Stockholms universitet

A study of age, period and cohort effects applied to Swedish mortality rates

Ellinor Krona

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2019:6
Försäkringsmatematik
Juni 2019

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2019:6**
http://www.math.su.se

A study of age, period and cohort effects applied to Swedish mortality rates

Ellinor Krona*

June 2019

Abstract

Preceding studies demonstrate that the inclusion of cohort specific effects could improve the fit of stochastic mortality models. To determine the potential improvements, the aim of this study was to quantitatively compare the Lee Carter (LC) and Renshaw Haberman (RH) mortality models explaining the development in mortality rates in Sweden and the United Kingdom. An exploratory analysis was conducted to detect cohort effects while the robustness, interpretability and prediction accuracy was evaluated by periodic model fittings and projections. Ultimately, the fitting of the RH model decreased the occurrence of cross-year and cross-age correlations hence a part of the systematic effects in mortality data can be explained by including a cohort term. However, in contrast to the robust LC model, the RH model proved sensitive to outliers and changes in the underlying data. Although the mortality projections by the RH model improved in terms of forecast accuracy, the lack of robustness in the RH model emphasise that we can not completely rely on the predictions.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: ellinor.krona@gmail.com. Supervisor: Chun-Biu Li.

Acknowledgements

First and foremost, I would like to thank Chun-Biu Li for guidance throughout the writing of this report. I would further like to thank my colleagues at the Swedish Financial Supervisory Authority for their inspirational ideas. Last but not least, I would like to express my great appreciation to my family for always supporting me.

Contents

1 Introduction

Over the past decades, rising longevity and ageing populations have directed elevated interest to the study of development and projection of mortality rates. An accelerating improvement of mortality rates impacts financial institutions providing life insurance and pension. Unexpected improvements in longevity influence the pay out benefits by prolonging the pay out period longer than anticipated.

The historical development of mortality rates is commonly modelled by decomposing the mortality rates into variables of age, period and cohort. Each of these variables contributes to the understanding of how mortality rates evolve subject to the ageing of a population, the effect of medical advancements and socioeconomic progress over time as well as the lifelong influence that follow each individual from birth.

In 1992 Lee and Carter proposed a model widely used to forecast future mortality rates [21]. The Lee Carter (LC) model consists of a single age and period term explaining the evolution of mortality rates. In current literature the LC method is sometimes described as one of the most efficient methods to generate realistic forecasts of life expectancy [3]. Therefore, the method is widely used to produce reference values for other modelling methods. Ever since the LC model was introduced various extensions have been proposed such as inclusion of higher order and cohort specific effects [22].

The influence that follow an individual from birth is of significance within mortality research. In 1999 Willets published a paper that dramatically increased attention to the presence of cohort effects in mortality data [31]. He observed the phenomenon of a more rapid improvement in mortality for people born in the United Kingdom between 1925 and 1945 than the adjacent generations. Later on, the subject of cohort effects was also thoroughly investigated by the Continuous Mortality Investigation [7]. Shortly after, several other studies revealed similar mortality improvements in other European countries such as the Netherlands and Belgium [30]. Following a rising interest in cohort effects, Renshaw and Haberman adapted the original LC model by incorporating a cohort term which is known as the Renshaw Haberman (RH) model [25].

The purpose of this study is to investigate if cohort specific effects are present in the Swedish population and if the incorporation of a cohort term in the modelling could improve the projection of future mortality rates. UK data is used as referential data due to the evidence of cohort effects. The study

attempts a comparison based on quantitative and qualitative properties of the LC and RH models. Qualitative properties are assessed by considering the interpretability as well as the applicability of the models. Quantitative properties are evaluated by examining each model fit with regard to consistency with historical data and robustness of parameter estimates relative to the range of data employed.

To begin with, we provide initial theory and background on the LC and RH stochastic mortality models in Section 2. It is followed by specific theory that is relevant for the modelling and the comparative analysis in Section 3. In Section 4 we conduct an exploratory analysis and quantify the presence of cohort effects in the Swedish and UK population data. In Section 5 we perform a quantitative comparison of the models which is followed by the most important results. The results are later on discussed and analysed in Section 6.

2 Mortality models

In the following sections the general theory of the LC and RH model is provided along with explanation of cohort effects.

2.1 Notation

We define $\mu_{x,t}$ as the force of mortality which corresponds to the mortality rate for a specific age $x = x_1, x_2, ..., x_k$ given a certain calender year $t = t_1, t_2, ..., t_n$. We define the central mortality rate $m_{x,t}$ as the average number of deaths at year x during calender year t. Under the assumption that the force of mortality is constant over a one year interval of age $[x, x+1]$ and calendar year $[t, t+1]$, $m_{x,t}$ equals $\mu_{x,t}$ namely

$$m_{x,t} = \mu_{x,t} = \frac{D_{x,t}}{E_{x,t}}. \tag{1}$$

In the above equation $D_{x,t}$ refers to the number of deaths occurred at age x during calender year t. Moreover, $E_{x,t}$ correspond to the exposure of risk, thus the period under which the population where exposed to the risk of death at age x during calender year t [2]. In the following sections we will continuously use $\mu_{x,t}$ when referring to the mortality rate.

2.2 The Lee Carter model

In the original paper from 1992, Lee and Carter proposed a new model for forecasting mortality rates [21]. Nowadays, the model is commonly used to make long-run forecasts of age specific mortality.

The LC model integrates the trend that can be observed in a data set ranging over a long period, which Lee and Carter referred to as the index of mortality. The model is extrapolative and allows age specific mortality rates to decline exponentially without limit [22]. The changes in mortality are subject to both the time measured in calender years t and age k. The model is of the bilinear form

$$\ln[\mu_{x,t}] = \alpha_x + \beta_x^{(1)}\kappa_t + \epsilon_{x,t} \tag{2}$$

adopting a logarithm transformation of the force of mortality. The parameters α_x and $\beta_x^{(1)}$ are age functions and κ_t represent the main time trend index of the general mortality level. The latter function may also be referred to as period function.

In detail α_x is a static age function describing the general shape of the age specific mortality rates. Different age groups exhibit different mortality rates but in general the mortality varies as a function of age in a systematic way. If we consider e^{α_x} we find the general shape of the mortality curve across the age range without the effects of the overall time trend.

Parameters $\beta_x^{(1)}$ determine the pattern of mortality change across ages x. The coefficients measure the age specific deviations of mortality change from the overall trend. Therefore $\beta_x^{(1)}$ signify which rates decline rapidly or slowly in response to changes in κ_t. When $\beta_x^{(1)}$ is large at some age x, the mortality rate at that age is more sensitive to changes in the general level of mortality. On the contrary, when $\beta_x^{(1)}$ is small, the mortality rate at age x fluctuates modestly in response to changes in the general level of mortality. To illustrate, infant mortality is an example of volatile age specific mortality rates while mortality at older ages is more robust [23].

In theory $\beta_x^{(1)}$ are allowed to be both positive and negative indicating that mortality rates decline for some ages while rising for others. However, in practice all $\beta_x^{(1)}$ are assumed to be positive in the long run when the model is fitted to relatively long time periods. The age specific mortality rates are assumed to increase or decrease collectively due to the joint period effect κ_t.

However, the model allows the magnitude of the movements to differ across ages [23].

The κ_t explain how mortality effects evolve over time often referred to as the index of the general level of mortality. If κ_t is decreasing linearly over time then each age specific mortality rate will decline at its own exponential rate subject to the value of $\beta_x^{(1)}$. Thus, the slope of the κ_t curve multiplied by the value of $\beta_x^{(1)}$ determines the proportionate rate of change at any mortality rate. Consequently, the mortality rates $\mu_{x,t}$ are modelled as a stochastic process driven by the process of κ_t [22]. To summarise, the $\beta_x^{(1)}$ parameters describe how the mortality effects are distributed across ages and the κ_t parameters explain how they evolve over time.

The error terms $\epsilon_{x,t}$ are assumed to be $N(0, \sigma^2)$ distributed which reflect systematic and random variations such as age specific historical influences that are not captured in the model [21].

2.3 Cohort effects

If we consider the mortality rates for some subset of adjacent birth years, the cohort mortality refers to the mortality rates observed for that generation from birth to death. In this case, we refer to the individuals born during a specific amount of time as a cohort. The study of cohort mortality is time consuming which follow from the fact that it takes a long time before the last living individual in a generation has deceased. Nonetheless, it proves important to the insurance business [2].

A cohort effect is defined as variations within a population with some common characteristic. In this context, the cohort effect refers to the evolution of mortality rates in a specific generation. The year of birth defines the common characteristic and the cohort effect is examined by modelling the impact of birth year on rates of mortality change. Consequently, a cohort effect may be interpreted as an effect that reduces or increases the mortality at all ages for a specific generation [19].

Essentially, there exist a dependency between the mortality for a population and the period under which the population lived. The development of health care systems, new innovations and other events that are time specific such as catastrophes, wars or environmental disasters naturally influence longevity. On the one hand, certain events during a period affect mortality across all age groups. On the other hand, some alterations in mortality originate from

events during the birth year. The aftereffects stemming from such events remain throughout life and may be interpreted as cohort effects. As a result, a significant amount of information is gathered during these years that should not be neglected. Ultimately, the information concerning social and economic conditions under these years improves the understanding of the life expectancy of an individual at a specific age.

Willets observed a rapid mortality improvement for people born between 1925 and 1945 centred on the generation born in 1931 in the UK population. The observations was concluded as cohort effects [31]. As a result, the models allowing for cohort terms outperformed other models that excluded such terms [8]. Willets discussed several possible factors that may have been fundamental in actuating the observed mortality improvement. He emphasised that the prevalence of smoking along with trends in heart diseases and cancer where contributing factors to the aggregation of a cohort effect [31, 32].

2.4 The Renshaw Haberman model

As a consequence of the increased attention in cohort effects, Renshaw and Haberman generalised the LC model by incorporating a cohort term. The model proved to capture the underlying effects related to the birth cohorts in mortality data. If cohort effects were present in the data the model fit and the mortality projections improved significantly [25].

Presently, we define the RH model as

$$\ln[\mu_{x,t}] = \alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(0)}\gamma_c + \epsilon_{x,t}. \tag{3}$$

We let $c = t - x$ denote the cohort where $c \in \{t_n - x_k, ..., t_1 - x_1\}$ hence ranging from the oldest to the youngest cohort in the data. The notation in the model along with the interpretation of the parameters α_x, $\beta_x^{(1)}$ and κ_t is analogue to the one in Section 2.2. The final terms $\beta_x^{(0)}$ and γ_c are included in the LC model specification in order to incorporate the cohort effects. A decline in the γ_c suggest a mortality improvement for the generations born during that period. On the contrary, a pronounced incline in γ_c would suggest a negative effect on longevity to those born during that time period. The γ_c correspond to the cohort specific effect while the $\beta_x^{(0)}$ adjust the impact of the cohort effect on the age specific mortality rates.

In other papers, issues have been reported regarding robustness of (3) subject to deviations in the period of data used in the fitting [7, 9]. In addition,

the model have proved to suffer from convergence issues when the model was fitted to certain periods of data or ranges of age. As a solution, Renshaw and Haberman suggested a parsimonious model structure by setting $\beta_x^{(0)} = 1$ [16]. Lack of robustness in the parameters can be related to the number of parameters that are fitted during the fitting algorithm [15]. As the number of parameters is decreased, a higher rate of convergence is achieved. In following studies, increased robustness was observed when the simpler model was compared to the predecessor hence the simpler model proved less sensitive to changes in the period of data used [18]. In the continuous investigation, we refer to (4) as the RH model

$$\ln[\mu_{x,t}] = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \gamma_c + \epsilon_{x,t}. \tag{4}$$

The interpretation of the α_x, $\beta_x^{(1)}$, κ_t and $\epsilon_{x,t}$ parameters in (4) is analogue to those defined previously. However, the restriction of setting $\beta_x^{(0)} = 1$ allows the cohort terms to vary over a wider spectrum, thus representing consistently higher or lower mortality rates.

3 Model fitting

Lee and Carter applied a two-stage estimation procedure to fit the model. During the first stage Singular Value Decomposition (SVD) was used to obtain estimates of $\beta_x^{(1)}$ and κ_t. However, Lee and Carter noted that the total observed number of deaths was not guaranteed to equal the total number of deaths predicted by the model. A second stage estimation was necessary to compensate the eventual difference in the number of observed versus the predicted number of deaths. As a solution, they suggested a re-estimation of κ_t as a second stage in the estimation procedure. New estimates of κ_t were computed for each calendar year t by finding the corresponding value that equalled the observed and predicted total number of deaths [21].

The proposed two-stage estimation has advantages as well as disadvantages which have been discussed in several studies [14,33]. Lee and Carter mentions the occurrence of negative and positive values of the $\beta_x^{(1)}$ parameters, which indicate that mortality is increasing in some age groups while decreasing in others. However, all $\beta_x^{(1)}$ are assumed positive in the long run [21]. While the LC model was developed for all cause deaths, the occurrence of nonuniform signs is more common when investigating cause specific effects. In addition,

Girosi and King points out that the appearance of nonuniform signs could lead to multiple or even no solutions to the estimation problem [14].

In a later study, Wilmoth noted that the difference between the number of observed and predicted number of deaths originated from the minimising of the least square error over log mortality instead of the mortality. As a consequence, same weights were given to all age groups independent of their contribution to the number of total deaths [33].

As a solution, Wilmoth proposed that the problem could be avoided by fitting the model using weighted least squares (WLS) or maximum likelihood estimation (MLE) and setting weights equal to the number of deaths. Both procedures naturally manage the occurrence of observations where zero number of deaths are registered which pose as an advantage over SVD. In a discussion paper, it was mentioned that the SVD approach was not optimal. Alternatively, the author assumed a Poisson regression setting where the parameter estimates were obtained using MLE [1]. Similar topics have been discussed in related articles where the MLE has been proposed to achieve a superior estimation of the parameters along with statistical packages implementing strategies to model bilinear terms [5].

In the following sections, we choose to exclusively consider the MLE procedure. The model fitting algorithm is described in Section 3.2.

3.1 Parameter estimation

Since the number of deaths are a counting variable, a Poisson distribution has been proved to be plausible [5]. Assuming that the number of deaths $D_{x,t}$ follow a Poisson distribution, it follows that

$$D_{x,t} \sim Po(E_{x,t} \cdot \mu_{x,t}) \tag{5}$$

where the expected value of the number of deaths is defined as

$$\hat{d}_{x,t} = E[D_{x,t}] = \hat{\mu}_{x,t} \cdot E_{x,t}. \tag{6}$$

The interpretation of $\mu_{x,t}$ and $E_{x,t}$ remains as previously defined in Section 2.1. As previously discussed, parameter estimates of stochastic mortality models can be obtained by implementing the maximum likelihood method,

which essentially maximises the model log likelihood. The likelihood function is determined as

$$\mathcal{L}(d_{x,t}, \hat{d}_{x,t}) = \prod_{x,t} \frac{e^{-\hat{d}_{x,t}} \cdot \hat{d}_{x,t}^{\,d_{x,t}}}{d_{x,t}!} \tag{7}$$

Taking the logarithm of the likelihood function, the log likelihood function is obtained, namely

$$\ln\big(\mathcal{L}(d_{x,t}, \hat{d}_{x,t})\big) = \sum_{x,t} \omega_{x,t}\big\{ - \hat{d}_{x,t} + d_{x,t} \cdot \ln(\hat{d}_{x,t}) - \ln(d_{x,t}!)\big\} \tag{8}$$

where $\omega_{x,t}$ is introduced as a $\{0, 1\}$ weight matrix that determines which cohorts are to be given zero weight in order to exclude some observations. Since the last term in (8) is independent of $\hat{d}_{x,t}$ it is sufficient to maximise the following equation to obtain parameter estimates

$$l(d_{x,t}, \hat{d}_{x,t}) = \sum_{x,t} \omega_{x,t}\big\{ - \hat{d}_{x,t} + d_{x,t} \cdot \ln(\hat{d}_{x,t})\big\}. \tag{9}$$

3.1.1 Lee Carter model

In the case of the LC model defined in (2), the objective is to obtain estimates of α_x, $\beta_x^{(1)}$ and κ_t. Since mortality data is structured as a collection of elements, the mortality rate is dependent of row and column identified by the age x and the year t. Models for such array data structure often suffers from over-parametrization. Consequently, the parametrization in the model (2) is not unique since the model is invariant with respect to the transformations of the model parameters

$$\tilde{\alpha}_x \longrightarrow \alpha_x + c_1\beta_x^{(1)}, \quad \tilde{\beta}_x^{(1)} \longrightarrow \frac{\beta_x^{(1)}}{c_2}, \quad \tilde{\kappa}_t \longrightarrow c_2(\kappa_t - c_1). \tag{10}$$

This can be illustrated by substituting the transformations into (2) which will result in an identical value of $\ln[\mu_{x,t}]$. Thus, the over-parametrization results in an infinite number of solutions to the estimation which in turn generates equivalent predictions. In order to obtain unique parameter estimates, constraints are imposed on the parameters. We choose to impose

the constraints that were originally introduced by Lee and Carter [21]. The constraints are

$$\sum_t \kappa_t = 0, \quad \sum_x \beta_x^{(1)} = 1. \tag{11}$$

In order for the parameters to satisfy the restrictions defined in (11), a set of constants, c_1 and c_2, is determined to identify a set of parameters given by the transformations in (10). To illustrate, the constants c_1 and c_2 are determined by substituting the transformations (10) into the constraints (11) and solving for each constant. As such, the constants satisfy the constraints and ensure that unique parameter estimates are obtained in the model fitting. Thus, unique parameters are obtained using the following values of c_1 and c_2

$$c_1 = \frac{1}{n}\sum_t \kappa_t, \quad c_2 = \sum_x \beta_x^{(1)} \tag{12}$$

where n denotes the number of calendar years. The first constraint in (11) implies that the parameter α_x correspond to the main age effects averaged over time [25].

The second constraint in (11) refrains the $\beta_x^{(1)}$ parameters from taking large values which in turn modulates the influence of the κ_t parameters on the mortality rates. However, it does not restrict $\beta_x^{(1)}$ from taking negative values.

Finally, it is emphasised the identifiability constraints in (11) are not unique and it is important to note that other constraints might result in different interpretations of the parameters [18]. Regardless of the choice of constraints, the fitted values of the log mortality will be the same, but the parameter estimates will vary.

3.1.2 Renshaw Haberman model

The original RH model has been subject to criticism due to slow convergence and insufficient robustness to changes in the underlying data [11]. Other studies have identified that the model were unable to converge for some subsets of data and that the model were sensitive to the choice of starting values in the fitting algorithm [7]. As a solution, it was proposed that a parsimonious model (4) should be used for making predictions [16]. Consequently, we con-

duct our analysis using the model specification in (4) with $\beta_x^{(0)} = 1$. Thus, the objective is to obtain parameter estimates of α_x, $\beta_x^{(1)}$, κ_t and γ_c.

