Demographic Research a free, expedited, online journal
of peer-reviewed research and commentary
in the population sciences published by the
Max Planck Institute for Demographic Research
Konrad-Zuse Str. 1, D-18057 Rostock · GERMANY
www.demographic-research.org

DEMOGRAPHIC RESEARCH

VOLUME 25, ARTICLE 5, PAGES 173-214
PUBLISHED 15 JULY 2011
http://www.demographic-research.org/Volumes/Vol25/5/
DOI: 10.4054/DemRes.2011.25.5

Research Article

Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods

Han Lin Shang

Heather Booth

Rob J. Hyndman

Table of Contents

1	Introduction	174
2	Review of mortality forecasting methods	176
2.1	Lee-Carter (LC) method	176
2.2	Lee-Miller (LM) method	177
2.3	Booth-Maindonald-Smith (BMS) method	178
2.4	Hyndman-Ullah (HU) method	178
2.5	Robust Hyndman-Ullah (HUrob) method	180
2.6	Weighted Hyndman-Ullah (HUw) method	180
3	Data sets	181
4	Forecast evaluation	182
5	Comparisons of the point forecasts	182
5.1	Forecast log mortality rates	183
5.2	Forecast life expectancy	189
6	Review of interval forecast methods	194
6.1	LC methods	195
6.2	HU methods	195
6.3	Evaluating interval forecast accuracy	196
7	Comparisons of the interval forecasts	196
7.1	Forecast log mortality rates	196
7.2	Forecast life expectancy	198
8	Discussion	199
8.1	Point forecasts	199
8.2	Comparison with previous findings	203
8.3	Forecast trends in life expectancy	204
8.4	Interval forecasts	206
8.5	Effect of length of fitting period	206
8.6	Effect of robust estimation	207
8.7	Limitations	208
8.8	Implementation	209
9	Acknowledgments	209
	References	210

Demographic Research: Volume 25, Article 5
Research Article

Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods

Han Lin Shang[1]

Heather Booth[2]

Rob J. Hyndman[3]

Abstract

Using the age- and sex-specific data of 14 developed countries, we compare the point and interval forecast accuracy and bias of ten principal component methods for forecasting mortality rates and life expectancy. The ten methods are variants and extensions of the Lee-Carter method. Based on one-step forecast errors, the weighted Hyndman-Ullah method provides the most accurate point forecasts of mortality rates and the Lee-Miller method is the least biased. For the accuracy and bias of life expectancy, the weighted Hyndman-Ullah method performs the best for female mortality and the Lee-Miller method for male mortality. While all methods underestimate variability in mortality rates, the more complex Hyndman-Ullah methods are more accurate than the simpler methods. The weighted Hyndman-Ullah method provides the most accurate interval forecasts for mortality rates, while the robust Hyndman-Ullah method provides the best interval forecast accuracy for life expectancy.

[1] Department of Econometrics and Business Statistics, Monash University, Caulfield East, VIC 3145 Australia. E-mail: HanLin.Shang@monash.edu.

[2] Australian Demographic & Social Research Institute, Australian National University, ACT 0200 Australia. E-mail: Heather.Booth@anu.edu.au.

[3] Department of Econometrics and Business Statistics, Monash University, Clayton, VIC 3800 Australia. E-mail: Rob.Hyndman@monash.edu.

1. Introduction

In recent years, the rapid ageing of the population has been a growing concern for governments and societies. In many developed countries, the concerns are concentrated on the sustainability of pensions and health and aged care systems, especially given increased longevity. This has resulted in a surge of interest among government policy makers and planners in accurately modeling and forecasting age-specific mortality rates. Any improvements in the forecast accuracy of mortality rates would be beneficial for policy decisions regarding the allocation of current and future resources. In particular, future mortality rates are of great interest to the insurance and pension industries.

Several authors have proposed new approaches for forecasting mortality rates and life expectancy using statistical modeling (see Booth 2006; Booth and Tickle 2008, for reviews). Of these, a significant milestone in demographic forecasting was the work of Lee and Carter (1992). They used a principal component method to extract a single time-varying index of the level of mortality rates, from which the forecasts are obtained using a random walk with drift. Since then, this method has been widely used for forecasting mortality rates in various countries, including Australia (Booth, Maindonald, and Smith 2002; De Jong and Tickle 2006), Austria (Carter and Prskawetz 2001), Belgium (Brouhns, Denuit, and Vermunt 2002), Canada (Lee and Nault 1993), Chile (Lee and Rofman 1994), China (Lin 1995), England & Wales (Cairns et al. 2011), Finland (Alho 1998), Japan (Wilmoth 1996), Norway (Keilman, Pham, and Hetland 2002), Spain (Felipe, Guillén, and Pérez-Marín 2002; Debón, Montes, and Sala 2006), Sweden (Lundström and Qvist 2004; Tuljapurkar 2005), the U.K. (Renshaw and Haberman 2003a), the Nordic countries (Koissi, Shapiro, and Högnäs 2006), the seven most economically developed nations (G-7) (Tuljapurkar, Li, and Boe 2000), and United State (Cairns et al. 2011).

The strengths of the Lee-Carter (LC) method are its simplicity and robustness in situations where age-specific log mortality rates have linear trends (Booth et al. 2006). A weakness of the LC method is that it attempts to capture the patterns of mortality rates using only one principal component and its scores. To address this, Hyndman and Ullah (2007) propose a model that utilizes second and higher order principal components to capture additional dimensions of change in mortality rates. Although other methods have been developed (e.g., Renshaw and Haberman 2003a,b,c; Currie, Durban, and Eilers 2004; Bongaarts 2005; Girosi and King 2008; Renshaw and Haberman 2006; Haberman and Renshaw 2008; Ediev 2008), the LC method is often considered as the benchmark method. For example, the LC method is compared with other approaches in Cairns et al. (2011).

The LC method was designed for long-term forecasting on the basis of a lengthy time series of historical data. However, structural changes in mortality patterns have occurred during the twentieth century, reducing the relevance of data from the distant past for current forecasts (Booth, Maindonald, and Smith 2002). If a long fitting period cannot enhance forecast accuracy, the heavy data demands of the LC method can be somewhat relaxed. The question of whether or how the length of the fitting period affects point forecast accuracy was not evaluated until the works of Booth, Tickle, and Smith (2005) and Booth et al. (2006).

Several authors including Tuljapurkar, Li, and Boe (2000), Lee and Miller (2001) and Booth, Maindonald, and Smith (2002) have proposed variants of the LC method. There have also been several extensions of the LC method. Of these, the extension proposed by Hyndman and Ullah (2007) has been receiving increasing attention in the fields of demography and statistics. Their method combines the ideas of nonparametric smoothing, functional principal component regression and functional data analysis, in order to forecast mortality and fertility rates. This method has been applied by Erbas, Hyndman, and Gertig (2007) for forecasting breast cancer mortality rates in Australia. Furthermore, this method has been extended by Hyndman and Booth (2008) to improve the estimation of the variance and to include the forecasting of net migration numbers. Recently, Hyndman and Shang (2009) extended it to allow greater weight to be given to more recent data than data from the distant past.

The forecast accuracy of the LC method and its variants was first evaluated by Booth, Tickle, and Smith (2005), and further studied by Booth et al. (2006). In this article, we extend the results of Booth et al. (2006). First, we evaluate and compare the point forecast accuracy of the mortality rates and life expectancy from ten principal component methods, including methods proposed since the earlier comparison. Then, following the suggestion by Booth et al. (2006), we evaluate the forecast uncertainty of mortality rates and life expectancy. Although many authors have considered the estimation of prediction intervals, particularly for medium- to long-term forecasting (e.g., Alho 1997; Tayman, Schafer, and Carter 1998; Lutz and Goldstein 2004; Alho and Spencer 2005; Brouhns, Denuit, and Van Keilegom 2005; Koissi, Shapiro, and Högnäs 2006; Haberman and Renshaw 2008; Renshaw and Haberman 2008) they do not evaluate and compare the accuracy of prediction intervals. To our knowledge, with the exception of the limited evaluation in Booth, Tickle, and Smith (2005), an empirical comparison of the forecast uncertainty estimates has never been undertaken.

This article is organized as follows: In Section 2, we briefly describe the ten mortality forecasting methods that are included in our comparisons. Section 3 describes the data, while methods of forecast evaluation are presented in Section 4. In Section 5 we compare the relative point forecast accuracy of the ten methods. Section 6 presents the methods for calculating and evaluating interval forecasts, while Section 7 compares relative interval forecast accuracy of the ten methods. The evaluations include both age-specific mortality rates and life expectancy. A detailed discussion appears in Section 8.

2. Review of mortality forecasting methods

In this section, we review the ten methods for forecasting mortality rates and life expectancy, namely the LC method, the LC method without adjustment (LCnone), the Tuljapurkar-Li-Boe method (TLB), the Lee-Miller method (LM), the Booth-Maindonald-Smith method (BMS), the Hyndman-Ullah method (HU), the Hyndman-Ullah method using only data from 1950 onward (HU50), the robust Hyndman-Ullah method (HUrob), the robust Hyndman-Ullah method using only data from 1950 onward (HUrob50), and the weighted Hyndman-Ullah method (HUw).

We use the original notation of Lee and Carter (1992) and extend it as necessary for each method. In order to stabilize the high variance associated with high age-specific rates, it is necessary to transform the raw data by taking the natural logarithm. We denote by $m_{x,t}$ the observed mortality rate at age x in year t calculated as the number of deaths at age x in calendar year t, divided by the corresponding mid-year population aged x. The models and forecasts are all in log scale.

2.1 Lee-Carter (LC) method

The model structure proposed by Lee and Carter (1992) is given by

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

where a_x is the age pattern of the log mortality rates averaged across years; b_x is the first principal component reflecting relative change in the log mortality rate at each age; k_t is the first set of principal component scores by year t and measures the general level of the log mortality rates; and $\varepsilon_{x,t}$ is the residual at age x and year t.

The LC model in (1) is over-parametrized in the sense that the model structure is invariant under the following transformations:

$$\{a_x, b_x, k_t\} \mapsto \{a_x, b_x/c, ck_t\},$$
$$\{a_x, b_x, k_t\} \mapsto \{a_x - cb_x, b_x, k_t + c\}.$$

In order to ensure the model's identifiability, Lee and Carter (1992) imposed two constraints, given as:

$$\sum_{t=1}^{n} k_t = 0, \qquad \sum_{x=x_1}^{x_p} b_x = 1,$$

where n is the number of years and p is the number of ages in the observed data set.

In addition, the LC method adjusts k_t by refitting to the total number of deaths. This adjustment gives more weight to high rates, thus roughly counterbalancing the effect of using a log transformation of the mortality rates. The adjusted k_t is then extrapolated using ARIMA models. Lee and Carter (1992) used a random walk with drift model, which can be expressed as:

$$k_t = k_{t-1} + d + e_t,$$

where d is known as the drift parameter and measures the average annual change in the series, and e_t is an uncorrelated error. It is notable that the random walk model with drift provides satisfactory results in many cases (Tuljapurkar, Li, and Boe 2000; Lee and Miller 2001; Lazar and Denuit 2009). From this forecast of the principal component scores, the forecast age-specific log mortality rates are obtained using the estimated age effects a_x and b_x in (1).

Two other methods are closely related to the original LC method. The first is the LC method without adjustment of k_t, labelled LCnone. The second is the TLB method, which is without adjustment and also restricts the fitting period to 1950 onward.

2.2 Lee-Miller (LM) method

The LM method is a variant of the LC method, and it differs from the LC method in three ways.

1. The fitting period begins in 1950.
2. The adjustment of k_t involves fitting to the life expectancy $e(0)$ in year t.
3. The jump-off rates are the actual rates in the jump-off year instead of the fitted rates.

In their evaluation of the LC method, Lee and Miller (2001) found a mismatch between fitted rates for the last year of the fitting period and actual rates in that year; this jump-off error amounted to 0.6 years in life expectancy for males and females combined (Lee and Miller 2001, p.539). This jump-off error was eliminated by using actual rates in the jump-off year.

In addition, the pattern of change in mortality rates was not constant over time, which is a strong assumption of the LC method. Consequently, the adjustment of historical principal component scores resulted in a large estimation error. To overcome this, Lee and Miller (2001) adopted 1950 as the commencing year of the fitting period due to different age patterns of change for 1900-1949 and 1950-1995. This fitting period had previously been used in the TLB method.

In addition, the adjustment of k_t was done by fitting to observed life expectancy in year t, rather than by fitting to total deaths in year t. This has the advantage of eliminating the need for population data.

2.3 Booth-Maindonald-Smith (BMS) method

As a variant of the LC method, the BMS method differs from the LC method in three ways.

1. The fitting period is determined on the basis of a statistical 'goodness of fit' criterion, under the assumption that the principal component score k_t is linear.
2. The adjustment of k_t involves fitting to the age distribution of deaths rather than to the total number of deaths.
3. The jump-off rates are the fitted rates under this fitting regime.

A common feature of the LC method is the linearity of the best fitting time series model of the first principal component score, but Booth, Maindonald, and Smith (2002) found the linear time series to be compromised by structural change. By first assuming the linearity of the first principal component score, the BMS method seeks to achieve the optimal 'goodness of fit' by selecting the optimal fitting period from all possible fitting periods ending in year n. The optimal fitting period is determined based on the smallest ratio of the mean deviances of the fit of the underlying LC model to the overall linear fit.

Instead of fitting to the total number of deaths, the BMS method fits to the age distribution of deaths using the Poisson distribution to model deaths, and using deviance statistics to measure the 'goodness of fit' (Booth, Maindonald, and Smith 2002). The jump-off rates are taken to be the fitted rates under this adjustment.

