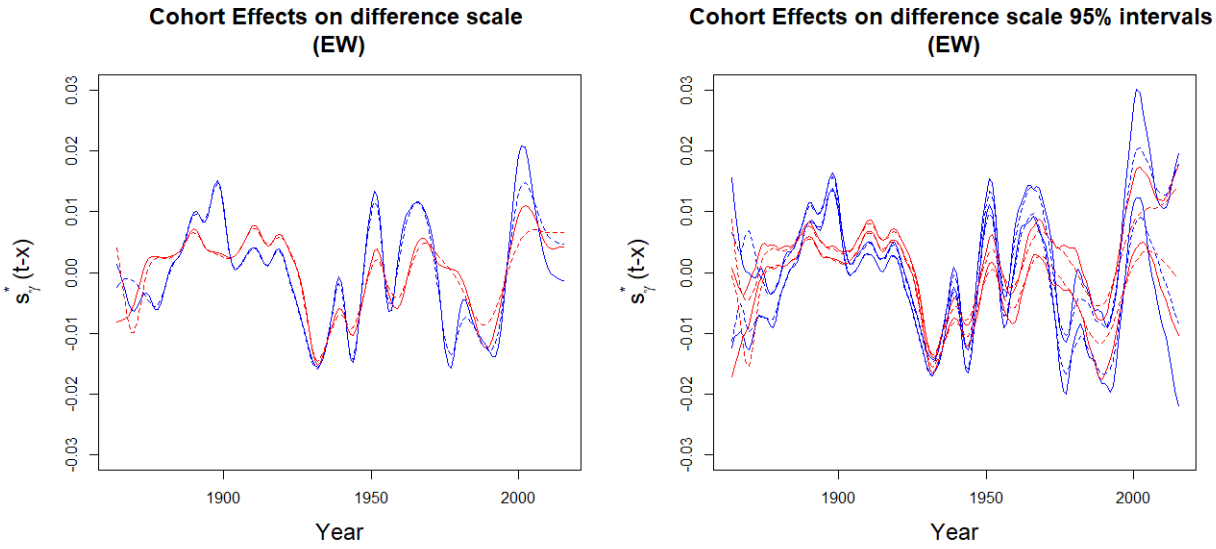
(d) cohort effects

Figure 7.3: Posterior means and 95% credible intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects (solid lines) of the Bayesian joint sex model for EW males (blue) and females (red). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

Figure 7.4 plots the posterior means of the first differences of the period and cohort effects. For the first differences of the period effects, the posterior means and the 95% intervals of the Bayesian model are almost the same as those from the frequentist approach. The posterior means and intervals of the differenced cohort effects from the Bayesian and classical approach are also similar, except at the earliest and latest cohorts.



(a) first differences of the period effects



(b) first differences of the cohort effects

Figure 7.4: Posterior means and 95% credible intervals of the first differences of the period effects and cohort effects (solid lines) of the Bayesian joint sex model for EW males (blue) and females (red). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

Figure 7.5 presents the posterior means and the 95% credible intervals of the 50-year ahead forecasts of EW male and female mortality rates at some selected ages. The maximum likelihood estimates and the interval estimates from the frequentist approach are also plotted for comparison. The posterior means are very similar to the maximum likelihood estimates obtained from the classical approach, while the 95% credible intervals are wider than the intervals from the classical approach, due to the fact that the smoothing parameters are variables instead of constants, although for EW females the intervals from the two approaches are very similar. Figure 7.6 shows the projections after incorporating expert opinion in the way discussed in Section 7.3.

The Bayesian intervals are much wider (solid lines) than the frequentist intervals (dotted lines) due to the fact that the target rates are now also given priors, instead of deterministic constants.

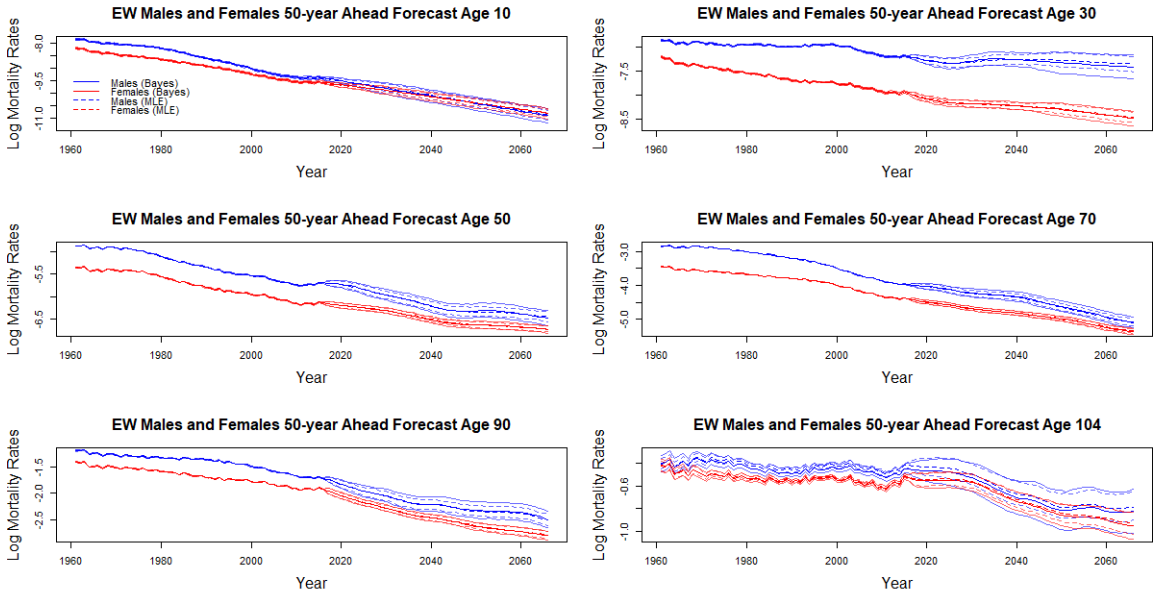


Figure 7.5: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW male and female mortality rates (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

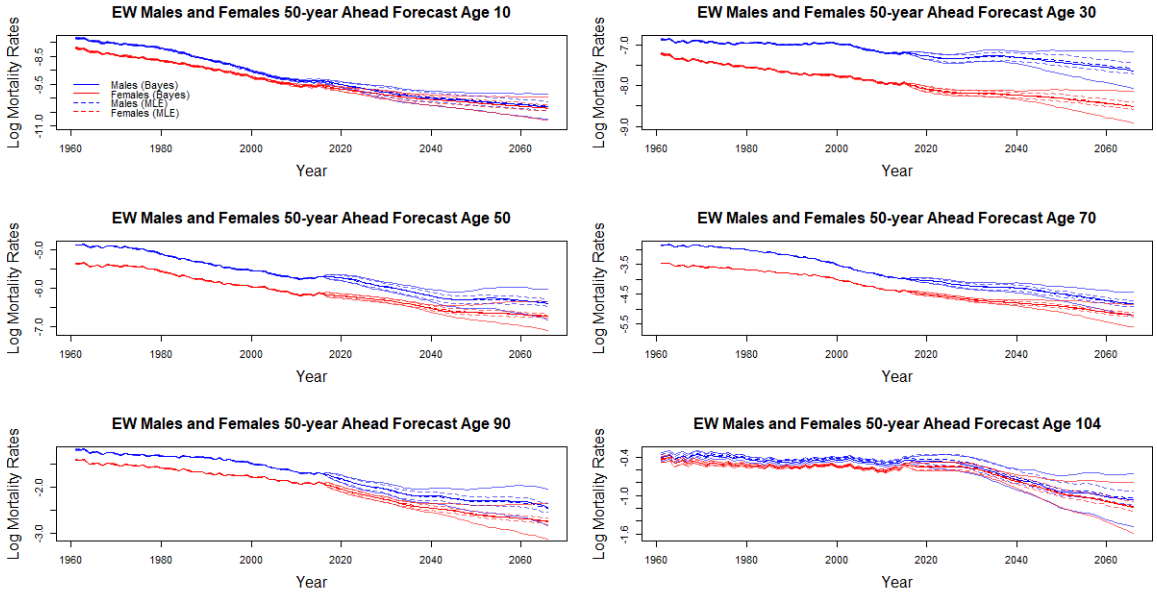
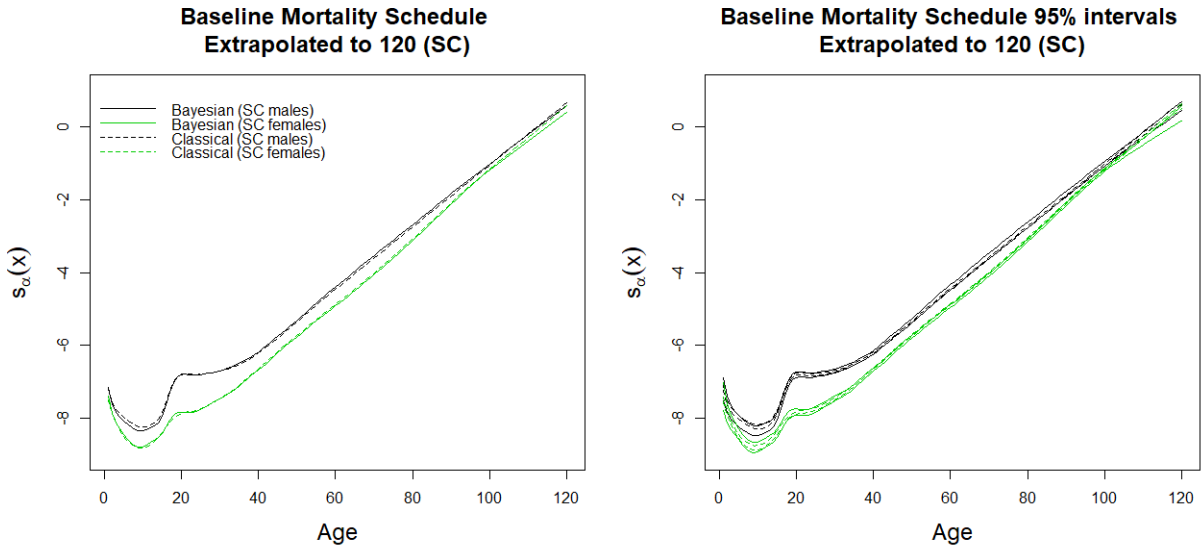


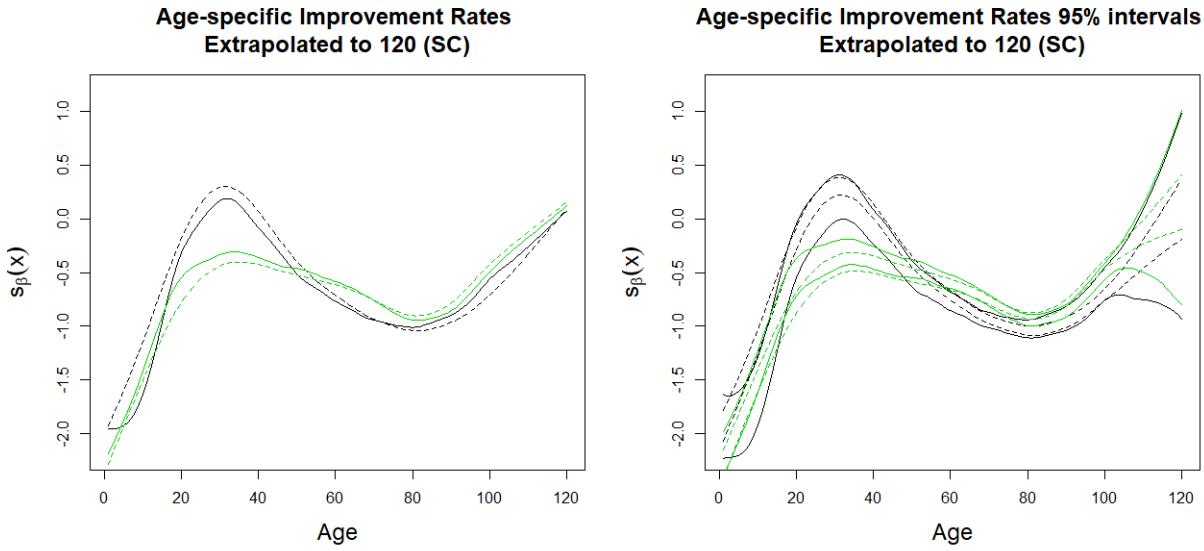
Figure 7.6: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW male and female mortality rates with the incorporation of expert opinion (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

7.4.2.2 Scotland

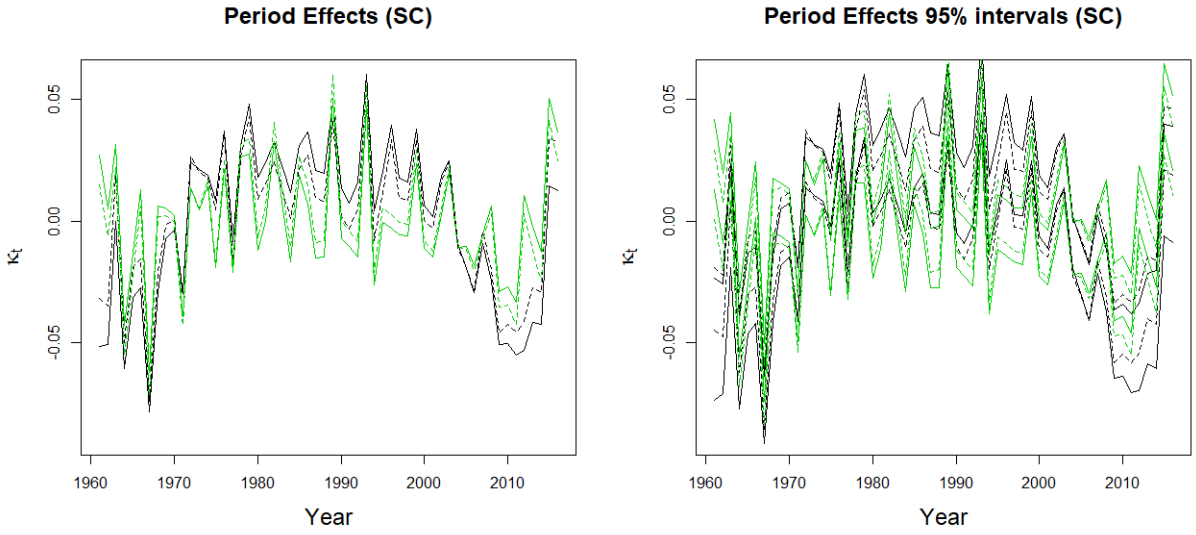
The posterior means and the 95% intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects of the SC male-and-female joint Bayesian model are shown in Figure 7.7. The posterior means of the baseline mortality schedules are very similar to the frequentist maximum likelihood estimates. However, the age-specific improvement rates show more discernible differences between the posterior means and the frequentist maximum likelihood estimates: the male and female mortality improvement rates are more similar at older ages in the Bayesian model. As in the independent models, the Bayesian joint sex model picks up a bigger mortality improvement for SC females born around 1988 that is not identified in the classical joint sex model. The posterior means of the period effects are also quite different from the maximum likelihood estimates. Nonetheless, if we inspect the first differences of the period effects (Figure 7.8), it can be seen that they are very similar in the Bayesian and classical model, meaning that period shocks to the annual mortality improvement rates are similar under the two models. The intervals of the splines are wider due to the fact that the smoothing parameters are now variables.



