Coherent mortality forecasting with Lee-Carter type models: Adjusting for smoothing and lifespan disparity

Ahbab Mohammad Fazle Rabbi[1], Stefano Mazzuco[1], Heather Booth[2], and Vladimir Canudas-Romo[2]

[1]*Department of Statistical Sciences, University of Padua, Italy*
[2]*School of Demography, Australian National University, Australia*

Abstract

Coherent mortality forecasting methods try to capture the influence of global improvement of health, communication, science on a specific population. The widely used coherent model is an hierarchical, extended form of the Lee-Carter method which assumes an invariant age component and a presumably linear time component to model the joint mortality data of "relational populations". Besides forecast inaccuracy due to estimation procedure, choosing the appropriate reference population remains an arbitrary process. We propose to apply the Lee-Carter method on smoothed mortality rates obtained by *LASSO* type regularization and hence to partially adjust the time component according to observed lifespan disparity to get the common factor of the relational populations. Time variability is also taken into consideration during obtaining the common factor. The reference group for making coherent forecast for a particular population is chosen from the set of available populations on the basis of closest lifespan disparity over time. The proposed methodology generates less forecast errors during out-of-sample forecast period and more optimistic forecast of life expectancy for most of the low-mortality countries. Moreover, choosing the relational populations on the basis of historical pattern of lifespan disparity made the choice of reference populations more systematic and competent.

Keywords: Lee-Carter method; Li-Lee method, Coherent mortality forecasting; Mortality smoothing; Lifespan disparity; LASSO

1 Introduction

Accurate forecasts of mortality and life expectancy are a core requirement for decision making in social, health-care and financial sectors. Fundamental changes of welfare policies of aging societies highly depend on the accurate forecasting of longevity (Booth et al. 2002). Stochastic modeling of mortality forecasting is rapidly gaining recognition in this context (Hyndman and Booth 2008). Among several different approaches, the most prominent method to date is that proposed by Lee and Carter (1992). This method decomposes the differences of log mortality rates and their temporal average into two parts; an invariant age component and a time component. Forecast is obtained by standard time series forecasting the time component. Thus, this model is a simple but powerful probabilistic forecasting approach which does not suffer from over-parametrization.

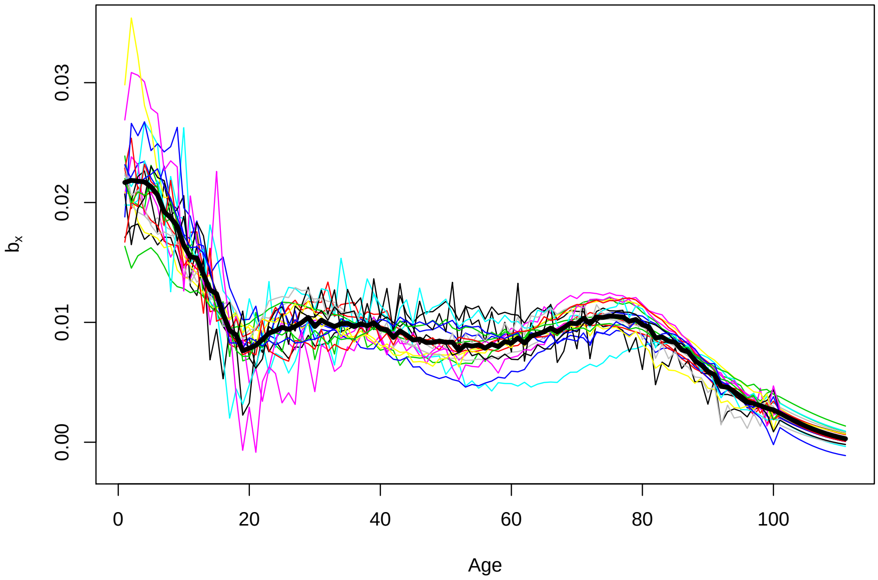
Owing to the straightforward interpretation of the model parameters, many low-mortality countries use different variants of the LC model for mortality forecasting. Among several proposed LC variants, multi-population forecasting is gaining attention as it seeks to ensure that forecasts for related population maintain certain structural relationships based on past mortality patterns and theoretical understanding. The first approach to coherent mortality forecasting was introduced by Li and Lee (2005) as an extended hierarchical interface of the LC method. Li and Lee (2005) modified the traditional LC method for forecasting mortality as the sum of a common trend and population specific rates which converge towards the common trend in the future. They referred to the product of the age and time component from the fitted model on joint mortality of related populations as the "common factor". Other approaches to coherent forecasting have subsequently been developed. Hyndman et al. (2013) extended the nonparametric multiple component approach of Hyndman and Ullah (2007) of which LC is a special case, Ahmadi and Li (2014) used generalized linear modeling and Bergeron-Boucher et al. (2017) used compositional data analysis of the distribution of deaths.

All these approaches are based on basic Lee and Carter (1992), so they also hold the limitations of LC models (Hyndman et al. 2013). Beside that, all coherent forecasting methods share the core problem of choosing an appropriate reference mortality for the population of interest. This includes the choice of an appropriate reference group of populations for the population of interest. The choice of reference depends on several consideration. First, the male-female gap in mortality exists all over the life span for which biological issue remained an important issue for choosing reference population in previous studies (Li and Lee 2005). Second, in the absence of mortality crises due to epidemics or war, mortality forecast can be expected to change steadily in keeping with related or regional populations. Similar economic and political frameworks among a group of countries, such as European Union, may also be taken into account (Kjærgaard et al. 2016). Environmental, geographical location may also have significant impact (Ahcan et al. 2014). Third, consideration should be given to the number of populations combined to form the reference population. Kjærgaard et al. (2016) found that a reference population made of a small number of countries tends to perform better in terms of forecast accuracy than that one comprising a larger group and choosing countries with closer life expectancy was found to be a better strategy for choosing the reference population. These findings of Kjærgaard et al. (2016) are important because choosing reference populations subjectively may introduce bias. However, merely adopting a specific size for the reference group does not make selection easier. Rather, if all countries are to be considered, it creates enormous computational complexities. For a particular population if one looks for the other populations to construct the reference group from another 39 populations, the best reference group of size 5 (including the particular population) would need to be identified from ${}^{39}C_4$= 82,251 possible combinations. Further, although low-mortality convergence is taking place, each populations still has a distinct pattern of mortality improvement. To illustrate, the estimated relative change (b_x) in the log-mortality rate at each age x from fitted Lee-Carter model for females of 20 low-mortality countries are illustrated in Figure 1. Comparing with the mean trend of b_x (marked with thick black line), clearly, each of the populations showed different pattern of mortality improvement in different parts of life span.

Combining populations from different mortality regimes may provide more optimistic forecasts of mortality but clearly it will be brought about by other populations with completely different mortality pattern rather than by population of interest and thus the real situation of the particular population will be dominated. To reduce computation complexities of combining all possible groups of populations, consideration may be given to the development/specification of robust prior assumptions about coherence among population. Further, the combination of populations takes no account of period effects such as the effect of the policy interventions on the reference population and hence the forecast. For instance, rapid and sharp increase may be observed in life expectancies of Portugal after joining European Union (HMD 2018), which is clearly results of adaptation and implementation of health policies of European Union.

In this paper, we propose to apply the Li and Lee (2005) methodology on smoothed mortality

Figure 1: Estimated relative change (b_x) in the log-mortality rate at each age x from fitted Lee-Carter model for females of 20 low-mortality countries (1956:2011)



Source: HMD (2018) and authors' calculations.
Note: The bold black line is the mean trend of b_x to show the distinct mortality improvement patterns.

rates obtained by *LASSO* type regularization and hence to partially adjust the time component of the common factor to match the observed lifespan disparity ($e_0^\dagger$). Smoothing by lasso produces less error during fitting period compared to other spline based smoothing techniques (Dokumentov et al. 2018). Also matching with $e_0^\dagger$ – a more informative indicator of longevity than e_0, made the time component more reflective of countries' mortality patterns (Rabbi and Mazzuco 2018a). Hence the common factor of coherent model is estimated utilizing only a subset of the available years (the best fitting period), and these same years are considered as country-specific fitting period as well. Instead of arbitrary of subjective choice, we formulate the reference group for a particular population by choosing populations having closest lifespan disparity, $e_0^\dagger$. Unlike life expectancy at birth, $e_0^\dagger$ provides more information about "expansion "or "shrinking"of mortality and can be utilized to measure mortality shift (Zhang and Vaupel 2009). Owing to its stability over time, Bohk-Ewald et al. (2017) employed $e_0^\dagger$ to evaluate forecast performances rather than just considering the fitted mortality rates or life expectancy.

2 Methods

Existing coherent forecasting technique: Li and Lee (2005) method

Li and Lee (2005) modified the standard Lee and Carter (1992) model to forecast mortality for several countries by taking into account their membership in a group, rather than forecasting individually. The two-factor LC model is,

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

where, $m_{x,t}$ is the central mortality rate at age x for year t; a_x represents the average of log-mortality at age x over time; b_x is the set of age-specific constants explaining the relative speed of change at each age x; k_t represents the overall level of mortality in year t; and $e_{x,t}$ is the model residual. Singular Value Decomposition (SVD) is used on $Z_{x,t} = [\ln(m_{x,t}) - \hat{a}_x]$ to obtain the OLS estimate of LC model. Symbolically,

$$\text{SVD}(Z_{x,t}) = ULV' = L_1 U_{x_1} V_{t_1} + \dots L_n U_{x_n} V_{t_n}. \tag{2}$$

For estimation of the age and time components, Lee and Carter (1992) considered the rank-1 approximation only as it explains most of the variance. Then the estimates of model parameters are,

$$\hat{k}_t = L_1 V_{t_1} \text{ and, } \hat{b}_x = U_{x_1}.$$

The original LC method incorporates a second stage estimate of k_t by finding the value of k_t which, for a given population age distribution and previously estimated a_x and b_x produces exactly the observed total number of deaths for the fitting period of the model (Lee and Carter 1992). Later variants of the LC method adopted different strategies to adjust the estimated k_t (Lee and Miller 2001; Booth et al. 2002). Most of the LC variants fit an *ARIMA(0,1,0)* to the adjusted $\hat{k}_t$, from which forecast are derived:

$$\hat{k}_t = c + \hat{k}_{t-1} + \xi_t; \tag{3}$$

where c is the drift term and ξ_t is the model residual. To extend the standard LC model to a coherent setting, Li and Lee (2005) first estimated the average mortality trend for the reference group (containing populations of interest and other "related"populations) and addressed it as common factor. Hence they added the historical particularities of particular population (unexplained by the common factor) in a hierarchical way to the standard LC model. Therefore, in the short term, inter-country mortality differences in trends may be preserved, but ultimately age-specific death rates within the group of countries are constrained to maintain a constant ratio to one another (Li and Lee 2005). This extended model can be formulated as,

$$\ln m_{x,t,i} = a_{x,i} + B_x K_t + b_{x,i} k_{t,i} + \epsilon_{x,t,i}, \tag{4}$$

where i denotes the specific country in the group, $a_{x,i}$ is country-specific average log mortality rates. The term B_x is the relative speed of change in mortality at each age x and K_t is the mortality index capturing the main time trend for the group. Li and Lee (2005) called the term $B_x K_t$ the common factor as this quantity is common for all countries of the group. The error term of equation (4) is country-specific. To obtain the country-specific estimates of the Li-Lee model, at first LC model is fitted to joint mortality data (combining all populations in the group including the population of interest) from which the common factor is estimated for use in equation (4). At this stage, K_t is adjusted for observed life expectancy of the joint mortality data. A second SVD is performed on the matrix of differences of country-specific mortality rates and $a_{x,i}$ and the common factor to estimate country-specific $b_{x,i}$ and $k_{t,i}$ without any adjustment of $k_{t,i}$. The forecast is obtained by applying standard time series models to both of the time components. In the current study, a random walk with drift is used for forecasting both for K_t and $k_{t,i}$. Following Lee and Miller (2001) and Li and Lee (2005), we also used the actual data as the jump-off rates to avoid jump–off error.

In the proposed methodology for coherent forecasting, we apply the Li and Lee (2005) method on mortality rates smoothed by *LASSO* type regularization (Dokumentov et al. 2018) and to adjust the fitted time component of the common factor according to observed lifespan disparity, $e_0^\dagger$ (Vaupel and Canudas-Romo 2003; Zhang and Vaupel 2009). Lasso smooths the data in a similar manner to spline-based techniques for the mortality age pattern, but is more accurate. Rabbi and Mazzuco (2018a) applied this methodology before for forecasting mortality in case of single population. In addition, we employed a variable fitting period to address the period effect on coherent forecasting.

2.1 Choice of the reference group

Unlike different subjective approaches or geographic proximity for choosing reference group (see Ahcan et al. 2014, for example), demographic perspectives like similar pattern of life expectancies seemed more realistic in terms of forecast accuracy (Kjærgaard et al. 2016). In our proposed methodology, the reference group for a particular population is chosen on the basis of closest lifespan disparity over time. There are several benefits for considering lifespan disparity in the context of mortality forecasting. For aging societies where survival is highly concentrated around older ages, the difference between the age at death and the expected remaining years decreases. As a result, lifespan disparity gets smaller over time for aging societies, and it may better capture premature mortality unlike life expectancy (Aburto and van Raalte 2018). Lifespan disparity may provide further information about the shrinking or expansion of mortality at different ages and it can be utilized as an alternative indicator of mortality shifting (Zhang and Vaupel 2009). To measure lifespan disparity, we used average number of life years lost at birth (Vaupel and Canudas-Romo 2003; Zhang and Vaupel 2009). Symbolically,

$$e_0^\dagger = \frac{\int_0^\omega e_x d_x \, dx}{l_0} \approx \frac{\sum_0^\omega e_x l_x m_x}{l_0}; \tag{5}$$

where, ω is the maximum attainable age, d_x is the distribution of death and l_x is the number of people alive at age x (l_0 is the life table radix) and m_x is the mortality rate at age x. Thus estimation of $e_0^\dagger$ is simple and straightforward. We base the choice of reference group populations on the observed trend of $e_0^\dagger$ with two assumptions: (i) a smaller number of populations is preferred for the sake of parsimony and (ii) future convergence of mortality among target populations as we did not distinguish male and female separately. For a particular population we choose the populations giving the smallest difference in observed $e_0^\dagger$. For a particular population i, another population j will be in the reference group if

$$\left|\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger\right| = min; \tag{6}$$

compared to other available populations. To implement this, we estimate $e_0^\dagger$ for all populations under consideration over the common fitting period and average over time to obtain the population specific $\bar{e}_0^\dagger$. For m years of available mortality rates and p populations the estimates can be presented in the following matrix notation:

$$\begin{array}{cc}
Year & \begin{array}{cccc} Pop_1 & Pop_2 & \cdots & Pop_p \end{array} \\
\begin{array}{c} t_1 \\ t_2 \\ \vdots \\ \vdots \\ t_{m-1} \\ t_m \end{array} & \begin{pmatrix}
e^\dagger_{0(t_1,1)} & e^\dagger_{0(t_1,2)} & \cdots & e^\dagger_{0(t_1,p)} \\
e^\dagger_{0(t_2,1)} & e^\dagger_{0(t_2,2)} & \cdots & e^\dagger_{0(t_2,p)} \\
\vdots & \vdots & & \vdots \\
e^\dagger_{0(t_{m-1},1)} & e^\dagger_{0(t_{m-1},2)} & \cdots & e^\dagger_{0(t_{m-1},p)} \\
e^\dagger_{0(t_m,1)} & e^\dagger_{0(t_m,2)} & \cdots & e^\dagger_{0(t_m,p)}
\end{pmatrix} \\
 & \begin{array}{cccc} \bar{e}^\dagger_{0(.,1)} & \bar{e}^\dagger_{0(.,2)} & \cdots & \bar{e}^\dagger_{0(.,p)}. \end{array}
\end{array}$$

For a particular population i, we sorted all $\bar{e}^{\dagger}_{0(t,j)}$ in ascending order. Populations j for which $\bar{e}^{\dagger}_{0(t,j)}$ is closest to $\bar{e}^{\dagger}_{0(t,i)}$ is selected into the reference group for population i. The pattern of ranked, temporal $\bar{e}^{\dagger}_{0(i,j)}$ may show inconsistent pattern depending on the available populations. For the low-mortality populations considered in this study, we obtained consistent patterns of sorted $\bar{e}^{\dagger}_{0(i,j)}$ starting from the period of 1982 as the sorted $\bar{e}^{\dagger}_{0(i,j)}$ have almost same pattern since then. Clearly, the number of populations and required gap between $\bar{e}^{\dagger}_{0(t,i)}$ and $\bar{e}^{\dagger}_{0(t,j)}$ may vary among reference groups as not all populations have symmetric distance of $\bar{e}^{\dagger}_{0}$ between each other. Following the findings of Kjærgaard et al. (2016), we did not pre-fix the number of populations in best reference group. Instead, we choose a best reference group consists of less number of populations and providing least forecast accuracies during out-of-sample evaluation period. We did not consider or compare different strategies for choosing reference population. Nevertheless, based on the strong relation between $e^{\dagger}_{0}$ and e_0, similar results may be obtained for a temporal average of life expectancies as well (Vaupel et al. 2011).

