ChenY.pdf

by Yongqing Chen

Joint Mortality Modelling and Projection of six Subpopulations in the UK under a 2-tier Coherent Framework

This dissertation is submitted as part of the requirements for the award of MSc in Actuarial Management at Cass Business School, City University London.

Submitted by: Yongqing (Ree) Chen; SID: 120057586
Supervised by: Dr. Pietro Millossovich

Acknowledgements

I would like to express my gratitude to my supervisor, Dr Pietro Millossovich for his generous guidance throughout this project, especially his constructive advice and great insights in the area of mortality projection. I want to thank all the staff in the Actuarial Management programme at Cass Business School, especially Douglas Wright, for teaching me very valuable knowledge throughout this year in the Actuarial Risk Management CA1 and Life Insurance ST2 subjects. I want to thank Jerry Song from Imperial College for peer reviewing my work thoroughly. I also would like to express my gratitude to my parents and my boyfriend Yiwang Xu for supporting and encouraging me during this year at Cass. Finally, I would like to thank an Australian couple, Leonie Tickle and Brad Louis, who are both actuaries, for their kind guidance and for making me a better and happier person.

Abstract

This work introduces a 2-tier Augmented Common Factor model (2-tier ACF) and applies it to the joint projection of United Kingdom mortality rates for two genders and three countries (England and Wales combined, Scotland, and Northern Ireland). The model is extended from the classic Lee-Carter (LC) model, with a common factor for the whole UK population, a sex specific factor for males and females, and a sex-country specific factor for each country within each gender. A Poisson framework is used, as death counts in each gender-country subpopulation are modelled directly. Our results show that the 2-tier ACF model improves the model fitting against past experience compared to the independent LC model fitted to each subpopulation of the UK. Mortality projection results also show that the 2-tier ACF model can produce coherent results for different genders within each country and different countries within each gender, which avoids the divergence issues in the independent LC projections. The 2-tier ACF is further extended to include a cohort factor, which takes into account the cohort effect of the UK and further improves the model fitting and randomness of residuals. The limitations of the 2-tier ACF and its application in the insurance industry and pension funds are also discussed.

Table of Contents

List of Tables

List of Figures

1. Introduction

Heterogeneity of mortality within the population has long been an area of interest for life insurance and pension companies (Su & Sherris 2012). Apart from age and gender being the traditional rating factors for life insurance and pension products, the social-economic differences between the three countries (England & Wales combined, Scotland, and Northern Ireland)[1] have led to notably different demographic trends at least in the short term while pertaining to the same larger population. This project aims to introduce a model that is capable of modelling and producing the mortality trends for all six subpopulations of the UK (two genders and three countries) jointly and coherently, and discuss its implication for life insurers and pension funds in the UK.

The last three decades have witnessed tremendous development in the area of mortality modelling and projection. This includes the Lee-Carter model (LC) proposed by Lee & Carter (1992), which is the most classic mortality forecast model due to its simplicity and ability to produce linear mortality index for many different countries. Over time, various extensions and variants of the basic LC model have been published (e.g. Lee & Miller 2001; Booth, Maindonald and Smith, 2002; Renshaw & Haberman 2003). All these models have their primary focus on a single population. When these models are used independently in modelling multiple subpopulations with similar demographic trends, such as different genders within a population, or different geographical areas within a population, the assumed independence would generally lead to divergence in forecasting results.

Diverging trend over time for closely related subpopulations is not a desirable outcome. For example, due to genetic and biological reasons, male mortality rates have constantly been observed to be higher than females (Kalben 2002). However, if male mortality improvements are faster than female and the two genders are projected independently, a model may forecast male mortality rates lower than the

[1] Please note that England and Wales are modelled as one country in this project, because the

corresponding female rates. As noted in Section 5.3 of working paper 15 of the Continuous Mortality Investigation Mortality Committe (2005), independent projection methodologies have to be adjusted to ensure divergence does not happen. It is also intuitively true that mortality rates of populations that are geographically close or politically related, are driven by a common set of factors such as social-economic conditions, health and care system, and the general environment. Therefore, non-divergent or "coherent" models are sought to address the issue of divergence. The augmented common factor model (ACF) suggested by Li & Lee (2005) is an important step to produce one model that captures both the short-term divergence and long-term coherence among related populations (subpopulations). The ACF (or 1-tier ACF) uses a common factor to depict the long-term common trend of the total population, with additional factors included to capture the short-term discrepancy from the common trend for each subpopulation.

The focus of this project is to introduce a new 2nd-tier extension (or a second dimension) to the ACF model - a common factor is used to model trend for the aggregated UK population, a sex specific factor to capture the discrepancy between each gender and the total population, and a country-sex specific factor to capture the discrepancy in mortality of a gender in a specific country from the overall trend of that gender. This is to ensure that coherence of forecasts is achieved in both dimensions – mortality of different genders within each country and mortality of different countries within each gender.

The 2-tier ACF model and the independent LC model are applied to the six subpopulations in the UK. In the LC model, the six subpopulations are modelled and projected independently. Both models are fitted to the period from 1975 to 2000, and various model criteria are compared. Out-of-sample forecast is performed for the period from 2001 to 2011, and the projected mortality is compared against actual observations to understand the level of accuracy of both models. Following the approach of Booth et al. (2005), the prediction accuracy of the two models is compared based on Mean Error and Mean Absolute Error in log-scale mortality rates and life expectancy.

To understand the long-term implications of the model, both the 2-tier ACF and the LC are then fitted to the period from 1975 to 2011, and projected for the period from 2012 to 2050, which is broadly consistent with the common approach that the length of fitting period is similar to the projection period. The projected mortality rates by age for the six subpopulations up to year 2050 are compared between the two models and the superiority of the 2-tier ACF over the independent LC forecasts is highlighted. For a multi-population model, coherence has been defined by Hyndman et al. (2013) as the convergence to reasonable constants of the ratios of age-specific mortality rates between any pair of subpopulations. As discussed later on, the proposed 2-tier ACF model satisfies this definition of coherence.

In the original construction of the models, both the original LC model and the ACF model are estimated by applying singular value decomposition (SVD) to the logarithms of mortality rates. In the ACF model, Li & Lee (2005) firstly apply SVD to the aggregate data to fit the common factor, and then fit the parameters of additional factors by applying SVD to the residuals. The SVD produces estimates in line with the method of least squares and the only assumption necessary under this method is that errors are normally distributed and homoscedastic. However, as pointed out by Brouhns et al. (2002), the model does not take into account the fact that the scarcity of number of deaths at older ages make mortality rates much more volatile. Brouhns et al. (2002) proposed a Poisson version of the Lee-Carter model, by specifying the death counts as Poisson variables and using maximum likelihood estimation. Li (2012) applied this Poisson framework to the ACF model and the Australian population with the addition of multiple sex-specific factors, and named this model as the Poisson Common Factor Model (PCFM). The Poisson framework is similar to the generalised linear model with a log link function; however the bilinear terms have to be estimated by minimising the deviance of a non-linear model structure through iterations. This technique has been developed into R packages such as StMoMo by Villegas et al. (2015) and Iterative Lee-Carter Package (Butt & Haberman 2009). In this project, the Poisson framework proposed by Brouhns et al. (2002) is applied to all the models, so that they are compared like-for-like under the robust statistical framework of the Poisson distribution.

When applying mortality models to the UK population, the research community has long recognised the fact that UK mortality in the past century does not only depend on age and calendar year, but also on the year of birth, i.e. the cohort. Willets (2004) explained the cohort effect with the prevalence of smoking, some major fatal diseases and social-economic classes. Renshaw & Haberman (2006) introduced the cohort extensions to the original LC model under both Gaussian and Poisson framework. Yang et al. (2014) introduced six different possible cohort extensions to the PCFM, which is an extension of the ACF under the Poisson framework. In this project, we also extend both the 1-tier and 2-tier ACF models by including a cohort factor to allow for the cohort effect of the UK population.

In Section 2, the details of models and methods are described including the basic LC model and the ACF model, and in particular the new 2-tier ACF model proposed by this project will be introduced. In Section 3.1, we compare the 2-tier ACF and the independent LC models when fitted to the six subpopulations in the UK during the period between 1975 and 2000, and produce out-of-sample forecasts for the period between 2001 and 2011. Section 3.2 further compares the long-term converging or diverging behaviours between the 2-tier ACF and the LC by projecting mortality of the six subpopulations up to 2050. In Section 4, cohort extensions to both the 1-tier and 2-tier ACF models are introduced, and results are critically appraised. The final section of this paper discusses the potential applications of the model in the UK life insurance and pension industry and also points out some limitations of the 2-tier ACF framework and its cohort extension.

The project is entirely based on open-source data from the Human Mortality Database, and it is noted that due to the high volatility of mortality rates at the very old age, ages above 100 have been excluded from analysis. The software R is used to perform all the analysis. In this project, the author is indebted to learning from previous involvement in the research project where the PCFM was fitted to OECD countries (Parr et al. 2014). In particular, when developing the new algorithms for this project, the author is indebted to previous guidance from Jackie Li on how to fit parameters in the PCFM.

2. Models and Methods

2.1 The Basic Lee Carter (LC) Model

The Lee-Carter model (1992) is defined as below. Define the central rate of mortality at age *x* and year *t* as $m_{x,t}$; the LC model is represented as:

$$ln(m_{x,t}) = a_x + b_x\, k_t + \varepsilon_{x,t} \ , \qquad (1)$$

where a_x represents the level of mortality at age *x*, k_t is an index of the mortality level at time *t*, b_x represents the relative speed of mortality decrease at age x, and $\varepsilon_{x,t}$ represents an error term that is Gaussian distributed with mean zero and variance σ_ϵ^2.

Two constraints $\sum_x b_x = 1$ and $\sum_t k_t = 0$ are imposed to ensure the identifiability of the model. Under these constraints, a_x can be calculated as the average of $ln(m_x)$ over different *t*. Lee and Carter originally proposed to use Singular Value Decomposition (SVD) to estimate b_x and k_t, and there is a second stage estimation to re-fit k_t in order to reconcile total observed death at time *t*. The term k_t is then extrapolated using a random walk time series.

$$k_t = k_{t-1} + d + e_t \ , \qquad (2)$$

where d is the drift term representing the annual change in k_t and e_t are uncorrelated normally distributed terms. Apart from its simplicity, the LC model is popular also because when k_t's are fitted to historical data of many different countries, an obvious linear trend can be observed.

2.2 The Augmented Common Factor Model

The augmented common factor (ACF) model, also known as 1-tier ACF in this study, proposed by Li & Lee (2005) is described below:

$$ln(m_{x,t,i}) = a_{x,i} + B_x\, K_t + \ b_{x,i}\, k_{t,i} + \varepsilon_{x,t,i}\ , \tag{3}$$

where $m_{x,t,i}$ is the central mortality rate in year t at age x for gender i, $B_x\, K_t$ is the common factor for the aggregated population including both genders, $b_{x,i}\, k_{t,i}$ is the sex-specific factor for gender i, and $\varepsilon_{x,t,i}$ is the normally distributed error term. The term K_t is designed to capture the overall trend of the aggregated population over time, and B_x measures the sensitivity to decrease in mortality at age x. The fact that subpopulations share the same component $B_x\, K_t$ forms a necessary and sufficient condition to avoid divergence in central forecast of subpopulations (Debón et al. 2011). Similarly, $k_{t,i}$ is the mortality time index of a specific gender, and $b_{x,i}$ is the age sensitivity factor. The component $b_{x,i}\, k_{t,i}$ hence captures the trend in mortality of the specific gender i on top of the overall trend of the aggregated population.

The constraints $\sum_x B_x = 1$, $\sum_t K_t = 0$, $\sum_x b_{x,i} = 1$ and $\sum_t k_{t,i} = 0$ for each i are used to ensure the identifiability of the model, and $a_{x,i}$ are calculated by averaging $ln(m_{x,i})$ over time. In the original model proposed by Li and Lee, the estimation is done in a similar fashion to the LC model. SVD is applied to the total population to estimate B_x and K_t, and then $b_{x,i}$ and $k_{t,i}$ are estimated by applying SVD to the residuals $\{ln(m_{x,t,i}) - (a_{x,i} + B_x\, K_t)\}$.

For the common factor time index K_t, a random walk time series is fitted to the model and used to extrapolate future forecasts similar to the LC model as below

$$K_t = K_{t-1} + d + e_t\ . \tag{4}$$

For $k_{t,i}$, independent auto-regressive AR(1) models are fitted to the two genders. When $\alpha_{1,i}$ below has absolute value less than 1, the process is weakly stationary, and $k_{t,i}$ will converge. In this case the expected value of male-to-female death ratio will converge to a constant over time. The term $z_{t,i}$ is the normally distributed error term for the AR(1) process, which is independent of $\varepsilon_{x,t,i}$ and e_t above, and also independent of other $z_{t,j}$ when $j \neq i$,

$$k_{t,i} = \alpha_{0,i} + \alpha_{1,i} k_{t-1,i} + z_{t,i} \ . \tag{5}$$

Li (2012) further extends the ACF model (3) so that multiple sex-specific factors $b_{x,i}\, k_{t,i}$ can be included. It was also proposed that $k_{t,i}$ can be fitted by AR(p). However, for simplicity purpose, such possibility is not explored in this project. In some situations, it is impossible to fit a stationary AR(1) model, so $k_{t,i}$ should be extrapolated as a random walk without drift, which takes the form of (4) but with d equals 0 (Li & Lee 2005).