2.4 Hyndman-Ullah (HU) method

Using the functional data analysis technique of Ramsay and Silverman (2005), Hyndman and Ullah (2007) proposed a nonparametric method for modeling and forecasting log mortality rates. It extends the LC method in three ways.

1. The log mortality rates are first smoothed using penalized regression splines with a partial monotonic constraint (see Ramsay 1988, for detail). It is assumed that there is an underlying continuous and smooth function $f_t(x)$ that is observed with error at discrete ages. To emphasize that age, x, is now considered as a continuous variable and incorporating the log transformation, we use $m_t(x)$ rather than $\ln m_{x,t}$ to represent the log mortality rates for age $x \in [x_1, x_p]$ in year t. Then, we can write

$$m_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \quad i = 1, \dots, p, \ t = 1, \dots, n \tag{2}$$

   where $m_t(x_i)$ denotes the log of the observed mortality rate for age x_i in year t; $\sigma_t(x_i)$ allows the amount of noise to vary with x_i in year t, thus rectifying the assumption of homoscedastic error in the LC model; and $\varepsilon_{t,i}$ is an independent and identically distributed standard normal random variable.

2. More than one principal component is used. Higher order terms of the principal component decomposition improve the LC model because these additional components capture non-random patterns, which are not explained by the first principal component (Booth, Maindonald, and Smith 2002; Renshaw and Haberman 2003b; Koissi, Shapiro, and Högnäs 2006). Using functional principal component analysis (FPCA), a set of curves is decomposed into orthogonal functional principal components and their uncorrelated principal component scores. That is,

$$f_t(x) = a(x) + \sum_{j=1}^{J} b_j(x) k_{t,j} + e_t(x), \tag{3}$$

   where $a(x)$ is the mean function estimated by $\hat{a}(x) = \frac{1}{n}\sum_{t=1}^{n} f_t(x)$; $\{b_1(x), \ldots, b_J(x)\}$ is a set of the first J functional principal components; $\{k_{t,1}, \ldots, k_{t,J}\}$ is a set of uncorrelated principal component scores; $e_t(x)$ is the residual function with mean zero; and $J < n$ is the number of principal components used. Following Hyndman and Booth (2008), we chose $J = 6$, which should be larger than any of the components required. The conditions for the existence and uniqueness of $k_{t,j}$ are discussed by Cardot, Ferraty, and Sarda (2003).

3. A wider range of univariate time series models may be used to forecast the principal component scores. By conditioning on the observed data $\boldsymbol{\mathcal{I}} = \{m_1(x), \ldots, m_n(x)\}$ and the set of functional principal components $\boldsymbol{B} = \{b_1(x), \ldots, b_J(x)\}$, the h-step-ahead forecast of $m_{n+h}(x)$ can be obtained by:

$$\hat{m}_{n+h|n}(x) = \mathrm{E}[m_{n+h}(x)|\boldsymbol{\mathcal{I}}, \boldsymbol{B}] = \hat{a}(x) + \sum_{j=1}^{J} b_j(x)\hat{k}_{n+h|n,j},$$

   where $\hat{k}_{n+h|n,j}$ denotes the h-step-ahead forecast of $k_{n+h,j}$ using a univariate time series model, such as the optimal ARIMA model selected by the automatic algorithm of Hyndman and Khandakar (2008), or an exponential smoothing state space model (Hyndman et al. 2008). We have tried both methods, finding only a marginal difference in forecast accuracy. In this paper, we use the exponential smoothing state-space models to forecast principal component scores.

The HU50 method is identical to the HU method except that the fitting period is restricted to 1950 onward.

2.5 Robust Hyndman-Ullah (HUrob) method

Since the presence of outliers can seriously affect the performance of modeling and forecasting, it is important to identify possible outliers and eliminate their effect. The robust HU method utilizes the reflection based principal component analysis (RAPCA) algorithm of Hubert, Rousseeuw, and Verboven (2002) to obtain projection-pursuit estimates of principal components and their associated scores. It then calculates the integrated squared error for each year, and this provides a measure of the accuracy of the principal component approximation for each year (Hyndman and Ullah 2007). Outlying years would result in a larger integrated squared error than the critical value obtained by assuming normality of $e_t(x)$ (see Hyndman and Ullah 2007, for details). By assigning zero weight to outliers, we can then apply the HU method to model and forecast mortality rates without possible influence of outliers.

The HUrob50 method is identical to the HUrob method except that the fitting period is restricted to 1950 onward.

2.6 Weighted Hyndman-Ullah (HUw) method

The HUw method adopts the same smoothing technique as the HU method, but it uses geometrically decaying weights in the estimation of $a(x)$ and $b_j(x)$, thus allowing these quantities to be based more on recent data than on data from the distant past.

The HUw method differs from the HU method in three ways.

1. The weighted functional mean $a^*(x)$ is estimated using the weighted average

$$\hat{a}^*(x) = \sum_{t=1}^{n} w_t f_t(x), \tag{4}$$

   where $\{w_t = \lambda(1-\lambda)^{n-t}, t = 1, \dots, n\}$ denotes a set of weights, and $0 < \lambda < 1$ denotes a geometrically decaying weight parameter. Hyndman and Shang (2009) describe how to estimate λ from data.

2. Using FPCA, a set of weighted curves $\{w_t[f_t(x) - \hat{a}^*(x)]; t = 1, \dots, n\}$ is decomposed into orthogonal weighted functional principal components and their uncorrelated principal component scores. That is,

$$f_t(x) = \hat{a}^*(x) + \sum_{j=1}^{J} b_j^*(x) k_{t,j} + e_t(x), \tag{5}$$

   where $\hat{a}^*(x)$ is the weighted functional mean, and $\{b_1^*(x), \dots, b_J^*(x)\}$ is a set of weighted functional principal components.

3. By conditioning on the observed data $\boldsymbol{\mathcal{I}} = \{m_1(x), \dots, m_n(x)\}$ and the set of weighted functional principal components $\boldsymbol{B}^* = \{b_1^*(x), \dots, b_J^*(x)\}$, the h-step-ahead forecast of $m_{n+h}(x)$ can be obtained by:

$$\hat{m}_{n+h|n}(x) = \mathrm{E}[m_{n+h}(x)|\boldsymbol{\mathcal{I}}, \boldsymbol{B}^*] = \hat{a}^*(x) + \sum_{j=1}^{J} b_j^*(x)\hat{k}_{n+h|n,j}.$$

3. Data sets

The data sets used in this study were taken from the Human Mortality Database (2009). Fourteen developed countries were selected, and thus 28 sex-specific populations were obtained for all analyses. The fourteen countries selected all have reliable data series commencing before 1950. Note that it was desirable to use only countries that have data prior to 1950, in order to maintain full and consistent comparisons of the ten methods. The selected countries are shown in Table 1. Age is in single years and we restrict the age range to $0 - 89+$, in order to avoid excessive fluctuations at older ages.

Table 1: Commencing year of the initial fitting period for each country and method

Country	LC	LCnone	TLB	LM	BMS[f]	BMS[m]	HU	HU50	HUrob	HUrob50	HUw
Australia	1921	1921	1950	1950	1953	1954	1921	1950	1921	1950	1921
Canada	1921	1921	1950	1950	1952	1948	1921	1950	1921	1950	1921
Denmark	1835	1835	1950	1950	1948	1948	1835	1950	1835	1950	1835
England	1841	1841	1950	1950	1952	1948	1841	1950	1841	1950	1841
Finland	1878	1878	1950	1950	1954	1954	1878	1950	1878	1950	1878
France	1816	1816	1950	1950	1947	1954	1816	1950	1816	1950	1816
Iceland	1838	1838	1950	1950	1838	1838	1838	1950	1838	1950	1838
Italy	1872	1872	1950	1950	1954	1954	1872	1950	1872	1950	1872
Netherlands	1850	1850	1950	1950	1946	1947	1850	1950	1850	1950	1850
Norway	1846	1846	1950	1950	1951	1948	1846	1950	1846	1950	1846
Scotland	1855	1855	1950	1950	1936	1948	1855	1950	1855	1950	1855
Spain	1908	1908	1950	1950	1952	1952	1908	1950	1908	1950	1908
Sweden	1751	1751	1950	1950	1948	1952	1751	1950	1751	1950	1751
Switzerland	1876	1876	1950	1950	1950	1950	1876	1950	1876	1950	1876

4. Forecast evaluation

We divide each data set into a fitting period and a forecasting period. The commencing year of the fitting periods differs by method, as seen in Table 1 and also described in Section 2. We use a rolling origin as follows: The forecasting period is initially set to be the last 30 years, ending in 2004. Using the data in the fitting period, we compute one-step-ahead and ten-step-ahead point forecasts, and determine the forecast errors by comparing the forecasts with the actual out-of-sample data. Then, we increase the fitting period by one year, and compute one-step-ahead and ten-step-ahead forecasts, and calculate the forecast errors. This process is repeated until the fitting period extends to 2003.

For the BMS method, the initial optimal fitting period is based on the goodness of fit criterion using data up to 1974; Table 1 shows the resulting commencing years. Commencing years differ between the sexes as a result of independently selecting the optimal fitting period. As the fitting period is increased by one year, the optimal fitting period is re-estimated for each increment (commencing years not shown). This may result in substantial changes in the commencing year, which may increase or decrease the length of the fitting period.

Using the *demography* package for R (Hyndman 2011), we calculate the point and interval forecasts of each method, and evaluate and compare their forecast accuracy. To measure point forecast accuracy, we utilize the mean absolute forecast error (MAFE) and the mean forecast error (MFE). The MAFE is the average of absolute errors, |actual − forecast|, and measures forecast precision, regardless of sign. The MFE is the average of errors, (actual − forecast), and is a measure of bias. These measures are used to evaluate forecasts of log mortality rates and life expectancy.

5. Comparisons of the point forecasts

For simplicity, we will refer to the LC method and its variants (all based on a single principal component) as the LC methods, and refer to the HU method and its variants (all using a smoothing technique and several principal components) as the HU methods. Results are presented by country and for two averages: the simple average and a weighted average using weights based on population size in 2004. For each country, the weight is calculated as the country's population size in 2004 divided by the sum of each country's population size and scaled to sum to 14. For males, the weights are 0.89, 1.42, 0.24, 2.33, 0.23, 2.63, 0.01, 2.52, 0.72, 0.20, 0.22, 1.87, 0.40, 0.32 in the country order of Table 4. For females, the weights are 0.87, 1.39, 0.24, 2.33, 0.23, 2.67, 0.01, 2.56, 0.71, 0.20, 0.23, 1.86, 0.39, 0.32.

5.1 Forecast log mortality rates

Tables 2 and 3 provide summaries of the point forecast accuracy based on the MAFEs for one-step-ahead forecasts of mortality rates averaged over different ages and years in the forecasting period for males and females respectively. As measured by the simple and weighted averages of MAFE over countries, the HU methods tend to perform better than the LC methods, and the HUw method performs the best in the male and female data. The HUw method also performs at least as well as any other method in 26 of the 28 populations.

Tables 4 and 5 show the corresponding MFEs for one-step-ahead forecasts. For both male and female rates, all HU methods overestimate mortality consistently for all countries. Among these methods, the HU50 method performs best for male rates, according to both the simple and weighted averages, while the HUw method performs best for female rates. Among the LC methods, there is less consistency. The LM method performs best overall, though the BMS method has the lowest simple average for male rates. The LC method underestimates female rates for all 14 countries and male rates for 11 of the 14 countries.

Figures 1a and 1b show the MAFEs for one-step-ahead forecasts for different methods averaged over unweighted countries and years in the forecasting period for male and female mortality rates respectively. Larger errors occur at younger ages (less than 40 at least) for all methods, reflecting the difficulty in capturing both the nadir of the mortality schedule and the accident hump; errors from the LC method are particularly large at these ages. At older ages, errors from the LCnone method are large relative to all other methods. Contributing to their lower average MAFEs in Tables 2 and 3, the HU methods tend to be more accurate than most LC methods at most ages.

Figures 1c and 1d display the corresponding MFEs. According to this measure, the HU methods exhibit a distinct pattern of bias by age in both the male and female data, indicative of difficulties in forecasting childhood mortality and the accident hump. With the exception of the LM method (see Section 8.1), the LC methods tend to produce generally larger errors with underestimation at early adult ages and overestimation at later adult ages, while patterns vary at childhood ages. The greater underestimation by the LC method in Tables 4 and 5 is seen to stem from marked underestimation at younger ages with only minimal counterbalancing at older ages, particularly for female rates. It is noteworthy that all methods tend to overestimate mortality rates at older ages; in other words, all forecasting methods tend to underestimate the speed of decline in old-age mortality rates.

