Validation and mortality forecasting: when to choose Lee-Carter over more complex methods?

Ricarda Duerst[1]
Christina Bohk-Ewald[1]

[1]*Max Planck Institute for Demographic Research
Rostock, Germany*

Submission to the EPC 2020

Abstract

The Lee-Carter model is widely used to forecast mortality by age over time. Since its introduction in 1992, a variety of methods with increasing methodological complexity have been proposed to fit and predict particular mortality developments more closely. Considering that so many simple and also complex methods are available our study has two main goals. First, to establish validation as test prior to mortality forecasting to confirm that a selected method is suitable. Second, to show that the simple Lee-Carter model is broadly applicable, particularly in mortality contexts without substantial trend changes and decisive shifts in the age at death distribution. We adopt a detailed and objective validation design to assess the accuracy and bias of Lee-Carter forecasts 20 years ahead. For our analysis, we use age-specific mortality rates of all available calendar years for 24 highly-developed countries from the Human Mortality Database. We quantify Lee-Carter's forecast performance with the percentage error for life expectancy and life span disparity at birth in three analytical settings. Most strikingly, we find that the simple Lee-Carter model is indeed suitable to forecast mortality for many highly-developed countries in the most recent years. More complex methods, however, would have been more suitable than Lee-Carter's model between 1960 and 2000, when discontinuous mortality developments have been prevalent in many countries. Finally, we show that validation serves as a meaningful first test to decide whether a method is likely to be appropriate to forecast mortality in a country of interest.

1 Introduction

The model of Lee and Carter (1992) is a golden standard in mortality forecasting due to its methodological simplicity and robust forecast performance across various mortality settings. It forecasts mortality in a country of interest assuming that mortality will change over age and time in the future as it has changed in the past. Since the introduction of the Lee-Carter model in 1992, several extensions have been proposed to fit and predict particular mortality developments more closely (e.g. Booth et al., 2006; Booth, 2006; Booth and Tickle, 2008). For example, considering the advancement of survival improvements to increasingly older ages (e.g. Rau et al., 2008), more recent approaches account for trends in rates of mortality improvement and in the distribution of ages at death (e.g. Haberman and Renshaw, 2012; Li et al., 2013; Ševčíková et al., 2016; Bohk-Ewald and Rau, 2017; de Beer et al., 2017; Bardoutsos et al., 2018; Basellini and Camarda, 2019; Camarda, 2019). Other methodological trends in mortality forecasting are to account for mortality that is attributable to health behavior such as smoking (e.g. Vogt et al., 2017; Janssen et al., 2013; Wang and Preston, 2009) and for mortality developments in other countries (e.g. Li and Lee, 2005; Hyndman et al., 2013; Raftery et al., 2013).

Method complexity has increased over time—but do we really need this complexity every time we forecast mortality in a country of interest? And how can we be sure if a particular method is suitable to forecast mortality and, even more important, if it is not suitable? Looking at validation and mortality forecasting from a new angle, we use validation to assess a method to be (or to be not) suitable to forecast mortality in a country of interest. So far, few research has focused on comprehensive validation of demographic forecast methods in general and mortality forecast methods in particular (e.g. Bohk-Ewald et al., 2018; Shang, 2015, 2012; Shang et al., 2011; Booth et al., 2006). Although assessing forecast performance of introduced methods has recently become more popular (e.g. Li et al., 2013; Bohk-Ewald and Rau, 2017; Basellini and Camarda, 2019; Camarda, 2019). However, validating a method's forecast performance has to rely on a large data basis (covering many mortality levels and patterns over age) to be truly informative. We can only evaluate a method as being robust and broadly applicable if its overall forecast performance across different mortality settings is high.

To reliably evaluate a model's forecast performance we distinguish three analytical settings for two common measures of mortality: life expectancy and life span disparity at birth (e.g. van Raalte et al., 2018; Vaupel and Romo, 2003). Specifically, we analyze by how much a model's forecast performance is influenced by

1. average level of life expectancy and life span disparity in forecast years (as they represent mortality settings with distinct patterns of mortality over age that may (or may not) be difficult to capture by a forecast method).
2. annual rate of change in life expectancy and life span disparity in forecast years (as they represent mortality settings with different levels of mortality improvement that may (or may not) be difficult to capture by a forecast method).
3. trend change in life expectancy and life span disparity between recently observed years and forecast years (as they represent mortality settings with abrupt changes in mortality improvement from recently observed to forecast years that are difficult to capture for each forecast method).

