Alternative parameterisations of the Poisson common factor model for modelling mortality jointly for both sexes

Kenneth Wong

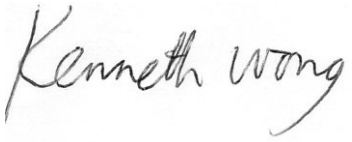
A thesis submitted for the degree of Master of Research

Department of Applied Finance and Actuarial Studies
Faculty of Business and Economics
Macquarie University

October 2016

Declaration

I certify that the work contained in this thesis has not been submitted for a higher degree to any other university or institution. All sources of information and literature utilised within have been appropriately disclosed and acknowledged.



Kenneth Wong

Summary

In a global environment where mortality continues to decline, it is becoming increasingly important to develop mortality models which are able to account for global trends and relationships while also producing reasonable forecasts. In recent years there has been a growing interest in the co-modelling of multiple populations to address this. One such example is the Poisson common factor model proposed by Li (2013) for modelling mortality jointly for both sexes.

This thesis expands on the Poisson common factor model by proposing two alternative parameterisations which relax some of the original assumptions. One variation allows a different number of sex-specific factors for each sex, providing more flexibility in taking into account differing features and trends between females and males. The other variation considers a common age effect shared by both sexes, potentially improving the parsimony of the model's optimal use of parameters.

The two extended models are then tested using mortality data from six populations. Model performance is measured using goodness-of-fit and forecasting accuracy. The results indicate that both of the two modifications improve fitting compared to the original model, and slightly improve forecasting accuracy in many cases.

Acknowledgements

I would like to thank my supervisor, Jackie Li, for his invaluable help and support. This thesis would not have been possible without his advice pointing me in the right direction. I am also grateful to Kenny Mok and Colin Zhang for their suggestions for improvement. I thank my family and friends for their support and providing me with motivation.

I also wish to thank Agnieszka Baginska, Lin Bai, and Angela Chow for their help and support in administrative matters throughout the year. Finally, I would like to thank the Department of Applied Finance and Actuarial Studies for providing me with this opportunity.

Table of Contents

Chapter 1

Introduction

In a world that is becoming increasingly connected, changes in population demographics and trends have the potential to cause a large impact on future economic and environmental planning. In particular, the continual improvement in life expectancy over the last few decades presents a significant challenge for government pension schemes, superannuation funds and other retirement income providers. With the additional factor of a rapidly developing longevity risk market, it is of utmost importance to develop suitable mortality models for measuring the trends and also to examine their projection results for potential use in future planning exercises.

There have been many developments in the field of mortality modelling and projection in the last two decades. Ever since Lee and Carter (1992) laid the foundation for stochastic mortality modelling, much discussion has surrounded model selection, modelling procedures and improving mortality forecasts. The Lee-Carter model is elegant and straightforward, but it also suffers from various limitations. Many extensions, modifications and alternatives to the Lee-Carter model have been proposed and tested in the literature – see Cairns et al. (2008) and Cairns et al. (2009) for a comprehensive review and comparison of the prevalent mortality models.

A research area that has recently gained much attention is the co-modelling of multiple populations. In general, examples of related populations include geographically and socioeconomically close countries, female and male populations

of the same region, and regional compared to national mortality of the same country. It can be argued that these subdivided groups or neighbouring populations are linked by certain common driving forces, and mortality models should be developed to capture the underlying relationships properly. One of the main criteria of a good multi-population mortality model is biological consistency. For instance, it is a common feature that females have a higher life expectancy than that of males. A model that forecasts the opposite situation would be difficult to justify due to inconsistency with historical data.

Early implementations of the Lee-Carter model have treated related populations separately (or simply in aggregate). However, studies have found that this practice can lead to future projections that are inconsistent with historical data. Carter and Lee (1992) and Li and Lee (2005) noted that treating the two sexes as independent when applying the Lee-Carter model could result in divergent mortality projections, and possibly a mortality crossover – Li (2013) also expressed similar concerns when modelling Australian mortality. Without considering related populations jointly, mortality forecasts in the long run could become nonsensical and unusable. This thesis focuses on the female-male mortality relationship and explores some new ways to improve modelling mortality of both sexes together.

In order to produce coherent or non-divergent mortality forecasts, Li and Lee (2005) proposed an extension of the Lee-Carter method, known as the augmented common factor model. This multi-population model allows for a common population-wide factor for the main long-term trend as well as an additional sex-specific factor for short-term deviations of each sex from the main trend. The specific time series modelling of the additional factors ensures convergence in projected male-female death rate ratios at each age in the long term. This helps avoid undesirable effects such as a mortality crossover or continual divergence between the sexes. Li (2013) presented an improvement on the augmented common factor model by modifying the homoscedastic error term assumption, incorporating a Poisson model to cater for the total number of deaths rather than the death rate, and generalising the model to incorporate multiple sex-specific factors. This Poisson common factor model (PCFM) maintains a male-female death rate ratio convergence at each age, serves as a formal

model framework for statistical analysis, and provides more flexibility in capturing higher order effects in the data.

Nevertheless, there is still room for improvement in the PCFM. Yang et al. (2016) extended the PCFM by incorporating the cohort effect in a number of variations, and found that the new structures improve model fitting, reduce the optimal number of additional factors and maintain coherent mortality forecasts. This thesis, on the other hand, seeks to modify the PCFM in a different fashion, as very briefly noted in Li et al. (2016). One suggestion arises from the fact that the original PCFM does not allow for a different number of male- and female-specific factors. In principle, relaxing this limitation should allow the model to capture more different features or trends between the sexes, resulting in a more flexible model that is applicable to more situations. Another alternative is to impose a common age-to-period sensitivity effect on the sex-specific factors in the PCFM. This has the potential to improve the parsimony of the model, in terms of reducing the number of parameters required. These two suggestions work in an opposite way to some extent – one allows for differences in the period effects, whereas the other exploits similar age sensitivity patterns (if any) between females and males. But the main purpose is the same – to develop more ways to adapt the PCFM to data from different populations with diverse features.

In summary, this thesis aims to test the two extensions of the PCFM and determine if the alternatives show a performance improvement over the baseline model. Given the need to focus on coherent forecasting, this improvement is measured by investigating the accuracy of the model fit, as well as forecast accuracy and checking long-term forecast trends.

The remainder of this thesis is structured as follows. Chapter 2 presents a literature review of mortality modelling and projection. Chapter 3 reviews the PCFM and provides details of the two proposed model variations. Chapter 4 applies the models to datasets of six populations and analyses the fitting results. Chapter 5 compares the models in terms of forecasting performance, and also performs out-of-sample testing. Finally, Chapter 6 sets forth concluding remarks and comments on future research.

Chapter 2

Literature Review

This chapter provides a brief review of existing mortality models in the literature and the precursors to the PCFM. It also highlights the issues facing mortality forecasting and the need for joint mortality models.

2.1. Outline

Mortality models used for forecasting can be broadly classified into three categories: expectation[1], explanation and extrapolation (Booth and Tickle, 2008). Of these three categories, extrapolative forecasting methods are the most widely used. Expectation models are based on subjective opinions from experts to predict future mortality trends. Explanation methods rely on structural or epidemiological models to connect mortality rates to causes of death. Extrapolative methods make use of age patterns and trends over time found in mortality data, under the assumption that past mortality trends will continue in the future.

Of the three approaches, extrapolative models are the most common and widely used. Expectation methods suffer from the drawback of requiring subjective input. This tends to result in overly pessimistic estimates of mortality decline (Alho and Spencer, 1990; Lee and Carter, 1992; Lee and Miller, 2001). Explanatory models are difficult to use in practice, because the connections between mortality and risk factors are

[1]The expectation models are also known as judgmental models.

constantly changing and hard to quantify. Moreover, it is also required to produce forecasts of the risk factors themselves, which may be just as difficult as forecasting mortality. Indeed, forecasts with structural models have generally not performed very accurately, especially in the long term (Booth, 2006; Booth and Tickle, 2008; Keyfitz, 1982). In comparison, extrapolative models are relatively straightforward to implement, possessing no reliance on theories or hypotheses except that the future will reflect past patterns. This is a key strength but also a fundamental weakness. Nevertheless, extrapolative approaches have proven to be widely popular in the literature (Booth, 2006). The most prominent extrapolative approach is the Lee-Carter model; the remainder of this review focuses on this model and the developments that lead to the PCFM.

2.2. The Lee-Carter model and extensions

Lee and Carter (1992) introduced an age-period-specific model to forecast mortality. The log central death rate is expressed as a function of two age factors and one time factor:

$$\ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}$$

where a_x represents the base age effect, k_t describes the overall change in mortality over time, and b_x is an age-specific sensitivity measure. $\varepsilon_{x,t}$ is a homoscedastic error term. Parameters are estimated via singular value decomposition (SVD). The Lee-Carter model can be used for forecasting by modelling the time component k_t as a time series such as a random walk with drift.

The main advantages of the Lee-Carter model lie in its simplicity and ability to produce mortality forecasts without relying on subjective inputs. Due to this, it has been a popular choice for academics and practitioners for many years. However, the Lee-Carter model is not without drawbacks. For example, the assumption of a constant b_x over time is unrealistic when it comes to forecasting mortality. The model assumes that the rates of mortality decline across different ages always maintain the same ratios to one another over time, but in practice this is not the case (Lee, 2000). Another issue is that the estimates of b_x tend to be jagged across different ages,

leading to uneven mortality forecasts which should be expected to be smooth (De Jong and Tickle, 2006). A final example is that with only one factor, the Lee-Carter model cannot incorporate cohort effects – residual analysis for certain populations show clear evidence of clustering, violating the assumption of independence (Cairns et al., 2008).

Many extensions to the Lee-Carter model have been proposed to improve its shortcomings. Lee and Miller (2001) noted that forecasting performance can be improved by focusing on goodness-of-fit in the final, jump-off year as opposed to the entire dataset. Brouhns et al. (2002) demonstrated a Poisson regression approach to estimating parameters, implementing more formal statistical methods such as maximum likelihood estimation and making the model more intuitively acceptable. Enhanced models with higher-order terms and cohort-specific terms have also been considered (Booth et al., 2002; Renshaw and Haberman, 2003, 2006). Others have formulated approaches to tackle the Lee-Carter model's smoothing issues (De Jong and Tickle, 2006; Delwarde et al., 2007). Indeed, many of the base Lee-Carter model's flaws have been examined and improved over the years. However, many of these models have remained focused on applying the model to a single population.

2.3. Mortality trends and joint modelling

In the context of globalisation, countries are becoming more closely linked in terms of lifestyle, technology and other socio-economic factors. As such, it appears reasonable to assume that similarities in mortality patterns will also begin to emerge among closely related populations. In order to collate information and patterns across groups of populations, various joint mortality models have been proposed. These joint models attempt to improve model fit and produce more accurate, coherent forecasts compared to implementing separate individual models.

Potential issues regarding separate modelling of individual populations have been identified as early as Carter and Lee (1992), where it was demonstrated that forecasting US mortality for the two sexes separately results in a long-term divergence of mortality rates. Such a large difference between male and female death

rates is illogical from a biological perspective. Moreover, the forecasts also indicated a mortality crossover – where at certain ages, females were projected to have higher mortality than males. This conclusion would be inconsistent with the historical sex differential. Similar inconsistency issues also exist for other scenarios – for example, comparing countries of vastly different trends, or modelling mortality of an insurance portfolio against nationwide mortality to account for longevity basis risk. While Carter and Lee (1992) suggested some approaches to model the two sexes, an explicit joint model extension of the Lee-Carter model was not proposed until Li and Lee (2005).

2.4. The augmented common factor model

Li and Lee (2005) noted that while the Lee-Carter model works well for a single population (either one sex or both sexes combined), dealing with each sex separately would result in the divergence problem described in Carter and Lee (1992).
To tackle this issue, Li and Lee (2005) proposed the augmented common factor model. It is a multi-population extension of the Lee-Carter model:

$$\ln m_{x,t,i} = a_{x,i} + B_x K_t + b_{x,i} k_{t,i} + \varepsilon_{x,t,i}$$

where $B_x K_t$ represents the common factor, and $a_{x,i}$, $b_{x,i}$, $k_{t,i}$ and $\varepsilon_{x,t,i}$ hold similar meanings to the Lee-Carter model for specific population i. The common factor describes the main long-term trend in mortality change for the combined population as a whole, while the additional population-specific factors represent short-term deviations from the main trend. For forecasting, the common factor K_t is modelled as a random walk with drift, and the additional factors $k_{t,i}$ are assumed to be stationary AR(1) processes. These assumptions cause the projected ratio of death rates between two populations to tend to a constant in the long run, thus allowing for short-term discrepancies but avoiding mortality divergence.

While the augmented common factor model manages to deal with the issue of divergence, it inherits some of the shortcomings of the original Lee-Carter model. Also, incorporating multiple time components can result in increased uncertainty in short-term forecasts. Some extensions to the model have been proposed – one

example is the product-ratio method (Hyndman et al., 2013), which can be viewed as a generalisation of the augmented common factor model. The product-ratio method allows for multiple specific factors and incorporates more dynamic time series processes. Using this model, Hyndman et al. (2013) focused on a more precise definition of "coherent" forecasting, as opposed to simple non-divergence, and demonstrated a marked improvement in forecasting accuracy compared to other methods. A second extended model is the PCFM (Li, 2013), which applied Poisson regression as in Brouhns et al. (2002) to the augmented common factor model, resulting in similar benefits. A more detailed review of the PCFM follows in the next chapter.

2.5. Other models

There are a number of other multi-population mortality models that have been proposed in the literature. Li et al. (2014) provided a comprehensive review of these models. Many of these models are designed for modelling a large population with a much smaller sub-population in insurance hedging applications. In contrast, this thesis focuses on co-modelling females and males within a population, with potential use in government policy planning and insurance pricing. As such, the other models are only briefly mentioned here.

The Cairns-Blake-Dowd (CBD) model (Cairns et al., 2006) is a single-population model that focuses on how old-age mortality changes over time. While it may fit older ages better in certain cases compared to the Lee-Carter model, it comes at the cost of a poorer fit if it is applied to the whole age range. Nevertheless, the CBD model can be considered another starting branch for many extensions and modified models. Joint extensions of the CBD model also exist – one such example is detailed in Tan et al. (2014).

Some other models reviewed in Li et al. (2014) are based on Lee-Carter but have been developed separately to the augmented common factor model. For example, Russolillo et al. (2011) incorporated a population effect as a third dimension in their joint Lee-Carter extension. Debón et al. (2011) also proposed a joint Lee-Carter

extension with an extra factor. These can be loosely described as a group of models that incorporate common and specific factors (Li et al., 2014).

2.6. Summary

There have been many developments in mortality modelling and forecasting in the past few decades. Most of the mortality models in the literature can be said to belong to either the Lee-Carter family or CBD family of models. However, the motivations of this thesis lead to a focus on the Lee-Carter branch – specifically, extensions of the PCFM for co-modelling male and female populations. The potential usefulness of the PCFM has already been demonstrated – examples of applications to demographic and insurance problems can be found in Li and Haberman (2015), Li et al. (2016), Parr et al. (2016), and Yang et al. (2016). It is hoped that the proposed extensions to the PCFM result in a model that is more suitable to a wide variety of applications.

Chapter 3

Data and Methods

3.1. Review of the Poisson common factor model

In the PCFM (Li, 2013), the force of mortality $\mu_{x,t,i}$ at age x in year t for sex i is assumed to be constant over an integer age-period interval. As a result, the central death rate $m_{x,t,i} = \mu_{x,t,i}$ and the number of deaths can be modelled directly as a Poisson random variable:

$$D_{x,t,i} \sim \text{Pn}(E_{x,t,i} m_{x,t,i})$$

where $D_{x,t,i}$ is the number of deaths and $E_{x,t,i}$ is the corresponding exposure to risk. While $E_{x,t,i}$ is a known quantity, $m_{x,t,i}$ is an unknown parameter that requires estimation.

The Poisson assumption has several advantages. As argued in Li (2013), this assumption leads to a rigorous statistical framework for analysing mortality data. Also, treating the number of deaths $D_{x,t,i}$ as a counting random variable is a more natural choice compared to modelling the death rate with a homoscedastic error term in earlier models such as Lee and Carter (1992) and Li and Lee (2005). This Poisson framework is widely used in the literature – see Brouhns et al. (2002) and Cairns et al. (2009) for previous applications. When assessing uncertainty in mortality changes, however, death counts in population data appear to be over-dispersed for many countries, with a higher variance than mean (Cairns et al., 2009). In such cases, the

Poisson assumption can readily be modified as over-dispersed Poisson, in which the mean $\mathrm{E}(D_{x,t,i}) = E_{x,t,i} m_{x,t,i}$ remains the same while the variance is defined as $\mathrm{Var}(D_{x,t,i}) = \phi E_{x,t,i} m_{x,t,i}$ instead, with $\phi > 1$ as the dispersion parameter (Renshaw and Haberman, 2006). There is no change needed in the computation algorithm and the parameter estimates would stay the same, except that the extra dispersion parameter has to be calculated separately from the deviance function.

In line with the augmented common factor model in Li and Lee (2005), the log central death rate is modelled as:

$$\ln m_{x,t,i} = a_{x,i} + B_x K_t + \sum_{j=1}^{n} b_{x,i,j} k_{t,i,j}$$

where $a_{x,i}$ represents the overall age effect, $B_x K_t$ is the common factor for both sexes, and $b_{x,i,j} k_{t,i,j}$ is the *j*th additional sex-specific factor for sex *i*. In more detail, K_t is the mortality index of the common factor, and B_x measures the sensitivity of the log central death rate to changes in K_t for each age category. Similarly, $k_{t,i,j}$ is the time component of the *j*th sex-specific factor for sex *i*, with corresponding age sensitivity measure $b_{x,i,j}$. Compared to Li and Lee (2005), the PCFM allows for the incorporation of multiple sex-specific factors where necessary, resulting in improved modelling results (Li, 2013; Li et al., 2016).