Table 2: Point forecast accuracy of male log mortality rates by method and country, as measured by the MAFE for one-step-ahead forecasts

Country	LC	LCnone	TLB	LM	BMS	HU	HU50	HUrob	HUrob50	HUw
Australia	0.388	0.183	0.110	0.092	0.093	0.079	0.079	0.105	0.092	0.076
Canada	0.246	0.133	0.088	0.069	0.079	0.060	0.061	0.090	0.077	0.057
Denmark	0.185	0.180	0.152	0.157	0.145	0.127	0.127	0.143	0.136	0.123
England	0.545	0.175	0.085	0.058	0.088	0.062	0.054	0.083	0.063	0.052
Finland	0.445	0.211	0.150	0.156	0.140	0.141	0.133	0.147	0.137	0.131
France	0.450	0.180	0.083	0.054	0.102	0.064	0.050	0.115	0.061	0.050
Iceland	0.328	0.326	0.328	0.407	0.327	0.332	0.350	0.337	0.344	0.335
Italy	0.283	0.168	0.111	0.065	0.111	0.075	0.062	0.113	0.083	0.061
Netherlands	0.148	0.133	0.110	0.093	0.111	0.081	0.078	0.095	0.084	0.078
Norway	0.235	0.171	0.151	0.150	0.140	0.118	0.120	0.133	0.127	0.120
Scotland	0.649	0.215	0.141	0.152	0.131	0.124	0.121	0.130	0.121	0.118
Spain	0.243	0.158	0.112	0.067	0.113	0.065	0.064	0.072	0.075	0.058
Sweden	0.191	0.171	0.142	0.140	0.137	0.118	0.115	0.146	0.123	0.114
Switzerland	0.225	0.183	0.131	0.136	0.131	0.116	0.113	0.132	0.122	0.112
Average	0.326	0.185	0.135	0.128	0.132	0.112	0.109	0.131	0.118	0.106
Weighted average	0.351	0.168	0.103	0.075	0.105	0.074	0.067	0.102	0.080	0.065

Mean is taken over ages and years in the forecasting period.

Table 3: Point forecast accuracy of female log mortality rates by method and country, as measured by the MAFE for one-step-ahead forecasts

Country	LC	LCnone	TLB	LM	BMS	HU	HU50	HUrob	HUrob50	HUw
Australia	0.288	0.179	0.104	0.114	0.103	0.091	0.091	0.104	0.095	0.089
Canada	0.201	0.128	0.074	0.084	0.075	0.069	0.072	0.073	0.073	0.068
Denmark	0.508	0.203	0.175	0.201	0.177	0.159	0.156	0.183	0.161	0.152
England	0.372	0.119	0.083	0.071	0.069	0.066	0.063	0.083	0.069	0.057
Finland	0.636	0.254	0.193	0.213	0.224	0.171	0.173	0.184	0.174	0.168
France	0.517	0.168	0.081	0.066	0.131	0.058	0.059	0.109	0.063	0.055
Iceland	0.381	0.367	0.375	0.414	0.434	0.353	0.358	0.354	0.355	0.343
Italy	0.363	0.138	0.095	0.078	0.098	0.072	0.068	0.103	0.075	0.066
Netherlands	0.377	0.159	0.101	0.114	0.110	0.091	0.089	0.099	0.091	0.088
Norway	0.564	0.204	0.169	0.189	0.199	0.152	0.153	0.180	0.156	0.151
Scotland	0.526	0.203	0.176	0.192	0.166	0.150	0.154	0.174	0.155	0.145
Spain	0.437	0.162	0.130	0.083	0.140	0.072	0.073	0.084	0.082	0.068
Sweden	0.389	0.176	0.145	0.181	0.285	0.147	0.138	0.223	0.141	0.139
Switzerland	0.400	0.232	0.166	0.189	0.172	0.146	0.148	0.157	0.150	0.145
Average	0.426	0.192	0.148	0.156	0.170	0.129	0.128	0.151	0.131	0.124
Weighted average	0.398	0.155	0.102	0.094	0.117	0.080	0.079	0.105	0.084	0.075

Mean is taken over ages and years in the forecasting period.

Table 4: MFEs for one-step-ahead point forecasts of male log mortality rates by method and country

Country	LC	LCnone	TLB	LM	BMS	HU	HU50	HUrob	HUrob50	HUw
Australia	0.311	−0.076	−0.030	−0.013	−0.008	−0.025	−0.016	−0.034	−0.020	−0.024
Canada	0.171	−0.058	−0.035	−0.011	−0.020	−0.017	−0.017	−0.026	−0.023	−0.015
Denmark	0.018	0.030	−0.046	−0.008	−0.024	−0.034	−0.046	−0.037	−0.060	−0.036
England	0.495	−0.065	−0.011	−0.006	0.028	−0.031	−0.014	−0.033	−0.018	−0.021
Finland	0.365	−0.076	−0.024	−0.004	−0.003	−0.064	−0.044	−0.058	−0.048	−0.058
France	0.389	−0.082	−0.028	−0.010	0.020	−0.036	−0.013	−0.041	−0.017	−0.030
Iceland	−0.059	−0.059	−0.061	−0.011	−0.044	−0.153	−0.210	−0.164	−0.196	−0.171
Italy	0.210	−0.033	−0.039	−0.007	0.042	−0.035	−0.015	−0.037	−0.018	−0.027
Netherlands	−0.039	0.018	−0.053	−0.008	−0.028	−0.035	−0.019	−0.033	−0.024	−0.031
Norway	0.011	0.004	−0.053	−0.009	−0.035	−0.041	−0.036	−0.044	−0.047	−0.039
Scotland	0.584	−0.051	−0.000	−0.004	0.010	−0.037	−0.042	−0.039	−0.040	−0.038
Spain	0.192	−0.015	0.003	0.004	0.055	−0.007	−0.027	−0.009	−0.025	−0.009
Sweden	−0.039	0.015	−0.045	−0.010	0.005	−0.055	−0.044	−0.058	−0.050	−0.055
Switzerland	0.151	−0.027	−0.014	−0.009	−0.001	−0.041	−0.043	−0.047	−0.049	−0.041
Average	0.197	−0.034	−0.031	−0.007	0.000	−0.044	−0.042	−0.047	−0.045	−0.043
Weighted average	0.272	−0.045	−0.025	−0.007	0.019	−0.030	−0.020	−0.033	−0.023	−0.025

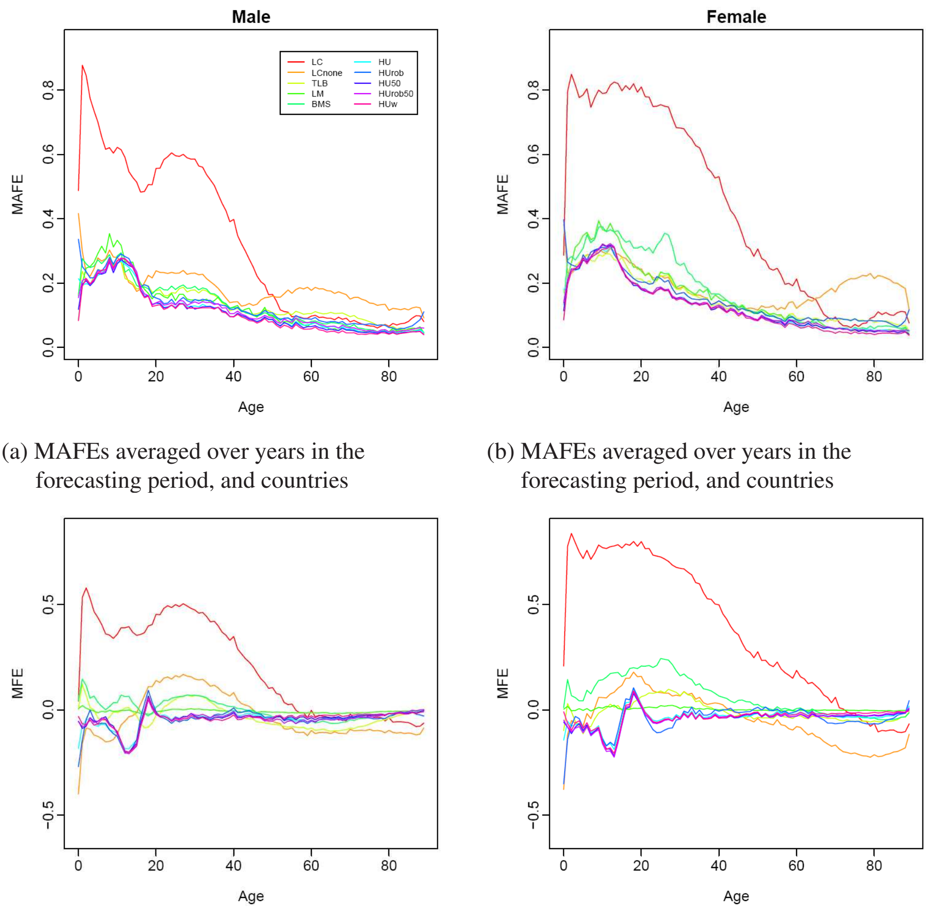
Mean is taken over ages and years in the forecasting period.

Table 5: MFEs for one-step-ahead point forecasts of female log mortality rates by method and country

Country	LC	LCnone	TLB	LM	BMS	HU	HU50	HUrob	HUrob50	HUw
Australia	0.245	−0.057	−0.012	−0.007	0.011	−0.019	−0.020	−0.026	−0.023	−0.022
Canada	0.170	−0.030	0.002	−0.002	−0.024	−0.008	−0.019	−0.009	−0.018	−0.009
Denmark	0.462	−0.033	0.005	0.000	0.009	−0.051	−0.046	−0.055	−0.057	−0.047
England	0.344	−0.037	−0.007	−0.001	−0.004	−0.008	−0.019	−0.013	−0.021	−0.007
Finland	0.591	−0.077	−0.002	0.007	0.098	−0.054	−0.062	−0.058	−0.059	−0.052
France	0.488	−0.077	−0.006	−0.002	0.085	−0.022	−0.013	−0.031	−0.013	−0.016
Iceland	0.100	−0.100	−0.158	0.007	0.082	−0.133	−0.161	−0.131	−0.138	−0.123
Italy	0.332	−0.043	−0.015	0.001	0.046	−0.017	−0.019	−0.019	−0.020	−0.016
Netherlands	0.339	−0.032	0.012	0.001	0.028	−0.019	−0.013	−0.019	−0.016	−0.017
Norway	0.528	−0.037	0.005	0.004	0.103	−0.038	−0.053	−0.050	−0.046	−0.038
Scotland	0.485	−0.045	−0.009	0.006	0.015	−0.050	−0.067	−0.056	−0.063	−0.046
Spain	0.398	−0.051	−0.019	0.010	0.081	−0.004	−0.030	−0.005	−0.032	−0.004
Sweden	0.349	−0.034	0.005	0.001	0.216	−0.063	−0.042	−0.087	−0.041	−0.060
Switzerland	0.350	−0.045	0.006	0.002	0.029	−0.044	−0.050	−0.048	−0.050	−0.041
Average	0.370	−0.050	−0.014	0.002	0.055	−0.038	−0.044	−0.043	−0.043	−0.035
Weighted average	0.364	−0.049	−0.007	0.001	0.045	−0.018	−0.023	−0.023	−0.024	−0.016

Mean is taken over ages and years in the forecasting period.

Figure 1: MAFEs and MFEs for one-step-ahead point forecasts of log mortality rates by sex and method



(a) MAFEs averaged over years in the forecasting period, and countries

(b) MAFEs averaged over years in the forecasting period, and countries

(c) MFEs averaged over years in the forecasting period, and countries

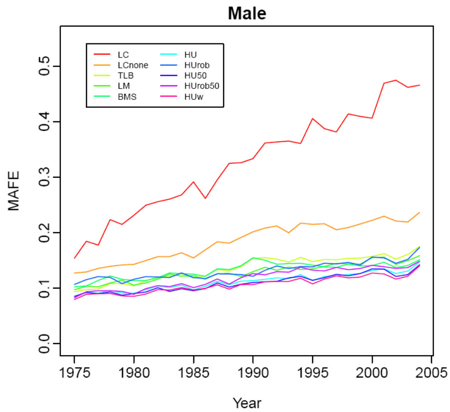
(d) MFEs averaged over years in the forecasting period, and countries

Figures 2a and 2b show the MAFEs for one-step-ahead point forecasts from different methods averaged over countries and ages for male and female mortality respectively. The trend in these measures indicates that it is marginally increasingly difficult to forecast the mortality for later years. Further, differences in the patterns of fluctuation by method indicate that the degree of difficulty is not only dependent on the particular annual mortality experience.

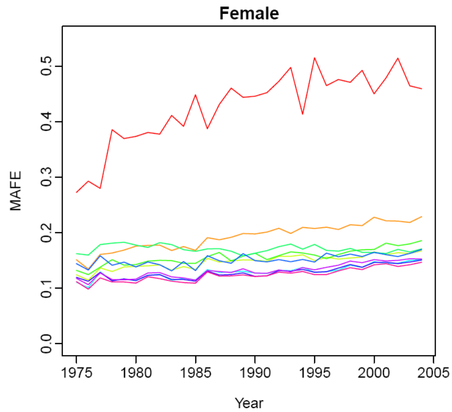
Figures 2c and 2d similarly show the MFEs for one-step-ahead forecasts. For most methods, bias is very small though a slight trend towards overestimation in later years occurs for the HU methods and some LC methods. In contrast, the LC method tends strongly towards underestimation in later years. The LM method displays no trend for either sex. In interpreting these one-step-ahead values, it should be borne in mind that they are significantly influenced by jump-off error and by random error in the particular year (seen in the common pattern of fluctuation). The absence of jump-off error in the LM method helps to explain the relative accuracy of this method (see also Section 8.1).

Figures 2e and 2f show the MFEs for ten-step-ahead forecasts; these refer to years 1984 to 2004 corresponding to fitting periods ending in 1974 to 1994. Compared with the one-step-ahead measures, these are influenced less by jump-off error and more by the trend in the forecast. Nevertheless, the relative position of methods is broadly the same, though the range is wider among the ten-step-ahead MFEs (excluding the LC method). For both males and females, ten-step-ahead forecasts based on fitting periods ending in about 1980-84 tend to be more accurate than those based on periods ending in earlier or later years. This is due to the particular mortality trends over time. Only the LC method does not display this feature. For male mortality, all methods except the LC method tend to overestimate mortality rates at a horizon of ten years regardless of the particular year in the forecasting period. For female mortality, all HU methods tend to overestimate mortality but there is less consistency among the LC methods.