Similar to the LC model, the RH model is subject to over-parametrization and to ensure identifiability in the model three constraints are imposed. The constraints select one set out of a range of equivalent parameter sets which yields the same fitted mortality rates. Consequently, the parametrization in the model is not unique since the model (4) is invariant with respect to the parameter transformations

$$\begin{aligned} \tilde{\alpha}_x &\longrightarrow \alpha_x + c_1\beta_x^{(1)} + c_2, & \tilde{\beta}_x^{(1)} &\longrightarrow \frac{\beta_x^{(1)}}{c_3} \\ \tilde{\kappa}_t &\longrightarrow c_3(\kappa_t - c_1), & \tilde{\gamma}_c &\longrightarrow \gamma_c - c_2 \end{aligned} \tag{13}$$

As discussed in the previous section, there exists several potential constraints that may be imposed to ensure identifiability. We choose to impose the constraints that was proposed in a recent paper [19]. In the aforementioned study, the constraints were developed based on observations in an earlier study [8]. The constraints are

$$\sum_t \kappa_t = 0, \quad \sum_x \beta_x^{(1)} = 1 \quad \sum_{\substack{x,t \\ c=t-x}} \gamma_c = 0. \tag{14}$$

The third constraint in (14), which is imposed on the cohort terms, yields the interpretation that the average level of the cohort parameters fitted by the model should be zero. The constraint on the cohort term does however not change the interpretation of the other parameters [5].

Analogous to the LC model, a number of constants, c_1, c_2 and c_3, is determined by substituting the transformations in (13) into the constraints in (14) and solving for each constant. Thus, unique parameter estimates are obtained using the following values of c_1, c_2 and c_3

$$c_1 = \frac{1}{n}\sum_t \kappa_t, \quad c_2 = \frac{1}{n+k-1}\sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c, \quad c_3 = \sum_x \beta_x^{(1)}. \tag{15}$$

In (15), the n and k refers to the number of calendar years and the number of cohorts respectively.

3.2 Fitting algorithm

There are various R packages available in the literature that have been developed over the years to fit stochastic mortality models. We utilised the R packages ilc and StMoMo to fit the LC and RH model respectively [6, 29].

In order to fit the LC model we applied a fitting procedure which is based on an iterative algorithm that minimises the deviance [25]. The procedure was adapted using previous work on fitting algorithms based on the log likelihood [5, 33]. The aforementioned fitting procedure is incorporated in the ilc package [6]. Similarly, we applied a fitting procedure which is based on the maximising of the log likelihood in (9) to fit the RH model [5]. The fitting methodology was implemented by using the StMoMo package [29]. Since both procedures follow a similar updating process, we describe the general structure of the fitting algorithm with regard to the minimising of the deviance in this section. Later on, in Appendix B, both procedures are described in detail.

As previously stated, the LC model fitting was conducted by implementing an iterative algorithm that minimises the deviance. The iterative fitting process generates the maximum likelihood estimates of the parameters. The updating mechanism followed a Newton Raphson iteration technique where the parameters were updated subject to the minimisation of the deviance. Given the parameter θ and the iteration step v to $v+1$, the parameters were updated according to

$$\hat{\theta}^{\{v+1\}} = \hat{\theta}^{\{v\}} - \frac{\delta D^{\{v\}}/\delta\theta}{\delta^2 D^{\{v\}}/\delta\theta^2} \tag{16}$$

where D is the deviance of the model. Assuming Poisson errors, the deviance can be expressed as

$$D(d_{x,t}, \hat{d}_{x,t}) = \sum_{x,t} \omega_{x,t} \cdot dev(x,t) \tag{17}$$

$$= \sum_{x,t} 2\omega_{x,t} \Big\{ d_{x,t} \ln(\frac{d_{x,t}}{\hat{d}_{x,t}}) - (d_{x,t} - \hat{d}_{x,t}) \Big\} \tag{18}$$

where $\hat{d}_{x,t} = E_{x,t} \exp\{\hat{\alpha}_x + \hat{\beta}_x^{(1)} \hat{\kappa}_t\}$ and $\hat{d}_{x,t} = E_{x,t} \exp\{\hat{\alpha}_x + \hat{\beta}_x^{(1)} \hat{\kappa}_t + \hat{\gamma}_c\}$ in the LC and RH model fitting respectively. In the above notation $dev(d_{x,t}, \hat{d}_{x,t})$ denotes the unit deviance. The deviance $D(d_{x,t}, \hat{d}_{x,t})$ ultimately compare the

log likelihood of the fitted model with the saturated model that is, the model corresponding to a perfect fit. In Appendix B, the updating schemes for the LC and RH fitting are described in detail.

3.3 Forecasting

In order to produce forecasts of mortality rates using the LC and RH model, the age specific parameters α_x and $\beta_x^{(1)}$ are assumed to remain invariant of time. As a consequence, the forecasting of mortality rates relies on time series projections of the period and cohort parameters, κ_t and γ_c. The mortality rate projections are calculated by using

$$\hat{\mu}_{x,t_n+s} = \exp\{\hat{\alpha}_x + \hat{\beta}_x^{(1)}\hat{\kappa}_{t_n+s} + \hat{\beta}_x^{(0)}\hat{\gamma}_{t_n+s-x}\} \tag{19}$$

where $\hat{\kappa}_{t_n+s}$ and $\hat{\gamma}_{t_n+s-x}$ are projected values of the period and cohort effects. The range of s is determined by the number of years that are included in the prediction and t_n correspond to the last calendar year in the model fitting. Note that $\hat{\beta}_x^{(0)}$ are set to 1 and 0 to obtain the forecasts of the RH and LC model respectively.

In practice, it is common to extrapolate the period and cohort terms by stochastic Auto Regressive Integrated Moving Average (ARIMA) processes, which are dependent of the parameters (p, d, q) [21, 25]. The ARIMA model is an extension of an ARMA model that is necessary to consider when a non-stationary time series is investigated. The component p correspond to the number of lag observations in the model, commonly recognised as the lag order. The value of p implies the number of times that the variable of interest is regressed on its prior values. The d represents the degree of differencing which indicates the number of nonseasonal differences. Differencing is applied to a non-stationary time series in order to make it stationary. Lastly, the component q is the number of lagged forecast errors in the prediction equation [28].

In the original paper, Lee and Carter considered several univariate ARIMA specifications to model κ_t. However, they found that the random walk with drift, ARIMA$(0, 1, 0)$, deemed most appropriate [21]. In agreement with most practices, the κ_t is therefore modelled and projected as a random walk with drift which is recognised as

$$\hat{\kappa}_t = \theta + \hat{\kappa}_{t-1} + \epsilon_t \tag{20}$$

where θ equals the drift term that captures the steady trend between time periods and the $\epsilon_t \sim N(0, \sigma^2)$. In other words the time trend index at time t equals last years value κ_{t-1} increased by a slow steady change θ and some noise ϵ_t. If the drift term is positive then the κ_t are subject to an upward trend while a negative value implies a declining trend. An estimate of the drift term can be obtained by using MLE, namely

$$\hat{\theta} = \frac{\hat{\kappa}_{t_n} - \hat{\kappa}_{t_1}}{t_n - t_1} \tag{21}$$

where t_1 and t_n correspond to the first and last calendar year within the investigated interval [28]. Consequently, the estimate is dependent of the first and last estimate of $\hat{\kappa}_t$.

In agreement with previous studies, the cohort effects are extrapolated by an ARIMA$(1, 1, 0)$ process under the assumption that there exists independence between the period and cohort effects [10,25]. Thus, we regress the difference of γ_c according to

$$\hat{\gamma}_c = \theta + \hat{\gamma}_{c-1} + \phi_1(\hat{\gamma}_{c-1} - \hat{\gamma}_{c-2}) + \epsilon_c \tag{22}$$

where ϕ_1 equals the autoregressive coefficient. The models in (20) and (22) are used to produce forecasted values of the period and cohort terms which in turn are used to derive the forecasted mortality rates in (19). For further information on the technical details of the parameter estimation and forecasting of ARIMA models, we refer the reader to prior literature [4, 28].

As a remark, despite previous findings, it is reasonable to examine the trend of the above effects during the model fitting. If there exists a linear tendency in the annual mortality improvements, the ARIMA$(0, 1, 0)$ remains appropriate in accordance with previous studies. However, if non-linear structures appear then other ARIMA processes should be considered. Similar argumentation applies to the fit of an ARIMA$(1, 1, 0)$ process in the case of cohort effects.

Accuracy of a forecast can be evaluated by considering the level of consistency between the realised and the predicted mortality rates associated with each forecast. In Section 5 we choose to forecast the mortality rates between 2000 and 2017 thus for every predicted value we have a corresponding realised value. As a result, it is possible to measure forecast accuracy by comparing

each set of values. The forecast error (FE) relative to the realised value is quantified by

$$FE = \frac{|\mu_{x,t_n+s} - \hat{\mu}_{x,t_n+s}|}{\mu_{x,t_n+s}}, \quad s > 0. \tag{23}$$

Ultimately, the forecast accuracy can be measured by taking the inverse of the forecast error.

4 Data

Data on human mortality is available from the Human Mortality Database (HMD) [17]. We specifically use the death rates and central exposures of risk of Swedish and UK mortality data in 1×1 intervals corresponding to age and time intervals respectively.

Given the population we divide the data into two sub-populations of males and females respectively. Their mortality rates differ hence the analysis should be held separate of gender. Due to legislation implemented from 2012 insurers are not allowed to use gender-based factors in the calculation of premiums and benefits [12]. However, every insurance business may internally distinguish females and males in order to ensure that a sufficient amount of capital is deposited to cover future claims.

Furthermore, it is important to distinguish between the observed mortality for the population as a whole and for the part of the population with insurance coverage. Naturally, every individual is not obliged to buy private insurance and the ability to purchase insurance cover relies on macroeconomic as well as socioeconomic factors. As a result, mortality data for the insured population is a subset of the population as a whole wherefore it is not sufficient to accurately determine the mortality trend. One proposed solution is to conduct the analysis using data for the population as a whole and then calibrate the results using the information obtained from the data on the insured population [13].

In the following two sections Swedish and UK data are discussed emphasising the properties of mortality data in each country. We choose to include figures depicting the life expectancy of each population. The development of mortality rates across time and age was also investigated but the summarised measures of life expectancy is included to illustrate the general development.

4.1 Swedish mortality data

The rates are available for the Swedish population between 1751 to 2007 and may be extracted subject to sex, age and time. Due to uncertain quality of the collected data during the early years we will exclude data prior to 1860. The collection and administration of data has subsequently improved such that data from the beginning of the 20th century is considered of good quality [27]. Consequently, we will confine the analysis to data collected between 1900-2017.

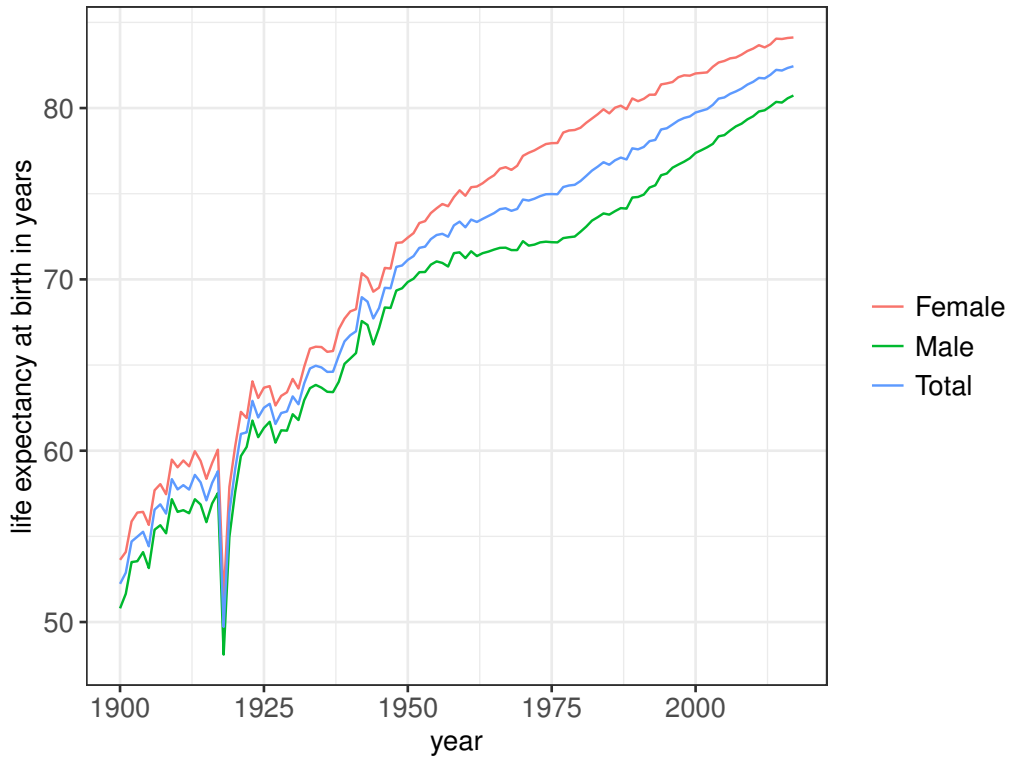


Figure 1: *Sweden: Development of life expectancy at birth in years between 1900 and 2017.*

We refrain from analysing infant mortality since the trends in mortality rates for infants differ from the trends observed among older ages. Additionally, the infant mortality levels observed for developed countries have been steadily decreasing and are currently very low. This trend is also observed in Sweden which is shown in Figure 19. Therefore, they have little impact on the summary measures such as life expectancy and for investigating trends for adults. Further, we restrict the analysis to the age range 20-95 which is of

most interest to providers of pensions and annuities.

In Figure 1 life expectancy at birth indicate a steady augmenting trend for female and males in Sweden between 1900 and 2017. Initially, most of the increasing life expectancy is due to improved health in the younger population especially for infants. As time progresses, the health improvements shift to the older population. During the war period in the first part of the 20th century we observe larger fluctuations in life expectancy. One major downturn is observed around year 1918 which originates from the Spanish flu pandemic that spread shortly after the ending of the World War 1. Individuals at younger ages between 20 and 40 were more susceptible to dying in the pandemic, thus the peak in the mortality rate originate mainly from those ages. Over time, there is a consistent augmenting trend in life expectancy and the fluctuations slowly diminishes toward the latter part of 1900.

Consequently, a general decline is observed in the overall mortality rates during this period but the trend of decline is not constant over the years. Instead the degree of decline varies at different ages whereas a significant amount of variation is visible at older ages.

4.2 UK mortality data

The HMD data is available for the UK population since 1922. During the war period between 1939 to 1950 the data comprises only the civilian population. Furthermore, we note that the UK population is more than six times larger than the Swedish population. As a result, there exist more underlying data to calculate the mortality rates, especially at higher ages.

As seen in Figure 2, the observations for Swedish life expectancy applies to the UK life expectancy. The general mortality trend is similar to the Swedish data. We observe large fluctuations during the earlier years in the data. There are two revealing downturns observed shortly before 1930 and in 1940. In 1930, an increase in mortality rates was probably caused by a higher frequency of deaths from lung cancer and heart disease. A similar argument was presented regarding the downturn in 1940 [32]. Around 1950 the trend changes and the slope of the curve becomes less steep during the latter part of the 20th century. In contrast to the Swedish data, there is a larger difference between the male and female life expectancy during the years shortly after the war period. However, the mortality rates in both populations seem to have evolved similarly over time. During 2016 the mortality rates were 82.84 and 79.18 years respectively for females and males in the UK

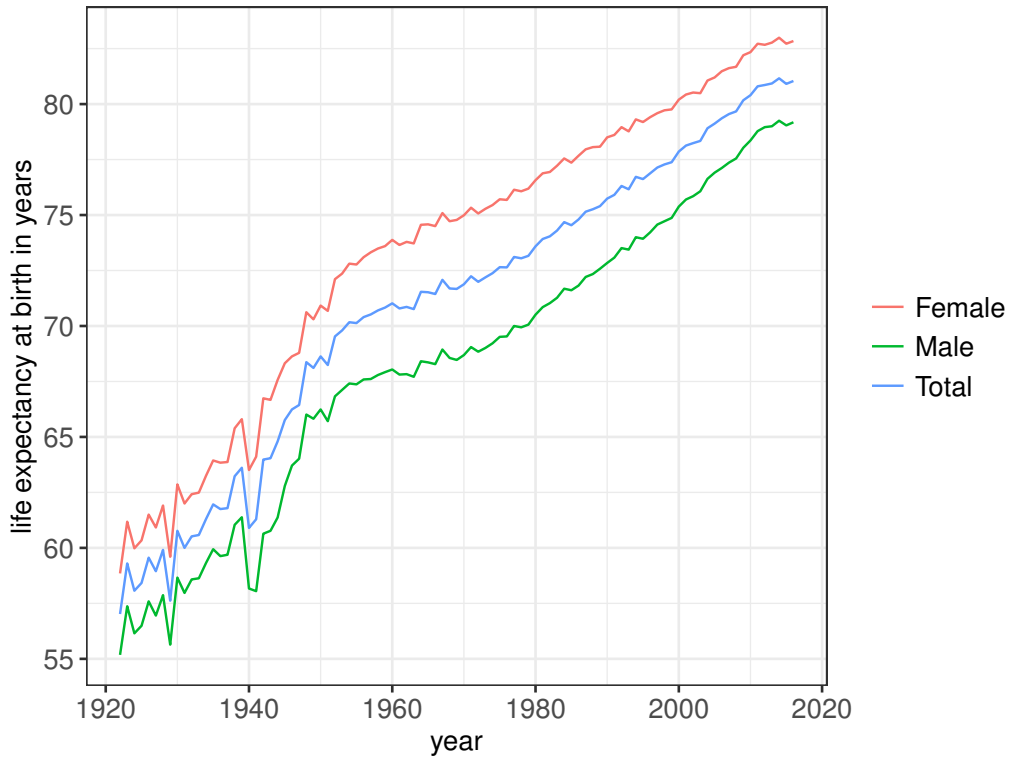


Figure 2: *UK: Development of life expectancy at birth in years between 1922 and 2016.*

population. The corresponding rates for the Swedish population are 84.09 and 80.57 respectively.

5 Analysis

In the following sections a comparative analysis of the LC and RH models is conducted. Initially, we explored the full data from the Swedish and UK population in order to evaluate if cohort effects were present. Later on, the focus was to investigate the model fit to historical data. Each model fit was assessed by inspection of the standardised residuals. Under the Poisson distributional assumption that the number of deaths are independent of age and calender year, the standardised residuals are independent and identically distributed. In addition, the robustness of the parameter estimates was investigated by exploring the parameter estimates relative to changes in the period of data used to fit the models. Thereafter, we investigated how

robustness issues impact the mortality projections. Based on initial our observations, we proceeded to analyse a subset of the population data in depth in order to closely evaluate the error of prediction.

5.1 Standardised deviance residuals

Residuals can be utilised to assess the model fit to the data. Diagnostic checks of the fitted model is conducted by visual inspection of the plotted residuals. In order to investigate if cohort effects are present in the Swedish and UK data the LC model is fitted to the full range of data. Failure to capture any effect emerges as non-random patterns in the residuals. Therefore, residual patterns demonstrate eventual features in the data which are not successfully captured by the fitted model.

The standardised deviance residuals were computed according to the following formula

$$r_{Di} = \sqrt{\frac{\omega_{x,t} \cdot dev(d_{x,t}, \hat{d}_{x,t})}{\phi}} \cdot sign(d_{x,t} - \hat{d}_{x,t}) \tag{24}$$

where ϕ is the dispersion parameter and $dev(d_{x,t}, \hat{d}_{x,t})$ denotes the unit deviance which was defined in (17). Essentially, the dispersion parameter is estimated based on the deviance

$$\hat{\phi} = \frac{D(d_{x,t}, \hat{d}_{x,t})}{n - r} \tag{25}$$

where n and r are the numbers of observations and parameters respectively. Additionally, the definition of $D(d_{x,t}, \hat{d}_{x,t})$ in (17) remains. Residuals are graphed and analysed based on the pattern of standardised residuals against age, year of birth and calendar year. Heat maps are constructed by plotting age against calendar year.