The (conditional) maximum-likelihood parameter estimates of the PCFM are calculated via an iterative updating scheme (see Appendix for details). In order to ensure model identification, the model is subject to (4*n* + 2) constraints $\sum_x B_x = 1$, $\sum_t K_t = 0$, $\sum_x b_{x,i,j} = 1$ and $\sum_t k_{t,i,j} = 0$. To determine the optimal number of additional factors, the Bayesian Information Criterion (BIC)[2] is used as the main statistical measure to balance between model fit and over-parameterisation. Other indicators include the patterns of the residual plots, the trends of the additional parameters and the volume of data under investigation.

[2] The BIC is calculated as $-2\hat{l} + n_p \ln(n_d)$, where $\hat{l}$ is the computed log-likelihood, n_p is the effective number of parameters being estimated, and n_d is the number of observations.

After calibration of the model, the time components K_t and $k_{t,i,j}$ need to be projected into the future. Previous studies have shown that the common mortality index K_t tends to be linear and decreasing for various countries (Li et al., 2016). Hence, K_t can be modelled as a random walk with drift:

$$K_t = \mu + K_{t-1} + e_t$$

where μ is the drift term and e_t is a normally distributed random variable with mean 0 and variance σ^2. On the other hand, the sex-specific terms $k_{t,i,j}$ are intended to represent short-term deviations from the main trend for each sex, so a mean-reverting process is an ideal model. Hence, each $k_{t,i,j}$ is assumed to follow a weakly stationary AR(p) process:

$$k_{t,i,j} = \alpha_{0,i,j} + \sum_{l=1}^{p} \alpha_{l,i,j} k_{t-1,i,j} + \varepsilon_{t,i,j}$$

where $\alpha_{0,i,j}$ and $\alpha_{l,i,j}$ are autoregressive model parameters and $\varepsilon_{t,i,j}$ is a normal error term with mean 0 and variance $\omega_{i,j}^2$. The order p is chosen based on the partial autocorrelation function (PACF) of the time components and the autocorrelations of the residuals. Additionally, each $k_{t,i,j}$ is assumed to be independent to the others.[3] Under these conditions, future death rates (in year $t > T$) can be projected as:

$$\hat{m}_{x,t,i} = \dot{m}_{x,T,i} \exp\left(B_x(\hat{K}_t - K_T) + \sum_{j=1}^{n} b_{x,i,j}(\hat{k}_{t,i,j} - k_{T,i,j}) \right)$$

where $\hat{K}_t$ and $\hat{k}_{t,i,j}$ are projected values from their respective time series setting the error terms to zero, and the starting point of the projection $\dot{m}_{x,T,i}$ is calculated from the latest set of data in year T. This helps to avoid significant bias in the beginning of the projection period (Lee and Miller, 2001). Using these projections, the ratio of male-to-female death rates can be expressed as:

$$\frac{\hat{m}_{x,t,2}}{\hat{m}_{x,t,1}} = \frac{\dot{m}_{x,t,2}}{\dot{m}_{x,t,1}} \exp\left(\sum_{j=1}^{n} (b_{x,2,j}(\hat{k}_{t,2,j} - k_{T,2,j}) - b_{x,1,j}(\hat{k}_{t,1,j} - k_{T,1,j})) \right).$$

[3] An alternative is to model the $k_{t,i,j}$ terms as multivariate time series when assessing uncertainty of mortality changes, but this approach may complicate the projection exercise with a small data period. This is left as an option for future research.

This ratio only converges to a constant if $\hat{k}_{t,1,j}$ and $\hat{k}_{t,2,j}$ also converge. This is ensured as long as each $k_{t,i,j}$ is weakly stationary. In cases where the fitted model is not weakly stationary, an alternative model such as a random walk without drift can also be used.

3.2. Poisson common factor model with variable sex-specific factors

This thesis now proposes a modified version of the PCFM with variable sex-specific factors (PCFM-VSF), which relaxes the initial assumption that the number of additional sex-specific factors is the same for each sex. This approach can be interpreted as allowing for the existence of potential factors or trends that only affect one sex or impact each sex differently. For example, life expectancy figures have shown their own distinct trends in recent decades. While both sexes have continually improved, male life expectancy has increased faster than female life expectancy over time. Li (2013) demonstrated that for Australian data, the observed difference between sexes was approximately 7 years in 1968, reduced to 4.5 years in 2007, and can be projected to narrow to 3.1 years in 2050. This lends credence to the theory that there are some factors that are impacting male mortality rates more than females.

Utilising the same notation as above, the log central death rate is modelled as:

$$\ln m_{x,t,i} = a_{x,i} + B_x K_t + \sum_{j=1}^{n_i} b_{x,i,j} k_{t,i,j}$$

where there are n_i additional sex-specific factors for sex i. The estimation procedure remains the same with slight modifications (see Appendix for details), while future death rates under this model are projected as:

$$\hat{m}_{x,t,i} = \dot{m}_{x,T,i} \exp\left(B_x(\hat{K}_t - K_T) + \sum_{j=1}^{n_i} b_{x,i,j}(\hat{k}_{t,i,j} - k_{T,i,j}) \right)$$

and the ratio of male-to-female death rates becomes:

$$\frac{\hat{m}_{x,t,2}}{\hat{m}_{x,t,1}} = \frac{\dot{m}_{x,t,2}}{\dot{m}_{x,t,1}} \exp\left(\sum_{j=1}^{n_2} b_{x,2,j}(\hat{k}_{t,2,j} - k_{T,2,j}) - \sum_{j=1}^{n_1} b_{x,1,j}(\hat{k}_{t,1,j} - k_{T,1,j}) \right)$$

3.3. Poisson common factor model with a common age effect

Hyndman and Ullah (2007) proposed a generalised version of the Lee-Carter model using a functional data analysis approach. One of the extensions for modelling multiple groups considered a common age effect (CAE), where different populations share the same age-period function. Kleinow (2015) also proposed a CAE extension of the Lee-Carter model, and found that instead of considering an individual age effect for each country, adding an extra common age-period effect resulted in a better model fit. This common age concept can readily be adapted to the PCFM. As shown in Figures 1 and 2 below, some of the computed $b_{x,i,j}$ values from the original PCFM for the countries under consideration indeed show fairly similar peaks and troughs between females and males. This interesting observation provides a strong incentive to seek a more efficient use of model parameters.

Accordingly, this thesis proposes a modification of the PCFM with a common age effect (PCFM-CAE), in which the age sensitivity measures for each additional sex-specific factor are assumed to be equal between the sexes, that is, $b_{x,1,j} = b_{x,2,j}$. This assumption allows a more parsimonious use of parameters and may lead to a lower BIC value in some cases. As a result, the log central death rate is modelled as

$$\ln m_{x,t,i} = a_{x,i} + B_x K_t + \sum_{j=1}^{n} b_{x,j} k_{t,i,j}$$

Again, only slight changes are needed for the estimation procedure (see Appendix for details). Future death rates are projected using the same equation as the baseline PCFM, with $b_{x,i,j}$ replaced by $b_{x,j}$:

$$\hat{m}_{x,t,i} = \dot{m}_{x,T,i} \exp\left(B_x (\hat{K}_t - K_T) + \sum_{j=1}^{n} b_{x,j} (\hat{k}_{t,i,j} - k_{T,i,j}) \right)$$

and the ratio of male-to-female death rates becomes:

$$\frac{\hat{m}_{x,t,2}}{\hat{m}_{x,t,1}} = \frac{\dot{m}_{x,t,2}}{\dot{m}_{x,t,1}} \exp\left(\sum_{j=1}^{n} b_{x,j} ((\hat{k}_{t,2,j} - k_{T,2,j}) - (\hat{k}_{t,1,j} - k_{T,1,j})) \right)$$

3.4. Description of data

Datasets for six populations – Australia, France, West Germany, England and Wales, the United States, and Canada – are obtained from the Human Mortality Database (HMD 2016). The six developed countries are selected on the basis that they are good representatives of the major continents including Australasia, Europe, and North America. As shown in the next chapter, the data of these countries call for different model choices amongst the alternatives considered, highlighting the importance to have more flexibility in the modelling approach. These datasets are separated by sex and single year of age. As the exposed-to-risk and death counts are generally too volatile for more advanced ages, the age range 0-89 is chosen to allow ~~for~~ more precise analysis.[4] Moreover, in line with previous studies (Li, 2013; Yang et al., 2016), the year 1970 is chosen as the start of the sample period in order to avoid the structural changes in mortality improvement that occurred around that time.[5] This ensures that the data used are relevant and helps to make projections more straightforward. The ending year of 2011 is the latest year that all six populations have data available.

[4] The volatile patterns of advanced ages tend to distort the model fitting and require a further separate analysis, e.g. see Thatcher (1999).

[5] Booth et al. (2002) developed a statistical measure for selecting the optimal fitting period for the original Lee-Carter model. Li et al. (2011) investigated the detection of structural changes and their impact on forecasting.

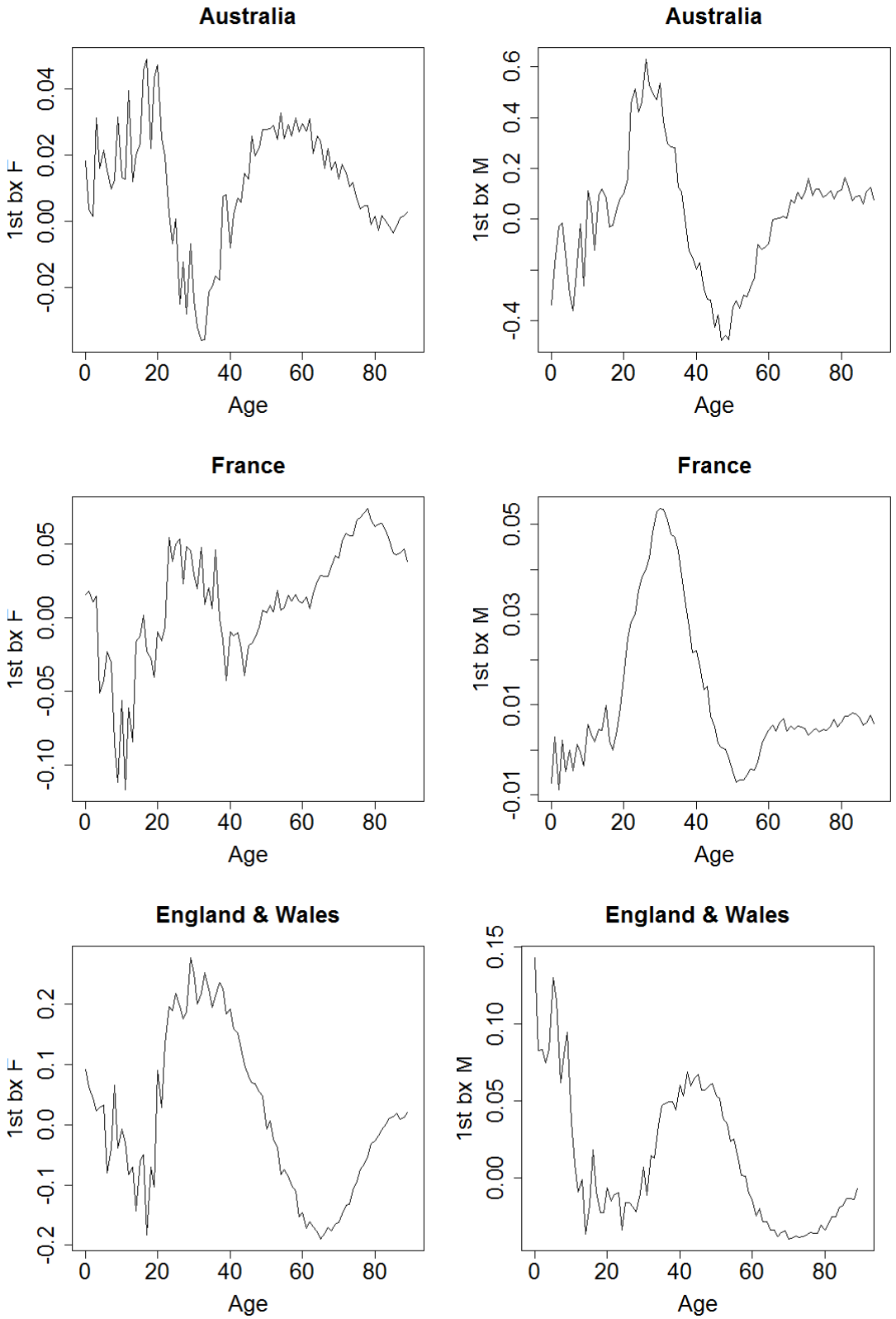


Figure 1: Computed $b_{x,i,1}$ values from the baseline PCFM for females (left) and males (right), Australia, France, and England and Wales.

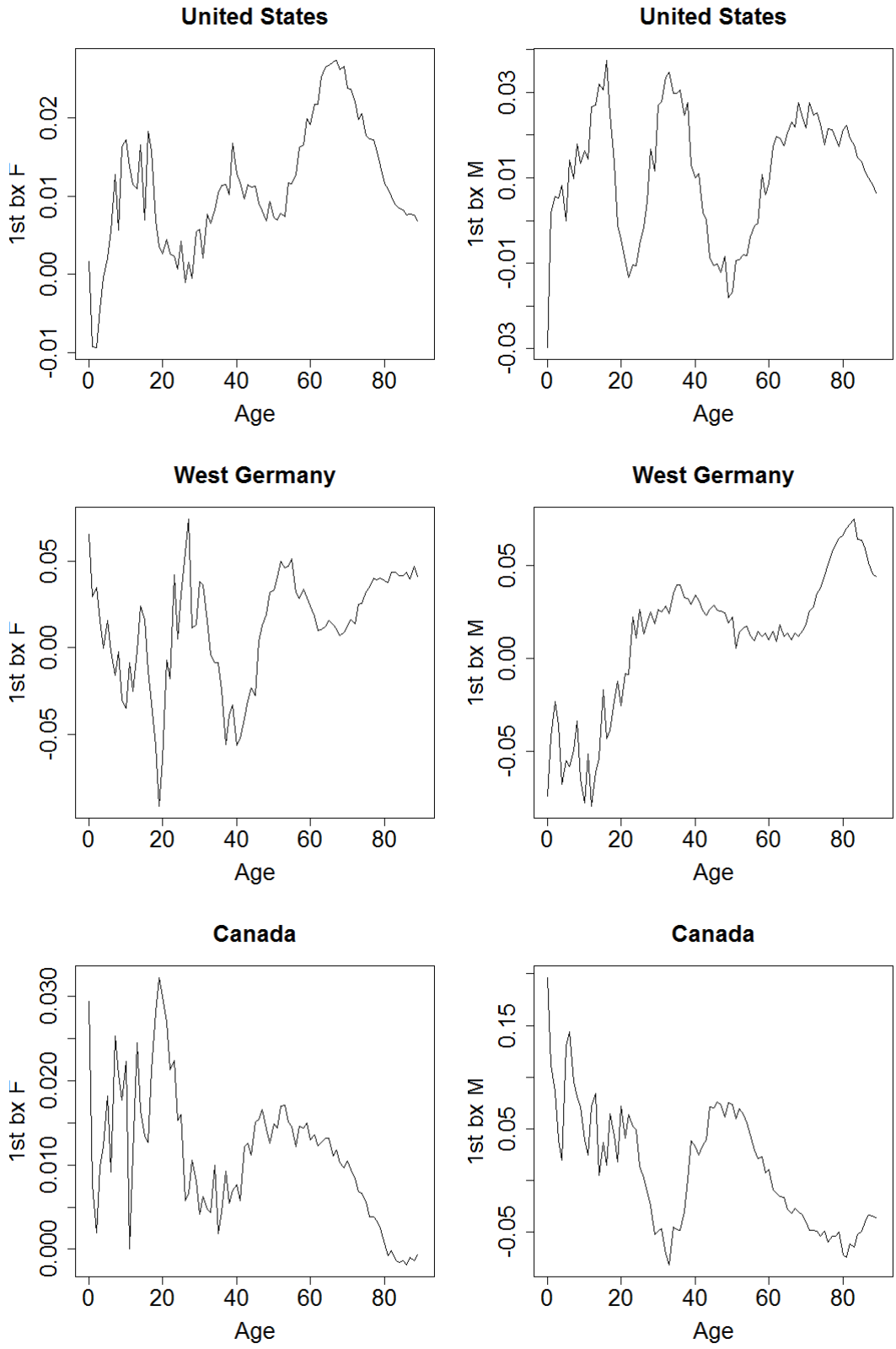


Figure 2: Computed $b_{x,i,1}$ values from the baseline PCFM for females (left) and males (right), United States, West Germany, and Canada.

Chapter 4

Analysis of Model Fitting Results

This section details the results of fitting the three models discussed in the previous chapter to the six population datasets. The R statistical software package (R Core Team, 2015) is used to perform all the computations[6]. The optimal model fit is decided by analysing the BIC values and standardised residual plots. Table 1 shows the BIC results for the PCFM-VSF and PCFM-CAE for Australia. Baseline PCFM results are also included on the diagonal, as the PCFM-VSF is effectively an expansion of the base model. The baseline model agrees with the results from Li (2013) and Yang et al. (2016) that two sex-specific factors is the optimal choice, with the lowest BIC value (72,718) on the diagonal. The PCFM-CAE results suggest a further improvement in model fit (71,838), where there is no change in the optimal number of factors. In contrast, the PCFM-VSF results indicate that the optimal choice is to eliminate one factor from the optimal baseline model – still using two male-specific factors but incorporating only one female-specific factor (72,674). For Australian data, both new models produce a better fit than the baseline model, with the PCFM-CAE as the most optimal choice amongst all.

[6] The R code used is available from the author on request.

Table 1: BIC values for the PCFM, PCFM-VSF and PCFM-CAE, Australia.

