Complementing the evaluation toolkit of mortality forecasts with measures of lifespan disparity

Christina Bohk[1] Marcus Ebeling[1,2] Roland Rau[1,2]

Last modification: `December 14, 2015`

Abstract

Evaluating the predictive ability of mortality forecasts is important and yet, at the same time, difficult. Average lifespan and death rates are basic life table functions that are typically used to analyze how much forecasts deviate from their realized values. While these parameters are useful to specify how precisely mortality has been forecasted at a certain point in time, they cannot be used to indicate whether the underlying mortality developments are plausible, too. We therefore propose to look in addition at lifespan disparity to examine whether the forecasted variability of the age at death is a plausible continuation of past trends. Validating mortality forecasts for Italy, Japan, and Denmark demonstrate that their predictive performance can be evaluated more comprehensively when analyzing average lifespan and lifespan disparity at the same time, i.e. jointly analyzing mean and dispersion of mortality. Approaches accounting for dynamic age-shifts in survival improvements outperform others that enforce invariant patterns.

[1] University of Rostock, Rostock, Germany
[2] Max Planck Institute for Demographic Research, Rostock, Germany

1 Introduction

Mortality forecasts typically aim to predict how many additional years of life people are likely to gain. Basic life table functions such as life expectancy at birth, a measure of central tendency, and age-specific death rates, a measure of mortality intensity, are usually applied to evaluate how precise such forecasts are. Evaluations focus on *ex post* quantitative aspects; they quantify, for instance, how many years forecasted average lifespans deviate from their realized values (Keilman, 1997; Cairns et al., 2011; Smith et al., 2001; Booth et al., 2006; Koissi et al., 2006; Shang et al., 2011; Dowd et al., 2010). Although basic life table functions are useful to specify how precisely mortality has been forecasted, evaluations could be incomplete, even wrong, if they rely on them only. Small forecast errors of average lifespan do not necessarily indicate that the forecasted underlying mortality developments are plausible, too. Figure 1 illustrates this issue in more detail with a scatterplot that displays the negative correlation between life expectancy at birth on the horizontal axis, and lifespan disparity, measured by average life years lost at birth, $e(0)^{\dagger}$ (e.g. Vaupel and Romo, 2003), on the vertical axis for women in Italy (green), Denmark (blue), and Japan (red) from 1950 to 2012.

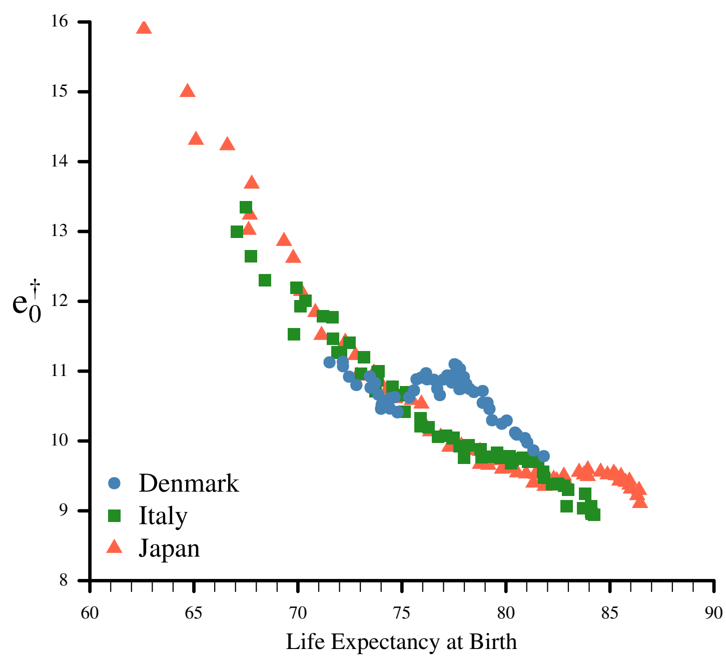

Figure 1. Scatterplot of life expectancy at birth (horizontal axis) and average life years lost at birth due to death (vertical axis) for women in Denmark, Italy, and Japan from 1950 to 2012.