In Figure 3 and 4 we observe the heat maps for Swedish and UK female data. The corresponding heat maps for male data is observed in Figure 21 and 22. The different colours designate the value of the residuals ranging from negative to positive. A negative sign demonstrates a larger fitted than observed value whereas a positive sign demonstrates the opposite. In Figure 3 there are wider diagonal bands along with narrow diagonal lines. The

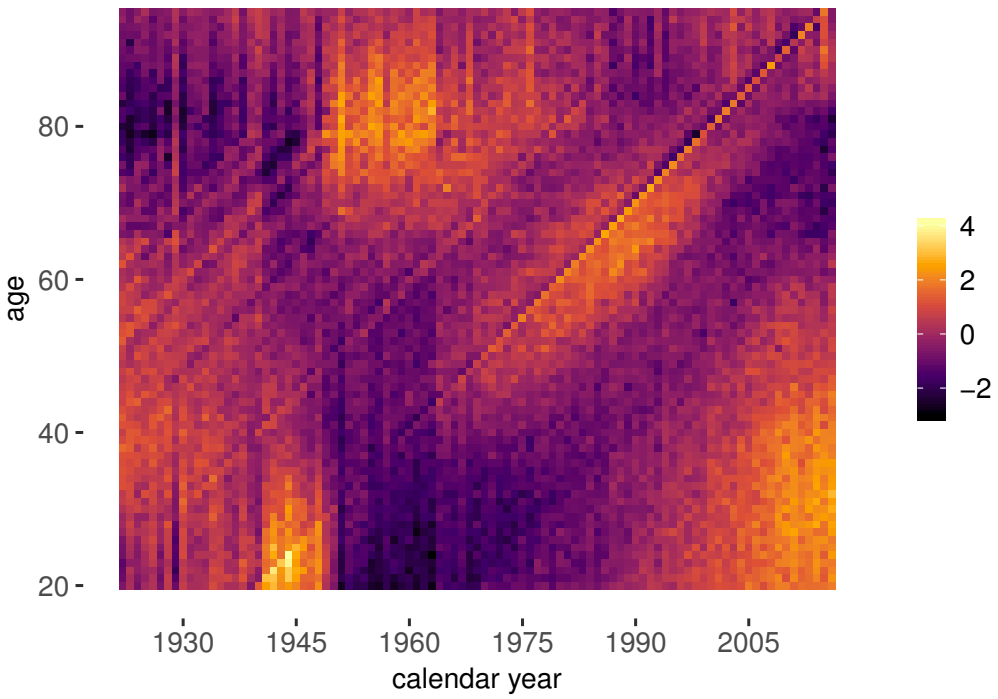


Figure 3: *UK: Standardised residual plot depicting the LC model fit to female data between 1922 and 2016. Diagonal clusters of residuals signify cohort effects.*

existence of two distinct lines centred around 1920 and 1940 reflects linear relationships that the model failed to capture. The diagonal clustering of negative and positive residuals in Figure 3 provide evidence of a cohort effect. In addition, there seems to be a bias toward negative residuals which indicate a prevalent underestimation. Additional light clustering for ages 20 and 80 is present which demonstrates an overestimation.

In contrast to the distinct patterns observed in Figure 3, the residual patterns in Figure 4 are less defined. Firstly, we observe a vertical line at 1918 depicting an overestimation at younger ages and an underestimation at older ages. As previously observed, a peak in mortality rates during this period were presumably emanating from the outbreak of the Spanish flu. Secondly, it appears to be a clustering of the residuals in the upper centre, suggesting some variability that the model is incapable to capture. In the upper centre this is viewed as a lighter shaded area which implies an underestimation of

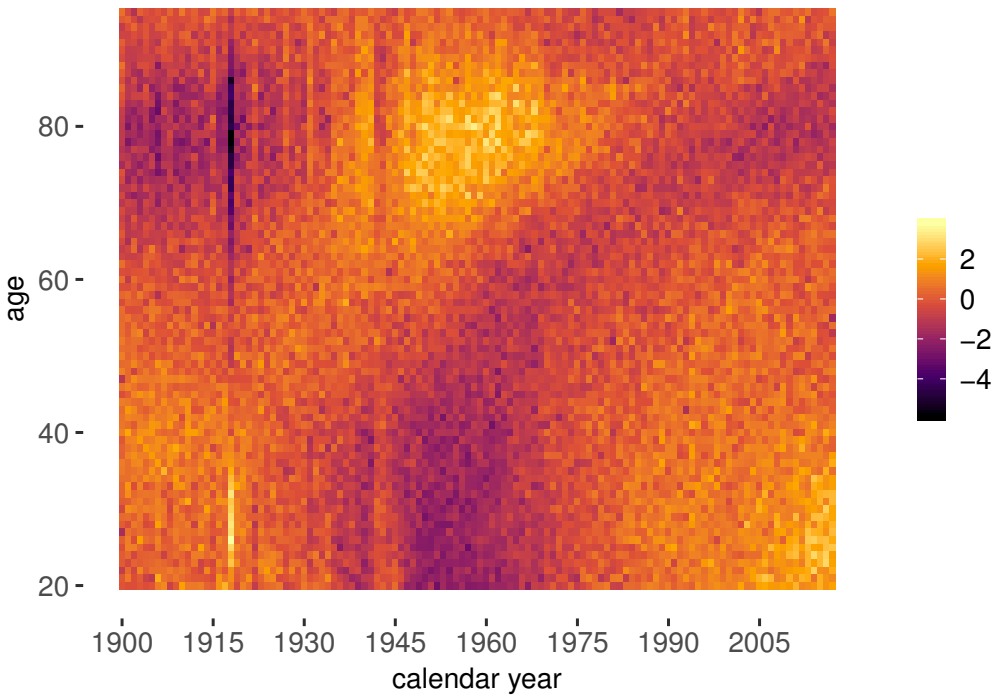


Figure 4: *Sweden: Standardised residual plot depicting the LC model fit to female data between 1900 and 2017. Diagonal clusters of residuals signify cohort effects.*

the observed values. However, the overall bias toward positive residuals is substantially different from Figure 3. A slightly curved diagonal clustering is apparent from age 40 from the 1945.

Similar conclusions are drawn in response to the observations in Figure 21 and 22. The major difference between the male and female Swedish data appears as an increased bias toward negative residuals. In the UK male data a clear underestimation is present at younger ages between 1945 and 1990.

As a complement to the heat maps we investigated the residual patterns in scatter plots which are displayed in Figure 23, 24, 25 and 26. Especially, the residual patterns in sub figures 23a, 24a, 25a and 26a are particularly revealing. The patterns suggest that there exists some variability due to year of birth that is not accounted for in the model. From the 1950, the plots exhibit positive trends. It becomes evident that the LC model is unable to integrate this feature. The patterns in Figure 24a and 26a are however

notably different from the corresponding plots in Figure 23a and 25a. Clearly, the cohort effect is stronger in the UK data.

In the preceding analysis, different non-random patterns appeared in the plotted residuals. The observations suggest that some features in the data are not incorporated by the LC model. Furthermore, the occurrence of diagonal clusters in the heat maps suggests some dependence between the mortality rate and the year of birth. Although the residual patterns in the heat maps of the Swedish data are less distinct, the patterns reflect a similar trend as in the corresponding plots for the UK data. Therefore, we can not neglect that the LC model fails to capture all the predictive information in the data.

5.2 Robustness in the parameter estimates

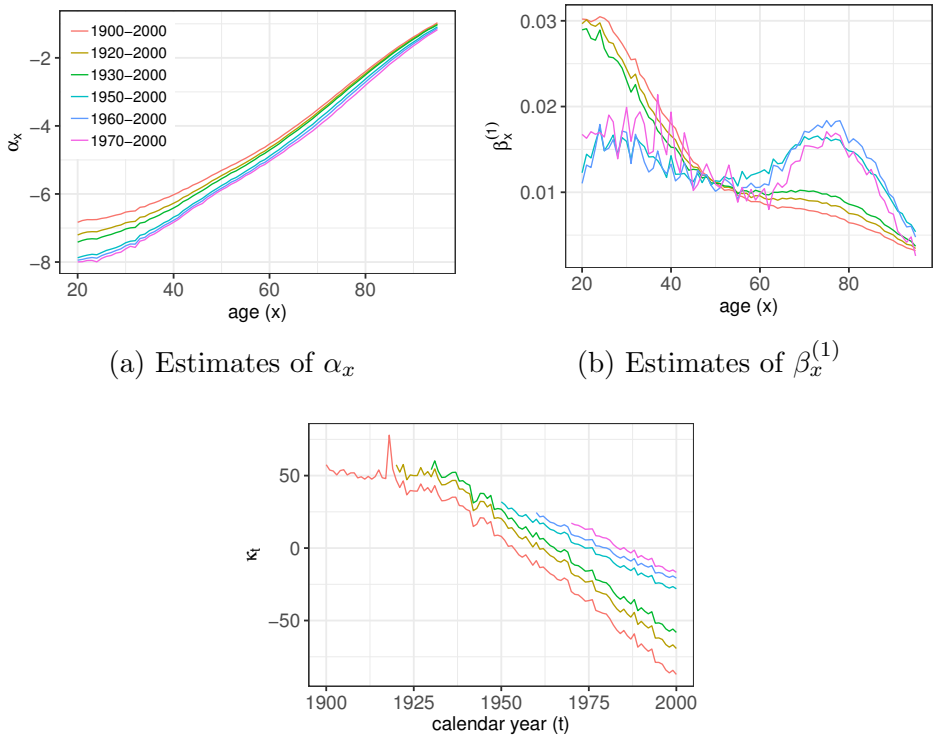
Since mortality rate time series span over a long time period it is necessary to consider the non-stationarity in the data. Over time, each decade possesses particular characteristics that is reflected indirectly in the mortality rate. As a result, the effect of certain time specific events persists over a longer period of time.

In order to sufficiently capture the mortality trend there is a trade-off between the amount of data used and the non-stationarity, making it difficult to forecast. In particular, the decision to use more recent data or to widen the range of data is highly relevant. The wider scope of the study, the more probable it is to include effects that may not be relevant to the current mortality. Either the fitting provides better estimates of the time trend or a less stable estimate, but nonetheless more accurate.

In the following sections, the robustness of the estimated parameters are assessed by considering their values relative to different time periods in the fitting procedure. The parameter estimates are considered robust if shorter data periods induce marginal deviations in the estimated values. Based on our observations in Section 4, we divide the Swedish and UK data into subsets. As such, we subsequently exclude the most volatile years. Thereafter, the LC and RH model is fitted to each subset. The aim is to investigate how sensitive the parameter estimates are to trends or events observed during certain decades.

5.2.1 Models fitted to Swedish data

Earlier in Section 4.1, we discussed the development of life expectancy at birth. Two discontinuities was observed in 1918 and 1940 where the latter was not as severe as the first. The adjacent years did not display a similar increase whereas the event did not affect the overall mortality rate. In addition, we observed larger fluctuations during the years predating the world wars than the years that followed. The point marking the clear trend change seemed to be around year 1950. As a consequence, data from the earlier part of the 20th century might not be representative of the current mortality trend. In Figure 5 and 6, we observe the results of respective model fitting.



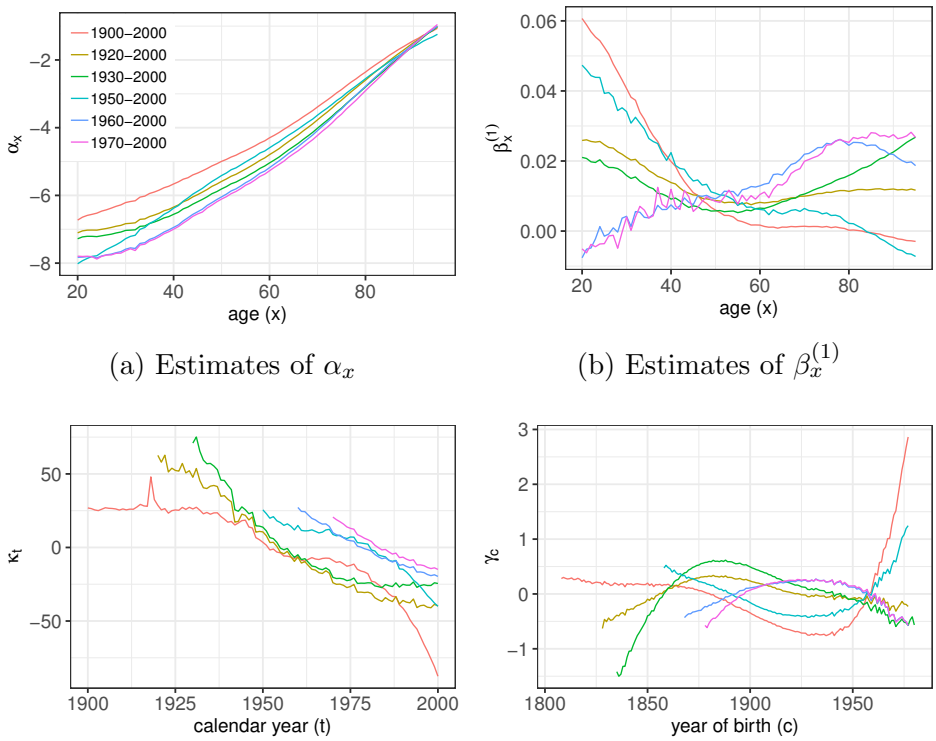
(a) Estimates of α_x

(b) Estimates of $\beta_x^{(1)}$

(c) Estimates of κ_t

Figure 5: *Sweden: Parameter estimates of LC model fitted to different time periods of female data.*

In Figure 5c, the general trend of the mortality index κ_t appears consistent over the model fittings. The general trend is linear and decreasing which confirms the expectation that mortality declines over time. The discontinuity at year 1918 clearly marks the impact of the Spanish flu. In the full data the



(a) Estimates of α_x

(b) Estimates of $\beta_x^{(1)}$

(c) Estimates of κ_t

(d) Estimates of γ_c

Figure 6: *Sweden: Parameter estimates of RH model fitted to different time periods of female data.*

estimates of κ_t seem to accelerate toward the latter part of the 20th century. As κ_t goes to a large negative number, the age specific effects decline and converges to zero. The overall linear trend concludes that the age specific mortality has fallen nearly exponentially over time. During the second half of the 20th century, the development of the mortality index is smooth and linear.

As seen in Figure 5b, there are particular differences in the estimated values of $\beta_x^{(1)}$ across all ages. The estimates are all positive indicating a decreasing mortality. In the models fitted to the wider data ranges, the figure illustrates a pronounced decline at younger ages between 20-60. Thereafter, the curve flattens before a modest decline at older ages between 70-95. As a result, the age specific mortality rates were subject to larger reduction at younger than at older ages. The pattern of the estimates obtained from the models fitted to the shorter periods is different from the previous fits. For ages

between 30-60 the estimates are decreasing while increasing at older ages with a peak at approximately age 75. The increase in $\beta_x^{(1)}$ at older ages signifies an abundant fall in mortality. Large fluctuations in the estimates are observed for the models fitted to the shorter time periods whereas a less volatile diminishing trend is observed for the estimates based on the full time period. Clearly, the estimates are sensitive to the time period considered in the fitting. However, the results seem to be consistent over adjacent time periods.

In Figure 5a, which depicts the estimates of $\hat{\alpha}_x$, the main difference appears at younger ages. The general trend of α_x is similar in every model which confirms the assumption that the general level of age specific mortality increases as age increases.

Corresponding plots of the parameter estimates obtained from the RH model are found in Figure 6. There is a strong resemblance between the estimates of α_x for the RH and LC model, whereas previous observations apply. However, the deviation in between time periods is larger than previously observed.

In Figure 6b, the produced estimates of the interaction terms $\beta_x^{(1)}$ appears to be smoother than observed previously in Figure 5b. All estimates are positive except for some of the youngest and oldest ages, which indicate that mortality is increasing for some ages while decreasing for others. The estimates of the model fitted to wider data ranges follow a decreasing slope similar to the observation in Figure 5b. The other estimates follow an increasing slope which signify an abundant decrease in mortality at older ages.

Large deviations are observed in between the fitted models. This is especially prominent in Figures 6b and 6d where the models fitted to the shorter data ranges exhibit dissimilar patterns. In Figure 6d, the estimates of γ_c obtained from the longer time periods show a modest decrease during the early birth years with a more prominent decrease between 1875 and 1900. For the youngest cohorts born in 1950 and above, the estimates of γ_c significantly increase in some cases, while a decrease occurs in others. Over all time periods, both negative and positive estimates of γ_c occur. The negative sign of the cohort terms results in lower age specific mortality rates. The estimates of the cohort term obtained from the shorter time periods, show opposite results. The estimates follow a concave shape between the birth years from the late 19th century to the latter part of the 20th century. Thus, the results indicate a decrease in mortality for the youngest and oldest cohorts.

Opposite to the findings in Figure 5c, the estimates of the period effect in Figure 6c appear to be non-linear. Extending further back in time the

linearity of decline cease to hold. In Figure 6c the fitted values from the widest data range exhibits a flat region between 1900 and 1935 except for the peak in 1918. A similar flat region is observed between 1955 and 1970. Since the κ_t reflects the annual changes in the general level of mortality, this interprets as if no annual changes occurred during these years. Moreover, it is observed that as the scope of the time series decreases the linear decline returns.

5.2.2 Models fitted to UK data

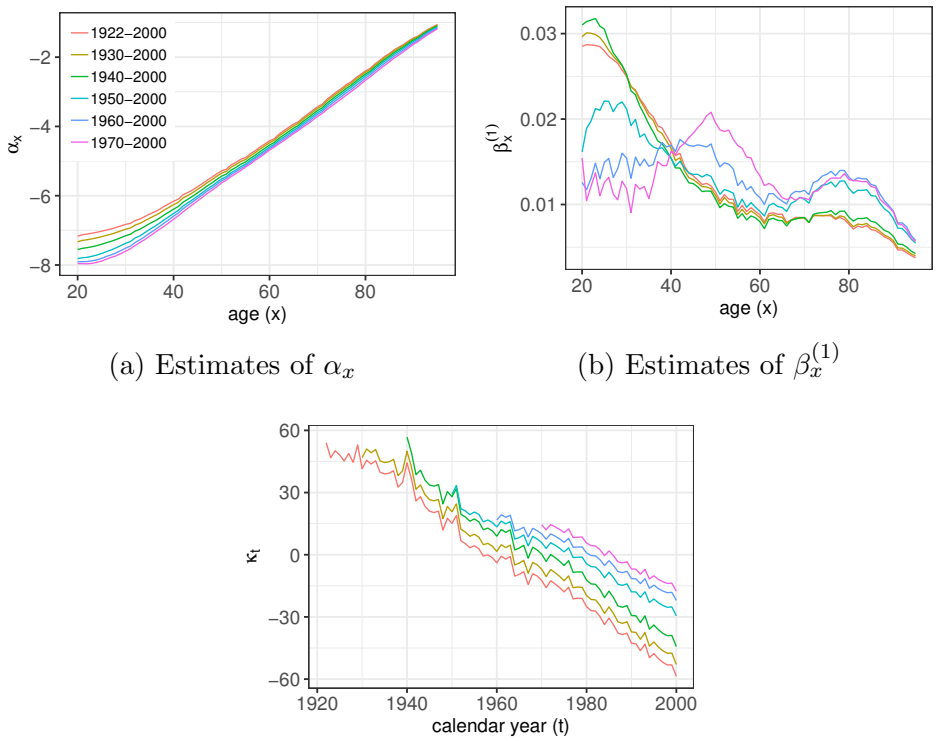
As observed in Section 4.2, the UK mortality trend indicate two discrepancies present in year 1930 and 1940. The overall trend of the life expectancy in Figure 2 implies a volatile period before year 1950, followed by a less volatile period. The estimated parameters obtained from each of the LC and RH model fit are illustrated in Figure 7 and 8.