		No. of female factors							
		0	1	2	3	4	5	6	CAE
No. of male factors	0	75,495	75,189	75,233	75,902	76,727	77,581	78,498	75,495
	1	73,246	72,941	72,985	73,654	74,478	75,332	76,249	72,983
	2	72,980	**72,674**	*72,718*	73,387	74,212	75,066	75,983	**71,838**
	3	73,632	73,326	73,370	74,039	74,864	75,717	76,634	72,277
	4	74,456	74,151	74,195	74,864	75,688	76,542	77,459	73,203
	5	75,329	75,023	75,067	75,736	76,561	77,414	78,331	74,237
	6	76,245	75,939	75,983	76,652	77,477	78,331	79,248	75,351

The BIC values for France are presented in Table 2. The baseline model points to the use of four factors (94,176). Similar to the results above, the PCFM-VSF suggests that the optimal model choice is to remove one factor from the optimal baseline model, resulting in three female-specific factors and four male-specific factors (93,743). On the other hand, the PCFM-CAE leads to a different conclusion of adding more factors instead, where the optimal choice is using five factors for each sex (92,980). For French data, both new models deliver a better fit than the original model, again with the PCFM-CAE being the most optimal one.

Table 2: BIC values for the PCFM, PCFM-VSF and PCFM-CAE, France.

		No. of female factors							
		0	1	2	3	4	5	6	CAE
No. of male factors	0	119,537	115,127	111,205	111,007	111,440	111,986	112,663	119,537
	1	111,222	106,812	102,890	102,692	103,125	103,671	104,348	107,222
	2	105,437	101,027	97,106	96,907	97,340	97,886	98,563	100,074
	3	103,298	98,888	94,967	94,768	95,201	95,747	96,424	96,158
	4	102,273	97,863	93,941	**93,743**	*94,176*	94,722	95,399	93,998
	5	102,570	98,160	94,238	94,040	94,473	95,019	95,696	**92,980**
	6	103,127	98,717	94,795	94,597	95,030	95,576	96,253	93,290

The BIC values for England and Wales are set forth in Table 3. Here, there is no improvement in the PCFM-VSF over the baseline – both versions choose $n = 3$ for both male- and female-specific factors (90,810). In comparison, the PCFM-CAE suggests an additional factor for each sex which results in a better model fit under the common age effect (89,174).

Table 3: BIC values for the PCFM, PCFM-VSF and PCFM-CAE, England & Wales.

		No. of female factors							
		0	1	2	3	4	5	6	CAE
No. of male factors	0	123,144	112,825	108,014	106,555	106,723	107,446	108,210	123,144
	1	113,530	103,211	98,400	96,941	97,109	97,832	98,596	104,438
	2	107,799	97,480	92,669	91,210	91,378	92,101	92,865	95,180
	3	107,400	97,081	92,270	***90,810***	90,979	91,701	92,465	90,150
	4	107,558	97,239	92,428	90,969	91,137	91,860	92,624	**89,174**
	5	108,199	97,880	93,069	91,609	91,778	92,500	93,264	89,791
	6	108,866	98,547	93,735	92,276	92,445	93,167	93,931	90,545

Table 4 lists the BIC values for the United States. There is a slight departure in the model fitting results here compared to other populations. While the baseline model recommends $n = 5$ (114,904), the PCFM-VSF, rather than removing a factor from the optimal baseline, suggests adding an extra male-specific factor, leading to five female-specific and six male-specific factors (114,788). The PCFM-CAE also supports adding more factors over the optimal baseline model, resulting in six factors for each sex as the optimal choice (115,012). For United States data, the PCFM-VSF shows the best model fit, whereas the PCFM-CAE is the least optimal among the three candidates.

Table 4: BIC values for the PCFM, PCFM-VSF and PCFM-CAE, United States.

		No. of female factors							
		0	1	2	3	4	5	6	CAE
No. of male factors	0	323,884	269,642	247,122	233,627	231,845	231,682	232,047	323,884
	1	272,805	218,563	196,043	182,547	180,766	180,603	180,968	222,510
	2	234,615	180,372	157,852	144,357	142,575	142,412	142,777	170,417
	3	214,808	160,565	138,045	124,550	122,768	122,605	122,970	138,098
	4	209,124	154,882	132,362	118,866	117,085	116,921	117,287	121,819
	5	207,106	152,864	130,344	116,848	115,067	*114,904*	115,269	117,425
	6	206,991	152,748	130,228	116,733	114,951	**114,788**	115,153	**115,012**

Table 5 presents the BIC values for West Germany. The PCFM-VSF shows no improvement over the baseline model here (101,580), but the PCFM-CAE suggests adding two extra factors for each sex, rather than just one like previously, and leads to a lower BIC value (99,922).

Table 5: BIC values for the PCFM, PCFM-VSF and PCFM-CAE, West Germany.

		No. of female factors							
		0	1	2	3	4	5	6	CAE
No. of male factors	0	120,661	116,638	114,439	113,910	113,724	113,747	113,890	120,661
	1	113,223	109,199	107,000	106,471	106,285	106,308	106,451	109,367
	2	110,722	106,699	104,500	103,971	103,785	103,808	103,951	105,193
	3	109,297	105,274	103,074	102,545	102,359	102,382	102,526	102,647
	4	108,518	104,494	102,295	101,766	***101,580***	101,603	101,746	101,039
	5	108,757	104,734	102,535	102,006	101,820	101,843	101,986	100,153
	6	109,071	105,048	102,849	102,320	102,134	102,157	102,300	**99,922**

Finally, the BIC values for Canada are given in Table 6. The results here are similar to those of England and Wales, in which the PCFM-VSF makes no improvement over the baseline (77,159), while the PCFM-CAE requires one more factor for each sex and produces a better model fit (76,381).

Table 6: BIC values for the PCFM, PCFM-VSF and PCFM-CAE, Canada.

		No. of female factors							
		0	1	2	3	4	5	6	CAE
No. of male factors	0	88,740	85,687	85,184	85,282	86,167	87,066	87,990	88,740
	1	83,266	80,213	79,710	79,808	80,693	81,592	82,516	82,576
	2	80,715	77,662	***77,159***	77,257	78,142	79,042	79,965	77,046
	3	81,064	78,011	77,508	77,606	78,491	79,390	80,314	**76,381**
	4	81,841	78,788	78,285	78,383	79,268	80,167	81,090	77,061
	5	82,681	79,628	79,125	79,223	80,108	81,007	81,931	78,075
	6	83,583	80,530	80,027	80,125	81,010	81,909	82,833	79,166

Overall, the PCFM-CAE leads to the best model fit for five of the countries, the PCFM-VSF is the best one for one case, and the baseline PCFM is the least optimal amongst the three options for five countries. These results clearly show that the two proposed extensions outperform the original model in terms of fitting population mortality data from a number of different countries. It appears that the age sensitivity is rather similar between the sexes for certain countries and so setting common age sensitivity can make the model more parsimonious in these cases. Moreover, a different number of factors for each sex can provide more flexibility for modelling mortality data in some cases.

Next, the standardised deviance residuals[7] for the three models and six datasets are displayed in Figures 3-8. Overall, there do not appear to be any significant systematic patterns in the residuals plotted against age or calendar year. However, there are signs of some weak patterns in the residuals plotted against cohort year. More importantly, there are slight differences in the magnitude of these patterns between the models. For Australia, although the PCFM-VSF improves the model fit over the baseline PCFM, it comes at the cost of a more obvious pattern in the residuals for females. This feature has also been found for France, though to a lesser extent. These observations are likely due to the optimal PCFM-VSF model having one less female-specific factor compared to the baseline PCFM. However, the same is not evident for the United States, owing to the large number of factors used in both models. For the PCFM-CAE, there are no noticeable differences in cohort residual randomness for Australia, France and the United States when compared to the baseline PCFM. For England and Wales, as well as Canada, the PCFM-CAE shows a slight improvement in the residuals over the baseline model. In contrast, for West Germany, the PCFM-CAE has a slightly more pronounced pattern in the residuals. In fact, these cohort patterns may be addressed by modifying the PCFM-VSF and PCFM-CAE to include a cohort factor in a similar approach to Yang et al. (2016) – this modification is left as an option for future research.

To conduct a more thorough investigation on the deviance residuals, the autocorrelation function (ACF) can be used. For each age and sex combination, if the sample ACF is less than twice the estimated standard error in magnitude, it is considered insignificant. While the results are not shown here, there are significant autocorrelations at multiple lags for many age and sex combinations. As the usage of the BIC in model fitting assumes that serial correlations in the data have been adequately captured, this result is slightly problematic for the proposed models.

[7] To account for possible over-dispersion in the data, the residuals are standardised with respect to the dispersion parameter (Li, 2013; Yang et al., 2016). The equation is shown in the Appendix.

In order to conduct another check of the suitability of the model fit, the mean absolute percentage error (MAPE) values of the fitted log central death rates are displayed in Table 7. The MAPE is defined as:

$$\frac{1}{n_d}\sum_{x,t,i}\left|\frac{\ln \widetilde{m}_{x,t,i} - \ln(d_{x,t,i}/e_{x,t,i})}{\ln(d_{x,t,i}/e_{x,t,i})}\right|$$

where n_d is the number of data points, $\widetilde{m}_{x,t,i}$ is the fitted log central death rate, and $d_{x,t,i}$ and $e_{x,t,i}$ are observed values of the number of deaths and exposed-to-risk respectively. The MAPE values are expressed in percentage below. As an alternative measure, the MAPE values on the fitted actual central death rates are also provided.

Table 7: MAPE values for fitted log (actual) central death rates

Country	PCFM	PCFM-VSF	PCFM-CAE
Australia	1.12% (7.14%)	1.18% (7.40%)	1.17% (7.53%)
France	0.66% (3.96%)	0.67% (4.08%)	0.67% (4.07%)
England & Wales	0.73% (4.57%)	-	0.72% (4.56%)
United States	0.35% (1.97%)	0.34% (2.06%)	0.37% (1.94%)
West Germany	0.72% (4.36%)	-	0.73% (4.42%)
Canada	0.90% (5.66%)	-	0.88% (5.61%)

All the MAPE values for log rates are quite small in general, and the differences are very small between the three models. None of the models has a clear advantage over the others. In terms of the goodness-of-fit, the performances of all three models are satisfactory. This pattern also holds for considering the MAPE values on the actual rates.

As a final note, it is interesting to see that there seems to be some trade-off between setting common age sensitivity (fewer parameters) and adding more factors (more parameters). For five of the countries being considered, the PCFM-CAE applications require adding one or two extra factors compared to the baseline model. Consequently, there are more time components in the structure and so more time series models are needed to perform future projections. Despite a better model fit, using more time series models complicates the projection exercise and does not necessarily lead to better projection results. This issue will be investigated in the next chapter.

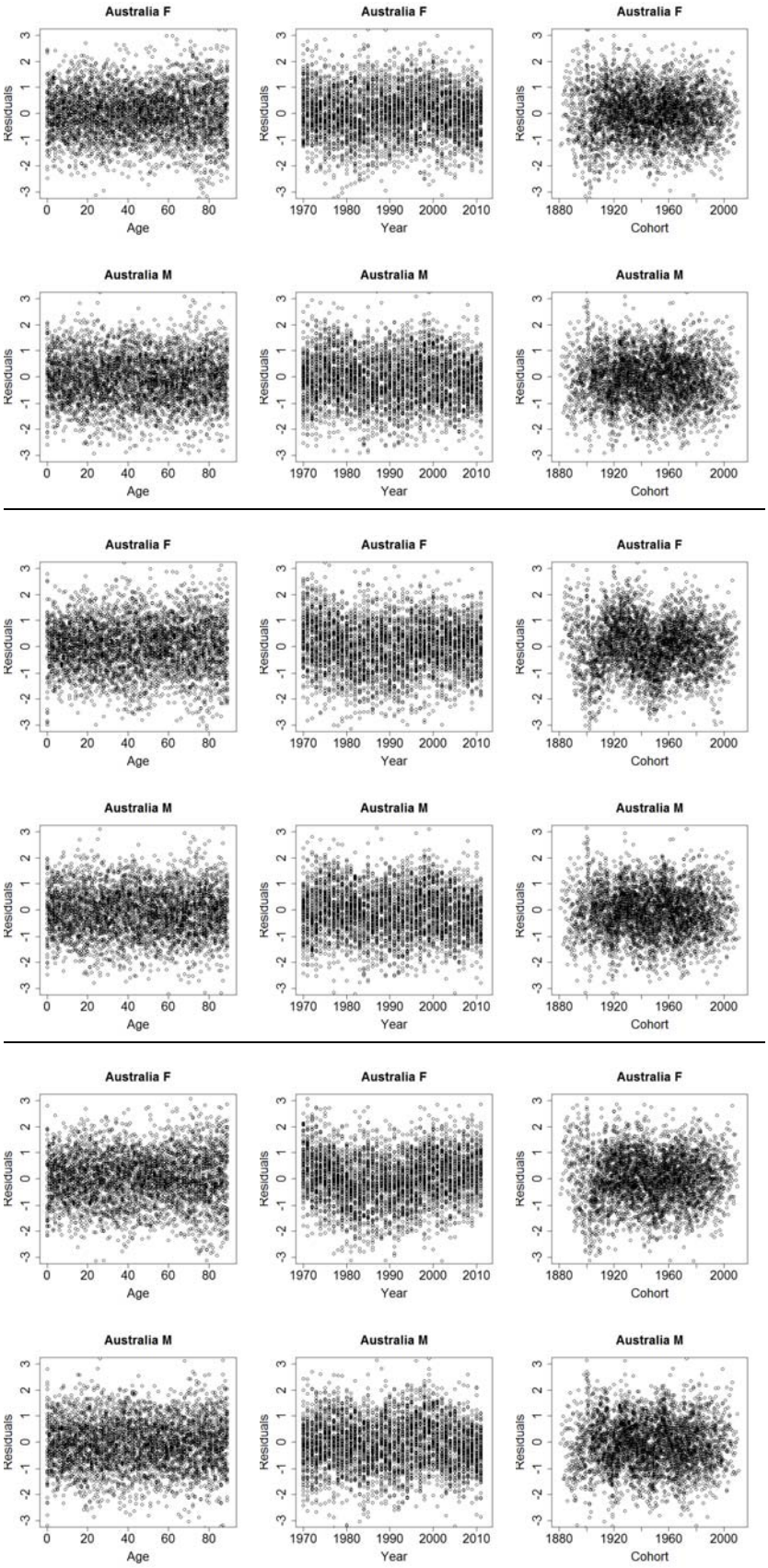


Figure 3: Standardised deviance residuals for the PCFM (top panel), PCFM-VSF (middle panel) and PCFM-CAE (bottom panel), Australia. The first row is for females and the second for males. Residuals are plotted against age (left), calendar year (middle), and then cohort year (right).

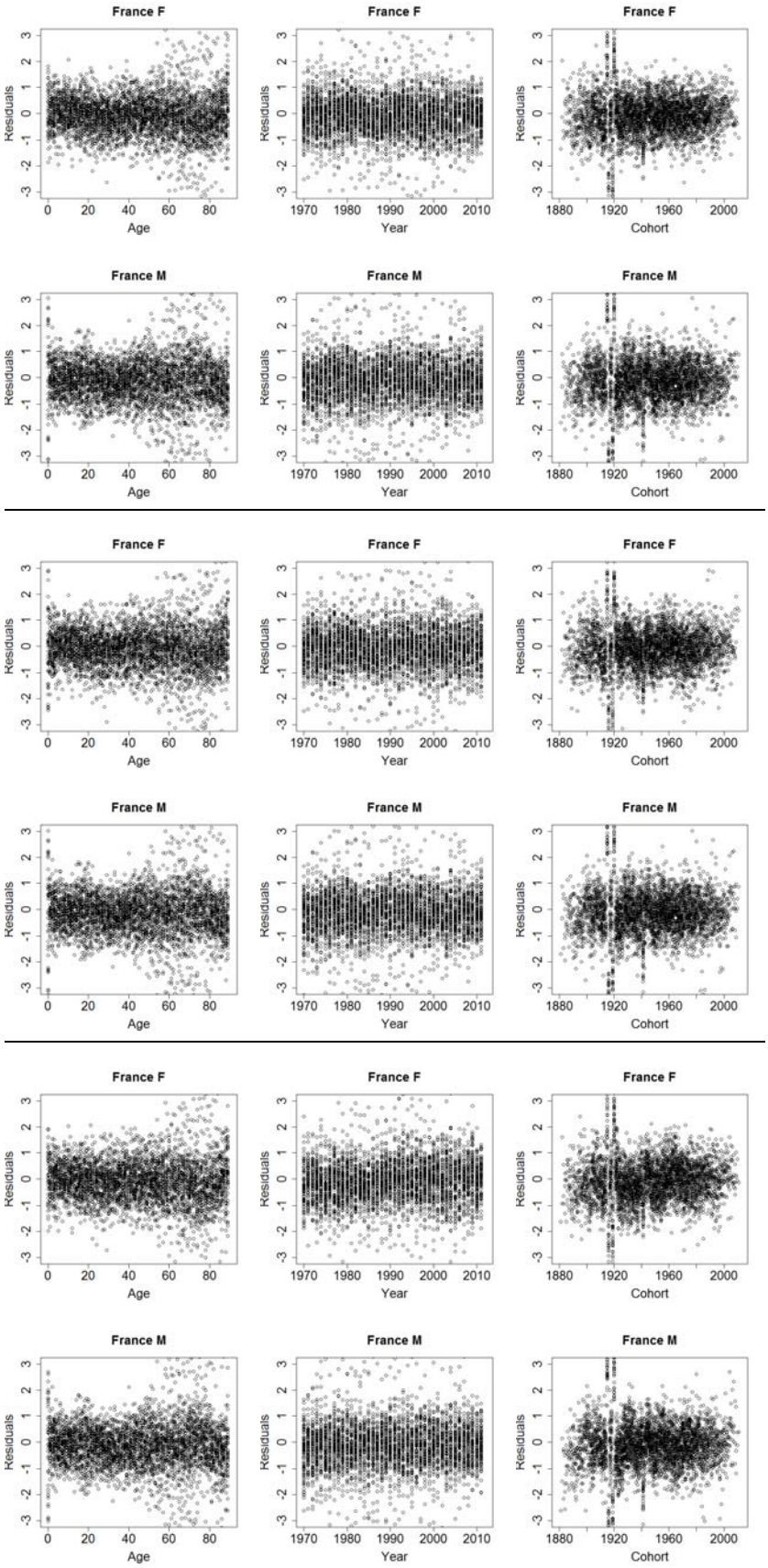


Figure 4: Standardised deviance residuals for the PCFM (top panel), PCFM-VSF (middle panel) and PCFM-CAE (bottom panel), France. The first row is for females and the second for males. Residuals are plotted against age (left), calendar year (middle), and then cohort year (right).