We analyze a model's forecast performance over many countries and time periods to discover stages of mortality improvement that may be challenging to capture. Based on the comprehensive validation, we would recommend to use a method only if its forecast performance turns out to be consistently high across all three analytical settings for life expectancy and life span disparity in most recent years. For a country of interest, we focus on the most recent years as they indicate how mortality will probably evolve and if a selected method is likely to capture that.

The remainder of this paper describes data and methods in section 2 and depicts preliminary validation results for the simple Lee-Carter model in section 3. Finally, we draw main conclusions in section 4.

2 Data & Methods

We use data from the Human Mortality Database (HMD, www.mortality.org). This open access database is maintained by the Max Planck Institute for Demographic Research (Rostock, Germany) and the Department of Demography at the University of California, Berkeley. The HMD provides high-quality population and mortality data for 41 developed countries. For women and men, we extract mortality rates (m) by single ages (x), 0 to 110 and above, from life table data. Based on our validation design, which will be explained later on, we need a minimum of 60 consecutive years of data for a country in our analysis. This reduces the number of eligible countries to 24 (see Table 1 in the Appendix for HMD coverage of mortality data by country and calendar year).

For the beginning, we forecast mortality only with the Lee-Carter model (Lee and Carter, 1992). It is defined as:

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

a_x is the average of log death rates ($m_{x,t}$) over time t, b_x is the response at age x to change in the overall level of mortality over time t, k_t is the overall level of mortality in year t, and $\epsilon_{x,t}$ is an error term. We use singular value decomposition to identify independent age patterns in $m_{x,t}$ and their importance over time in the base period. To forecast mortality ($m_{x,t}$) with the Lee-Carter model, we fix the estimated age-parameters a_x and b_x and extrapolate the time-varying parameter k_t using a time series model. The main assumption of the Lee-Carter method (and extrapolation approaches in general) is that past trends will persist in the future.

We adopt an out-of-sample validation design to assess Lee-Carter's forecast performance. Meaning, we withhold observed mortality data for some calendar years to later compare them with corresponding mortality forecasts. Specifically, for each of the 24 countries, we forecast death rates (m_x) one to 30 years ahead as often as the available mortality data allows. We set the base period (of data used as input for the forecasts) to 30 years. For example, for a country with mortality data available for 70 calendar years ($t_1, \ldots, t_{70}$), we generate a total of 12 forecasts, starting with jump-off years $t_{30}, t_{31}, \ldots, t_{41}$.

We calculate two different mortality measures based on the m_x-forecasts: life expectancy at birth (e_0) and life span disparity at birth ($e_0^\dagger$). Bohk-Ewald et al. (2017) argue that mortality forecast methods not only need to capture the change in life expectancy, but also the variation in the life span distribution, e.g. the compression, shifting, and expansion of mortality. Therefore, we included $e_0^\dagger$ in our analysis, in addition to e_0. Prospectively, we will extend the analysis to mortality measures at further ages, e.g. life expectancy and life span disparity at age 65, to assess the performance of the forecast methods at higher ages.

To quantify accuracy and bias of Lee-Carter forecasts, we calculate the forecast error $e_t = F_t - Y_t$, where F_t is the forecast value and Y_t is the observed value in year t. From this we derive the forecast percentage error $pe_t = 100 * e_t/Y_t$. A percentage error smaller than 0 indicates underestimation of observed mortality, while a percentage error larger than 0 indicates overestimation. Therefore, the closer the percentage error is to 0, the more accurate is a forecast. PE has the advantage of being scale-independent, allowing us to compare a model's forecast performance across different mortality levels (Hyndman and Koehler, 2006). We analyze the overall forecast percentage error across all 24 countries by jump-off year.

We use three analytical settings to quantify the sensitivity of overall forecast accuracy and bias in terms of e_0 and $e_0^\dagger$ to the different mortality levels and trends in the 24 countries. First, we plot the forecast percentage error in relation to the mean value of mortality in the forecast horizon. This gives us insight into whether there is at all a relationship between the overall level of mortality and forecast accuracy and bias. Our goal is to see whether a forecast method can capture different levels of mortality and their associated age patterns. Second, we assess whether the mortality trend over time plays a role in the forecast performance. Meaning, how well can a forecast method capture moderate to strong mortality changes (which can be positive or negative) in the forecast years? To do so, we calculate the annual rate of change of mortality in the forecast horizon and plot it in relation to the percentage error. However, this does not portray any changes in the development of mortality before and after the jump-off year that might have occurred. Therefore, in a third step, we plot the percentage error in relationship to the trend change. We define the trend change as the difference between the annual rate of change in the forecast horizon and the annual rate of change in the base period. We perform all analyses separately for females and males.