2.3 The Poisson Framework

As suggested earlier on, the original Lee-Carter (1992) and Li & Lee (2005) models share a major drawback - they assume the error terms are normally distributed and homoscedastic, which is unrealistic as the volatility of mortality is much higher at older ages. The Poisson framework introduced by Brouhns et al. (2002) does not model the logarithms of mortality rates directly, but models the number of deaths as a Poisson variable instead. As pointed out by Brouhns et al. (2002), Li (2012), and various others, the Poisson choice provides a solid statistical framework where the estimates can be based on maximum likelihood methods and information criteria can also be allowed.

Assuming constant force of mortality $\mu_{x+s,t+u,i}$ for $0 \leq u, s < 1$, we can model number of deaths at age x, time t, subpopulation i, $D_{x,t,i}$ as:

$$D_{x,t,i} \sim Poisson\ (E_{x,t,i}\, m_{x,t,i})\ , \tag{6}$$

where $E_{x,t,i}$ is the exposure corresponding to the same age, period and subpopulation. The algorithm of iteratively updating parameters under a Poisson Lee-Carter model is clearly illustrated by Brouhns et al. (2002) while Li (2012) applied the algorithm to the ACF model. For both models, iterations are performed until deviance is minimised. The deviance for the Poisson Lee-Carter model was given by Renshaw & Haberman (2003) and was extended to the ACF by Li (2012) as below:

$$deviance = \sum_{x,t,i} 2\left[d_{x,t,i} ln\left(\frac{d_{x,t,i}}{\hat{d}_{x,t,i}}\right) - d_{x,t,i} + \hat{d}_{x,t,i}\right], \quad (7)$$

where $d_{x,t,i}$ is the observed number of deaths, and $\hat{d}_{x,t,i}$ is the fitted number of deaths.

To compare like-for-like, this project uses a Poisson framework for all models to compare various fitting criteria of models under the robust statistical framework of Poisson distribution. The results produced by the LC model can be validated using the StMoMo R package, which automates the fitting and forecasting of the LC model under Poisson framework.

2.4 The 2-Tier Augmented Common Factor Model

In this section, we introduce a new second-tier extension to the above ACF model (2-tier ACF) to include a second additional factor for each specific country within each gender, so that a two dimensional framework is achieved when modelling subpopulations of different sex and countries jointly. The Poisson framework is used:

$$D_{x,t,i,j} \sim Poisson\ (E_{x,t,i,j}\ m_{x,t,i,j}), \quad (8)$$

$$ln(m_{x,t,i,j}) = a_{x,i,j} + B_x\ K_t + \ b_{x,i}\ k_{t,i} + b_{x,i,j}\ k_{t,i,j}, \quad (9)$$

where $D_{x,t,i,j}$, $E_{x,t,i,j}$ and $m_{x,t,i,j}$ represent the death counts, exposure and central mortality rates respectively at age *x*, time *t*, for the i^{th} gender and j^{th} country. The term $B_x\ K_t$ describes the general trend and random fluctuation for the whole population, and the term $b_{x,i}\ k_{t,i}$ depicts the trend of each gender departing away from the total population. The meaning of B_x, K_t, $b_{x,i}$, and $k_{t,i}$ is the same as in Section 2.2 above. The extended factor $k_{t,i,j}$ captures the mortality index of in country *j* gender *i* on top of the trends allowed for by K_t and $k_{t,i}$; $b_{x,i,j}$ is the sensitivity to $k_{t,i,j}$ at age *x* in the subpopulation. Therefore, $b_{x,i,j}\ k_{t,i,j}$ is used to capture the trend and random fluctuation specific to the subpopulation (gender i and country j) on top of overall trends for that gender.

Another way to look at this model is that when we only consider the mortality of one country, that is, when j is fixed in (9), the 2-tier ACF model reduces to the ACF model. The third bilinear term can be left out, as there is no idiosyncratic country-specific trend when the scope of modelling is only one country. Hence the original 1-tier ACF model can be viewed as a special case of the 2-tier ACF model.

To ensure the identifiability of the model, apart from the constraints of the ACF model, we also have to restrict $\sum_x b_{x,i,j} = 1$ and $\sum_t k_{t,i,j} = 0$ for each i and j.

For the extrapolation of common factor K_t and gender specific factor $k_{t,i}$, (4) and (5) above are used. For the extrapolation of $k_{t,i,j}$, there is no reason to exclude the possibility of fitting higher order ARIMA models, but for simplicity purpose, this project still uses weakly stationary AR(1) model or random walk without drift. The terms $\alpha_{0,i,j}$, $\alpha_{1,i,j}$ and $z_{t,i,j}$ represent the intercept, slope and error term of the AR(1) model of $k_{t,i,j}$ respectively,

$$k_{t,i,j} = \alpha_{0,i,j} + \alpha_{1,i,j}\ k_{t-1,i,j} + z_{t,i,j}\ . \tag{10}$$

Also for simplicity purpose, the mortality indices in (9) are all assumed to be independent and separately extrapolated in our projection.

To fit the 2-tier ACF model, we update our parameters using the Newton-Raphson method

$$\theta^* = \theta - \frac{\partial l/\partial\theta}{\partial^2 l/\partial\theta^2}\ , \tag{11}$$

where l is the log likelihood and θ represents any parameter to be fitted. Let $d_{x,t,i,j}$ be the observed number of death for age x, year t, sex i and country j, and $\hat{d}_{x,t,i,j}$ is the corresponding fitted number of deaths

$$\hat{d}_{x,t,i,j} = E_{x,t,i,j} \exp\left(\hat{a}_{x,i,j} + \hat{B}_x \hat{K}_t + \hat{b}_{x,i} \hat{k}_{t,i} + \hat{b}_{x,i,j} \hat{k}_{t,i,j}\right), \quad (12)$$

$$l = \sum_{x,t,i,j}\left[d_{x,t,i,j} \ln\left(\frac{d_{x,t,i,j}}{\hat{d}_{x,t,i,j}}\right) - d_{x,t,i,j} + \hat{d}_{x,t,i,j}\right]. \quad (13)$$

Adapting Brouhns et al. (2002) and Li (2012), parameters in (9) above are updated by the steps I to XI as follows,

Step I: Initialise parameter values $\hat{a}_{x,i,j}$ as the mean of $ln(m_{x,t,i,j})$, $\hat{K}_t = \hat{k}_{t,i} = \hat{k}_{t,i,j} = 0$, and $\hat{B}_x = \hat{b}_{x,i} = \hat{b}_{x,i,j} = 1/101$

Step II: Update $\hat{a}_{x,i,j}{}^* = \hat{a}_{x,i,j} + \sum_t (d_{x,t,i,j} - \hat{d}_{x,t,i,j}) / \sum_t \hat{d}_{x,t,i,j}$ for all x, i and j, and recalculate $\hat{d}_{x,t,i,j}$;

Step III: Update $\hat{K}_t{}^* = \hat{K}_t + \sum_{x,i,j}(d_{x,t,i,j} - \hat{d}_{x,t,i,j}) \hat{B}_x / \sum_{x,i,j} \hat{d}_{x,t,i,j} \hat{B}_x^2$ for all t, adjusted by the constraint $\sum_t K_t = 0$ and recalculate $\hat{d}_{x,t,i,j}$;

Step IV: Update $\hat{B}_x{}^* = \hat{B}_x + \sum_{t,i,j}(d_{x,t,i,j} - \hat{d}_{x,t,i,j}) \hat{K}_t / \sum_{t,i,j} \hat{d}_{x,t,i,j} \hat{K}_t^2$ for all x, adjusted by the constraint $\sum_x B_x = 1$, and recalculate $\hat{d}_{x,t,i,j}$;

Step V: Repeat Step II to IV till deviance converges[2];

Step VI: Update $\hat{k}_{t,i}{}^* = \hat{k}_{t,i} + \sum_{x,j}(d_{x,t,i,j} - \hat{d}_{x,t,i,j}) \hat{b}_{x,i} / \sum_{x,j} \hat{d}_{x,t,i,j} \hat{b}_{x,i}^2$ for all t and i, adjusted by the constraint $\sum_t k_{t,i} = 0$, and recalculate $\hat{d}_{x,t,i,j}$;

Step VII: Update $\hat{b}_{x,i}{}^* = \hat{b}_{x,i} + \sum_{t,j}(d_{x,t,i,j} - \hat{d}_{x,t,i,j}) \hat{k}_{t,i} / \sum_{t,j} \hat{d}_{x,t,i,j} \hat{k}_{t,i}^2$ for all x and i, adjusted by the constraint $\sum_x b_{x,i} = 1$, and recalculate $\hat{d}_{x,t,i,j}$;

Step VIII: Repeat step VI to VII till deviance converges;

---

[2] The author used "repeat" function in R to iterate the parameter updating steps, and break the loop if the difference in deviance from the previous iteration is less than 10^{-10} as the condition of convergence. Matrix operations are used in the "repeat" loop when updating parameters to improve efficiency of the codes.

Step IX: Update $\hat{k}_{t,i,j}{}^{*} = \hat{k}_{t,i,j} + \sum_x (d_{x,t,i,j} - \hat{d}_{x,t,i,j}) \, \hat{b}_{x,i,j} / \sum_x \hat{d}_{x,t,i,j} \; \hat{b}^2_{x,i,j}$ for all *t*, *i* and *j*, adjusted by the constraint $\sum_t k_{t,i,j} = 0$, and recalculate $\hat{d}_{x,t,i,j}$;

Step X: Update $\hat{b}_{x,i,j}{}^{*} = \hat{b}_{x,i,j} + \sum_t (d_{x,t,i,j} - \hat{d}_{x,t,i,j}) \, \hat{k}_{t,i,j} / \sum_t \hat{d}_{x,t,i,j} \, \hat{k}^2_{t,i,j}$ for all x, i and j, adjusted by the constraint $\sum_x b_{x,i,j} = 1$, and recalculate $\hat{d}_{x,t,i,j}$;

Step XI: Repeat step IX and X till the deviance converges.

Similar to (7), the deviance for the 2-tier ACF model is given by

$$deviance = \sum_{x,t,i,j} 2 \left[d_{x,t,i,j} ln \left(\frac{d_{x,t,i,j}}{\hat{d}_{x,t,i,j}} \right) - d_{x,t,i,j} + \hat{d}_{x,t,i,j} \right]. \qquad (14)$$

Step II to XI can be thought of as three major stages as below:

- Fit $\hat{a}_{x,i,j} + \hat{B}_x \, \hat{K}_t$, corresponding to Steps II to V;
- Conditional on that, fit $\hat{b}_{x,i} \, \hat{k}_{t,i}$, which corresponds to Steps VI to VIII;
- Conditional on above two stages, fit $\hat{b}_{x,i,j} \, \hat{k}_{t,i,j}$, which corresponds to Steps IX to XI.

The algorithm is designed in three stages so that for each bilinear component, the mortality time index and age sensitivities are fitted in a way that best explain the overall trend of an aggregated population, leaving any trends particular to a subpopulation to the next stage of model fitting. This also ensures the identifiability and convergence of the model under the constraints above.

In the later sections, this 2-tier ACF model is applied to model the six subpopulations of the UK (two genders and three countries within each gender). Section 4 also introduces the cohort extensions to the above ACF and the 2-tier ACF models.

3. Comparison between 2-tier ACF and Independent LC

3.1 Model Fitting and Short-Term Forecast Evaluation

In this section, we focus on the comparison of model-fitting and short-term forecast accuracy between the following two models when fitted to the six subpopulations of the UK:

- The 2-tier ACF with a common factor for the total UK population, a gender specific factor, and a gender-country specific factor
- The Lee-Carter model for each gender and each country separately.

Both models are fitted to the period between 1975 and 2000, and statistical measures, including BIC, AIC, Mean Absolute Percentage Error (MAPE), and Explanation Ratios[3] are examined for the two models (Table 1). The smaller BIC, AIC, MAPE are, the higher the explanation ratio, the better a model fits the past experience. From Table 1, it is clear that the 2-tier ACF outperforms the LC model by all statistical measures. While the differences are slight, it can be concluded that the 2-tier ACF fits the past data better than the LC model. All numbers in the tables are exact values rather than percentages.

It should be noted that both models fit better to the mortality experience in England & Wales, less so to Scotland, and fit least well to Northern Ireland. This is due to the fact that populations with larger exposures have more stable historical mortality patterns hence easier to fit using simplified mathematical models. England & Wales

[3] BIC = $-2l\,(\hat{m}) + \ n_p \ln\,(n_d)$ based on Schwarz (1978);
AIC = $-2l\,(\hat{m}) + \ 2\,n_p$ based on Akaike (1974);
MAPE $= \frac{1}{n_d}\sum_{x,t,i,j}\left|\frac{\hat{d}_{x,t,i,j}-d_{x,t,i,j}}{d_{x,t,i,j}}\right|$ and Explanation Ratio $= 1 - \frac{\sum_{x,t,i,j}\{d_{x,t,i,j}-e_{x,t,i,j}*\exp(\hat{m}_{x,t,i,j})\}^2}{\sum_{x,t,i,j}\{d_{x,t,i,j}-e_{x,t,i,j}*\exp(a_{x,i,j})\}^2}$ similar to Li (2012), where n_p is the number of parameters netted of number of constraints and n_d is the number of actual observations. The term $\hat{m}_{x,t,i,j}$ is the fitted mortality rate, and $l(\hat{m})$ is the log-likelihood with all the fitted parameters.

is largest population among the three countries; therefore the model best fits its experience, followed by Scotland and then Northern Ireland.