2.2 Coherent forecast of mortality rates

After identifying the populations to be used as the reference group, we smoothed the mortality rates of each population using *LASSO* (Dokumentov et al. 2018). The rationale and advantages for using LASSO are illustrated in earlier study (Rabbi and Mazzuco 2018a). In next step, we combined respective populations in the reference group and followed the standard LC methodology to obtain the initial estimates of the age and time component. In this stage, we applied same weight on all populations to overcome problem of combining larger exposure to smaller exposure (also for Li and Lee (2005)),

$$m_{x,t} = \frac{1}{p}\sum_{i=1}^{p} m_{x,i,t}. \tag{7}$$

These $m_{x,i,t}$ are standardized mortality rates, thus it has almost null influence of population size. In this stage, fitted two-factor LC model over joint mortality rates (reference group) will be,

$$\ln(\hat{m}_{x,t}) = \hat{a}_x + \hat{B}_x\hat{K}_t; \tag{8}$$

which is known as common factor model (Li and Lee 2005). Following Lee and Miller (2001), Li and Lee (2005) method made a second stage estimate of K_t by finding the value of K_t which produces exactly the observed life expectancy for the fitting period of the model. Following Rabbi and Mazzuco (2018a), we adjust the estimated K_t by solving the following equation:

$$e^{\dagger}_{0\ observed} = \sum_{0}^{\omega} \exp(\hat{a}_x + \hat{B}_x.K_{t\ adj})e_x l_x/l_0. \tag{9}$$

The e_x and l_x of equation (9) are obtained from life table estimated from the fitted common factor model. After obtaining the adjusted $\hat{K}_t$, we identified the most appropriate period for which this common factor should be considered. Most of the LC variants obtained linear trend of k_t for vast majority of the populations (Lee 2000). We considered this findings for the K_t to obtain the best fitting period by choosing the fitting period which maximize the linearity (Booth et al. 2002). It is also based on the idea that some more distant past history may not be relevant for the future, which is more sensible coherent forecasting. In order to get the best fitting period, we obtain the close approximation of deviance(t) which is equal to $\chi^2(t)$ statistic of the lack of fit in observed distribution of death $D_{x,t}$,

$$\chi^2(t) = \sum_{x}\frac{\left[D_{x,t} - D'_{x,t}\right]^2}{D'_{x,t}}; \tag{10}$$

where $D'_{x,t}$ are fitted deaths which can be obtained from observed exposure $N_{x,t}$ as follow:

$$D'_{x,t} = N_{x,t}\left[\exp(\hat{a}_x + \hat{B}_x.K_{t\ adj})\right]. \qquad (11)$$

Following Booth et al. (2002), the total lack of fit to the log-linear model derives from two sources: the base lack of fit from the log-additive model or LC model with $\hat{K}_{t\ adj}$ and the additional lack of fit from the imposition of the ARIMA(0,1,0) model on $\hat{K}_{t\ adj}$. The base lack of fit for the period S years prior to last year of the fitting period is measured by

$$\chi^2_{\text{logadd}}(S) = \sum_t \chi^2_{\text{logadd}}(t);$$

where the $D'_{x,t}$ are derived from $\hat{K}_{t\ adj}$. For the log-linear model,

$$\chi^2_{\text{loglin}}(S) = \sum_t \chi^2_{\text{loglin}}(t);$$

here the $D'_{x,t}$ are derived from the linear fit of $\hat{K}_{t\ adj}$. This total lack of fit will be greater than or equal to the base lack of fit. According to Booth et al. (2002), to compare $\chi^2_{\text{loglin}}(S)$ and $\chi^2_{\text{logadd}}(S)$ they are divided by the corresponding degrees of freedom to produce mean-χ^2 statistic. For n age categories and m years in the fitting period, the df for $\chi^2_{\text{loglin}}(S)$ is $n(m-2)$ and df for $\chi^2_{\text{logadd}}(S)$ is $(n-1)(m-1)$. The length of S is determined by the extent of the additional lack of fit relative to the total lack of fit. The additional lack of fit will be small for a good fit of the ARIMA(0,1,0) model. The first statistical measure of ratio of the total to base lack of fit can be obtained as,

$$R(S) = \frac{\chi^2_{\text{loglin}}(S)/[n(m-2)]}{\chi^2_{\text{logadd}}(S)/[(n-1)(m-2)]}. \qquad (12)$$

The marginal effect of including one more year in S can be obtained from the ratio of the differences in total and base mean-χ^2 statistics for S and $S+1$,

$$RD(S) = \frac{\left[\chi^2_{\text{loglin}}(S) - \chi^2_{\text{loglin}}(S+1)\right]/n}{\left[\chi^2_{\text{logadd}}(S) - \chi^2_{\text{logadd}}(S+1)\right]/(n-1)}. \qquad (13)$$

Small values of $R(S)$ and $RD(S)$ indicate that the additional lack of fit is relatively small. The best fitting period is identified by the value of S for which $R(S)$ and $RD(S)$ are substantially smaller than corresponding statistics for preceding values of S, indicating that the inclusion of $S-1$ years (and preceding) prior to last year of fitting period in the fitting period results in a relatively large reduction in goodness of fit of the ARIMA model (Booth et al. 2002).

For country-specific coherent forecast, the basic LC model is then fitted to country-specific mortality rates without the common factor. To obtain the country-specific ordinary least square estimates of $b_{x,i}$ and $k_{t,i}$, SVD is performed on

$$Z_{x,i,t} = \ln(m_{x,i,t}) - \hat{a}_{x,i} - \hat{B}_x\hat{K}_{t\ adj}. \qquad (14)$$

The estimation procedure is as before. However, at this stage the LC model is fitted without any adjustment for country-specific $k_{t,i}$ and it is fitted for the best fitting period obtained during estimation of the common factor. A random walk with drift is then fitted to both $\hat{K}_{t\ adj}$ and $\hat{k}_{t,i}$. To eliminate jump-off error, we used the actual data as the jump-of rates for the forecast.

2.3 Forecast accuracy

We consider the following two measures for checking the forecast accuracy of mortality rates:

mean absolute forecast error,

$$\mathrm{MAE} = \frac{1}{(p+1)q}\sum_{r=1}^{q}\sum_{x=0}^{p}\left|y_{x,r} - \widehat{y}_{x,r|r-h}\right|; \tag{15}$$

mean squared forecast error,

$$\mathrm{MSE} = \frac{1}{(p+1)q}\sum_{r=1}^{q}\sum_{x=0}^{p}\left(y_{x,r} - \widehat{y}_{x,r|r-h}\right)^2; \tag{16}$$

and for life expectancy at birth, we consider the mean error,

$$\mathrm{ME} = \frac{1}{q}\sum_{r=1}^{q}\left(\hat{e}_{0,r} - e_{0,r}\right). \tag{17}$$

Here $y_{x,r}$ represents the observed mortality rate for age x in year r and $\widehat{y}_{x,r}$ represents the forecast; $e_{0,r}$ represents the observed life expectancy at birth in year r and $\hat{e}_{0,r}$ represents the forecast. From the available mortality data, we used the last 10 years as the period for forecasting and the previous years as the fitting period. Using the data in the fitting period, we made ten-step-ahead forecasts, and determined the forecast accuracy by comparing the forecasts with the observed data in the hold-out period. Limitation of the forecasting period to ten years was based on the use of data from 1956 in order to avoid the disruptions of epidemics and WWII. We skipped the comparison of smoothing techniques in this paper as it is well documented by Dokumentov et al. (2018) and Rabbi and Mazzuco (2018a). We kept our comparison with Li and Lee (2005) only in this paper as that one is the most used coherent forecast technique till now. Through out the paper we denoted the proposed method as $\mathrm{LL}_{e_0^\dagger}$ and Li and Lee (2005) by LL.

3 Data

The data used in this study came from Human Mortality Database (HMD, 2018). The HMD is one of the best sources in terms of data quality and it strives to provide mortality data for any population for which death registration and census data are virtually complete. We analyzed the mortality of 40 populations, namely the male and female populations of the following 20 low-mortality countries: Australia (AUS), Austria (AUT), Belgium (BEL), Canada (CAN), Denmark (DNK), Finland (FIN), France (FRA), Germany (DEU), Ireland (IRL), Italy (ITA), Japan (JPN), The Netherlands (NLD), New Zealand (NZL), Norway (NOR), Portugal (PRT), Spain (ESP), Sweden (SWE), Switzerland (CHE), United Kingdom (UK) and USA (USA). As data of Germany are not available before 1990, we combined the data for East and West Germany together. Total populations, rather than smaller subpopulations, are considered for France, New Zealand and the United Kingdom. For all of these countries, HMD covers the period 1956 to 2011, which thus defines the data used in coherent setting (required common fitting period for all populations). We did not consider high mortality populations from Central and Eastern Europe to avoid the data quality issues, inconsistent mortality pattern over time and a shorter fitting period (Rabbi and Mazzuco 2018b). The data are available for ages 0 to 110+, and we constructed life tables for those ages. Owing the presence of missing values after age 100 for several years, missing values at ages older than 100 were estimated using the Kannisto model (Thatcher et al. 1998). Details of Kannisto model are attached in the Appendix.

4 Results

4.1 Best reference population

We calculated $\bar{e}_0^\dagger$ and sorted from smallest to largest for all 40 populations for period starting in 1956 to 2011. The ranking of populations remained stable for starting years since 1982. Although future convergence of male and female mortality is assumed, the sorted $\bar{e}_0^\dagger$ are ranked naturally by sex with the lowest values being for females followed by males although some overlapping occurs in the middle of the range. Table 1 shows the sorted values of $\bar{e}_0^\dagger$ for the period 1982 to 2011. We performed this step before smoothing but the ranking remained the same for smoothed data.

Table 1: Sorted $\bar{e}_0^\dagger$ over the period 1982:2011 for the 20 low-mortality countries

Population	ESP_F	CHE_F	JPN_F	SWE_F	FIN_F	ITA_F	AUT_F	NOR_F
$\bar{e}_0^\dagger$	9.888	9.922	9.933	9.969	10.011	10.043	10.101	10.150
Population	DEU_F	NLD_F	FRA_F	BEL_F	AUS_F	IRL_F	PRT_F	UK_F
$\bar{e}_0^\dagger$	10.246	10.326	10.378	10.454	10.467	10.493	10.618	10.791
Population	CAN_F	SWE_M	NLD_M	NZL_F	DNK_F	JPN_M	NOR_M	IRL_M
$\bar{e}_0^\dagger$	10.881	10.911	11.039	11.100	11.103	11.310	11.340	11.341
Population	CHE_M	UK_M	ITA_M	AUS_M	DEU_M	BEL_M	USA_F	DNK_M
$\bar{e}_0^\dagger$	11.464	11.495	11.547	11.648	11.717	11.800	11.802	11.831
Population	CAN_M	AUT_M	ESP_M	NZL_M	FIN_M	FRA_M	PRT_M	USA_M
$\bar{e}_0^\dagger$	11.857	11.949	11.990	11.999	12.154	12.516	12.808	13.127

Source: HMD (2018) and authors' calculations.
Note: Country codes are listed in section 3. M and F indicate males and females respectively.

For a particular population, the populations are closest in terms of $\bar{e}_0^\dagger$ can be easily identified from Table 1. To illustrate, the closest populations for female populations of France are given in Table 2. It is seen that the closest 15 populations are female, underlying the significance of the sex difference in mortality. The order of closest populations obtained in Table 2 is important for coherent forecasting as we added closest population simultaneously in reference group following this order. Kjærgaard et al. (2016) mentioned the importance of choosing the reference population in coherent settings as it clearly effects the forecast accuracy and future mortality.

We did not fix the number of populations in best reference group prior to model fitting. To determine the best reference group, we analyzed the forecast accuracy both for Li and Lee (2005) and the proposed method ($\text{LL}_{e_0^\dagger}$). As reference group, a combination of countries which produces lowest forecast error during out-of-sample evaluation is chosen. Details of forecast accuracies for proposed methods are discussed in the next section; here we discuss only the findings for French Females for illustration of identification of the best reference group. We added populations one at a time into the reference group with previous combination for both $\text{LL}_{e_0^\dagger}$ and LL and used same weight for both methods. The forecast accuracy for French Female during hold-out period is plotted in Figure 2.