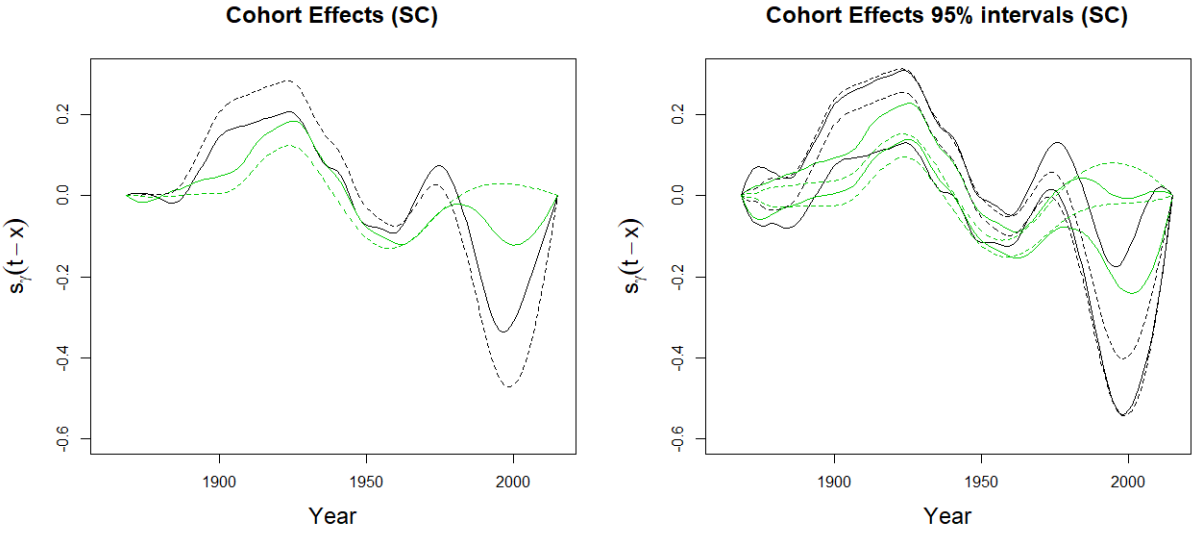
(a) baseline mortality schedules



(b) age-specific improvement rates



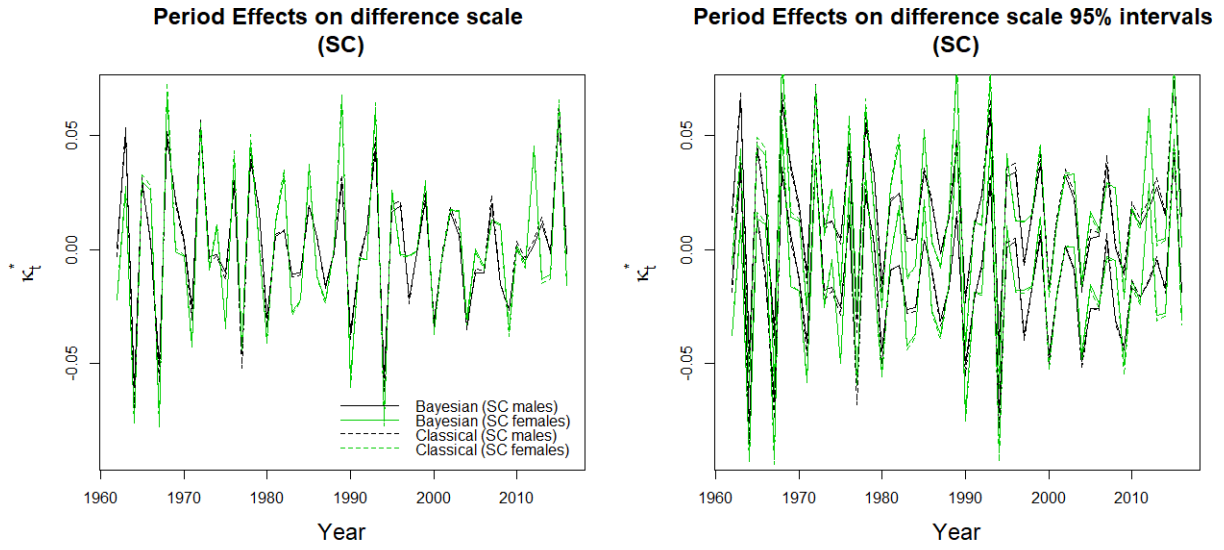
(c) period effects



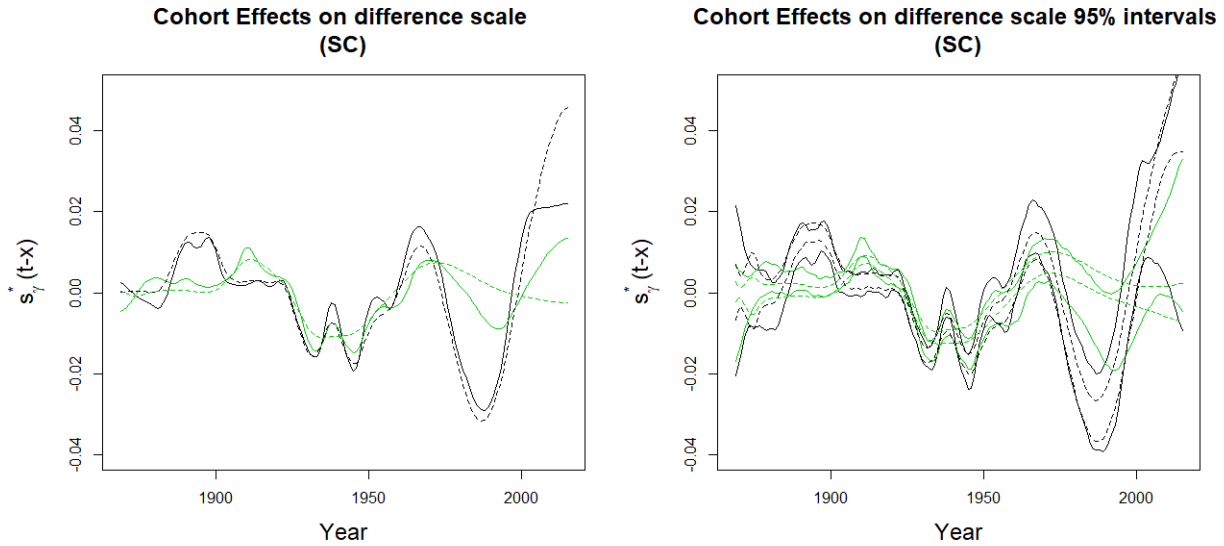
(d) cohort effects

Figure 7.7: Posterior means and 95% credible intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects (solid lines) of the Bayesian joint sex model for SC males (black) and females (green). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

The posterior means of the first differences of the period and cohort effects are shown in Figure 7.8. In the Bayesian approach, there are two distinct troughs in the posterior means of the differenced cohort effects (around years 1932 and 1945), meaning that these cohorts have better mortality improvement, while in the classical approach these are merged into one big trough.



(a) first differences of the period effects



(b) first differences of the cohort effects

Figure 7.8: Posterior means and 95% credible intervals of the first differences of the period effects and cohort effects (solid lines) of the Bayesian joint sex model for SC males (black) and females (green). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

Figure 7.9 presents the posterior means of the 50-year ahead forecasts of SC males and females and their corresponding 95% intervals. The intervals produced by the Bayesian approach are much wider. Figure 7.10 shows the projections after incorporating expert opinion. Similar to the EW joint sex model, the Bayesian intervals are wider (solid lines) than the frequentist intervals (dotted lines).

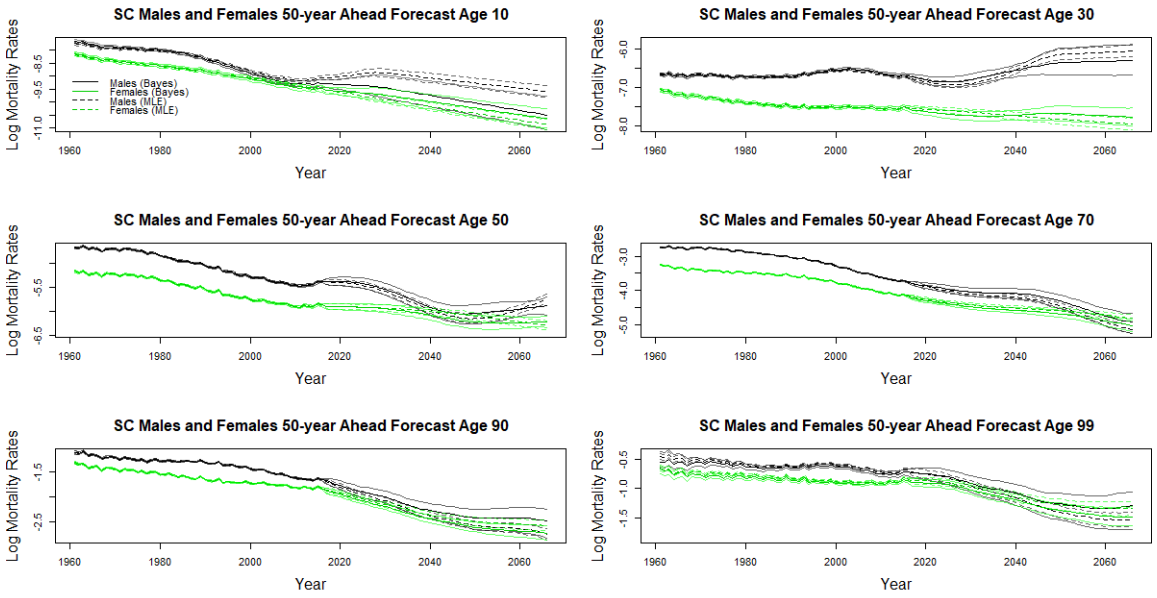


Figure 7.9: Posterior means and 95% credible intervals of the 50-year ahead forecasts of SC male and female mortality rates (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

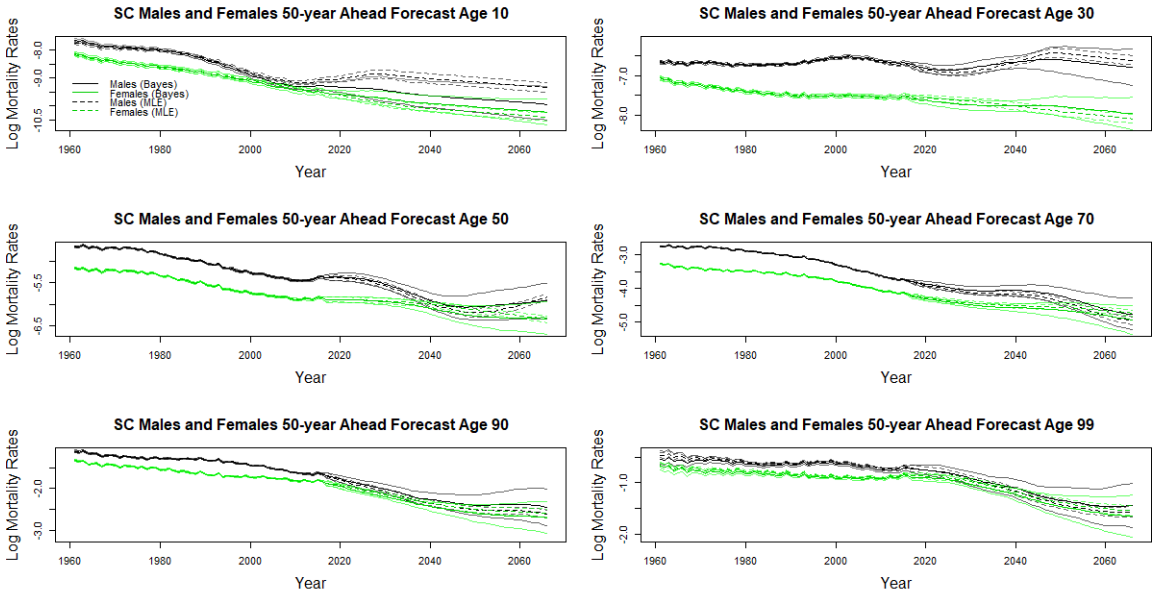


Figure 7.10: Posterior means and 95% credible intervals of the 50-year ahead forecasts of SC male and female mortality rates with the incorporation of expert opinion (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

Table 7.1 shows the empirical coverage of the estimated 95% intervals from the Bayesian approach, backtested on data in the most recent decade. The coverages of the frequentist intervals are also shown for comparison. As expected, the coverages of the intervals under the Bayesian approach are higher as the Bayesian intervals are wider than the frequentist intervals.

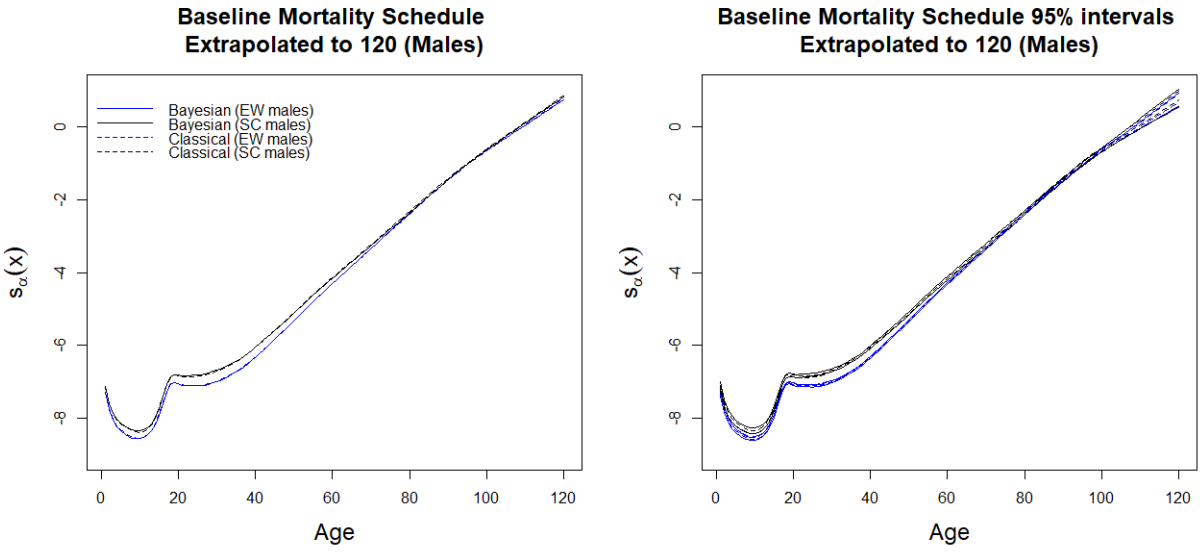
Age	95% intervals coverage							
	EW males		EW females		SC males		SC females	
	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist
all ages	0.6009615	0.4576923	0.5442308	0.4865385	0.3757576	0.302020191	0.4151515	0.331313127
61-90	0.9433333	0.7666667	0.7	0.63	0.6	0.4833333	0.54	0.4133333
91+	0.7785714	0.6285714	0.5357143	0.4571429	0.4	0.3111111	0.5777778	0.4444444

Table 7.1: The coverage of the 95% intervals of the Bayesian joint sex model

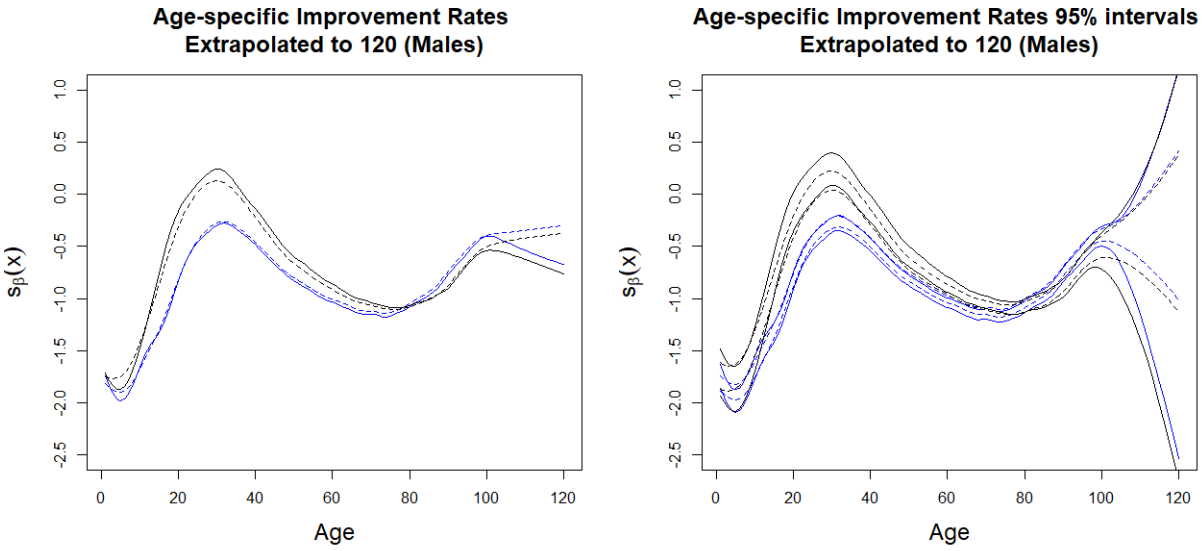
7.4.3 England-and-Wales and Scotland Joint Country Model