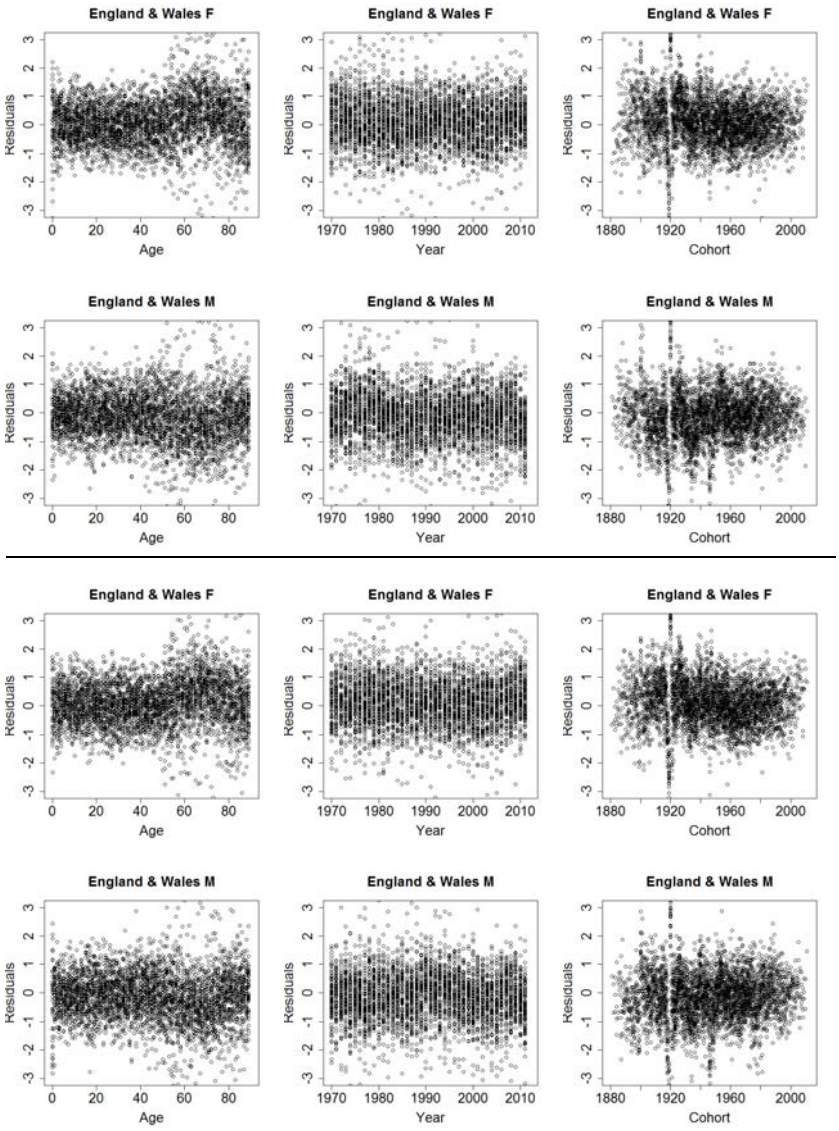


Figure 5: Standardised deviance residuals for the PCFM (top panel) and PCFM-CAE (bottom panel), England and Wales. The first row is for females and the second for males. Residuals are plotted against age (left), calendar year (middle), and then cohort year (right).

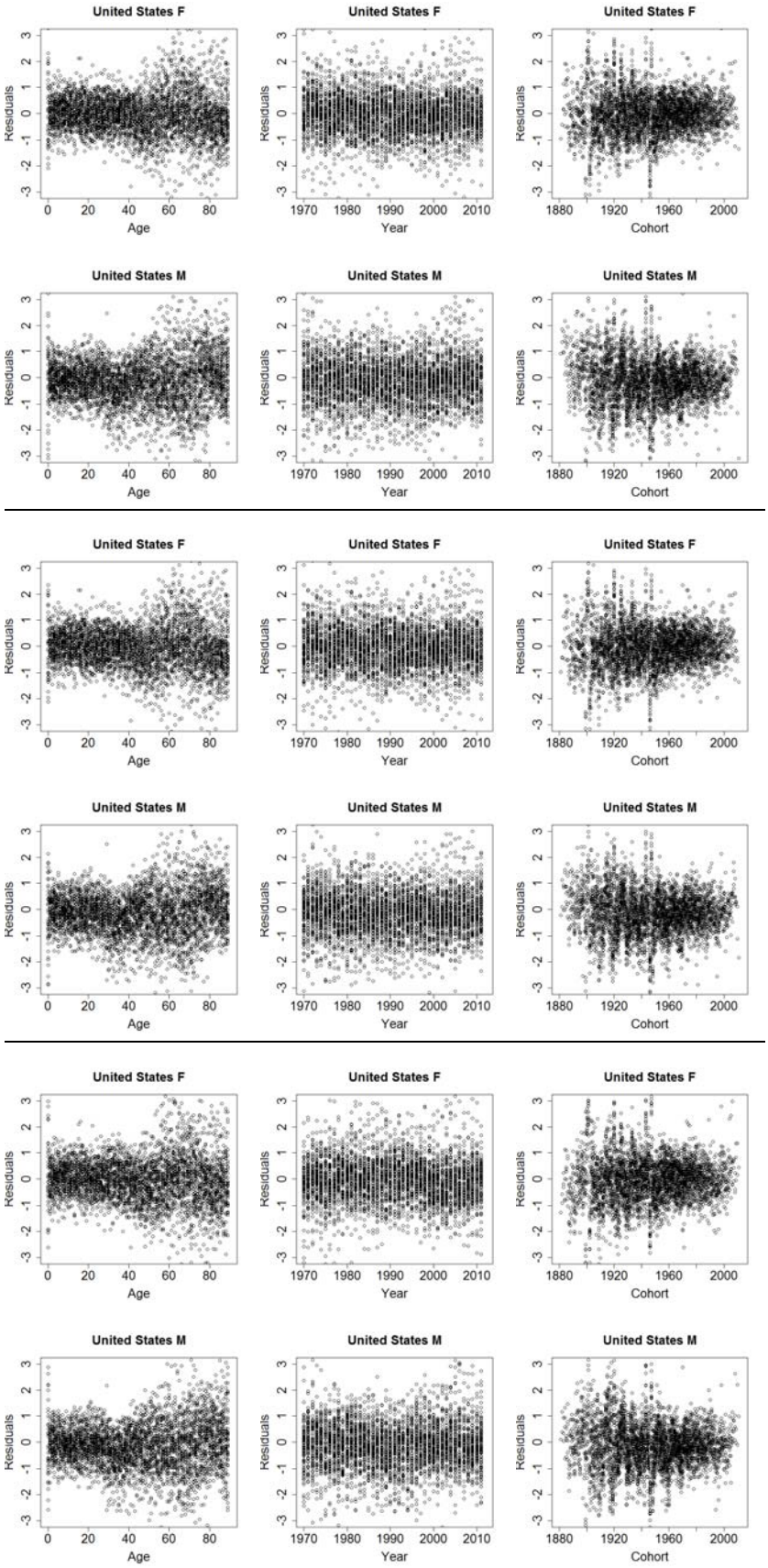


Figure 6: Standardised deviance residuals for the PCFM (top panel), PCFM-VSF (middle panel) and PCFM-CAE (bottom panel), United States. The first row is for females and the second for males. Residuals are plotted against age (left), calendar year (middle), and then cohort year (right).

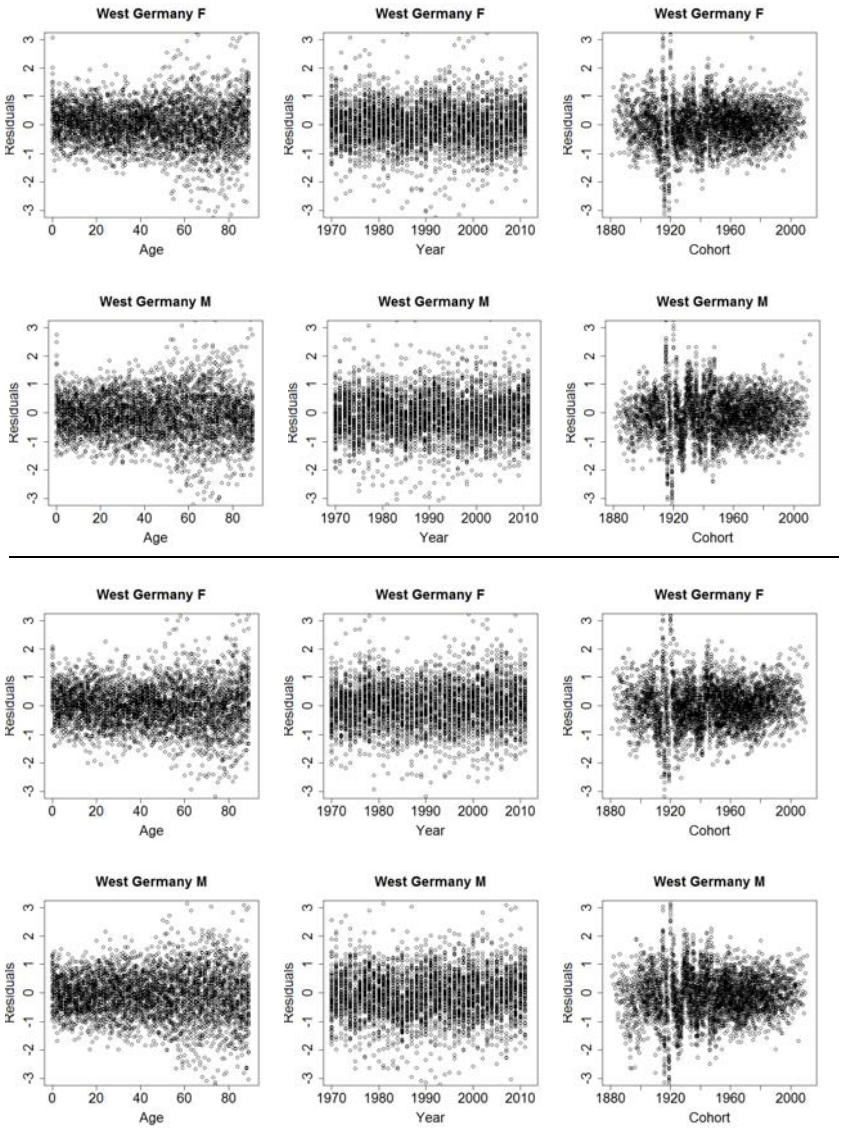


Figure 7: Standardised deviance residuals for the PCFM (top panel) and PCFM-CAE (bottom panel), West Germany. The first row is for females and the second for males. Residuals are plotted against age (left), calendar year (middle), and then cohort year (right).

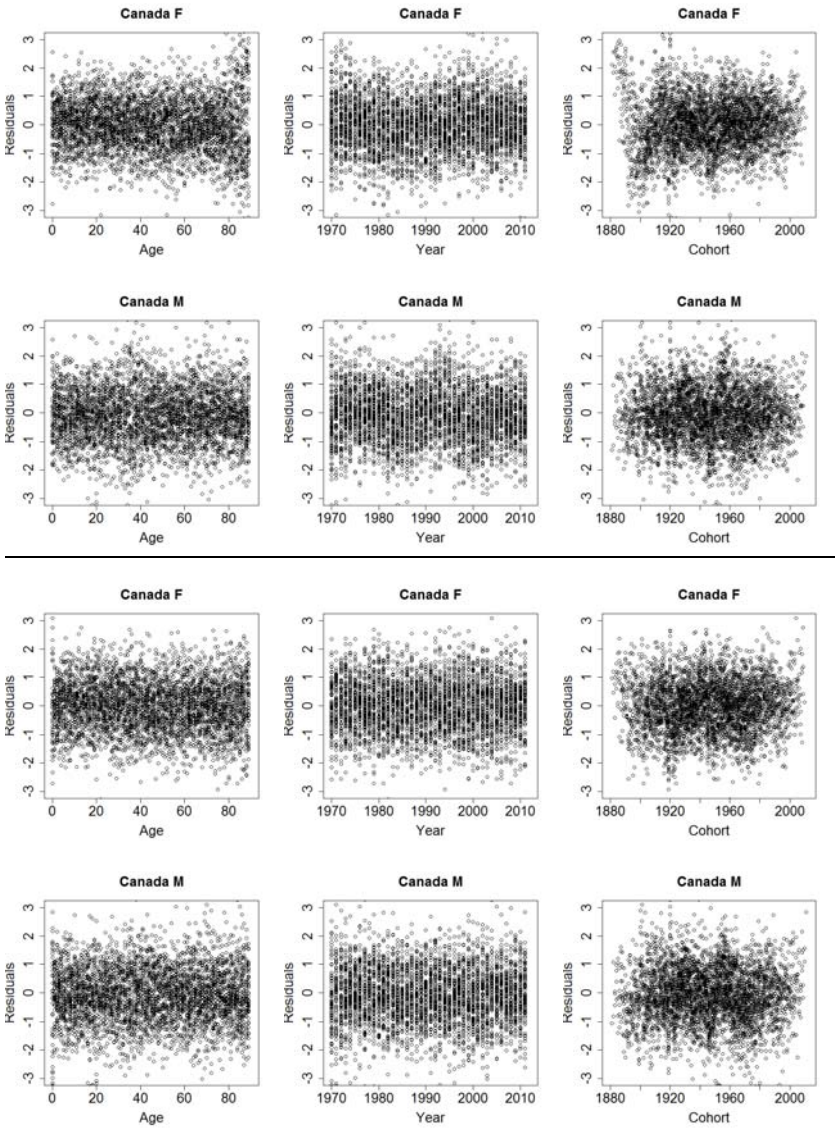


Figure 8: Standardised deviance residuals for the PCFM (top panel) and PCFM-CAE (bottom panel), Canada. The first row is for females and the second for males. Residuals are plotted against age (left), calendar year (middle), and then cohort year (right).

Chapter 5

Model Projection

5.1. Analysis of forecasting results

This section investigates the accuracy and convergence of the mortality forecasts produced by the optimal model for each of the three variations. As mentioned in Chapter 3, the common mortality index K_t is modelled as a random walk with drift, while the sex-specific terms $k_{t,i,j}$ are modelled as AR(p) processes. The order p is chosen based on the PACF of the time components, the autocorrelations of the residuals[8], and whether the fitted model is weakly stationary. In cases where a weakly stationary AR(p) model does not exist, a random walk without drift is used as a substitute. The purpose is to ensure that the forecast ratio of male-to-female death rates at each age eventually converges to a constant, thus avoiding the potential problems of unrealistic projected ratios and mortality crossover as demonstrated in Li (2013) and Yang et al. (2016).

In line with the 42 years of data from 1970-2011, the projection period is chosen to be 42 years long ranging from 2012-2053. Figures 9-14 display the projected ratios of male-to-female death rates for the three models. In general, the three models all show similar projection properties. There are some slight differences in the direction and speed of convergence across the age groups, but the projected ratios all converge ultimately and appear roughly in line with the past trends. One possible issue arises

[8] See Appendix for details.

with the PCFM-CAE projection for West Germany – because the optimal model has two more factors than the baseline model, there are four extra time components to be projected. Using more time series models for the extra components results in forecast ratios with much stronger autoregressive effects, most evident in the 0-9 and 10-29 age groups.[9] Although the ratios still converge and the projection reflects the past data well, the frequent fluctuations may nevertheless be undesirable for forecasting purposes. This problem could be mitigated by using AR(1) models as opposed to AR(p) models for the time components, but this alternative approach is rather arbitrary.

5.2. Analysis of out-of-sample testing results

Another method of checking the accuracy of mortality forecasts is to perform out-of-sample testing, also known as backtesting, to compare the PCFM, PCFM-VSF and PCFM-CAE mortality projections against observed data.[10] The sample period is split into two parts – the first 30 years (1970-1999) are used to fit each model and the remaining 12 years (2000-2011) are used to evaluate the forecasting performance of that model. The projection accuracy is measured using the MAPE of the projected log (actual) central death rates of each sex i, defined in a similar manner to Chapter 4. The MAPE of the projected ratios of male-to-female death rates is also considered, defined as:

$$\frac{1}{n_d}\sum_{x,t}\left|\frac{\hat{m}_{x,t,2}/\hat{m}_{x,t,1}-(d_{x,t,2}/e_{x,t,2})/(d_{x,t,1}/e_{x,t,1})}{(d_{x,t,2}/e_{x,t,2})/(d_{x,t,1}/e_{x,t,1})}\right|.$$

[9] It is worth noting that three of the four extra time components in the PCFM-CAE for West Germany are of order 3 or higher.

[10] As before, cases where the optimal PCFM-VSF selects the same number of factors as the PCFM are omitted.