In contrast to basic life table measures, $e(0)^\dagger$ informs about underlying mortality developments. While life expectancy at birth has increased due to reductions of mortality at progressively higher ages in recent decades, $e(0)^\dagger$ has decreased due to predominant survival improvements at premature ages, shifting deaths towards the end of the lifespan. This effect is also known as the compression of mortality (Fries, 1980). In Figure 1, it is striking that the average lifespan of Italian, Danish, and Japanese women has been similar in recent decades, whereas the decline in their variability of the age at death differed considerably as soon as their average lifespan exceeded 75 years: The lifespan dispersion (1) declined regularly for Italian women, (2) leveled off for Japanese women, and (3) increased and decreased for Danish women. These findings illustrate that similar average lifespans can originate from different underlying mortality trajectories. That is why we propose to expand the toolkit of conventional evaluation procedures. Basic life table functions should be complemented by measures of lifespan dispersion to improve the assessment of *ex post* quantitative aspects, and to evaluate the plausibility of underlying mortality trends.

2 Data and Methods

To measure lifespan dispersion, we take the average number of life years lost at birth (Vaupel and Romo, 2003; Zhang and Vaupel, 2009), $e(0)^\dagger$, estimated by

$$e(0)^\dagger = \frac{\int_0^\omega e(a)d(a)\,da}{l(0)}, \tag{1}$$

with $e(a)$ being remaining life expectancy at age a and $d(a)$ being life table deaths at age a, both integrated from age 0 to ω, which is the highest age at death. $l(0)$ is the radix of the life table. There are also several other measures, such as the standard deviation or the inter-quartile-range, that can be used instead of $e(0)^\dagger$ to analyze lifespan disparity. All of these measures are highly correlated (Wilmoth and Horiuchi, 1999), but their influence on the results is marginal. A major reason why we choose $e(0)^\dagger$ is that it is demographically interpretable, and that it indicates the potential for further progress in survival.

To test the usefulness of $e(0)^\dagger$ as an evaluation measure we forecast mortality for women in Italy, Japan, and Denmark who substantially differ in their level of lifespan disparity in recent decades (see Figure 1). We use the original model of Lee and Carter (1992), its rotating variant of Li et al. (2013), as well as the model of Bohk and Rau (2014) to predict mortality from 1991 to 2009, using data from 1965 to 1990 as the reference period. Since the three approaches differ primarily in their ability to capture dynamic age shifts in the distribution of deaths, they are particularly suitable to evaluate how well they forecast (actual) developments in lifespan disparity. As input data in the forecasts, we use deaths and exposures by single age for women in Italy, Japan, and Denmark from the Human Mortality Database (2015). To enable the forecasting approaches to shift deaths to ages beyond 110+, we extend the age range of mortality data to 130+ with the model of Kannisto (Thatcher et al., 1998); this is similar to the revised UN forecasting approach of Ševčíková et al. (2015), which also implemented the rotating Lee-Carter model to account for dynamic age shifts of survival improvement.

3 Preliminary Results

Figure 2 displays average lifespan, $e(0)$ (left panels), and average life years lost, $e(0)^{\dagger}$ (right panels), for women in Italy (top), Japan (center), and Denmark (bottom). Observed data are depicted in black, forecasted data are illustrated in red (Lee-Carter model), green (rotating variant), and blue (Bohk-Rau model). Moreover, the forecast years, from 1991 to 2009, are highlighted in gray, and the reference period, from 1965 to 1990, is highlighted in beige.

Case of Italy (stable trends for $e(0)$ and $e(0)^{\dagger}$):

If mortality develops regularly without any trend changes in the forecast years, the predictions of all three approaches are close to the realized values. In Italy, the forecasts of all approaches capture the regular increase in the average lifespan as well as the regular decline in the lifespan disparity with only small deviations.