7.4.3.1 Males

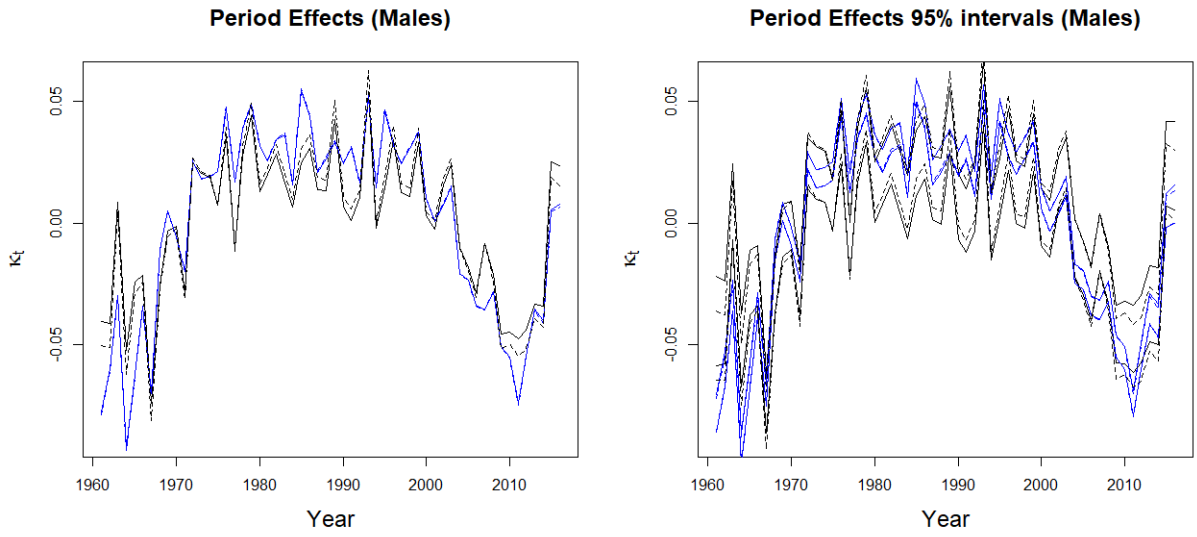
The posterior means and the 95% intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects of the Bayesian joint country model for EW and SC males are shown in Figure 7.11. The posterior means of the parameter estimates for EW males are very similar to the frequentist maximum likelihood estimates, except for the age-specific improvement rates at extrapolated ages where the posterior means from the Bayesian approach are lower (i.e. bigger improvements). For SC males, the differences between the posterior means from the Bayesian approach and the maximum likelihood estimates from the classical approach are relatively larger.



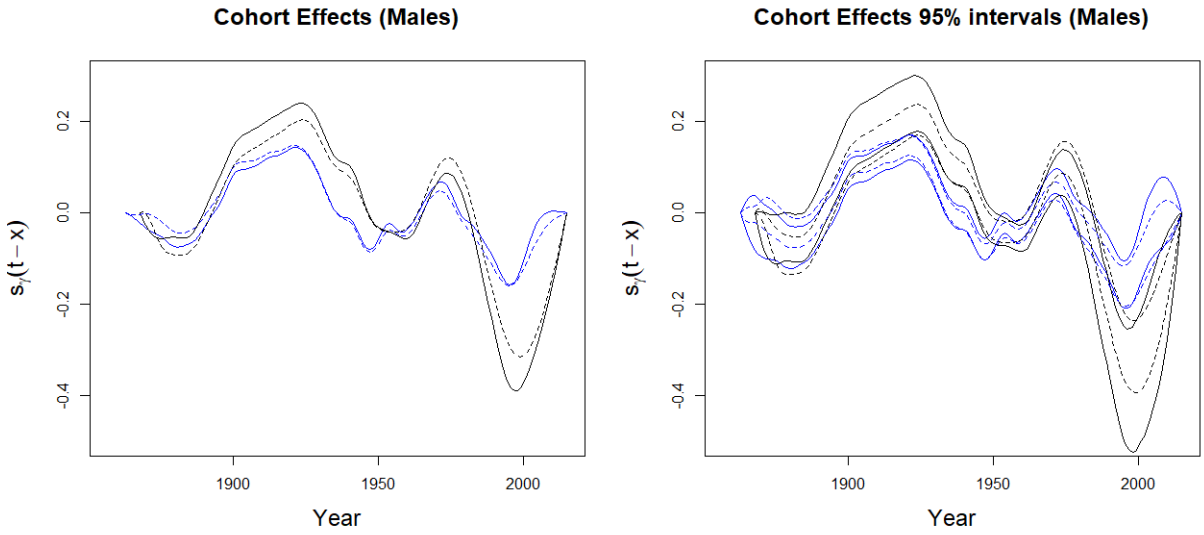
(a) baseline mortality schedules



(b) age-specific improvement rates



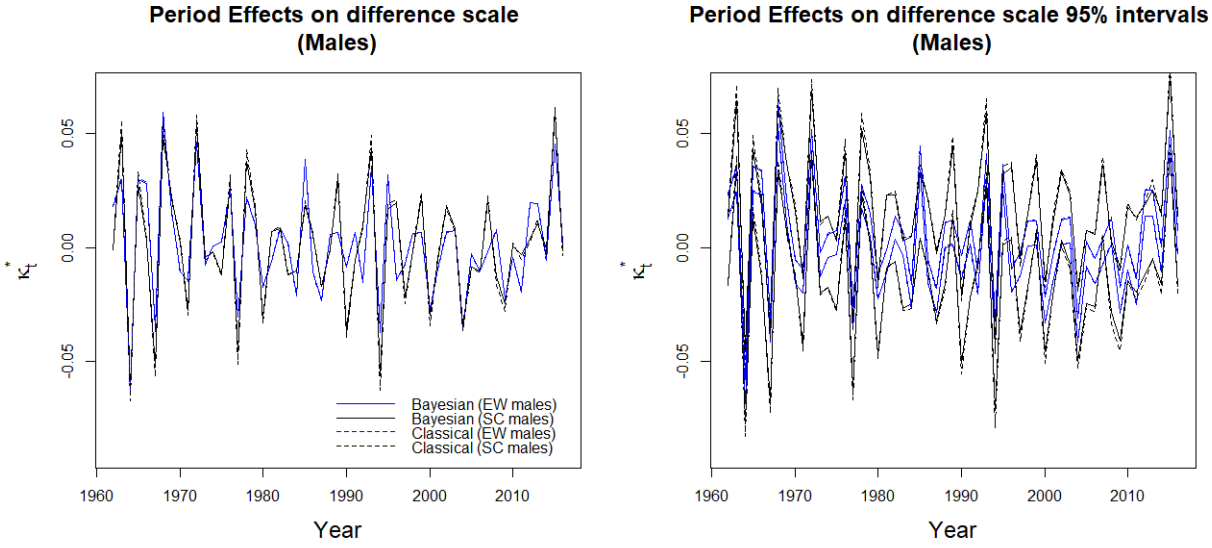
(c) period effects



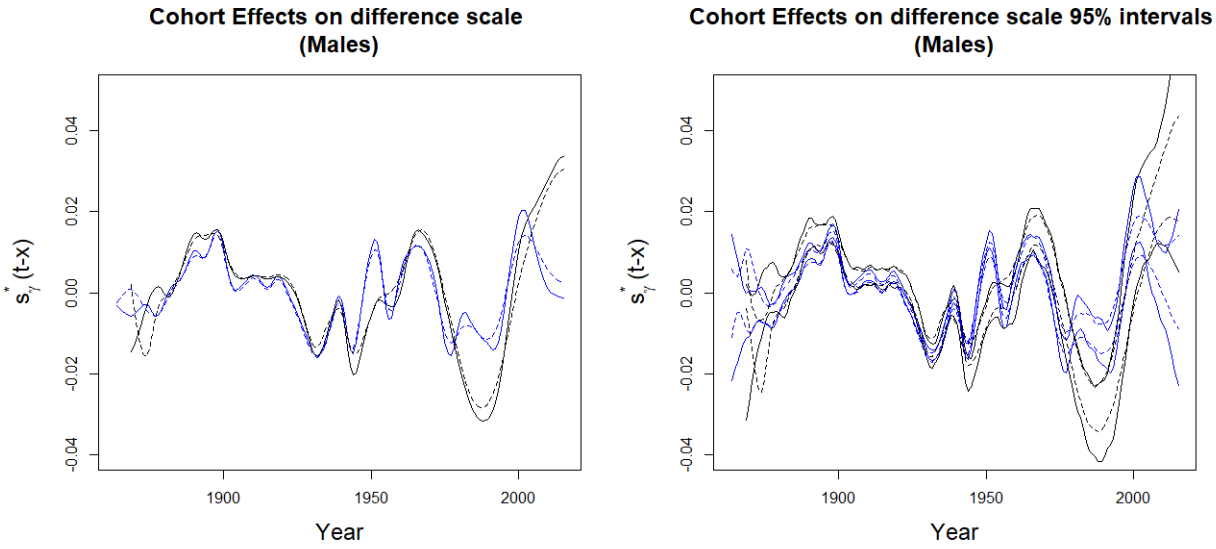
(d) cohort effects

Figure 7.11: Posterior means and 95% credible intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects (solid lines) of the Bayesian joint country model for EW (blue) and SC (black) males. The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

The posterior means of the first differences of the period and cohort effects are shown in Figure 7.12. Although there are discernible differences in the Bayesian posterior means and the frequentist maximum likelihood estimates of the period and cohort effects of SC males, the differenced period and cohort effects are from the two approaches are quite similar.



(a) first differences of the period effects



(b) first differences of the cohort effects

Figure 7.12: Posterior means and 95% credible intervals of the first differences of the period effects and cohort effects (solid lines) of the Bayesian joint country model for EW (blue) and SC(black) males. The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

Figure 7.13 presents the posterior means and 95% intervals of the 50-year ahead forecasts of the EW and SC males at some selected ages. As expected, the Bayesian approach produces wider interval forecasts. Figure 7.14 shows the projections after incorporating expert opinion. Similar to the joint sex models, the Bayesian intervals are wider (solid lines) than the frequentist intervals (dotted lines). The long-term projections are non-divergent as the mortality improvement rates converge to the expert advised rates.

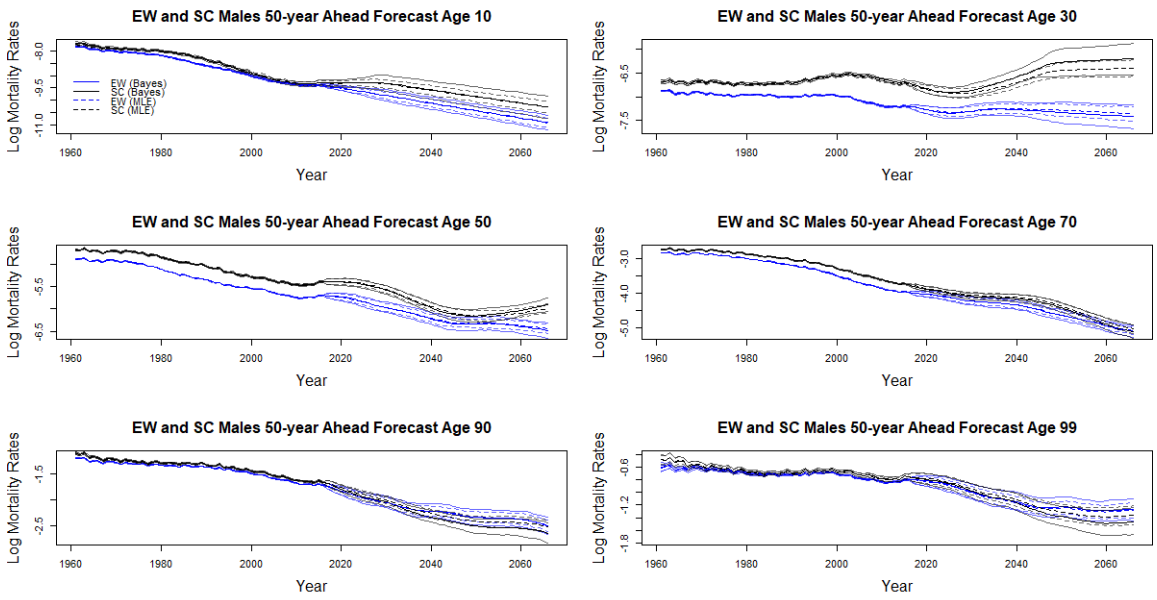


Figure 7.13: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW and SC male mortality rates (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

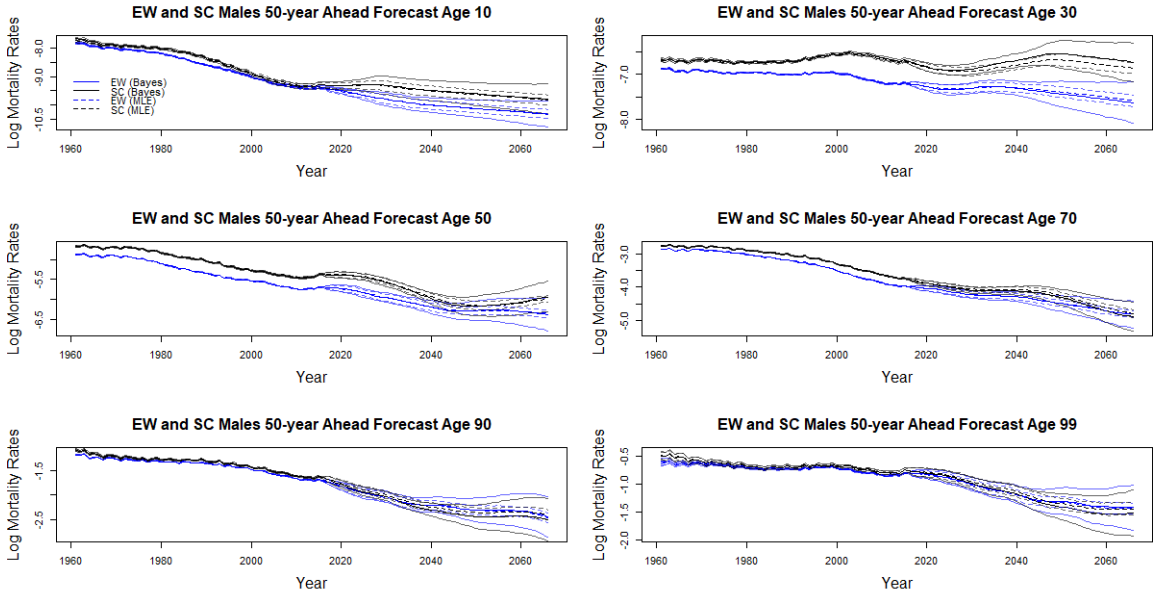
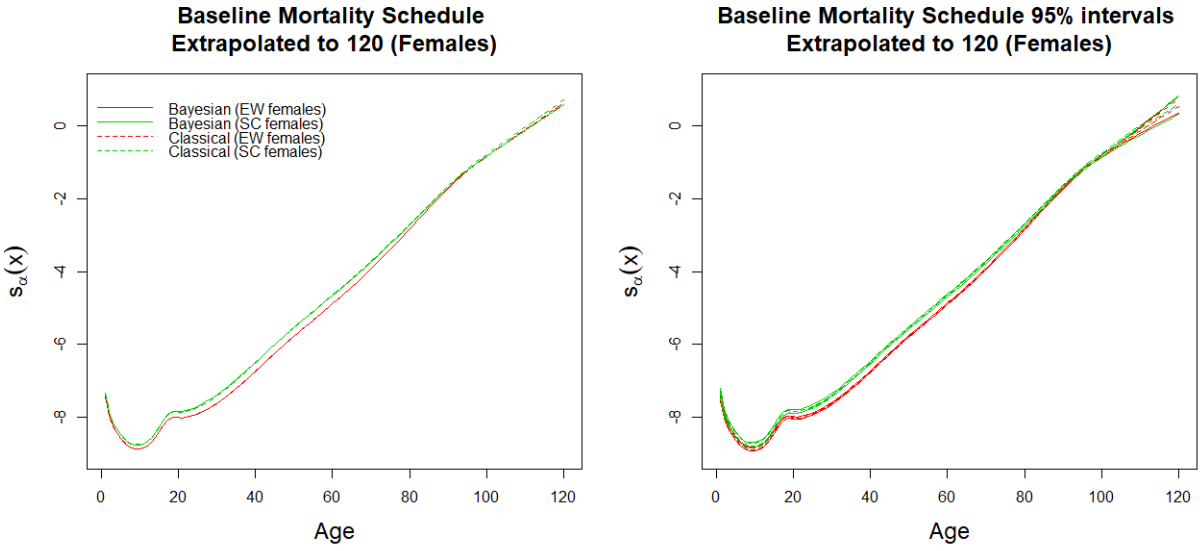


