The importance of the reference population for coherent mortality forecasting models

Søren Kjærgaard†, Vladimir Canudas-Romo†, and James W. Vaupel†

†Max Planck Odense Center on the Biodemography of Aging,Institute of Public Health,University of Southern Denmark,

December 14, 2015

We compare medium and long term life expectancy forecasts for two multi-population mortality models in order to find the optimal set of countries to use a reference population. The reference set population is calculated taking into account all the possible combinations of available countries. These different reference populations possibilities are compared by their forecast performance using the Root Mean Squared Forecast Error. The two multi-population mortality models used are the Li-Lee model and the Double-Gap Life Expectancy Forecasting model. The preliminary results show that the selection of countries for multi-population mortality models has a huge effect on the models life expectancy forecasts for Danish females, and lesser relevance for the Spanish women.

Keywords: Mortality forecasting, Coherent forecast, Life expectancy, Li-Lee

1 Introduction

In recent years, life expectancy has increased with around 3 months per year for many low mortality countries (Oeppen and Vaupel, 2002). The high increase in life expectancy introduces planning problems for pensions funds and governments as a higher life expectancy implies higher pension payments for most European countries (Stoeldraijer et al., 2013). Accurate forecasts of the life expectancy is, therefore, an extremely important issue.

A very large number of different models have been suggested to forecast the life expectancy. A fairly used and studied model is the Lee-Carter (1992) model. The model has been subject to a lot of critique and various modifications and extensions have been suggested in order to improve it (Lee and Miller, 2001; Li and Lee, 2005; Hyndman and Booth, 2008; Cairns et al., 2009; Hyndman et al., 2013). The original Lee-Carter model is a single population model and often used for females and males independently. Hence, the Lee-Carter model can produce forecasts where females and males in the same country are predicted to follow different trends. The Li-Lee (2005) model provides a solution to this problem by modelling a single population in reference to another coherent population. Several other models followed the idea of the Li-Lee model and we refer to these models as multi-population models or coherent mortality models.

An important factor in the coherent mortality models is the reference population or group of populations that are modelled together. Little work has been done on the importance or selection of the reference group and there is no standard procedure on how to select populations for the reference population. Li and Lee (2005) select the reference group based on the level of explained variation. Hatzopoulos and Haberman (2013) select based on a Fuzzy C-means cluster analysis and find 2 clusters among the countries in the Human Mortality Database.

In this paper we analyse the selection of the reference group based on the ability to forecast the life expectancy. We compare different forecasts by calculating the Root Mean Squared Forecast Error(RMSE) and compare the RMSE for different models and reference populations. Hence, we focus on the models ability to forecast instead of the in-sample fit as Li and Lee (2005) and Hatzopoulos and Haberman (2013). The study is carried by both varying the number of countries in the reference group and which countries to include. We only consider the life expectancy for females as the life expectancy forecast for males can be computed by forecasting the gap between females and males see Pascariu et al. (2015) for more details.

The remainder of this paper is organised as follows. First, section 2 presents the models used in the analysis. Section 3 explains how different reference populations are selected for the analysis.

Section 4 presents the data used for the analysis. Finally, section 5 presents and discusses the preliminary results.

2 The methods

In this paper we chose to focus on two multi-population models that is the Li-Lee (2005) model and the Double-gap life expectancy forecasting model (DG) by Pascariu et al. (2015). We choose the Li-Lee model as it is a natural extension of the Lee-Carter model which has become a standard mortality forecasting tool. Both the Lee-Carter model and the Li-Lee model use death rates to forecast the life expectancy. Another approach is to forecast the life expectancy directly as the DG model does.

2.1 The Lee-Carter model

The Lee-Carter model is a single population model which by the use of one factor fits and forecasts central death rates. The Lee-Carter model was suggested by Lee and Carter (1992) and several modifications have been suggested afterwards, see Booth et al. (2006) for a comparison. We choose to use the modification suggested by Lee and Miller (2001). The Lee-Carter model can be written as,

$$ln(m_{x,t}) = \alpha_x + \beta_x k_t + \epsilon_{x,t}, \tag{1}$$

where $m_{x,t}$ is the central death rate in year t at age x, α_x represents the general age-pattern of mortality, k_t is a mortality index and describes average level of mortality over time, β is a loading parameter which adjust age specific changes in mortality in proportion to k_t, and $\epsilon_{x,t}$ is the estimation error (Booth et al., 2006).

The model can be estimated using a Singular Value Decomposition (SVD) procedure which minimizes the squared estimation errors. Lee and Miller (2001) suggest that the k_t estimated are refitted using life expectancy at birth (e_0).

The Lee-Carter model forecasts the central deaths rates by extrapolating k_t using an ARIMA(p,d,q) model and multiply the k_t forecast on the α and β estimates according to the model. The random walk with drift is often found to be a suitable ARIMA model (Tuljapurkar et al., 2000).