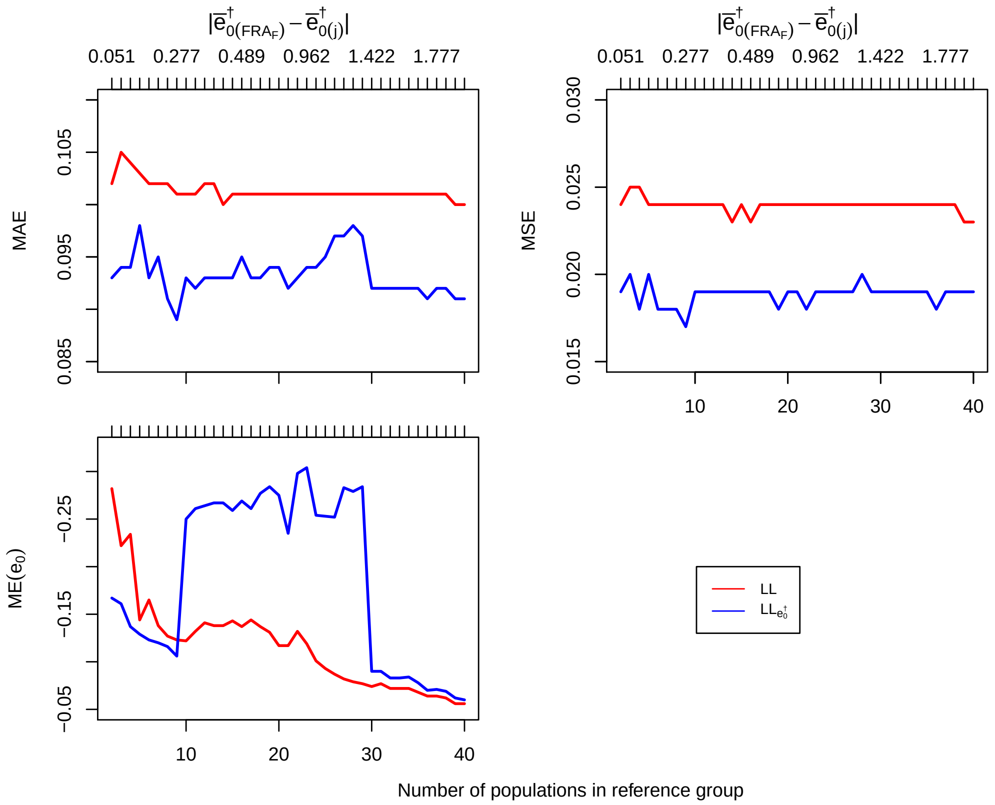
For all three measures of forecast accuracy there is a distinct fall (rise) in accuracy (error) level after adding certain countries using $\text{LL}_{e_0^\dagger}$ but LL did not produce this sharp threshold in the case of $\text{ME}(e_0)$. All three measures of forecast accuracy indicated the same best reference group for French females in the case of $\text{LL}_{e_0^\dagger}$. For LL, MAE and MSE indicate the same best reference group, whereas different best reference group was indicated by $\text{ME}(e_0)$. Different combinations of populations as best reference group from different measures of forecast accuracies are also

Table 2: Reference population for French Females based on closest difference in $\bar{e}_0^\dagger$ during 1982:2011

Population	$\mathrm{FRA_F}$	$\mathrm{NLD_F}$	$\mathrm{BEL_F}$	$\mathrm{AUS_F}$	$\mathrm{IRL_F}$	$\mathrm{DEU_F}$	$\mathrm{NOR_F}$	$\mathrm{PRT_F}$
$\left\lvert \bar{e}^\dagger_{0,\mathrm{FRA_F}} - \bar{e}^\dagger_{0j} \right\rvert$	-	0.051	0.076	0.088	0.115	0.131	0.227	0.240
Population	$\mathrm{AUT_F}$	$\mathrm{ITA_F}$	$\mathrm{FIN_F}$	$\mathrm{SWE_F}$	$\mathrm{UK_F}$	$\mathrm{JPN_F}$	$\mathrm{CHE_F}$	$\mathrm{ESP_F}$
$\left\lvert \bar{e}^\dagger_{0,\mathrm{FRA_F}} - \bar{e}^\dagger_{0j} \right\rvert$	0.276	0.334	0.367	0.408	0.412	0.444	0.455	0.489
Population	$\mathrm{CAN_F}$	$\mathrm{SWE_M}$	$\mathrm{NLD_M}$	$\mathrm{NZL_F}$	$\mathrm{DNK_F}$	$\mathrm{JPN_M}$	$\mathrm{NOR_M}$	$\mathrm{IRL_M}$
$\left\lvert \bar{e}^\dagger_{0,\mathrm{FRA_F}} - \bar{e}^\dagger_{0j} \right\rvert$	0.502	0.533	0.660	0.721	0.725	0.932	0.962	0.963
Population	$\mathrm{CHE_M}$	$\mathrm{UK_M}$	$\mathrm{ITA_M}$	$\mathrm{AUS_M}$	$\mathrm{DEU_M}$	$\mathrm{BEL_M}$	$\mathrm{USA_F}$	$\mathrm{DNK_M}$
$\left\lvert \bar{e}^\dagger_{0,\mathrm{FRA_F}} - \bar{e}^\dagger_{0j} \right\rvert$	1.086	1.117	1.169	1.270	1.339	1.421	1.424	1.453
Population	$\mathrm{CAN_M}$	$\mathrm{AUT_M}$	$\mathrm{ESP_M}$	$\mathrm{NZL_M}$	$\mathrm{FIN_M}$	$\mathrm{FRA_M}$	$\mathrm{PRT_M}$	$\mathrm{USA_M}$
$\left\lvert \bar{e}^\dagger_{0,\mathrm{FRA_F}} - \bar{e}^\dagger_{0j} \right\rvert$	1.478	1.571	1.612	1.620	1.776	2.137	2.430	2.749

Source: HMD (2018) and authors' calculations.
Note: Country codes are listed in section 3. M and F indicate males and females respectively.

Figure 2: Number of countries for achieving lowest MAE, MSE and ME(e_0) for French Females

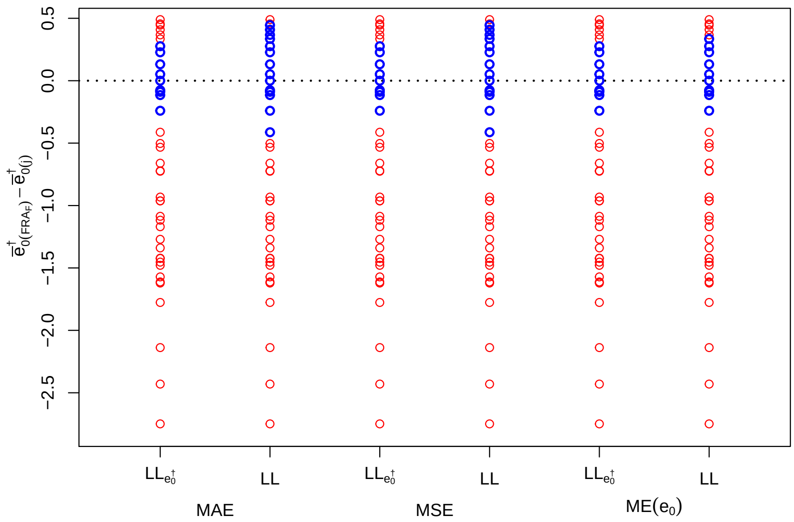


Source: HMD (2018) and authors' calculations.

observed for many other populations. However, this threshold in level of forecast accuracy is important, because it clearly shows which countries belongs to the best reference group. For several other populations, we observed that ME(e_0) diminishes after adding many populations even though those populations are not so close to that of interest in terms of observed mortality level. Adding these populations with huge gap in level of population-specific $\bar{e}_0^\dagger$ does not have high impact on MAE or MSE (Figure 2). The best reference populations obtained for French

Females according to distance of $\bar{e}_0^\dagger$ with other population is plotted in Figure 3.

Figure 3: Closest population to be included as reference for French Females according to different measures of forecast accuracy

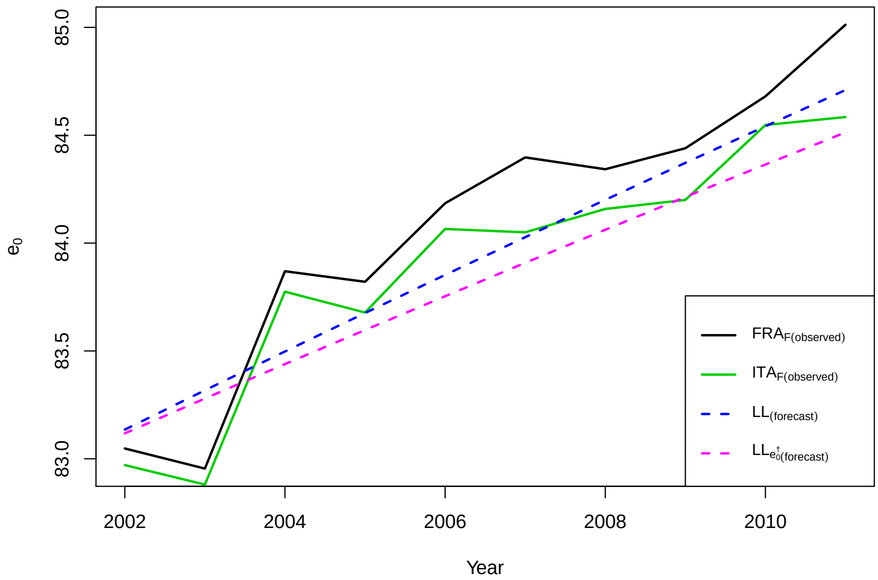


Source: HMD (2018) and authors' calculations.
Note: The horizontal black dotted line represents French Females. The Red dots are set available populations while blue dots represent the ones forming the best reference group in terms of highest forecast accuracy.

For French Females, the $ME(e_0)$ showed a sharp rise after adding Italian Females into the reference group. The $\bar{e}_0^\dagger$ for French Females was closer to populations added in reference group prior to Italian Females (see Table 4 for details). The observed e_0 of French and Italian Females along with forecast of e_0 from both LL and $LL_{e_0^\dagger}$ are plotted in Figure 4. Both for LL and $LL_{e_0^\dagger}$, the reference group consists of 10 populations which includes up to Austrian Females (see Table 4 for details). The observed life expectancies for both of the populations were close to each other until 2004. Irregular divergence is visible afterwards and French e_0 is closer to the LL forecast than to the $LL_{e_0^\dagger}$ forecast. Further insight into the reference populations for French females can be obtained from the trend in $e_0^\dagger$ for some of the populations in the possible best reference group (from Table 2), seen in Figure 5 .

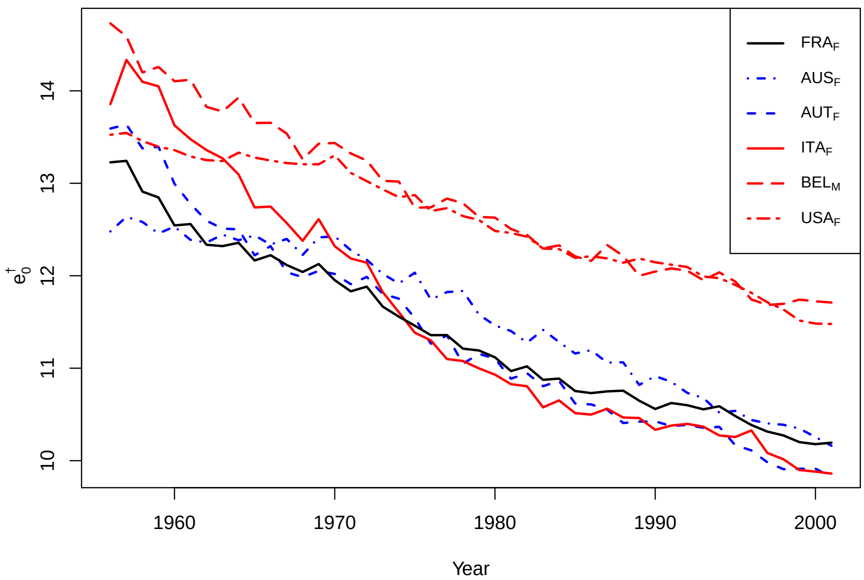
For Italian Females, therecent trend of $e_0^\dagger$ is close to other populations in the best reference group for French Females, but the past trend was quite different. This explains the change in forecast accuracy after adding many populations. For instance, Belgian Males are ranked as 29-th closest population to French Females in terms of lifespan disparity with a much higher level of mortality (Table 2). Combining many such divergent populations in the reference group finally increases accuracy because of using the same weight in the common factor model, but, this attained without any similarities in mortality patterns. Although adding a large number of countries increases the level of accuracy for life expectancy, it does not change MAE or MSE substantially from the values obtained for a smaller number of populations in the reference group.

Figure 4: Observed and forecast of e_0 for French Females (2002:2011)



Source: HMD (2018) and authors' calculations.
Note: Observed Italian Females e_0 are also added for comparison.

Figure 5: Trend of $e_0^\dagger$ for French Females and some other populations (1956:2001)

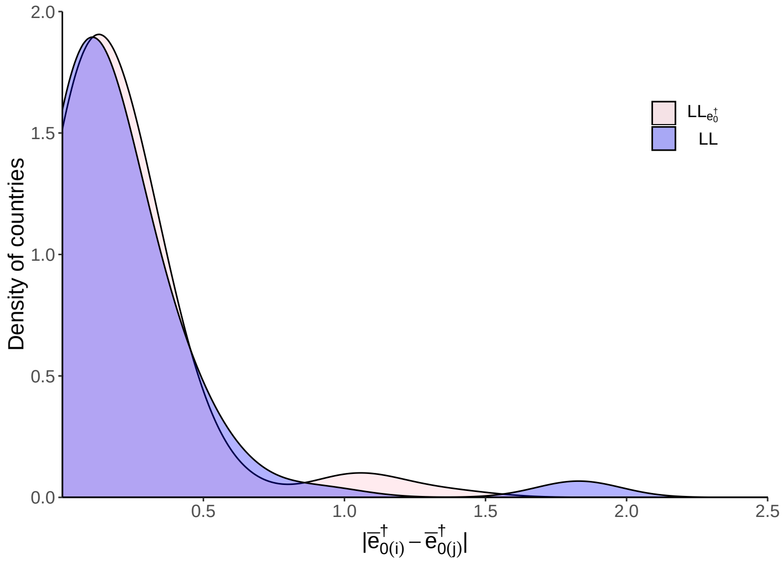


Source: HMD (2018) and authors' calculations.
Note: The populations in blue lines are later found to be in best reference group for French Females and red are for those populations which are not in best reference group for $\text{LL}_{e_0^\dagger}$.

4.2 Optimal size of best reference group

Unlike previous approaches we did not restrict the number of populations in the best reference group; rather we choose to observe the threshold number of populations in reference group for which the forecast accuracy is highest. From empirical analysis, we obtained different results for LL and $\mathrm{LL}_{e_0^\dagger}$. The best reference groups obtained from LL and $\mathrm{LL}_{e_0^\dagger}$ using different measures of forecast accuracy for all 40 appear in the Appendix. The distribution of number of countries in the best reference group according to $\left|\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger\right|$ is plotted in Figure 6, based on the three measures of forecast accuracy together as different measures of forecast accuracy indicate different populations as the best reference group.

Figure 6: Distribution of countries obtaining reference population according to $\left|\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger\right|$.



Source: HMD (2018) and authors' calculations.
Note: We plotted a kernel density instead of histogram as it shows the skewness properly.

In the case of same accuracy level for two or more combinations, the combination having smaller number of countries is chosen as best reference group for a parsimonious model. The optimal number of populations in reference group and corresponding differences in $\left|\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger\right|$ considering all forecast errors are summarized in Table 3. The errors are considered separately for males, females and both sexes together.

Table 3: Summary statistics for best reference group

	Summary statistics	$\mathrm{LL}_{e_0^\dagger}$ Male	$\mathrm{LL}_{e_0^\dagger}$ Female	$\mathrm{LL}_{e_0^\dagger}$ All	LL Male	LL Female	LL All
Number	Mean	5	5	5	5	5	5
of	Median	5	5	5	4	4	4
populations	IQR	6	4	5	4	4	4
Difference	Mean	0.25	0.16	0.21	0.28	0.19	0.24
in	Median	0.15	0.14	0.15	0.15	0.09	0.12
$\left\|\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger\right\|$	IQR	0.26	0.20	0.24	0.26	0.20	0.25

Source: HMD (2018) and authors' calculations.
Note: IQR stands for interquartile range. We choose it over other measure of dispersion as it gives more idea about spread of the distribution in this context.