Figure 7.14: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW and SC male mortality rates after the incorporation of expert opinion (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

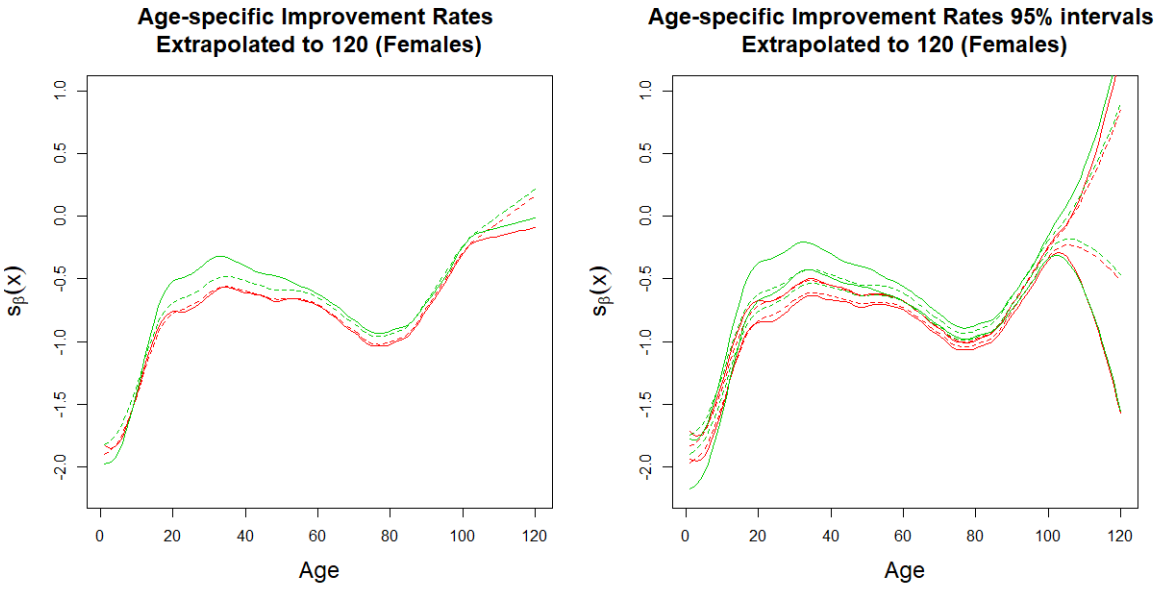
7.4.3.2 Females

The posterior means and the 95% intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects of the Bayesian joint country model for

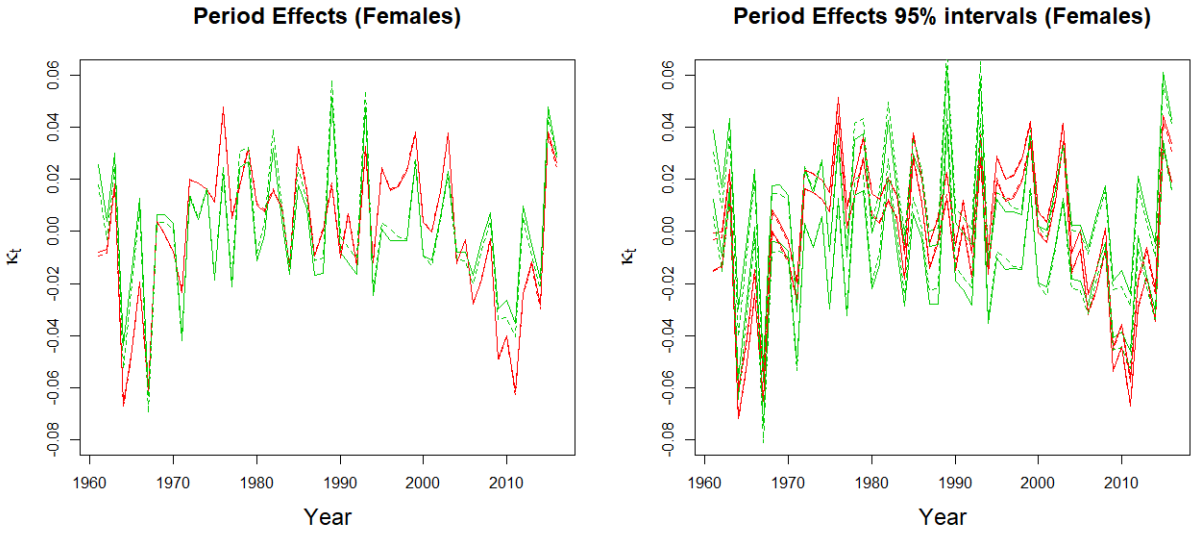
EW and SC females are shown in Figure 7.15. Similar to the male joint country model, the posterior means of the parameter estimates for EW females are very similar to the frequentist maximum likelihood estimates, except for the age-specific improvement rates at extrapolated ages. On the other hand, the differences between the two approaches are more apparent, and the sampled smoothing parameters for the cross-country penalty for $s_\beta(\cdot)$ are much lower than that of the maximum likelihood estimates, hence the shape of the age-specific improvement rates of SC females are less similar to that of EW females under the Bayesian model.



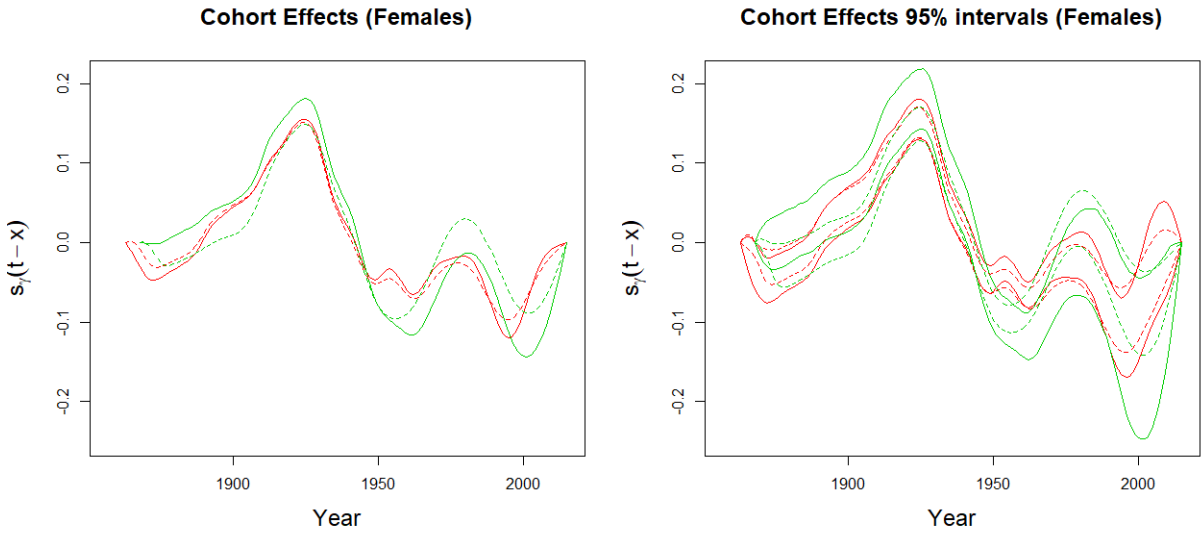
(a) baseline mortality schedules



(b) age-specific improvement rates



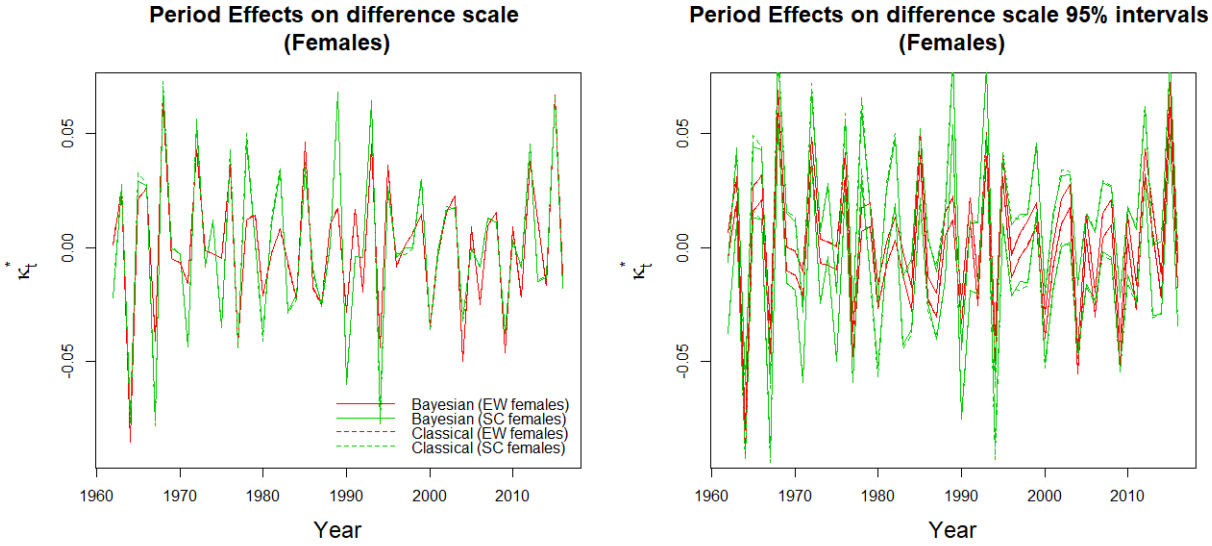
(c) period effects



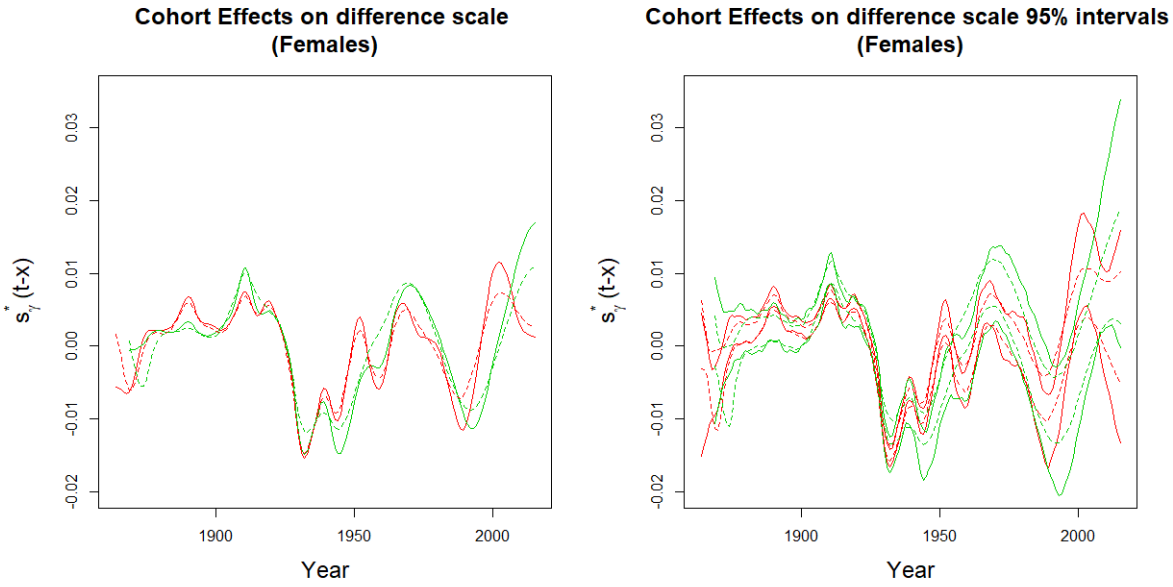
(d) cohort effects

Figure 7.15: Posterior means and 95% credible intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects (solid lines) of the Bayesian joint country model for EW (red) and SC (green) females. The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

The posterior means of the first differences of the period and cohort effects are shown in Figure 7.16. The first differences of the period and cohort effects between the two approaches are similar.



(a) first differences of the period effects



(b) first differences of the cohort effects

Figure 7.16: Posterior means and 95% credible intervals of the first differences of the period effects and cohort effects (solid lines) of the Bayesian joint country model for EW (red) and SC(green) females. The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

Figure 7.17 presents the posterior means and 95% intervals of the 50-year ahead forecasts of the EW and SC females at some selected ages. As expected, the Bayesian approach produces wider interval forecasts. Figure 7.18 shows the projections after incorporating expert opinion. Similar to the joint sex models, the Bayesian intervals are wider (solid lines) than the frequentist intervals (dotted lines). The long-term projections are non-divergent as the mortality improvement rates converge to the expert advised rates.

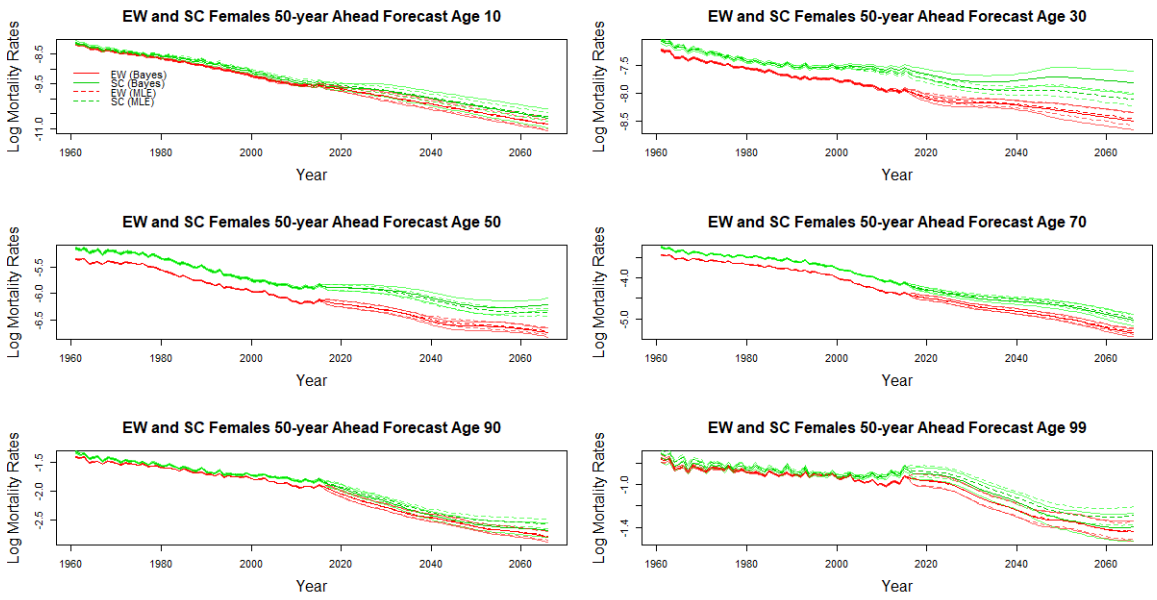


Figure 7.17: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW and SC female mortality rates (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