In Figure 7a, it is observed that the estimates of α_x are steadily augmenting. A larger variation in between the model fits appears for the younger ages while the estimates converge toward the older ages. The observations correspond to the familiar characteristics of the α_x. The results of α_x obtained from the models fitted to UK data indicate that the overall mortality level is similar to the Swedish trend.

In Figure 7b, it is found that the estimates of $\beta_x^{(1)}$ obtained from the model fitted to the wider data ranges exhibit similar trends as in the Swedish data. The mortality declines faster at ages 20-60 before slowing down at ages between 60-95. The estimates obtained from the shorter time periods show two distinctive peaks, one in between 40-50 and the other at approximately age 80. Thus, the observations indicate a more abundant decrease in mortality at those ages.

As seen in Figure 7c, the κ_t estimates is subject to discontinuities at year 1930, 1940 and 1950. The peak at year 1940 is significantly prominent. The estimates obtained from the model fitted to the wider data ranges follow a steeper curve in contrast to the shorter data ranges. However, the mortality index is clearly linearly decreasing.

In Figure 8d we observe particularly distinguishing patterns of the cohort term. The overall trend seem to be similar to that observed for the Swedish models. However, we observe a pronounced decline in γ_c during the interwar years between 1925 and 1945 as well as a discontinuity around year 1918. The discontinuity at year 1918 can be identified as effects stemming



(a) Estimates of α_x

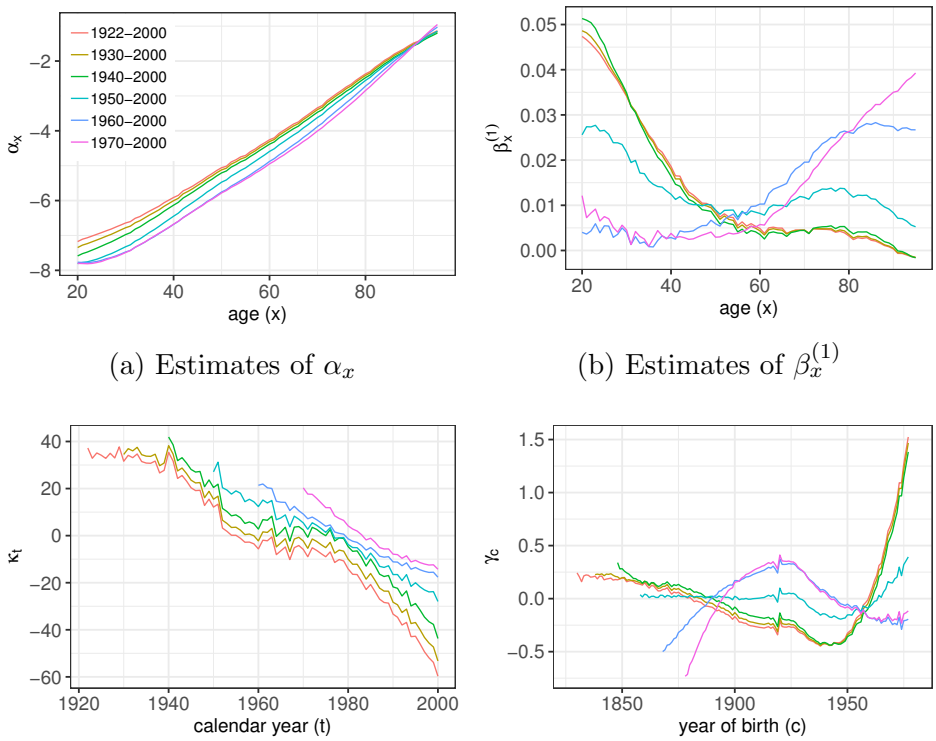
(b) Estimates of $\beta_x^{(1)}$

(c) Estimates of κ_t

Figure 7: *UK: Parameter estimates of LC model fitted to different time periods of female data.*

from the influenza epidemic. The impact of the epidemic was not integrated in the mortality index as in the Swedish data, thus the influence emerges in the cohort terms. However, the estimates obtained from the model fitted to the wider data ranges show a modest decline. The findings of a decline between 1925 to 1945 are consistent with the observed mortality improvement that occurred for the generations born during this period. For the youngest cohorts and the wider period ranges, an abundant increase in the estimates is observed which in turn influence the mortality negatively. A similar increase was observed in Figure 6d. Consequently, additional analysis is required to establish the potential causes that may have induced an increase. In absence of such an analysis, it may be argued if the results are reasonable relative to the observed level of mortality.

In Figure 8c, the estimates of κ_t depicts nonlinear patterns. However, the curvature is most prominent for the wider data ranges where the mortality



(a) Estimates of α_x

(b) Estimates of $\beta_x^{(1)}$

(c) Estimates of κ_t

(d) Estimates of γ_c

Figure 8: *UK: Parameter estimates of RH model fitted to different time periods of female data.*

index seem to decelerate at year 1960 followed by an acceleration toward the later part of the 20th century. Since the overall trend is decreasing, the observations conclude that the age specific mortality has decreased exponentially over time.

Lastly, we observe similar patterns in the estimates of $\beta_x^{(1)}$ in Figure 8b, as was observed in the model fittings to the Swedish data. A rapid decline is observed at younger ages which devolves into a modest decline at older ages. As the data range in the fitting decreases the slope of the $\beta_x^{(1)}$ changes to an upward trend indicating a faster mortality decline at the older ages.

5.3 Forecasting mortality rates

In the following sections we investigate the sensitivity of the mortality projections due to changes in the fitting period. We forecast mortality rates

between 2000 and 2017 using the fitted models in the previous section.

Mortality projections for the UK and Swedish data by the LC and RH models are generated and presented in Figure 9 to 12. As the period range of the fitted model decreases, the difference between mortality rates per projected year decreases. Due to a shorter time period in the fitting procedure, the volatility of the projections increase.

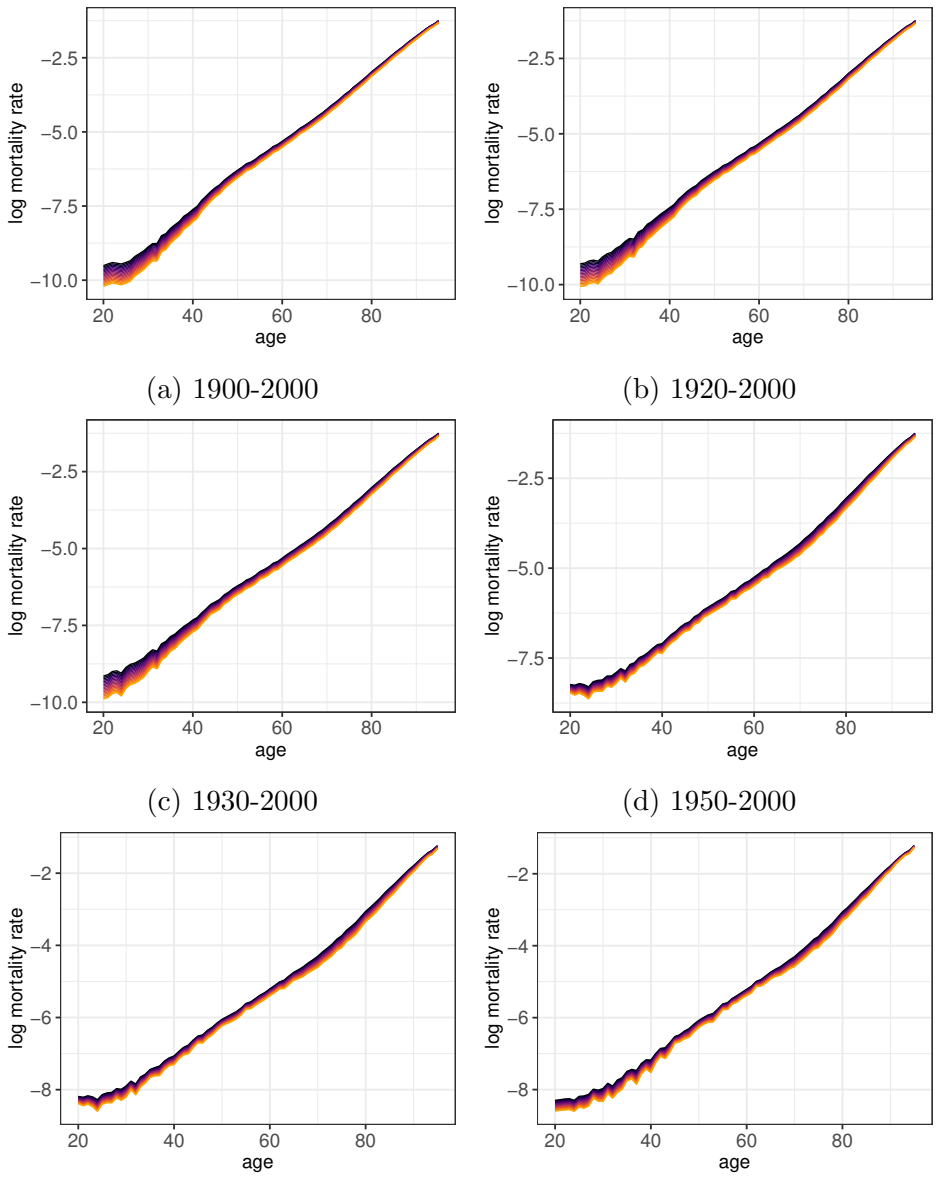
In Figure 9 it appears that the LC model produces fairly similar mortality projections subject to all time periods. The main difference appears in Figures 9a, 9b and 9c which depict lower mortality rates especially at the younger ages. The broader band of projections at the younger ages also suggest that they are subject to a more rapid mortality improvement than the older ages. This is especially apparent up to age 40. In the other projections the projected mortality improvement is modest.

In particular, there are significant differences in the projections produced by the LC and RH model. The LC model generates over all smooth mortality projections except at younger ages. On the contrary, the RH model generates fluctuating projections as observed in Figure 10 and 12.

In Figure 10a, the mortality rates are predicted to increase at ages ranging between 30-70. Instead in Figure 10b, the exclusion of the first 20 years including 1918 eliminates the effect which suggests that the model is sensitive to outliers. However, the pattern is restricted to the first projection using the full data. Due to the extrapolation of the period and cohort effects, the previous steep declining and inclining trends observed in Figures 6c and 6d are reflected by a projected continuous decrease and increase of both effects. As a result, an increase in future mortality is observed for the youngest cohorts. As previously discussed, such results may not be reasonable which suggest that the wider data ranges may not be suitable to predict current mortality. A similar, but less distinct pattern, is observed in Figure 10f. The development diminishes as the period range decreases and the reverse is observed for the shorter data ranges.

In Figure 11, the mortality projections based on the LC model appear to be less sensitive to the period range. In Figures 11a, 11b and 11c the predictions exhibit similar patterns and the trend slowly changes as the period range decreases. Consequently, it can be concluded that the LC model appears to be robust to changes in the period of data used. Similar to the observations in Figure 9, the band of mortality projections subsequently contracts implying a lower rate of mortality improvement over time.

In Figure 12, an increase in mortality rates is observed at younger ages in

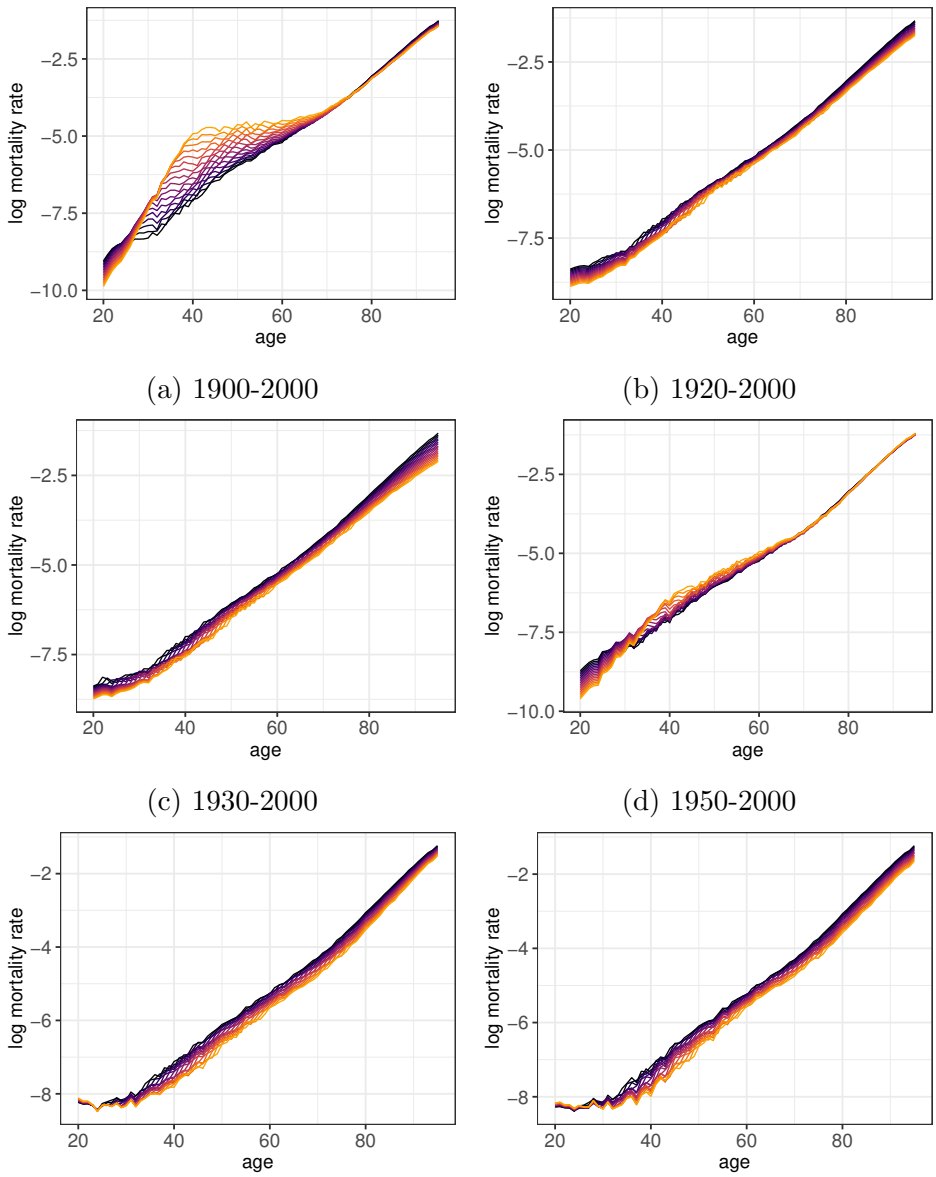


(a) 1900-2000 (b) 1920-2000

(c) 1930-2000 (d) 1950-2000

(e) 1960-2000 (f) 1970-2000

Figure 9: *Sweden: Log mortality rate projections for 2001-2016 from LC models fitted to different periods of data. Each line represent a 1-year forecast indexed by age. Colours are ordered from darkest to lightest representing 2000 to 2017.*

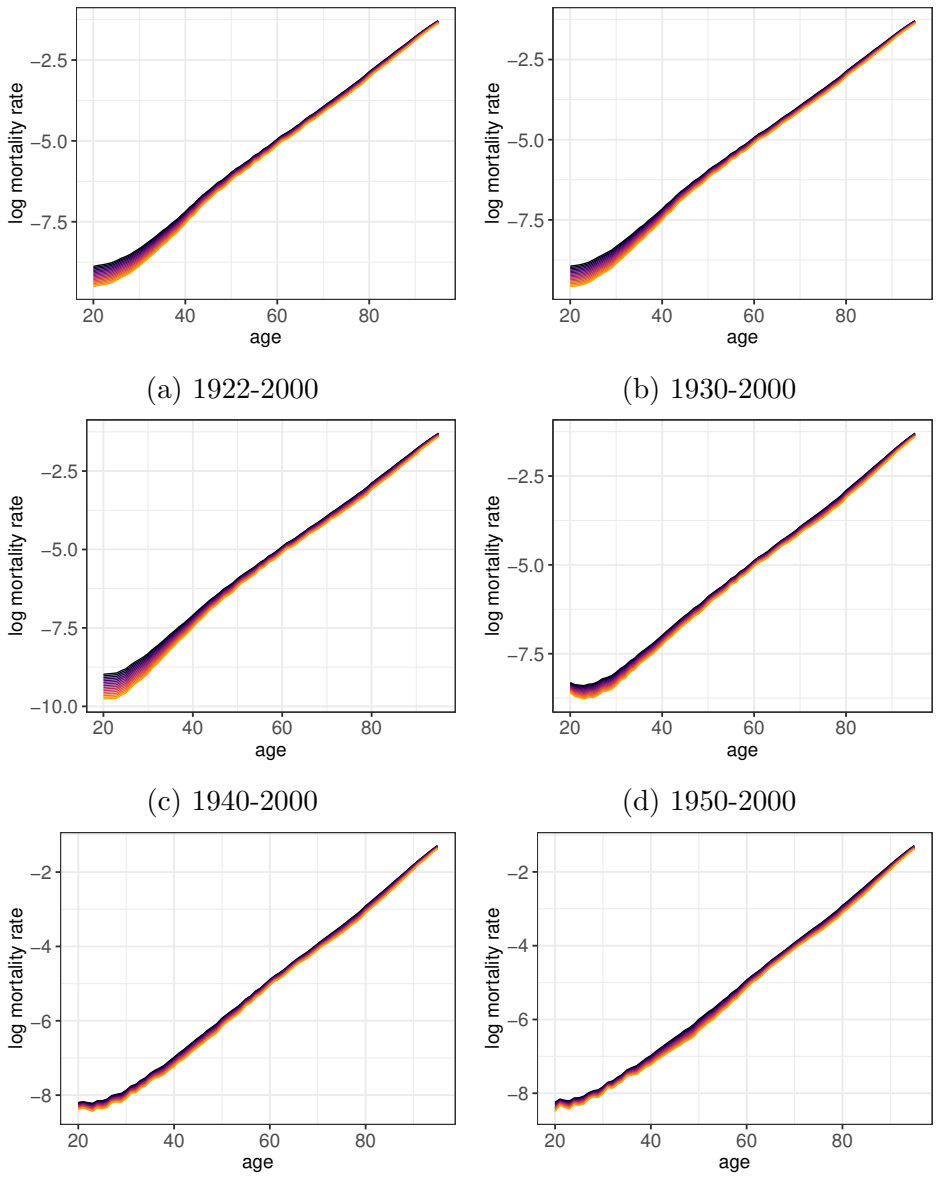


(a) 1900-2000 (b) 1920-2000

(c) 1930-2000 (d) 1950-2000

(e) 1960-2000 (f) 1970-2000

Figure 10: *Sweden: Log mortality rate projections for 2001-2017 from RH models fitted to different periods of data. Each line represent a 1-year forecast indexed by age. Colours are ordered from darkest to lightest representing 2000 to 2017.*