4.3 Forecast accuracy

$\mathrm{LL}_{e_0^\dagger}$ returns lower MAE and MSE than LL for all populations except for US Females. The forecast accuracy for French Females is given in Table 4.

Table 4: Comparison of forecast accuracy for French female during out of sample evaluation period (2002-2011)

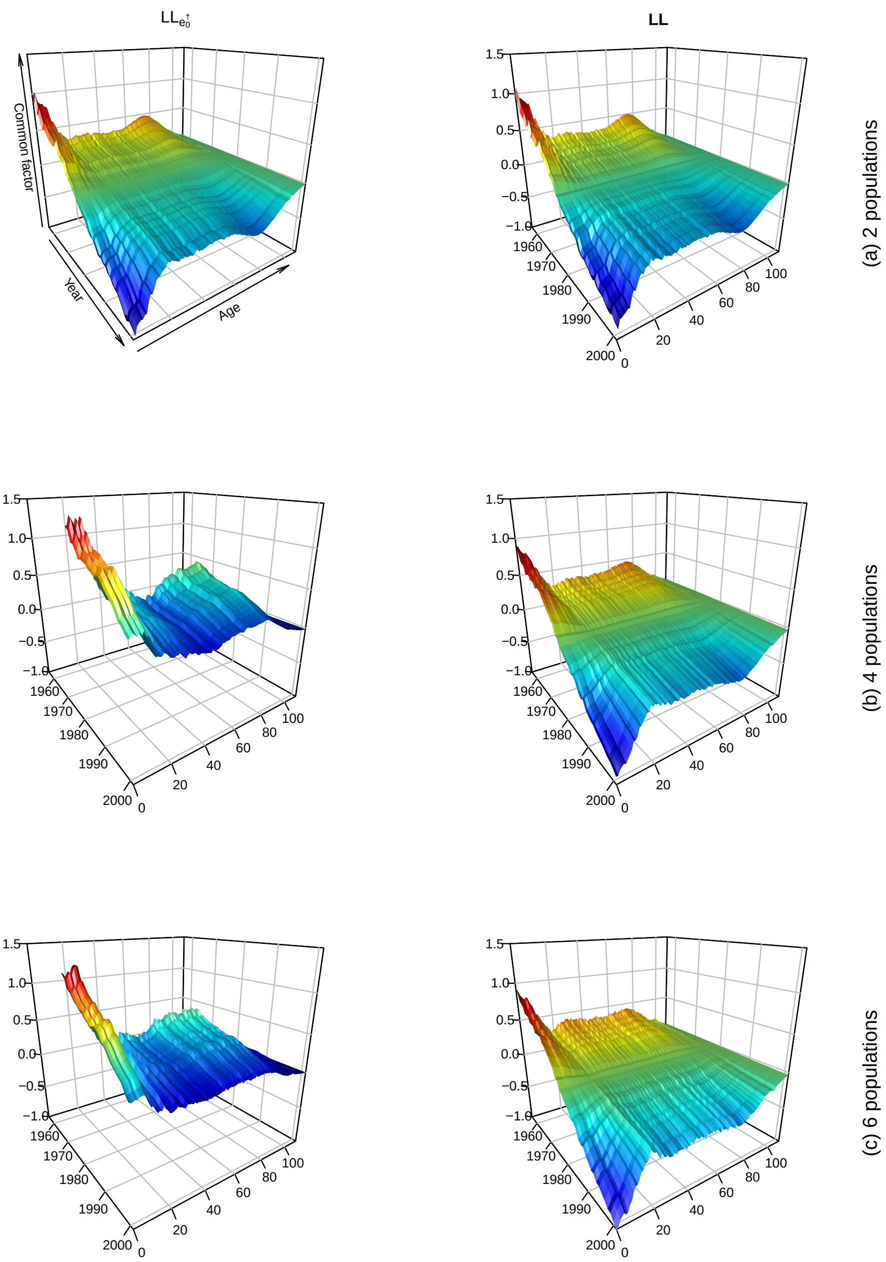
Reference population	MAE		MSE		ME(e_0)		Best fitting
	LL	$\mathrm{LL}_{e_0^\dagger}$	LL	$\mathrm{LL}_{e_0^\dagger}$	LL	$\mathrm{LL}_{e_0^\dagger}$	period ($\mathrm{LL}_{e_0^\dagger}$)
$\mathrm{FRA_F}$+$\mathrm{NLD_F}$	0.103	0.093	0.024	0.019	-0.282	-0.167	1956:2001
Above+$\mathrm{BEL_F}$	0.105	0.094	0.025	0.020	-0.222	-0.161	1956:2001
Above+$\mathrm{AUS_F}$	0.105	0.094	0.025	0.018	-0.235	-0.137	1979:2001
Above+$\mathrm{IRL_F}$	0.103	0.098	0.024	0.020	-0.144	-0.129	1979:2001
Above+$\mathrm{DEU_F}$	0.102	0.093	0.024	0.018	-0.165	-0.123	1977:2001
Above+$\mathrm{NOR_F}$	0.102	0.095	0.024	0.018	-0.138	-0.120	1979:2001
Above+$\mathrm{PRT_F}$	0.101	0.091	0.024	0.018	-0.127	-0.116	1974:2001
Above+$\mathbf{AUT_F}$	0.101	**0.089**	0.023	**0.017**	-0.124	**-0.105**	1974:2001
Above+$\mathrm{ITA_F}$	0.101	0.093	0.024	0.019	*-0.123*	-0.250	1964:2001
Above+$\mathrm{FIN_F}$	0.101	0.092	0.024	0.019	-0.133	-0.261	1960:2001
Above+$\mathrm{SWE_F}$	0.102	0.093	0.024	0.019	-0.141	-0.264	1960:2001
Above+$\mathrm{UK_F}$	0.102	0.093	0.024	0.019	-0.139	-0.267	1960:2001
Above+JPN_F	*0.100*	0.093	*0.023*	0.019	-0.138	-0.212	1957:2001
Above+$\mathrm{CHE_F}$	0.101	0.093	0.024	0.019	-0.144	-0.208	1957:2001
Above+$\mathrm{ESP_F}$	0.101	0.095	0.024	0.019	-0.138	-0.201	1957:2001
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Source: HMD (2018) and authors' calculations.
Note: Bold texts are used for showing the lowest errors obtained by$\mathrm{LL}_{e_0^\dagger}$, while italic texts are used for showing lowest error obtained by LL.

It has been already mentioned before that different measures of forecast accuracy often leads to different best reference groups. The lowest errors were obtained for same combination of populations in the case of $\mathrm{LL}_{e_0^\dagger}$. For LL, lowest MAE and MSE was obtained from same combination while lowest ME(e_0) were obtained with a combination of fewer populations. One important feature of proposed $\mathrm{LL}_{e_0^\dagger}$ is the concept of best fitting period. We did not find best fitting period for all of the combinations (Table 4). The best reference group (bold texts in Table 4) obtained for French Females consists of 9 populations and it has the best fitting period for 1974 to 2001. The reason of better accuracy of $\mathrm{LL}_{e_0^\dagger}$ is already explained in Figure 4 and 5. The proposed modifications produce different common factor for $\mathrm{LL}_{e_0^\dagger}$ than that of LL. The common factor obtained from $\mathrm{LL}_{e_0^\dagger}$ and LL for French Females are plotted in Figure 7. Here we showed 3 different reference groups consisting 2, 4 and 6 populations respectively.

For small number of populations in reference group, the effect of adjusting K_t according to $e_0^\dagger$ is highly visible (Figure 7a). However, for group consisting large number of populations clearly the best fitting period is responsible for higher forecast accuracy in the case of $\mathrm{LL}_{e_0^\dagger}$ (Figure 7b,c). However, with adding more populations to the reference group, this best fitting period slowly shifts to full fitting period eventually. The reference group consisting of all 40 low-mortality populations considers full observational period (1956:2001) as best fitting period in $\mathrm{LL}_{e_0^\dagger}$. The comparison of different measures of forecast accuracy during hold-out period is

Figure 7: Common factor of the fitted models with different sizes of reference group



Source: HMD (2018) and authors' calculations.

given in Table 5.

Since $LL_{e_0^\dagger}$ produces lower error than LL for all combinations, the population-specific lowest forecast error for both $LL_{e_0^\dagger}$ and LL are shown in Table 5. $LL_{e_0^\dagger}$ and LL produced same level of MAE and MSE for US Females. For ME(e_0), LL performed better than that of $LL_{e_0^\dagger}$; LL produced lower error for 11 populations. Unusual rise in ME(e_0) is observed in the case of $LL_{e_0^\dagger}$ for several other countries, for all of them we noticed same pattern as observed for French Females (Figure 5). Best fitting period corresponding to best reference groups from all these

Table 5: Comparison of minimum values of different measures of forecast accuracy for female populations of 20 low-mortality countries during hold-out period (2002-2011)

	MAE		MSE		ME(e_0)	
Country	LL	LL$_{e_0^\dagger}$	LL	LL$_{e_0^\dagger}$	LL	LL$_{e_0^\dagger}$
Australia	0.121	*0.096*	0.038	*0.022*	-0.011	*-0.004*
Austria	0.193	*0.156*	0.100	*0.066*	0.007	*0.003*
Belgium	0.153	*0.135*	0.070	*0.053*	0.002	-0.009
Canada	0.080	*0.068*	0.016	*0.013*	-0.042	*-0.041*
Denmark	0.231	*0.200*	0.129	*0.102*	-0.672	*-0.542*
Finland	0.223	*0.199*	0.125	*0.080*	-0.017	-0.072
France	0.100	*0.089*	0.023	*0.017*	-0.123	*-0.106*
Germany	0.087	*0.082*	0.017	*0.014*	0.049	0.082
Ireland	0.251	*0.242*	0.143	*0.130*	-0.998	-1.063
Italy	0.085	*0.078*	0.019	*0.015*	-0.005	-0.032
Japan	0.120	*0.116*	0.034	*0.031*	0.431	0.465
The Netherlands	0.149	*0.138*	0.068	*0.056*	-0.516	*0.514*
New Zealand	0.224	*0.184*	0.118	*0.838*	-0.298	-0.301
Norway	0.230	*0.188*	0.150	*0.100*	-0.036	-0.217
Portugal	0.174	*0.140*	0.074	*0.047*	-0.597	*-0.281*
Spain	0.100	*0.096*	0.023	*0.021*	0.008	0.011
Sweden	0.169	*0.149*	0.079	*0.063*	-0.037	*-0.012*
Switzerland	0.222	*0.165*	0.161	*0.074*	0.105	0.105
United Kingdom	0.079	*0.069*	0.014	*0.010*	-0.343	*-0.331*
USA	0.054	0.054	0.005	0.005	-0.156	-0.221

Source: HMD (2018) and authors' calculations.
Note: Values in italic For LL$_{e_0^\dagger}$ means the accuracy were higher than that of LL.

measures of forecast accuracies are shown in the appendix.

4.4 Forecast of life expectancy

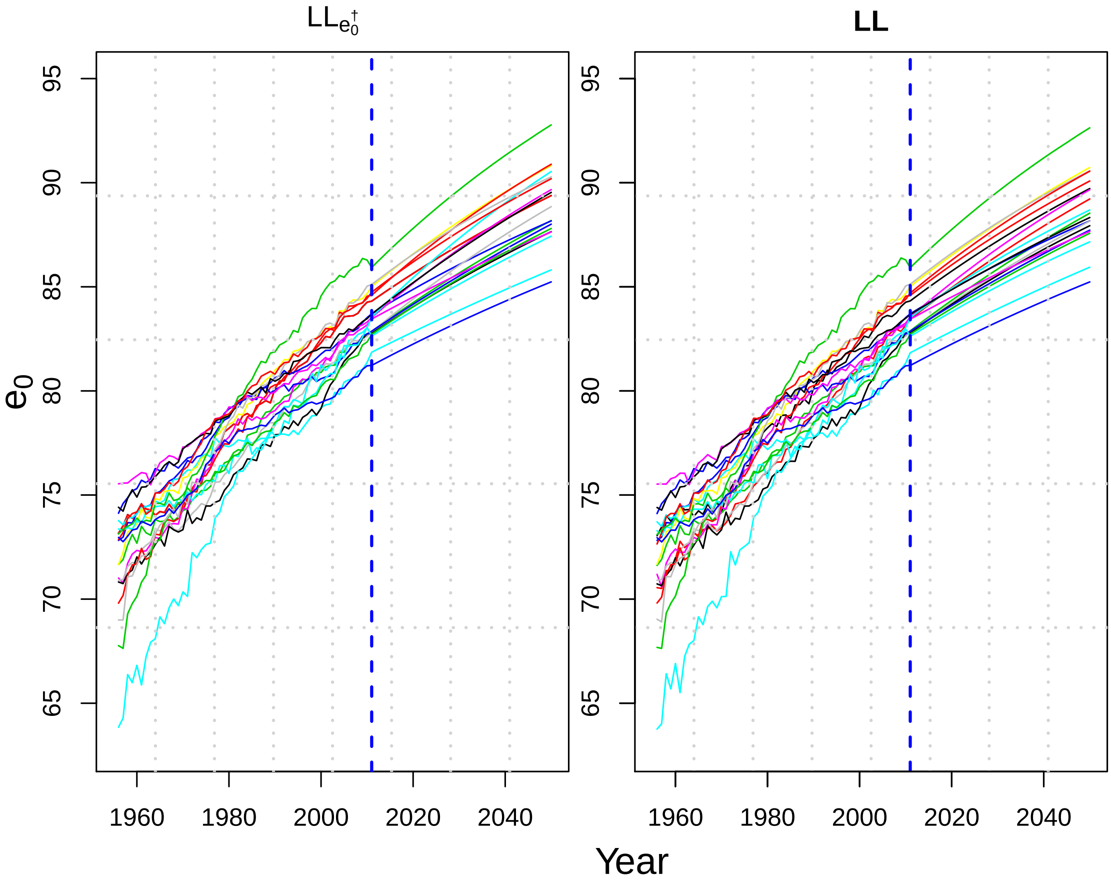
The forecasts of female life expectancy at birth for all 20 low-mortality countries are plotted in Figure 8. For both of the methods, we checked forecast accuracy during out-of-sample evaluation and from three possible best reference groups we choose the one having lower number of populations to make coherent forecast.

The coherent forecast of life expectancy at birth in 2050 for all of these populations are given in Table 6. Except for Australia, Belgium, Denmark, Ireland, Norway, Spain and Switzerland, the forecast of life expectancy were higher for LL$_{e_0^\dagger}$ than that of LL.

To gain greater insight of the obtained results we further identified the populations for which LL$_{e_0^\dagger}$ was most and least optimistic than that of LL. For Portugal and Sweden the forecast obtained from LL$_{e_0^\dagger}$ have highest positive difference with that of LL whereas the highest negative difference were observed for Belgium and Switzerland. The forecast of these four countries are plotted in Figure 9.

Among these four countries, Portugal is mentioned before for remarkable improvement in health status and rapid increase of life expectancy (Van Oyen et al. 2013). Between 2000

Figure 8: Observed (1956:2011) and forecast (2012:2050) of female life expectancy at birth for 20 low-mortality countries



Source: HMD (2018) and authors' calculations.