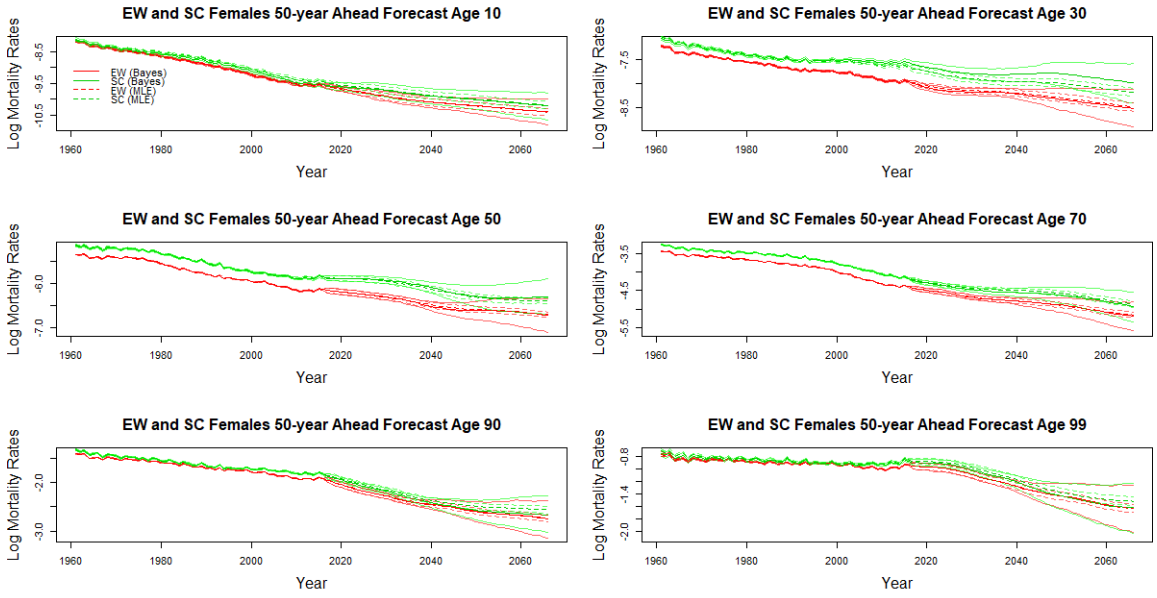


Figure 7.18: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW and SC female mortality rates after the incorporation of expert opinion (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

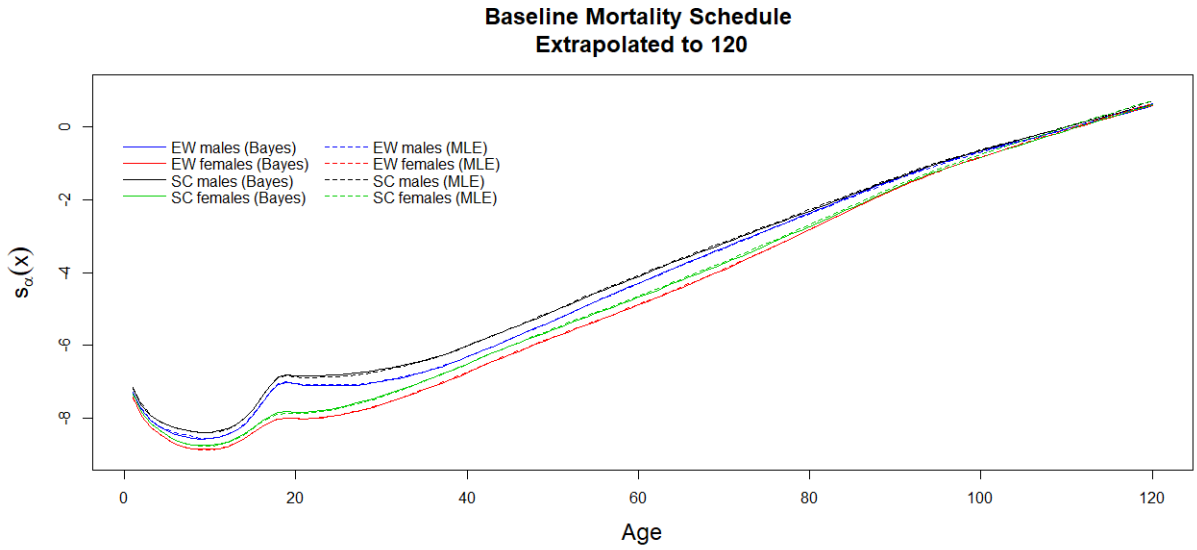
Table 7.2 shows the empirical coverage of the estimated 95% intervals of the joint country model from the Bayesian approach, backtested on data in the most recent decade. The coverages of the frequentist intervals are also shown for comparison. For all populations except SC males, the coverage has improved in the Bayesian setting.

Age	95% intervals coverage							
	EW males		EW females		SC males		SC females	
	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist
all ages	0.5846154	0.4567308	0.5365385	0.4634615	0.3555556	0.369696982	0.3929293	0.280808073
61-90	0.9333333	0.7633333	0.6966667	0.6166667	0.5766667	0.6	0.5466667	0.4333333
91+	0.7785714	0.65	0.4857143	0.45	0.3444444	0.4777778	0.4888889	0.1777778

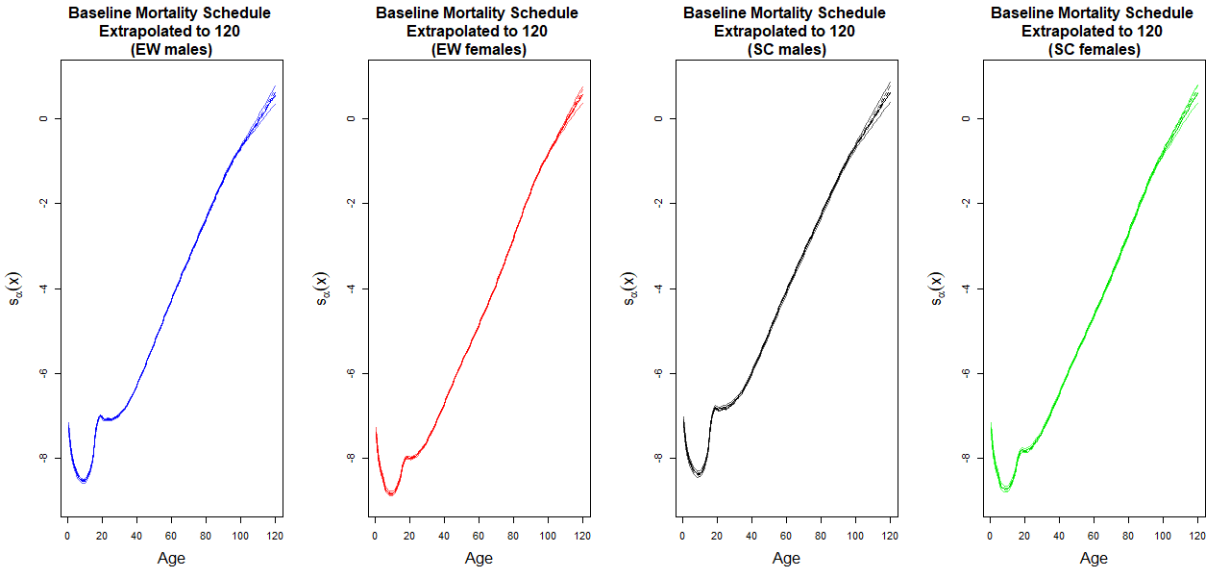
Table 7.2: The coverage of the 95% intervals of the Bayesian joint country model

7.4.4 England-and-Wales and Scotland males and females Joint Model

Figures 7.19, 7.20, 7.21 and 7.22 present the posterior means and 95% intervals of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects of the Bayesian four-population joint model. The intervals from the Bayesian approach are wider than that from the frequentist approach except for the period effects. The posterior means of the baseline mortality schedules and period effects from the Bayesian approach are very similar to the frequentist maximum likelihood estimates. However, the differences between the two approaches are larger in the age-specific improvement rates and cohort effects, most notable for SC males and females. For EW males and females, the fitted terms are still consistent between the two approaches, however, the age-specific improvement rates for SC males and females at ages around 20 to 50 in the Bayesian approach are higher (smaller improvement) than the frequentist maximum likelihood estimates.

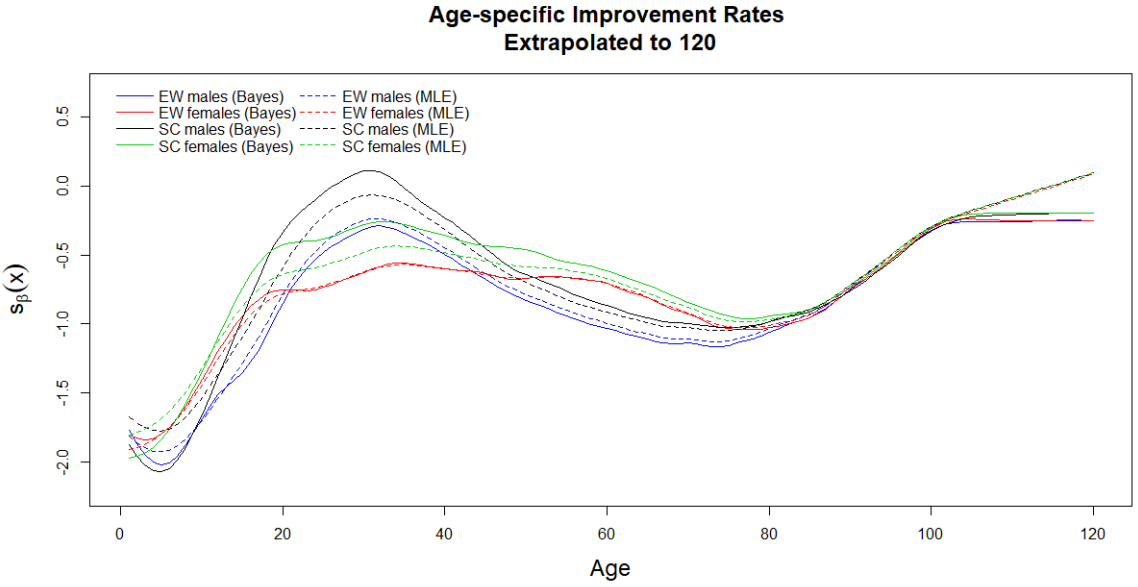


(a) baseline mortality schedules extrapolated to age 120

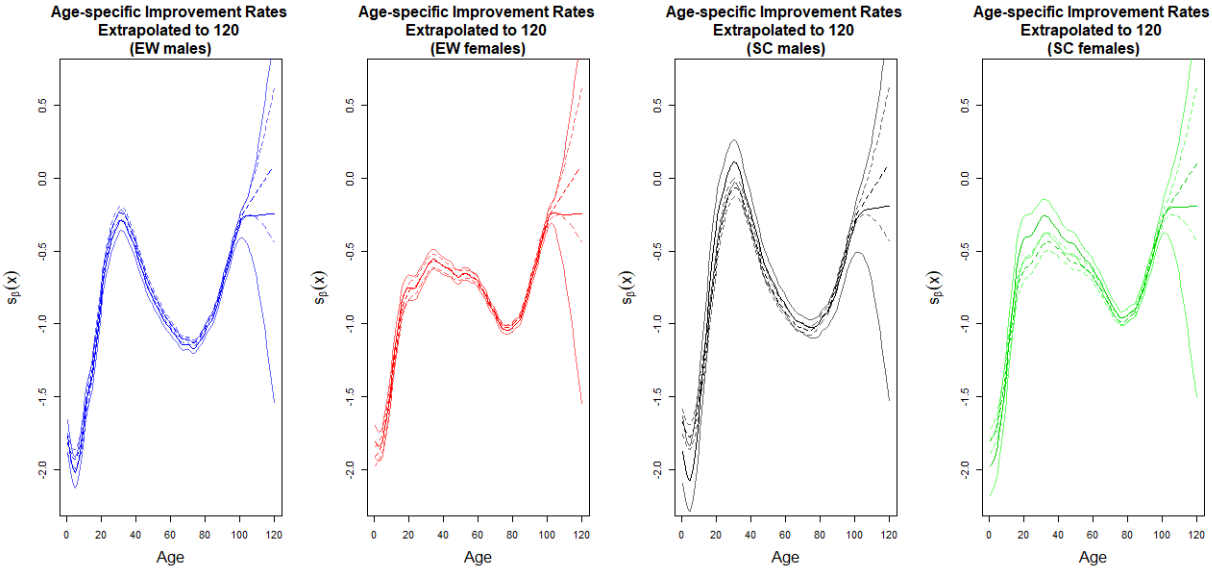


(b) individual baseline mortality schedules

Figure 7.19: Posterior means and 95% intervals of the baseline mortality schedules extrapolated to age 120. The baselines for each population are also plotted individually for a clearer illustration.



(a) age-specific mortality improvement rates extrapolated to age 120



(b) individual age-specific mortality improvement rates

Figure 7.20: Posterior means and 95% intervals of the age-specific mortality improvement rates extrapolated to age 120. The baselines for each population are also plotted individually for a clearer illustration.

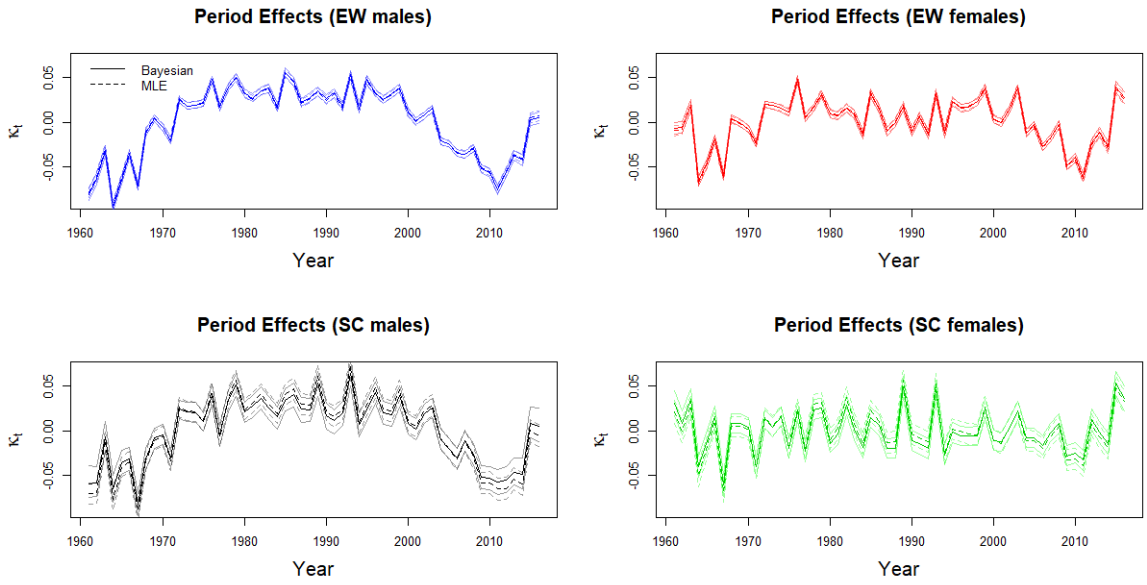


Figure 7.21: Posterior means and 95% intervals of the period effects of each population plotted individually.