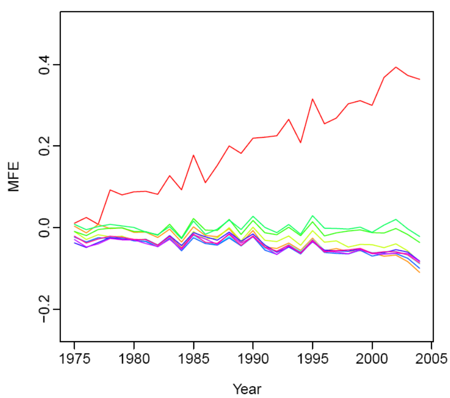
Figure 2: MAFEs and MFEs for point forecasts of log mortality rates by sex and method



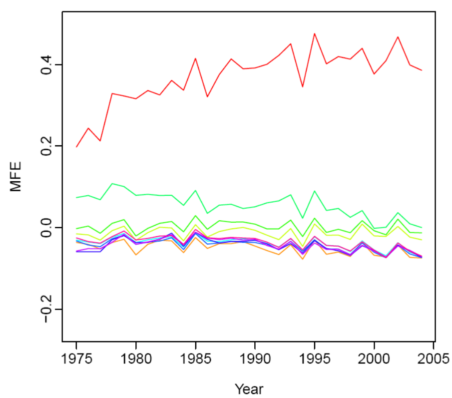
(a) One-step-ahead MAFEs averaged over ages and countries



(b) One-step-ahead MAFEs averaged over ages and countries

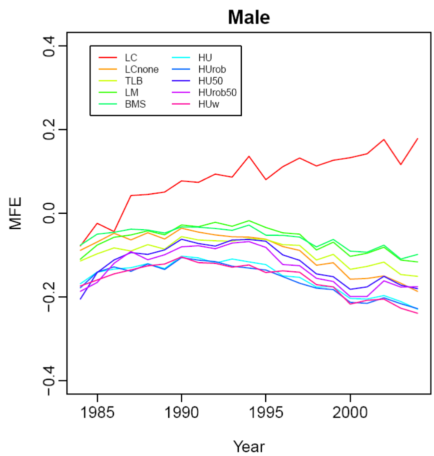


(c) One-step-ahead MFEs averaged over ages and countries

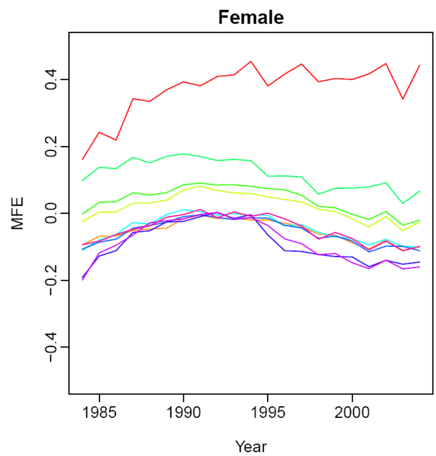


(d) One-step-ahead MFEs averaged over ages and countries

Figure 2: (Continued)



(e) Ten-step-ahead MFEs averaged over ages and countries



(f) Ten-step-ahead MFEs averaged over ages and countries

5.2 Forecast life expectancy

Tables 6 and 7 present the MAFEs for one-step-ahead life expectancy point forecasts averaged over years in the forecasting period for males and females respectively. Based on both the simple and weighted averages, the LM method achieves the best point forecast accuracy for male mortality. This is also the case for female mortality when the simple average is used, but the HUw method is superior according to the weighted average. The LM method is superior among the LC methods, both on average and across almost all countries, while the HUw method is superior among HU methods.

Table 6: Point forecast accuracy of male life expectancy by method and country, as measured by the MAFE for one-step-ahead forecasts

Country	LC	LCnone	TLB	LM	BMS	HU	HU50	HUrob	HUrob50	HUw
Australia	0.616	1.836	0.693	0.282	0.371	0.330	0.313	0.513	0.322	0.286
Canada	0.194	1.326	0.484	0.150	0.263	0.190	0.223	0.395	0.288	0.157
Denmark	0.381	0.380	0.574	0.223	0.420	0.297	0.339	0.560	0.554	0.230
England	0.936	1.837	0.486	0.176	0.235	0.491	0.198	0.633	0.274	0.255
Finland	0.587	1.861	0.580	0.188	0.282	0.545	0.327	0.475	0.412	0.390
France	0.921	2.298	0.285	0.128	0.172	0.522	0.188	1.032	0.210	0.291
Iceland	0.832	0.854	1.032	0.854	0.915	1.512	2.081	1.636	1.921	1.526
Italy	0.608	1.411	0.921	0.201	0.244	0.606	0.229	0.937	0.310	0.286
Netherlands	0.348	0.534	0.769	0.190	0.575	0.327	0.235	0.333	0.284	0.244
Norway	0.606	0.725	0.895	0.212	0.559	0.336	0.329	0.413	0.434	0.259
Scotland	1.303	1.728	0.446	0.204	0.285	0.448	0.357	0.473	0.329	0.246
Spain	0.526	1.264	0.661	0.186	0.254	0.302	0.432	0.441	0.419	0.177
Sweden	0.518	0.657	0.626	0.148	0.235	0.409	0.303	0.626	0.397	0.315
Switzerland	0.472	0.921	0.296	0.178	0.213	0.326	0.283	0.454	0.399	0.297
Average	0.632	1.259	0.625	0.237	0.359	0.474	0.417	0.637	0.468	0.354
Weighted average	0.657	1.554	0.586	0.179	0.266	0.432	0.260	0.677	0.311	0.254

Mean is taken over ages and years in the forecasting period.

Table 7: Point forecast accuracy of female life expectancy by method and country, as measured by the MAFE for one-step-ahead forecasts

Country	LC	LCnone	TLB	LM	BMS	HU	HU50	HUrob	HUrob50	HUw
Australia	0.412	1.493	0.371	0.253	0.275	0.261	0.245	0.308	0.264	0.240
Canada	0.294	1.056	0.351	0.111	0.105	0.147	0.194	0.198	0.192	0.101
Denmark	1.366	1.199	0.543	0.246	0.432	0.203	0.236	0.457	0.251	0.200
England	0.828	1.208	0.279	0.171	0.197	0.189	0.183	0.464	0.207	0.144
Finland	0.901	1.997	0.499	0.211	0.288	0.235	0.285	0.310	0.285	0.216
France	0.983	2.212	0.293	0.186	0.288	0.221	0.208	0.950	0.209	0.154
Iceland	0.655	1.777	2.289	0.869	1.030	1.724	2.074	1.566	1.755	1.393
Italy	0.571	1.655	0.639	0.184	0.247	0.270	0.222	0.526	0.290	0.171
Netherlands	0.811	1.216	0.396	0.177	0.238	0.220	0.206	0.241	0.219	0.172
Norway	1.065	1.254	0.494	0.181	0.242	0.287	0.347	0.586	0.298	0.206
Scotland	1.158	1.442	0.593	0.242	0.293	0.363	0.401	0.781	0.355	0.231
Spain	0.680	1.720	1.031	0.194	0.238	0.245	0.258	0.377	0.330	0.181
Sweden	0.942	0.958	0.325	0.185	0.304	0.287	0.214	1.074	0.213	0.206
Switzerland	0.793	1.575	0.404	0.180	0.244	0.202	0.230	0.308	0.174	0.188
Average	0.818	1.483	0.608	0.242	0.316	0.347	0.379	0.582	0.360	0.272
Weighted average	0.733	1.572	0.488	0.183	0.239	0.229	0.223	0.528	0.248	0.167

Mean is taken over ages and years in the forecasting period.

Corresponding MFEs for one-step-ahead point forecasts of life expectancy are shown in Tables 8 and 9. In general, average underestimation in mortality rates does not necessarily translate into overestimation in life expectancy and vice versa, because of the implicit weights applied to errors by age. However, there is a clear association between differences in the age patterns in forecast errors and differences in the size and sign of forecast errors in life expectancy. For both sexes, all HU methods and the LCnone (in particular) and TLB methods tend to underestimate life expectancy, both on average and almost consistently across countries, while the LC method overestimates life expectancy on average and for most countries. Based on the simple and weighted averages, the LM method is superior for male life expectancy, while for female life expectancy the HUw method is superior according to the weighted average and the BMS method is superior according to the simple average.

Figures 3a and 3b display the MAFEs for one-step-ahead point forecasts of life expectancy by years in the forecasting period for males and females respectively. For both sexes, all methods except LCnone are roughly equally accurate over years in the forecasting period, and LM is the most accurate in most years. For life expectancy, the outlier is LCnone rather than LC. This arises from the relatively large errors at older ages which have a greater effect on life expectancy (at low levels of mortality), whereas the LC adjustment produces relatively large errors at young ages.

Figures 3c and 3d show the MFEs for one-step-ahead point forecasts of life expectancy. For most methods, there is little discernible trend. For LC and LCnone, bias is larger (females) or becomes larger in later years (males): LC tends towards overestimation while LCnone tends more strongly towards underestimation. Again, these MFE values are significantly influenced by jump-off error. Values of MFE for ten-step-ahead, shown in Figures 3e and 3f, mirror the curvature found in MFE for mortality rates. At a horizon of ten years, all forecasting methods tend to underestimate male life expectancy, whereas female life expectancy is likely to be underestimated by most methods, particularly LCnone. The LC, LM and BMS methods tend towards slight overestimation of female life expectancy. It is notable that the LC method performs well overall at this horizon.

Table 8: MFEs for one-step-ahead point forecasts of male life expectancy by method and country

Country	LC	LCnone	TLB	LM	BMS	HU	HU50	HUrob	HUrob50	HUw
Australia	−0.570	1.836	0.687	0.171	0.235	0.256	0.124	0.493	0.228	0.210
Canada	−0.179	1.326	0.484	0.120	0.251	0.161	0.187	0.394	0.265	0.124
Denmark	0.216	0.036	0.527	0.092	0.360	0.253	0.276	0.518	0.497	0.110
England	−0.914	1.837	0.486	0.113	0.151	0.491	0.136	0.633	0.244	0.225
Finland	−0.571	1.861	0.569	0.077	0.166	0.540	0.248	0.471	0.361	0.367
France	−0.921	2.298	0.282	0.079	0.109	0.522	0.137	1.032	0.145	0.291
Iceland	0.447	0.377	0.918	0.589	0.835	1.494	2.081	1.630	1.921	1.468
Italy	−0.047	1.405	0.913	0.123	0.167	0.606	0.177	0.916	0.283	0.257
Netherlands	0.310	−0.213	0.769	0.121	0.555	0.314	0.188	0.212	0.259	0.219
Norway	0.364	0.094	0.895	0.141	0.527	0.309	0.223	0.382	0.361	0.190
Scotland	−1.257	1.728	0.429	0.096	0.068	0.412	0.323	0.456	0.288	0.223
Spain	−0.381	1.264	0.661	−0.008	−0.064	0.247	0.349	0.381	0.374	0.062
Sweden	0.291	−0.311	0.573	0.108	0.197	0.409	0.292	0.626	0.385	0.315
Switzerland	−0.468	0.921	0.187	0.078	0.058	0.291	0.253	0.433	0.368	0.210
Average	−0.263	1.033	0.598	0.136	0.258	0.450	0.357	0.613	0.427	0.305
Weighted average	−0.445	1.472	0.578	0.095	0.161	0.414	0.196	0.655	0.269	0.213

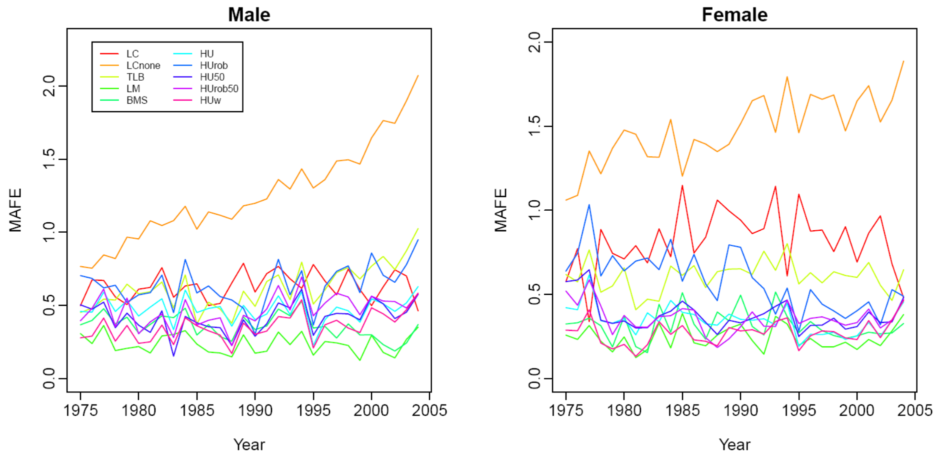
Mean is taken over ages and years in the forecasting period.