Table 1: BIC, AIC, MAPE & Explanatory Ratio of 2-tier ACF and LC

1975-2000 fitted model		2-tier ACF	Lee-Carter
	BIC	145227	146205
	AIC	131959	135811
MAPE	Female England &Wales	0.0446	0.0506
	Female Scotland	0.1371	0.1394
	Female Northern Ireland	0.2678	0.2706
	Male England & Wales	0.0394	0.0437
	Male Scotland	0.1131	0.1149
	Male Northern Ireland	0.2013	0.2018
	Overall	0.1339	0.1368
Explanation Ratio	Female England &Wales	0.9674	0.9493
	Female Scotland	0.8472	0.8152
	Female Northern Ireland	0.7502	0.7490
	Male England & Wales	0.9727	0.9682
	Male Scotland	0.9017	0.8990
	Male Northern Ireland	0.8027	0.8021
	Overall	0.9697	0.9601

The second part of this section is to compare how well the model predicts when we use the two models to project mortality experience during the forecast period from 2001 to 2011. Adopting Booth et al. (2005), Mean Error (forecast-observed) and Mean Absolute Error (|forecast – observed|) of log-scale mortality rates were computed to measure the predictive accuracy of models. The Mean Absolute Errors of log mortality rates are compared in Table 2. The 2-tier ACF is marginally better on average in forecasting the mortality rates during this period. The 2-tier ACF model is at least as accurate as if not better than the separate LC model for forecasting the mortality experience of the six subpopulations in the UK.

Table 2: Mean Absolute Error (log-mx) of 2-tier ACFC vs. LC

Forecast Period 2001-2011		2-tier ACF	Lee-Carter
Mean Absolute Error (log-mx), mean across all ages and years	Female England &Wales	0.1074	0.0977
	Female Scotland	0.1638	0.1849
	Female Northern Ireland	0.2342	0.2535
	Male England & Wales	0.1185	0.1221
	Male Scotland	0.1719	0.1999
	Male Northern Ireland	0.2199	0.2773
	Overall	0.1693	0.1892

Figure 1 plots the Mean Error in logarithm of mortality rates (mean over all years in the forecast period) for each subpopulation against age. For both modelss, the scale of mean errors are within a reasonable range (mostly between -0.5 and 0.5). Neither of the models systematically over or under estimates mortality experience, suggesting the time indices k's have captured mortality improvement sufficiently well. Both models underestimate mortality at very young ages and overestimate mortality for young adults and those aged between 60 and 80. As explained by Booth et al. (2005), this issue is due to the fact that the age sensitivity terms fitted do not adequately capture the age pattern changes in the forecast period. However, for most of ages, the forecast is sufficiently close to actual observations. As the fitting period is relatively short, there is no consistent underestimation of mortality rate across all different ages such as what was observed by Booth et al. (2005). The errors in projection shown in Figure 1 also include the jump-off bias as the jump-off point in this project is chosen to be the fitted value at the end of fitting period instead of the actual observation. This is based on the fact that when the fitting period is relatively short, the predictive accuracy is better when jumping off from fitted value for the final year in the base period, as features specific to the jump-off year would otherwise be extrapolated (Booth et al. 2006).

Figure 1: Mean Error (log-mx) by age (mean over all years) for the six subpopulations

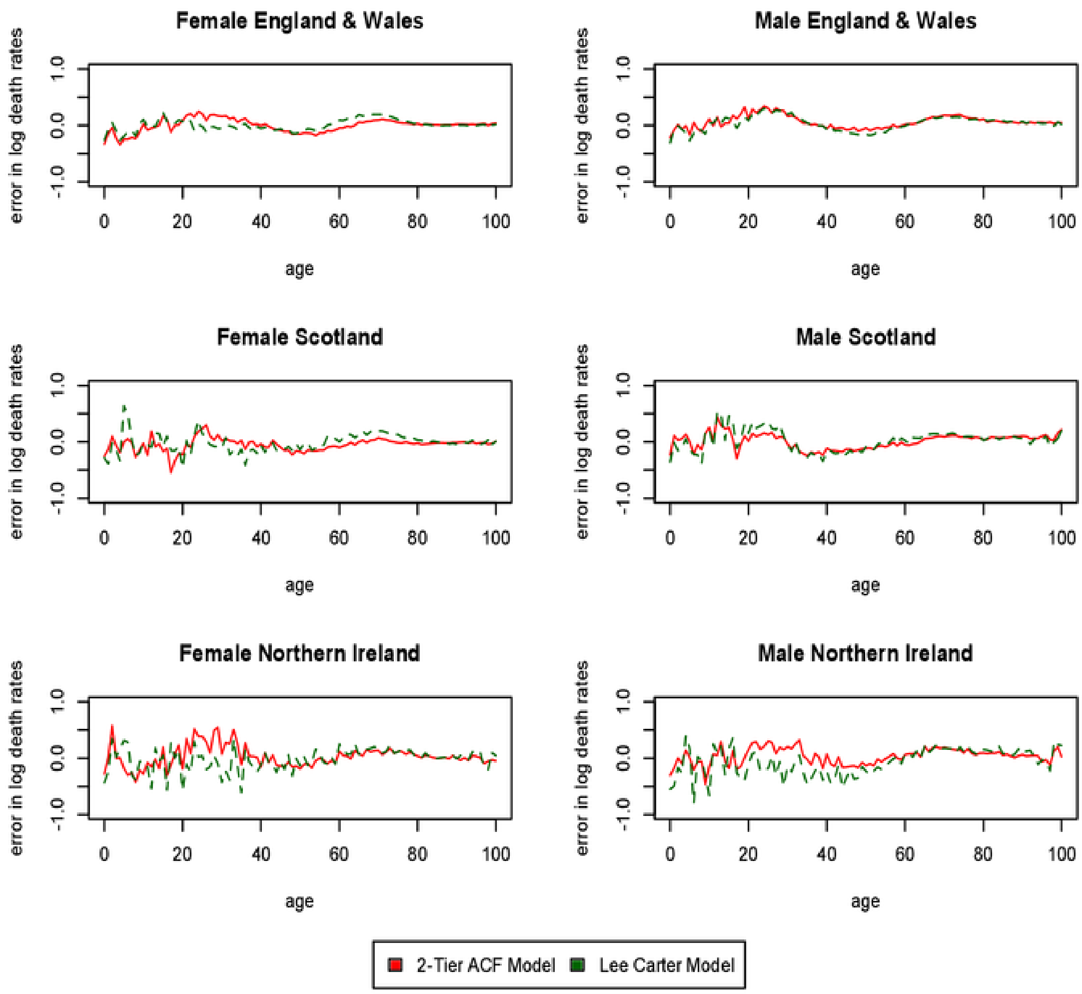


Table 3 also compares the two models using the Mean Absolute Error (life expectancy[4]) that represents the average of (|forecast – observed|) life expectancy. The 2-tier ACF is much more accurate when estimating female and male Scotland life expectancy, and is slightly weaker when estimating male England & Wales and male Northern Ireland. Overall, The 2-tier ACF is more accurate than independent

[4] Life expectancy always refers to period life expectancy in this work.

projection using the LC. As both models produce Mean Absolute Error of life expectancy less than 0.7 year, we can conclude that both models perform reasonably well in terms of their short-term predictive accuracy, and the 2-tier ACF performs better than the independent LC forecast in this analysis.

Table 3: Mean Absolute Error (Life Expectancy) of 2-tier ACFC vs. LC

Forecast Period 2001-2011		2-tier ACF	Lee-Carter
Mean Absolute Error life expectancy (in years), mean across all ages and years	Female England &Wales	0.1169	0.3497
	Female Scotland	0.0981	0.4862
	Female Northern Ireland	0.1757	0.2984
	Male England & Wales	0.6306	0.5291
	Male Scotland	0.3589	0.4935
	Male Northern Ireland	0.6049	0.5247
	Overall	0.3308	0.4469

3.2. Long-term Projection

In this section, the long-term projection behaviours of the models are studied to highlight the various merits of the 2-tier ACF when producing long-term projections over the independent LC model. In particular, the 2-tier ACF is more capable in producing coherent and smooth mortality projections. The 2-tier ACF and the independent LC are fitted to the past data of six subpopulations of the UK between 1975 and 2011, and then projected to 2050.

Figure 2 is the central estimates for log mortality rates by age in year 2050 by the 2-tier ACF and the independent LC. Firstly, the 2-tier ACF produces forecast much smoother from age to age as compared to the LC. In Figure 2, the LC projection for Scotland male even shows decreasing mortality by age at around age 40. Lack of cross-age smoothness of the LC model has long been highlighted in research (Cairns et al. 2007), as it uses only one age modulator b_x to measure the age sensitivity to mortality improvement for the specific subpopulation and assumes that it remains constant. Over time, small differences between nearby b_x causes leads to large discrepancy of mortality forecasts between nearby ages, causing lack of smoothness.

However, in the 2-tier ACF model, for each subpopulation, the mortality improvement trend is decomposed into three components – the common trend of total population, the trend of a specific gender, and the trend of the specific subpopulation. In the 2-tier ACF (9), B_x, $b_{x,i}$, and $b_{x,i,j}$ captures the sensitivity of each age to the three different trends separately. Jointly, they create an age pattern that does not jump between nearby ages as age sensitivities are captured more finely from the historical data, thereby displaying smoother cross-age mortality improvement.

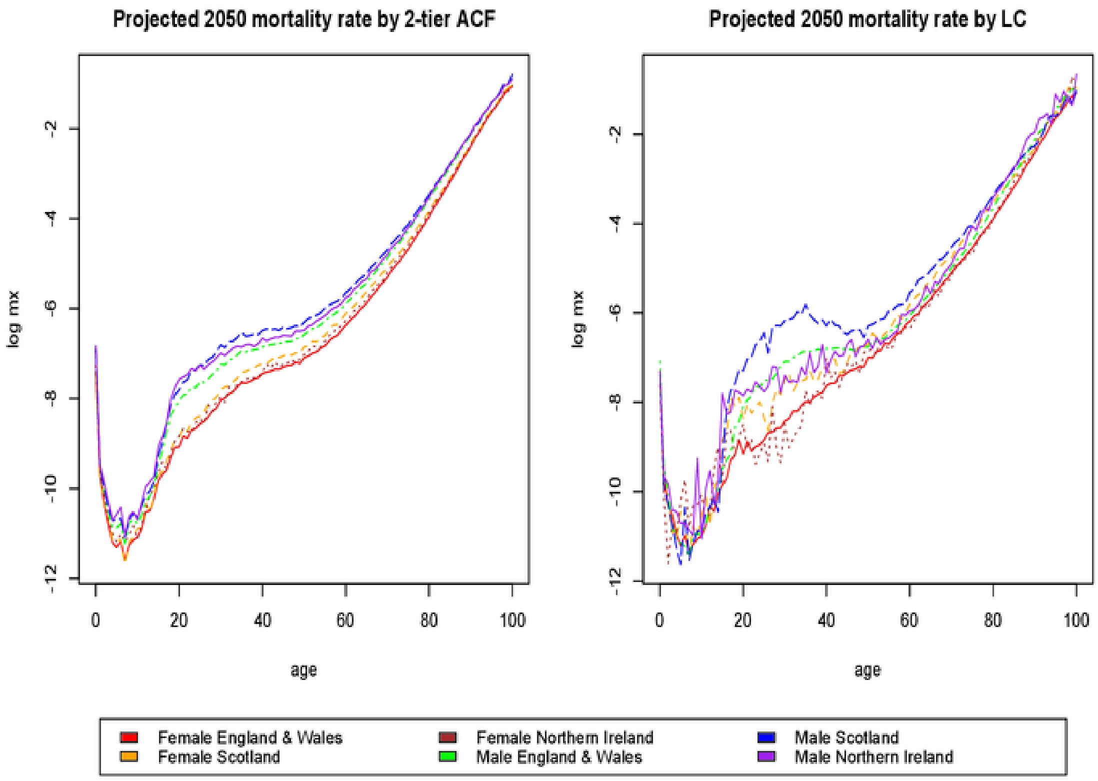
Secondly, Figure 2 shows that the LC model produces much larger differences among countries within each gender, and between different genders within each country, especially for the age range between 20 and 60. This is consistent with our expectation that independent extrapolations of different subpopulations under the LC model produce divergent mortality rates for related populations, whereas the ACF framework avoids such issue. As pointed out by Cairns et al. (2011), under the ACF framework, the global improvement trend will dominate over time, due to the fact that the subpopulation-specific components are mean reverting. The 2-tier ACF further extends the ACF model, so that the projections for different countries are dominated by the trend of the same gender. In other words, this extension ensures that the ratios of different subpopulations of the same gender converge over time, because the trend of the gender as a whole dominates over the trend in the specific subpopulation.