(a) 1922-2000

(b) 1930-2000

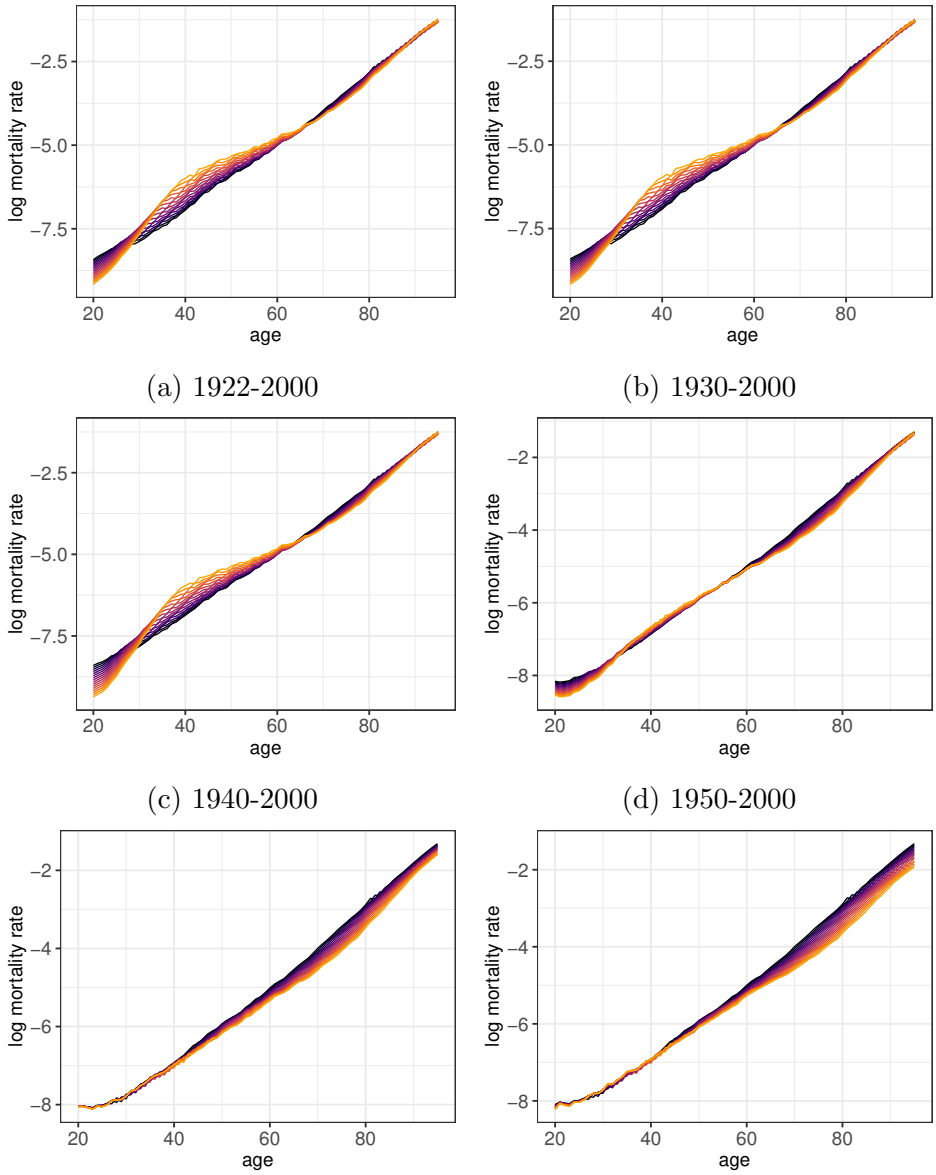
(c) 1940-2000

(d) 1950-2000

(e) 1960-2000

(f) 1970-2000

Figure 11: *UK: Log mortality rate projections for 2001-2016 from LC models fitted to different periods of data. Each line represent a 1-year forecast indexed by age. Colours are ordered from darkest to lightest representing 2001 to 2016.*



(a) 1922-2000 (b) 1930-2000

(c) 1940-2000 (d) 1950-2000

(e) 1960-2000 (f) 1970-2000

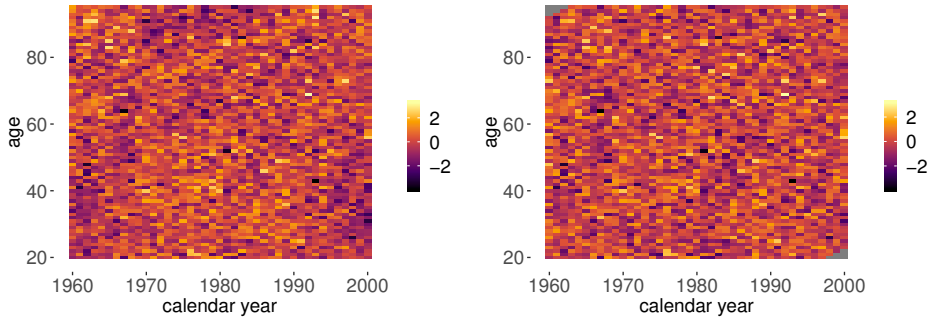
Figure 12: *UK: Log mortality rate projections for 2001-2016 from RH models fitted to different periods of data. Each line represent a 1-year forecast indexed by age. Colours are ordered from darkest to lightest representing 2001 to 2016.*

the first three figures of mortality projections. This effect can be traced to the trends of the period and cohort term preceding the extrapolation, as was observed previously for the Swedish data. In Figure 12e and 12f, the development due to decreasing period range is notably different. The older ages are subject to a higher mortality improvement while the younger ages induce a modest improvement which is consistent with the observations in Figure 8b and 8d.

5.4 Fitting period 1960-2000

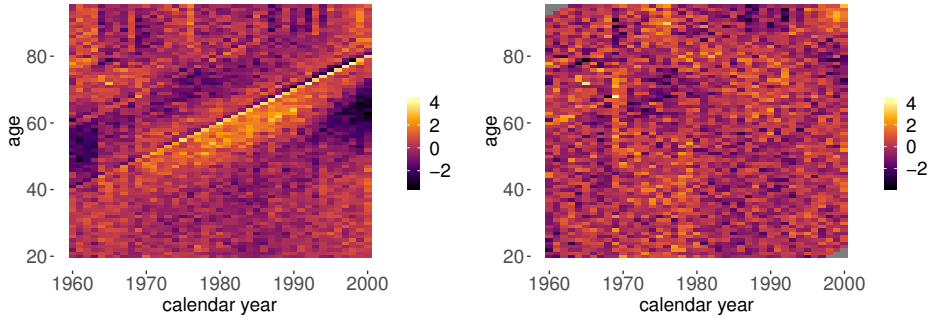
In order to investigate the mortality projections further we focus on one period range. The data range need to be sufficiently long in order to generate reliable estimates. Since the overall trend in the mortality projections seem to stabilise as a shorter data range is used, further analysis is conducted using the data between 1960 and 2000. The data range is also chosen in order to achieve a result that is comparable to previous studies [25].

To begin with, the deviance residuals are observed in Figure 13 and 14. In Figure 14 there is a significant reduction in residual patterns for the RH model in contrast to the patterns observed for the LC model. On the contrary, the residual patterns look more or less the same in Figure 13. Since the assessment depends on a visual inspection, a simple correlation test was performed in order to quantify the existence of significant correlations across year and age groups. A t-test was used to determine if the correlation coefficients were significantly different from zero. The null hypothesis denotes that the correlation coefficient equals zero and the alternative hypothesis denotes that the correlation coefficient is different from zero. The test follows a t-distribution under the null hypothesis with $N-2$ degrees of freedom and the test statistic used was $\rho(\sqrt{N-2})/(1-\rho^2)$ where ρ denotes the Pearson correlation coefficient. The null hypothesis is rejected on a 99% significance level if the p value < 0.01. Thereafter, we computed the fraction of significant correlations and summarised the results in Table 1. A significant reduction in both cross-age and cross-year correlations is observed for the UK. A less distinct reduction is present for Sweden. However, the results quantifies a reduction in correlations even in Sweden but nonetheless it proves to be small.



(a) LC Model (b) RH model

Figure 13: *Sweden: Standardised deviance residuals of LC and RH model.*



(a) LC Model (b) RH model

Figure 14: *UK: Standardised deviance residuals of LC and RH model.*

Country	UK		Sweden	
Model	LC	RH	LC	RH
Cross-age	12.59	3.62	2.22	1.73
Cross-year	32.06	6.90	3.69	3.21

Table 1: *Percentage of significant correlations on a 99% significance level.*

In addition, a Shapiro-Wilks normality test was conducted to investigate if the residuals are consistent with the normality assumption. We found that the null hypothesis could not be rejected hence the normality assumption held in any of the tests.

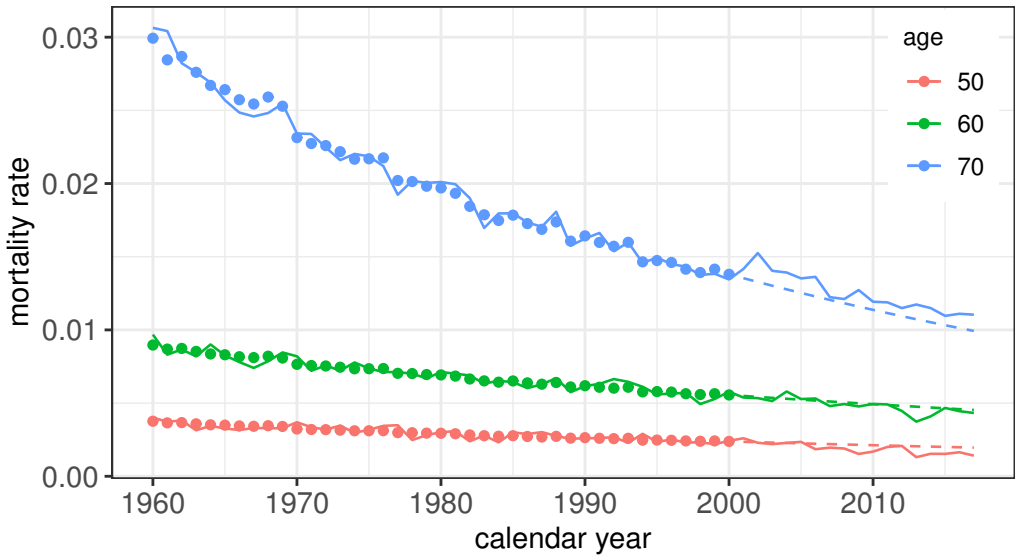
In Figure 15 and 16, the projected mortality rates from 2000 to 2017 for ages 50, 60 and 70 is graphed along with the historical and fitted rates for

the LC and RH model respectively. In Figure 15a, the projected Swedish mortality rates appears to be a good approximation of the realised values. In particular, the model captures the volatilities observed for age 80. However, the model underestimates the mortality rates at that age. As seen in Figure 15b, the mortality projections from the RH model clearly underestimates the mortality rates at age 70. However, the model seems to approximate the observed mortality rates at age 50 and 60. None of the models captured the upward turn observed for age 70 which is observed exclusively at that age.

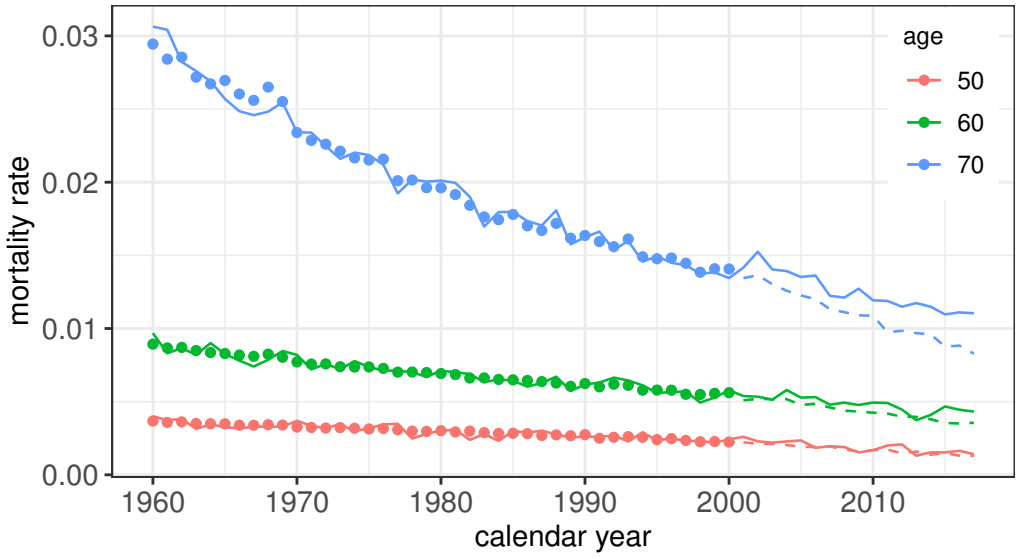
In Figure 16a, the LC model fitted to the UK data overestimates the future mortality rates at age 60 and 70. In particular, the difference between the forecasted and observed values is larger at age 70. In addition, it appears that the result is subject to increased volatility which is result of to the fluctuations in the observed values between 1960 and 2000. On the contrary, the forecasted mortality rates at age 50 correspond to the observed mortality rates. Furthermore, the observed mortality rates for the lower ages are less volatile year to year.

As seen in Figure 16b, the projected mortality rates by the RH model appear to be a better approximate of the observed mortality rates at every age. Especially, the RH model captures the pronounced decline in mortality that is eminent for age 70 during the first projected years. From year 2005 the decline in the observed mortality rates cease and shifts upward. At this point, the projected mortality rates continue the previous diminishing trend which result in an overestimation of the future mortality rates. However, the projected mortality rates at age 50 and 60 clearly capture the observed trend during these years.

In addition, the short term forecasting accuracy of each model was assessed by relating the predicted mortality rates to the realised mortality rates. In Section 3.3, we defined the forecast error as the difference between the predicted and observed values, normalised by the observed values. Intuitively the forecast error can be addressed as a percentage of deviation measuring the degree of fault in each prediction. The inverse of the forecast error, the forecast accuracy, is demonstrated in Figures 17 and 18.

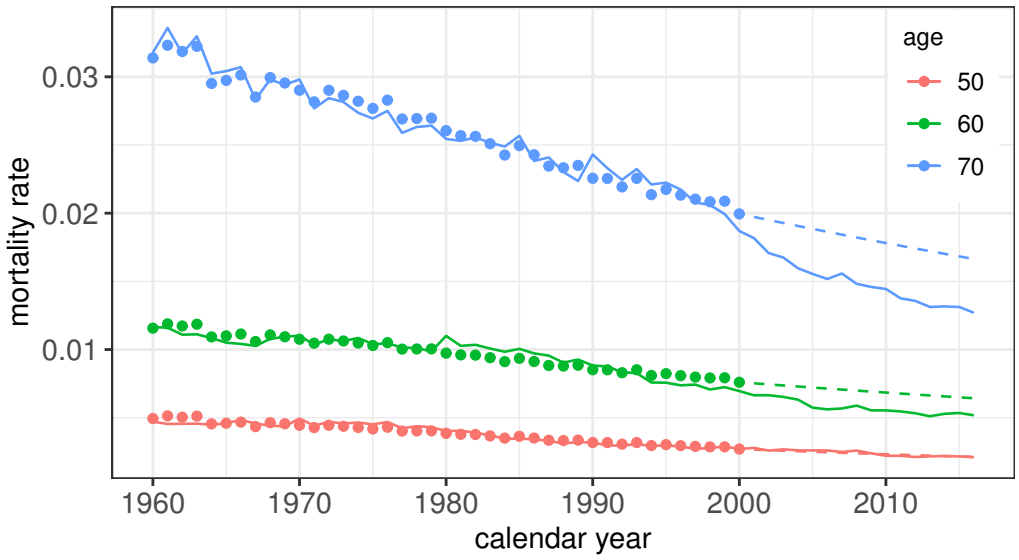


(a) LC Model

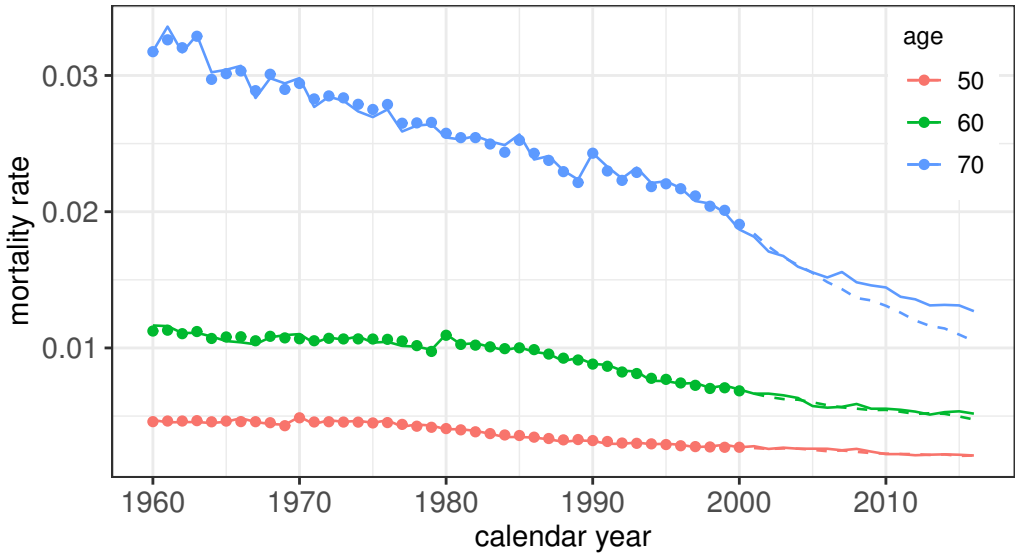


(b) RH model

Figure 15: *Sweden: Projected mortality rates from 2000 for ages 50, 60 and 70 indicated by the dashed lines. The solid lines correspond to the historical mortality rates between 1960 and 2017. The dots demonstrate the fitted rates for the respective model.*



(a) LC Model



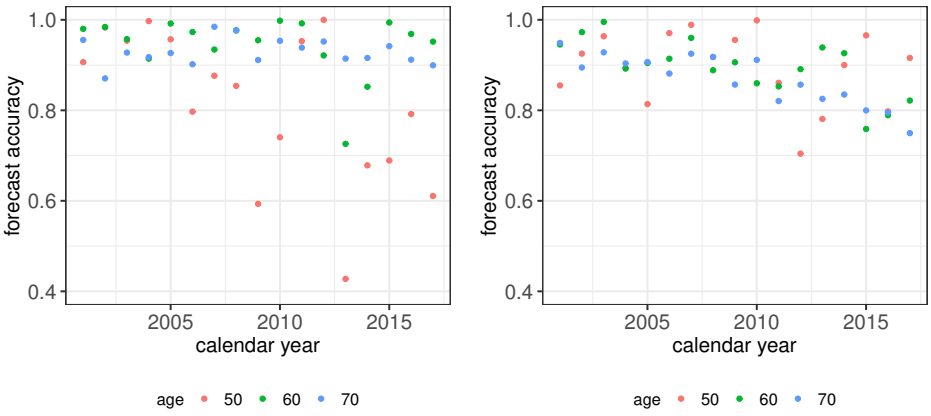
(b) RH model

Figure 16: *UK: Projected mortality rates from 2000 for ages 50, 60 and 70 indicated by the dashed lines. The solid lines correspond to the historical mortality rates between 1960 and 2017. The dots demonstrate the fitted rates for the respective model.*

The forecast accuracy in Figure 17a fluctuates between 42.76 and 99.98 percent while the observations in Figure 17b ranges between 70.43 and 99.87 percent. The highest errors are observed for the forecast at age 50 in Figure 17a. The forecast error clearly improves when the RH model is used to project the future mortality rates at that age. However, the pattern in Figure 17b suggests that the forecast accuracy slightly declines over time. In Figure 17a, the forecast accuracy seem to fluctuate similarly over the age groups 60 and 70. The pattern confirms that the LC model produce a highly accurate projections for age 70 as was observed in Figure 15a.