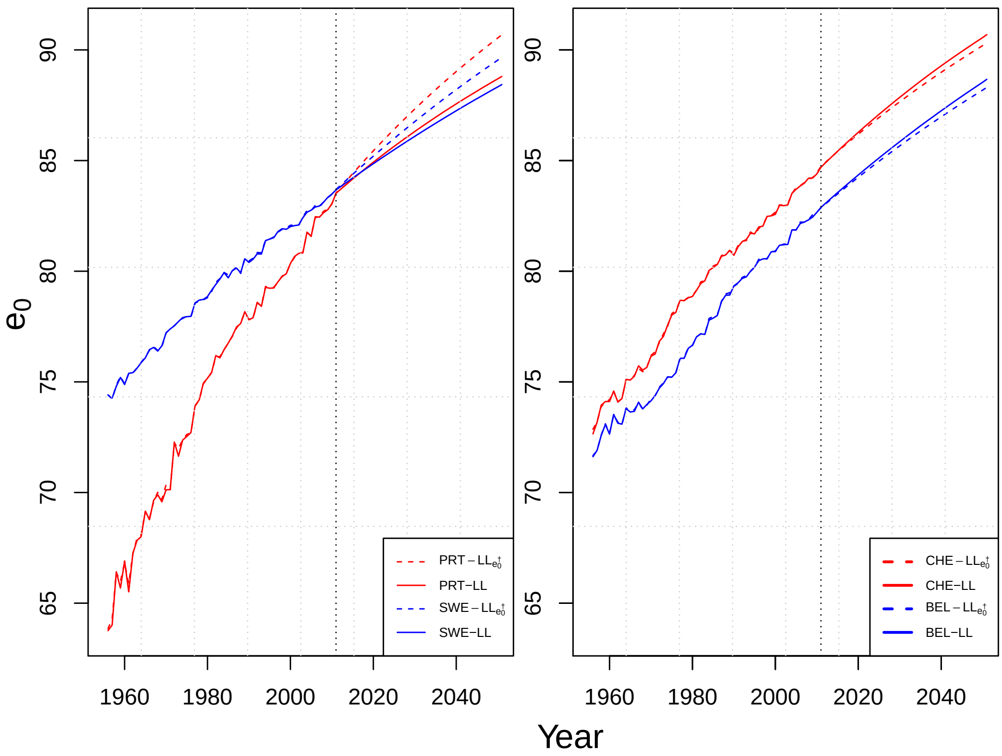
Table 6: Comparison of coherent forecast of life expectancy at birth in 2050 for female populations of 20 low-mortality countries

Country	$LL_{e_0^\dagger}$	LL	Country	$LL_{e_0^\dagger}$	LL
Australia	89.382	89.719	Japan	92.777	92.636
Austria	89.382	89.216	The Netherlands	88.009	87.721
Belgium	88.184	88.538	New Zealand	87.432	87.159
Canada	88.166	88.166	Norway	87.636	87.662
Denmark	85.813	85.943	Portugal	90.537	88.692
Finland	89.668	89.665	Spain	90.278	90.567
France	90.824	90.730	Sweden	89.535	88.327
Germany	88.851	88.154	Switzerland	90.194	90.557
Ireland	87.649	87.934	United Kingdom	87.802	87.571
Italy	90.887	90.082	USA	85.237	85.235

Source: HMD (2018) and authors' calculations.

and 2015 the female life expectancy increased by almost four years for Portugal, almost 5 years for males (HMD 2018). However, these improvements have not occurred at the same pace for different income groups and disparities exist for other important dimensions of health.

Figure 9: Observed (1956:2011) and forecast (2012:2050) of female life expectancy at birth for Portugal, Sweden, Denmark and Switzerland



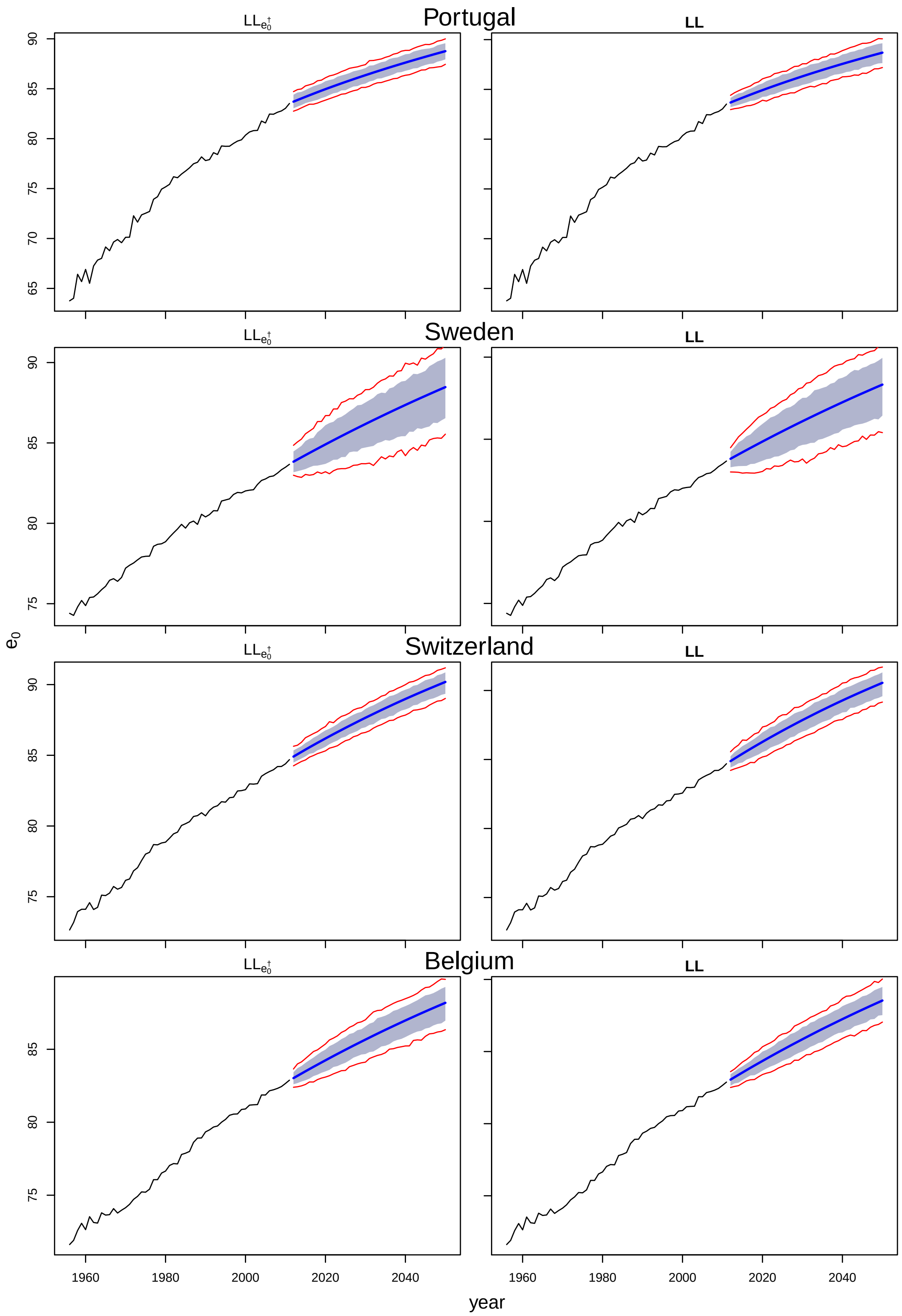
Source: HMD (2018) and authors' calculations.
Note: For Italy and Sweden the difference of forecast of life expectancy were highest at 2050 while it was lowest for Denmark and Switzerland.

Cardiovascular diseases and cancer are the largest contributors to mortality(Van Oyen et al. 2013). On the other hand, females of Sweden and Switzerland have steady pattern of mortality improvement since long (HMD 2018). However, this scenario is is not the case for Danish Females. During the past decades, the life expectancy of Danish women has lagged behind that of women in neighboring Western European countries (Jacobsen et al. 2002). Among various causes-of-deaths, ischaemic heart diseases followed by lung cancer are responsible for lower life expectancy of Danish Females. Danish female mortality is mentioned before for having distinct pattern of remarkable middle-aged mortality (Juel et al. 2000).

4.5 Interval forecast

To construct prediction interval of forecast of life expectancy at birth, we followed the procedure employed by Hyndman and Booth (2008). In this procedure, the fitted mortality rates from forecasting technique are simulated a large number of times to add disturbance to the time component of the model. Life expectancies are then calculated for each set of the simulated log-mortality rates. Prediction intervals are then constructed by 80% or 95% percentiles of the simulated sets of the life expectancies. Following the results of forecast of life expectancy till 2050, the prediction interval of e_0 is plotted in Figure 10 for Portugal, Sweden, Switzerland and Belgian Females. For Sweden and Belgium, the prediction interval of the $LL_{e_0^\dagger}$ is slightly wider than that of LL, whereas prediction interval from LL are slightly wider for Portugal and Switzerland. Interval forecast for other female populations are added in Appendix for interested readers.

Figure 10: Prediction interval of Portuguese, Swedish, Swiss and Belgian female life expectancy at birth till 2050



Source: HMD (2018) and authors' calculations.
Note: The blue area represents 80% prediction interval and red lines are for 95% prediction interval.

5 Discussion and Conclusion

In this line of research on Lee-Carter framework, we introduced a new methodology for coherent forecasting. Choosing the appropriate reference group is an old puzzle for coherent forecasting and different reference populations bring about quite different results. We addressed this problem by proposing a robust definition of reference population based on closest trend of lifespan disparity. This definition is found to be applicable for existing coherent forecasting technique as well. We incorporated lifespan disparity during parameter estimation of the coherent forecasting along with application of Lasso type smoothing prior to fitting the model to overcome the problem of a jagged trend of age-component over the lifespan. Along with consideration of best fitting period in proposed setup; the coherent forecast became more accurate during out-of-sample evaluation and provides more optimistic forecast.

Despite of promising results, there are still many open questions that deserve further investigation. For starter, interval forecast of life expectancy at birth is narrow for several of the populations for both methods. Although $\text{LL}_{e_0^\dagger}$ produced slightly wider prediction interval than LL in many cases, still, this is an old criticism regarding Lee-Carter variants (Hyndman and Ullah 2007). In the proposed model it happened due to application of smoothing and new adjustment technique which made the time component more linear. As a consequence it reduces the variance of the ARIMA model. In addition, variance of the model is lower in the proposed method, which also affects the interval forecast (Rabbi and Mazzuco 2018a).

In proposed definition of reference population, we considered both genders together. Although several approaches considered males and females together for coherent forecasting, still, consideration of males and females together in reference group is a topic of debate due to different pattern of mortality over the lifespan. There are two responses regarding this issue. First, during sorting out the closest populations, males and females were separated naturally from the value of $\bar{e}_0^\dagger$. Getting best reference population from opposite gender did not happen very often except for few populations. Second, most of these countries consider same policy for both genders regarding old health care system, as aging is common issue for both genders. Thus, practical implementation is not a big problem.

In coherent forecasting, we applied equal weight on mortality rates of all populations to construct joint mortality matrix (both for $\text{LL}_{e_0^\dagger}$ and LL). This resolved the problem of mixing population with large exposure with smaller one, however, using equal weight also has consequences. Different mortality pattern of different populations are result of both exposure size and different distribution of causes-of-deaths. Defining an appropriate population-specific weight to adjust the problem of different distribution of causes-of-deaths will be complicated because different data sources will be needed to obtain harmonized data for causes-of-death. Although the low-mortality countries are converging in terms of aging, still, each of these populations are distinct in terms of distribution of deaths (Figure 1). Equal weight also reduces the benefit of lasso in case of reference group consists of larger number of populations.

As most of the populations do not have data after age 100 for several years, we extrapolated the mortality rates of age 100:110+ using Kannisto model. Although we obtained missing mortality rates keeping the fitted rates at the closest distance to observed data, still it reduced the variability in centenarian mortality (see Figure 1 for an example). Also, it is clear that smoothing can create significant differences in mortality forecasting (Figure **??**). For coherent forecasting model, we separately smoothed the populations using Lasso and then combined them in common factor model. Instead of smoothing prior to combine the mortality rates, applying Lasso after combining populations may produce different results. However, smoothing after combining populations brings about a higher computational complexity because in this way Lasso should be run every time a new population is added in reference group to check the forecast accuracy. In addition to all of these issues, we did not present the forecast-accuracy or forecast by bringing two classes of models together (single and coherent), though the new coherent method found to be more accurate then Lee and Carter (1992) and Lee and Miller (2001) for most of the populations. We are leaving this issue for personal subjective opinion of getting the forecast independently for a particular population or from a coherent point of view.

Based on the results and limitations of the current research, we might point out some future scope of research on mortality forecasting. The first issue will be to overcome to problem of invariant b_x in Lee-carter framework. One possible solution to do that is to adopt Bayesian approaches on parameter estimation. Secondly, adaptation of cohort effect might get more insight of the mortality scenario of a population. Thirdly, we introduced a new systematic approach to obtain best reference group for a population. Although $e_0^\dagger$ better reflects the distribution of death for a population, further research on this field may produce better results than proposed method. Fourthly, although Lasso is found to be more effective smoothing technique for our data, it is a slightly time consuming method. Faster algorithm for getting optimal results for Lasso will be helpful. Fifthly, for sake of coherent forecasting, it is wise to consider life table of longer time series and longer life span. Future research on mortality forecasting may consider to make a more generalized definition for length of lifespan to be considered for coherent setup. Beside, current estimation technique for life expectancy at birth is not free from age-specific bias. Revision of its definition may change the current limitation of less accurate forecast of life expectancy.

References

Aburto, J. M. and van Raalte, A. (2018). Lifespan dispersion in times of life expectancy fluctuation: the case of central and eastern europe. *Demography*, 55(6):2071–2096.

Ahcan, A., Medved, D., Olivieri, A., and Pitacco, E. (2014). Forecasting mortality for small populations by mixing mortality data. *Insurance: Mathematics and Economics*, 54:12–27.

Ahmadi, S. S. and Li, J. S.-H. (2014). Coherent mortality forecasting with generalized linear models: A modified time-transformation approach. *Insurance: Mathematics and Economics*, 59:194–221.

Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J., and Vaupel, J. W. (2017). Coherent forecasts of mortality with compositional data analysis. *Demographic Research*, 37:527–566.

Bohk-Ewald, C., Ebeling, M., and Rau, R. (2017). Lifespan disparity as an additional indicator for evaluating mortality forecasts. *Demography*, 54(4):1559–1577.

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

Dokumentov, A., Hyndman, R. J., and Tickle, L. (2018). Bivariate smoothing of mortality surfaces with cohort and period ridges. *Stat*, 7(1):e199.

HMD (2018). *Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).

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.

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

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.

Jacobsen, R., Keiding, N., and Lynge, E. (2002). Long term mortality trends behind low life expectancy of danish women. *Journal of Epidemiology & Community Health*, 56(3):205–208.

Juel, K., Bjerregaard, P., and Madsen, M. (2000). Mortality and life expectancy in denmark and in other european countries: What is happening to middle-aged danes? *The European Journal of Public Health*, 10(2):93–100.