2.2 Li-Lee Model

The Li-Lee model extends the Lee-Carter model by modelling the central death rate for a single country in reference to a reference population (Li and Lee, 2005). The Le-Lee model allows for differences between the population under consideration and the reference population in the short-run but binds the two populations to a constant ratio in the long-run (Stoeldraijer et al., 2013).

The Li-Lee model can be written as,

$$ln(m_{x,t,i}) = \alpha_{x,i} + B_x K_t + \beta_{x,i} k_{t,i} + \epsilon_{x,t,i} \qquad (2)$$

where i is a country indicator, $\alpha_{x,i}$ represents the general age-pattern of mortality averaged over time for each country, K_t is a mortality index capturing the main time trend for the reference population, and B_x is the relative speed of change in mortality at each age when the index K_t changes. The term $B_x K_t$ is know as the common factor which is common to all countries considered in the analysis. The parameter $k_{t,i}$ together with $\beta x, i$ correspond to country specific fluctuations with $k_{t,i}$ as the time variation, and $\beta_{x,i}$ as the age specific pattern for each country. Finally, $\epsilon_{x,t,i}$ is the age and country specific estimation error.

The Li-Lee model is estimated by first estimating a Lee-Carter model using death rates from the reference population which provides estimates for K_t and B_x. $\alpha_{x,i}$ can be found as the average over the log death rates for each country and, finally $\beta_{x,i}$ and $\alpha_{x,i}$ can be estimated by running a Lee-Carter model on the residuals from the log mortality rates minus estimates for the 3 other parameters. That is, $ln(m_{x,t,i}) - \hat{\alpha_{x,i}} - \hat{K_t}\hat{B_x} = \hat{\beta_{x,i}}\hat{k_{x,i}} + \epsilon_{x,t,i}$.

Construction of death rates for the reference population

The death rates for the reference population can be composed in different ways e.g as a simple average or population weighted average over the country specific death rates. Li and Lee (2005) note that if the goal is to fit the experience of each country the simple average should be applied.

Hyndman et al. (2013) suggest, in a extension of the Li-Lee model, that the reference population should be calculated using the geometric mean as it indicates the central tendency of the data. We follow this suggestion and compute the death rates for the reference population as $\bar{m}_{x,t} = \left(\prod_{i=1}^{I} m_{x,t,i}\right)^{1/I}$.

Li-Lee record life expectancy variant

In this paper, we also suggest a new variant of the Li-Lee model. The variant constructs the death rates for the reference population as a weighted average using one minus the percentage difference from the record life expectancy at birth as weight. This means that the death rates from the country with the record life expectancy, in a specific year, receives the highest weight. The other countries receives weights according to the difference from the record. Table 1 illustrates how the weights are calculated for three countries.

Table 1: Weight table illustration

Country	Sweden	France	Norway
Life expectancy	82	79	80
One minus percentage difference	1	0.975	0.98
Weight	0.338	0.332	0.33

2.3 The Double Gap life expectancy model (DG)

The Double Gap life expectancy forecasting model(DG) suggested by Pascariu et al. (2015) forecasts the life expectancy for a specific country in reference to the record life expectancy introduced by Oeppen and Vaupel (2002). The record life expectancy shows a close to linear pattern and, hence, Pascariu et al. (2015) estimates a linear model.

$$e_{x,t}^{record} = \alpha_{x0} + \alpha_{x1} t + \epsilon_{x,t}^{(0)},$$

where $t = 1, 2, 3, ...$, $e_{x,t}^{record}$ is the record life expectancy, α_{x0} and α_{x1} constant parameters, and $\epsilon_{x,t}^{(0)}$ is the estimation error.

From the linear regression the best-practise trend is defined as

$$e_{x,t}^{bp} = \hat{\alpha}_{x0} + \hat{\alpha}_{x1} t.$$

The best-practise trend is forecasted by extrapolating it linearly. The gap between the best practise trend and life expectancy in a specific country is defined as

$$D_{x,t,i} = e_{x,t}^{bp} - e_{x,t,i}, \tag{3}$$

where $D_{x,t,i}$ is the country-specific gap and $e_{x,t,i}$ is the country-specific life expectancy. The gap is fitted and forecasted using an ARIMA(p,d,q) model. The most appropriated ARIMA model can be found by standard Box-Jenkins methods.

The country specific life expectancy forecasts are, thereby, given by

$$\tilde{e}_{x,t} = \tilde{e}_{x,t}^{bp} - \tilde{D}_{x,t,i}, \tag{4}$$

where tilde indicates forecasted values.

Table 2 Summarizes important differences between the models.