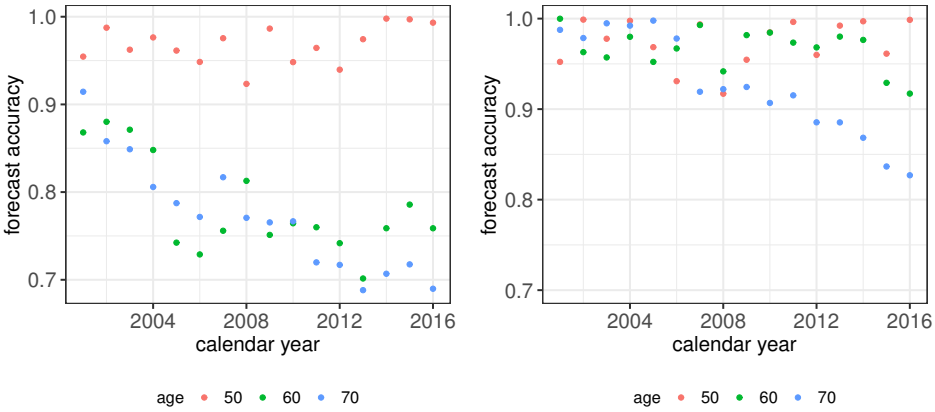
The observations in Figure 18a is notably different from Figure 17a. The LC model produce a highly accurate forecast of the mortality rates at age 50, while increasing errors are observed for the other ages. The observations in Figure 18b demonstrates a significant increase in forecast accuracy at ages 60 and 70 when the RH model is used in the projections. Additionally, the high forecast accuracy at age 50 is maintained.

In Figure 27 and 28, the forecast error is indexed by age and each line represents one annual forecast between 2000 and 2017. In the Swedish projections observed in Figures 27a and 27b, the forecast error converges to zero at ages above 60 while most of the variability is restricted to the younger ages. Shifting from the LC to RH model, an improvement of the forecast accuracy is obtained. In the UK projections, the forecast error in Figure 28a fluctuates around zero, with a higher variability at the youngest and oldest ages. In Figure 28b the forecast error converges to zero at ages above 80. For ages between 50 and 60, the forecast error increased compared to the LC model.



(a) LC Model (b) RH model

Figure 17: *Sweden: Forecast accuracy of 17 annual predictions of mortality rates for ages 50, 60 and 70 from the LC and RH models fitted to the period of data 1960-2000.*



(a) LC Model (b) RH model

Figure 18: *UK: Forecast accuracy of 16 annual predictions of mortality rates for ages 50, 60 and 70 from the LC and RH models fitted to the period of data 1960-2000.*

6 Discussion

Ever since the introduction of mortality models the overall development demonstrates that simplicity and interpretability in the model is premiered before complexity. Even though more complex models surpass the simpler models the question remains whether the incorporation of more effects are relevant to the forecasting of future mortality trends. Essentially, the main objective is to improve the forecasts by selecting a model that captures the persisting trend and represents the fundamentals of mortality patterns.

In agreement with preceding studies, the UK cohort effect was observed for the population born between year 1925 and 1945 centred on 1931 [31, 32]. The residual patterns observed in Section 5 supports the findings reviewed in earlier papers. Less distinct residual patterns was observed during the investigation of Swedish data which suggested that some variability was not accounted for by the LC model. As such, the findings supported the inclusion of an additional term incorporating cohort effects. The declaration of a Swedish cohort effect of such magnitude as observed in the UK is however unfeasible. Nonetheless, models that include cohort effects such as the RH model, persist as competitive candidates to model mortality rates.

Initially, robustness of the parameter estimates was investigated for each of the models. The analysis showed that the period of data used affects the variability of the estimates. Moreover, the results proved that the LC model was less sensitive to outliers and to changes in the period of data used than the RH model. Similar conclusions regarding robustness applied to both Sweden and the UK. The parameter estimates obtained from the RH model were subject to larger deviations in relation to adjacent periods of data used. Although, the deviations diminished between the periods 1960-2000 and 1970-2000 which suggested some stability, the overall results declared a lack of robustness in the model. The inconsistency in between time periods aggravates the interpretation of the estimates. Therefore, the plausibility of the RH model is uncertain. To summarise, the conducted analysis suggested that in terms of robustness in the parameters the LC model would be preferred.

Based on each fitting period, mortality rates were forecasted 17 years ahead. To begin with, it is important to note the difference between short term and long term forecasts. In terms of long term forecasts, irregular short term trends might not be relevant. Thus, the period used during the model fitting depends on the objective of the forecast of life expectancy. In general, a longer forecast requires a longer period of data to fit the model. In this study, the results implied that the more recent data that is used, the lower are the

forecasted mortality rates, which is in agreement with previous observations [1, 18, 20].

The LC model produced stable forecasts that were resistant to outliers and changes in the period of data used in the fitting. On the contrary, the observed robustness issues of the RH model was reflected in the mortality projections. As the period of data used decreased, a subsequent change was observed in the projected mortality rates. In terms of reliability of the mortality projections, the RH model proved less stable and therefore the LC model would be preferred.

In order to obtain additional information on the appropriateness of the RH and LC models an in-depth analysis was conducted based on the period of data 1960 to 2000. The Shapiro-Wilks test determined that the underlying normality assumption was fulfilled. In addition, cross-age and cross-year correlations was computed which quantified a decrease in the residual patterns. Despite a slight decrease in correlations for Sweden, the RH model overestimated the projected mortality improvement specifically at age 70. However, it is noted that the mortality at higher ages are subject to larger fluctuations making them difficult to forecast. Therefore, a similar but not as severe overestimation was obtained from the LC model. Moreover, the RH model proved to be a significantly better fit to the UK data which demonstrated the advantage of the RH model when a cohort effect is detected. In that case, the LC model failed to capture the predictive information dependent of year of birth, thus incorporating a cohort term improved the future mortality projections. This was especially eminent in the projections at age 70 where the RH model accurately captured the diminishing trend that was observed during the first forecasted years.

In terms of forecast accuracy a general improvement was obtained shifting from the LC to the RH model in the analysis of both Swedish and UK data. In the projections based on the Swedish data, the convergence of the forecast error at older ages improved thus demonstrating a reduced amount of variability. In the UK projections, the forecast error in the RH model exhibited decreased fluctuations in between forecast years but increased deviance in between age groups. Consequently, the RH model expressed improved forecast accuracy in comparison with the LC model.

Altogether, it was concluded that the RH model increased the understanding of the systematic effects in Swedish and UK mortality data. However, when a weak cohort effect is present, such as in the Swedish data, the inclusion of a cohort term do not exclusively result in improved prediction accuracy. The results points out the necessity of diversifying the analysis of mortality data

by considering a number of competitive models.

From an insurance business point of view, the reliability and accuracy of mortality projections is of great importance. The assumptions regarding mortality improvement directly affects pricing and reserving especially for life annuities. In the case of life and pension benefits, the calculation requires a suitable mortality projection such that underestimation of future costs are avoided. The consequences of a rapid and unexpected mortality improvement for one or several generations could lead to a potential loss of business capital. Essentially, life insurers aspire to reserve a correct amount of capital hence avoiding to reserve more than necessary. Ultimately, close monitoring of mortality and longevity risk is required in order to balance between too low and high assumptions.

Furthermore, it has been concluded that the tolerance level in the convergence criterion in the model fitting influence the parameter estimates [18,25]. An important note is to consider how the estimates change relative to a decreasing tolerance level. A disruptive pattern as compared to other tolerance levels may act as an indicator of problems in the fitting procedure. In this study, the tolerance level did not influence the parameter estimates. However, convergence issues occurred during the model fitting procedure of the RH model which was resolved by imposing an additional constraint in the fitting algorithm. Similar observations have been made in previous studies which has a negative influence on the usage of the RH model due to the amount of excessive work that is required [7, 18].

This study demonstrates that the field of mortality modelling and forecasting requires extensive research. Possible variations of the study could be to investigate grouped cohort effects by assigning a cohort factor to 5 year incremental data. Furthermore, it is noted that this study was conducted exclusively on population statistics to allow for a referential study against the UK. Since mortality rates are assumed to be lower among the population with private insurance coverage, this study could be continued by extending the investigation to analyse mortality data among such a population. Both the LC and RH models seem to underestimate the observed mortality rates thus additional investigation of other models could elaborate on the underlying factors.

A competitive model named Cairns-Blake-Dowd (CBD) was proposed in 2006 which included multiple age period terms [8]. Similar to the development of the LC model, various extensions to the CBD model has been proposed [9]. In addition, the Plat model was created as a hybrid using the advantages from the LC, RH and CBD models. The model incorporated the static age

function from LC model, the cohort term from RH model and the multiple parametric age function in CBD model [24]. Furthermore, the assessment of uncertainty in the mortality projections is an important topic which could be another possible extension of this study. The extrapolated time dependent parameters might not capture the full amount of variability in the future predicted values since it does not allow for uncertainty in the other model parameters. Other measures, such as bootstrapping, have later been proposed in order to explore the uncertainty in mortality risk [15, 26].

Finally, we remark that the results reported are restricted to the analysis of female mortality data. Difference in mortality data of female and males occur due to features in respective data. In male data larger variations are observed in older ages than in female data. Consequently, a similar analysis on male data yields a different outcome.

7 Conclusion and further remarks

The purpose of this study was to explore the possibility to obtain improved mortality rate projections by incorporating a cohort term in the modelling of Swedish morality data. We performed an exploratory analysis to investigate the presence of cohort effects followed by a quantitative comparison of the Lee Carter and Renshaw Haberman mortality models. Due to the well documented cohort effects in the UK population, we utilised UK mortality data as a reference to evaluate the findings in the Swedish mortality data.

A significant cohort effect of such magnitude as observed in the UK data was not present in the Swedish data. The ability to declare the presence of a cohort effect requires further research of socioeconomic and macroeconomic factors that influence the mortality rates in a population. However, the analysis demonstrated that a part of the systematic effects in the Swedish mortality data can be explained by including a cohort term. To demonstrate, the impact of the cohort effect was quantified by a decrease in cross-year and cross-age correlations.

The analysis proved that the LC model was less sensitive to outliers and changes in the data which resulted in stable predictions, while the opposite was observed for the RH model. Nonetheless, the comparison of forecasts against the realised outcomes revealed a good performance of the RH model. As a result, the RH model should not be neglected when analysing future mortality improvements, especially in terms of mortality projections. How-

ever, the lack of robustness in the RH model implies that we can not wholly rely on the predictions.

A good fit to the historical data and high forecast quality is of great importance to achieve reliable projections of mortality rates. In purpose of using mortality models in actuarial applications, a model with good forecasting performance is suitable. Therefore, the LC model may be preferred to the RH model due to the prominent stability in the projections. However, models without the inclusion of cohort effects may neglect important information in the underlying data wherefore the LC model may not be preferable in case of forecasting mortality. In conclusion, there is no model that outperforms the other in every aspect.

The conducted analysis emphasises the challenges of predicting future mortality rates. It is likely that the declining mortality trend will continue. However, the persistence and the proportion of declination as well as the question which effects that will best explain the trend remains ambiguous.

References

[1] Alho, J. M. (2000). The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications, Ronald Lee, January 2000. North American Actuarial Journal, 4(1), 91-93.

[2] Andersson, G. (2013). Livförsäkringsmatematik. Svenska Försäkringsföreningen. Stockholm, Sweden.

[3] Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. International Journal of Forecasting, 22(3):547-581.

[4] Brockwell, P. J., Davis, R. A., & Calder, M. V. (2002). Introduction to time series and forecasting (Vol. 2). New York: Springer.

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

[6] Butt, Z., Haberman, & S., Shang, H. L. The ilc Package.

[7] CMI Bureau Mortality Sub-Committee. (2002). Working paper 1: An interim basis for adjusting the '92 series' mortality projections for cohort effects. CMI, London.

[8] Cairns, A. J. G., Blake, D., & Dowd, K. (2006). A two-factor model for stochas- tic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance 73(4), 687-718.

[9] Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North American Actuarial Journal, 13(1), 1-35.

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

[11] Currie, I. D. (2016). On fitting generalized linear and non-linear models of mortality. Scandinavian Actuarial Journal, 2016(4), 356-383.

[12] Directive, C. (2004). 113/EC implementing the principles of equal treatment between men and women in the access to and supply of goods and services. Gender Goods and Services Directive.

[13] Forskningsnämnden, F., & Försäkringsförbund, S. (2014). Försäkrade i Sverige - livslängder och dödlighet, prognoser 2014 - 2070, baserade på data 2001 - 2012. Stockholm, Sweden.

[14] Girosi, F., & King, G. (2007). Understanding the Lee-Carter mortality forecasting method. Gking. Harvard. Edu.

[15] Haberman, S., & Renshaw, A. E. (2009). On age-period-cohort parametric mortality rate projections. 45(2), 255-270.

[16] Haberman, S., & Renshaw, A. (2011). A comparative study of parametric mortality projection models. Insurance: Mathematics and Economics, 48(1), 35-55.

[17] Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 11 February 2018).

[18] Hunt, A., & Villegas, A. M. (2015). Robustness and convergence in the Lee-Carter model with cohort effects. Insurance: Mathematics and Economics, 64, 186-202.

[19] Hunt, A., & Blake, D. (2018). On the structure and classification of mortality models. North American Actuarial Journal.

[20] Koissi, M. C., Shapiro, A., & Högnäs, G. (2005). Fitting and Forecasting Mortality Rates for Nordic Countries Using the Lee-Carter method. Actuarial Research Clearing House, 1, 21.

[21] Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. Journal of the American statistical association, 87(419), 659-671.

[22] Lee, R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. North American actuarial journal, 4(1), 80-91.

[23] Lee, R., & Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. Demography, 38(4), 537-549.

[24] Plat, R. (2009). On stochastic mortality modeling. Insurance: Mathematics and Economics, 45(3), 393-404.

[25] Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insurance: Mathematics and economics, 38(3), 556-570.

[26] Renshaw, A. E., & Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling. 42(2), 797-816.

[27] Statistics Sweden. (2010). Cohort mortality in Sweden. Mortality statistics since 1861. Stockholm, Sweden.

[28] Stoffer, D. S., Shumway, R. H. (2006). Time series analysis and its applications: With R examples.

[29] Villegas, A., Kaishev, V. K., & Millossovich, P. (2015). StMoMo: An R package for stochastic mortality modelling.

[30] van Berkum, F., Antonio, K., & Vellekoop, M. (2013). Structural changes in mortality rates: with an application to Dutch and Belgian data. AFI Research Report, (AFI-1379).

[31] Willets, R. (1999). Mortality in the next millennium. Staple Inn Actuarial Society, 7.

[32] Willets, R. C. (2004). The cohort effect: insights and explanations. British Actuarial Journal, 10(4), 833-877.

[33] Wilmoth, J. R. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical report, Department of Demography, University of California, Berkeley.

Appendix A Figures

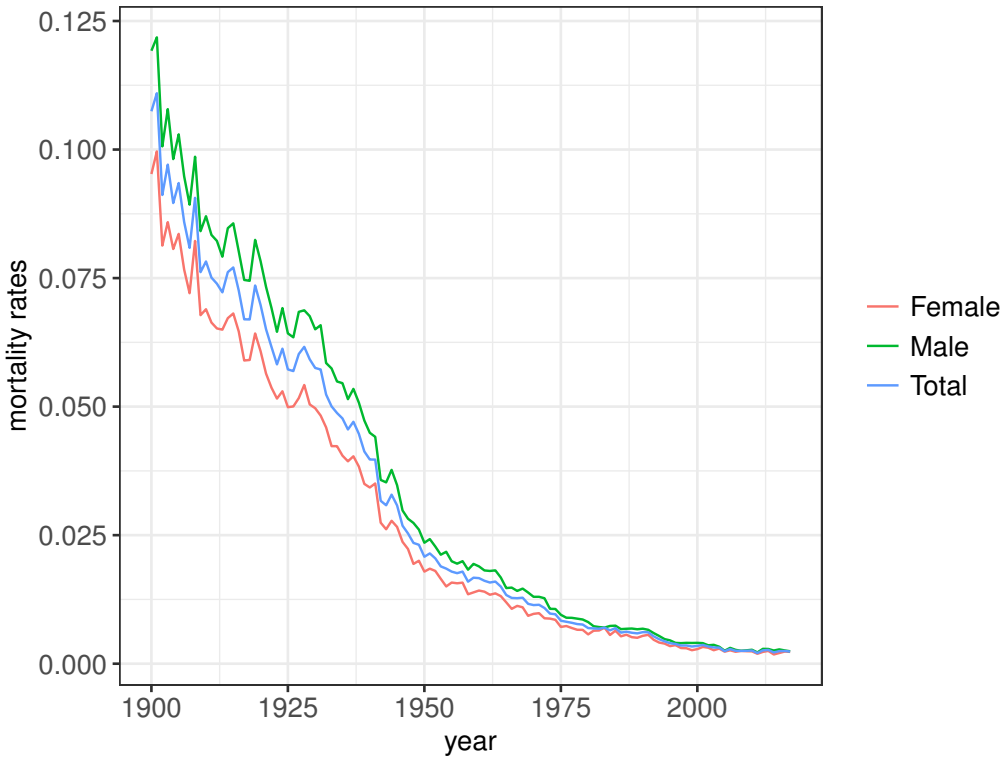


Figure 19: *Sweden: Development of infant mortality rates between 1900 and 2017.*