Kjærgaard, S., Canudas-Romo, V., and Vaupel, J. W. (2016). The importance of the reference populations for coherent mortality forecasting models. In *European Population Conference*, Mainz, Germany.

Lee, R. (2000). The lee-carter method for forecasting mortality, with various extensions and applications. *North American actuarial journal*, 4(1):80–91.

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

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

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

Rabbi, A. M. F. and Mazzuco, S. (2018a). A modified lee-carter method adjusting for smoothing and lifespan disparity. In *European Population Conference*, Brussels, Belgium.

Rabbi, A. M. F. and Mazzuco, S. (2018b). Mortality and life expectancy forecast for (comparatively) high mortality countries. *Genus*, 74(1):18.

Thatcher, A. R., Kannisto, V., and Vaupel, J. W. (1998). The force of mortality at ages 80 to 120.

Van Oyen, H., Nusselder, W., Jagger, C., Kolip, P., Cambois, E., and Robine, J.-M. (2013). Gender differences in healthy life years within the eu: an exploration of the "health–survival" paradox. *International journal of public health*, 58(1):143–155.

Vaupel, J. W. and Canudas-Romo, V. (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of nathan keyfitz's 90th birthday. *Demography*, 40(2):201–216.

Vaupel, J. W., Zhang, Z., and van Raalte, A. A. (2011). Life expectancy and disparity: an international comparison of life table data. *BMJ open*, pages bmjopen–2011.

Zhang, Z. and Vaupel, J. W. (2009). The age separating early deaths from late deaths. *Demographic Research*, 20:721–730.

Appendix

A1. Old age mortality

Mortality rates are not available for some of the later age-groups for the whole fitting period. Considering shifting mortality of almost all the populations, we fitted Kannisto's model at later age group for coherent forecasting. For ages $x = 80, 81, \dots 110+$, let the observed death counts are noted as D_x and exposure as E_x. We extrapolate the unavailable mortality rates for later age groups by fitting the Kannisto's model of old age mortality on observed death rates M_x to estimate the underlying hazards function μ_x as,

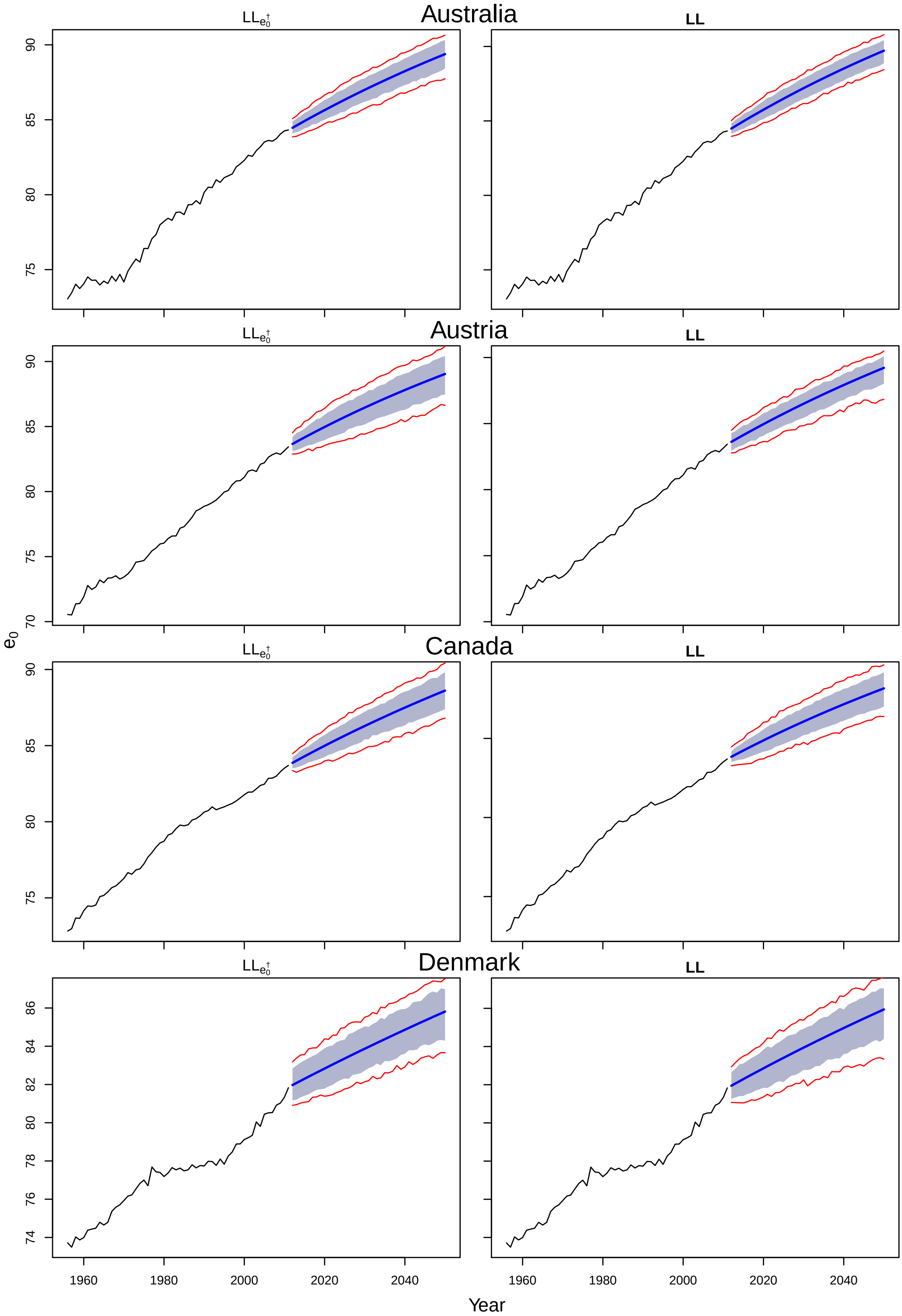
$$\mu_{x,(a,b)} = \frac{ae^{b(x-80)}}{1 + ae^{b(x-80)}}; \qquad a, b \geq 0.$$

Sensitivity of $e_0^\dagger$ due to Kannisto fitted mortality rates

We tried different combinations of fitting period and then added different combination of smoothed data with observed data. In this analysis we used the data obtained from fitting period at age 80:100 and adding the smoothed data of age 100:100+ with observed data till age 99. Among various combination we tried, this combination was the closest to real data and the difference of estimated and observed $e_0^\dagger$ during the fitting period (1956-2011) were lowest for this combination.

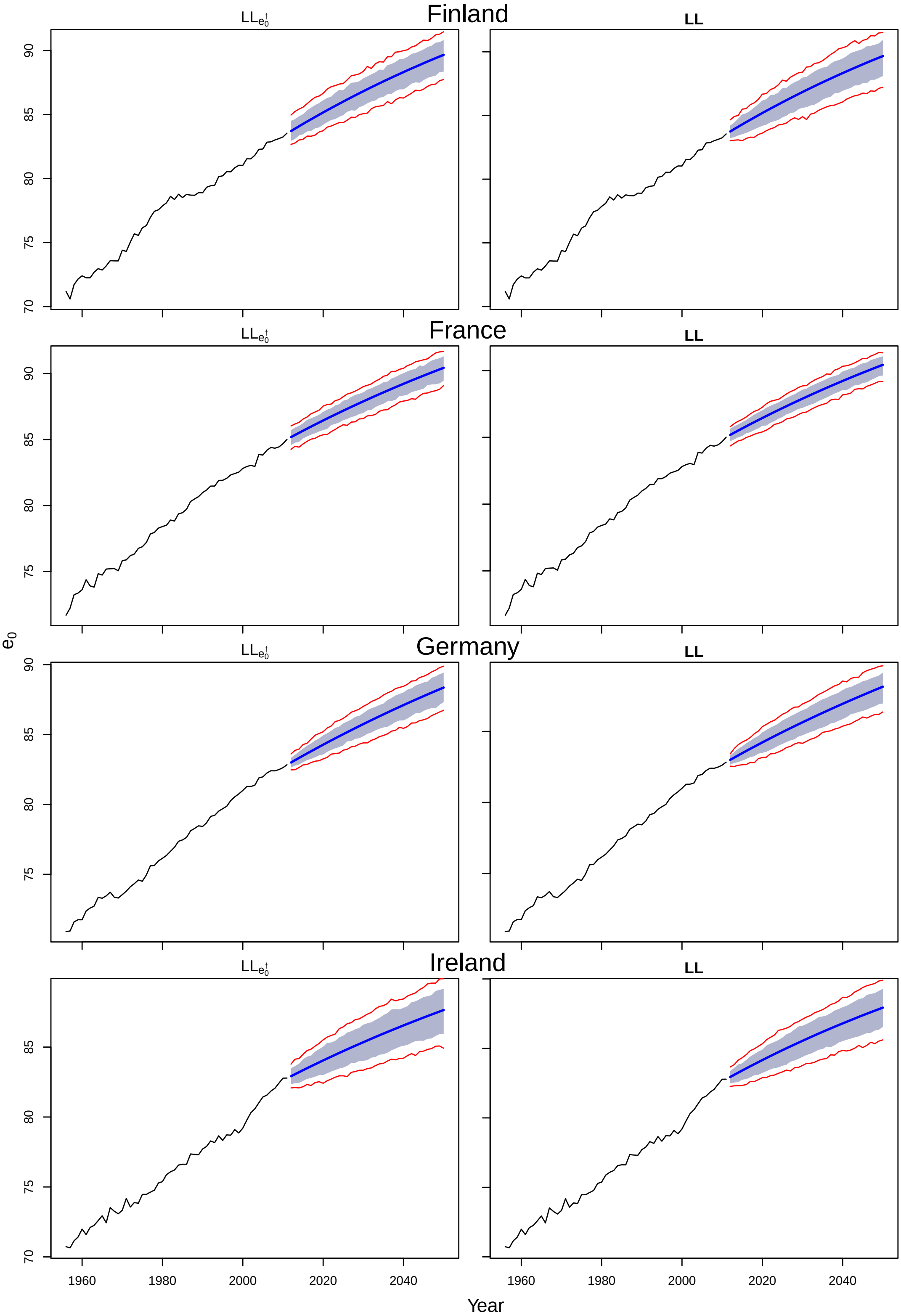
A2. Interval forecast of life expectancy

Figure 11: Prediction interval of female life expectancy at birth till 2050 for Australia, Austria, Canada and Denmark



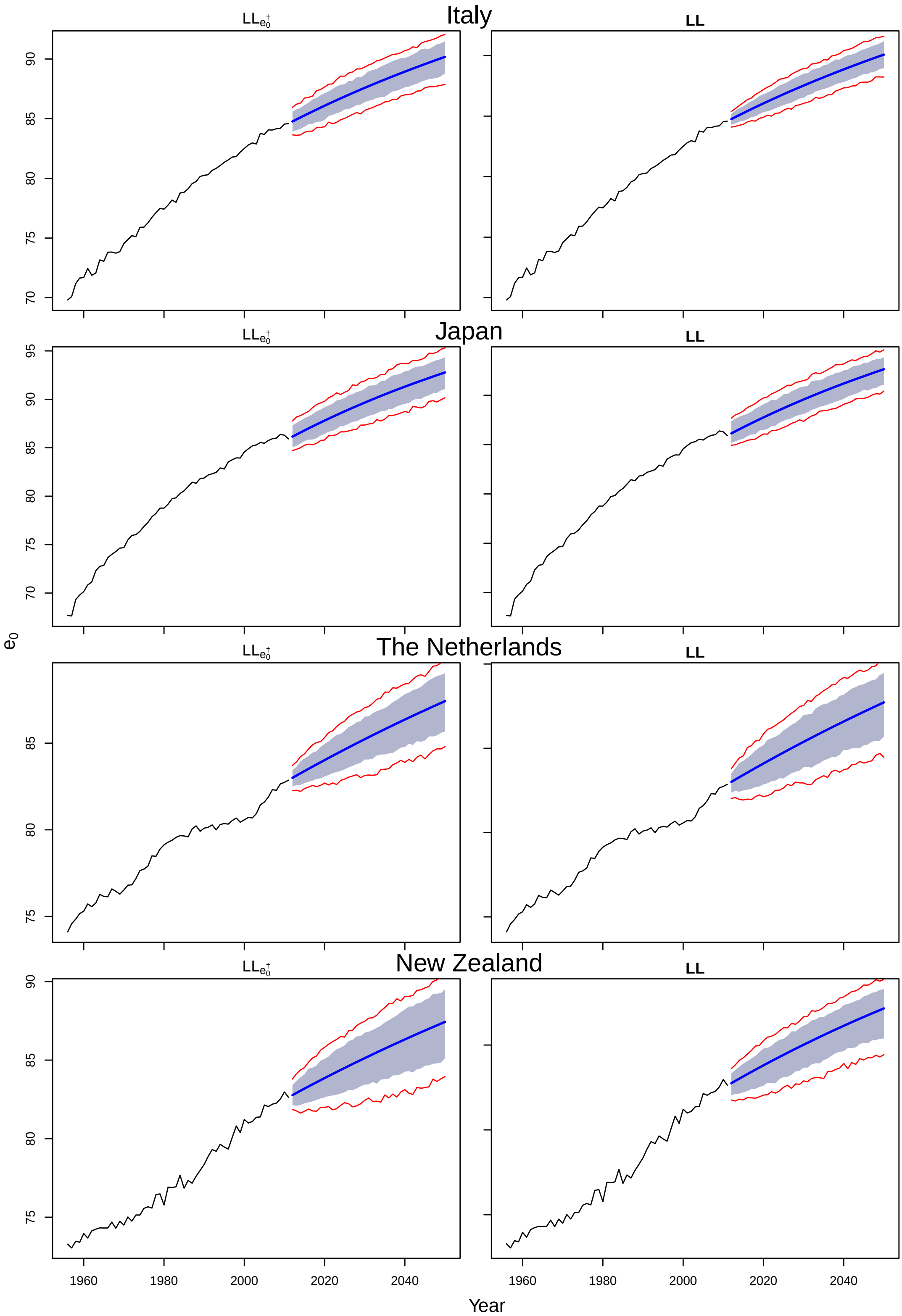
Source: HMD (2018) and authors' calculations.
Note: The blue area represents 80% prediction interval and red lines are for 95% prediction interval.