Case of Japan (stable trend for $e(0)$ and unstable trend for $e(0)^{\dagger}$):

If the trend of the average lifespan is stable, but the trend of the lifespan disparity is not, differences in the predictive ability of the three approaches are present, but they become visible only if we complement standard approaches with a measure of dispersion. In Japan, the forecasts of the three models are close to the realized average lifespan, suggesting that the forecasts were precise. Only the analysis of the lifespan disparity shows that all approaches overestimate the observed decline in the variability of the age at death; the deviations are greater for the Lee-Carter models than for the Bohk-Rau model. As a consequence, all three approaches predict the concentration of deaths at higher ages to be much greater than it actually is.

Case of Denmark (unstable trends for $e(0)$ and $e(0)^{\dagger}$):

If the trends of the average lifespan and its disparity are unstable, both evaluation measures indi-

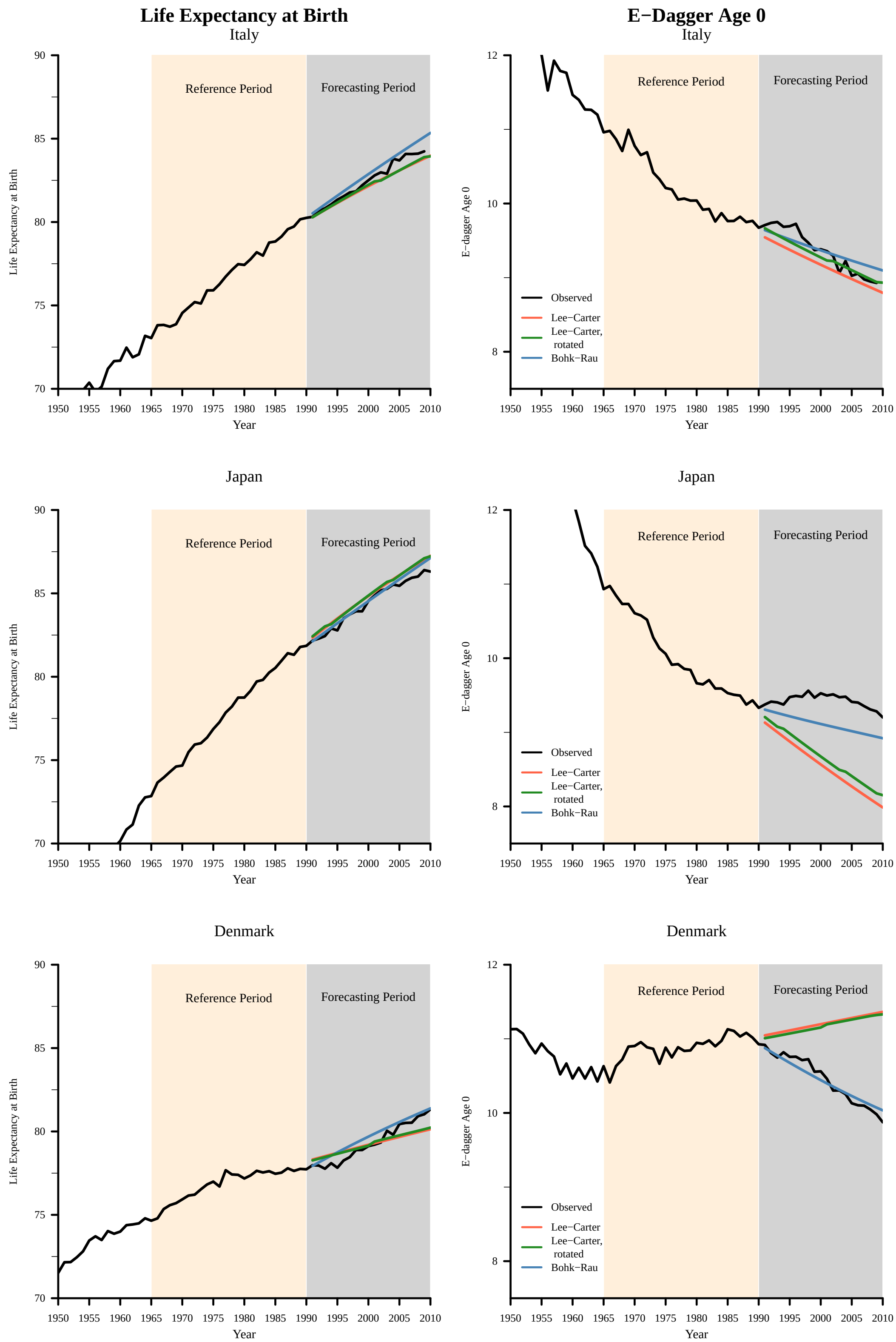

Figure 2. Life expectancy at birth (left panels) and life years lost at birth (right panels) for women in Italy (top), Japan (center), and Denmark (bottom); observed data are depicted in black, forecasted data are illustrated in red (Lee-Carter model), green (rotating variant), and blue (Bohk-Rau model).

cate forecast errors. In Denmark, the forecasts of the three models capture the increasing trend of the average lifespan (after a period of stagnation in the 1980s and early 1990s) quite well, although they deviate marginally more from the realized values than in the Italian and Japanese forecasts. However, the situation is different for lifespan disparity: The Lee-Carter models predict that it increases in the forecast years, although it actually declines. This outcome does not only deviate substantially from the realized values, it also appears to be rather implausible, given the general negative correlation between (rising) life expectancy at birth and (declining) lifespan disparity (see Figure 1). In contrast, the Bohk-Rau model captures the changing trend of the lifespan disparity in the forecast years quite well, resulting in a more plausible forecast with only small deviations from the realized values. The model of Bohk and Rau appears to have a greater adaptability than the original Lee-Carter model; this is probably due to the modeling of time-variant survival improvements for each age rather than time-invariant changes for the death rates (as assumed by the original Lee-Carter model). The rotating Lee-Carter model tries to overcome this inflexibility with dynamic age shifts for the survival improvements, too, but this methodological refinement cannot come into action as long as the lifespan is below 80 years. That is also the reason why the validating forecasts of the Lee-Carter models do not differ much. This will change in prospective mortality forecasts that belong to future work of this project.