Table 9: MFEs for one-step-ahead point forecasts of female life expectancy by method and country

Country	LC	LCnone	TLB	LM	BMS	HU	HU50	HUrob	HUrob50	HUw
Australia	−0.339	1.493	0.275	0.026	0.018	0.043	0.047	0.131	0.081	0.054
Canada	−0.294	1.056	−0.189	−0.078	−0.025	−0.106	−0.032	−0.145	−0.060	−0.073
Denmark	−1.366	1.199	−0.471	−0.076	−0.143	0.100	−0.048	0.436	0.059	0.012
England	−0.828	1.208	0.268	0.008	0.073	0.151	0.107	0.464	0.148	0.026
Finland	−0.901	1.997	0.484	−0.043	−0.079	0.149	0.152	0.266	0.124	0.078
France	−0.983	2.212	0.177	−0.046	−0.206	0.181	0.031	0.905	0.040	0.031
Iceland	0.142	1.663	2.289	0.728	0.978	1.724	2.074	1.562	1.732	1.374
Italy	−0.419	1.655	0.622	−0.008	−0.075	0.238	0.110	0.479	0.154	0.094
Netherlands	−0.811	1.216	−0.252	−0.079	−0.167	−0.027	−0.071	−0.031	−0.123	−0.016
Norway	−1.065	1.254	0.314	−0.029	−0.176	0.236	0.282	0.586	0.204	0.060
Scotland	−1.158	1.442	0.517	−0.003	0.110	0.331	0.368	0.749	0.314	0.108
Spain	−0.587	1.720	1.031	−0.051	−0.028	0.135	0.168	0.248	0.272	−0.036
Sweden	−0.942	0.958	−0.226	−0.066	−0.248	0.280	0.027	0.948	−0.059	0.176
Switzerland	−0.793	1.575	0.040	−0.067	−0.164	0.091	0.086	0.268	0.012	−0.032
Average	−0.739	1.475	0.349	0.015	−0.009	0.252	0.236	0.490	0.207	0.133
Weighted average	−0.688	1.572	0.324	−0.031	−0.069	0.136	0.078	0.429	0.100	0.027

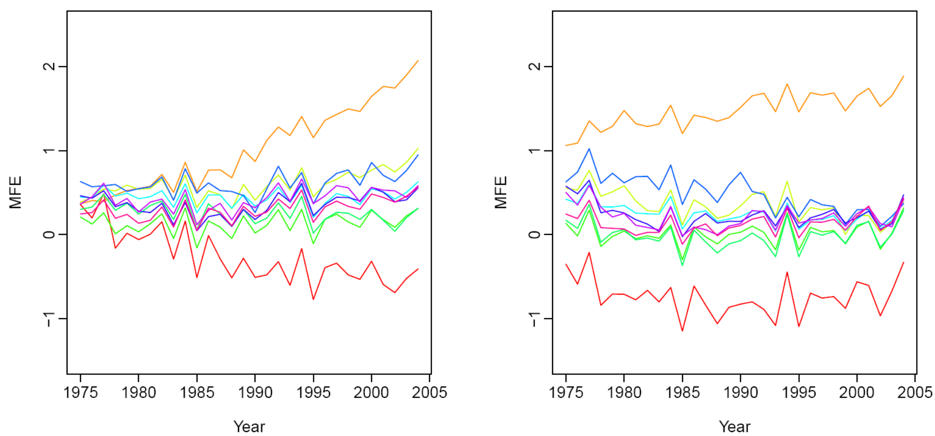
Mean is taken over ages and years in the forecasting period.

Figure 3: MFEs and MAFEs for point forecasts of life expectancy by sex and method



(a) One-step-ahead MAFEs averaged over countries

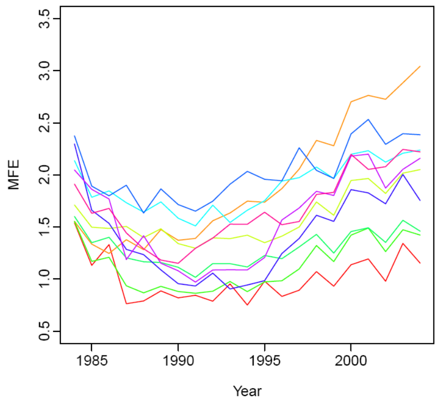
(b) One-step-ahead MAFEs averaged over countries



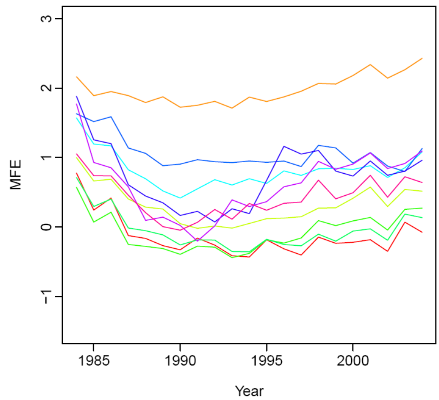
(c) One-step-ahead MFEs averaged over countries

(d) One-step-ahead MFEs averaged over countries

Figure 3: (Continued)



(e) Ten-step-ahead MFEs averaged over countries



(f) Ten-step-ahead MFEs averaged over countries

6. Review of interval forecast methods

Prediction intervals are a valuable tool for assessing the probabilistic uncertainty associated with point forecasts. As was emphasized by Chatfield (1993, 2000), it is important to provide interval forecasts as well as point forecasts, so as to

1. assess future uncertainty levels;
2. enable different strategies to be planned for the range of possible outcomes indicated by the interval forecasts;
3. compare forecasts from different methods more thoroughly; and
4. explore different scenarios based on different assumptions.

In Section 6.1, we first briefly describe a parametric approach to the construction of prediction intervals for the LC methods. In Section 6.2, we review a parametric approach for the HU methods proposed by Hyndman and Ullah (2007). Section 6.3 presents the comparisons of the interval forecast accuracy for both the log mortality rates and life expectancy.

6.1 LC methods

In Lee and Carter (1992), the LC method considers only the uncertainty in the innovations. Although Lee and Carter (1992, p.665) acknowledged that the inclusion of uncertainty in the drift would increase the standard error of their forecasts by 25% after 50 years (based on the period from 1900 to 1989), they did not include it. Booth, Maindonald, and Smith (2002) included the uncertainty in the drift, and compared the variances for the LC, LM and BMS methods with and without this additional uncertainty.

We consider two sources of uncertainty: errors in the parameter estimation of the LC models and forecast errors in the projected time series coefficients. Because of the orthogonality between the first principal component and the error term in (1), the overall forecast variance can be approximated by the sum of the two variances. Conditioning on the observed data $\boldsymbol{\mathcal{I}}$ and the first principal component b_x, we obtained the overall forecast variance of $\ln m_{x,t}$,

$$\text{Var}[\ln m_{x,t}] \approx b_x^2 u_{n+h|n} + v_x,$$

where b_x^2 is the variance of the first principal component; $u_{n+h|n} = \text{Var}(k_{n+h}|k_1, \dots, k_n)$ can be obtained from the time series model; and the model residual variance v_x is estimated by averaging the residual squares $\{\varepsilon_{x,1}^2, \dots, \varepsilon_{x,n}^2\}$ for each x in (1).

6.2 HU methods

The forecast variance follows from (2) and (3). Due to the orthogonality between the principal components and the error term, the overall forecast variance can be approximated by the sum of four variances. Conditioning on the observed data $\boldsymbol{\mathcal{I}}$ and the set of fixed principal components $\boldsymbol{B} = \{b_1(x), \dots, b_J(x)\}$, we obtained the overall forecast variance of $m_{n+h}(x)$,

$$\text{Var}[m_{n+h}(x)|\boldsymbol{\mathcal{I}}, \boldsymbol{B}] \approx \hat{\sigma}_a^2(x) + \sum_{j=1}^{J} b_j^2(x) u_{n+h|n,j} + v(x) + \sigma_{n+h}^2(x),$$

where $\hat{\sigma}_a^2(x)$ (the variance of the smooth estimate $\hat{a}(x)$) can be obtained from the smoothing method used in (2); $b_j^2(x)$ is the variance of the j^{th} principal component; $u_{n+h|n,j} = \text{Var}(k_{n+h,j}|k_{1,j}, \dots, k_{n,j})$ can be obtained from the time series model; the model residual variance $v(x)$ is estimated by averaging $\{e_1^2(x), \dots, e_n^2(x)\}$ for each x; and the observational error variance $\sigma_{n+h}^2(x)$ is estimated by averaging $\{\hat{\sigma}_1^2(x), \dots, \hat{\sigma}_n^2(x)\}$ for each x (Hyndman and Ullah 2007).

By assuming that each of the four sources of uncertainty has a normal distribution and that they are uncorrelated, the $100(1-\alpha)\%$ prediction intervals of $m_{n+h}(x)$ are constructed as $\hat{m}_{n+h|n}(x) \pm z_\alpha \sqrt{\text{Var}[m_{n+h}(x) \mid \boldsymbol{\mathcal{I}}, \boldsymbol{B}]}$, where z_α is the $(1-\alpha/2)$ standard normal quantile.

6.3 Evaluating interval forecast accuracy

The method of evaluation of interval forecast accuracy for mortality rates is as follows: Variances for the LC methods were calculated as described in Section 6.1, and variances for all HU methods were calculated as described in Section 6.2. For each year in the forecasting period, one-step-ahead prediction intervals were calculated at the 0.8 nominal coverage probability, and were then tested against the actual proportion of out-of-sample data that fell within the calculated prediction intervals (Swanson and Beck 1994; Tayman, Smith, and Lin 2007). The empirical coverage probability is defined as the proportion of observations that fall into the calculated 80% prediction intervals, where the denominator is the total number of observations in the forecasting period (for example, 90 ages $\times$ 30 years = 2700 observations). We calculated the coverage probability deviance, which is the absolute difference between the nominal coverage probability and the empirical coverage probability, and use this measure to evaluate the interval forecast accuracy of each method. With a nominal coverage probability of 0.8, the maximum coverage probability deviance is 0.8 (i.e., when the empirical coverage probability is 0), while the minimum coverage probability deviance is 0 (i.e., when the empirical coverage probability is 0.8). Results are presented in terms of the coverage probability deviance.

To obtain prediction intervals for future life expectancies, we simulated the forecast log mortality rates as described in Hyndman and Booth (2008). In short, the simulated forecasts of log mortality rates were obtained by adding disturbances to the forecast time series coefficients which were then multiplied by the fixed principal components. Life expectancy was then calculated for each set of simulated log mortality rates. Prediction intervals were constructed from the 80% percentiles of the simulated life expectancies. In this case, the empirical probability coverage and the coverage probability deviance were based on 30 observations (one for each year in the forecasting period).

7. Comparisons of the interval forecasts

7.1 Forecast log mortality rates

The simple and weighted average coverage probability deviances for male and female log mortality rates are shown in Tables 10 and 11, respectively. Based on both averages, the HUw method performs the best for both male and female data. The results show that all methods tend to underestimate the coverage probability. In other words, the prediction intervals are too narrow. Furthermore, this is consistently the case across countries for all methods except the HUw method, for which an occasional positive result occurs (HU also has one slightly positive result). In addition, we find that the HU methods provide more accurate forecasts of uncertainty than the LC methods.

Of the HU methods, the HU50 and HUrob50 methods perform least well. Their comparison with the HU and HUrob methods indicates that restricting the fitting period to 1950 onward does not provide a more accurate estimate of forecast uncertainty.

Table 10: Coverage probability deviances of forecasted male log mortality rates by method and country

Country	LC	LCnone	TLB	LM	BMS
Australia	0.710(−)	0.710(−)	0.720(−)	0.509(−)	0.605(−)
Canada	0.739(−)	0.667(−)	0.690(−)	0.642(−)	0.710(−)
Denmark	0.577(−)	0.594(−)	0.747(−)	0.735(−)	0.756(−)
England	0.640(−)	0.467(−)	0.663(−)	0.463(−)	0.597(−)
Finland	0.650(−)	0.538(−)	0.602(−)	0.423(−)	0.668(−)
France	0.684(−)	0.409(−)	0.673(−)	0.562(−)	0.640(−)
Iceland	0.421(−)	0.490(−)	0.590(−)	0.710(−)	0.497(−)
Italy	0.465(−)	0.425(−)	0.638(−)	0.427(−)	0.633(−)
Netherlands	0.208(−)	0.274(−)	0.625(−)	0.562(−)	0.535(−)
Norway	0.584(−)	0.580(−)	0.710(−)	0.706(−)	0.746(−)
Scotland	0.642(−)	0.649(−)	0.688(−)	0.642(−)	0.542(−)
Spain	0.567(−)	0.441(−)	0.602(−)	0.404(−)	0.629(−)
Sweden	0.390(−)	0.364(−)	0.668(−)	0.657(−)	0.706(−)
Switzerland	0.591(−)	0.580(−)	0.650(−)	0.678(−)	0.679(−)
Average	0.562(−)	0.514(−)	0.662(−)	0.580(−)	0.639(−)
Weighted average	0.589(−)	0.478(−)	0.658(−)	0.514(−)	0.635(−)
Country	**HU**	**HU50**	**HUrob**	**HUrob50**	**HUw**
Australia	0.226(−)	0.306(−)	0.259(−)	0.324(−)	0.134(−)
Canada	0.389(−)	0.454(−)	0.445(−)	0.464(−)	0.205(−)
Denmark	0.219(−)	0.470(−)	0.236(−)	0.471(−)	0.163(−)
England	0.066(−)	0.358(−)	0.013(−)	0.397(−)	0.144(+)
Finland	0.226(−)	0.510(−)	0.179(−)	0.512(−)	0.164(−)
France	0.116(−)	0.419(−)	0.154(−)	0.381(−)	0.130(+)
Iceland	0.287(−)	0.479(−)	0.294(−)	0.470(−)	0.270(−)
Italy	0.060(−)	0.328(−)	0.037(−)	0.379(−)	0.165(+)
Netherlands	0.167(−)	0.536(−)	0.184(−)	0.516(−)	0.120(−)
Norway	0.222(−)	0.405(−)	0.233(−)	0.442(−)	0.205(−)
Scotland	0.451(−)	0.590(−)	0.433(−)	0.582(−)	0.352(−)
Spain	0.178(−)	0.451(−)	0.133(−)	0.443(−)	0.133(−)
Sweden	0.310(−)	0.626(−)	0.402(−)	0.657(−)	0.032(+)
Switzerland	0.437(−)	0.554(−)	0.440(−)	0.531(−)	0.283(−)
Average	0.239(−)	0.463(−)	0.246(−)	0.469(−)	0.179(−)
Weighted average	0.166(−)	0.413(−)	0.165(−)	0.422(−)	0.153(−)

Mean is taken over different ages and years in the forecasting period. Minus sign indicates that the empirical coverage probability is less than the nominal coverage probability. Plus sign indicates that the empirical coverage probability is greater than the nominal coverage probability. Note that 0.710 is equivalent to 71% in the table.