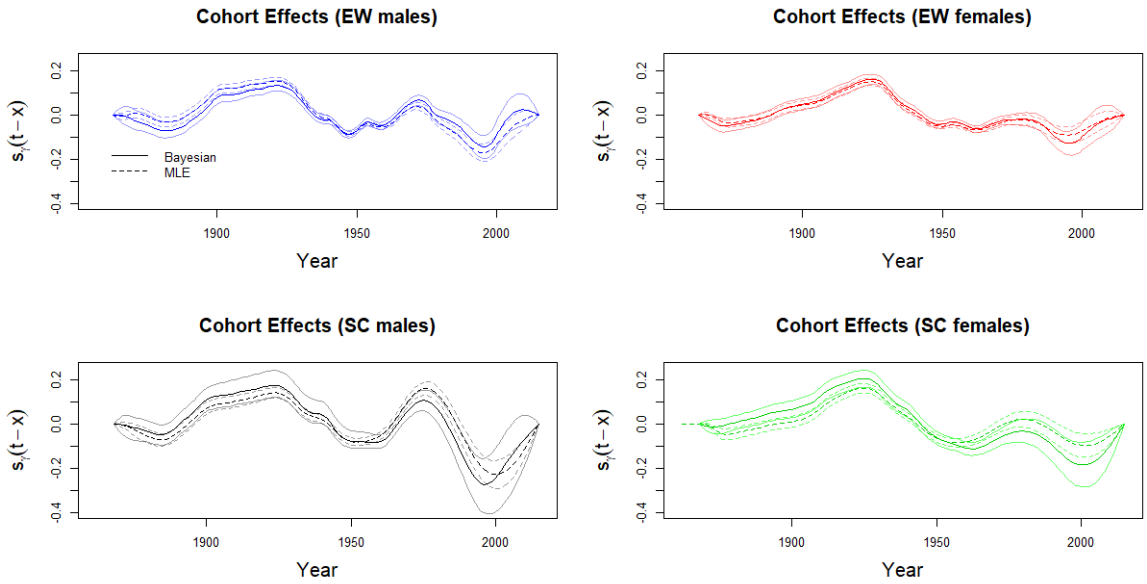
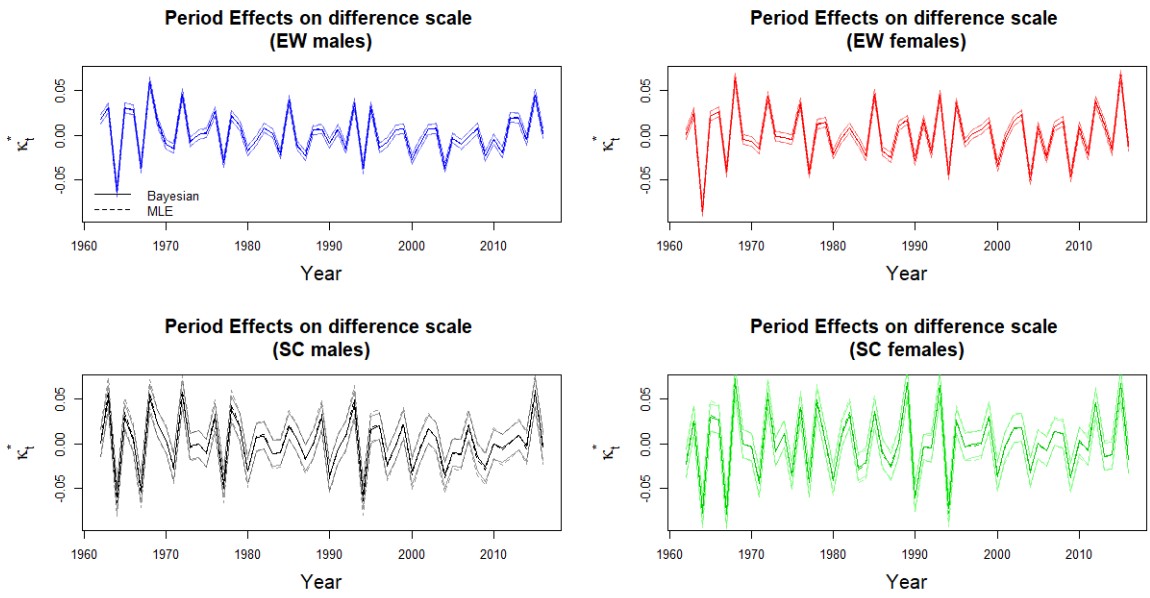
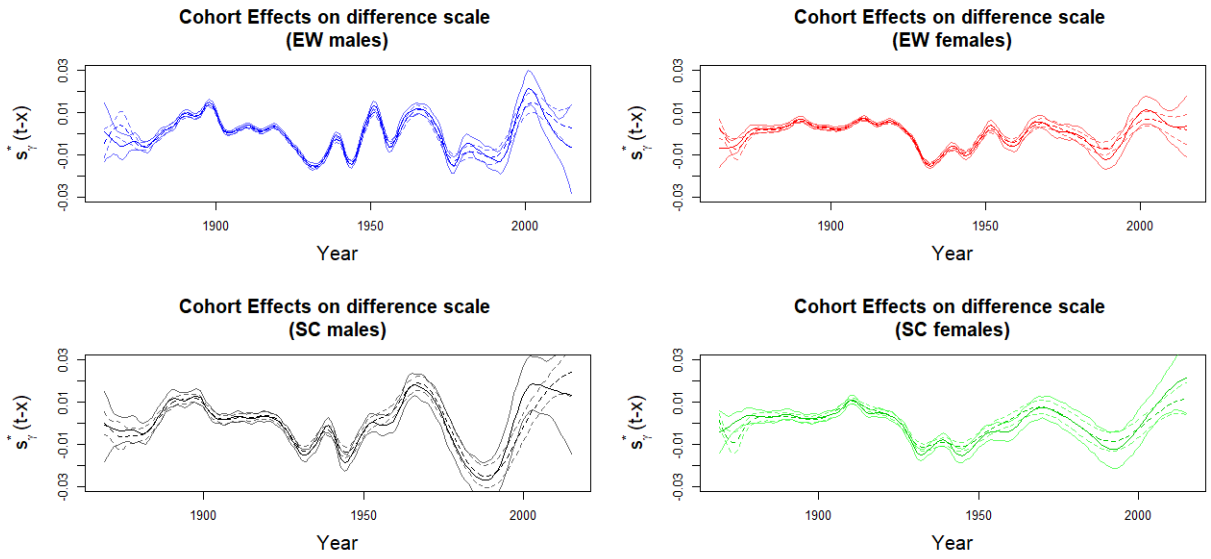


Figure 7.22: Posterior means and 95% intervals of the cohort effects of each population plotted individually.

Figure 7.23 shows the first differences of the period and cohort effects. The posterior means and the 95% intervals from the Bayesian approach almost superpose the maximum likelihood estimates from the classical approach. The means of the first differences of the cohort effects from the Bayesian approach are also very similar to the maximum likelihood estimates, except at the first and last few cohorts.



(a) first differences of the period effects



(b) first differences of the cohort effects

Figure 7.23: Posterior means and 95% credible intervals of the first differences of the period effects and cohort effects (solid lines) of the Bayesian approach. The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

For clarity, the forecasts obtained under the Bayesian model are shown pairwise. Figures 7.24 and 7.25 plot the 50-year ahead mortality forecasts of the male-and-female pairs of EW and SC respectively. Figures 7.26 and 7.27 plot the forecasts of the EW-and-SC pairs of males and females respectively. The posterior means of the forecasts under the Bayesian approach are consistent with the classical approach and intervals are generally wider. For EW females, the forecast intervals from the two approaches are very similar. In addition, the Bayesian model

produces forecasts that are decreasing at a faster rate than the maximum likelihood estimates for SC males and females at old ages.

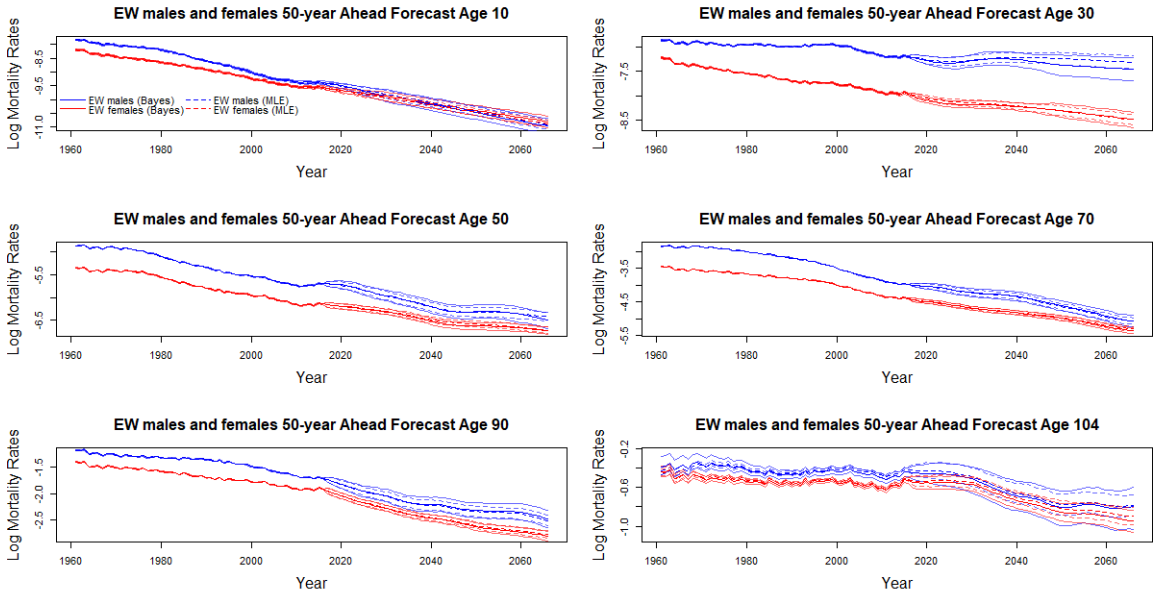


Figure 7.24: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW male and female mortality rates (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

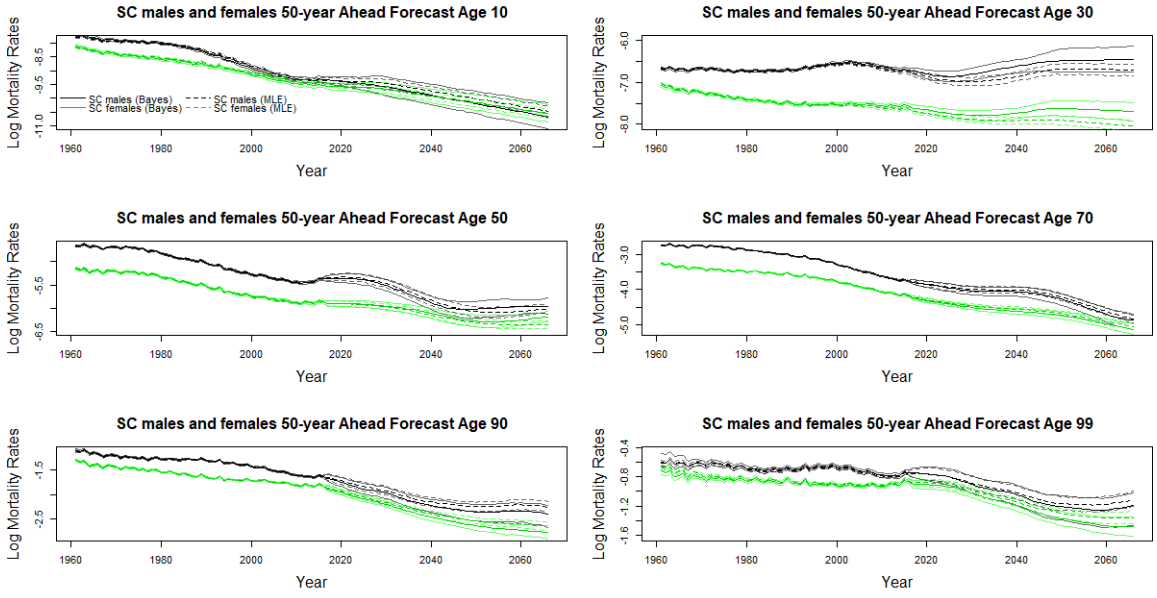


Figure 7.25: Posterior means and 95% credible intervals of the 50-year ahead forecasts of SC male and female mortality rates (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

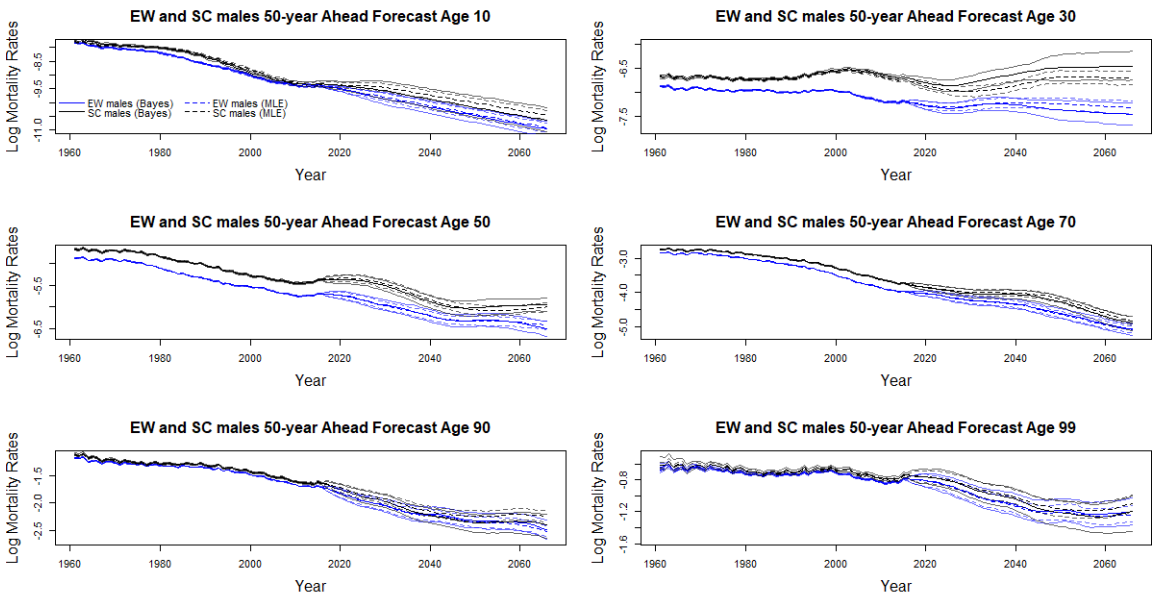


Figure 7.26: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW and SC male mortality rates (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

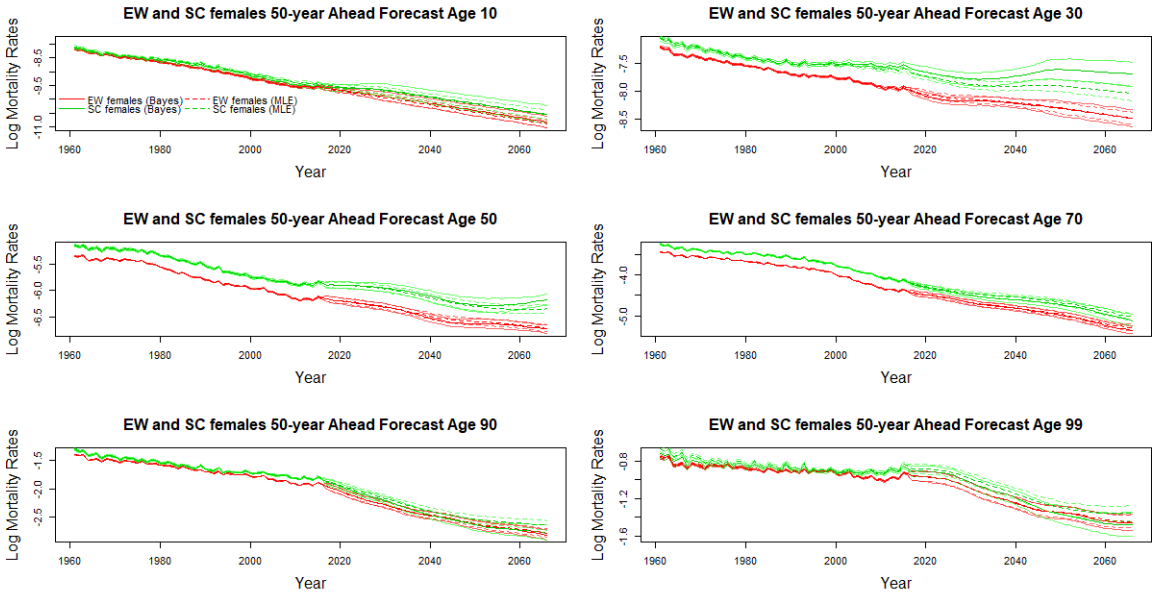


Figure 7.27: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW and SC female mortality rates (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

Figures 7.28, 7.29, 7.30 and 7.31 show the projections pairwise after incorporating expert opinion. The Bayesian intervals are much wider (solid lines) than the frequentist intervals (dotted lines). The long-term projections are non-divergent as the mortality improvement rates converge to the expert advised rates.