Our analyses illustrate that the joint evaluation of the lifespan, $e(0)$, and life years lost, $e(0)^{\dagger}$, provide new insights that we deem necessary to evaluate the predictive performance of mortality forecasts comprehensively. We also suggest to use these new insights when improving or developing new methods to forecast mortality.

Acknowledgments

The European Research Council has provided financial support under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 263744.

References

Bohk, C. and R. Rau (2014). Probabilistic Mortality Forecasting with Varying Age-Specific Survival Improvements. arXiv:1311.5380v2 [stat.AP].

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*(1–2), 289–310.

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

Dowd, K., A. J. G. Cairns, D. Blake, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah (2010). Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics 47*, 255–265.

Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine 303*(3), 130–135.

Keilman, N. (1997). Ex-post Errors in Official Population Forecasts in Industrialized Countries. *Journal of Official Statistics 13*(3), 245–277.

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

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

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.

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*, 173–214.

Smith, S. K., J. Tayman, and D. A. Swanson (2001). *State and Local Population Projections: Methodology and Analysis*. The Springer Series on Demographic Methods and Population Analysis. Springer Netherlands.

Thatcher, A. R., V. Kannisto, and J. W. Vaupel (1998). *The Force of Mortality at Ages 80 to 120*. Monographs on Population Aging, 5. Odense University Press.

University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock, (Germany) (2015). Human Mortality Database. Available at `www.mortality.org`.

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

Ševčíková, H., N. Li, V. Kantorová, P. Gerland, and A. E. Raftery (2015). Age-Specific Mortality and Fertility Rates for Probabilistic Population Projections. arXiv:1503.05215v1 [stat.AP].

Wilmoth, J. and S. Horiuchi (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography 36*, 475–495.

Zhang, Z. and J. W. Vaupel (2009). The age separating early deaths from late deaths. *Demographic Research S8*(29), 721–730.