3 Preliminary Results

Figures 1 and 2 show the three analytical settings for Lee-Carter forecasts 20 years ahead according to our validation design for females and males, respectively. The color gradient from black over purple and orange to yellow depicts the range of jump-off years from 1780 to 1997. The time span of available data differs by country (see Table 1 in the Appendix). Therefore, only a few countries (Sweden, Denmark, Norway, Netherlands) represent the earliest jump-off years while a larger number of countries represents more recent jump-off years. In all 6 individual panels of each Figure, we display the forecast percentage error (PE) on the vertical axis. A PE larger than 0 indicates overestimation, while a value smaller than 0 shows underestimation. In each Figure, we show the PE for life expectancy at birth (e_0) in the left column and for life span disparity ($e_0^\dagger$) in the right column.

PE versus average level of e_0 and $e_0^\dagger$ in forecast years In the top row of Figures 1 and 2, we plot the PE in relation to the mean value of mortality in the forecast horizon of 20 years (horizontal axis) for life expectancy at birth and life span disparity. Overall, we find that the mean level of e_0 appears to have a smaller effect on the PE of Lee-Carter forecasts than the mean level of $e_0^\dagger$.

Specifically, the overall small PE of e_0 shifts horizontally to the right from earlier to recent jump-off years (purple to yellow) and from low to high average levels of e_0 (35 to 85 years). At the same time, we find that the variance of the PE for e_0 decreases for later jump-off years: while the Lee-Carter method mostly underestimates life expectancy at birth slightly for mid-level jump-off years (red and orange) it performs well in terms of forecast accuracy and bias for the most recent jump-off years (yellow).

Regarding the PE for $e_0^\dagger$, we find a backward shift, in the form of an arch, from high levels (25 years) to low levels (10 years) of life years lost from earlier to recent jump-off years (black to yellow). We find that the Lee-Carter method can not properly capture the often strong decrease in life span disparity in the mid-level jump-off years (red and orange) and, consequently, tends to overestimate $e_0^\dagger$. Here, bias and inaccuracy are even more accentuated than for e_0 (PE up to +80% as opposed to +20%). However, for the earliest and most recent jump-off years (purple and yellow), the Lee-Carter method produces accurate and unbiased results for $e_0^\dagger$.

PE versus annual rate of change of e_0 and $e_0^\dagger$ in forecast years The middle row of Figures 1 and 2 shows the relationship of the PE with the annual rate of change of e_0 and $e_0^\dagger$ in the forecast horizon (fh) 20 years ahead. If the rate is smaller than 0, e_0 resp. $e_0^\dagger$ have decreased in the forecast horizon. For values larger than 0, e_0 and $e_0^\dagger$ have increased. The larger the absolute value of the annual rate of change, the stronger is the change in e_0 and $e_0^\dagger$ (positive or negative) in the forecast years. Overall, we observe a strong effect of the annual rate of change of e_0 and $e_0^\dagger$ on accuracy and bias of the Lee-Carter forecasts.

For early and mid-level jump-off years (purple to orange), the Lee-Carter method can not capture the strong increase in life expectancy (+0.5 to +1 years of life) in the forecast horizon. This results in forecast inaccuracy and mostly underestimation of the true annual rate of change of e_0. However, the Lee-Carter method performs well for the most recent jump-off years (yellow) where there is a more moderate increase in life expectancy (+0 to +0.3 years of life) in many countries.

We see a similar relationship between the annual rate of change and PE for life span disparity at birth. The larger the decrease in $e_0^\dagger$ is (up to −0.5 years of life lost), the heavier is the overestimation of the Lee-Carter forecasts (PE up to 80%). Moving towards an annual rate of change of 0 in the earliest and most recent jump-off years (purple and yellow), the Lee-Carter forecasts become more accurate.