Table 2: Models,input and reference population

Models	Life table value	Reference population
Lee-Carter	Central death rates	single population
Li-Lee	Central death rates	average
Li-Lee record	Central death rates	weighted average
DG	Life expectancy	record

3 Selection of reference population

This paper seeks to find the optimal reference population for forecasting the life expectancy for Danish females. Hence, the most appropriate reference population is chosen by its ability to minimize the Root Mean Squared Forecast Error (RMSE). The RMSE is calculated using a 15 and 20 years out-of-sample forecast period meaning that we analyse the models ability to forecast these latter years. That is,

$$RMSE = \left(\frac{1}{H} \sum_{h=1}^{H} (e_{x,t+h} - \tilde{e}_{x,t+h})^2 \right)^{\frac{1}{2}},$$

where h is the different forecast years and H the total number of forecast years. For example for the Danish females h is running from 1991 to 2011 as 2011 is the last available year for Denmark.

The optimal reference population could vary with respect to which and how many input countries are included. Hence, the RMSE is calculated by varying both the number and which countries that enter the reference group. We only vary the number of countries between 2,3,5,10 and 15 countries. Table 3 shows the number of combinations for each number of countries.

Table 3: Number of combinations for the reference population based on selection of 21 countries

Number of countries	Number of combinations
2	210
3	1330
5	20349
10	352716
15	54264
21	1

4 Data

Data is from the Human Mortality Database (Human Mortality Database, 2015). We include only data with the highest possible standard and where long time series are available without warnings. Hence the following 21 countries are considered: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, UK, and USA.

5 Preliminary Results

Due to the large number of possible combinations of reference populations the RMSE is not calculated for all combinations yet. We show in this version some preliminary results for selected reference populations using the Lee-Carter model, Li-Lee model with unweighed average and the DG model.

Figure 1 shows 50 years ahead forecast using the Li-Lee model and a Nordic (Denmark, Norway, Sweden), a European(Germany, Denmark, France, Norway, Sweden, UK) and an International (all 21 countries) reference population. Forecasts using the DG model are only calculated for the Nordic and international reference population as forecasts for the European reference population were very similar to those for the Nordic.

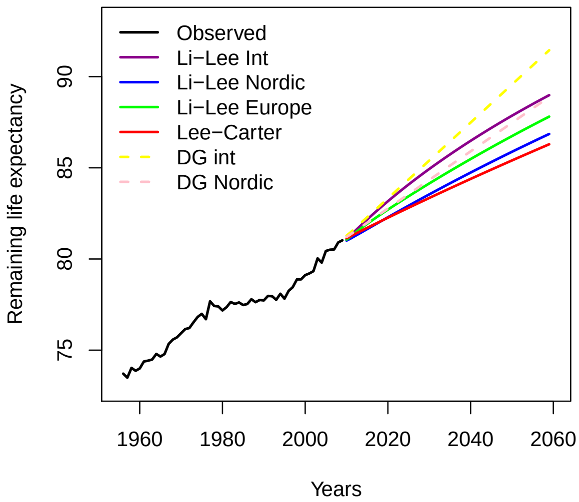

Figure 1: 50 years forecast for Danish females using the Lee-Carter model, the Li-Lee model and the DG model.

The 50 years ahead forecasts show that the three models produce very different forecast and that the coherent models highly depend on the used reference population. The DG model using an international reference population has the highest life expectancy forecast and the Le-Carter model the lowest with a different between the two at around 6 years in 2059.

The 50 years forecasts differ also across references-populations for each of the models. For the DG model the difference in forecasted life expectancy is around three years in 1959 between the Nordic and the International reference population. For the Li-Lee model the difference from the International reference population and to the Nordic and the European is respectively 1 and 1.5 years.

The International reference population produces the highest forecast for both models and the Nordic the lowest. This results is a natural consequence of mortality development for the included countries. The Nordic reference population produces low forecast for the Li-Lee model because the Danish women has had a low increase in the life expectancy from 1950 to 2009 compared to other of the included countries. The International reference population includes Japanese and French females which have had a large increase in the life expectancy from 1950 to 2009 compared to Danish females and, hence, extrapolated trends from the International reference population will lead to higher forecast.

The DG model uses the record life expectancy and hence a large reference population will always lead to a higher or as high forecasts compared to a small reference. Given that the small reference population includes populations that are in the large. This is the case because, inclusion of a record country increase the best practice trend in the DG model and inclusion of the lower than record country does not affect the model.

Table 4 shows the RMSE for a 15 and 20 years out-of-sample forecast for Danish females using the Lee-Carter model, the Li-Lee model and the DG model and the similar reference populations as in Figure 1. The last available data year is 2011 for the Danish females. Hence, the 20 years out-of-sample forecast uses data from 1950 to 1991 and the 15 years data from 1950 to 1996.