Table 11: Coverage probability deviances of forecasted female log mortality rates by method and country

Country	LC	LCnone	TLB	LM	BMS
Australia	0.572(−)	0.680(−)	0.633(−)	0.448(−)	0.541(−)
Canada	0.624(−)	0.669(−)	0.544(−)	0.601(−)	0.628(−)
Denmark	0.720(−)	0.572(−)	0.592(−)	0.644(−)	0.668(−)
England	0.704(−)	0.404(−)	0.573(−)	0.338(−)	0.452(−)
Finland	0.731(−)	0.625(−)	0.656(−)	0.622(−)	0.667(−)
France	0.701(−)	0.387(−)	0.583(−)	0.426(−)	0.499(−)
Iceland	0.396(−)	0.540(−)	0.558(−)	0.673(−)	0.476(−)
Italy	0.574(−)	0.269(−)	0.565(−)	0.403(−)	0.473(−)
Netherlands	0.601(−)	0.394(−)	0.439(−)	0.499(−)	0.517(−)
Norway	0.739(−)	0.572(−)	0.625(−)	0.639(−)	0.653(−)
Scotland	0.722(−)	0.617(−)	0.656(−)	0.589(−)	0.547(−)
Spain	0.622(−)	0.437(−)	0.616(−)	0.453(−)	0.639(−)
Sweden	0.641(−)	0.349(−)	0.566(−)	0.546(−)	0.644(−)
Switzerland	0.670(−)	0.592(−)	0.587(−)	0.651(−)	0.664(−)
Average	0.644(−)	0.508(−)	0.585(−)	0.538(−)	0.576(−)
Weighted average	0.646(−)	0.438(−)	0.577(−)	0.455(−)	0.538(−)
Country	**HU**	**HU50**	**HUrob**	**HUrob50**	**HUw**
Australia	0.262(−)	0.307(−)	0.298(−)	0.302(−)	0.046(−)
Canada	0.278(−)	0.388(−)	0.246(−)	0.403(−)	0.079(−)
Denmark	0.326(−)	0.471(−)	0.359(−)	0.486(−)	0.197(−)
England	0.051(−)	0.349(−)	0.120(−)	0.360(−)	0.097(+)
Finland	0.445(−)	0.534(−)	0.423(−)	0.503(−)	0.316(−)
France	0.080(−)	0.434(−)	0.169(−)	0.416(−)	0.137(+)
Iceland	0.307(−)	0.474(−)	0.301(−)	0.460(−)	0.300(−)
Italy	0.003(+)	0.376(−)	0.070(−)	0.383(−)	0.144(+)
Netherlands	0.194(−)	0.430(−)	0.216(−)	0.388(−)	0.161(−)
Norway	0.317(−)	0.429(−)	0.392(−)	0.382(−)	0.129(−)
Scotland	0.373(−)	0.452(−)	0.399(−)	0.482(−)	0.310(−)
Spain	0.139(−)	0.462(−)	0.122(−)	0.484(−)	0.089(−)
Sweden	0.277(−)	0.561(−)	0.283(−)	0.491(−)	0.054(−)
Switzerland	0.395(−)	0.518(−)	0.418(−)	0.507(−)	0.379(−)
Average	0.246(−)	0.442(−)	0.272(−)	0.432(−)	0.174(−)
Weighted average	0.137(−)	0.408(−)	0.178(−)	0.407(−)	0.125(−)

Mean is taken over different ages and years in the forecasting period. Minus sign indicates that the empirical coverage probability is less than the nominal coverage probability. Plus sign indicates that the empirical coverage probability is greater than the nominal coverage probability.

7.2 Forecast life expectancy

Based on the simple and weighted average coverage probability deviances for life expectancy shown in Tables 12 and 13, the HUrob method performs the best for both the male and female data. In most cases, the LC methods produce narrower prediction intervals than the HU methods, resulting in large negative coverage probability deviances relative to the HU methods. The superior performance of the HU methods can be attributed to the consideration of four sources of uncertainty. Most HU methods also have negative average coverage probability deviances, but positive country-specific values are much more frequent. The HUw method is unique in producing prediction intervals that are consistently too wide across all populations. It is noted that restriction of the fitting period to 1950 onwards results in a loss of accuracy for male life expectancy (as might be expected from the similar result for log mortality rates) but not for female life expectancy.

8. Discussion

The above comparative analysis of mortality forecasting methods is the most comprehensive to date. It constitutes an evaluation of point and interval forecasts for log mortality rates and life expectancy based on ten principal component methods and 28 populations. The methods include the LC method and four LC variants, and the HU method, itself an extension of LC method, and four HU variants. The evaluation of point forecasts is an expansion of Booth et al. (2006). The evaluation of interval forecast accuracy is novel in demographic forecasting, though we acknowledge the early contribution of Booth, Maindonald, and Smith (2002); Booth, Tickle, and Smith (2005) in relation to the LC, LM and BMS methods.

8.1 Point forecasts

Our overall findings regarding point forecasts of mortality rates are that the HUw method is more accurate than any other method for one-step-ahead forecasts (Tables 2 and 3), and that the LM method is the least biased (Tables 4 and 5). The success of the HUw method suggests that attributing greater weight to the recent past leads to generally smaller age-specific errors, given that such errors are cumulated in MAFE.

In general, the HU methods lead to smaller MAFE values than the LC methods. This is probably due to two factors. First, the smoothing of mortality rates means that the observational error is treated separately from dynamic changes over time. Second, the additional principal components allow more complicated dynamics to be modelled, rather than the restriction to simple age-specific time trends that result from a single principal component.

Table 12: Coverage probability deviances of forecasted male life expectancy by method and country

Country	**LC**	**LCnone**	**TLB**	**LM**	**BMS**
Australia	0.400(−)	0.800(−)	0.800(−)	0.067(−)	0.267(−)
Canada	0.133(−)	0.800(−)	0.800(−)	0.700(−)	0.700(−)
Denmark	0.400(−)	0.167(−)	0.733(−)	0.633(−)	0.733(−)
England	0.400(−)	0.700(−)	0.800(−)	0.333(−)	0.400(−)
Finland	0.033(−)	0.800(−)	0.800(−)	0.367(−)	0.600(−)
France	0.067(+)	0.800(−)	0.800(−)	0.667(−)	0.733(−)
Iceland	0.167(−)	0.100(−)	0.433(−)	0.533(−)	0.167(−)
Italy	0.067(−)	0.567(−)	0.800(−)	0.167(+)	0.167(+)
Netherlands	0.067(−)	0.033(+)	0.800(−)	0.700(−)	0.700(−)
Norway	0.433(−)	0.400(−)	0.700(−)	0.667(−)	0.800(−)
Scotland	0.700(−)	0.800(−)	0.800(−)	0.600(−)	0.267(−)
Spain	0.367(−)	0.800(−)	0.800(−)	0.267(−)	0.600(−)
Sweden	0.300(−)	0.067(−)	0.733(−)	0.700(−)	0.733(−)
Switzerland	0.033(−)	0.700(−)	0.700(−)	0.667(−)	0.667(−)
Average	0.255(−)	0.538(−)	0.750(−)	0.505(−)	0.538(−)
Weighted average	0.216(−)	0.661(−)	0.793(−)	0.429(−)	0.513(−)
Country	**HU**	**HU50**	**HUrob**	**HUrob50**	**HUw**
Australia	0.067(+)	0.067(+)	0.167(+)	0.100(+)	0.200(+)
Canada	0.033(−)	0.200(−)	0.333(−)	0.233(−)	0.167(+)
Denmark	0.100(−)	0.300(−)	0.067(−)	0.333(−)	0.200(+)
England	0.200(+)	0.100(−)	0.100(+)	0.233(−)	0.200(+)
Finland	0.067(−)	0.567(−)	0.067(+)	0.567(−)	0.100(+)
France	0.200(+)	0.267(−)	0.067(−)	0.067(−)	0.200(+)
Iceland	0.200(+)	0.400(−)	0.200(+)	0.400(−)	0.200(+)
Italy	0.173(+)	0.167(+)	0.171(+)	0.167(+)	0.200(+)
Netherlands	0.067(+)	0.633(−)	0.000(+)	0.600(−)	0.200(+)
Norway	0.067(−)	0.233(−)	0.033(+)	0.367(−)	0.200(+)
Scotland	0.700(−)	0.800(−)	0.600(−)	0.767(−)	0.200(+)
Spain	0.067(−)	0.333(−)	0.000(+)	0.233(−)	0.200(+)
Sweden	0.200(+)	0.700(−)	0.000(+)	0.633(−)	0.200(+)
Switzerland	0.400(−)	0.533(−)	0.400(−)	0.500(−)	0.200(+)
Average	0.181(−)	0.379(−)	0.157(−)	0.371(−)	0.190(+)
Weighted average	0.152(−)	0.261(−)	0.126(−)	0.236(−)	0.195(+)

Mean is taken over different ages and years in the forecasting period. Minus sign indicates that the empirical coverage probability is less than the nominal coverage probability. Plus sign indicates that the empirical coverage probability is greater than the nominal coverage probability.

Table 13: Coverage probability deviances of forecasted female life expectancy by method and country

Country	LC	LCnone	TLB	LM	BMS
Australia	0.167(−)	0.800(−)	0.667(−)	0.000	0.200(−)
Canada	0.067(+)	0.800(−)	0.267(−)	0.133(−)	0.333(−)
Denmark	0.600(−)	0.800(−)	0.000	0.267(−)	0.433(−)
England	0.467(−)	0.767(−)	0.800(−)	0.000	0.067(−)
Finland	0.300(−)	0.800(−)	0.800(−)	0.533(−)	0.433(−)
France	0.267(−)	0.800(−)	0.700(−)	0.200(−)	0.133(−)
Iceland	0.067(−)	0.200(−)	0.233(−)	0.333(−)	0.133(−)
Italy	0.067(−)	0.767(−)	0.800(−)	0.100(−)	0.133(−)
Netherlands	0.400(−)	0.800(−)	0.100(−)	0.100(−)	0.200(−)
Norway	0.567(−)	0.800(−)	0.767(−)	0.433(−)	0.533(−)
Scotland	0.800(−)	0.800(−)	0.800(−)	0.333(−)	0.367(−)
Spain	0.500(−)	0.800(−)	0.800(−)	0.400(−)	0.400(−)
Sweden	0.133(−)	0.767(−)	0.567(−)	0.367(−)	0.333(−)
Switzerland	0.033(−)	0.800(−)	0.500(−)	0.533(−)	0.467(−)
Average	0.317(−)	0.750(−)	0.557(−)	0.267(−)	0.298(−)
Weighted average	0.287(−)	0.788(−)	0.657(−)	0.175(−)	0.217(−)
Country	**HU**	**HU50**	**HUrob**	**HUrob50**	**HUw**
Australia	0.033(−)	0.167(−)	0.133(+)	0.033(+)	0.200(+)
Canada	0.200(−)	0.400(−)	0.300(−)	0.333(−)	0.133(+)
Denmark	0.200(+)	0.100(+)	0.200(+)	0.067(+)	0.200(+)
England	0.200(+)	0.200(+)	0.200(+)	0.200(+)	0.200(+)
Finland	0.033(+)	0.300(−)	0.033(+)	0.167(−)	0.100(+)
France	0.200(+)	0.033(+)	0.133(+)	0.033(+)	0.200(+)
Iceland	0.200(+)	0.033(−)	0.200(+)	0.033(−)	0.200(+)
Italy	0.200(+)	0.067(+)	0.033(+)	0.133(+)	0.200(+)
Netherlands	0.167(+)	0.067(+)	0.167(+)	0.067(+)	0.167(+)
Norway	0.200(+)	0.033(−)	0.100(+)	0.133(+)	0.200(+)
Scotland	0.033(+)	0.300(−)	0.067(−)	0.333(−)	0.133(+)
Spain	0.100(+)	0.000	0.033(+)	0.100(+)	0.200(+)
Sweden	0.200(+)	0.233(−)	0.167(+)	0.133(−)	0.200(+)
Switzerland	0.133(+)	0.133(+)	0.067(+)	0.133(+)	0.167(+)
Average	0.150(+)	0.148(−)	0.131(+)	0.136(+)	0.179(+)
Weighted average	0.167(+)	0.127(−)	0.127(+)	0.129(+)	0.188(+)

Mean is taken over different ages and years in the forecasting period. Minus sign indicates that the empirical coverage probability is less than the nominal coverage probability. Plus sign indicates that the empirical coverage probability is greater than the nominal coverage probability.