For subpopulations in country j and k of the same gender i, the difference of age specific mortality (log scale) are given by:

$$ln(m_{x,t,i,j}) - ln\left(m_{x,t,i,k}\right) = (a_{x,i,j} - a_{x,i,k}) + (b_{x,i,j}\ k_{t,i,j} - b_{x,i,k}\ k_{t,i,k})\ . \qquad (15)$$

As $k_{t,i,j}$ and $k_{t,i,k}$ are mean reverting process, it is clear that the difference in log-scale mortalities is a mean-reverting process too. Hence the differences in mortality rates between countries are more constrained in the 2-tier ACF projection compared to the LC.

Figure 2: Projected Year 2050 log-mx by 2-tier ACF vs. LC



The extension also ensures the expected male-female ratios of different countries converge over time in a similar trend. Let *f* represents the female population and *m* represents the male population, then for country *j*, the male-female mortality ratio on a log scale is:

$$ln(m_{x,t,m,j}/m_{x,t,f,j}) = (a_{x,m,j} - a_{x,f,j}) + (b_{x,m}\, k_{t,m} - b_{x,f}\, k_{t,f}) + (b_{x,m,j}\, k_{t,m,j} - b_{x,f,j}\, k_{t,f,j}) \quad (16)$$

From (16), it can be seen that, the male-female ratio for all countries will share the common component $b_{x,m}\, k_{t,m} - b_{x,f}\, k_{t,f}$, which is reverting to a positive mean, and this component captures the overall trend in gender differences for the whole UK. Although the component $b_{x,m,j}\, k_{t,m,j} - b_{x,f,j}\, k_{t,f,j}$ is specific to each country and could possibly revert to a non-zero mean, but after fitting the overall trend and gender trends, $k_{t,m,j}$ and $k_{t,f,j}$ are normally best fitted by AR(1) process with zero long-term mean - the results actually show that the trend in male-female ratio of each country are dominated by the male-female ratio of the whole UK.

Figure 3: Projected Life Expectancy at Birth by 2-tier ACF vs. LC

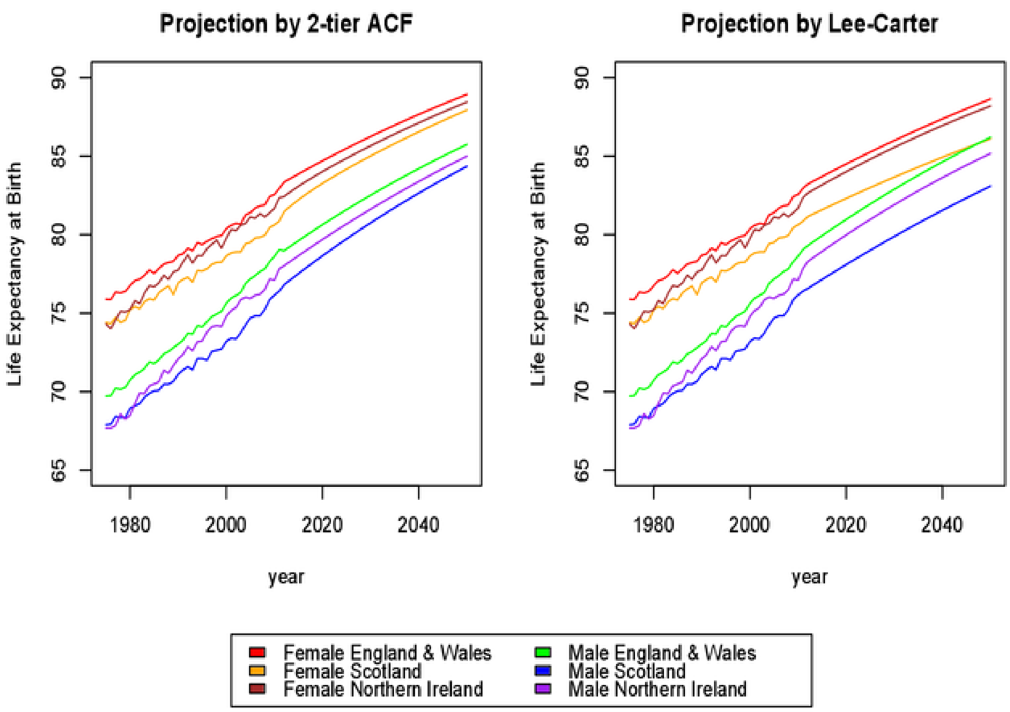


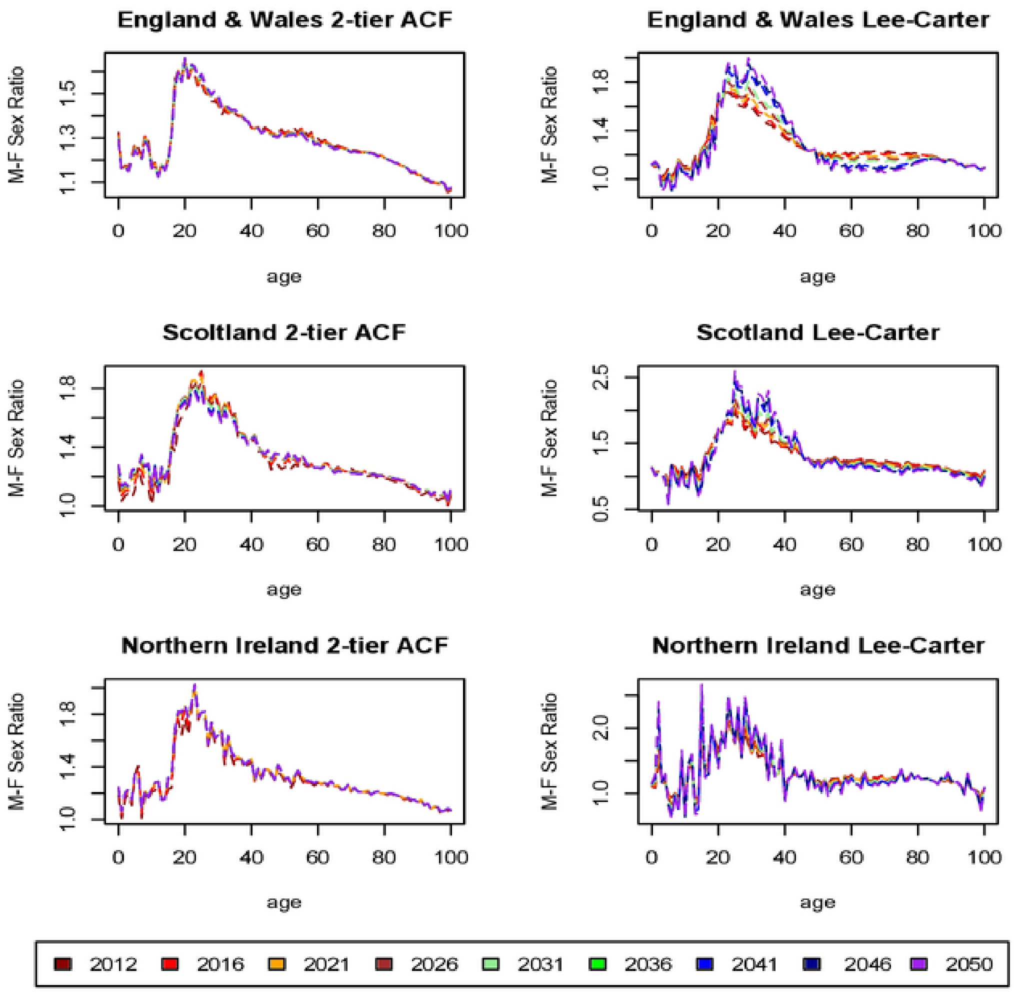
Figure 3 shows the projected life expectancy at birth for all six subpopulations, using the 2-tier ACF and LC. Under the 2-tier ACF framework, the life expectancy forecasts are more constrained, whereas for the independent LC model, life expectancy forecasts are diverging. Under the 2-tier ACF framework, subpopulations of the same gender show a similar trend over time, and both genders converge to the common overall trend. With the LC projection, there is an increasing gap in life expectancy between Scotland and the rest of the UK for both genders. As suggested by Mccartney et al. (2011), before 1980, the higher mortality experienced by Scotland is most likely contributed by the deprivation and poverty linked to the industrial employment patterns. Since 1980, the higher mortality in Scotland is most likely due to community disruption caused by deindustrialisation, which affected the West of Scotland more than the rest of UK. These essential historical factors may continue to cause higher mortality in Scotland compared to the rest of the UK. However, it is difficult to justify an increasingly widening gap in mortality between countries in four decade's time, when the countries are related in terms of politics, economics and healthcare. Scotland is the only country so far providing free personal social care for those aged 65 or above (COSLA & The Scottish Government n.d.), and has a level of health funding per head much above England (Bevan et al. 2014). Latest research has also shown that gap of health system performance indicators has narrowed between Scotland and rest of UK due to dramatic improvement in Scotland since 2010 (Bevan et al. 2014). Greater regional equality across the UK is an objective underlying all the public policies and the 2-tier ACF framework allows both the short-term disparities among countries and a more reasonable future outlook.

In the remaining part of this section, the coherence of projection results are examined against the definition of coherence, proposed by Hyndman et al. (2013), that is the expected ratios of age-specific mortality rates of any two subpopulations should converge to some appropriate constants over time. In particular, projected male-female ratio in death rates (on a square-root scale, as in Hyndman et al. 2013) are examined for each country and compared between the two models. In Figure 4, the sex ratios in the projected years are plotted against age for a selection of years in the projection period.

From Figure 4, it can be seen that for the LC projection, as also found by Hyndman et al. (2013), at the very young and old ages, when the number of deaths is very small, undesirable projection outcomes of sex ratios less than 1 may occur. However, the projection under the 2-tier ACF does not have such issues. The sex ratios remain quite stable and constrained over time under the 2-tier ACF, whereas the independent LC produces very large sex ratios (as high as 2.5) in some years, showing the undesirable feature of a divergent projection model. The 2-tier ACF also produces smoother cross-age sex ratios, and the results are in line with the understanding that sex differences in mortality is mainly contributed by the high mortality of very young and middle aged males (Kalben 2002).

It is concluded from above analyses that the 2-tier ACF model shows strong coherence property for long-term projections with reasonable forecasts for different countries within each gender and stable sex ratios within each country, and is superior to the independent LC in this aspect.

Figure 4: Sex Ratio of Death Rates: sqrt(M/F) over 39 Years of Projection



4. Cohort Extension of ACF (1-tier and 2-tier)

The previous sections have demonstrated the desirable coherence property of the 2-tier ACF model as compared to the independent LC. However, the models introduced so far have not taken into account the cohort effect of mortality – the mortality experience does not only depend on the year *t*, but also relate to the year of birth *t-x*. For people born in the same year, we refer them as a cohort. In the UK, the cohort effect has a more narrowed meaning referring to the more rapid improvement and lower death rates in mortality for the golden generation born between 1925 and 1945 (Willets, 2004). Such pattern can be observed while plotting residuals against cohorts. According to Renshaw & Haberman (2003) and Li (2012), because the model fitting uses over-dispersed Poisson distributions, the standard deviance residual is given by

$$sgn\left(d_{x,t,i,j} - \hat{d}_{x,t,i,j}\right)\sqrt{\left\{2\left[d_{x,t,i,j}ln\left(\frac{d_{x,t,i,j}}{\hat{d}_{x,t,i,j}}\right) - d_{x,t,i,j} + \hat{d}_{x,t,i,j}\right]/\hat{\phi}\right\}}\ . \tag{17}$$

where

$$\hat{\phi} = deviance/(n_d - n_p)\ , \tag{18}$$

where n_d is the number of observations and n_p is the effective number of parameters - the number of parameters netted off the number of constraints. It is noted that, the index factor *j* representing country should be left out for the 1-tier ACF.

The purpose of this section is to introduce cohort extensions into the 2-tier ACF framework in a simple and reasonable way. However, since the original 1-tier ACF (Li & Lee, 2005) does not consider the cohort effect either, the first step is to introduce a cohort factor to the 1-tier ACF in Section 4.1, before extending the cohort component to the 2-tier ACF in Section 4.2.

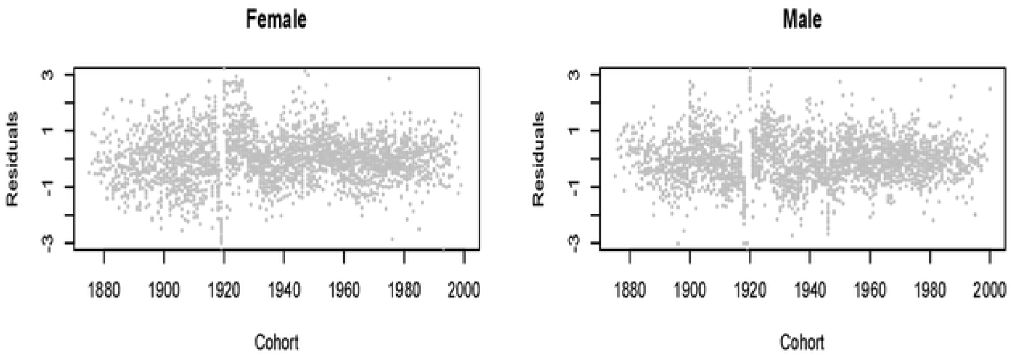
4.1 Cohort Extension of 1-tier ACF

From Figure 5 below, the cohort effect can be observed when the 1-tier ACF is fitted to the mortality experience of total females and males in the UK from 1975 to 2000, particularly apparent for the female population. This shows a limitation of the original ACF model proposed by Li & Lee (2005) that it only captures trends in mortality via relations with age *x* and period *t*, but not cohort *t-x*. Yang et al. (2014) also highlighted this issue, and they introduced six possible variants of cohort extensions to the PCFM (Li 2012).