PE versus trend change in e_0 and $e_0^\dagger$ between recently observed and forecast years In the bottom row of Figures 1 and 2, we show the relationship of the PE with the trend change in mortality from the base period (bp) to the forecast horizon (fh). If the annual rate of change in the forecast horizon is smaller than the rate in the base period, the trend change is smaller than 0. If the annual rate of change in the forecast horizon exceeds the rate in the base period, the trend change is larger than 0. Trend changes in mortality are especially hard to capture in forecasts as they are always unexpected. Overall, we find that trend changes appear to have a strong impact on the PE of e_0 and $e_0^\dagger$.

As we see from the plots, trend changes have happened for both life expectancy and life span disparity, resulting in higher forecast inaccuracy especially in the early and middle jump-off years (purple to orange). However, a trend change from smaller increases in life expectancy in the base period to larger increases in life expectancy in the forecast horizon (positive trend change) effects the PE of e_0 more strongly than a negative trend change. Further, positive trend changes in e_0 are related to underestimation (PE up to −70%) while negative trend changes rather cause overestimation (PE up to +40%). For the most recent jump-off years (yellow), the Lee-Carter method gives more accurate results as the trend change is closer to 0.

Trend changes in life span disparity take place particularly for mid-level jump-off years and range from -0.4 to 0.4 (red to orange). Here, the negative relationship between PE of $e_0^\dagger$ and trend change is even more accentuated than for the PE of e_0. The Lee-Carter method penalizes strong negative trend

Figure 1: Performance of Lee-Carter model when forecasting female mortality 20 years ahead