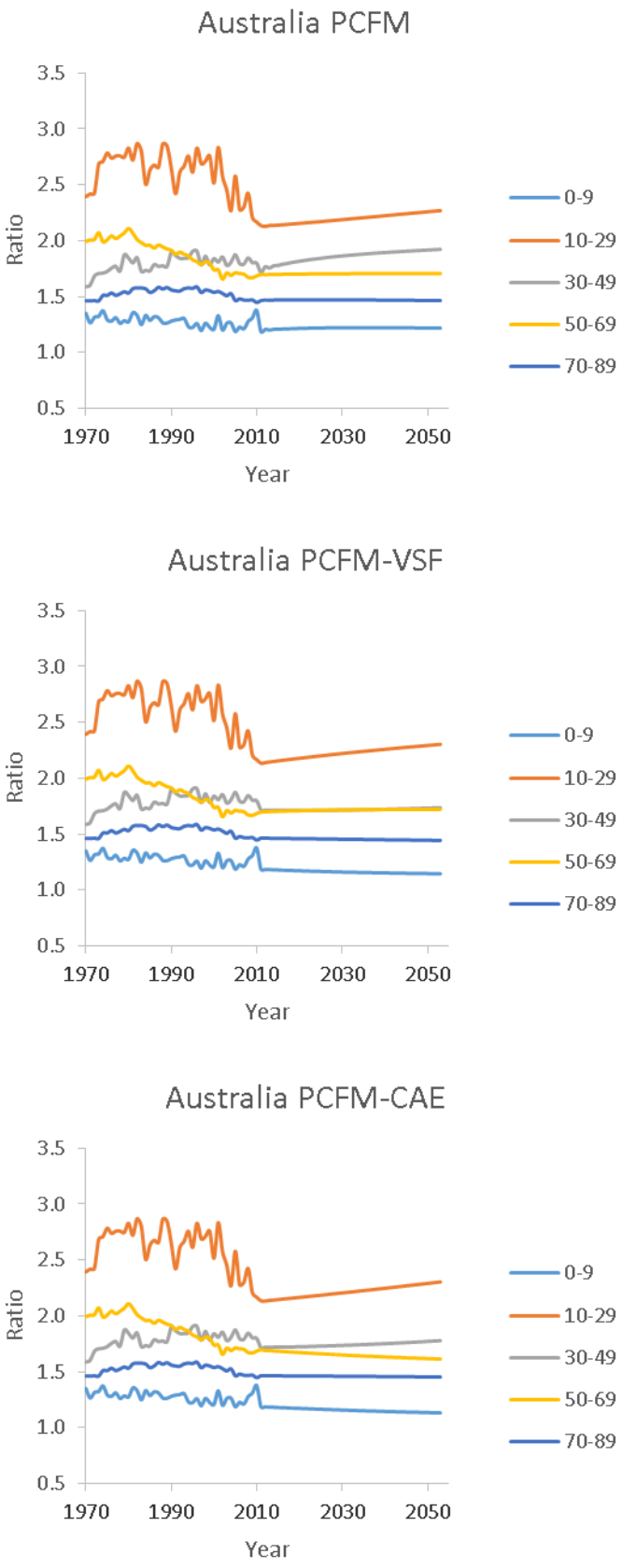


Figure 9: Projected ratios of male-to-female death rates, Australia.

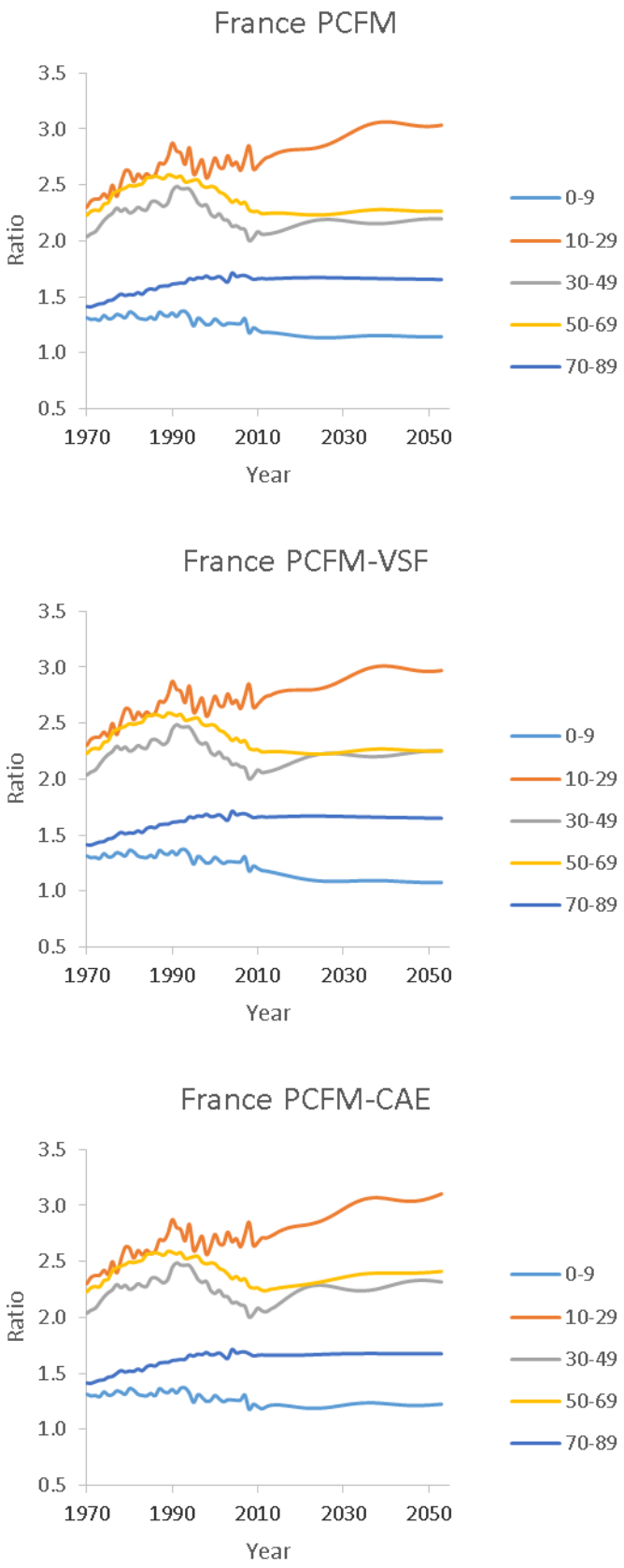


Figure 10: Projected ratios of male-to-female death rates, France.

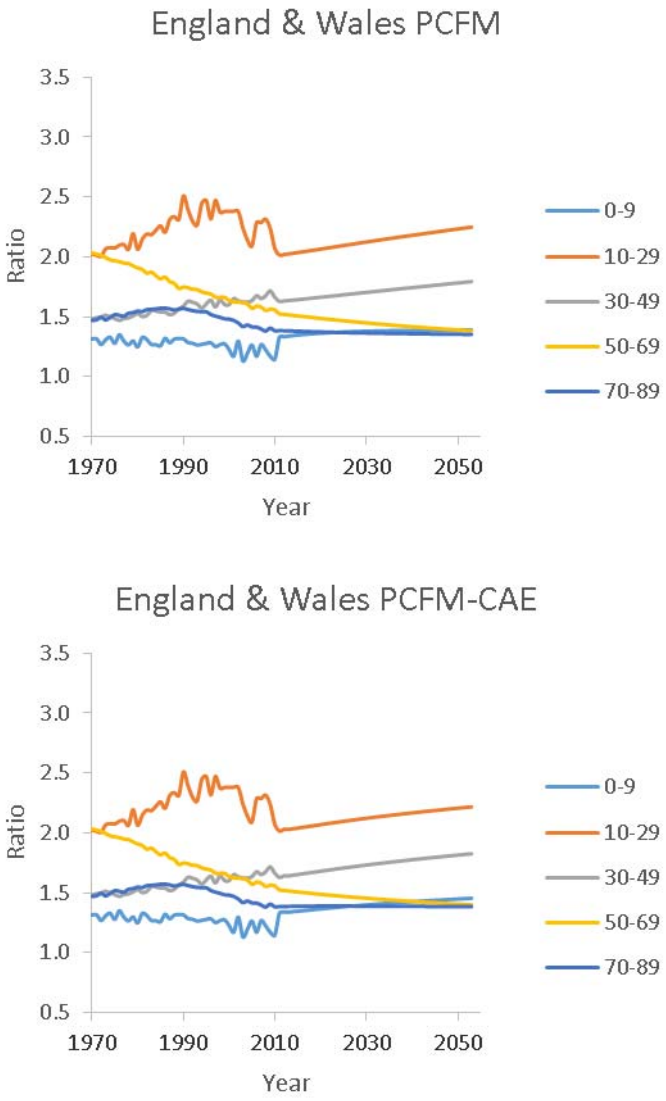


Figure 11: Projected ratios of male-to-female death rates, England and Wales.

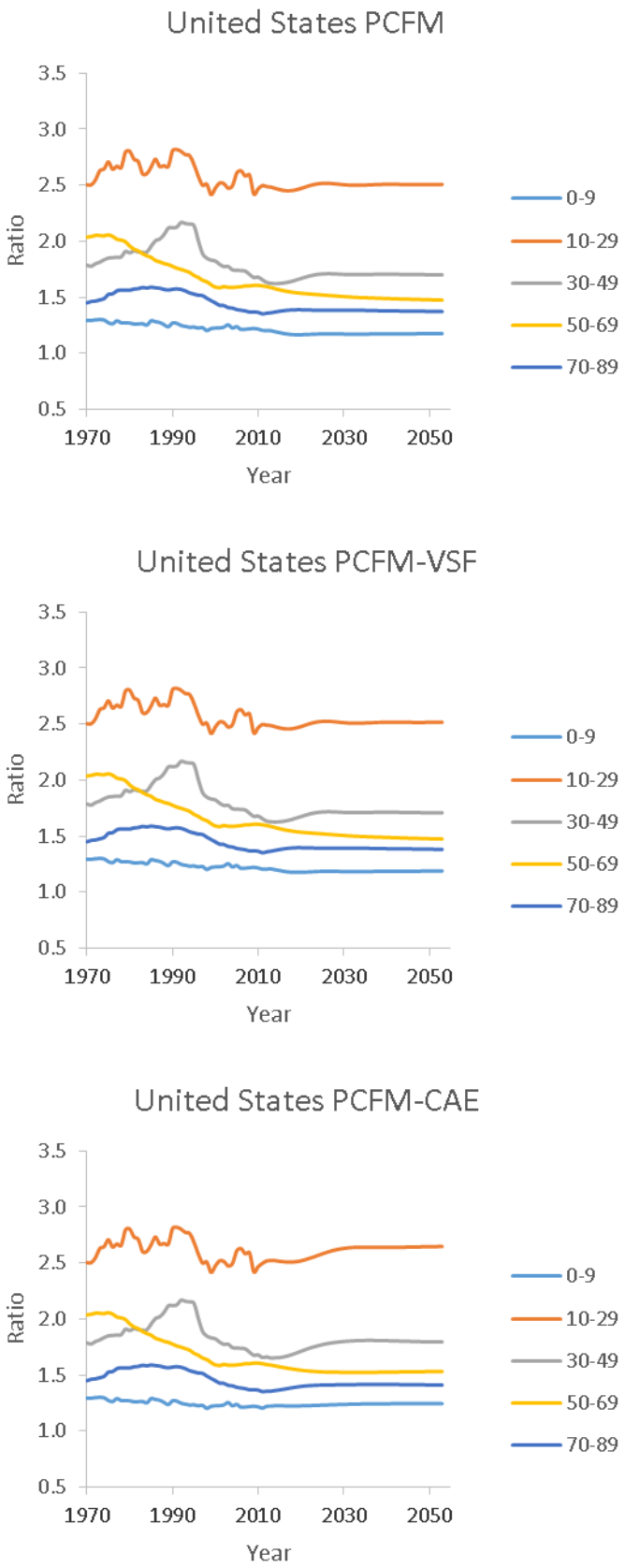


Figure 12: Projected ratios of male-to-female death rates, United States.

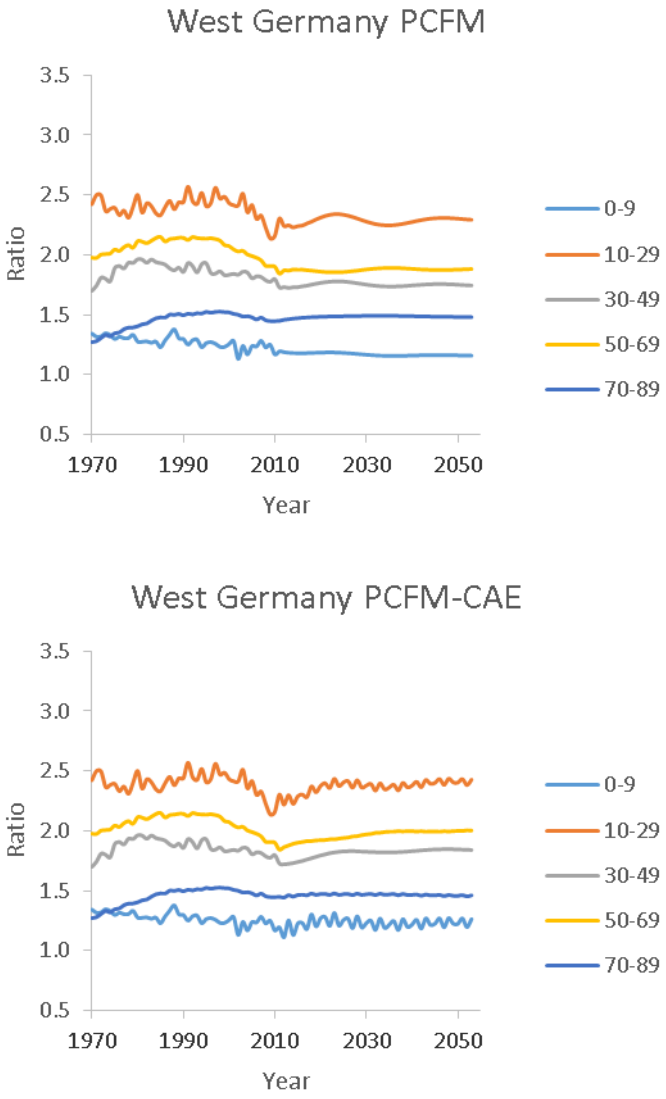


Figure 13: Projected ratios of male-to-female death rates, West Germany.

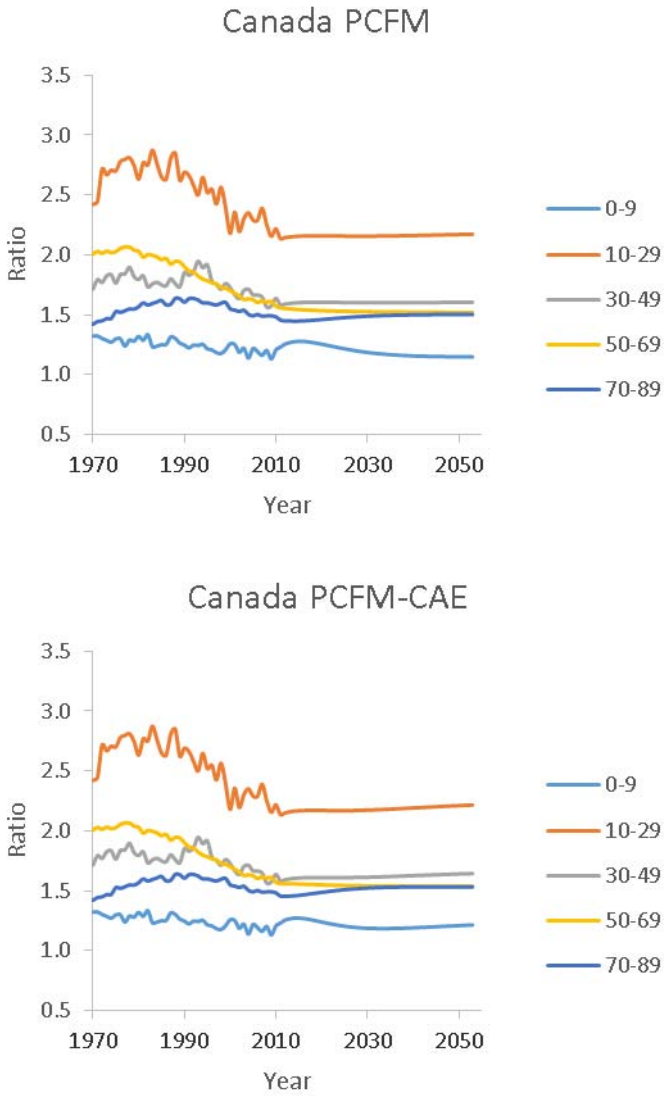


Figure 14: Projected ratios of male-to-female death rates, Canada.

Table 8 presents the MAPE values of the projected estimates of the six countries under the three models, and Figures 15-20 display the projected male-to-female ratios compared to observed values. Similar to the results in the previous chapter, the MAPE values for male and female log central death rates are very small, and the differences between models are negligible. The MAPE values for actual death rates are quite a bit larger, especially for Australia, but again, no single model has clear advantages compared to the others. However, when considering male-to-female ratios, the PCFM-CAE noticeably outperforms the baseline model in all cases except one. The more obvious examples are France (5.38% compared to 6.02%) and West Germany (5.01% compared to 6.77%). This outcome is further emphasised by looking at the graphs – for France (Figure 16), age groups 0-9 and 30-49 are the main contributors to the difference, while for West Germany (Figure 19), age group 10-29 is the key difference. [11] These results suggest that the PCFM-CAE may be a more accurate model for forecasting in regards to considering male-to-female death rate ratios.

For the PCFM-VSF compared to the baseline model, the results are not as clear. Regarding projected male-to-female ratios, the PCFM-VSF slightly outperforms the base model for Australia, is slightly outperformed instead for France, and the difference between the models for the United States is negligible. From a graphical perspective, only France appears to have a noticeable difference – the PCFM-VSF is slightly worse at the very end of the projection in age groups 0-9 and 10-29. With fewer populations to compare, it is difficult to come to any conclusions as to whether the PCFM-VSF outperforms the baseline model.

It is important to consider that the results of out-of-sample testing can change substantially if a different split between the sample and projection periods is used. To further investigate the forecast accuracy of the three models, the analysis is repeated with a sample period of 25 years (1970-1994) and a projection period of 17 years (1995-2011). The results of this second out-of-sample test are presented in Table 9 and Figures 21-26. Comparing the two tests in regards to MAPE values for individual sexes, there does not appear to be an overall pattern in performance changes. However,

[11] Again, it is worth noting that the optimal models for West Germany contain some higher order AR(p) processes, resulting in a highly fluctuating projection that overly reflects past data. Additionally, in this out-of-sample test, the optimal PCFM-CAE has six additional factors compared to the PCFM's three.

when looking at the MAPE values for projected male-to-female death rate ratios, the longer projection period results in lower accuracy in all cases except the PCFM for West Germany. Comparing the graphs between the two analyses, France (Figure 22) performs noticeably worse with the shorter data period, where the 10-29 age group is visibly less accurate. For the United States (Figure 24), the projected male-to-female death rate ratios are constantly higher than the actual values for the 10-29 and 30-49 age groups, in contrast to the first analysis where the projected ratios are constantly lower. Overall, however, there does not appear to be a significant decrease in accuracy for the larger projection period.