From Figure 5, it is also clear that cohort effect differs between the two genders. For females, the cohort effect is more prominent for the golden generation (1925-1945), while for males cohort pattern is more volatile and seem to present in a few different generations.

Figure 5: Standard Deviance Residuals by Cohort, Fitted by ACF

Female (left) and Male (right) of the Total UK Population, Fitted by ACF (1975-2000 fitting period)



In this project, we introduce one simple cohort extensions to the original ACF:

$$ln(m_{x,t,i}) = a_{x,i} + B_x K_t + b_{x,i} k_{t,i} + g_{t-x,i} \ . \tag{19}$$

All the definitions above are the same as in Section 2.2, but a new cohort parameter $g_{t-x,i}$ is introduced for each gender i, which is based on the observation that the cohort patterns are fairly dissimilar between the two genders. The uniqueness of the cohort factor is guaranteed by $\sum_{h=t-x} g_{h,i} = 0$ for each gender i.

To ensure convergence between the two genders over time, a sufficient and necessary condition is that $g_{t-x,i}$ should be extrapolated as mean-reverting time series. Although Yang et al. (2014) suggested to use an AR(p) model, for simplicity purpose, $g_{t-x,i}$ are extrapolated as independent AR(1) process similar to $k_{t,i}$, such that the sex ratio between the two genders is a mean-reverting process on its own:

$$\ln \{\frac{m_{x,t,m}}{m_{x,t,f}}\} = (a_{x,m} - a_{x,f}) + (b_{x,m}\ k_{t,m} - b_{x,f}\ k_{t,f}) + (g_{t-x,m} - g_{t-x,f})\ . \qquad (20)$$

Yang et al. (2014)'s approach is to fit the original ACF model first and fit $g_{t-x,i}$ to the residuals to ensure consistency with the ACF model. However, this approach is fundamentally different from Renshaw & Haberman (2006) when they first introduced the cohort extension to the LC model. Renshaw & Haberman (2006) used a two-step fitting process, where the static age effect a_x is specified first as the average of log scale mortality rates (same as the SVD method), conditional on which the cohort factors are fitted simultaneously with the period factor k_t and the age modulating indices, i.e. b_x in the basic LC model. Using Yang et al. (2014) approach, the cohort factor fitted turn out to be much more erratic than the Renshaw and Haberman (2006) approach, and this will be discussed later in Section 4.3 when the cohort extension is critically appraised. For now, the model given by (20) is still fitted to the UK female and male population using Yang et al. (2014)'s approach, for fitting period 1975 to 2000, age 0 to 100, and residual plots by cohorts are given by Figure 6 below. Moreover, it should be noted that because Yang et al. (2014)'s approach ensures a coherent framework by fitting residuals of the ACF, so the cohort factor is a stationary process. Comparing Figure 6 to Figure 5, it can be seen that the randomness of residuals is dramatically improved by including the cohort factor, and there is no identifiable systematic pattern in Figure 6.

We then compare the various statistical measures of model fitting between the ACF and the ACF with Cohort Extension (ACFC) below in Table 4. It is clear that the ACFC outperforms the ACF, when measured by BIC and AIC after penalising the increase in parameters.

Figure 6: Standard Deviance Residuals by Cohort, Fitted by ACFC

Female (left) and male (right) of the Total UK Population, fitted by ACF with cohort factor $g_{t-x,i}$ (1975-2000 fitting period)

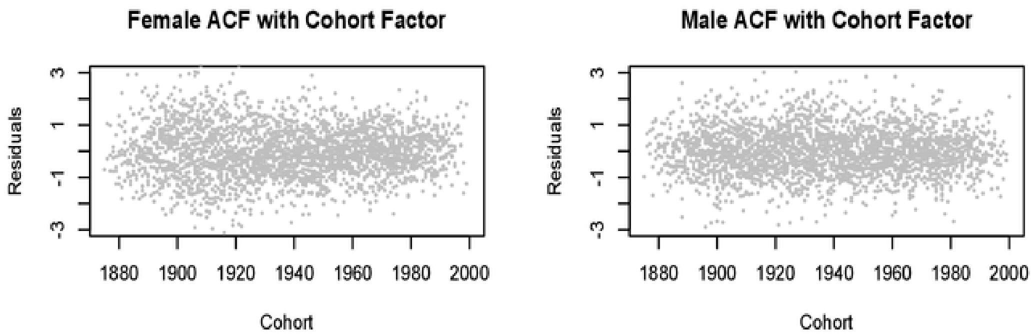


Table 4: BIC, AIC, MAPE, & Explanation Ratio of ACFC vs. ACF

Fitting Period 1975-2000	BIC	AIC	MAPE			Explanation Ratio		
			Female	Male	Overall	Female	Male	Overall
ACF	65738	61949	0.0460	0.0394	0.0427	0.9469	0.9711	0.9621
ACFC	61750	56320	0.0429	0.0360	0.0394	0.9717	0.9899	0.9831

Similar to Section 3.1, the results of back testing for the out-of-sample period from 2001 to 2011 are shown in Table 5 and 6. It can be seen that the ACFC gives marginally smaller mean absolute error of log scale mortality rates than the ACF. As for mean absolute life expectancy, the ACFC gives slightly larger error for females, but predicts better for males. Hence it can be concluded that the ACFC model improves the model fitting of the ACF, maintains the coherence property, and performs reasonably well in the out-of-sample tests for log-scale mortality and life expectancy.

Table 5: Mean Absolute Error (log-mx) of ACFC vs. ACF

Forecast Period 2001-2011	Mean Absolute Error log-mx (mean across all ages and years)	
	Female	Male
ACF	0.10586	0.11250
ACFC	0.10351	0.11038

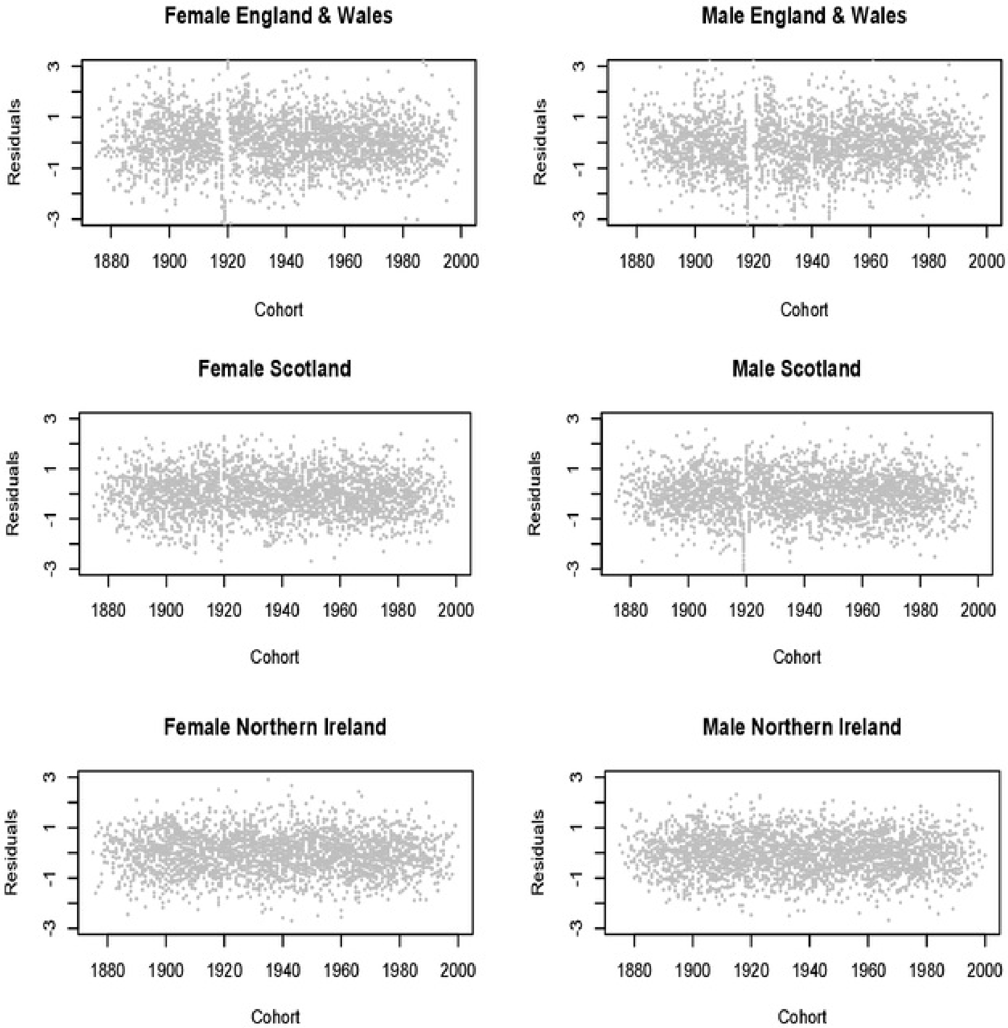
Table 6: Mean Absolute Error (Life Expectancy) of ACFC vs. ACF

Forecast Period 2001-2011	Mean Absolute Error life expectancy (Year) (mean across all ages and years)	
	Female	Male
ACF	0.1988	0.5858
ACFC	0.2093	0.5549

4.2 Cohort Extension of 2-tier ACF

The 2-tier ACF introduced in Section 2.4 has not considered the cohort effect when modelling the mortality of the six subpopulations in the UK, which is an issue particularly for the England and Wales population that displays the most significant cohort trends among the three countries considered. Figure 7 shows the plots of standard deviance residuals after fitted by the 2-tier ACF against cohort for all six subpopulations of the UK.

Figure 7: Standard Deviance Residuals by Cohort, Fitted by 2-tier ACF



The first finding from Figure 7 is that the cohort effect is most prominent in England & Wales, and least in Northern Ireland. This is not surprising as England & Wales population has the largest exposures and therefore the greatest weighting under a Poisson framework when death counts are modelled directly. Hence the major source of cohort effect of the UK should be contributed by England & Wales.

When comparing Figure 7 to Figure 5, it is also obvious that the cohort effect of each individual country is less significant when compared to the aggregated UK population. Firstly, England and Wales has the largest exposures, so their cohort effect will naturally drive the aggregated population. Secondly, this could be explained by the fact that if the cohort effect of each individual country coincides and reinforces each other, then on an aggregate level the cohort effect becomes more prominent. Lastly but most importantly, in the 2-tier ACF model, there are three bilinear components for each subpopulation; the simple fact that cohort is merely time period minus age means that most of the cohort effect for each subpopulation may already be captured by the 2-tier ACF model internally without an explicit cohort factor, and Figure 7 shows what is left as the residual cohort effect. It should be emphasized that Figure 5 shows the residual plots of the 1-tier ACF, so it is reasonably expected that residual cohort effect is more prominent, because the 1-tier ACF is simpler and consists only two bilinear components.

The above findings suggest the possibility of fitting a cohort factor $g_{t-x,i}$ in the 2-tier model for each gender on a national level to capture the overall cohort effect of all three countries. Therefore, we introduce the 2-tier ACF model with cohort extension (2-tier ACFC) below:

$$ln(m_{x,t,i,j}) = a_{x,i,j} + B_x\,K_t + \ b_{x,i}\,k_{t,i} + \ b_{x,i,j}\,k_{t,i,j} + g_{t-x,i}\ , \qquad (21)$$

where all other parameters have the same meaning as in the 2-tier ACF model defined in Section 2.4. As the same cohort factor is fitted for population of the same gender in all three countries, there is no issue of divergence among countries. For the same reason explained by (20), this 2-tier ACFC model also implies the gender differences within each country is a mean-reverting stochastic process.

As Section 2.4 suggests, when fitting the 2-tier ACF model, we estimate each bilinear component in steps and ensure deviance is minimised when fitting each component. This is intended to set priorities among all the different bilinear terms reflecting the idea that common trends are prioritised in model fitting relatively to trends of a particular subpopulation. However, when the cohort term $g_{t-x,i}$ is included, there is a choice between:

- $g_{t-x,i}$ should be fitted after fitting the whole 2-tier ACF model and taking all parameters as given to ensure consistency between the 2-tier ACFC and the 2-tier ACF
- or $g_{t-x,i}$ should be fitted prior to fitting $b_{x,i,j}\ k_{t,i,j}$, but after fitting $B_x\ K_t$ because $g_{t-x,i}$ is part of the common trend of gender i at the aggregated national level.