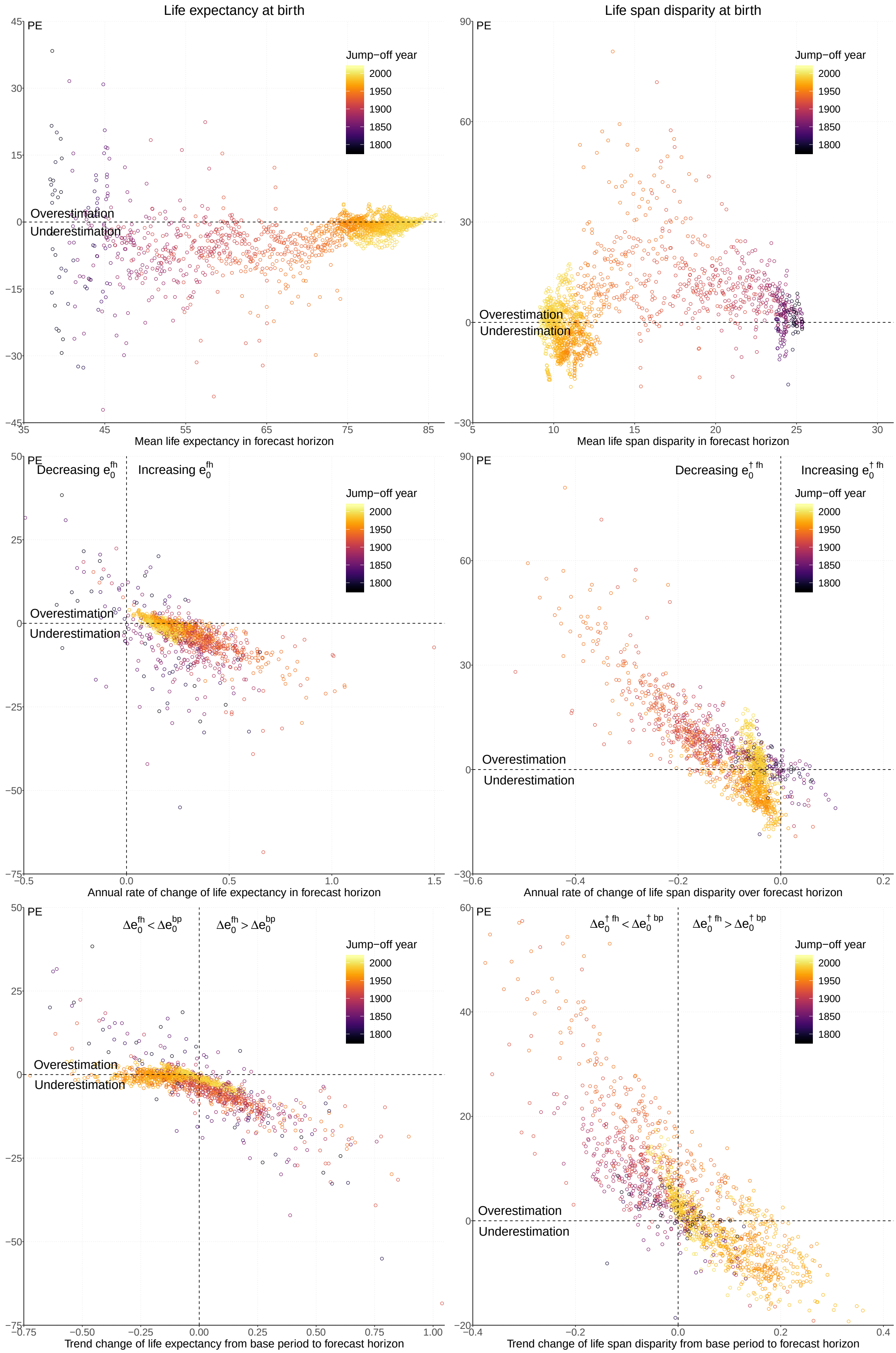
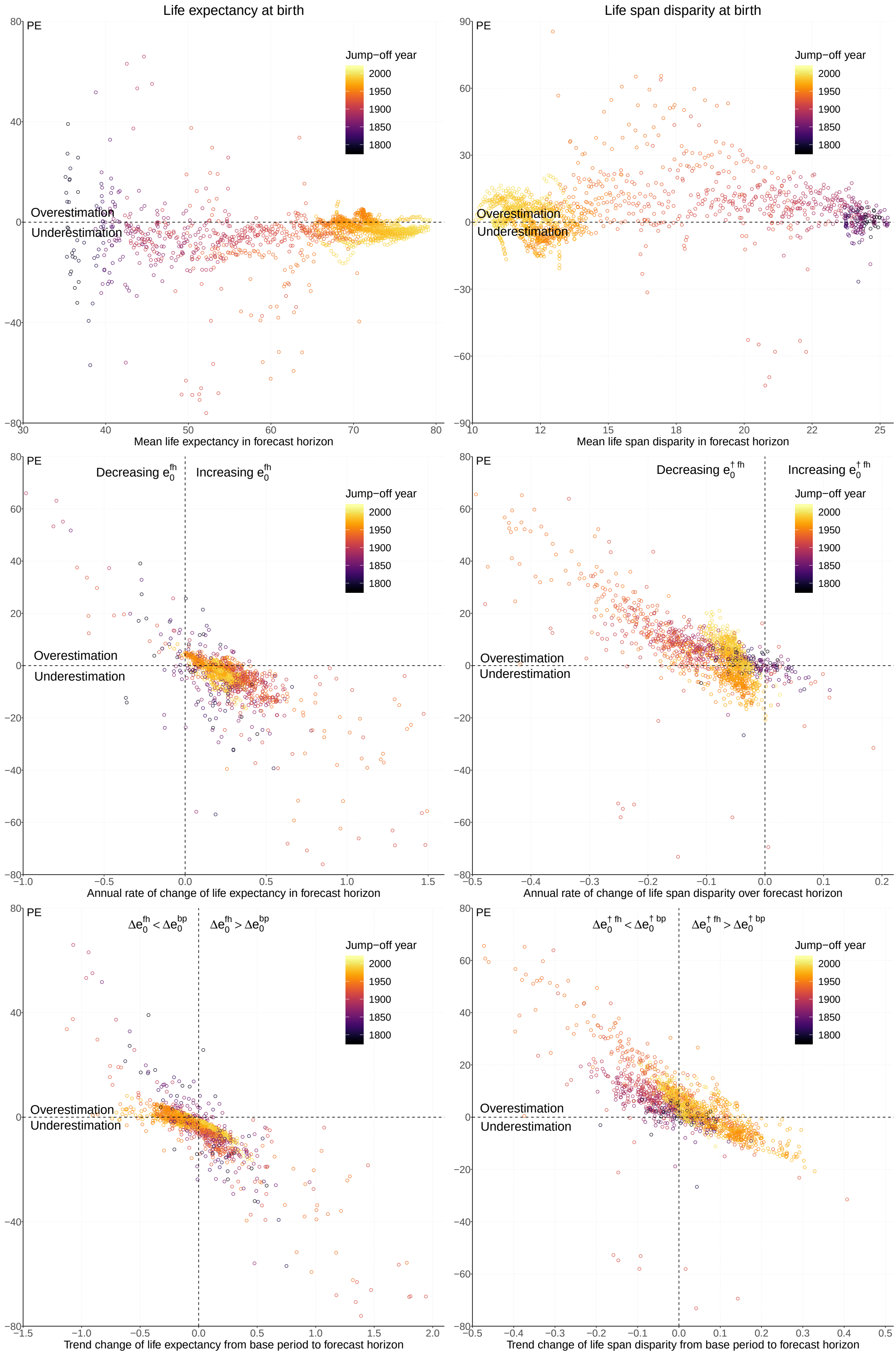