One final thing to note is that all the models considered sometimes produce death rate ratio forecasts that are fairly different to what have actually happened, including ages 10 to 29 for Australia and West Germany and ages 30 to 69 for France. Nevertheless, the results presented here and in the previous chapter broadly reaffirm that the PCFM-CAE and PCFM-VSF have potential advantages over the base PCFM, without significantly sacrificing fitting or projection accuracy.

Table 8: MAPE values from the PCFM, PCFM-VSF and PCFM-CAE

	PCFM			PCFM-VSF			PCFM-CAE		
Country	Female[a]	Male[a]	Male to female ratios[b]	Female[a]	Male[a]	Male to female ratios[b]	Female[a]	Male[a]	Male to female ratios[b]
Australia	1.99% (15.55%)	2.56% (18.15%)	7.71%	2.06% (16.09%)	2.56% (18.15%)	7.43%	2.02% (15.87%)	2.55% (17.97%)	7.21%
France	1.95% (14.45%)	2.20% (13.97%)	6.02%	1.86% (13.44%)	2.20% (13.97%)	6.32%	1.90% (14.09%)	2.08% (12.77%)	5.38%
England & Wales	1.85% (11.29%)	2.53% (13.16%)	5.22%	-	-	-	1.76% (11.17%)	2.60% (13.25%)	5.70%
United States	1.27% (7.10%)	1.89% (8.38%)	4.39%	1.27% (7.08%)	1.89% (8.38%)	4.41%	1.21% (6.97%)	1.70% (7.82%)	3.84%
West Germany	1.65% (10.56%)	1.76% (10.25%)	6.77%	-	-	-	1.63% (10.88%)	1.79% (10.80%)	5.01%
Canada	1.80% (11.19%)	2.38% (12.32%)	6.38%	-	-	-	1.80% (11.20%)	2.39% (12.38%)	6.34%

This out-of-sample analysis, applies the PCFM, PCFM-VSF (where applicable) and PCFM-CAE to the first period, projects death rates for the second period, and compares the projected values against the observed rates. The first period is 30 years and the second period is 12 years.

[a]: MAPE values are based on fitted log (actual) central death rates.

[b]: The MAPE values of the projected male-to-female ratios of death rates are based on 10-year age groups, since the ratios can be volatile for younger individual ages.

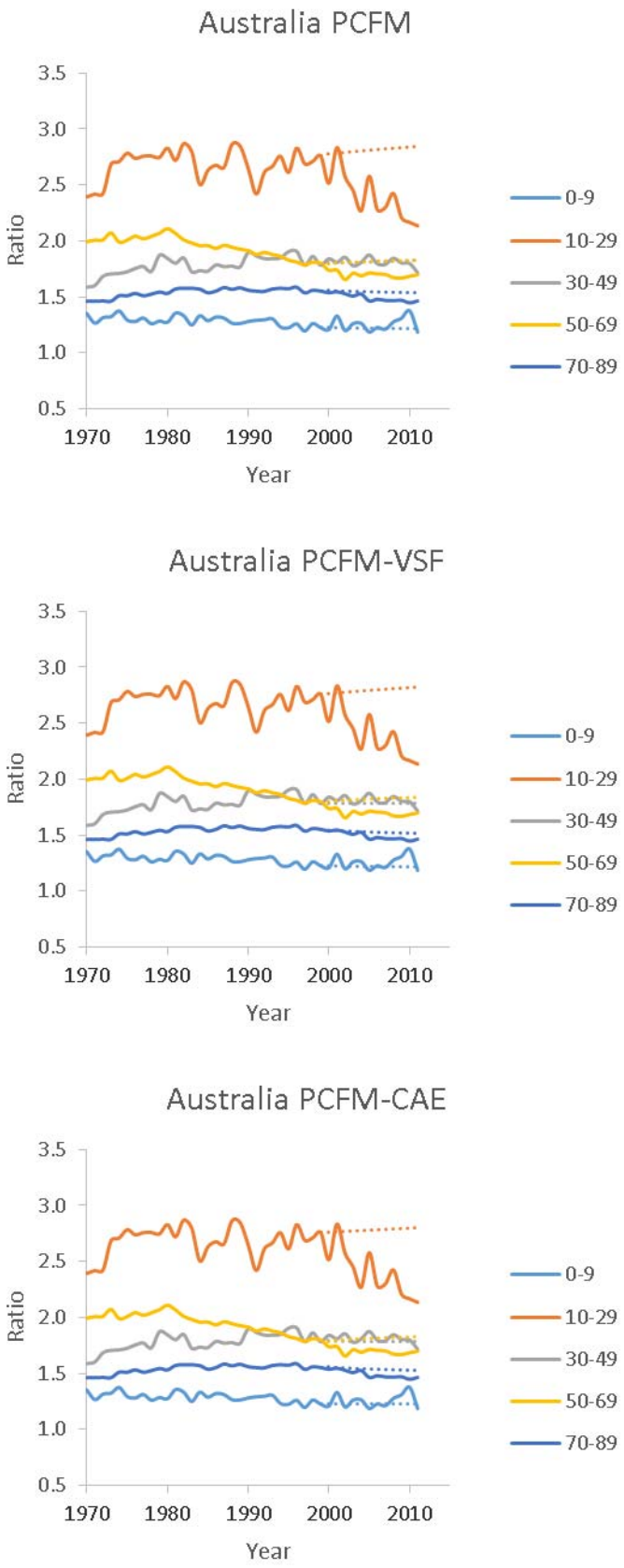


Figure 15: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, Australia. The sample period is 30 years and the projection period is 12 years.

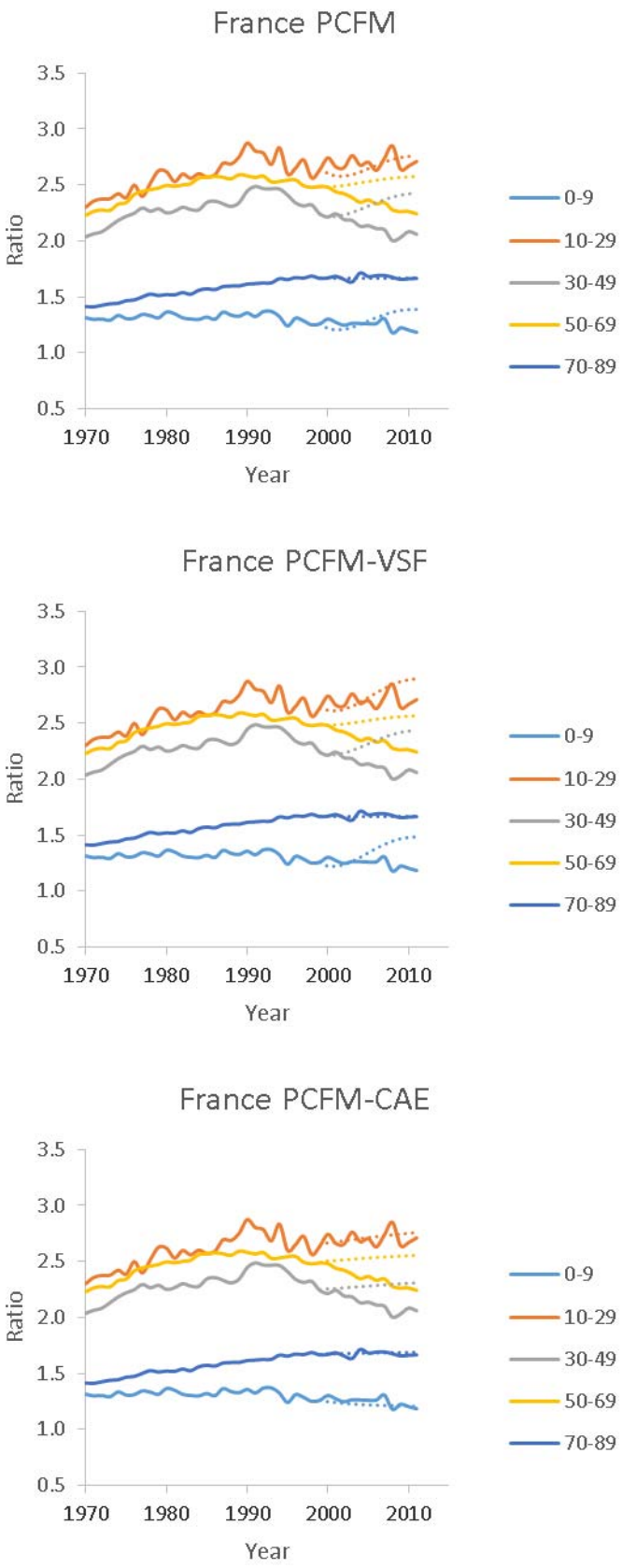


Figure 16: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, France. The sample period is 30 years and the projection period is 12 years.

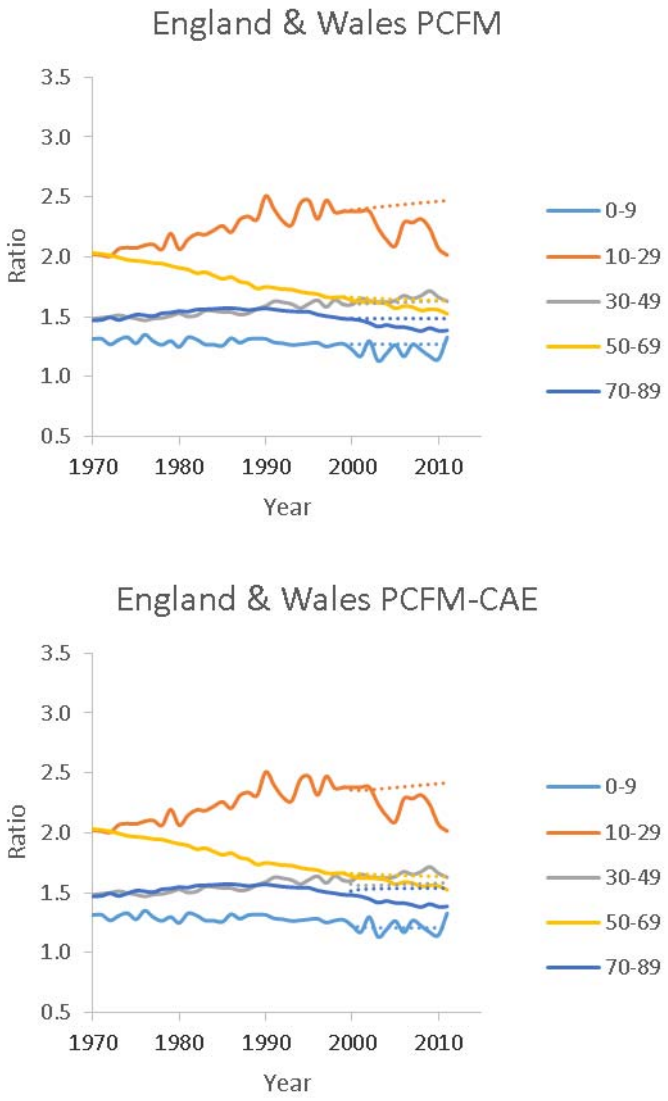


Figure 17: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, England and Wales. The sample period is 30 years and the projection period is 12 years.

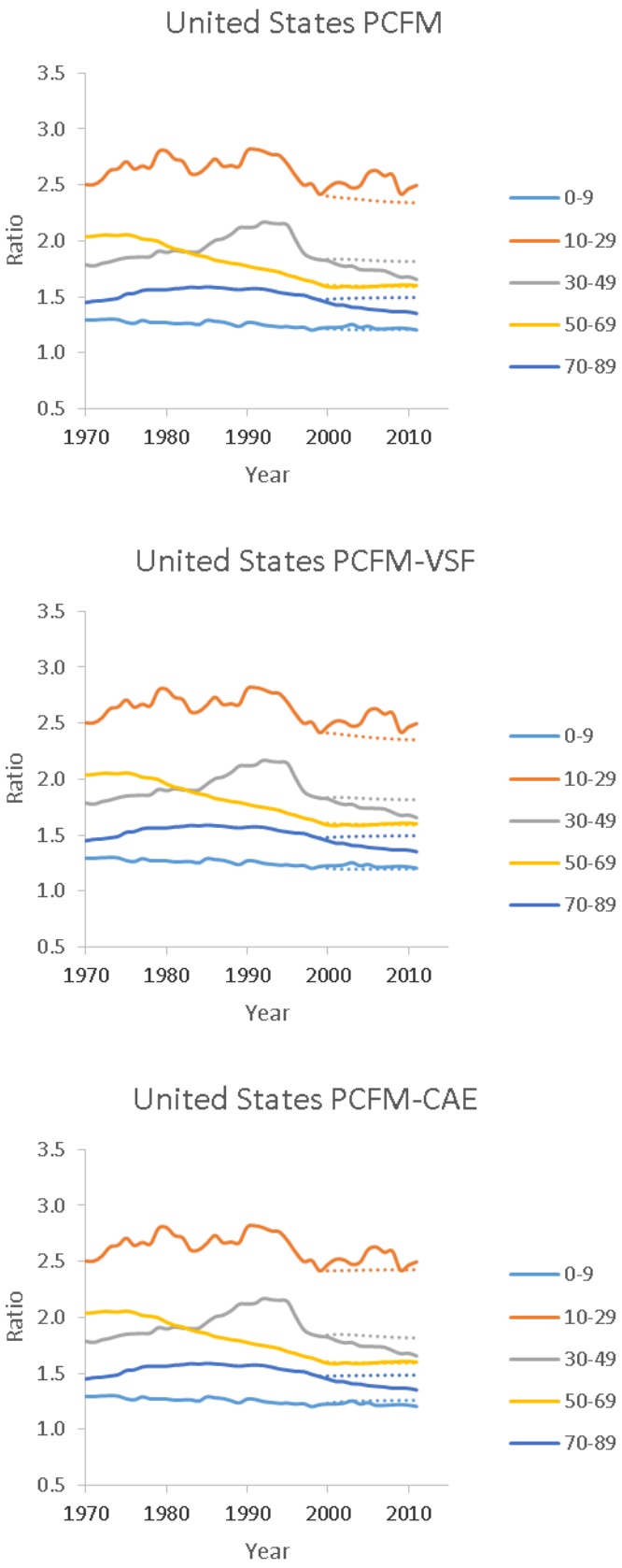


Figure 18: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, United States. The sample period is 30 years and the projection period is 12 years.

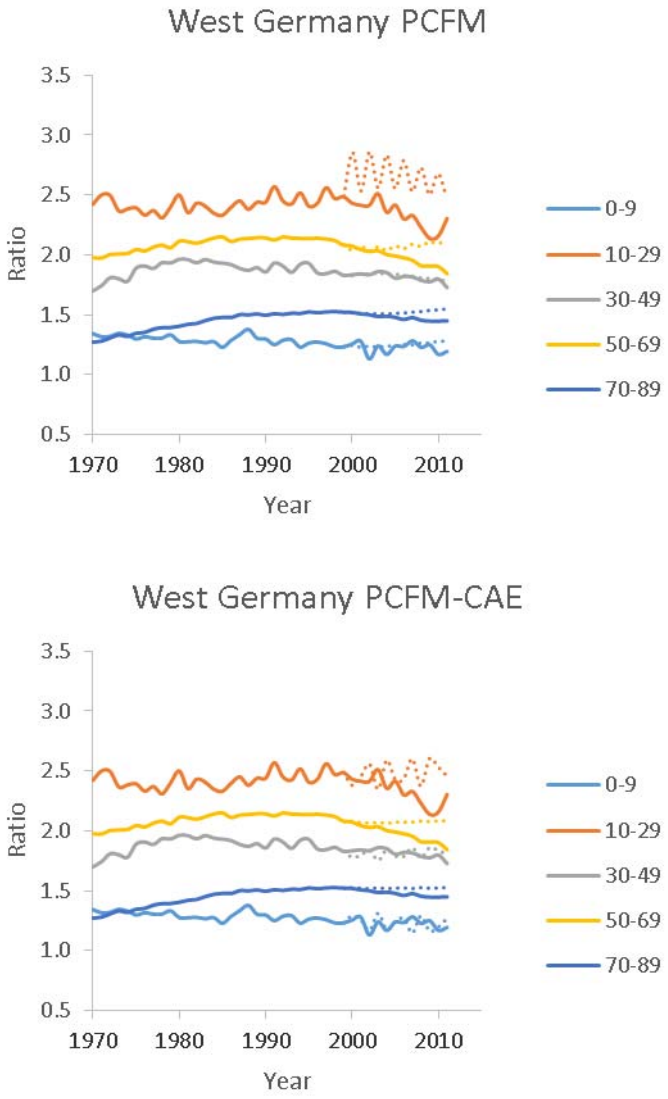


Figure 19: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, West Germany. The sample period is 30 years and the projection period is 12 years.

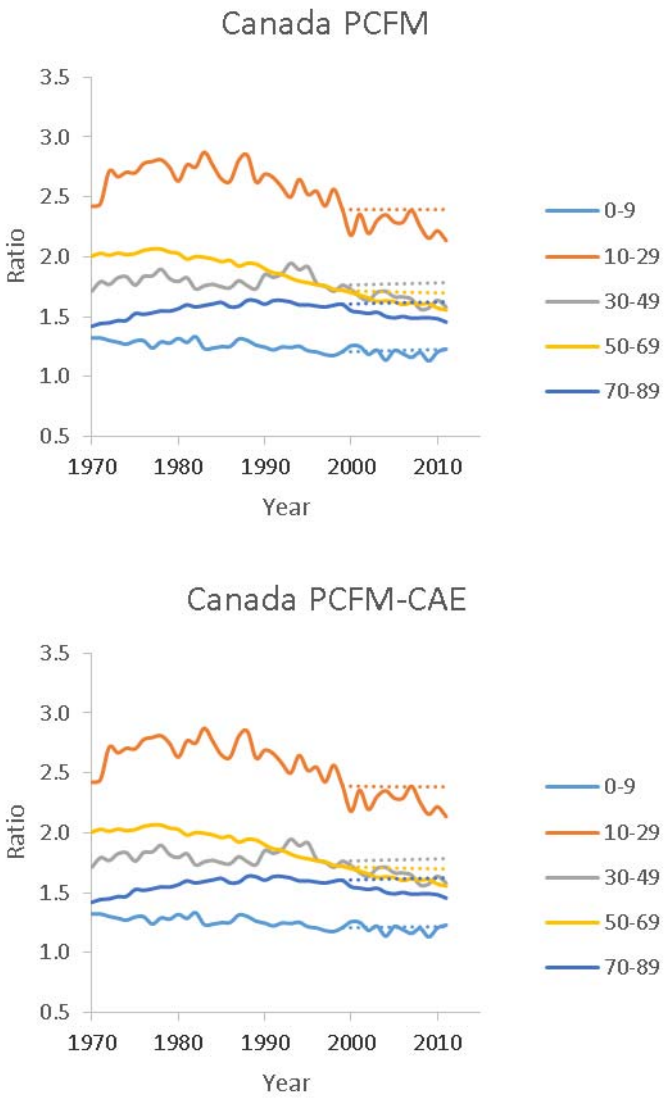


Figure 20: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, Canada. The sample period is 30 years and the projection period is 12 years.