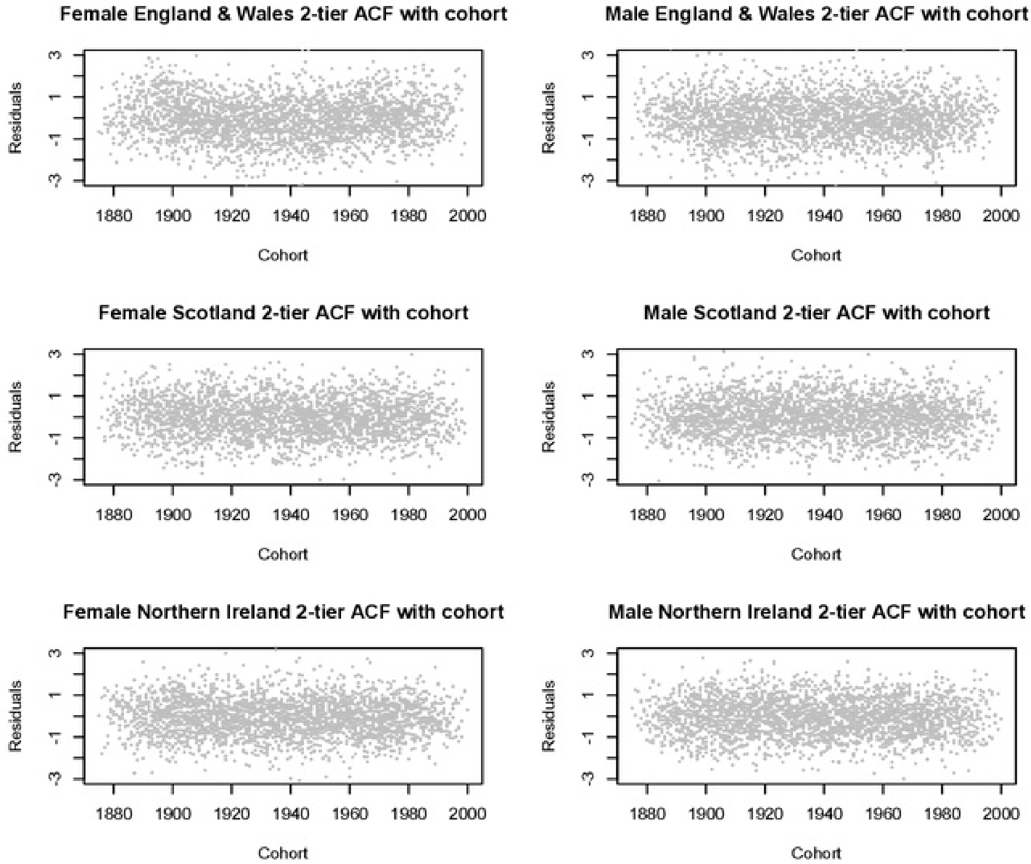
Although the first approach might be more in line with Yang et al. (2014), the second method is chosen for this project as it aligns better with the principle that common factor is prioritised to ensure convergence, before fitting any subpopulation specific factors. In the Steps I to XI in Section 2.4, we insert after step VIII the updating step for $g_{t-x,i}$ given by:

Step VIII a:

Update $\hat{g}_{h,i}^{*} = \hat{g}_{h,i} + \sum_{x,t,t-x=h,j} w_{x,t}(d_{x,t,i,j} - \hat{d}_{x,t,i,j}) / \sum_{x,t,t-x=h,j} w_{x,t}\ \hat{d}_{x,t,i,j}$ for all h and i, adjusted by the constraints $\sum_{h=t-x} g_{h,i} = 0$ for each i,

where $w_{x,t}$ = 0 for the first the last five cohorts of the fitting period population due to the scarcity of data in those cohorts. All other definitions are the same as Section 2.4 and (21). This single step is repeated till deviance converges, and then we continue with Step IX in Section 2.4.

Figure 8: Standard Deviance Residuals by Cohort, Fitted by 2-tier ACFC



Again the 2-tier ACFC is fitted to the period between 1975 and 2000 and back-tested for the period from 2001 to 2011. From Figure 8, it can be seen that after including one cohort factor for each gender on an aggregated national level, the cohort effect can no longer be observed in residual plots for England & Wales, and Scotland. Residuals by cohort year are randomly distributed within a desirable range and display no systematic pattern.

The statistical model-fitting measures of the 2-tier ACF and the 2-tier ACFC are then compared in Table 7 below. The results are consistent with our expectation that the 2-tier ACFC fits better than the 2-tier ACF by every standard.

Table 7: BIC, AIC, MAPE, & Explanation Ratio of 2-tier ACFC vs. 2-tier ACF

1975-2000 fitted model		2-tier ACF	2-tier ACFC
	BIC	145227	142600
	AIC	131959	127417
MAPE	Female England &Wales	0.0446	0.0423
	Female Scotland	0.1371	0.1354
	Female Northern Ireland	0.2678	0.2662
	Male England & Wales	0.0394	0.0361
	Male Scotland	0.1131	0.1115
	Male Northern Ireland	0.2013	0.2009
	Overall	0.1339	0.1321
Explanation Ratio	Female England &Wales	0.9674	0.9853
	Female Scotland	0.8472	0.8617
	Female Northern Ireland	0.7502	0.7548
	Male England & Wales	0.9727	0.9903
	Male Scotland	0.9017	0.9175
	Male Northern Ireland	0.8027	0.8040
	Overall	0.9697	0.9873

Table 8: Mean Absolute Error (log-mx) of 2-tier ACFC vs. 2-tier ACF

Forecast Period 2001-2011		2-tier ACF	2-tier ACFC
Mean Absolute Error (log-mx), mean across all ages and years	Female England &Wales	0.1074	0.1074
	Female Scotland	0.1638	0.1619
	Female Northern Ireland	0.2342	0.2340
	Male England & Wales	0.1185	0.1162
	Male Scotland	0.1719	0.1738
	Male Northern Ireland	0.2199	0.2189
	Overall	0.1693	0.1687

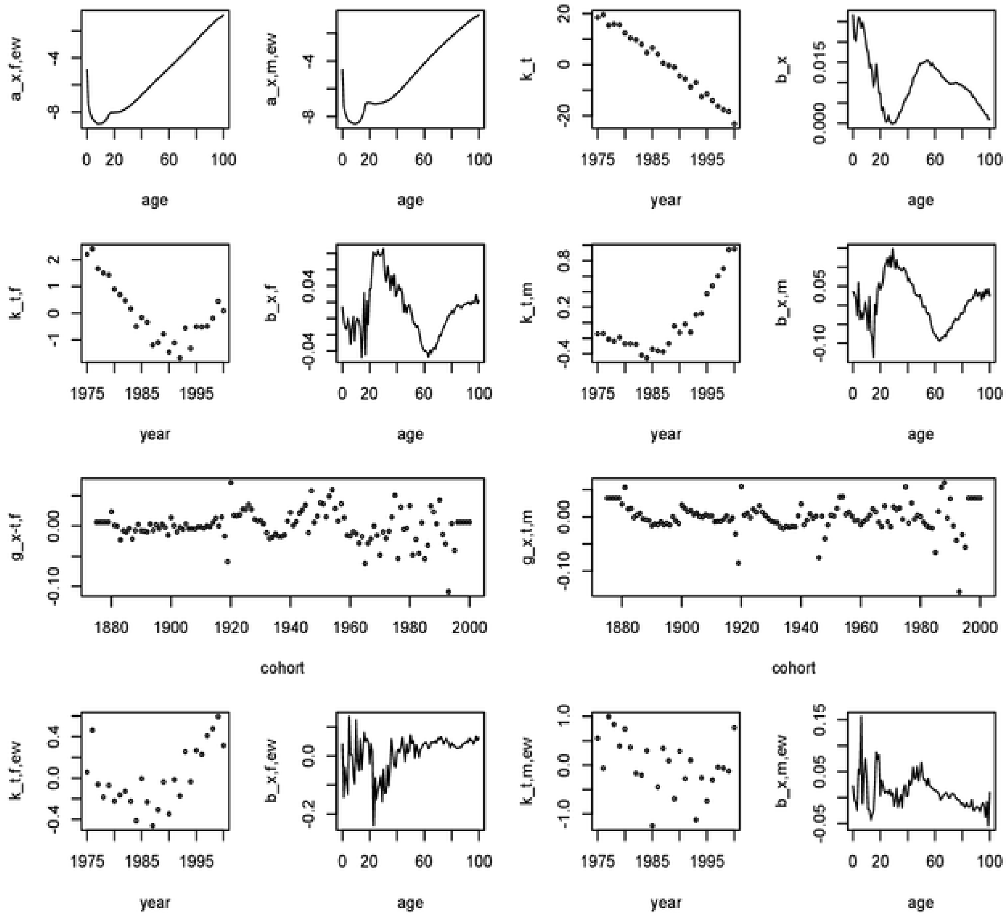
Table 8 and Table 9 compare the out-of-sample forecast results between the 2-tier ACF and the 2-tier ACFC. It can be seen that in terms of both mean absolute error of log mortality and mean absolute error life expectancy, the 2-tier ACFC perform better than the 2-tier ACF on an overall basis, but the difference is slight. However, because the out-of-sample tests are only carried out for a period as short as 11 years, for longer-term forecast, the cohort factor might contribute significantly to the overall predictive power of the model.

Table 9: Mean Absolute Error (Life Expectancy) of 2-tier ACFC vs. 2-tier ACF

Forecast Period 2001-2011		2-tier ACF	2-tier ACFC
Mean Absolute Error Life Expectancy (in Years), mean across all ages and years	Female England &Wales	0.1169	0.1212
	Female Scotland	0.0981	0.0875
	Female Northern Ireland	0.1757	0.1989
	Male England & Wales	0.6306	0.5974
	Male Scotland	0.3589	0.3701
	Male Northern Ireland	0.6049	0.5674
	Overall	0.3308	0.3237

Figure 9: Plots of 2-tier ACFC Parameters

Plots of 2-tier ACFC fitted to period between 1975 and 2000 for the UK dataset, and parameter plots are shown for the England & Wales only ($i=f$ or m; $j=ew$ in (21)).



4.3 Critical Appraisal of the Cohort Extension to 2-tier ACF model

As discussed earlier, this project chooses to fit the cohort factor to residuals after fitting the term $a_{x,i,j} + B_x K_t + b_{x,i} k_{t,i}$ but before fitting $b_{x,i,j} k_{t,i,j}$ in the 2-tier ACFC model. The method is analogous to Yang et al. (2014)'s approach to fit cohort extensions to PCFM (Li 2012). As stated earlier, the approach taken is fundamentally different from the method proposed by Renshaw & Haberman (2006) when introducing the cohort extension to the LC model, where they fitted the cohort factor simultaneously with other parameters. However, Renshaw & Haberman (2006) approach cannot be readily applied into the ACF framework, as the multiple bilinear components of the ACF are arranged in hierarchy so that common trends are fitted prior to fitting individual subpopulation trends. Therefore, $g_{t-x,i}$ would have to be placed within this hierarchy, and for the model to make sense, $g_{t-x,i}$ should be specified after fitting $a_{x,i,j} + B_x K_t$ and before fitting $b_{x,i,j} k_{t,i,j}$. As $g_{t-x,i}$ is part of the common trend of the aggregated population of the same gender, which comes after the common trend of the entire population, but before the trend of each country within the gender. This is another key feature of the 2-tier ACFC model that including $g_{t-x,i}$ would still give raise to a coherent forecast in terms of differences in mortality among subpopulations, because the common trend of the entire population is prioritised while $g_{t-x,i}$ is modelled as stationary auto-regressive process. It may be argued that a common cohort factor g_{t-x} can be fitted together with $B_x K_t$ and coherence property can still be maintained. However, the residual plots from the 1-tier ACF suggests that cohort patterns do differ between different genders, which is consistent with the findings by Willets (2004).

However, the approach used in this project also fits $b_{x,i} k_{t,i}$ prior to fitting $g_{t-x,i}$. This predetermines the priority of time period factors over cohort factors, that is, it follows the assumption that mortality is more dependent on the time of death (period) than the time of birth (cohort) when fitting the idiosyncratic trend for each gender. Some research findings, however, disagrees with this assumption. Richards et al. (2006) suggested that when fitting mortality rates of the elderly population in the UK, the cohort effects are more prominent than the period effects. This may suggest alternative orders in fitting the different components of the ACFC model - one might choose to fit the cohort factor $g_{t-x,i}$ prior to fitting any $b_{x,i}\, k_{t,i}$ terms or at least fit them in one step when minimising the deviance function. Haberman & Renshaw (2009) also suggests that the order of model fitting in age-period-cohort models make a substantial difference to parameter shapes. Further research may therefore be able to identify more elegant way of including the cohort extensions within the 2-tier ACF hierarchy.

It should also be noted that it only makes sense to extrapolate $g_{t-x,i}$ as stationary process if $a_{x,i,j} + B_x\, K_t + b_{x,i}\, k_{t,i}$ is prioritised in the model fitting process, as it is the residuals after fitting these components that is driving $g_{t-x,i}$. The plots of cohort factors produced by Yang et al. (2014) are much more erratic compared to Renshaw & Haberman (2006). This is primarily because the PCFM model (Li 2012) used by Yang et al. has up to five sex-specific bilinear terms to capture the trends of a gender departing from the overall combined population, and if the whole PCFM model is fitted prior to fitting any cohort extension, the residuals used to fit the cohort extension are already very erratic. However, since we impose that only $a_{x,i,j}$, $B_x\, K_t$, and $b_{x,i}\, k_{t,i}$ are fitted before fitting $g_{t-x,i}$, and $b_{x,i,j}\, k_{t,i,j}$ is fitted after $g_{t-x,i}$, the cohort factor turns out to be less erratic (Figure 9) and easily interpretable. One can easily observe from the $g_{t-x,i}$ plots in Figure 9 the golden period from 1925 to 1945, especially for females, the $g_{t-x,i}$ terms are negative, indicating lower mortality than expected from the age-period model, i.e. the 2-tier ACF. Moreover, for the period around 1931, the slope of $g_{t-x,i}$ terms is negative for both genders, suggesting a faster pace of improvement, which is consistent with Willets (2004). Another merit of the current approach is that it avoids the issues in the 2-step method adopted by Renshaw & Haberman (2006) that the model may not converge for certain age period combinations of data and varying parameter patterns under different identifiability constraints, which makes the cohort factor harder to interpret, as is pointed out by Hunt & Villegas (2015).