Figure 2: Performance of Lee-Carter model when forecasting male mortality 20 years ahead



changes with severe overestimation (PE up to +60%) and positive trend changes with underestimation (PE up to −20%). As the trend change for the most recent jump-off years is close to 0 (yellow), the Lee-Carter model performs well in terms of forecast accuracy. However, we observe slight overestimation of $e_0^\dagger$.

Similar patterns for women and men With respect to the male population (see Figure 2), we observe similar patterns in the relationship of the PE with the mean level, the annual rate of change, and the trend change of e_0 and $e_0^\dagger$. However, the overall level of mortality is higher for males than for females. Further, the variance in the PE for both e_0 and $e_0^\dagger$ is larger for males compared to females.

Overall, our analyses show that the Lee-Carter method is suitable for forecasting mortality of the 24 HMD countries in the most recent years. For these cases, judging from the forecast percentage error, the forecast results are mostly accurate and unbiased. The Lee-Carter method can not capture drastic changes in the annual rate of change and strong trend changes, resulting in bias and loss of accuracy. These characteristics apply to the mid-level jump-off years, were a transition from low-level to high-level mortality has taken place.

4 Conclusion & Outlook

How can we objectively decide if a simple method is suitable to forecast mortality in a country of interest? In recent decades, researchers have developed numerous mortality forecast approaches with increasing levels of methodological complexity. Still, the Lee-Carter model (1992) is handled as a golden standard due to robust forecast performance, despite being methodologically simple. Considering that so many simple and also complex methods are available our study has two main goals. First, to establish validation as test prior to mortality forecasting to confirm that a selected method is suitable. Second, to show that the simple Lee-Carter model is broadly applicable, particularly in mortality contexts without substantial trend changes and decisive shifts in the age at death distribution.

In this paper we have validated if the basic assumption of the Lee-Carter method holds in a country of interest. Namely, that mortality changes in the forecast horizon will develop in the same way they have had in the base period. Therefore, we have applied the Lee-Carter method to all available mortality data for 24 countries of the Human Mortality Database according to our own validation design. We have assessed forecast accuracy and bias of life expectancy and life span disparity at birth over all 24 countries by jump-off year in three analytical settings. First, we have analyzed the forecast percentage error in relation to the mean level of mortality in the forecast horizon. Second, we have examined how the percentage error reacts to mortality changes in the forecast horizon. Third, we have shown the relationship between the percentage error and mortality trend changes from base period to forecast horizon.

Based on our extensive validation results, we have found, most strikingly, that the Lee-Carter method is indeed suitable to forecast mortality for many highly-developed countries in most recent years. More complex methods, however, would have been more suitable than Lee-Carter's model for jump-off years 1960 through 2000, when discontinuous mortality developments have been prevalent in many of those countries. Caused by, for example, mortality developments such as the advancement of large survival improvements from younger to increasingly older ages with ongoing time. Finally, we have shown that validation serves as a meaningful first test to decide whether a method is likely to be appropriate to forecast mortality in a country of interest.

Prospectively, we will deepen and extend our analyses in several ways. First, we will apply the validation design to additional forecasting methods that are more complex. Second, we will use UN data to increase the number of countries and to include, for example distinctive mortality levels and patterns over age of less developed populations. Third, we will assess how sensitive a method's forecast performance is to the lengths of base period and forecast horizon. We will deepen the evaluation of accuracy and bias using further error measures and we will also look at the coverage of prediction intervals. Finally, we will explore forecast performance patterns by age and region. Doing so, we wish to define rules as guidance for deciding whether a method is likely to be suitable to forecast mortality for a country of interest.