The success, in terms of bias, of the LM method stems from the cancellation in MFE of errors across age, years in the forecasting period, and country (if applicable). Cancellation across age (and country) is seen to be advantageous (Figures 2c and 2d), while cancellation across years in the forecasting period (and country) is seen to be even more advantageous (Figures 1c and 1d). However, the cancellation of one-step-ahead errors over years in the forecasting period occurs to a greater degree in the LM method than in any other method because of the simple linear time series model and the absence of jump-off error. Thus the superior performance of the LM method with respect to bias is mainly an artefact of averaging over one-step-ahead errors based on a rolling origin, and may not be true in other circumstances (e.g., longer forecast horizons). Among the nine remaining methods, the BMS method performs best for male mortality and the TLB method for female mortality.

For life expectancy, the cancellation of errors across age occurs in its calculation, thus affecting MAFE as well as MFE, prior to averaging over years in the forecasting period and country. Further, for the LM method, averaging over years in the forecasting period is less directly instrumental in reducing bias (than in the case of rates). On balance, the LM method is ranked first on both MAFE and MFE for male life expectancy, while the HUw method is ranked first on both MAFE and MFE for female life expectancy based on weighted averages.

Regarding the direction of bias, all HU methods show a tendency towards one-step-ahead overestimation of mortality rates (Tables 4 and 5), and a related tendency towards underestimation of life expectancy, particularly for male mortality (Tables 8 and 9). Similar findings occur for the LCnone and TLB methods. In contrast, the LC method exhibits a strong tendency towards one-step-ahead underestimation of mortality rates, usually producing the overestimation of life expectancy, especially for female mortality. The small LM biases tend towards overestimation in male mortality and underestimation in female mortality. For ten-step-ahead forecasts of mortality rates, there is a tendency towards greater overestimation in all HU methods, and for male mortality in all LC methods.

These biases are better understood in terms of age patterns of errors (Figures 1c and 1d). For all methods, one-step-ahead errors are smaller at older than younger ages, and overestimation occurs at older ages. No method captures the rapidity of mortality decline at older ages, contributing to the tendency to underestimate life expectancy. The consistent overestimation of mortality across countries by all HU methods (Tables 4 and 5) is clearly related to their remarkably similar age patterns. That this pattern differs from those of the LC methods suggests that they are influenced by a common feature not present in the LC methods; this could be the smoothing of the logarithm of rates or the modeling of second (and higher) order principal components. Further research is needed to examine this finding in greater detail.

The age patterns of errors in the LC methods are clearly related to the different features of these methods. The overall underestimation of mortality by the LC method arises from the very large errors at younger ages, which stem from the fact that the LC adjustment gives greater weight to older ages. This adjustment compensates for the weighting implicit in the use of logarithms, which produces large errors at older ages for the LC-none method. The TLB method also has no adjustment, but the shorter fitting period leads to smaller errors through a more appropriate b_x pattern. A shorter fitting period also contributes to smaller errors for the BMS and LM methods.

8.2 Comparison with previous findings

In broad terms, our analysis supports the main finding of Booth et al. (2006): all nine variants and extensions of the LC method are substantially more accurate (based on MAFE) than the original LC method in forecasting mortality rates, but this is not the case in forecasting life expectancy. However, comparison of our findings with those of Booth et al. (2006) is complicated by several factors. First, we use a rolling origin, from 1975 to 2003, and examine only one-step-ahead MAFE and one- and ten-step-ahead MFE, rather than a fixed origin (1985) and the average of one- to fifteen-step-ahead measures used by Booth et al. (2006). Second, the number of countries is greater, and includes Iceland, a clear outlier in terms of accuracy. Third, the fitting periods are often longer. Fourth, we use data for ages 0 to 89+, rather than 0–95+. Fifth, we base our conclusions on weighted averages across countries, though the simple averages are also shown. It should also be noted that Booth et al. (2006) adopted the $(\text{forecast} - \text{actual})$ definition of errors, necessitating a change of sign in comparing the MFE.

For the BMS method, comparison is particularly difficult. In Booth et al. (2006), the BMS fitting period was fixed once selected; commencing between 1962 and 1976, BMS fitting periods were shorter than for methods commencing in 1950. In our analysis, the initial commencing year varies between 1936 and 1954 (except 1838 for Iceland), no longer producing a short period. Further, the rolling origin and consequent re-estimation of the commencing year can result in abrupt changes. These factors severely limit comparison, not only with Booth et al. (2006), but also with the other methods which are all based on fitting periods that are fixed except for incremental change resulting from the rolling origin. A similar but lesser issue arises for the HUw method, where weights are re-estimated for each origin.

8.3 Forecast trends in life expectancy

The least biased methods for forecasting male and female life expectancy (the LM and HUw methods respectively) both produce underestimation. Moreover, underestimation is greater for the ten-step-ahead forecasts than for the one-step-ahead forecasts, indicating deceleration in the rate of increase in life expectancy. Given findings of linear life expectancy (White 2002; Oeppen and Vaupel 2002) and debate about its continuation (Bengtsson 2003), it is pertinent to compare the best performing method in terms of both accuracy and bias with the naïve linear extrapolation of life expectancy. The linear extrapolation was achieved by applying the random walk with drift (RWD) model:

$$Y_t = c + Y_{t-1} + Z_t, \quad t = 1, \dots, n,$$

where Z_t is a normal i.i.d error with $\mathrm{E}(Z_t) = 0$. Forecasts are given by

$$Y_{n+h} = ch + Y_n, \qquad h = 1, 2, \dots.$$

Computationally, the forecasts are obtained via the `rwd` function in the *forecast* package in R.

Tables 14 and 15 provide a comparison of the RWD extrapolation with the best performing principal component method for male and female life expectancy in terms of MAFE and MFE for one- and ten-step-ahead forecasts. The linear extrapolation is shown on two bases: the full fitting period for each country (RWD), and from 1950 (RWD50).

For male life expectancy, linear extrapolation is more accurate and less biased than the LM method regardless of fitting period. Greater underestimation by the LM method can be attributed to the curvature in forecast life expectancy arising from fixed b_x despite linear k_t. Bias in the linear extrapolations suggests that forecasts based on increases in life expectancy over the longer period will tend to produce overestimation, whereas forecasts based on the shorter period will tend to produce underestimation. The male mortality stagnation of the 1960s, which occurred in many countries, accounts for the underestimation by the RWD50 method.

For female life expectancy, the HUw method outperforms the linear extrapolation on all measures. The linear extrapolation produces overestimation in almost all populations especially when based on the full fitting period. For female mortality, there is a clear advantage in adopting the more sophisticated method.

Compared with the HU and HUrob methods, for which comparison of fitting period length is possible (see Section 8.5), the linear extrapolation is relatively stable with respect to forecast accuracy and bias for country-specific forecasts based on the two fitting periods. This may be advantageous in practical applications. The obvious disadvantage in directly forecasting life expectancy is that forecast rates are not made available.

Table 14: MAFE and MFE of male life expectancy by country for the best performing principal component method and the RWD model

	MAFE $h=1$			MFE $h=1$			MFE $h=10$		
	RWD	RWD50	LM	RWD	RWD50	LM	RWD	RWD50	LM
Australia	0.265	0.278	0.282	0.122	0.160	0.171	1.011	1.474	1.751
Canada	0.083	0.119	0.150	0.002	0.073	0.120	−0.149	0.625	1.273
Denmark	0.220	0.217	0.223	−0.093	0.054	0.092	−0.987	0.598	1.114
England	0.147	0.164	0.176	0.031	0.091	0.113	0.234	0.910	1.216
Finland	0.169	0.170	0.188	−0.019	−0.000	0.077	−0.433	−0.263	0.751
France	0.125	0.120	0.128	0.067	0.030	0.079	0.585	0.218	0.942
Iceland	0.703	0.700	0.854	−0.037	0.100	0.589	−0.878	0.404	1.435
Italy	0.173	0.166	0.201	−0.098	0.041	0.123	−1.072	0.403	1.480
Netherlands	0.195	0.183	0.190	−0.077	0.090	0.121	−0.950	0.837	1.264
Norway	0.188	0.200	0.212	0.002	0.100	0.141	−0.121	0.953	1.588
Scotland	0.197	0.195	0.204	0.005	0.059	0.096	−0.070	0.494	0.997
Spain	0.249	0.231	0.186	−0.191	−0.154	−0.008	−2.375	−2.091	−0.227
Sweden	0.130	0.137	0.148	0.045	0.082	0.108	0.577	1.042	1.405
Switzerland	0.186	0.175	0.178	−0.074	0.050	0.078	−0.900	0.417	0.850
Average	0.216	0.218	0.237	−0.023	0.055	0.136	−0.395	0.430	1.131
Weighted average	0.168	0.171	0.179	−0.024	0.037	0.095	−0.393	0.253	1.051

The RWD model uses all historical data, whereas the RWD50 model uses data from 1950 onward.

Table 15: MAFE and MFE of female life expectancy by country for the best performing principal component method and the RWD model

	MAFE $h=1$			MFE $h=1$			MFE $h=10$		
	RWD	RWD50	LM	RWD	RWD50	LM	RWD	RWD50	LM
Australia	0.243	0.253	0.240	0.017	0.056	0.054	−0.186	0.238	0.519
Canada	0.150	0.104	0.101	−0.142	−0.065	−0.073	−1.784	−1.009	−0.483
Denmark	0.259	0.238	0.200	−0.147	−0.071	0.012	−1.693	−0.994	−0.069
England	0.171	0.166	0.144	−0.055	0.006	0.026	−0.653	−0.036	0.347
Finland	0.227	0.213	0.216	−0.130	−0.070	0.078	−1.771	−1.233	0.318
France	0.163	0.176	0.154	0.009	−0.056	0.031	−0.030	−0.756	0.376
Iceland	0.480	0.500	1.393	−0.142	−0.000	1.374	−1.994	−0.775	1.839
Italy	0.221	0.171	0.171	−0.162	−0.057	0.094	−1.804	−0.763	0.862
Netherlands	0.214	0.163	0.172	−0.156	−0.063	−0.016	−1.921	−1.050	−0.007
Norway	0.182	0.175	0.206	−0.066	−0.030	0.060	−0.877	−0.517	0.276
Scotland	0.246	0.243	0.231	−0.055	−0.035	0.108	−0.679	−0.495	0.436
Spain	0.249	0.221	0.181	−0.221	−0.171	−0.036	−2.554	−2.133	−0.068
Sweden	0.166	0.174	0.206	−0.011	−0.046	0.176	−0.132	−0.540	1.474
Switzerland	0.200	0.169	0.188	−0.144	−0.050	−0.032	−1.645	−0.703	0.160
Average	0.227	0.212	0.272	−0.100	−0.047	0.133	−1.266	−0.769	0.427
Weighted average	0.198	0.179	0.167	−0.098	−0.055	0.027	−1.194	−0.796	0.323

The RWD model uses all historical data, whereas the RWD50 model uses data from 1950 onward.

8.4 Interval forecasts

A major finding is that all methods underestimate the variability in rates; in other words, on average the prediction intervals are too narrow. That underestimation occurs for almost all populations and methods underlines the robustness of this finding. The HUw method consistently produces wider (only occasionally too wide) predictions intervals than all other methods, and is more accurate on average for mortality rates. This greater accuracy does not extend to life expectancy: for this measure, the HUrob method performs best overall for male and female mortality.

For both mortality rates and life expectancy, the HU methods produce more accurate interval forecasts than the LC methods. The main reason for their narrower prediction intervals is that the LC methods model only two sources of uncertainty, whereas the HU methods model two additional sources of uncertainty (see also Cairns 2000; Alho and Spencer 2005, p.255). First, the HU methods allow more principal components to be included in the model. These additional components capture patterns, which are not explained by the first principal component (Booth, Maindonald, and Smith 2002; Renshaw and Haberman 2003a). Second, the HU methods use a nonparametric smoothing technique, namely penalized regression splines with the partial monotonic constraint, to smooth the noisy log mortality rates. In most cases, this can be expected to reduce the overall variability (Ramsay 1988). In the LC methods, these additional sources of uncertainty are treated as error, with potential counterbalancing.

It should be noted that the prediction intervals for rates and life-expectancy are asymmetric and the coverage probability deviance does not distinguish between the upper and lower prediction limits. Further research is needed to examine the performance of interval forecasts in greater detail.

8.5 Effect of length of fitting period

The effect of length of fitting period on forecast accuracy has been previously discussed (e.g., Lee and Miller 2001; Booth, Maindonald, and Smith 2002; Bengtsson 2003). Booth, Tickle, and Smith (2005) and Booth et al. (2006) found shorter fitting periods, commencing in 1950 or later, to be generally advantageous. Our analysis provides an opportunity to evaluate the effect of truncating the fitting period at 1950, when the longer fitting periods commence in 1751 to 1921, by comparing TLB with LCnone, HU50 with HU, and HU50rob with HUrob. For male mortality rates, truncation at 1950 is advantageous in terms of accuracy and bias (Tables 2 and 4), whereas for female rates (Tables 3 and 5) the advantage is somewhat restricted in that truncation is disadvantageous for bias in the HU methods. For life expectancy, truncation is advantageous for accuracy and bias for both sexes (Tables 6 to 9). These results support the findings of Booth et al. (2006). Overall, there is convincing support in one-step-ahead forecasting for truncation of the fitting period at 1950. At a ten-year horizon (Figures 2e, 2f, 3e and 3f), the advantage remains for male mortality, but is lost for female mortality rates when the relevant HU methods are compared.