Model	Reference population	RMSE 20 years (ranking)	RMSE 15 years (ranking)
Lee-Carter	Denmark	0.59(3)	0.88(6)
Li-Lee	Nordic	0.39(1)	0.55(5)
Li-Lee	Europe	0.74(5)	0.23(3)
Li-Lee	International	0.98(6)	0.21(2)
DG	Nordic	0.43(2)	0.37(4)
DG	International	0.61(4)	0.20(1)

Table 4: Forecast performance for the Lee-Carter model, the Li-Lee model and the DG model for 15 and 20 years out-of-sample forecast for Danish females

Table 4 shows that the Li-Lee model with a Nordic reference population fits the Danish life expectancy best on a 20 years horizon and the DG model on a 15 years horizon. Hence, non of the models or reference populations is best for both forecast horizons.

Figure 2 shows 50 years forecast for the Spanish females using the different models and reference populations. The different reference populations are the same as for the Danish females except for the Nordic group where France and Portugal are used in the Li-Lee model and the European for the DG model.

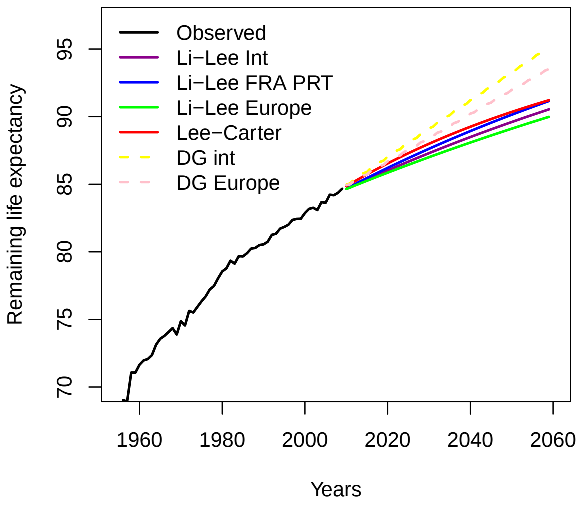

Figure 2: 50 years forecast for Spainsh females using the Lee-Carter model, the Li-Lee model and the DG model. FRA = France and PRT = Portugal

Figure 2 shows that, as for the Danish females, that life expectancy forecasts depend on the model and the reference population. The reference population seems to be less important for the Spanish females as the forecast generally lies closer.

Model	Reference population	RMSE 20 years (ranking)	RMSE 15 years (ranking)
Lee-Carter	Spain	0.17(2)	0.20(2)
Li-Lee	France,Portugal	0.53(4)	0.43(4)
Li-Lee	Europe	0.93(6)	0.71(6)
Li-Lee	International	0.67(5)	0.55(5)
DG	Europe	0.33(3)	0.21(3)
DG	International	0.17(1)	0.19(1)

Table 5: Forecast performance for the Lee-Carter model, the Li-Lee model and the DG model for 15 and 20 years out-of-sample forecast for Spanish females

Table 5 shows the forecast performance for the different models and reference populations when forecasting the life expectancy for Spanish females. Here, the Lee-Carter model and DG model with international reference populations fits the life expectancy for Spanish females best.

The preliminary results, in this section, shows that different coherent forecasting models and reference populations produce very different life expectancy forecasts. The forecast performance for the models might also depend on the country under consideration and, hence, the final version of the paper will focus on France, the USA and Spain besides Denmark. Similar is the the forecast horizon important for the conclusions and ,hence, the final version of the paper will also include 25 and 30 out-of-sample forecasts.

References

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*.

Cairns, A. J. G., Blake, D., Down, K., Coughland, G. D., Epstein, D., Ong, A., and Balevich, I. 2009. A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, 13:1–35.

Hatzopoulos, P. and Haberman, S. 2013. Common mortality modeling and coherent forecasts. an empirical analysis of worldwide mortality data. *Insurance: Mathematics and Economics*, 52(2):320 – 337.

Human Mortality Database 2015. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Available: www.mortality. org or www.humanmortality.de (accessed: 01/08/2015).

Hyndman, J. R. and Booth, H. 2008. Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24:323–342.

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.

Lee, R. D. and Carter, L. R. 1992. Modeling and forecasting u. s. mortality. *Journal of the American Statistical Association*, 87(419):659–671.

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

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

Oeppen, J. and Vaupel, J. W. 2002. Demography - broken limits to life expectancy. *Science*, 296:1029–1031.

Pascariu, M., Canudas-Romo, V., and Vaupel, J. W. 2015. The double-gap life expectancy forecasting model. *Forthcomning*, unknown.

Stoeldraijer, L., van Duin, C., van Wissen, L. J. G., and Janssen, F. 2013. Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research*, 29(13):323–354.

Tuljapurkar, S., Li, N., and Boe, C. 2000. A universal pattern of mortality decline in the g7 countries. *Nature*, 405(405):789–792.