Table 9: MAPE values from the PCFM, PCFM-VSF and PCFM-CAE

	PCFM			PCFM-VSF			PCFM-CAE		
Country	Female[a]	Male[a]	Male to female ratios[b]	Female[a]	Male[a]	Male to female ratios[b]	Female[a]	Male[a]	Male to female ratios[b]
Australia	2.14% (14.42%)	2.65% (15.93%)	8.12%	2.15% (14.47%)	2.69% (15.73%)	7.59%	2.15% (14.47%)	2.73% (15.81%)	7.51%
France	2.05% (15.52%)	2.51% (18.41%)	7.81%	2.05% (15.52%)	2.49% (18.10%)	7.08%	2.07% (15.62%)	2.42% (17.52%)	7.53%
England & Wales	1.71% (10.04%)	2.53% (13.18%)	6.46%	-	-	-	1.68% (9.96%)	2.53% (13.19%)	6.14%
United States	1.42% (7.68%)	2.21% (13.29%)	12.95%	1.41% (7.68%)	2.21% (13.22%)	12.75%	1.54% (8.41%)	2.24% (13.27%)	11.20%
West Germany	1.67% (10.60%)	1.85% (10.67%)	5.83%	-	-	-	1.62% (10.25%)	1.86% (10.79%)	6.50%
Canada	1.69% (10.77%)	2.32% (13.13%)	10.24%	-	-	-	1.69% (10.79%)	2.32% (13.13%)	10.17%

This out-of-sample analysis, applies the PCFM, PCFM-VSF (where applicable) and PCFM-CAE to the first period, projects death rates for the second period, and compares the projected values against the observed rates. The first period is 25 years and the second period is 17 years.

[a]: MAPE values are based on fitted log (actual) central death rates.

[b]: The MAPE values of the projected male-to-female ratios of death rates are based on 10-year age groups, since the ratios can be volatile for younger individual ages.

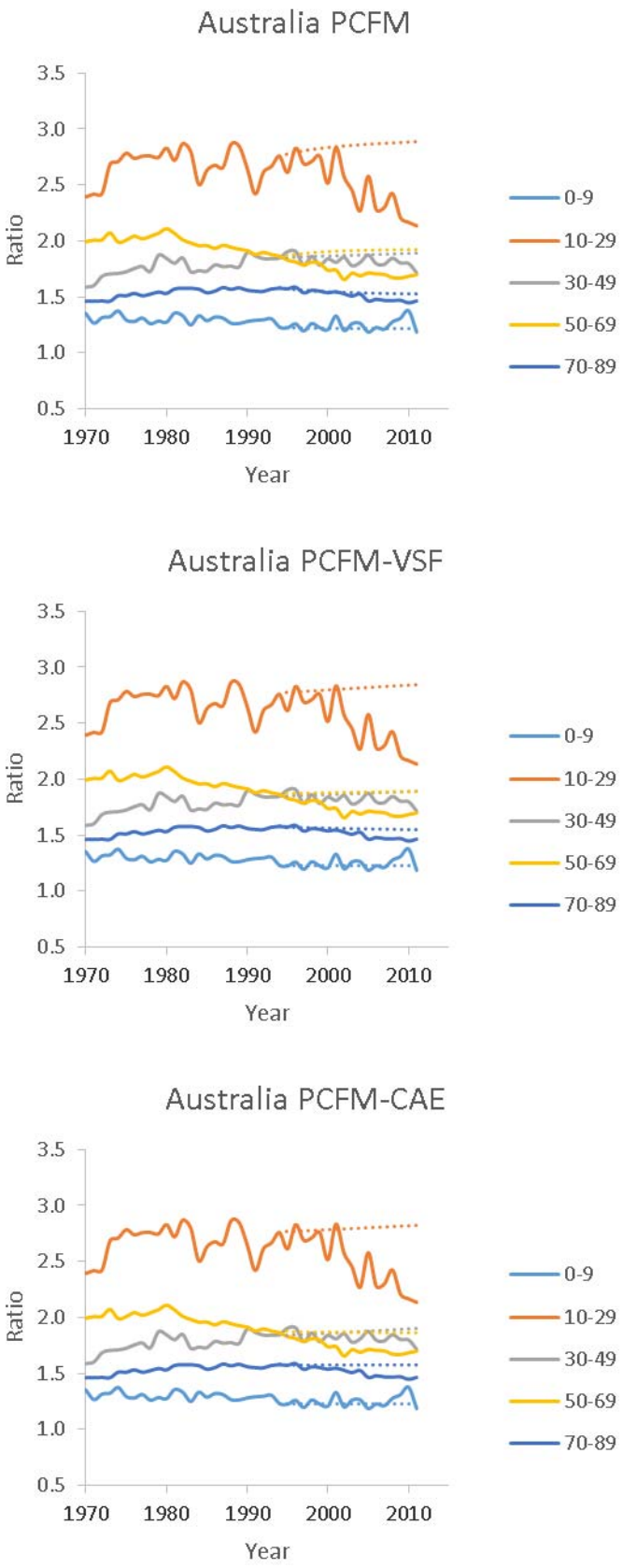


Figure 21: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, Australia. The sample period is 25 years and the projection period is 17 years.

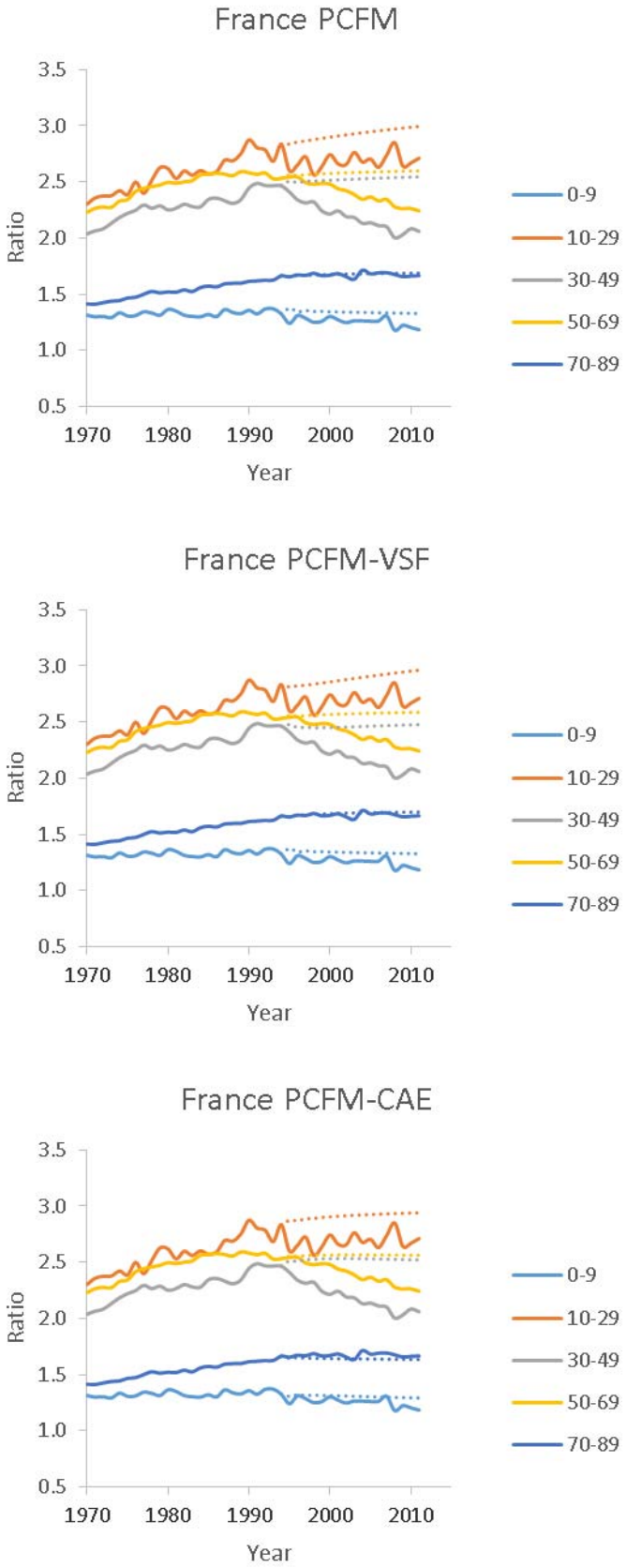


Figure 22: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, France. The sample period is 25 years and the projection period is 17 years.

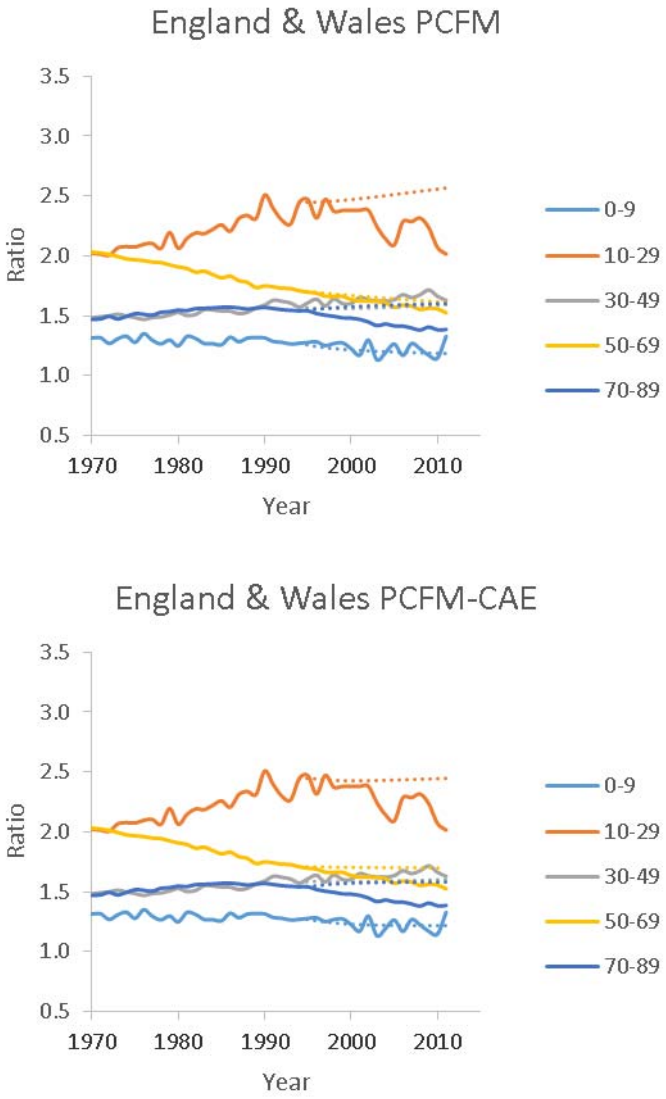
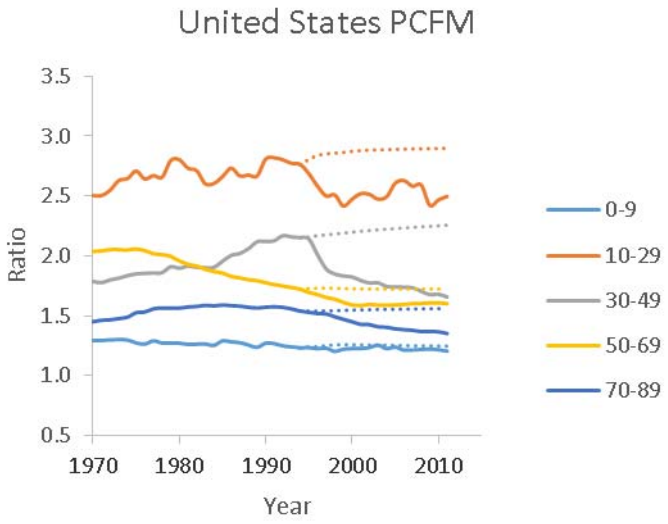
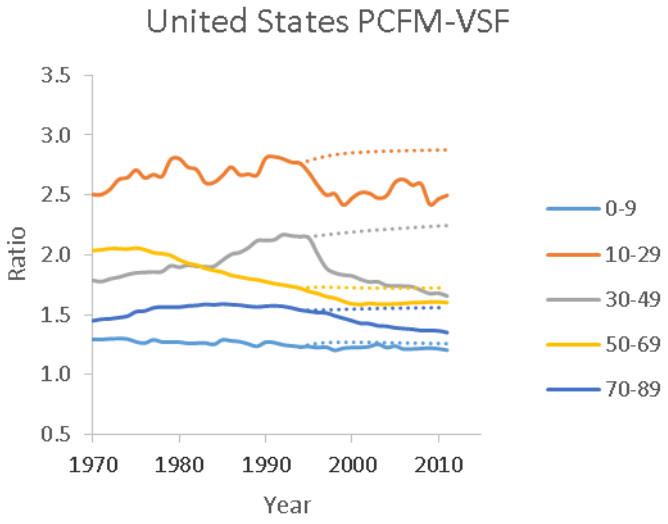


Figure 23: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, England and Wales. The sample period is 25 years and the projection period is 17 years.





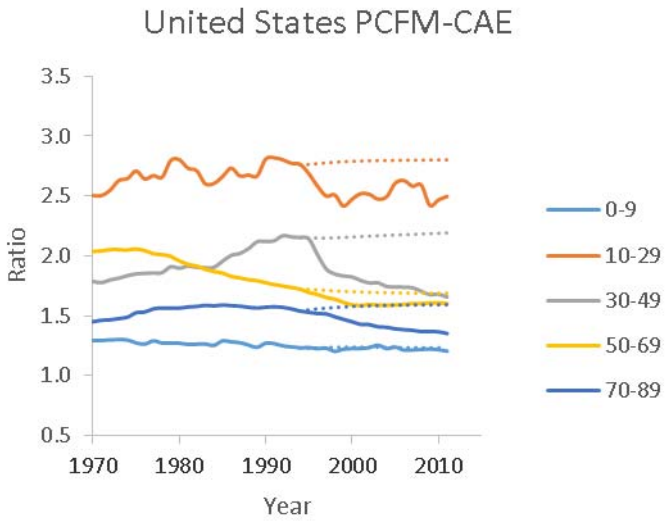


Figure 24: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, United States. The sample period is 25 years and the projection period is 17 years.

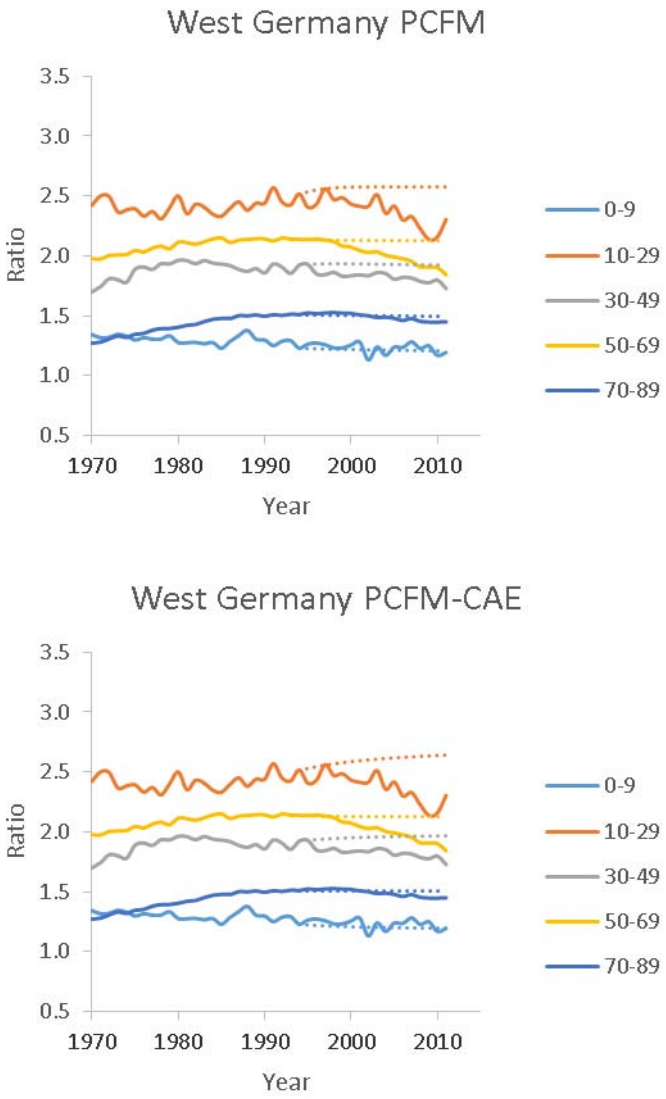


Figure 25: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, West Germany. The sample period is 25 years and the projection period is 17 years.