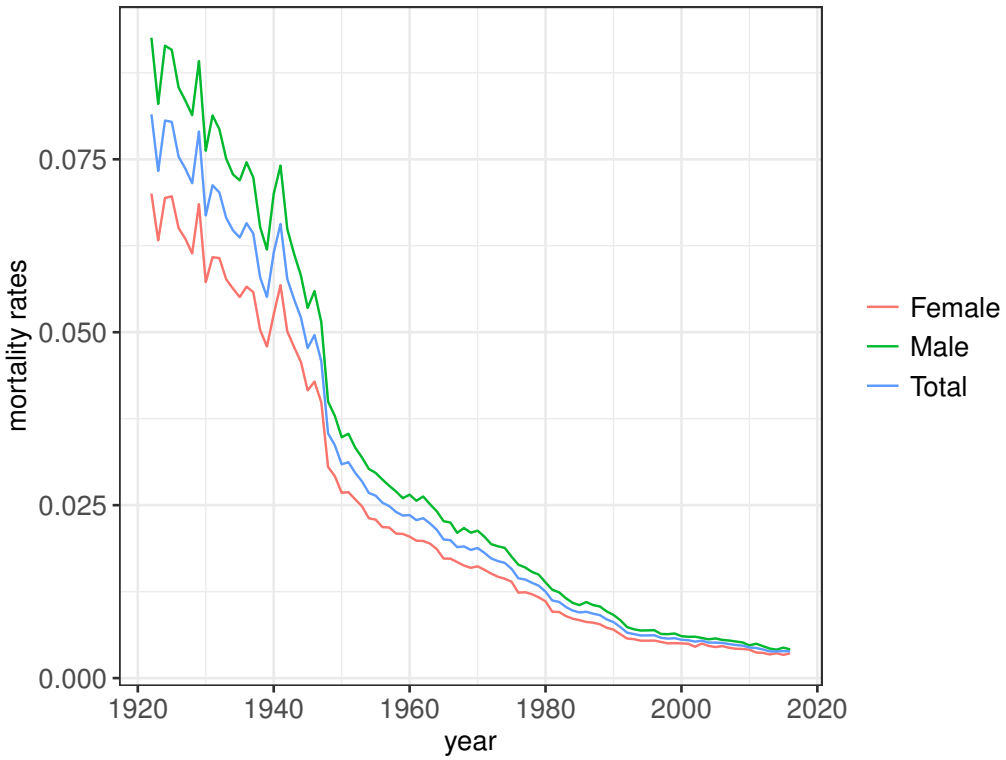


Figure 20: *UK: Development of infant mortality rates between 1922 and 2016.*

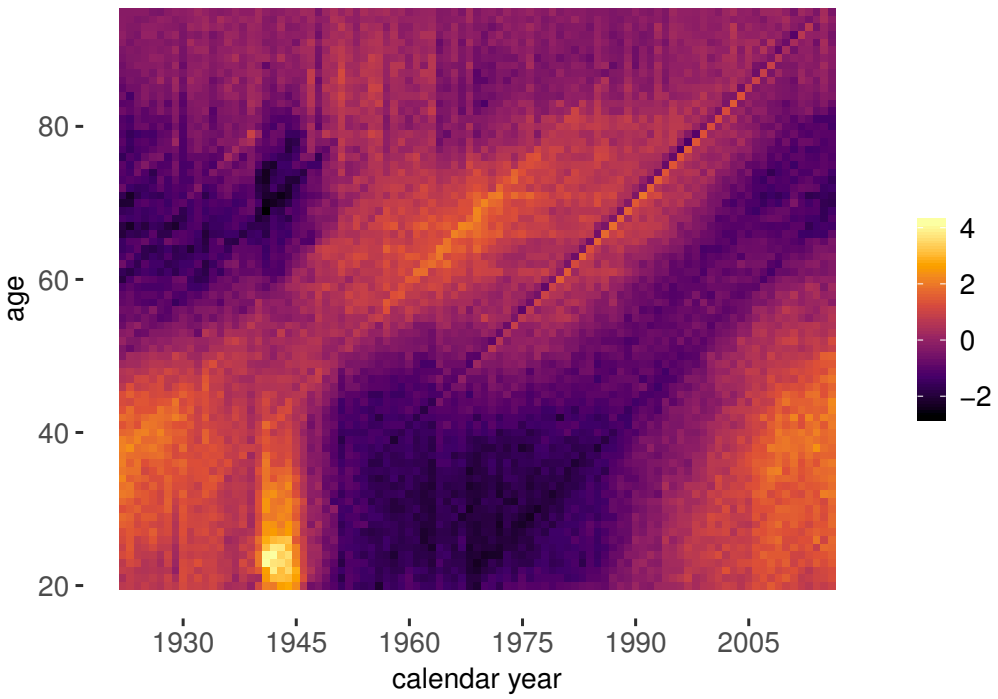


Figure 21: *UK: Standardised deviance residual plot depicting the LC model fit to male data between 1922 and 2016. Diagonal clusters of residuals signify cohort effects.*

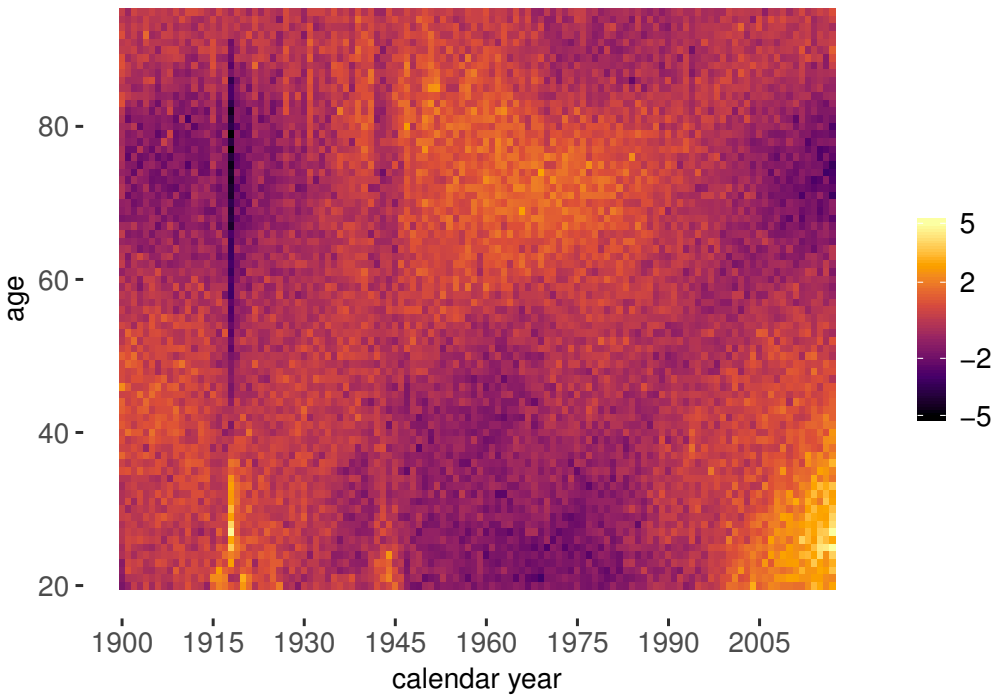
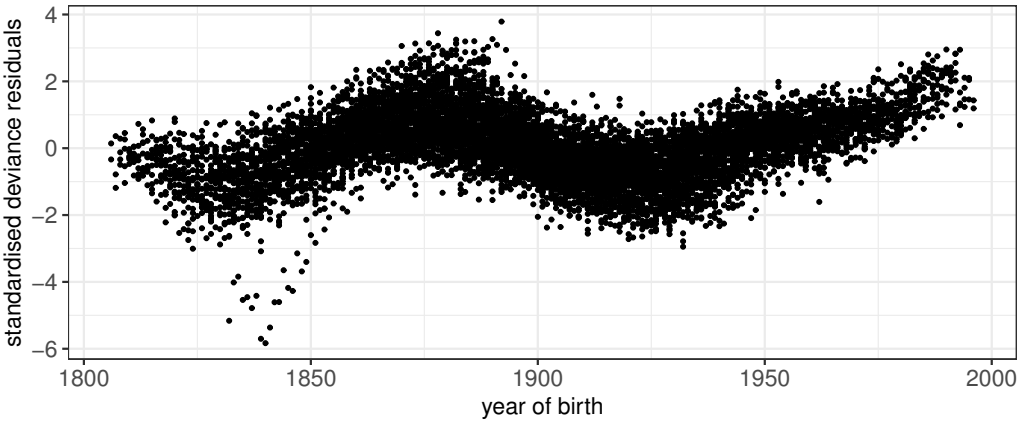
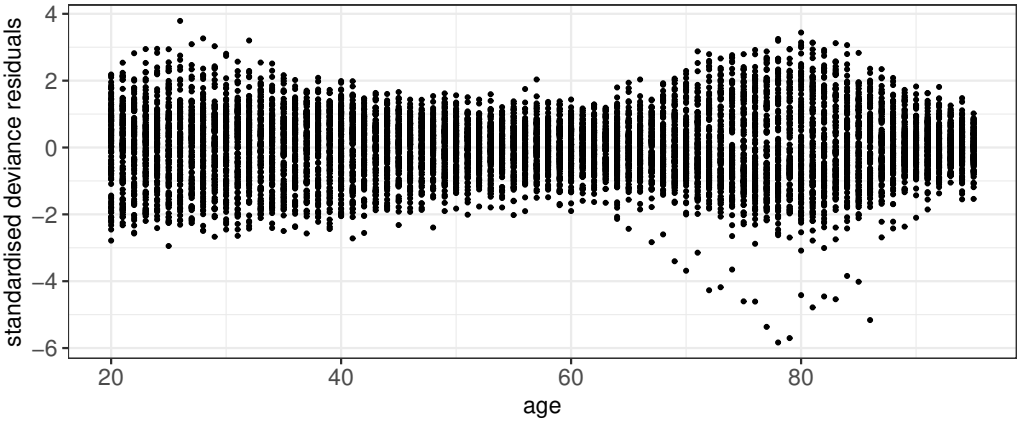


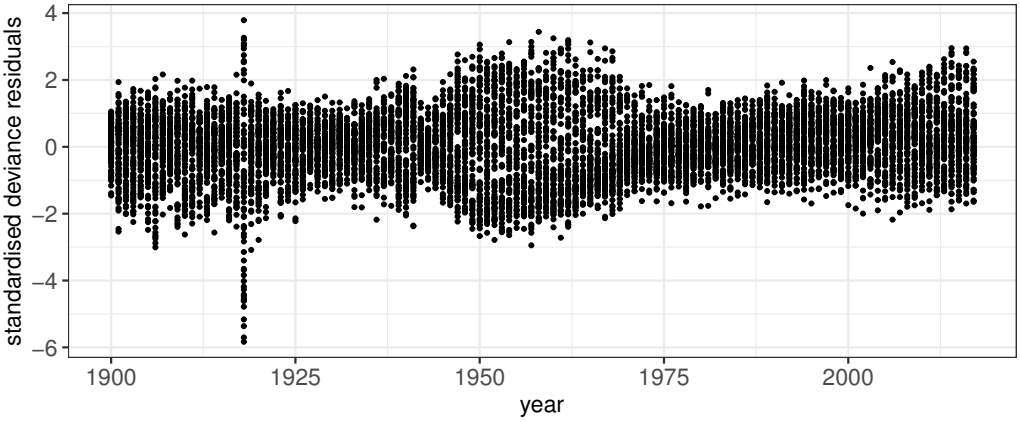
Figure 22: *Sweden: Standardised residual plot depicting the LC model fit to male data between 1900 and 2017. Diagonal clusters of residuals signify cohort effects.*



(a) Year of birth

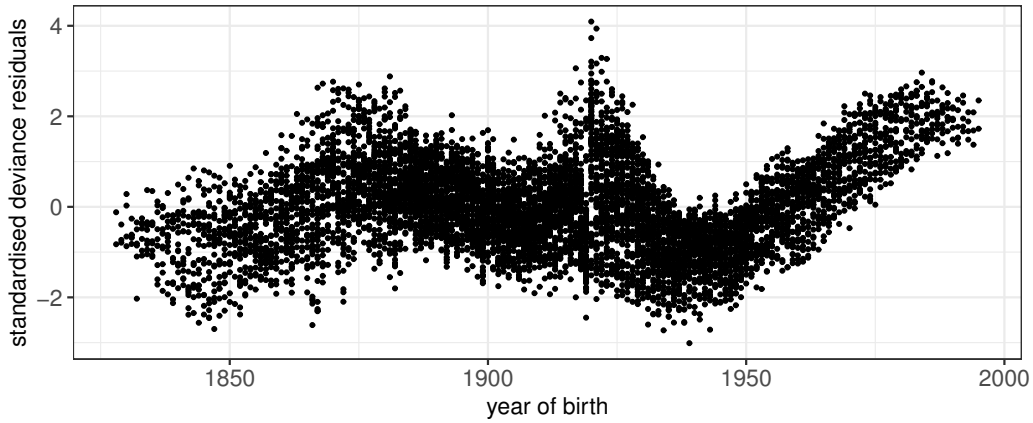


(b) Age

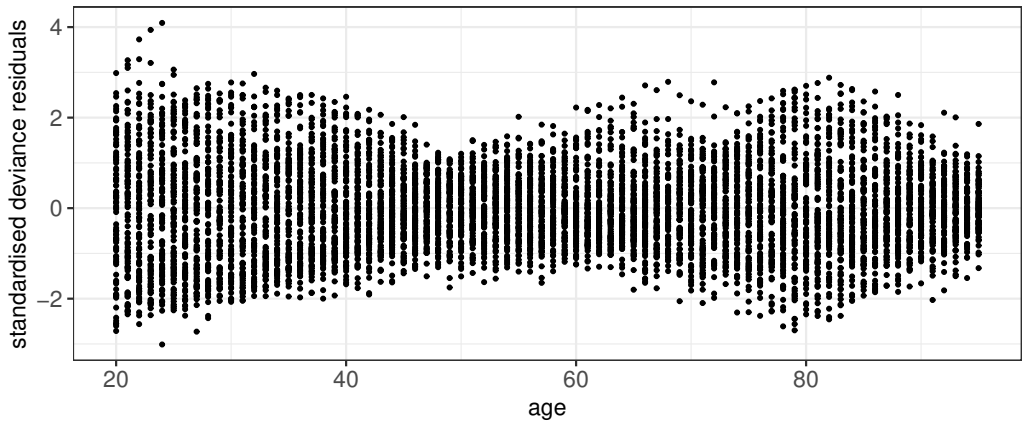


(c) Calendar year

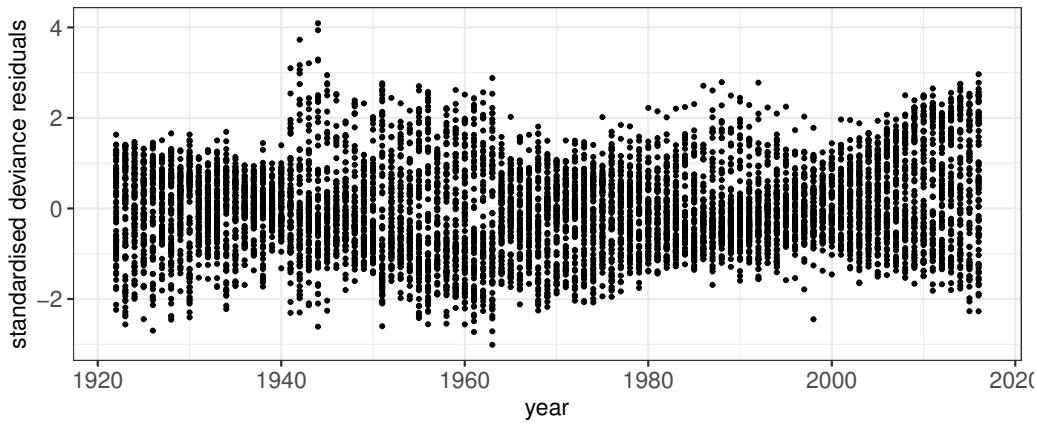
Figure 23: *Sweden: Standardised deviance residuals plotted against year of birth, age and calendar year of LC fit to female data between 1900 and 2017.*



(a) Year of birth

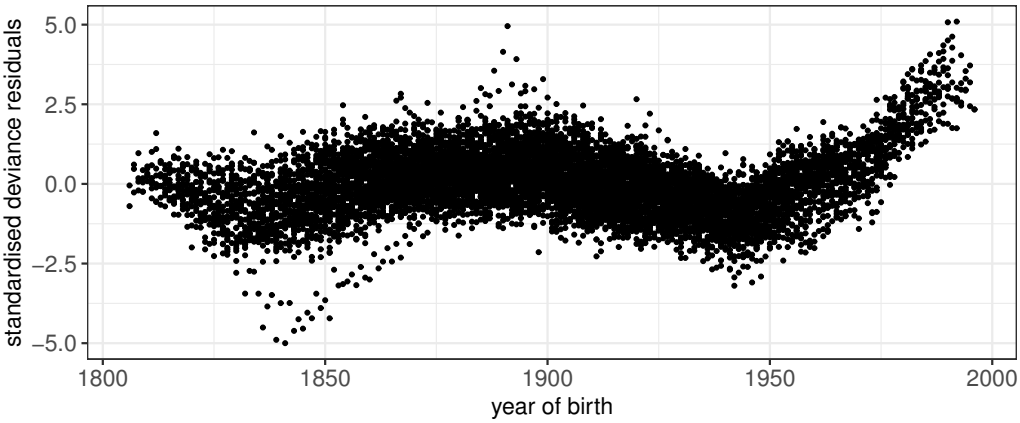


(b) Age

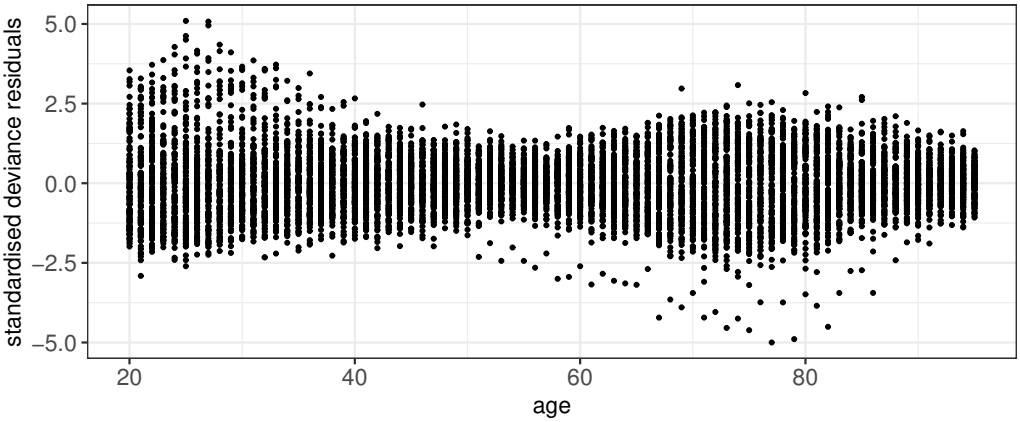


(c) Calendar year

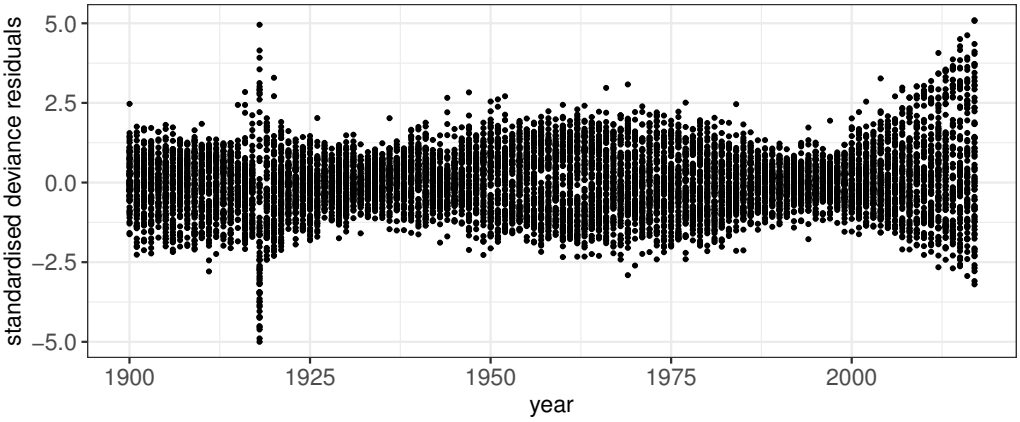
Figure 24: *UK: Standardised deviance residuals plotted against year of birth, age and calendar year of LC fit to female data between 1922 and 2016.*