The cohort factor is conventionally considered as a non-stationary process, as by definition it should capture the structural changes in mortality patterns by cohort.

However, because the approach taken here prioritises the model fitting of certain age and period terms, some cohort patterns may already be implicitly captured before fitting $g_{t-x,i}$, due to the simple fact that cohort is merely age x netted off time period t, and the cohort terms are intrinsically related to the prioritised age and period terms. Nevertheless, whether the residual cohort effect represented by $g_{t-x,i}$ in the ACFC models really represents structure trends in mortality and should be extrapolated into the future using AR(1) process are areas involving a lot of subjective judgements. What we can conclude from the above analysis is the cohort factors fitted under this method display reasonable trends over time and can be easily interpreted, although the pattern gets more erratic in later cohorts (after 1975s); the cohort factors can also improve the fitting of the model, evidenced by the lower BIC and AIC.

4.4 Application: Projecting UK Population to 2050 Using 2-tier ACFC

In previous sections, evidences were found that the 2-tier ACFC is a model that can produce both coherent forecasts between different subpopulations of the UK and capture the cohort effect that the 2-tier ACF fails to account for. Now, the 2-tier ACFC is fitted to the period between 1975 and 2011 (latest observations on HMD), and projected to year 2050.

Figure 10 shows the male-female mortality ratios projected for the three countries by the 2-tier ACFC for a selection of years. Although projecting the cohort factor of each gender independently introduces some variants over the forecast years for the 2-tier ACFC compared to the 2-tier ACF in Figure 4, the sex ratios for each country produced by the 2-tier ACFC still remain constrained in a stable and reasonable range. There is no crossover between male mortality and female mortality at any age in the projections and it is also reasonable that most gender differences in mortality fall within the age range 20-40. The forecasted sex ratios are higher in Northern Ireland and Scotland as compared to England & Wales. As suggested by Longevity Science Advisory Panel (2012), narrowing gender gap in mortality rates are attributed by strong downward trends in lifestyle factors such as tobacco and alcohol consumptions, which affects men more severely. However it may also signals other concerning social trends such as obesity, which affects females more than males. It is reasonable to assume that these factors are more significant in countries with overall better economic development such as England & Wales, as compared to Scotland and Northern Ireland.

Figure 10: Sex Ratios of Death Rates: sqrt(M/F) over 39 Years of Projection by 2-tier ACFC

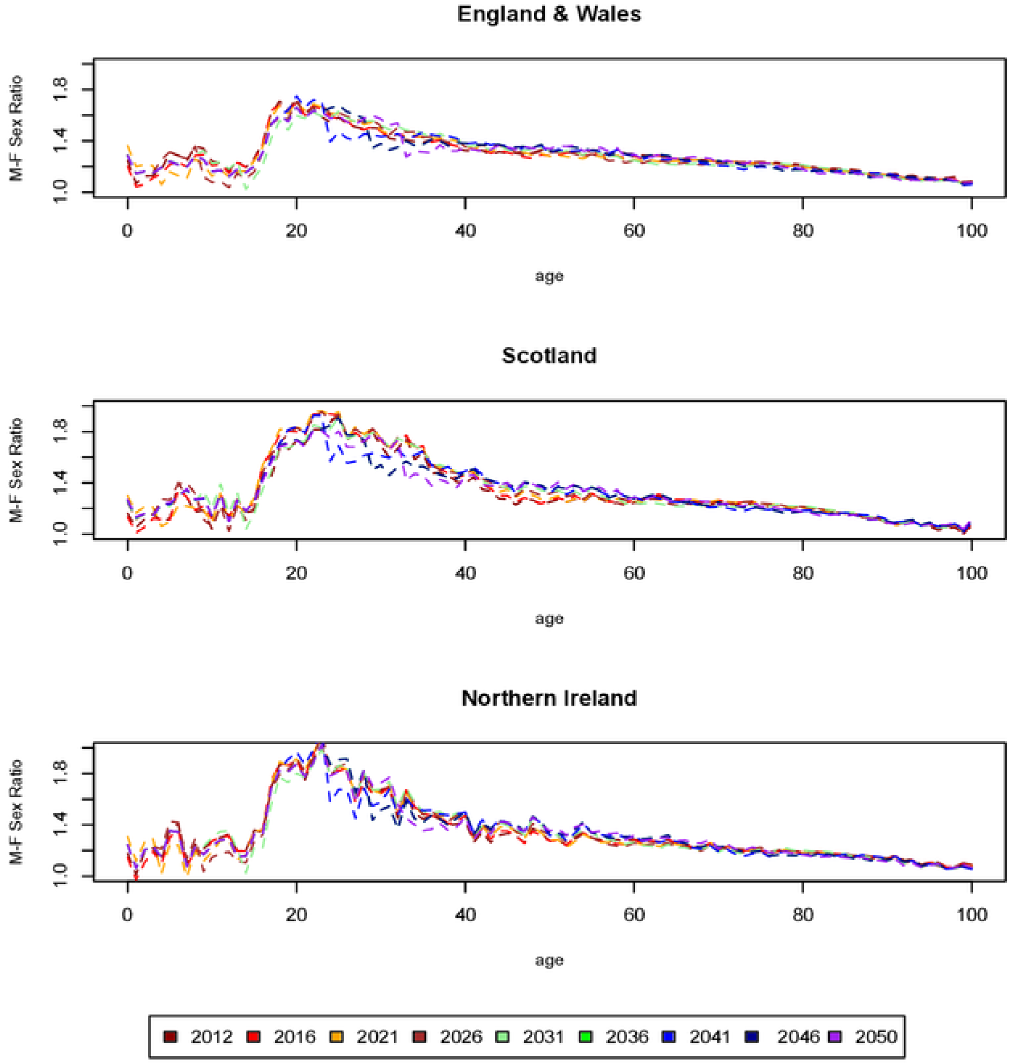


Figure 11 shows the projected life expectancy by the 2-tier ACFC for the six subpopulations of the UK at different ages. At all ages, female life expectancy is consistently higher than male life expectancy, and within each gender, those from England & Wales have highest life expectancy, followed by Northern Ireland and then Scotland (Figure 11a – 11d). At the very old ages, there is some overlapping among male mortality of different countries (Figure 11f). As reflected in the shape of the curves in Figure 11, life expectancy increases, and the trends are inevitably similar among all six subpopulations, since they are primarily dominated by the common factor $B_x K_t$. All subpopulations are also governed by the common trend of each gender, as subpopulations of the same gender is dominated by both mortality improvement trend $b_{x,i}\, k_{t,i}$ and cohort factor $g_{t-x,i}$. It is also visible from Figure 11 that, except for the oldest ages, there is a converging trend between the two genders and also among three countries of the same gender. Therefore, the 2-tier ACFC not only factors in the cohort effect of the UK elegantly, as evidenced in Section 5.2, but also maintains the 2-tier ACF's property to produce coherent long-term forecasts.

Figure 11: 2-tier ACFC Projected Life Expectancy

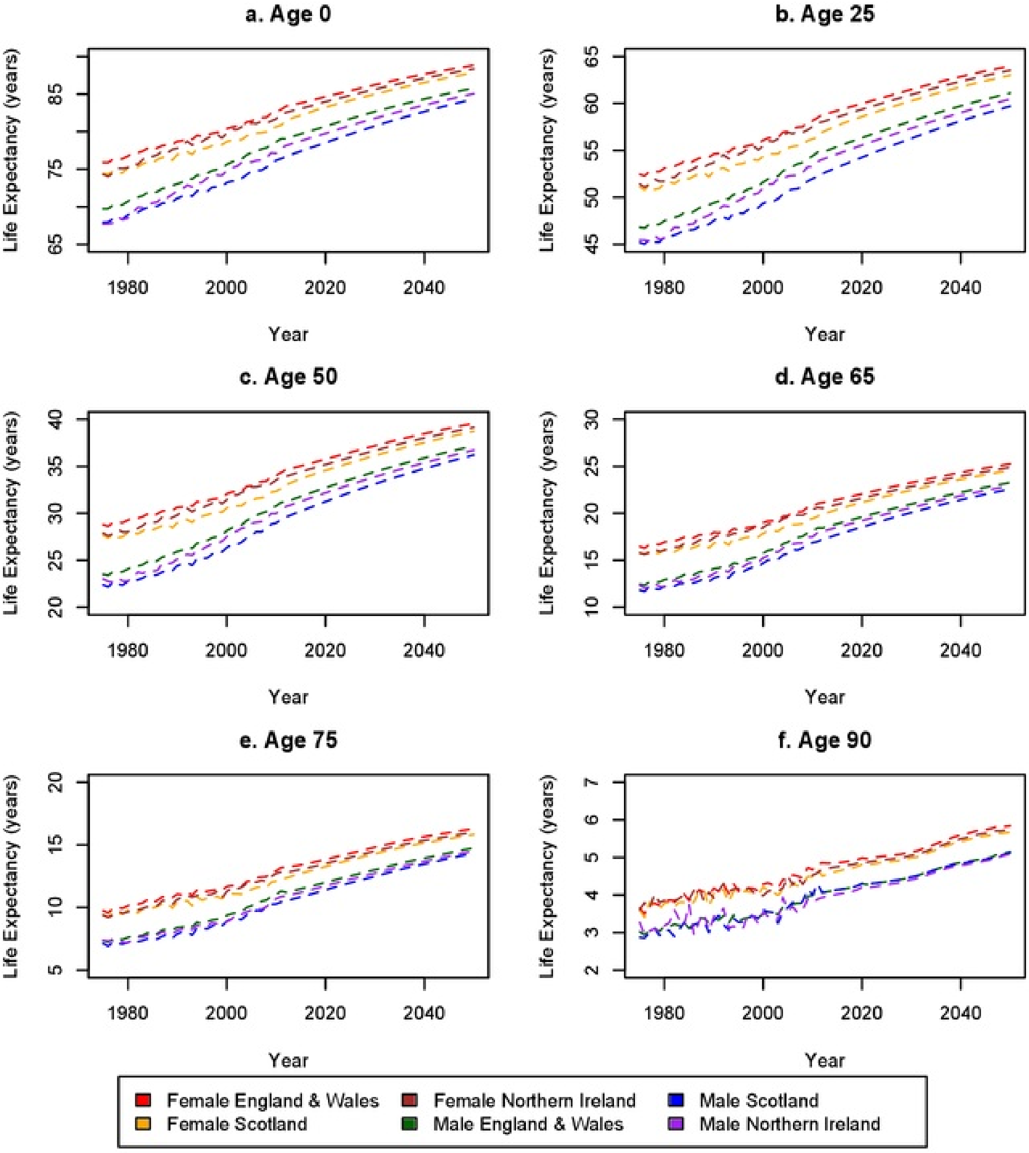


Table 10 below compares the projected life expectancy at birth of the 2-tier ACFC to the published ONS projection for year 2037 (Office for National Statistics 2013). It can be seen that the results are fairly similar between the two, although the 2-tier ACFC projects slightly higher life expectancy for Scotland and lower life expectancy for Northern Ireland. The forecasted life expectancy for England & Wales combined by the 2-tier ACFC is halfway between the ONS estimates for England and Wales. Considering the 2-tier ACFC implicitly gives higher weighting to England than Wales, the 2-tier ACFC estimates for England & Wales are actually lower than the weighted average of ONS estimates. This again shows that the 2-tier ACFC produces a set of reasonable mortality rates while emphasizes more on narrowing regional inequalities in mortality as compared to other official projection methods.

Table 10: Comparison to ONS Life Expectancy Projections

2037 Life Expectancy at Birth (in Years)		
	ONS	**2-tier ACFC**
Females		
England	87.6	87.3
Wales	87.0	87.3
Northern Ireland	86.9	86.7
Scotland	85.5	86.1
Males		
England	84.4	83.9
Wales	83.6	83.9
Northern Ireland	83.3	83.0
Scotland	82.0	82.1

5. Further Discussion and Conclusion Remarks

5.1 General Application in Pension Funds and Life Insurance Industry

Mortality/longevity risk has long been an area of concern for pension funds and life insurers. Underestimating mortality leads to greater than expected liabilities for annuity business and defined-benefit pension funds, while overestimating mortality leads to overstating the solvency position in life insurers' protection books.

While the industry is developing hedging tools such as index-linked longevity swaps, trying to hedge mortality/longevity risk, basis risk still exists because individual funds' mortality experience differs from the population mortality index on which the derivatives are based. Recent research have developed tools and methodologies to model the extent of basis risk by two-populations models such as "M7-M5" and "CAE + cohorts" (Haberman et al. 2014). The research carried out by Li & Hardy (2011) has proved that the ACF framework performs better in modelling related populations and basis risk in longevity swaps over independent LC models and a couple of other LC variations. However, the potential of ACF framework in basis risk modelling may be limited for small populations due to lack of data to estimate the bilinear terms (Haberman et al. 2014).