Figure 12: Prediction interval of female life expectancy at birth till 2050 for Finland, France, Germany and Ireland



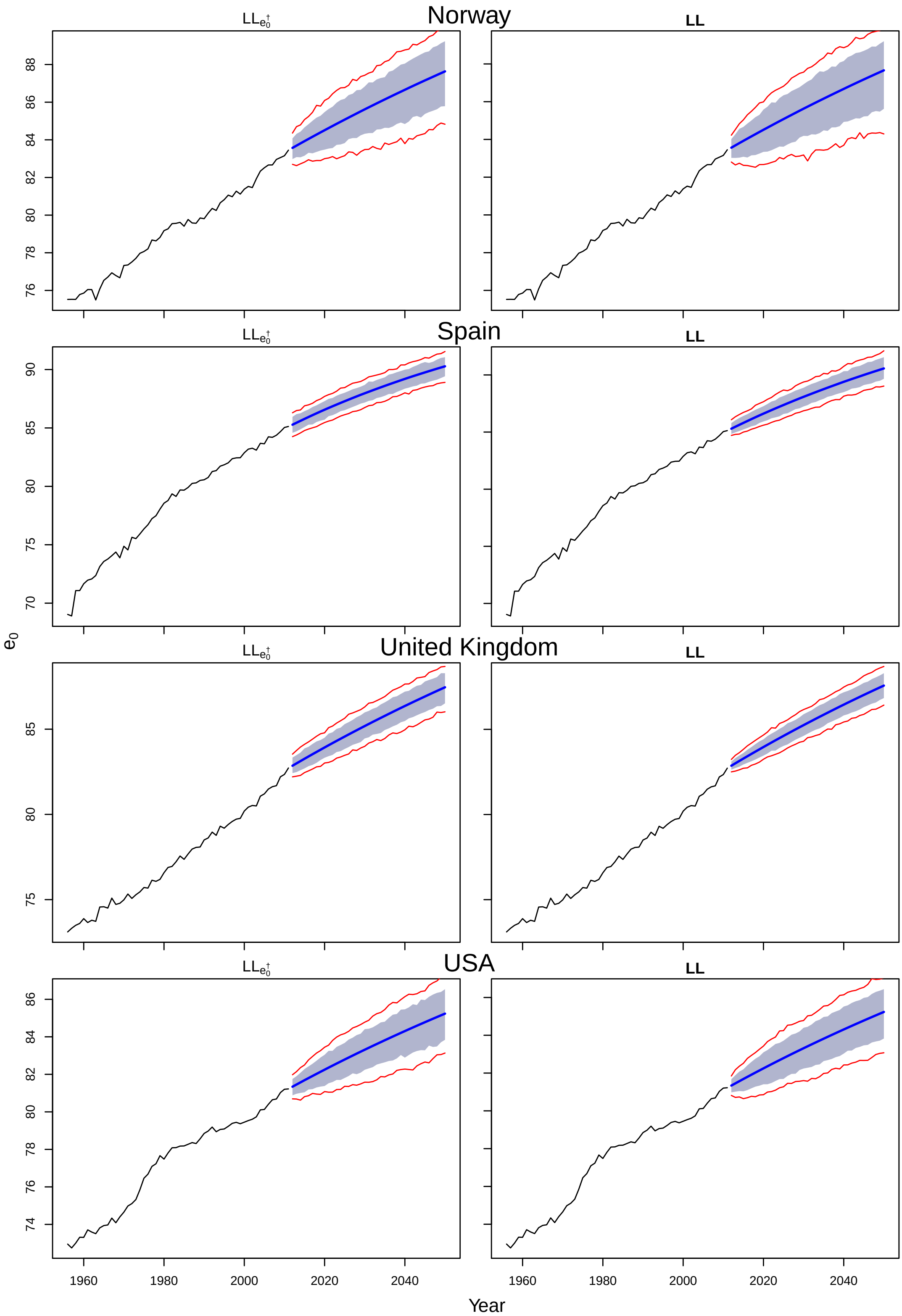
Source: HMD (2018) and authors' calculations.
Note: The blue area represents 80% prediction interval and red lines are for 95% prediction interval.

Figure 13: Prediction interval of female life expectancy at birth till 2050 for Italy, Japan, the Netherlands and New Zealand



Source: HMD (2018) and authors' calculations.
Note: The blue area represents 80% prediction interval and red lines are for 95% prediction interval.

Figure 14: Prediction interval of female life expectancy at birth till 2050 for Norway, Spain, United Kingdom and USA



Source: HMD (2018) and authors' calculations.
Note: The blue area represents 80% prediction interval and red lines are for 95% prediction interval.

A3. Best reference group for coherent forecasting

The best reference groups obtained from three different measures of forecast accuracy during out-of-sample evaluations are given in this section for LL and $\mathrm{LL}_{e_0^\dagger}$. The combinations without mentioning best fitting period utilized the full available date in case of $\mathrm{LL}_{e_0^\dagger}$. German male forecast accuracies were increasing indefinitely, so it is omitted for LL. For $\mathrm{LL}_{e_0^\dagger}$, only the combinations with different best fitting period rather than full observed time are mentioned.

Table 7: Best reference group for males of low-mortality countries according to lowest MAE in LL

Country	Other populations in best reference group	$\lvert\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger\rvert$
Australia	$\mathrm{DEU_M}$	0.06
Austria	$\mathrm{ESP_M}$	0.04
Belgium	$\mathrm{USA_F, DNK_M, CAN_M, DEU_M, AUT_M, AUS_M}$	0.15
Canada	10 populations	0.30
Denmark	$\mathrm{CAN_M}$	0.025
Finland	$\mathrm{NZL_M, ESP_M, AUT_M}$	0.204
France	$\mathrm{PRT_M, FIN_M, NZL_M, ESP_M, AUT_M}$	0.56
Germany	-	-
Ireland	$\mathrm{NOR_M}$	0.00048
Italy	$\mathrm{UK_M, CHE_M, AUS_M}$	0.1
Japan	$\mathrm{NOR_M}$	0.03
Netherlands	$\mathrm{NZL_M, DNK_F, SWE_M}$	0.12
New Zealand	$\mathrm{ESP_M, AUT_M}$	0.049
Norway	$\mathrm{IRL_M}$	0.0004
Portugal	$\mathrm{FRA_M, USA_M, FIN_M, NZL_M, ESP_M, AUT_M}$	0.858
Spain	$\mathrm{NZL_M, AUT_M, CAN_M}$	0.13
Sweden	CAN_M	0.03
Switzerland	$\mathrm{UK_M, ITA_M, IRL_M, NOR_M, JPN_M, AUS_M}$	0.184
UK	$\mathrm{CHE_M, ITA_M, AUS_M, IRL_M, NOR_M, JPN_M, DEU_M, BEL_M}$	0.335
USA	$\mathrm{PRT_M}$	0.31

Source: HMD (2018) and authors' calculations.

Table 8: Best reference group for males of low-mortality countries according to lowest MSE in LL

Country	Other populations in best reference group	$\lvert\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger\rvert$
Australia	$\mathrm{DEU_M}$	0.06
Austria	$\mathrm{ESP_M}$	0.04
Belgium	$\mathrm{USA_F, DNK_M, CAN_M, DEU_M, AUT_M, AUS_M}$	0.15
Canada	$\mathrm{DNK_M, USA_F, BEL_M, AUT_M, ESP_M, DEU_M, NZL_F, AUS_M, FIN_M}$	0.29
Denmark	$\mathrm{CAN_M}$	0.029
Finland	$\mathrm{NZL_M, ESP_M, AUT_M, CAN_M}$	0.29
France	$\mathrm{PRT_M, FIN_M, NZL_M, ESP_M, AUT_M}$	0.56
Germany	-	-
Ireland	$\mathrm{NOR_M, JPN_M, CHE_M, UK_M}$	0.15
Italy	$\mathrm{UK_M, CHE_M, AUS_M, DEU_M, IRL_M}$	0.20
Japan	$\mathrm{NOR_M, IRL_M, CHE_M, UK_M, DNK_M}$	0.206
Netherlands	$\mathrm{NZL_M, DNK_F, SWE_M, CAN_F}$	0.15
New Zealand	$\mathrm{ESP_M, AUT_M, CAN_M, FIN_M, DNK_M, USA_F, BEL_M}$	0.199
Norway	$\mathrm{IRL_M}$	0.0004
Portugal	$\mathrm{FRA_M, USA_M, FIN_M, NZL_M, ESP_M, AUT_M, CAN_M}$	0.95
Spain	13 populations	0.52
Sweden	$\mathrm{CAN_M}$	0.03
Switzerland	$\mathrm{UK_M, ITA_M, IRL_M, NOR_M, JPN_M, AUS_M, DEU_M}$	0.25
UK	$\mathrm{CHE_M, ITA_M, AUS_M, IRL_M, NOR_M, JPN_M, DEU_M}$	0.304
USA	$\mathrm{PRT_M}$	0.31

Source: HMD (2018) and authors' calculations.

Table 9: Best reference group for males of low-mortality countries according to lowest ME(e_0) in LL

Country	Other populations in best reference group	$\bar{e}^{\dagger}_{0i} - \bar{e}^{\dagger}_{0j}$
Australia	$\mathrm{DEU_M, ITA_M}$	0.10
Austria	$\mathrm{ESP_M}$	0.04
Belgium	$\mathrm{USA_F, DNK_M, CAN_M, DEU_M, AUT_M, AUS_M, ESP_M}$	0.19
Canada	10 populations	0.30
Denmark	$\mathrm{CAN_M, USA_F}$	0.029
Finland	$\mathrm{NZL_M, ESP_M, AUT_M}$	0.204
France	$\mathrm{PRT_M, FIN_M, NZL_M, ESP_M, AUT_M}$	0.56
Germany	-	-
Ireland	$\mathrm{NOR_M, JPN_M, CHE_M, UK_M}$	0.15
Italy	$\mathrm{UK_M, CHE_M, AUS_M}$	0.1
Japan	$\mathrm{NOR_M}$	0.03
Netherlands	$\mathrm{NZL_M, DNK_F, SWE_M}$	0.12
New Zealand	$\mathrm{ESP_M}$	0.008
Norway	$\mathrm{IRL_M}$	0.004
Portugal	$\mathrm{FRA_M}$	0.29
Spain	$\mathrm{NZL_M, AUT_M}$	0.04
Sweden	$\mathrm{CAN_M, UK_M, NLD_M}$	0.127
Switzerland	$\mathrm{UK_M, ITA_M, IRL_M, NOR_M, JPN_M, AUS_M}$	0.184
UK	$\mathrm{CHE_M}$	0.03
USA	$\mathrm{PRT_M, FRA_M}$	0.61

Source: HMD (2018) and authors' calculations.

Table 10: Best reference group for females of low-mortality countries according to lowest MAE in LL

Country	Other populations in best reference group	$\bar{e}^{\dagger}_{0i} - \bar{e}^{\dagger}_{0j}$
Australia	$\mathrm{BEL_F, IRL_F, FRA_F}$	0.088
Austria	13 populations	0.39
Belgium	$\mathrm{AUS_F, IRL_F, NLD_F}$	0.12
Canada	$\mathrm{SWE_M, UK_F, NLD_F}$	0.15
Denmark	$\mathrm{NZL_F, NLD_M}$	0.06
Finland	$\mathrm{ITA_F, SWE_F, JPN_F, CHE_F, AUT_F, ESP_F, NOR_F, DEU_F, NLD_F, FRA_F}$	0.36
France	13 populations	0.44
Germany	-	-
Ireland	$\mathrm{AUS_F, BEL_F}$	0.03
Italy	$\mathrm{FIN_F, AUT_F, SWE_F, NOR_F, JPN_F, CHE_F, ESP_F}$	0.15
Japan	$\mathrm{CHE_F, SWE_F}$	0.01
Netherlands	$\mathrm{FRA_F}$	0.05
New Zealand	$\mathrm{DNK_F, NLD_M, , SWE_M}$	0.06
Norway	$\mathrm{AUT_F}$	0.048
Portugal	$\mathrm{IRL_F, AUS_F}$	0.151
Spain	$\mathrm{CHE_F, JPN_F, SWE_F}$	0.08
Sweden	$\mathrm{JPN_F, FIN_F, CHE_F}$	0.046
Switzerland	$\mathrm{JPN_F, ESP_F, SWE_F, FIN_F, ITA_F, AUT_F, NOR_F, DEU_F, NLD_F}$	0.403
UK	$\mathrm{CAN_F, SWE_M, PRT_F, NLD_M, IRL_F}$	0.297
USA	$\mathrm{BEL_M}$	0.002

Source: HMD (2018) and authors' calculations.

Table 11: Best reference group for females of low-mortality countries according to lowest MSE in LL

Country	Other populations in best reference group	$\bar{e}_{0i}^{\dagger} - \bar{e}_{0j}^{\dagger}$
Australia	$\mathrm{BEL_F, IRL_F, FRA_F}$	0.088
Austria	13 populations	0.39
Belgium	$\mathrm{AUS_F, IRL_F, NLD_F}$	0.12
Canada	$\mathrm{SWE_M, UK_F, NLD_F, NZL_F, DNK_F}$	0.22
Denmark	$\mathrm{NZL_F, NLD_M}$	0.06
Finland	13 populations	0.44
France	13 populations	0.44
Germany	11 populations	0.31
Ireland	$\mathrm{AUS_F, BEL_F}$	0.03
Italy	$\mathrm{FIN_F, AUT_F, SWE_F, NOR_F, JPN_F}$	0.11
Japan	$\mathrm{CHE_F, SWE_F}$	0.01
Netherlands	$\mathrm{FRA_F}$	0.05
New Zealand	$\mathrm{DNK_F}$	0.003
Norway	$\mathrm{AUT_F, DEU_F}$	0.09
Portugal	$\mathrm{IRL_F, AUS_F}$	0.151
Spain	$\mathrm{CHE_F, JPN_F, SWE_F}$	0.08
Sweden	$\mathrm{JPN_F, FIN_F, CHE_F, ITA_F}$	0.07
Switzerland	$\mathrm{JPN_F, ESP_F, SWE_F, FIN_F}$	0.088
UK	$\mathrm{CAN_F, SWE_M, PRT_F, NLD_M}$	0.248
USA	$\mathrm{BEL_M}$	0.002

Source: HMD (2018) and authors' calculations.

Table 12: Best reference group for females of low-mortality countries according to lowest ME(e_0) in LL

Country	Other populations in best reference group	$\bar{e}_{0i}^{\dagger} - \bar{e}_{0j}^{\dagger}$
Australia	$\mathrm{BEL_F, IRL_F, FRA_F, NLD_F, PRT_F, DEU_F}$	0.22
Austria	$\mathrm{NOR_F, ITA_F, FIN_F, SWE_F, DEU_F, JPN_F, CHE_F}$	0.17
Belgium	$\mathrm{AUS_F, IRL_F, NLD_F, PRT_F}$	0.16
Canada	$\mathrm{SWE_M}$	0.03
Denmark	$\mathrm{NZL_F, NLD_M}$	0.06
Finland	$\mathrm{ITA_F, SWE_F, JPN_F, CHE_F}$	0.04
France	$\mathrm{NLD_F, BEL_F, AUS_F, IRL_F, DEU_F, NOR_F, PRT_F, AUT_F, ITA_F}$	0.33
Germany	$\mathrm{NLD_F}$	0.079
Ireland	$\mathrm{AUS_F}$	0.02
Italy	$\mathrm{FIN_F}$	0.03
Japan	$\mathrm{CHE_F, SWE_F}$	0.01
Netherlands	$\mathrm{FRA_F}$	0.05
New Zealand	$\mathrm{DNK_F, , NLD_F, SWE_M, JPN_M, CAN_F, NOR_M, IRL_M}$	0.241
Norway	$\mathrm{AUT_F}$	0.048
Portugal	$\mathrm{IRL_F, AUS_F, BEL_F, UK_F, FRA_F}$	0.24
Spain	$\mathrm{CHE_F, JPN_F, SWE_F, FIN_F, ITA_F, AUT_F, NOR_F, DEU_F, NLD_F}$	0.43
Sweden	$\mathrm{JPN_F, FIN_F}$	0.041
Switzerland	$\mathrm{JPN_F, ESP_F, SWE_F, FIN_F, ITA_F, AUT_F, NOR_F, DEU_F, NLD_F}$	0.403
UK	$\mathrm{CAN_F, SWE_M}$	0.12
USA	$\mathrm{BEL_M, DNK_M, CAN_F, DEU_M}$	0.084

Source: HMD (2018) and authors' calculations.