The success of the HUw method suggests that even shorter fitting periods may be more advantageous, as found by Booth et al. (2006). The effect of giving greater weight to more recent data is examined by comparison of the HUw method with the HU method. The advantage for the accuracy of forecast rates is clear: in 26 of the 28 populations in Tables 2 and 3, the HUw method is more accurate than the HU method. Moreover, the HUw method is more accurate than the HU50 method in 26 of 28 populations. Similarly, the HUw method is less biased than the HU or HU50 methods in the majority of populations (Tables 4 and 5), and also performs better in terms of the accuracy and bias of life expectancy (Tables 6 to 9). These findings suggest that the ability of the HUw method to respond to the data through population- and period-specific geometrically decaying weights is highly advantageous in one-step-ahead forecasting. By giving the greatest weight to the most recent (jump-off) year, the HUw method reduces jump-off error, contributing to its overall success. However, this advantage is lost after ten years (Figures 2e, 2f, 3e and 3f).

It should be noted that all HU methods involve some degree of weighting in favour of more recent data by virtue of the exponential smoothing models used (Hyndman et al. 2008). Among time series models, only for the random walk with drift does this not apply. As all LC methods use the RWD model, the forecast linear trend in k_t (after adjustment) is determined with equal weight by the first and last years of the particular fitting period. Thus a 50% weight is attributed to data referring to as long ago as 1751. This factor may in part account for the generally better performance of the HU methods.

The finding that it is marginally increasingly difficult over time to forecast next years mortality rates (Figures 2a and 2b) might possibly be interpretable as indicative of a negative relationship between length of fitting period and forecast accuracy. It should be borne in mind, however, that the findings relate to the particular 30-year forecasting period and its relationship to earlier observations. Changing difficulty (in either direction) arises from non-linearity in mortality trends in the 30-year period. Non-linearity in mortality trends also accounts for the patterns in the ten-step-ahead MFEs in Figures 2e and 2f.

8.6 Effect of robust estimation

Comparisons between HUrob and HU and between HUrob50 and HU50 demonstrate the effect of adopting robust estimation procedures to minimise the effect of outliers on point forecasts. The robust method performs less well in all cases: for accuracy and bias, for male and female mortality, and for mortality rates and life expectancy. In each case, the reduction in forecast accuracy due to robust estimation is greatest when the fitting period is longer.

For the accuracy of interval forecasts, the findings are less clear. The robust method either reduces, or has minimal effect on, the accuracy of interval forecasts for mortality rates (Tables 10 and 11) but improves accuracy in three out of four instances for life expectancy (Tables 12 and 13). As already noted, the HUrob method performs best in terms of the accuracy of interval forecasts for life expectancy.

8.7 Limitations

Though including ten methods, this comparative analysis is limited to methods based on the principal component approach. Comparison of these methods with other forecasting approaches is beyond the scope of this paper. Nevertheless the analyses presented here provide a basis for the selection of principal components methods for comparison with other models, such as those considered by Cairns et al. (2011).

The comparison is largely based on one-step-ahead forecasts, though ten-step-ahead forecasts have also been presented for the MFE. The use of one-step-ahead evaluations is standard practice in forecasting, especially when a rolling origin is used, and has the advantage of avoiding confounding horizon with year in the forecasting period. However, the use of one-step-ahead measures does limit interpretation of the results. The influence of jump-off error is particularly strong in the first year of the forecast, limiting the value of one-step-ahead MFE as an indicator of bias in the trend. The effect of error in the trend is better seen at longer horizons and for this reason we include results for the ten-step-ahead MFE.

It must also be acknowledged that the comparison is specific to the particular forecasting period 1975-2004. It is possible that different results would be obtained for a different forecasting period, especially as the period used commences soon after the 1960s when mortality trends were relatively flat.

The conclusions of this comparative analysis are based on sex-specific averages and weighted averages across 14 countries, and do not necessarily hold in all situations. Indeed, examples can be seen in the tables of methods that perform relatively poorly on average but perform well for a particular country, and conversely of methods that perform relatively well on average but perform poorly for a particular country. Further research is needed to understand the circumstances in which this occurs.

It is also acknowledged that some of the comparisons discussed here may not be statistically significant. Statistical tests of the results are problematic because the countries and years included cannot be considered a random sample of any well-defined population of countries and years, and so no tests have been carried out.

8.8 Implementation

Implementation of the methods used in this article is straightforward using the readily-available R package *demography* (Hyndman 2011). The data requirements are historical mortality rates and population numbers in a complete matrix format by age and time. Such data are readily available for developed countries from the Human Mortality Database (2009).

9. Acknowledgments

The authors thank the editor Nico Keilman and two reviewers for their insightful comments. The first author acknowledges the financial support from the postgraduate publication award of Monash University.

References

Alho, J.M. (1997). Scenarios, uncertainty and conditional forecasts of the world population. *Journal of the Royal Statistical Society: Series A* 160(1): 71–85. doi:10.1111/1467-985X.00046.

Alho, J.M. (1998). A stochastic forecast of the population of Finland. Helsinki, Finland: Statistics Finland. (Reviews 1998/4).

Alho, J.M. and Spencer, B.D. (2005). *Statistical Demography and Forecasting*. New York: Springer.

Bengtsson, T. (2003). The need for looking far back in time when predicting future mortality trends. In: Bengtsson, T. and Keilman, N. (eds.) *Perspectives on mortality forecasting, Vol. 1*. Swedish National Social Insurance Agency.

Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. *Demography* 42(1): 23–49. doi:10.1353/dem.2005.0003.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting* 22(3): 547–581. doi:10.1016/j.ijforecast.2006.04.001.

Booth, H., Hyndman, R.J., Tickle, L., and De Jong, P. (2006). Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research* 15(9): 289–310. doi:10.4054/DemRes.2006.15.9.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies* 56(3): 325–336. doi:10.1080/00324720215935.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science* 3(1-2): 3–43. doi:10.1017/S1748499500000440.

Booth, H., Tickle, L., and 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): 13–34.

Brouhns, N., Denuit, M., and Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal* 2005(3): 212–224. doi:10.1080/03461230510009754.

Brouhns, N., Denuit, M., and Vermunt, J.K. (2002). A Poission log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* 31(3): 373–393. doi:10.1016/S0167-6687(02)00185-3.

Cairns, A.J.G. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* 27(3): 313–330. doi:10.1016/S0167-6687(00)00055-X.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics* 48(3): 355–367. doi:10.1016/j.insmatheco.2010.12.005.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* 13(3): 571–591.

Carter, L.R. and Prskawetz, A. (2001). Examining structural shifts in mortality using the Lee-Carter method. Rostock: Max Planck Institute for Demographic Research. (Working paper, 2001-007). http://www.demogr.mpg.de/Papers/Working/wp-2001-007.pdf.

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics* 11(2): 121–135. doi:10.2307/1391361.

Chatfield, C. (2000). *Time-Series Forecasting*. Boca Raton, Florida: Chapman & Hall/CRC. doi:10.1201/9781420036206.

Currie, I.D., Durban, M., and Eilers, P.H.C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4(4): 279–298. doi:10.1191/1471082X04st080oa.

De Jong, P. and Tickle, L. (2006). Extending Lee-Carter mortality forecasting. *Mathematical Population Studies* 13(1): 1–18. doi:10.1080/08898480500452109.

Debón, A., Montes, F., and Sala, R. (2006). A comparison of models for dynamic life tables. Application to mortality data from the Valencia Region (Spain). *Lifetime Data Analysis* 12(2): 223–244. doi:10.1007/s10985-006-9005-1.

Ediev, D.M. (2008). Extrapolative projections of mortality: Towards a more consistent method. Vienna Institute of Demography. (Working paper, 3/2008). http://www.oeaw.ac.at/vid/download/WP2008_03.pdf.

Erbas, B., Hyndman, R.J., and Gertig, D.M. (2007). Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine* 26(2): 458–470. doi:10.1002/sim.2306.

Felipe, A., Guillén, M., and Pérez-Marín, A.M. (2002). Recent mortality trends in the Spanish population. *British Actuarial Journal* 8(4): 757–786.

Girosi, F. and King, G. (2008). *Demographic Forecasting*. Princeton: Princeton University Press.

Haberman, S. and Renshaw, A. (2008). Mortality, longevity and experiments with the

Lee-Carter model. *Lifetime Data Analysis* 14(3): 286–315. doi:10.1007/s10985-008-9084-2.

Hubert, M., Rousseeuw, P.J., and Verboven, S. (2002). A fast method of robust principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems* 60(1-2): 101–111. doi:10.1016/S0169-7439(01)00188-5.

Human Mortality Database (2009). [electronic resource]. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Http://www.mortality.org/ (April 15, 2009).

Hyndman, R.J. (2011). Demography: Forecasting mortality, fertility, migration and population data. [electronic resource]. (with contributions from Heather Booth and Leonie Tickle and John Maindonald, R package version 1.09-1). http://CRAN.R-project.org/package=demography.

Hyndman, R.J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting* 24(3): 323–342. doi:10.1016/j.ijforecast.2008.02.009.

Hyndman, R.J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 27(3).

Hyndman, R.J., Koehler, A.B., Ord, J.K., and Snyder, R.D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer. doi:10.1007/978-3-540-71918-2.

Hyndman, R.J. and Shang, H.L. (2009). Forecasting functional time series (with discussion). *Journal of the Korean Statistical Society* 38(3): 199–221. doi:10.1016/j.jkss.2009.06.002.

Hyndman, R.J. and Ullah, M.S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* 51(10): 4942–4956. doi:10.1016/j.csda.2006.07.028.

Keilman, N., Pham, D., and Hetland, A. (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research* 6(15): 409–454. doi:10.4054/DemRes.2002.6.15.

Koissi, M., Shapiro, A.F., and Högnäs, G. (2006). Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics* 38(1): 1–20. doi:10.1016/j.insmatheco.2005.06.008.

Lazar, D. and Denuit, M.M. (2009). A multivariate time series approach to projected life tables. *Applied Stochastic Models in Business and Industry* 25(6): 806–823.

doi:10.1002/asmb.781.

Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* 87(419): 659–671. doi:10.2307/2290201.

Lee, R.D. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography* 38(4): 537–549. doi:10.1353/dem.2001.0036.

Lee, R.D. and Nault, F. (1993). *Modeling and forecasting provincial mortality in Canada*. Paper presented at the World Congress of the International Union for the Scientific Study of Population, Montreal, Canada.

Lee, R.D. and Rofman, R. (1994). Modeling and projecting mortality in Chile. *Notas de Poblacion* 22(59): 183–213.

Lin, J. (1995). Changing kinship structure and its implications for old-age support in urban and rural China. *Population Studies* 49(1): 127–145. doi:10.1080/0032472031000148286.

Lundström, H. and Qvist, J. (2004). Mortality forecasting and trend shifts: An application of the Lee-Carter model to Swedish mortality data. *International Statistical Review/Revue Internationale de Statistique* 72(1): 37–50. doi:10.1111/j.1751-5823.2004.tb00222.x.

Lutz, W. and Goldstein, J.R. (2004). Introduction: How to deal with uncertainty in population forecasting? *International Statistical Review/Revue Internationale de Statistique* 72(1): 1–4. doi:10.1111/j.1751-5823.2004.tb00219.x.

Oeppen, J. and Vaupel, J.W. (2002). Broken limits to life expectancy. *Science* 296(5570): 1029–1031. doi:10.1126/science.1069675.

Ramsay, J.O. (1988). Monotone regression splines in action. *Statistical Science* 3(4): 425–441. doi:10.1214/ss/1177012761.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis, 2nd*. New York: Springer. doi:10.1002/0471667196.ess3138.

Renshaw, A.E. and Haberman, S. (2003a). Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. *Journal of the Royal Statistical Society: Series C* 52(1): 119–137. doi:10.1111/1467-9876.00393.

Renshaw, A.E. and Haberman, S. (2003b). Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics* 33(2): 255–272. doi:10.1016/S0167-6687(03)00138-0.

Renshaw, A.E. and Haberman, S. (2003c). On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics* 32(3): 379–401. doi:10.1016/S0167-6687(03)00118-5.

Renshaw, A.E. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38(3): 556–570. doi:10.1016/j.insmatheco.2005.12.001.

Renshaw, A.E. and Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling. *Insurance: Mathematics and Economics* 42(2): 797–816. doi:10.1016/j.insmatheco.2007.08.009.

Swanson, D.A. and Beck, D.M. (1994). A new short-term county population projection method. *Journal of Economic and Social Measurement* 20(1): 25–50.

Tayman, J., Schafer, E., and Carter, L. (1998). The role of population size in the determination and prediction of population forecast errors: An evaluation using confidence intervals for subcounty areas. *Population Research and Policy Review* 17(1): 1–20. doi:10.1023/A:1005766424443.

Tayman, J., Smith, S.K., and Lin, J. (2007). Precision, bias, and uncertainty for state population forecasts: An exploratory analysis of time series models. *Population Research and Policy Review* 26(3): 347–369. doi:10.1007/s11113-007-9034-9.

Tuljapurkar, S. (2005). Stochastic forecasts of mortality, population, and pension systems. In: Keilman, N. (ed.) *Perspectives on Mortality Forecasting. II: Probabilistic Models*. Stockholm: Swedish Social Insurance Agency: 65–77.

Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature* 405(6788): 789–792. doi:10.1038/35015561.

White, K.M. (2002). Longevity advances in high-income countries, 1955-96. *Population and Development Review* 28(1): 59–76. doi:10.1111/j.1728-4457.2002.00059.x.

Wilmoth, J.R. (1996). Mortality projections for Japan: A comparison of four methods. In: Caselli, G. and Lopez, A.D. (eds.) *Health and Mortality among Elderly Populations*. Oxford: Clarendon Press: 266–287.