As suggested by Hyndman et al. (2013) when introducing their product-ratio method, which can be seen as a generalization of the ACF model, if mortality ratios of subpopulations to an aggregated population can be established with confidence, then applying these ratios to a standard table of the aggregated population may be useful in setting assumptions for a subpopulation with missing data. Lack of data is a practical issue faced by actuaries, and applying high-level adjustments to standard tables has long been a tradition. The high-level adjustments are often set in a broad-brushed way, such as taking the ratio of latest mortality rates of a subpopulation to a total population, and then adjust all projected future mortality rates by this constant ratio. Such approach is intuitively flawed, as it fails to take into account the trends of ratios into the future. For example, if the ratio of a subpopulation to a total population is actually decreasing in an annuity portfolio, neglecting the trends will cause underestimation of liabilities. The 2-tier ACF/ACFC framework provides a very simple way of creating more reasonable ratios in projecting mortality, as is demonstrated via the example below.

Suppose a start-up life insurance company plans to focus its sales in country j, and they are setting their mortality assumptions. The external standard table for insured lives is usually for the UK as a whole, split by gender, such as tables produced by CMI. We can possibly use the 2-tier ACFC framework to project ratios based on the population data, then apply these ratios to the projection based on standard tables of insured lives for the entire population (e.g. CMI tables). If we assume the mortality of the whole population with gender i follows (19) and the mortality of all lives with gender i in country j follows (21), we end up with (22), where all the definitions are as in Section 2:

$$\frac{m_{x,t,i,j}}{m_{x,t,i}} \approx \exp\left[(a_{x,i,j} - a_{x,i}) + b_{x,i,j}\ k_{t,i,j}\right]. \tag{22}$$

If we assume the ratio in (22) approximates the ratio of insured lives in country j to insured lives in the whole population, both with gender i, then the ratio can be applied to the standard mortality tables after adjustments are made to reflect time period t in the future. In this way we can estimate the insured lives mortality experience in country j, with trends in the differences between subpopulation in country j and the total population extrapolated. Meanwhile, this ratio above also provides some insights when insurers set their country loadings in their prices. Any flat price loadings should take into account the future period when the price will be offered, and the ratios above converges to some constant in the long term. Insurers should take into account both this long-term mean and short-term behaviours of $k_{t,i,j}$ in setting their country loadings. Simplicity of the ACF may make assumption setting a more intuitive exercise.

Furthermore, another merit of applying the 2-tier ACFC model is that the cohort adjustment is specified at the UK level for each gender. Companies are more likely to have sufficient data to fit and extrapolate the cohort factors aggregately rather than fitting cohort factors to each individual country. The CMI database, for example, is sufficient to fit cohort factors for each gender, but not further split down to different countries. Moreover, because cohort factors are fitted at an aggregated level, so (22) above does not have a cohort component, which further simplifies the ratios of a specific country to the total population under the 2-tier ACFC framework.

Another potential application of the ACF (ACFC) models is to understand the gender gap in mortality. As discussed earlier on, the ACF and the ACFC models, both 1-tier and 2-tier, have emphasized on the mean-reverting property of male-female differences in mortality. The design of all four models mandates that the projected mean differences in log-scale mortality to converge to a reasonably small positive constant to make sure there is no crossover between the two genders. The long-term mean of this difference may be useful in setting assumptions; especially after gender price discrimination was banned in 2012 (European Commission 2012). Knowing the long-term mean difference between genders implies that insurers only need sufficient information on gender mix in the policies sold after 2012 to set the unisex pricing assumptions confidently, especially for long-term policies.

5.2 Limitations of ACF/ACFC models

In this section, some drawbacks of the ACF/ACFC models are discussed. No model could form a perfect reflection of real-life experience. While these models generate coherent and relatively accurate predictions of future mortality experience, they are still guesses with limitations.

Firstly, the method fundamentally belongs to the class of models described as "extrapolative", so it can only capture trends well embedded in the historical data and lack the ability to project more up-to-date information such as medical factors, environmental factors and social-economic changes. For example, a new treatment for cancer or the increasing female workforce participation (Hudson 2007).

Secondly, all the ACF/ACFC models are extensions or modifications of the LC model, of which a major issue is that it neglects the age-time interaction. The LC model assumes rate of mortality change b_x, $b_{x,i}$, and $b_{x,i,j}$ all remain constant over time, whereas substantial age-time interactions have been identified in actual experience (Lee & Miller 2001). This results in the fact that the models tend to underestimate life expectancy. Carter & Prskawetz (2001) proposed a possible extension to the LC model to account for the changing age sensitivity to mortality improvement by applying the LC model to successive subsamples of the fitting period to account for structural changes in b_x's.

Another issue of the ACF framework is that it assumes homogeneity at different levels. When B_x K_t is fitted, homogeneity is assumed for all lives aged *x* in year *t*, but when $b_{x,i}$ $k_{t,i}$ is estimated, homogeneity is assumed for all lives aged *x* in year *t* with the same gender, and the assumption is further relaxed when the model is extended to the country dimension. It should be noted that homogeneity assumptions were embedded in the basic LC model, and methods to allow heterogeneity into the framework has been suggested by Li et al. (2009).

Throughout the project, we propose to fit AR(1) or random walk to all the mortality time index k's instead of other higher order ARIMA model for simplicity purpose, which may exclude models that may fit better to past experience. Moreover, the mortality indices in the model have been extrapolated independently. Despite the fact that $k_{t,i}$ and $k_{t,i,j}$ may be correlated statistically and a vector approach may further improve the model fitting (Hyndman et al. 2013), they are extrapolated independently to avoid overly complicating the model. Moreover, if a vector approach is taken, correlations among time indices would have to be estimated, compromising simplicity of the model. Similarly, AR(1) was chosen to fit cohort factors in the ACFC models, and the cohort models of the two genders are extrapolated independently. Although historically, females and males show different cohort patterns in their mortality improvements, there could be interactions between the cohort effects of two genders, since females and males born in the same year are inevitably exposed to similar social-economic context and healthcare facilities. Therefore, it may make more sense to fit and extrapolate the cohort factors using a vector approach.

Most of the results generated in this project are point estimates for future mortality rates. Further research should look into the statistical errors of estimates, which are primarily driven by standard errors of parameters in fitting the mortality time indices.

5.3 Conclusions

We have extended the ACF model by Li and Lee (2005) to a 2-tier structure in order to model subpopulations of different genders and countries jointly and coherently. A Poisson structure similar to Li (2012) is applied to introduce a robust statistical framework to evaluate the accuracy of model fitting. Results show that the 2-tier ACF model is superior to the independent LC model in terms of model fitting, short-term prediction accuracy, and the coherence property in long-term mortality forecasts. A cohort extension is further added to the 2-tier ACF to construct the 2-tier ACFC, which improves the model fitting and short-term prediction accuracy, maintains the coherent property, and removes systematic patterns displayed in the residual plots against cohort for the 2-tier ACF. The 2-tier ACFC model has great potential in forming mortality assumptions for pension funds and insurance companies focusing sales in a specific country or setting uni-sex prices for life insurance products.

References

Akaike, H., 1974. A New Look at the Statistical Model Identification. *IEEE Transactionson Automatic Control*, 19(6), pp. 716–723.

Bevan, G., Karanikolos, M., Exley, J., Nolte, E., Connolly, S. & Mays, N., 2014. *The four health systems of the United Kingdom: how do they compare?*, Available at: http://www.nuffieldtrust.org.uk/sites/files/nuffield/140411_four_countries_health_systems_full_report.pdf.

Booth, H., Hyndman, R.J., Tickle, L. & de Jong, P., 2006. Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research*, 15, pp. 289–310.

Booth, H., Maindonald, J. & Smith, L., 2002. Applying Lee-Carter under conditions of variable mortality decline. *Population studies: A Journal of Demography*, 56(3), pp. 325–336.

Booth, H., Tickle, L. & Smith, L., 2005. Evaluation of the variants of the Lee-Carter method of forecasting mortality: A multi-country comparison. *New Zealand Population Review*, 31(1), pp. 13–34.

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), pp. 373–393.

Butt, Z. & Haberman, S., 2009. *Actuarial Research Paper Ilc : A Collection of R Functions for Fitting a Class of Lee-Carter Mortality Models using Iterative Fitting Algorithms*, Available at: https://www.cass.city.ac.uk/__data/assets/pdf_file/0019/37180/190ARP.pdf.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A. & Balevich, I., 2007. A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *Pensions Institute Discussion Paper PI-0701*.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D. & Khalaf-Allah, M., 2011. Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, 41(1), pp. 29–59.

Carter, L. & Prskawetz, A., 2001. *Examining structural shifts in mortality using the Lee-Carter method*, Available at: http://www.demogr.mpg.de/Papers/working/wp-2001-007.pdf.

Continuous Mortality Investigation Mortality Committee, 2005. *Continuous Mortality Investigation Working Paper 15 - Projecting future Mortality: Towards a proposal for a stochastic methodology*, Available at: http://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/3fb42309a60823afc12570b5002dece7/$FILE/wp15.pdf.

COSLA & The Scottish Government, n.d. Public Information Leaflet: Free Personal & Nursing Care. Available at: http://www.gov.scot/Resource/Doc/305166/0095748.pdf.

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), pp. 291–308.

European Commission, 2012. Factsheet : EU rules on gender-neutral pricing in insurance. Available at: http://ec.europa.eu/justice/gender-equality/files/unisex_insurance_en.pdf.

Haberman, S., Kaishev, V., Millossovich, P., Villegas, A., Baxter, S., Gaches, A., Gunlaugsson, S. & Sison, M., 2014. *Longevity Basis Risk A Methodology for Assessing Basis Risk*, Available at: http://www.actuaries.org.uk/research-and-resources/documents/longevity-basis-risk-methodology-assessing-basis-risk.

Haberman, S. & Renshaw, A., 2009. On age-period-cohort parametric mortality rate projections. *Insurance: Mathematics and Economics*, 45(2), pp. 255–270.

Hudson, R., 2007. Mortality projections and unisex pricing of annuities in the UK. *Journal of Financial Regulation and Compliance*, 15(2), pp. 166–179.

Hunt, A. & Villegas, A.M., 2015. Robustness and convergence in the Lee-Carter model with cohorts. *Insurance: Mathematics and Economics*, 64, pp. 186–202.

Hyndman, R.J., Booth, H. & Yasmeen, F., 2013. Coherent Mortality Forecasting: The Product-Ratio Method With Functional Time Series Models. *Demography*, 50(1), pp. 261–283.

Kalben, B.B., 2000. Why men die younger: Causes of mortality differences by sex. *North American Actuarial Journal*, 4(4), pp. 83–111.

Lee, R. & Carter, L., 1992. Modelling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419), pp. 659–671.

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

Li, J., 2012. A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies: A Journal of Demography* , pp. 1–16.

Li, J.S.-H., Hardy, M. & Tan, K.S., 2009. Uncertainty in Mortality Forecasting. *ASTIN Bulletin*, 39(1), pp. 137–164.

Li, J.S.-H. & Hardy, M.R., 2011. Measuring Basis Risk in Longevity Hedges. *North American Actuarial Journal*, 15(2), pp. 177–200.

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

Longevity Science Advisory Panel, 2012. *Life expectancy: Past and future variations by gender in England & Wales LSAP Paper 2*, Available at: http://www.longevitypanel.co.uk/_files/life-expectancy-by-gender.pdf.

Mccartney, G., Collins, C., Walsh, D. & Batty, D., 2011. *Accounting for Scotland's Excess Mortality : Towards a Synthesis*, Available at: http://www.gcph.co.uk/publications/238_accounting_for_scotlands_excess_mortality_towards_a_synthesis.

Office for National Statistics, 2013. *Historic and Projected Mortality Data from the Period and Cohort Life Tables, 2012-based, UK, 1981-2062*, Available at: http://www.ons.gov.uk/ons/rel/lifetables/historic-and-projected-data-from-the-period-and-cohort-life-tables/2012-based/stb-2012-based.html.

Parr, N., Li, J. & Tickle, L., 2014. *The Cost of Living Longer: Projections of the Effects of Prospective Mortality Improvement on Economic Support Ratios for Fourteen More Advanced Economies*, Available at: http://epc2014.princeton.edu/papers/140226.

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

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

Richards, S.J., Kirkby, J.G. & Currie, I.D., 2006. The Importance of Year of Birth in Two-Dimensional Mortality Data. *British Actuarial Journal*, 12(01), pp. 5–38.

Schwarz, G., 1978. Estimating the Dimension of a Model. *The Annuals of Statistics*, 6(2), pp. 461–464.

Su, S. & Sherris, M., 2012. Heterogeneity of Australian population mortality and implications for a viable life annuity market. *Insurance: Mathematics and Economics*, 51(2), pp. 322–332.

The Human Mortality Database. Available at: http://www.mortality.org/ [Accessed June 28, 2015].

Villegas, A., Kaishev, V. & Millossovich, P., 2015. *StMoMo: An R Package for Stochastic Mortality Modelling*, Available at: http://cran.r-project.org/web/packages/StMoMo/vignettes/StMoMoVignette.pdf.

Willets, B.R.C., 2004. The Cohort Effect : Insights and Explanations. *British Actuarial Journal*, 10(4), pp. 833-877.

Yang, B., Li, J. & Balasooriya, U., 2014. Cohort extensions of the Poisson common factor model for modelling both genders jointly. *Scandinavian Actuarial Journal*, (May), pp. 1-20.