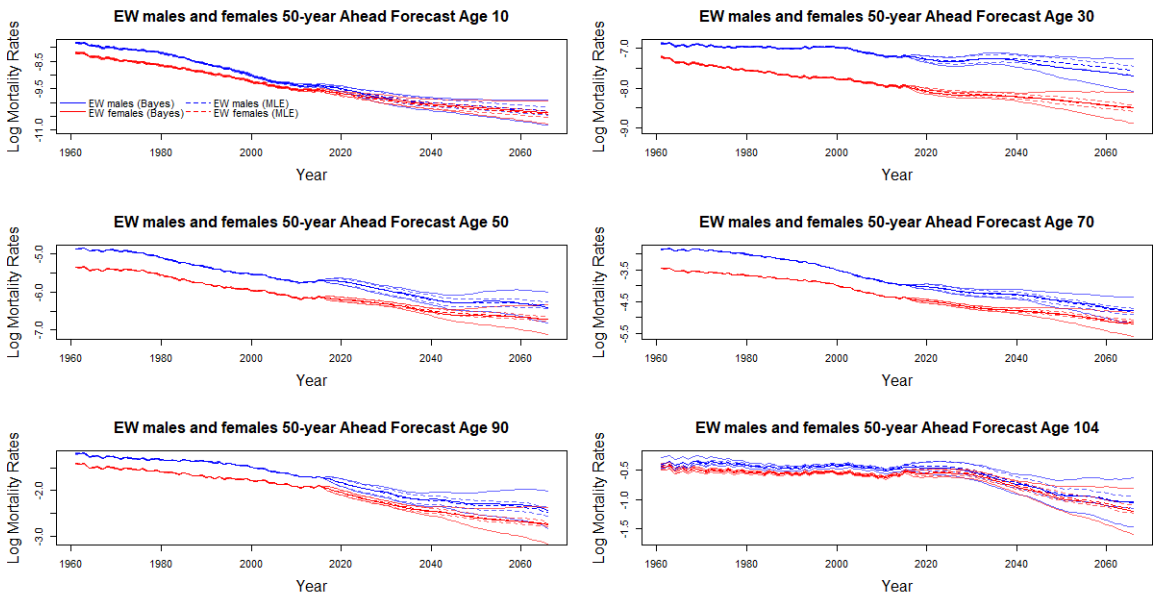


Figure 7.28: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW male and female mortality rates after the incorporation of expert opinion (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

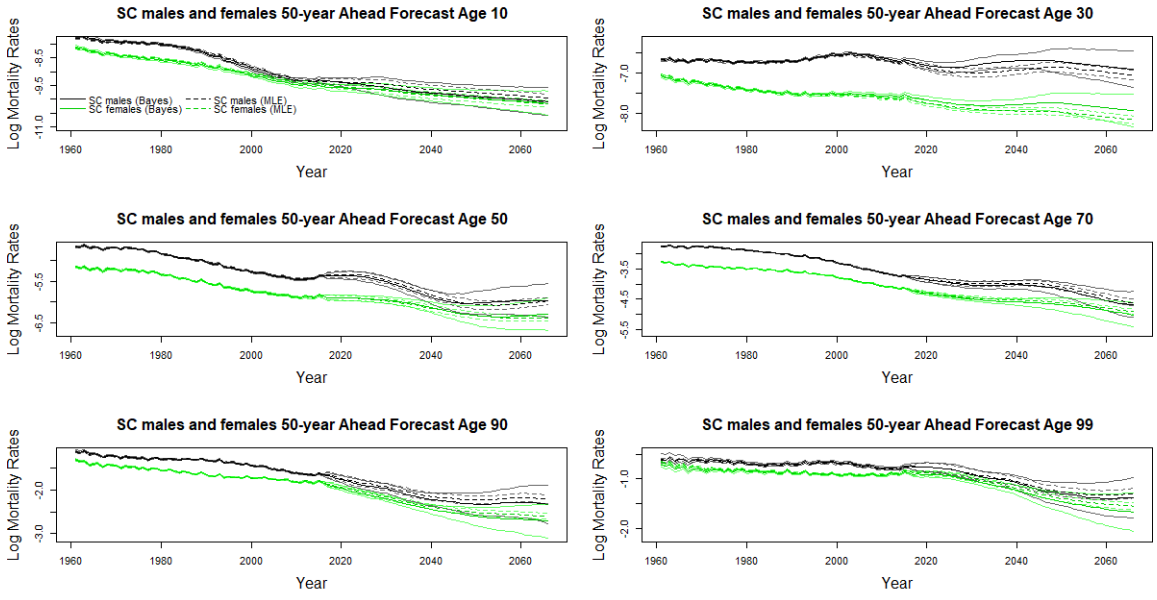


Figure 7.29: Posterior means and 95% credible intervals of the 50-year ahead forecasts of SC male and female mortality rates after the incorporation of expert opinion (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

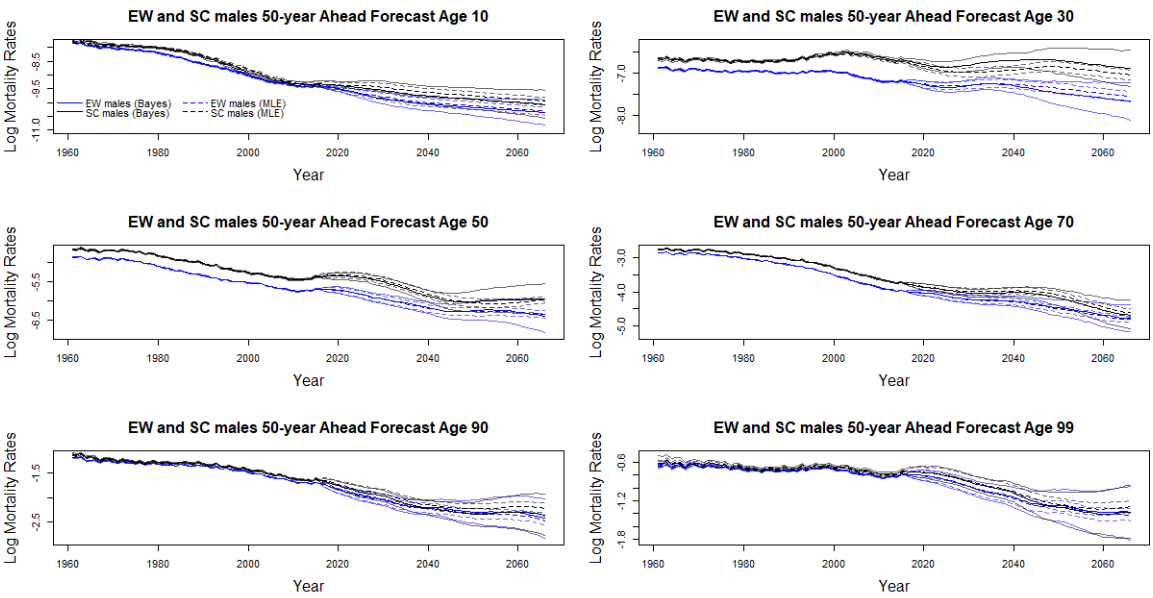


Figure 7.30: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW and SC male mortality rates after the incorporation of expert opinion (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

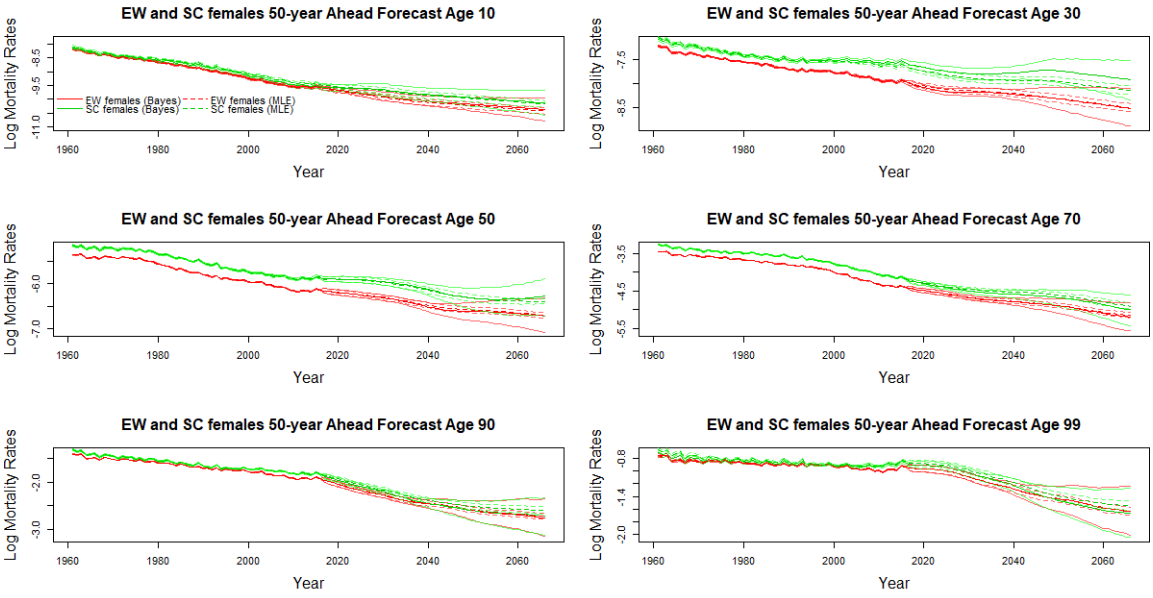


Figure 7.31: Posterior means and 95% credible intervals of the 50-year ahead forecasts of EW and SC female mortality rates after the incorporation of expert opinion (solid lines). The corresponding maximum likelihood estimates from the frequentist approach are also plotted for comparison (dotted lines).

Table 7.3 shows the empirical coverage of the estimated 95% intervals of the joint model from the Bayesian approach, backtested on data in the most recent decade. The coverages of the frequentist intervals are also shown for comparison. For all populations except SC males, the

general coverage has improved in the Bayesian setting, most notable for EW males aged 61-90 and SC females aged 91+. Nonetheless, the Bayesian intervals still under-state the uncertainty at most age groups. As mentioned, this might be an indication of over-dispersion and a negative binomial model could be explored instead which allows more variability.

	95% intervals coverage							
Age	EW males		EW females		SC males		SC females	
	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist
all ages	0.5875	0.4480769	0.5355769	0.4663462	0.3575758	0.351515139	0.4090909	0.289898979
61-90	0.93	0.77	0.7	0.6233333	0.59	0.64	0.5533333	0.4233333
91+	0.7785714	0.65	0.5071429	0.45	0.3777778	0.5222222	0.5666667	0.2111111

Table 7.3: The coverage of the 95% intervals of the Bayesian joint country model

7.5 Conclusion

In this chapter we used a Bayesian framework as it provides a natural and comprehensive manner to include parameter uncertainty and any prior beliefs. Compared to the estimates obtained under the classical approach, the interval forecasts are wider under the Bayesian approach. Although not directly comparable, the posterior means of the baseline mortality schedules, age-specific improvement rates, period effects and cohort effects are very similar to their maximum likelihood estimates for EW males and females. For SC males and females, the Bayesian approach gives slightly worse mortality improvement rates at around ages 20 to 50.

Chapter 8

Conclusion and Further Work

8.1 Conclusion

In this thesis we have proposed stochastic models for estimating and projecting mortality rates for different populations jointly using adaptive P-splines. In particular,we proposed models for mortality graduation as well as mortality projection in both the age and time directions that are suitable for the whole age range and robust.

Crude mortality rates often exhibit irregular and wiggly patterns, due to natural randomness. Therefore the crude rates have to be smoothed before they are used, this process is sometimes called mortality graduation. In Chapter 4 a model for mortality graduation for the whole age range is proposed. The model is flexible at younger ages where data is abundant and robust at the oldest ages where data is scarce. The local penalty also improves the adaptivity of the spline, as the age pattern of mortality is less smooth at younger ages, which then follows quite a smooth log-linear trend. We also borrowed strength at the oldest ages by jointly graduating male and female mortality rates, which further increases the robustness. In addition, utilising the locality of B-spline basis functions, we demonstrated an efficient and easy way to impose constraints on the spline coefficients such that graduated female mortality rates are always less than or equal to that of males, a feature that is noted and studied in the literature (Schünemann et al., 2017; Oksuzyan et al., 2008; Barford et al., 2006; Luy, 2003; Smith and Warner, 1989; Waldron, 1985; Lopez, 1983). In this way, our model does not require subjective/ad-hoc adjustments to the graduated rates in order to enforce non-intersection between male and female mortality rates.

In Chapter 5 the mortality graduation model is extended to a joint sex and joint country model, strength is borrowed between the two populations especially at the highest ages where exposures are small. In particular, the age effects between populations are related. We also demonstrated a way to incorporated expert opinion into projection so that the long term forecasts could be moderated by these judgments. Following Dodd et al. (2020), weights were assigned between the

current mortality improvement rates and the expert-determined target rates such that future mortality improvement rates will gravitate towards the expert-determined target rates. Specifically, at the 25-th forecast year, mortality improvement rates coincide with the expert-determined rates and remain constant thereafter. The incorporation of expert opinion avoids perpetual linear improvement trends at current rates, which is unlikely. It also ensures the projections are non-divergent as the target rates set by experts are the same for both sexes and countries. The joint sex models have shown to be able to produce more reasonable long term male and female mortality projections that are non intersecting, a quality that single sex models often fail to achieve. At the extrapolation age range the joint model also gives more plausible estimates, especially for the mortality improvement rates for females where a worsening mortality is otherwise projected under the single population model. One benefit of jointly modelling EW and SC populations is that information could be learned from the bigger EW population which also has data of a wider age range. In the extrapolation age range for SC, the estimates capture the decelerating rate of increase in mortality age patterns, a phenomenon that is often observed at the oldest ages. Backtests on the most recent 10 years of data show that the projection and projection intervals of the joint country model has improved, especially for SC males.

In Chapter 6, we modelled all of the four populations jointly, i.e. EW males, EW females, SC males and SC females. This model combines features of both the joint sex model and joint country model, information is borrowed across sexes, and across countries. For each male-and-female pair, the estimated baseline schedules and the mortality improvement rates tend to a common value and SC males and females are able to utilize information offered by the EW populations, especially when extending to ages beyond the available SC data range. The four-population joint model is able to produce forecasts that are more sensible, which are non-divergent and non-intersecting male and female mortality trends within the data range. Backtests have been conducted on data in the most recent decade and it is shown that the projection and projection intervals has improved in all but the younger age groups. A finer breakdown of the forecast errors reveals that the improvement in the forecast accuracy mainly comes from the SC populations, as the forecast errors of EW males and females have slightly increased.

Finally, in Chapter 7, the models in Chapters 5 and 6 are fitted in a Bayesian framework. The main advantage of this is that the parameter and forecast uncertainty can be incorporated in a natural and comprehensive manner. The estimated smoothing parameters under the Bayesian approach are consistent with those obtained under the classical approach. However, compared to the mortality estimates obtained under the classical approach, the interval forecasts are wider under the Bayesian approach, which is intuitive as the smoothing parameters are now variables instead of constants.

8.2 Discussions and further work

Even though the joint models are able to produce reasonable projections, there are still some shortcomings. First of all, as indicated by the backtests, the short term forecast accuracy may not necessarily improve, especially for younger age groups where there is relatively large data. However, note that the magnitude of the differences in the forecast errors is relatively small. For example, when the joint sex models are considered, the forecast accuracy of EW males and females aged 31-60 has slightly worsened in terms of MAE. Albeit having a low to none cross-sex penalty at the youngest ages, the projections are still influenced by the cross-sex penalty at later ages due to the presence of cohort effects that interacts with the age and period effects. In fact, some authors have also experienced difficulties in estimating and projecting the cohort effects in a robust way (Plat, 2009a; Renshaw and Haberman, 2006). Antonio et al. (2017) have even opted out not to include cohort effects into their mortality models for Netherlands and Belgium.

Another issue to consider is that while the linear mortality improvement trend appears to be suitable for most of the years, there seems to be a deviation in the most recent years, where the first differences of the estimated period effects drift away from zero. Some researchers have looked into structural changes and proposed methods of detecting and breaking down the time into periods of different mortality improvement trends (O'Hare and Li, 2015; Coelho and Nunes, 2011). Therefore, structural breaks in period effects maybe incorporated into the model.

The stochasticity of the forecasts of our model comes from the single period effect. In other words, mortality forecasts at all ages will have the same variability and perfect correlation. Therefore, the prediction intervals will have the same width if parameter uncertainty were ignored. One could forecast the cohort effects using time series for a non-trivial correlation structure. Inclusion of higher order period effects have also been considered in the literature.

The estimated period effects in all of the models we have investigated seem to be correlated among populations. Therefore, a possibility to improve forecast accuracy is then to use correlated time series for the period effects of different populations instead of forecasting each of them individually. Such work has been investigated by researchers, for example the two population model by Cairns et al. (2011b) and the gravity model by Dowd et al. (2011), among others discussed in Section 3.3.

The assumption of Poisson deaths often exhibits over-dispersion. The one-parameter Poisson distribution may be too restrictive on the variance, especially for large, heterogeneousus populations. One remedy would be to introduce an over-dispersion parameter to account for the extra variability. As shown in Section 1.1, a particular convenient choice of parametrisation leads to a negative binomial distribution, which is not of the exponential family.

Instead of jointly modelling multiple populations, the model could be fitted to a large population and its sub-populations (e.g. a national population and the insured population). The shape and

difference penalties are then simply intuitive and natural as the sub-populations and the main population would undoubtedly share characteristics in their mortality schedules. At higher ages the differences are also expected to wear off, hence the mortality levels (the baseline mortality rates) and the trends (the improvement rates) should be similar at these ages. This is in line with the relational model by Plat (2009b) (Section 3.3) where the authors enforced the ratio between the main and sub-population to be exactly 1 at the oldest age (i.e. the mortality rates of the main and sub-population converge to a common value at the oldest age).