Table 13: Best reference group for males of low-mortality countries according to lowest MAE in $\text{LL}_{e_0^\dagger}$

Country	Other populations in best reference group (best fitting period from:)	$\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger$
Australia	DEU_M	0.06
Austria	ESP_M	0.04
Belgium	9 populations	0.199
Canada	10 populations	0.30
Denmark	$\text{CAN}_\text{M}, \text{USA}_\text{F}, \text{BEL}_\text{M}$	0.03
Finland	$\text{NZL}_\text{M}, \text{ESP}_\text{M}, \text{AUT}_\text{M}$	0.204
France	PRT_M	0.29
Germany	12 populations	0.376
Ireland	$\text{NOR}_\text{M}, \text{JPN}_\text{M}, \text{CHE}_\text{M}, \text{UK}_\text{M}$	0.15
Italy	$\text{UK}_\text{M}, \text{CHE}_\text{M}, \text{AUS}_\text{M}$	0.1
Japan	NOR_M	0.03
Netherlands	$\text{NZL}_\text{M}, \text{DNK}_\text{F}, \text{SWE}_\text{M}, \text{CAN}_\text{F}$ (1974:)	0.15
New Zealand	ESP_M	0.008
Norway	IRL_M	0.0004
Portugal	9 populations (1965:)	1.00
Spain	$\text{NZL}_\text{M}, \text{AUT}_\text{M}, \text{CAN}_\text{M}, \text{CAN}_\text{M}$	0.13
Sweden	CAN_M	0.03
Switzerland	$\text{UK}_\text{M}, \text{ITA}_\text{M}$	0.08
UK	10 populations	0.33
USA	$\text{PRT}_\text{M}, \text{FRA}_\text{M}, \text{FIN}_\text{M}, \text{NZL}_\text{M}, \text{ESP}_\text{M}, \text{AUT}_\text{M}$ (1965:)	1.17

Source: HMD (2018) and authors' calculations.
Note: For combinations without mentioned best fitting period have best fit for (1956:2001).

Table 14: Best reference group for males of low-mortality countries according to lowest MSE in $\text{LL}_{e_0^\dagger}$

Country	Other populations in best reference group (best fitting period from:)	$\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger$
Australia	DEU_M	0.06
Austria	ESP_M	0.04
Belgium	$\text{USA}_\text{F}, \text{DNK}_\text{M}, \text{CAN}_\text{M}, \text{DEU}_\text{M}, \text{AUT}_\text{M}, \text{AUS}_\text{M}, \text{ESP}_\text{M}$	0.19
Canada	$\text{DNK}_\text{M}, \text{USA}_\text{F}, \text{BEL}_\text{M}, \text{AUT}_\text{M}, \text{ESP}_\text{M}, \text{DEU}_\text{M}, \text{NZL}_\text{F}, \text{AUS}_\text{M}, \text{FIN}_\text{M}$	0.29
Denmark	$\text{CAN}_\text{M}, \text{USA}_\text{F}, \text{BEL}_\text{M}, \text{DEU}_\text{M}$	0.11
Finland	$\text{NZL}_\text{M}, \text{ESP}_\text{M}, \text{AUT}_\text{M}, \text{CAN}_\text{M}$	0.29
France	PRT_M	0.29
Germany	13 populations	0.377
Ireland	$\text{NOR}_\text{M}, \text{JPN}_\text{M}, \text{CHE}_\text{M}, \text{UK}_\text{M}$	0.15
Italy	$\text{UK}_\text{M}, \text{CHE}_\text{M}, \text{AUS}_\text{M}, \text{DEU}_\text{M}$	0.16
Japan	NOR_M	0.03
Netherlands	$\text{NZL}_\text{M}, \text{DNK}_\text{F}, \text{SWE}_\text{M}, \text{CAN}_\text{F}$ (1974:)	0.15
New Zealand	ESP_M	0.008
Norway	$\text{IRL}_\text{M}, \text{JPN}_\text{M}, \text{CHE}_\text{M}, \text{UK}_\text{M}$	0.15
Portugal	9 populations (1965:)	1.00
Spain	13 populations	0.52
Sweden	CAN_M	0.03
Switzerland	$\text{UK}_\text{M}, \text{ITA}_\text{M}$	0.08
UK	10 populations	0.33
USA	$\text{PRT}_\text{M}, \text{FRA}_\text{M}, \text{FIN}_\text{M}, \text{NZL}_\text{M}, \text{ESP}_\text{M}, \text{AUT}_\text{M}, \text{DEU}_\text{M}$ (1965:)	1.40

Source: HMD (2018) and authors' calculations.
Note: For combinations without mentioned best fitting period have best fit for (1956:2001).

Table 15: Best reference group for males of low-mortality countries according to lowest ME(e_0) in LL$_{e_0^\dagger}$

Country	Other populations in best reference group (best fitting period from:)	$\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger$
Australia	$\mathrm{DEU_M, ITA_M}$	0.10
Austria	$\mathrm{ESP_M}$	0.04
Belgium	9 populations	0.25
Canada	10 populations	0.30
Denmark	$\mathrm{CAN_M, USA_F, BEL_M}$	0.03
Finland	$\mathrm{NZL_M, ESP_M, AUT_M, DNK_M, USA_F}$ (1964:)	0.35
France	$\mathrm{PRT_M}$	0.29
Germany	12 populations	0.376
Ireland	$\mathrm{NOR_M, JPN_M, CHE_M, UK_M}$	0.15
Italy	$\mathrm{UK_M, CHE_M, AUS_M}$	0.1
Japan	$\mathrm{NOR_M}$	0.03
Netherlands	$\mathrm{NZL_M, DNK_F, SWE_M, CAN_F}$ (1974:)	0.15
New Zealand	$\mathrm{ESP_M}$	0.008
Norway	$\mathrm{IRL_M}$	0.004
Portugal	$\mathrm{FRA_M, USA_M, FIN_M, NZL_M}$ (1965:)	1.00
Spain	$\mathrm{NZL_M, AUT_M}$	0.04
Sweden	$\mathrm{CAN_M}$	0.03
Switzerland	$\mathrm{UK_M, ITA_M}$	0.08
UK	10 populations	0.33
USA	$\mathrm{PRT_M, FRA_M, FIN_M, NZL_M}$ (1965:)	1.12

Source: HMD (2018) and authors' calculations.
Note: For combinations without mentioned best fitting period have best fit for (1956:2001).

Table 16: Best reference group for females of low-mortality countries according to lowest MAE in LL$_{e_0^\dagger}$

Country	Other populations in best reference group (best fitting period from:)	$\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger$
Australia	$\mathrm{BEL_F, IRL_F, FRA_F}$	0.088
Austria	$\mathrm{NOR_F, ITA_F, FIN_F, SWE_F, DEU_F, JPN_F, CHE_F, ESP_F, NLD_F}$ (1957:)	0.22
Belgium	$\mathrm{AUS_F, IRL_F, NLD_F, PRT_F, DEU_F, NOR_F}$ (1974:)	0.30
Canada	$\mathrm{SWE_M, UK_F, NLD_F, NZL_F}$	0.21
Denmark	$\mathrm{NZL_F, NLD_M, SWE_M,}$	0.19
Finland	$\mathrm{ITA_F, SWE_F, JPN_F, CHE_F}$	0.08
France	$\mathrm{NLD_F}, BEL_\mathrm{F}, \mathrm{AUS_F, IRL_F, DEU_F, NOR_F, PRT_F, AUT_F}$ (1974:)	0.33
Germany	$\mathrm{NLD_F, NOR_F, FRA_F, AUT_F, ITA_F}$	0.20
Ireland	$\mathrm{AUS_F, BEL_F}$	0.03
Italy	$\mathrm{FIN_F, AUT_F, SWE_F, NOR_F, JPN_F, CHE_F, ESP_F, DEU_F}$ (1957:)	0.20
Japan	$\mathrm{CHE_F, SWE_F, ESP_F, FIN_F}$	0.07
Netherlands	$\mathrm{FRA_F, DEU_F, BEL_F, AUS_F}$ (1977:)	0.14
New Zealand	$\mathrm{DNK_F, NLD_M, SWE_M, JPN_M, CAN_F, NOR_M}$ (1976:)	0.24
Norway	$\mathrm{AUT_F}$	0.048
Portugal	$\mathrm{IRL_F, AUS_F, BEL_F, UK_F, FRA_F}$ (1966:)	0.24
Spain	$\mathrm{CHE_F, JPN_F, SWE_F, FIN_F, ITA_F, AUT_F, NOR_F}$ (1957:)	0.26
Sweden	$\mathrm{JPN_F, FIN_F, CHE_F}$	0.04
Switzerland	13 populations (1958:)	0.57
UK	$\mathrm{CAN_F}$ (1972:)	0.08
USA	$\mathrm{BEL_M}$	0.002

Source: HMD (2018) and authors' calculations.
Note: For combinations without mentioned best fitting period have best fit for (1956:2001).

Table 17: Best reference group for females of low-mortality countries according to lowest MSE in $\text{LL}_{e_0^\dagger}$

Country	Other populations in best reference group (best fitting period from:)	$\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger$
Australia	BEL_F	0.012
Austria	$\text{NOR}_\text{F}, \text{ITA}_\text{F}, \text{FIN}_\text{F}, \text{SWE}_\text{F}$	0.13
Belgium	$\text{AUS}_\text{F}, \text{IRL}_\text{F}, \text{NLD}_\text{F}, \text{PRT}_\text{F}, \text{DEU}_\text{F}, \text{NOR}_\text{F}$ (1974:)	0.30
Canada	$\text{SWE}_\text{M}, \text{UK}_\text{F}, \text{NLD}_\text{F}, \text{NZL}_\text{F}$	0.21
Denmark	$\text{NZL}_\text{F}, \text{NLD}_\text{M}, \text{SWE}_\text{M},$	0.19
Finland	$IT_\text{F}, \text{SWE}_\text{F}, \text{JPN}_\text{F}, \text{CHE}_\text{F}$	0.08
France	$\text{NLD}_\text{F}, NEL_\text{F}, \text{AUS}_\text{F}, \text{IRL}_\text{F}, \text{DEU}_\text{F}, \text{NOR}_\text{F}, \text{PRT}_\text{F}, \text{AUT}_\text{F}$ (1974:)	0.33
Germany	$\text{NLD}_\text{F}, \text{NOR}_\text{F}, \text{FRA}_\text{F}, \text{AUT}_\text{F}, \text{ITA}_\text{F}, \text{BEL}_\text{F}, \text{AUS}_\text{F}$ (1977:)	0.22
Ireland	$\text{AUS}_\text{F}, \text{BEL}_\text{F}$	0.03
Italy	$\text{FIN}_\text{F}, \text{AUT}_\text{F}, \text{SWE}_\text{F}, \text{NOR}_\text{F}, \text{JPN}_\text{F}, \text{CHE}_\text{F}, \text{ESP}_\text{F}, \text{DEU}_\text{F}, \text{NLD}_\text{F}$	0.28
Japan	$\text{CHE}_\text{F}, \text{SWE}_\text{F}, \text{ESP}_\text{F}, \text{FIN}_\text{F}$	0.07
Netherlands	$\text{FRA}_\text{F}, \text{DEU}_\text{F}, \text{BEL}_\text{F}, \text{AUS}_\text{F}$ (1977:)	0.14
New Zealand	$\text{DNK}_\text{F}, \text{NLD}_\text{M}, , \text{SWE}_\text{M}, \text{JPN}_\text{M}, \text{CAN}_\text{F}, \text{NOR}_\text{M}$ (1976:)	0.24
Norway	AUT_F	0.04
Portugal	$\text{IRL}_\text{F}, \text{AUS}_\text{F}, \text{BEL}_\text{F}, \text{UK}_\text{F}, \text{FRA}_\text{F}$ (1966:)	0.24
Spain	$\text{CHE}_\text{F}, \text{JPN}_\text{F}, \text{SWE}_\text{F}, \text{FIN}_\text{F}, \text{ITA}_\text{F}, \text{AUT}_\text{F}, \text{NOR}_\text{F}$ (1957:)	0.26
Sweden	$\text{JPN}_\text{F}, \text{FIN}_\text{F}, \text{CHE}_\text{F}$	0.04
Switzerland	10 populations (1957:)	0.45
UK	CAN_F (1972:)	0.08
USA	BEL_M	0.002

Source: HMD (2018) and authors' calculations.
Note: For combinations without mentioned best fitting period have best fit for (1956:2001).

Table 18: Best reference group for females of low-mortality countries according to lowest ME(e_0) in $\text{LL}_{e_0^\dagger}$

Country	Other populations in best reference group (best fitting period from:)	$\bar{e}_{0i}^\dagger - \bar{e}_{0j}^\dagger$
Australia	$\text{BEL}_\text{F}, \text{IRL}_\text{F}, \text{FRA}_\text{F}$	0.088
Austria	NOR_F	0.04
Belgium	AUS_F	0.012
Canada	$\text{SWE}_\text{M}, \text{UK}_\text{F}, \text{NLD}_\text{F}, \text{NZL}_\text{F}, \text{DNK}_\text{F}$	0.22
Denmark	$\text{NZL}_\text{F}, \text{NLD}_\text{M}, \text{SWE}_\text{M},$	0.19
Finland	$\text{ITA}_\text{F}, \text{SWE}_\text{F}, \text{JPN}_\text{F}, \text{CHE}_\text{F}$	0.08
France	$\text{NLD}_\text{F}, \text{BEL}_\text{F}, \text{AUS}_\text{F}, \text{IRL}_\text{F}, \text{DEU}_\text{F}, \text{NOR}_\text{F}, \text{PRT}_\text{F}, \text{AUT}_\text{F}$	0.33
Germany	$\text{NLD}_\text{F}, \text{NOR}_\text{F}, \text{FRA}_\text{F}, \text{AUT}_\text{F}, \text{ITA}_\text{F}$	0.20
Ireland	$\text{AUS}_\text{F}, \text{BEL}_\text{F}, \text{FRA}_\text{F}$	0.11
Italy	FIN_F	0.03
Japan	$\text{CHE}_\text{F}, \text{SWE}_\text{F}, \text{ESP}_\text{F}, \text{FIN}_\text{F}$	0.07
Netherlands	FRA_F	0.05
New Zealand	DNK_F	0.003
Norway	AUT_F	0.04
Portugal	$\text{IRL}_\text{F}, \text{AUS}_\text{F}, \text{BEL}_\text{F}, \text{UK}_\text{F}, \text{FRA}_\text{F}, \text{CAN}_\text{F}, \text{NLD}_\text{F}, \text{SWE}_\text{M}, \text{DEU}_\text{F}$ (1966:)	0.37
Spain	$\text{CHE}_\text{F}, \text{JPN}_\text{F}, \text{SWE}_\text{F}, \text{FIN}_\text{F}, \text{ITA}_\text{F}, \text{AUT}_\text{F}, \text{NOR}_\text{F}$ (1957:)	0.26
Sweden	$\text{JPN}_\text{F}, \text{FIN}_\text{F}, \text{CHE}_\text{F}$	0.04
Switzerland	13 populations	0.57
UK	CAN_F (1972:)	0.08
USA	BEL_M	0.002

Source: HMD (2018) and authors' calculations.
Note: For combinations without mentioned best fitting period have best fit for (1956:2001).