(a) Year of birth

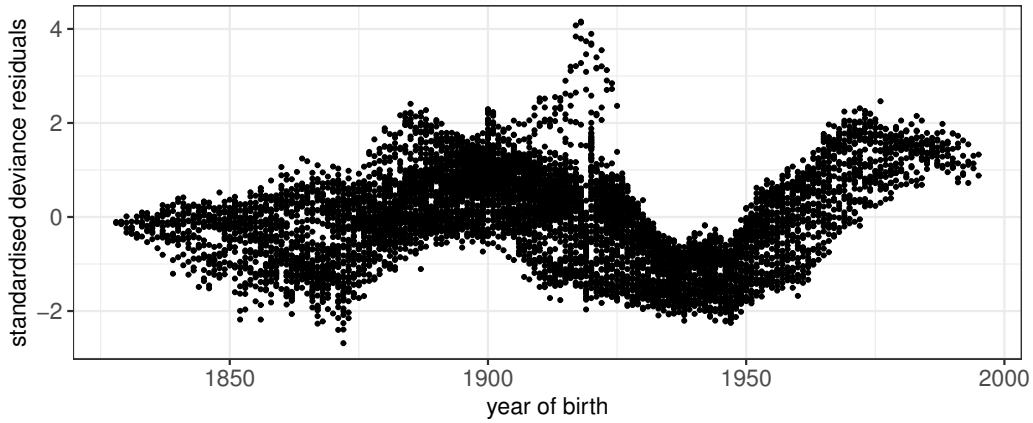


(b) Age

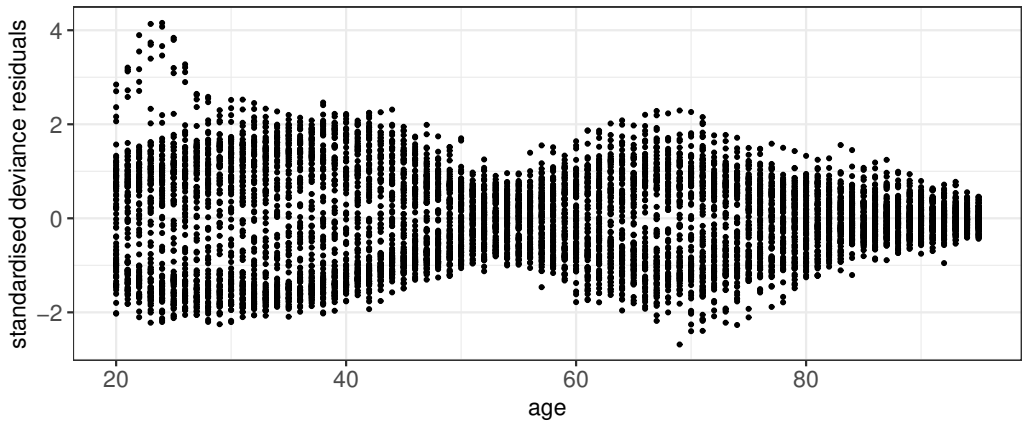


(c) Calendar year

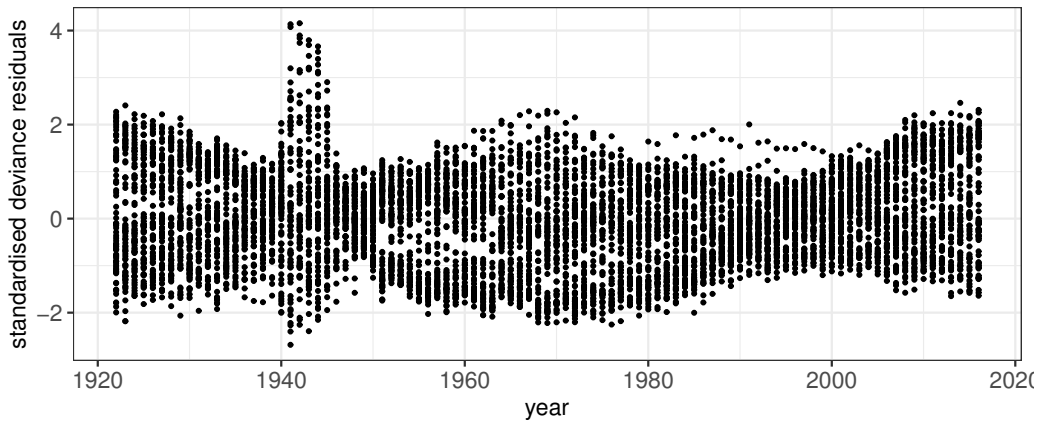
Figure 25: *Sweden: Standardised deviance residuals plotted against year of birth, age and calendar year of LC fit to male data between 1900 and 2017.*



(a) Year of birth

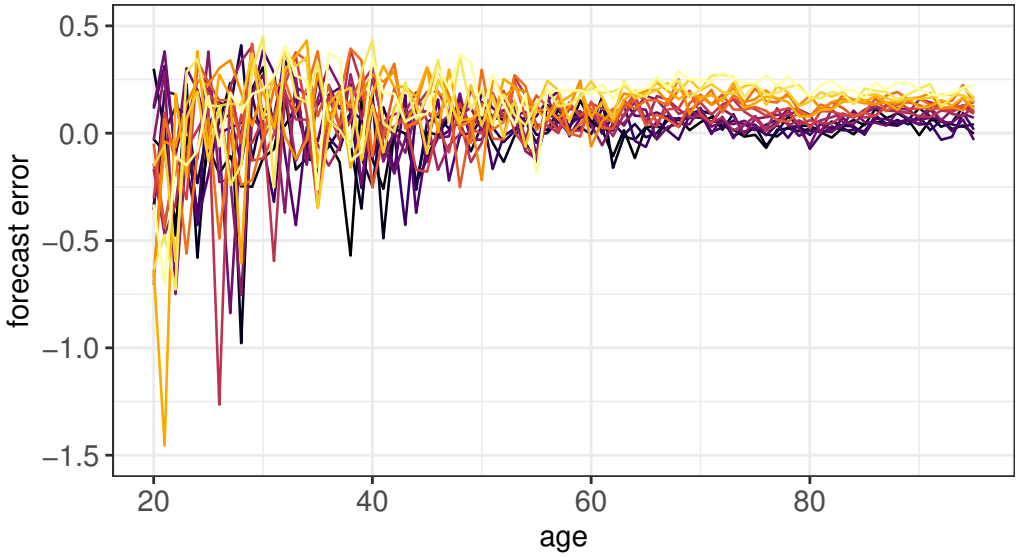


(b) Age

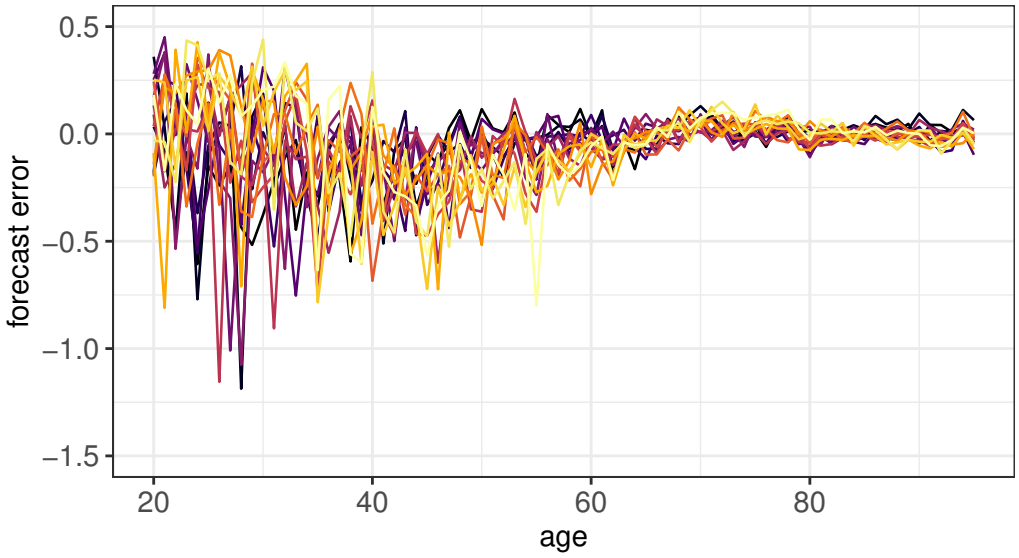


(c) Birth year

Figure 26: *UK: Standardised deviance residuals plotted against year of birth, age and calendar year of LC fit to male data between 1922 and 2016.*

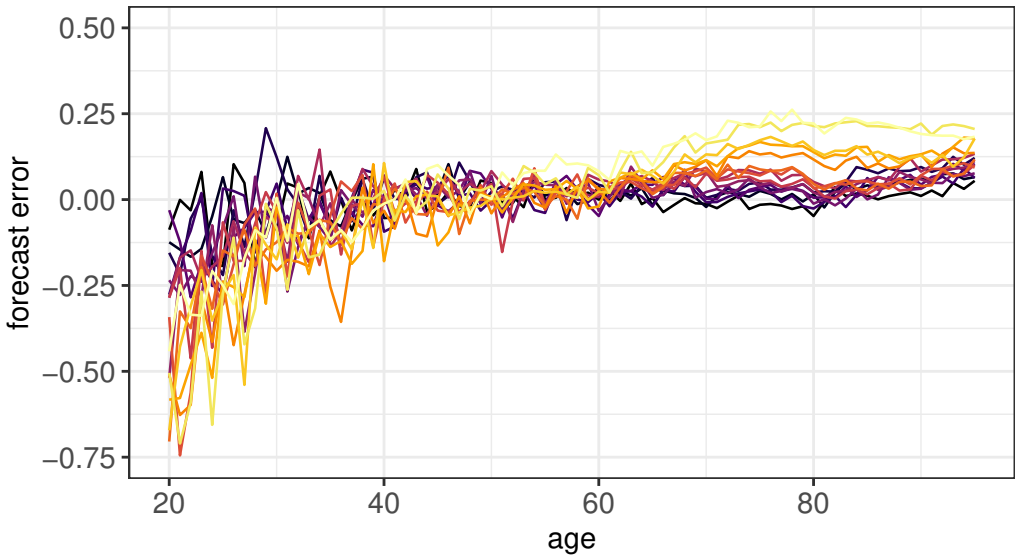


(a) LC Model

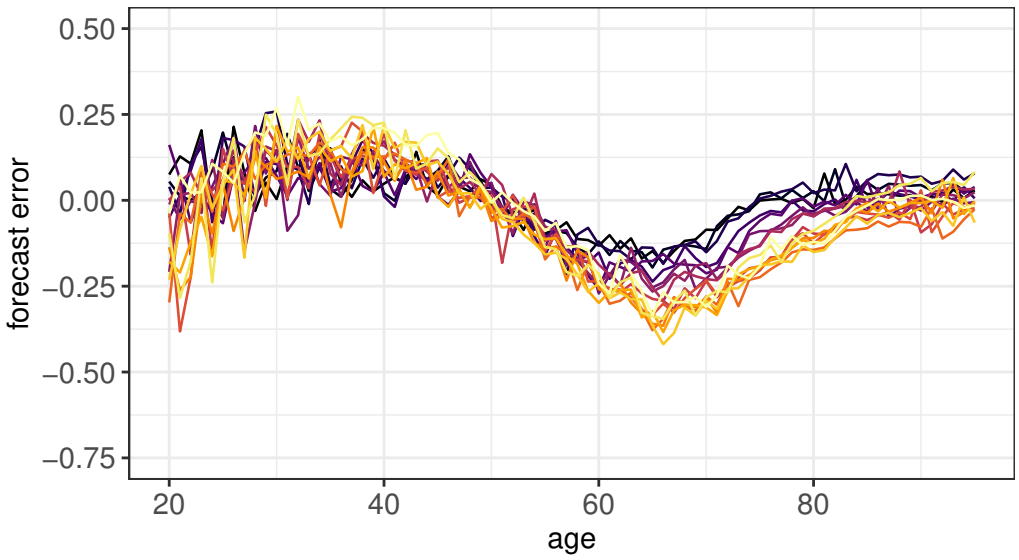


(b) RH model

Figure 27: *Sweden: Forecast error indexed by age for the LC and RH models. Each line represent one annual forecast between 2000 and 2017. As the colour fades, the forecasts are extending farther away in time.*



(a) LC Model



(b) RH model

Figure 28: *UK: Forecast error indexed by age for the LC and RH models. Each line represent one annual forecast between 2000 and 2017. As the colour fades, the forecasts are extending farther away in time.*

Appendix B Fitting Algorithms

B.0.1 Lee Carter model

The updating scheme for the LC fitting is determined by the following steps

1. Initial starting values conforming with the constraints in (11) are set to $\hat{\kappa}_t = 0$, $\hat{\beta}_x^{(1)} = \frac{1}{k}$ and $\hat{\alpha}_x = \frac{1}{n}\sum_t m_{x,t}$
2. Parameter $\hat{\alpha}_x$ is updated as

$$\hat{\alpha}_x^{\{v+1\}} = \hat{\alpha}_x^{\{v\}} + \frac{\sum_t 2\omega_{x,t}(d_{x,t} - \hat{d}_{x,t})}{\sum_t 2\omega_{x,t}\hat{d}_{x,t}}$$

   and the fitted values $\hat{d}_{x,t}$ is calculated along with the deviance $D(d_{x,t}, \hat{d}_{x,t})$.
3. Parameter $\hat{\kappa}_t$ is updated as

$$\hat{\kappa}_t^{\{v+1\}} = \hat{\kappa}_t^{\{v\}} + \frac{\sum_x 2\omega_{x,t}(d_{x,t} - \hat{d}_{x,t})}{2\omega_{x,t}\hat{d}_{x,t}(\hat{\beta}_x^{(1)})^2}$$

   and the fitted values $\hat{d}_{x,t}$ is calculated along with the deviance $D(d_{x,t}, \hat{d}_{x,t})$. After the updating of $\hat{\kappa}_t$ is adjusted to satisfy $\sum_t \kappa_t = 0$, which is imposed by centring $\hat{\kappa}_t = \hat{\kappa}_t - \overline{\hat{\kappa}_t}$.
4. Parameter $\hat{\beta}_x^{(1)}$ is updated as

$$\hat{\beta}_x^{(1)\{v+1\}} = \hat{\beta}_x^{(1)\{v+1\}} + \frac{\sum_t 2\omega_{x,t}(d_{x,t} - \hat{d}_{x,t})}{\sum_t 2\omega_{x,t}\hat{d}_{x,t}(\hat{\kappa}_x)^2}$$

   and the fitted values $\hat{d}_{x,t}$ is calculated along with the deviance $D(d_{x,t}, \hat{d}_{x,t})$.
5. The criterion to stop the iteration procedure is determined by the deviance decrease. If the deviance decrease is sufficiently small given a tolerance level set to 10^{-6}, the iteration procedure ends.
6. Once convergence is achieved the parameters are rescaled to satisfy the LC model constraints in (11) namely $\hat{\kappa}_t = \hat{\kappa}_t \cdot \sum_x \hat{\beta}_x^{(1)}$ and $\hat{\beta}_x^{(1)} = \frac{\hat{\beta}_x^{(1)}}{\sum_x \hat{\beta}_x^{(1)}}$.

B.0.2 Renshaw Haberman model

In order to circumvent convergence issues in the RH model we implement an alternative fitting procedure using an iterative Newton Raphson algorithm with an additional constraint. The alternative procedure was suggested to improve the fit to the data while resulting in faster convergence. We refrain from discussing the technical details of the derivation of the additional constraint. Profound information regarding the underlying study and mathematical derivation is made available in a recent paper [18]. The fitting procedure is however a replication of the previous section, but the parameters are instead updated using the log likelihood function.

The updating scheme for the RH fitting is determined by the following steps

1. Initial starting values of $\hat{\alpha}_x$, $\hat{\beta}_x^{(1)}$ and $\hat{\kappa}_t$ are set to the estimated values obtained from the LC fitting, while the $\hat{\gamma}_c$ parameters are set to $\hat{\gamma}_c = 0$.

2. Parameter $\hat{\alpha}_x$ is updated as

$$\hat{\alpha}_x^{\{v+1\}} = \hat{\alpha}_x^{\{v\}} + \frac{\sum_t \omega_{x,t}(d_{x,t} - \hat{d}_{x,t})}{\sum_t \omega_{x,t}\hat{d}_{x,t}}$$

   and the fitted values $\hat{d}_{x,t}$ is calculated.

3. Parameter $\hat{\kappa}_t$ is updated as

$$\hat{\kappa}_x^{\{v+1\}} = \hat{\kappa}_x^{\{v\}} + \frac{\sum_t \omega_{x,t}\hat{\beta}_x^{(1)}(d_{x,t} - \hat{d}_{x,t})}{\sum_t \omega_{x,t}\hat{d}_{x,t}(\hat{\beta}_x^{(1)})^2}$$

   and the fitted values $\hat{d}_{x,t}$ is calculated. After the updating of $\hat{\kappa}_t$, the estimates are adjusted to satisfy $\sum_t \kappa_t = 0$, which is imposed by centring $\hat{\kappa}_t = \hat{\kappa}_t - \overline{\hat{\kappa}_t}$ and setting $\hat{\alpha}_x = \hat{\alpha}_x + \overline{\hat{\kappa}_t}\hat{\beta}_x^{(1)}$.

4. Parameter $\hat{\beta}_x^{(1)}$ is updated as

$$\hat{\beta}_x^{(1)\{v+1\}} = \hat{\beta}_x^{(1)\{v\}} + \frac{\sum_t \omega_{x,t}(d_{x,t} - \hat{d}_{x,t})}{\sum_t \omega_{x,t}\hat{d}_{x,t}(\hat{\kappa}_x)^2}$$

   and the fitted values $\hat{d}_{x,t}$ is calculated. After the updating of $\hat{\beta}_x^{(1)}$, the estimates are adjusted to ensure $\sum_x \beta_x^{(1)} = 1$ by imposing $\hat{\kappa}_t =$

$\hat{\kappa}_t \cdot \sum_x \hat{\beta}_x^{(1)}$ and $\hat{\beta}_x^{(1)} = \frac{\hat{\beta}_x^{(1)}}{\sum_x \hat{\beta}_x^{(1)}}$.

5. Parameter $\hat{\gamma}_c$ is updated as

$$\hat{\gamma}_c^{\{v+1\}} = \hat{\gamma}_c^{\{v\}} + \frac{\sum_{c=t-x}^{x,t} \omega_{x,t}(d_{x,t} - \hat{d}_{x,t})}{\sum_{c=t-x}^{x,t} \omega_{x,t}\hat{d}_{x,t}}$$

   and the fitted values $\hat{d}_{x,t}$ is calculated. After the updating of $\hat{\gamma}_c$ is adjusted to satisfy $\sum_t \gamma_c = 0$, which is imposed by centring $\hat{\gamma}_c = \hat{\gamma}_c - \overline{\hat{\gamma}_c}$ and setting $\hat{\alpha}_x = \hat{\alpha}_x + \overline{\hat{\gamma}_c}$.

6. The estimates are adjusted through a final transformation to ensure that $\sum_{c=t_1-x_k}^{t_n-x_1}(c - \overline{c})\hat{\gamma}_c = 0$. As previously mentioned, this step is evaluated and described in detail in a recent paper whereas we refrain from including the details here [18]. Thereafter the fitted values $\hat{d}_{x,t}$ is calculated.

7. The criterion to stop the iteration procedure is determined by the increase in the log likelihood function. If the increase of the log likelihood function is sufficiently small given a tolerance level set to 10^{-6}, the iteration procedure ends and the parameter estimates are returned.