The incorporation of parameter uncertainty into the prediction intervals is also not a straightforward task under the classical approach, even though simulation techniques can be used. The intervals obtained for the estimated parameters are also not purely classical due to the presence of penalties. As discussed in Section 2.2.6, the interval estimates rely on a Bayesian point of view on the spline parameters. That said, the parameter intervals are somewhat midway between the classical and Bayesian approaches. An obvious way to capture parameter and forecast uncertainty in a more natural and systematic fashion would be to use a Bayesian approach. In this way the time series likelihood can also be incorporated as a prior, which is often neglected during the fitting process under the classical approach. In Chapter 7 we fitted the models in a Bayesian framework, however we have assumed an AR(1) prior on the period effects. It is shown in Chapters 5 and 6 that an ARIMA process may give better and more reasonable forecasts.

An attractive feature of the Bayesian framework is that prior beliefs can be incorporated into the model easily. For instance, if one believes that the baseline mortality schedule is smoother at older ages, one could put more weight in the prior of λ_2^α where $\lambda_2^\alpha > 0$. The magnitude of the smoothing parameters often depends on the range of the data and the spacing between the spline knots. Nonetheless, if prior knowledge is available this can also be reflected in the prior distributions. The smoothing parameters were given uniform priors in our model, however other distributions could be used. In the literature some authors have found that the prior specification of the smoothing parameters may affect the posterior (Jullion and Lambert, 2007) and suggested more robust prior specifications. For example, Jullion and Lambert (2007) proposed a mixture prior as well as adding another layer of hyperprior. Scheipl and Kneib (2009) on the other hand used the Normal-Exponential-Gamma prior for the spline coefficients that allows spatial adaptivity and computational convenience. Therefore other prior distributions could be considered for our model and the prior sensitivity could be explored.

In this thesis we proposed models for joint mortality graduation and projection. Splines have been used increasingly in mortality modelling, we proposed a way to flexibly model the age effects of different populations jointly using P-splines, therefore borrowing information across the populations. We showed that the estimates are more robust and the projections are more reasonable. We have also demonstrated a way to incorporate expert opinion into long-term forecasts, which helps moderate projections produced by the models and improves credibility.

References

Antero-Jacquemin, J. d. S., Berthelot, G., Marck, A., Noirez, P., Latouche, A., and Toussaint, J.-F. (2015). Learning from leaders: life-span trends in olympians and supercentenarians. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 70(8):944–949.

Antonio, K., Devriendt, S., de Boer, W., de Vries, R., De Waegenaere, A., Kan, H.-K., Kromme, E., Ouburg, W., Schulteis, T., Slagter, E., et al. (2017). Producing the dutch and belgian mortality projections: a stochastic multi-population standard. *European actuarial journal*, 7(2):297–336.

Baladandayuthapani, V., Mallick, B. K., and Carroll, R. J. (2005). Spatially adaptive bayesian penalized regression splines (p-splines). *Journal of Computational and Graphical Statistics*, 14(2):378–394.

Barford, A., Dorling, D., Smith, G. D., and Shaw, M. (2006). Life expectancy: women now on top everywhere.

Beard, R. E. (1959). Note on some mathematical mortality models. In *Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing)*, volume 5, pages 302–311. Wiley Online Library.

Biatat, V. D. and Currie, I. D. (2010). Joint models for classification and comparison of mortality in different countries. In *Proceedings of 25rd International Workshop on Statistical Modelling, Glasgow*, pages 89–94.

Booth, H., Hyndman, R. J., Tickle, L., and De Jong, P. (2006). Lee-carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, 15:289–310.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying lee-carter under conditions of variable mortality decline. *Population studies*, 56(3):325–336.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of actuarial science*, 3(1-2):3–43.

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, 31(3):373–393.

Cairns, A., Blake, D., Dowd, K., and Kessler, A. (2014). Phantoms never die: Living with unreliable mortality data. Technical report, Citeseer.

Cairns, A. J. (2013). Modeling and management of longevity risk.

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Cairns, A. J., Blake, D., and Dowd, K. (2008). Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal*, 2008(2-3):79–113.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2011a). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48(3):355–367.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, 13(1):1–35.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., and Khalaf-Allah, M. (2011b). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin: The Journal of the IAA*, 41(1):29–59.

Camarda, C. G. (2019). Smooth constrained mortality forecasting. *Demographic Research*, 41:1091–1130.

Carriere, J. F. (1992). Parametric models for life tables. *Transactions of the Society of Actuaries*, 44:77–99.

Coale, A. J. and Kisker, E. E. (1990). Defects in data on old-age mortality in the united states: new procedures for calculating mortality schedules and life tables at the highest ages.

Coelho, E. and Nunes, L. C. (2011). Forecasting mortality in the event of a structural change. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(3):713–736.

Continuous Mortality Investigation Bureau (CMI) (2006). Stochastic projection methodologies: Further progress and p-spline model features, example results and implications. Technical report.

Continuous Mortality Investigation Bureau (CMI) (2007). Stochastic projection methodologies: Lee-carter model features, example results and implications. Technical report.

Couzin-Frankel, J. (2011). A pitched battle over life span.

Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). Spatially adaptive bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics*, 16(2):265–288.

Currie, I. D. (2011). Modelling and forecasting the mortality of the very old. *ASTIN Bulletin: The Journal of the IAA*, 41(2):419–427.

Currie, I. D. and Durban, M. (2002). Flexible smoothing with p-splines: a unified approach. *Statistical Modelling*, 2(4):333–349.

Currie, I. D., Durban, M., and Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical modelling*, 4(4):279–298.

D'Amato, V., Piscopo, G., and Russolillo, M. (2011). The mortality of the italian population: Smoothing techniques on the lee—carter model. *The Annals of Applied Statistics*, pages 705–724.

De Boor, C., De Boor, C., Mathématicien, E.-U., De Boor, C., and De Boor, C. (1978). *A practical guide to splines*, volume 27. Springer-Verlag New York.

De Jong, P., Tickle, L., and Xu, J. (2016). Coherent modeling of male and female mortality using lee–carter in a complex number framework. *Insurance: Mathematics and Economics*, 71:130–137.

Delwarde, A., Denuit, M., and Eilers, P. (2007). Smoothing the lee–carter and poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. *Statistical modelling*, 7(1):29–48.

Delwarde, A., Denuit, M., Guillén, M., and Vidiella-i Anguera, A. (2006). Application of the poisson log-bilinear projection model to the g5 mortality experience. *Belgian Actuarial Bulletin*, 6(1):54–68.

Dodd, E., Forster, J., Bijak, J., and Smith, P. W. (2020). Stochastic modelling and projection of mortality improvements using a hybrid parametric/semiparametric age-period-cohort model. *Scandinavian Actuarial Journal*. doi: 10.1080/03461238.2020.1815238.

Dodd, E., Forster, J. J., Bijak, J., and Smith, P. W. (2018). Smoothing mortality data: the english life tables, 2010–2012. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3):717–735.

Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., and Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, 15(2):334–356.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, pages 89–102.

Fienberg, S. E. (2013). Cohort analysis' unholy quest: A discussion. *Demography*, 50(6):1981–1984.

Gallop, A. (2002). Mortality at advanced ages in the united kingdom. *Living to 100 and Beyond: Survival at Advanced Ages*.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, (115):513–583.

Gu, C. (1992). Cross-validating non-gaussian data. *Journal of Computational and Graphical Statistics*, 1(2):169–179.

Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(01):49–80.

Hilton, J., Dodd, E., Forster, J. J., and Smith, P. W. (2019). Projecting uk mortality by using bayesian generalized additive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(1):29–49.

Human Mortality Database (HMD). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at `www.mortality.org` or `www.humanmortality.de` (data downloaded on 22nd July, 2019).

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational statistics & data analysis*, 55(9):2579–2589.

Hyndman, R. J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50(1):261–283.

Jarner, S. F. and Kryger, E. M. (2011). Modelling adult mortality in small populations: The saint model. *ASTIN Bulletin: The Journal of the IAA*, 41(2):377–418.

Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive bayesian p-splines models. *Computational statistics & data analysis*, 51(5):2542–2558.

Kaishev, V. K., Dimitrova, D. S., Haberman, S., and Verrall, R. (2006). Geometrically designed, variable know regression splines: asymptotics and inference.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63:147–152.

Krivobokova, T., Crainiceanu, C. M., and Kauermann, G. (2008). Fast adaptive penalized splines. *Journal of Computational and Graphical Statistics*, 17(1):1–20.

Lang, S. and Brezger, A. (2004). Bayesian p-splines. *Journal of computational and graphical statistics*, 13(1):183–212.

Lee, R. and Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, 38(4):537–549.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.

Li, J., Dacorogna, M., Tan, C. I., et al. (2014). The impact of joint mortality modelling on hedging effectiveness of mortality derivatives. In *Tenth International Longevity Risk and Capital Markets Solutions Conference, Santiago, Chile*.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography*, 42(3):575–594.

Lindbergson, M. (2001). Mortality among the elderly in sweden 1988–1997. *Scandinavian Actuarial Journal*, 2001(1):79–94.

Liu, Z. and Guo, W. (2010). Data driven adaptive spline smoothing. *Statistica Sinica*, pages 1143–1163.

Lopez, A. D. (1983). The sex mortality differential in developed countries. *Sex differentials in mortality: trends, determinants and consequences*, pages 53–120.

Luo, L. (2013). Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. *Demography*, 50(6):1945–1967.

Luo, L. and Hodges, J. S. (2020). Constraints in random effects age-period-cohort models. *Sociological Methodology*, page 0081175020903348.

Luy, M. (2003). Causes of male excess mortality: insights from cloistered populations. *Population and Development Review*, 29(4):647–676.

Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries*, 8(6):301–310.

Mason, K. O., Mason, W. M., Winsborough, H. H., and Poole, W. K. (1973). Some methodological issues in cohort analysis of archival data. *American sociological review*, pages 242–258.

Mason, W. M. and Smith, H. L. (1985). Age-period-cohort analysis and the study of deaths from pulmonary tuberculosis. In *Cohort analysis in social research*, pages 151–227. Springer.

McNown, R. and Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8(3):413–432.

Murphy, M. (2009). The'golden generations'in historical context. *British Actuarial Journal*, pages 151–184.

Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy.

O'Hare, C. and Li, Y. (2014). Structural breaks in mortality models: An international comparison. *Available at SSRN 2515625*.

O'Hare, C. and Li, Y. (2015). Identifying structural breaks in stochastic mortality models. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 1(2).

Oksuzyan, A., Juel, K., Vaupel, J. W., and Christensen, K. (2008). Men: good health and high mortality. sex differences in health and aging. *Aging clinical and experimental research*, 20(2):91–102.

Olshansky, S. J., Carnes, B. A., and Cassel, C. (1990). In search of methuselah: estimating the upper limits to human longevity. *Science*, 250(4981):634–640.

O'Sullivan, F., Yandell, B. S., and Raynor Jr, W. J. (1986). Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association*, 81(393):96–103.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63(1):12–57.

Philipov, D., Jdanov, D., and Jasilionis, D. (2020). About mortality data for england and wales. human mortality database: Background and documentation. *University of California, Berkeley and Max Planck Institute for Demographic Research. http://www. mortality.org.*

Pintore, A., Speckman, P., and Holmes, C. C. (2006). Spatially adaptive smoothing splines. *Biometrika*, 93(1):113–125.

Pitacco, E. (2016). High age mortality and frailty. some remarks and hints for actuarial modeling.

Pitacco, E. (2019). Heterogeneity in mortality: a survey with an actuarial focus. *European Actuarial Journal*, 9(1):3–30.

Plat, R. (2009a). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3):393–404.

Plat, R. (2009b). Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics*, 45(1):123–132.

Renshaw, A. and Haberman, S. (2003). Lee–carter mortality forecasting: A parallel generalized linear modelling approach for england and wales mortality projections. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1):119–137.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the lee–carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570.

Richards, S. and Currie, I. (2009). Longevity risk and annuity pricing with the lee-carter model. *British Actuarial Journal*, 15(2):317–343.

Richards, S. J., Kirkby, J., and Currie, I. D. (2006). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, 12(1):5–38.

Rootzén, H. and Zholud, D. (2017). Human life is unlimited–but short. *Extremes*, 20(4):713–728.

Ruppert, D. and Carroll, R. J. (2000). Theory & methods: Spatially-adaptive penalties for spline fitting. *Australian & New Zealand Journal of Statistics*, 42(2):205–223.

Russolillo, M., Giordano, G., and Haberman, S. (2011). Extending the lee–carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, 2011(2):96–117.

Saikia, P. and Borah, M. (2014). A comparative study of parametric models of old-age mortality. *International Journal of Science and Research*, 3(5).

Scheipl, F. and Kneib, T. (2009). Locally adaptive bayesian p-splines with a normal-exponential-gamma prior. *Computational Statistics & Data Analysis*, 53(10):3533–3552.

Schmid, V. J. and Held, L. (2007). Bayesian age-period-cohort modeling and prediction-bamp. *Journal of Statistical Software*, 21(8):1–15.

Schünemann, J., Strulik, H., and Trimborn, T. (2017). The gender gap in mortality: How much is explained by behavior? *Journal of health economics*, 54:79–90.

Shang, H. L. et al. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: a multilevel functional data method. *The Annals of Applied Statistics*, 10(3):1639–1672.

Shang, H. L. and Hyndman, R. J. (2017). Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, 26(2):330–343.

Smith, D. W. and Warner, H. R. (1989). Does genotypic sex have a direct effect on longevity? *Experimental Gerontology*, 24(4):277–288.

Smith, H. L. (2004). Response: cohort analysis redux. *Sociological methodology*, 34(1):111–119.

Smith, H. L. (2008). Advances in age–period–cohort analysis. *Sociological Methods & Research*, 36(3):287–296.

Stan Development Team (2018). Rstan: the r interface to stan. r package version 2.19. 2.

Stan Development Team (2020). Stan modeling language users guide and reference manual. *Technical report*. 2.25. `https://mc-stan.org`.

Storlie, C. B., Bondell, H. D., and Reich, B. J. (2010). A locally adaptive penalty for estimation of functions with varying roughness. *Journal of Computational and Graphical Statistics*, 19(3):569–589.

Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):5–43.

Thiele, T. N. (1871). On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the danish life insurance company of 1871. *Journal of the Institute of Actuaries*, 16(5):313–329.

Tickle, L. and Booth, H. (2014). The longevity prospects of australian seniors: an evaluation of forecast method and outcome. *Asia-Pacific Journal of Risk and Insurance*, 8(2):259–292.

Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the g7 countries. *Nature*, 405(6788):789.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.

Villegas, A. M. and Haberman, S. (2014). On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in england. *North American Actuarial Journal*, 18(1):168–193.

Waldron, I. (1985). What do we know about causes of sex differences in mortality? a review of the literature. *Population Bulletin of the United Nations*, (18):59.

Wienke, A. (2014). Frailty models. *Wiley StatsRef: Statistics Reference Online*.

Willets, R. (2004). The cohort effect: insights and explanations. *British Actuarial Journal*, 10(4):833–877.

Wilson, C. (2011). Understanding global demographic convergence since 1950. *Population and Development Review*, 37(2):375–388.

Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.

Wood, S. N. (2016). Just another gibbs additive modeller: Interfacing jags and mgcv. *arXiv preprint arXiv:1602.02539*.

Yang, L. and Hong, Y. (2017). Adaptive penalized splines for data smoothing. *Computational Statistics & Data Analysis*, 108:70–83.

Yang, Y., Fu, W. J., and Land, K. C. (2004). A methodological comparison of age-period-cohort models: the intrinsic estimator and conventional generalized linear models. *Sociological methodology*, 34(1):75–110.

Yang, Y. and Land, K. C. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological methodology*, 36(1):75–97.

Yang, Y. and Land, K. C. (2008). Age–period–cohort analysis of repeated cross-section surveys: fixed or random effects? *Sociological methods & research*, 36(3):297–326.

Yue, Y. R., Speckman, P. L., and Sun, D. (2012). Priors for bayesian adaptive spline smoothing. *Annals of the Institute of Statistical Mathematics*, 64(3):577–613.

Zhou, R., Li, J. S.-H., and Tan, K. S. (2013). Pricing standardized mortality securitizations: A two-population model with transitory jump effects. *Journal of Risk and Insurance*, 80(3):733–774.