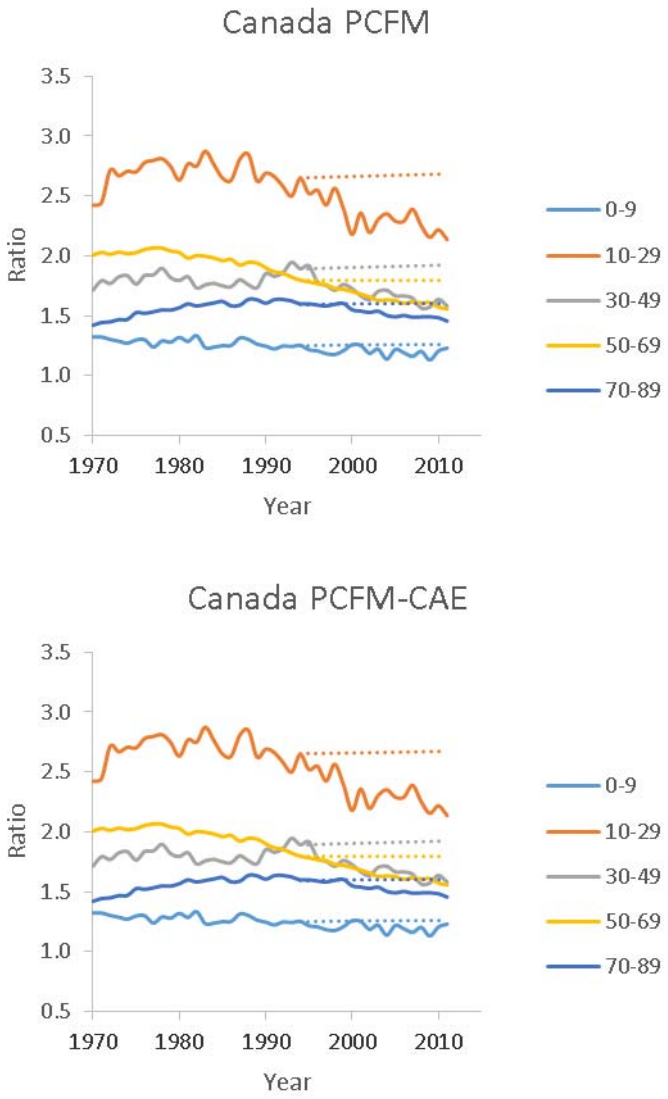


Figure 26: Observed (solid lines) and projected (dotted lines) ratios of male-to-female death rates, Canada. The sample period is 25 years and the projection period is 17 years.

6. Concluding Remarks

This thesis has suggested two modifications to the PCFM proposed by Li (2013) and evaluated the data fitting and forecasting performance when applied to multiple populations. The two modifications involve allowing for a different number of sex-specific factors, and incorporating a common age effect into the model in a similar fashion to Kleinow (2015). The results indicate that the modified PCFM-VSF and PCFM-CAE produce better fitting results compared to the base PCFM, while maintaining reasonable and accurate joint projections of male and female mortality. These new modifications offer more flexibility to deal with different features and patterns in different populations.

One potential issue for concern lies in the out-of-sample forecasting results. This thesis has shown that the time series dynamics of the PCFM family of models sometimes results in peculiar mortality projections. Further investigation into the sample period's impact on forecast accuracy is warranted. Moreover, there does not appear to be a universal relationship between the three models and the optimal choice of parameters. It would be interesting to apply these models to more population datasets and examine the results. Another limitation that needs to be addressed is the serial correlations in the data. By using the BIC, the model fitting procedure assumes that serial correlations have been adequately captured by the time series models, but analysis of the deviance residuals and the ACFs of the time series residuals demonstrate that this is not the case.

Ideas for future research include adding a cohort term to the PCFM-VSF and PCFM-CAE. Yang et al. (2016) extended the base PCFM with a cohort term and found that it reduced the need for additional period factors. The parsimony of the PCFM-VSF and PCFM-CAE could be further improved in a similar fashion. Moreover, it would be useful to apply these models to practical scenarios such as insurance pricing and government policy planning, since a realistic joint projection of females and males is very important for valuing annuities, pensions and social benefits correctly. Furthermore, a more detailed investigation of the use of time series models would be warranted. The choice of time series modelling affects not only the central projection but also the level of variability in the simulation. For insurance or demographic studies requiring probabilistic calculations, one should be cautious about the specific effects of assuming a particular time series model.

References

Alho, J. M., & Spencer, B. D. (1990). Error models for official mortality forecasts. *Journal of the American Statistical Association*, *85*(411), 609-616.

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

Booth, H., Maindonald, J., & Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, *56*(3), 325-336.

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

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.

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

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

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., 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.

Carter, L. R., & Lee, R. D. (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting*, *8*(3), 393-411.

De Jong, P., & Tickle, L. (2006). Extending Lee–Carter mortality forecasting. *Mathematical Population Studies*, *13*(1), 1-18.

Debón, A., Montes, F., & Martínez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. *European Actuarial Journal*, *1*(2), 291-308.

Delwarde, A., Denuit, M., & Eilers, P. (2007). Smoothing the Lee–Carter and Poisson log-bilinear models for mortality forecasting A penalized log-likelihood approach. *Statistical Modelling*, *7*(1), 29-48.

Human Mortality Database. (2016). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. (data downloaded on 27 Apr 2016).

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

Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, *51*(10), 4942-4956.

Keyfitz, N. (1982). Can knowledge improve forecasts?. *Population and Development Review*, 729-751.

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

Lee, R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, *4*(1), 80-91.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, *87*(419), 659-671.

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

Li, J. (2013). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies*, *67*(1), 111-126.

Li, J. S. H., Chan, W. S., & Cheung, S. H. (2011). Structural changes in the Lee-Carter mortality indexes: detection and implications. *North American Actuarial Journal*, *15*(1), 13-31.

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

Li, J., & Haberman, S. (2015). On the effectiveness of natural hedging for insurance companies and pension plans. *Insurance: Mathematics and Economics*, *61*, 286-297.

Li, J., Tickle, L., & Parr, N. (2016). A multi-population evaluation of the Poisson common factor model for projecting mortality jointly for both sexes. *Journal of Population Research*. doi:10.1007/s12546-016-9173-0

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

Parr, N., Li, J., & Tickle, L. (2016). A cost of living longer: Projections of the effects of prospective mortality improvement on economic support ratios for 14 advanced economies. *Population Studies*, *70*(2), 181-200.

R Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.

Renshaw, A. E., & Haberman, S. (2003). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, *33*(2), 255-272.

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

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

Tan, C. I., Li, J., Li, J. S. H., & Balasooriya, U. (2014). Parametric mortality indexes: From index construction to hedging strategies. *Insurance: Mathematics and Economics*, *59*, 285-299.

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

Yang, B., Li, J., & Balasooriya, U. (2016). Cohort extensions of the Poisson common factor model for modelling both genders jointly. *Scandinavian Actuarial Journal*, *2016*(2), 93-112.

Appendix

The iterative updating scheme used to estimate the parameters of the PCFM, PCFM-VSF and PCFM-CAE is designed to minimise the deviance function:

$$\text{deviance} = \sum_{x,t,i} 2\left[d_{x,t,i} \ln\left(\frac{d_{x,t,i}}{\hat{d}_{x,t,i}} \right) - d_{x,t,i} + \hat{d}_{x,t,i} \right]$$

or equivalently, maximise the log likelihood function:

$$l = \ln L = \sum_{x,t,i} d_{x,t,i} \ln(\hat{d}_{x,t,i}) - \hat{d}_{x,t,i} - \ln(d_{x,t,i}!)$$

where $d_{x,t,i}$ is the observed number of deaths at age x in year t for sex i, and $\hat{d}_{x,t,i}$ is the corresponding fitted number of deaths. For example, in the PCFM-CAE the fitted number of deaths is calculated as:

$$\begin{aligned} \hat{d}_{x,t,i} &= E_{x,t,i} \hat{m}_{x,t,i} \\ &= E_{x,t,i} \exp\left(\hat{a}_{x,i} + \hat{B}_x \hat{K}_t + \sum_{j=1}^{n} \hat{b}_{x,j} \hat{k}_{t,i,j} \right) \end{aligned}$$

The parameters are updated using the Newton-Raphson method:

$$\theta^* = \theta - \frac{\partial l / \partial \theta}{\partial^2 l / \partial \theta^2}$$

However, previous studies (Li, 2013; Yang et al., 2016) have shown that attempting to update all parameters simultaneously does not necessarily lead to an optimal solution, as the iterative procedure tends to head towards some local maxima. For practical purposes, a multi-stage estimation method is used to calculate a conditional maximum likelihood instead, as suggested in Booth et al. (2002).

Below is an illustration of the updating procedure as described in Li (2013), modified for the PCFM-CAE. First the model is fitted with the common factor only, i.e. $n = 0$. Assuming an age range of 0-89:

(1) Set up initial parameter values ($\hat{a}_{x,i}$ = mean of log death rates at age x of sex i over time, $\hat{B}_x = \hat{b}_{x,j} = 1/90$, $\hat{K}_t = \hat{k}_{t,i,j} = 0$), and calculate $\hat{d}_{x,t,i}$.

(2) Update $\hat{a}^*_{x,i} = \hat{a}_{x,i} + \sum_t (d_{x,t,i} - \hat{d}_{x,t,i}) / \sum_t \hat{d}_{x,t,i}$ for all x and i, and then recalculate all $\hat{d}_{x,t,i}$.

(3) Update $\hat{K}^*_t = \hat{K}_t + \sum_{x,i} (d_{x,t,i} - \hat{d}_{x,t,i}) \hat{B}_x / \sum_{x,i} \hat{d}_{x,t,i} \hat{B}^2_x$ for all t, adjusted with $\sum_t \hat{K}^*_t = 0$, and then recalculate all $\hat{d}_{x,t,i}$.

(4) Update $\hat{B}^*_x = \hat{B}_x + \sum_{t,i} (d_{x,t,i} - \hat{d}_{x,t,i}) \hat{K}_t / \sum_{t,i} \hat{d}_{x,t,i} \hat{K}^2_t$ for all x, and then recalculate all $\hat{d}_{x,t,i}$.

(5) For final parameter estimates, divide $\hat{B}_x$ by $\sum_x \hat{B}_x$ and multiply $\hat{K}_t$ by $\sum_x \hat{B}_x$.

(6) Compute the log-likelihood function or deviance function.

(7) Repeat (2) - (6) until the log-likelihood or deviance converges.

Now, treating the estimated parameters as given, the additional sex-specific factors are added one at a time and the new parameters are estimated as below:

(8) Update $\hat{k}^*_{t,i,j} = \hat{k}_{t,i,j} + \sum_x (d_{x,t,i} - \hat{d}_{x,t,i}) \hat{b}_{x,j} / \sum_x \hat{d}_{x,t,i} \hat{b}^2_{x,j}$ for all t, all i, and j = 1, adjusted with $\sum_t \hat{k}^*_{t,i,j} = 0$, and then recalculate all $\hat{d}_{x,t,i}$.

(9) Update $\hat{b}^*_{x,j} = \hat{b}_{x,j} + \sum_{t,i} (d_{x,t,i} - \hat{d}_{x,t,i}) \hat{k}_{t,i,j} / \sum_{t,i} \hat{d}_{x,t,i} \hat{k}^2_{t,i,j}$ for all x, and j = 1, and then recalculate all $\hat{d}_{x,t,i}$.

(10) Similarly to (5), divide $\hat{b}_{x,j}$ by $\sum_x \hat{b}_{x,j}$ and multiply $\hat{k}_{t,i,j}$ by $\sum_x \hat{b}_{x,j}$.

(11) Compute the log-likelihood function or deviance function.

(12) Repeat (8) – (11) until the log-likelihood or deviance converges.

(13) Treating the newly estimated $\hat{b}_{x,j}$ and $\hat{k}_{t,i,j}$ as given as well, repeat (8) – (12) for the other values of j.

This conditional maximum likelihood approach offers a more convenient modelling strategy, and has no problems with convergence of the parameters or the log-likelihood/deviance. The procedure is easily tweaked for the PCFM-VSF as well, by replacing $\hat{b}_{x,j}$ with $\hat{b}_{x,i,j}$ and omitting the relevant parts of (8) and (9) once all of one sex's factors have been added.

The standardised deviance residuals are calculated as in Li (2013):

$$r_{x,t,i} = \text{sign}(d_{x,t,i} - \hat{d}_{x,t,i})\sqrt{\frac{2(d_{x,t,i}/\hat{d}_{x,t,i}) - d_{x,t,i} + \hat{d}_{x,t,i})}{\hat{\phi}}}$$

where n_p is the effective number of parameters being estimated, n_d is the number of observations and the dispersion parameter $\hat{\phi} = \dfrac{\text{deviance}}{n_d - n_p}$.

The tables below illustrate the ACFs of the residuals of the selected time series processes for model projection. The symbols +, - and * are used to indicate whether the sample ACF value at a certain lag is larger than twice the estimated standard error, smaller than negative twice the estimated standard error, or is statistically insignificant. RW is used if the time series was modelled as a random walk without drift instead. Although there are some significant values in the chosen time series models, the alternative choices are non-stationary or otherwise unsuitable for projection purposes.

Australia PCFM								
Lag	1	2	3	4	5	6	7	8
kf1	RW							
kf2	*	+	+	*	*	*	*	*
km1	*	*	*	+	*	*	*	*
km2	*	*	*	*	*	*	*	*

Australia PCFM-VSF								
Lag	1	2	3	4	5	6	7	8
kf1	RW							
km1	*	*	*	+	*	*	*	*
km2	*	*	*	*	*	*	*	*

Australia PCFM-CAE								
Lag	1	2	3	4	5	6	7	8
kf1	-	*	*	+	*	*	*	*
kf2	*	*	*	*	*	*	*	*
km1	*	*	*	+	*	*	*	*
km2	*	*	*	*	*	*	*	*

France PCFM								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	*	*	*	*
kf2	*	+	*	*	*	*	*	*
kf3	*	*	*	*	*	*	*	*
kf4	*	+	+	*	*	*	*	*
km1	*	*	*	*	*	*	*	*
km2	*	+	*	*	*	*	*	*
km3	*	*	*	*	+	*	*	*
km4	*	*	+	*	*	*	*	*

France PCFM-VSF								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	*	*	*	*
kf2	*	+	*	*	*	*	*	*
kf3	*	*	*	*	*	*	*	*
km1	*	*	*	*	*	*	*	*
km2	*	+	*	*	*	*	*	*
km3	*	*	*	*	+	*	*	*
km4	*	*	+	*	*	*	*	*

France PCFM-CAE								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	*	*	*	*
kf2	*	*	*	*	*	*	*	*
kf3	*	+	*	*	*	*	*	*
kf4	*	*	*	*	*	*	*	*
kf5	*	*	*	-	*	*	*	*
km1	*	*	+	*	*	*	*	*
km2	*	*	*	*	*	*	*	*
km3	*	*	*	*	*	*	*	*
km4	*	*	*	*	+	*	*	*
km5	*	*	+	*	*	*	*	*

England & Wales PCFM								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	*	*	*	*
kf2	*	+	*	*	*	*	*	*
kf3	*	*	+	*	*	*	*	*
km1	RW							
km2	*	*	*	+	*	+	*	*
km3	*	*	*	*	*	*	*	*

England & Wales PCFM-CAE								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	*	*	*	*
kf2	RW							
kf3	*	*	*	*	*	*	*	*
kf4	*	*	*	*	*	*	*	*
km1	RW							
km2	RW							
km3	*	*	*	+	*	*	*	*
km4	*	*	*	+	*	*	*	*

United States PCFM								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	-	*	*	*
kf2	*	*	*	*	*	*	*	*
kf3	*	+	*	*	*	*	*	*
kf4	*	*	*	*	*	*	*	*
kf5	*	*	*	*	*	*	*	*
km1	*	*	*	*	*	*	*	*
km2	*	*	*	*	*	*	*	*
km3	*	*	*	*	*	*	*	*
km4	*	*	*	*	*	*	*	*
km5	*	*	*	*	*	*	*	*

United States PCFM-VSF								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	-	*	*	*
kf2	*	*	*	*	*	*	*	*
kf3	*	+	*	*	*	*	*	*
kf4	*	*	*	*	*	*	*	*
kf5	*	*	*	*	*	*	*	*
km1	*	*	*	*	*	*	*	*
km2	*	*	*	*	*	*	*	*
km3	*	*	*	*	*	*	*	*
km4	*	*	*	*	*	*	*	*
km5	*	*	*	*	*	*	*	*
km6	*	*	*	*	*	*	*	*

United States PCFM-CAE								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	*	*	*	*
kf2	*	*	*	*	*	*	*	*
kf3	*	*	*	*	*	*	*	*
kf4	*	+	*	*	*	*	*	*
kf5	*	*	*	*	*	*	*	*
kf6	*	*	*	*	*	-	*	*
km1	RW							
km2	*	*	*	*	*	*	*	*
km3	*	*	*	*	*	*	*	*
km4	*	*	*	*	*	*	*	*
km5	*	*	*	*	*	*	*	*
km6	*	*	*	*	*	*	*	*

West Germany PCFM								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	*	*	*	*
kf2	RW							
kf3	*	*	*	*	+	*	*	*
kf4	*	*	+	-	+	*	*	*
km1	*	*	+	*	+	*	*	*
km2	RW							
km3	*	*	*	*	*	*	*	*
km4	*	*	*	*	*	*	*	*

West Germany PCFM-CAE								
Lag	1	2	3	4	5	6	7	8
kf1	*	*	*	*	*	*	*	*
kf2	RW							
kf3	*	*	+	*	+	*	*	*
kf4	-	*	*	*	+	-	+	*
kf5	*	*	*	*	*	-	*	*
kf6	*	*	*	*	*	*	*	*
km1	*	*	+	*	+	*	*	*
km2	*	*	*	*	*	*	*	*
km3	*	*	*	*	*	*	*	*
km4	*	*	*	*	*	-	*	*
km5	*	*	*	*	+	*	*	*
km6	*	*	*	*	*	*	*	*

Canada PCFM								
Lag	1	2	3	4	5	6	7	8
kf1	RW	*	*	*	*	*	*	*
kf2	*	*	*	*	*	*	*	*
km1	*	*	*	*	*	*	*	*
km2	*	*	+	*	+	*	*	*

Canada PCFM-CAE								
Lag	1	2	3	4	5	6	7	8
kf1	*	+	*	*	*	*	*	*
kf2	RW							
kf3	*	*	+	*	*	*	*	*
km1	*	*	*	*	*	*	*	*
km2	*	*	+	*	*	*	*	*
km3	*	*	*	*	*	*	*	*