References

Bardoutsos, A., J. de Beer, and F. Janssen (2018, Oct). Projecting delay and compression of mortality. *Genus 74*(1), 17.

Basellini, U. and C. G. Camarda (2019). Modelling and forecasting adult age-at-death distributions. *Population studies 73*(1), 119–138.

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

Bohk-Ewald, C., P. Li, and M. Myrskylä (2018). Forecast accuracy hardly improves with method complexity when completing cohort fertility. *Proceedings of the National Academy of Sciences 115*(37), 9187–9192.

Bohk-Ewald, C. and R. Rau (2017). Probabilistic mortality forecasting with varying age-specific survival improvements. *Genus 73*(1), 1.

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

Booth, H., R. J. Hyndman, L. Tickle, and P. De Jong (2006). Lee-carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research 15*, 289–310.

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

Camarda, C. G. (2019). Smooth constrained mortality forecasting. *Demographic Research 41*, 1091–1130.

de Beer, J., A. Bardoutsos, and F. Janssen (2017). Maximum human lifespan may increase to 125 years. *Nature 546*(7660), E16.

Haberman, S. and A. Renshaw (2012). Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and economics 50*(3), 309–333.

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

Hyndman, R. J. and A. B. Koehler (2006). Another look at measures of forecast accuracy. *International journal of forecasting 22*(4), 679–688.

Janssen, F., L. J. van Wissen, and A. E. Kunst (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography 50*(4), 1341–1362.

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

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

Li, N., R. Lee, and P. Gerland (2013). Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography 50*(6), 2037–2051.

Raftery, A. E., J. L. Chunn, P. Gerland, and H. Ševčíková (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography 50*(3), 777–801.

Rau, R., E. Soroko, D. Jasilionis, and J. W. Vaupel (2008). Continued reductions in mortality at advanced ages. *Population and Development Review 34*(4), 747–768.

Ševčíková, H., N. Li, V. Kantorová, P. Gerland, and A. E. Raftery (2016). Age-specific mortality and fertility rates for probabilistic population projections. In *Dynamic demographic analysis*, pp. 285–310. Springer.

Shang, H. L. (2012). Point and interval forecasts of age-specific life expectancies: A model averaging approach. *Demographic Research 27*(21), 593–644.

Shang, H. L. (2015). Statistically tested comparisons of the accuracy of forecasting methods for age-specific and sex-specific mortality and life expectancy. *Population studies 69*(3), 317–335.

Shang, H. L., H. Booth, and R. Hyndman (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research 25*(5), 173–214.

van Raalte, A. A., I. Sasson, and P. Martikainen (2018). The case for monitoring life-span inequality. *Science 362*(6418), 1002–1004.

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

Vogt, T., A. van Raalte, P. Grigoriev, and M. Myrskylä (2017). The German East-West mortality difference: two crossovers driven by smoking. *Demography 54*(3), 1051–1071.

Wang, H. and S. H. Preston (2009). Forecasting united states mortality using cohort smoking histories. *Proceedings of the National Academy of Sciences 106*(2), 393–398.

Appendix

Table 1: HMD coverage of mortality data by country (group) and calendar years

Country group	Country	Code	Available years
Northern America	Canada	CAN	1921 - 2016
	USA	USA	1933 - 2017
Australia & New Zealand	Australia	AUS	1921 - 2016
	New Zealand total population	NZL_NP	1948 - 2013
Japan	Japan	JPN	1947 - 2017
Eastern Europe	Bulgaria	BGR	1947 - 2010
	Czechia	CZE	1950 - 2017
	Hungary	HUN	1950 - 2017
	Slovakia	SVK	1950 - 2017
Northern Europe	Denmark	DNK	1835 - 2016
	Finland	FIN	1878 - 2015
	Ireland	IRL	1950 - 2014
	Norway	NOR	1846 - 2014
	Sweden	SWE	1751 - 2017
	United Kingdom total population	GBR_NP	1922 - 2016
Southern Europe	Italy	ITA	1872 - 2014
	Portugal	PRT	1940 - 2015
	Spain	ESP	1908 - 2016
Western Europe	Austria	AUT	1947 - 2017
	France total population	FRATNP	1816 - 2017
	Germany East	DEUTE	1956 - 2017
	Germany West	DEUTW	1956 - 2017
	Netherlands	NLD	1850 - 2016
	Switzerland	CHE	1876 - 2016