A New Dimension to Mortality Forecasts?

Benjamin J. Seligman[1,2], Shripad Tuljapurkar[1,3]

1. Stanford University, Department of Biology
2. Stanford University School of Medcine
3. Morrison Institute for Population Research

Introduction

The Lee-Carter method has become the standard approach for forecasting mortality and life expectancy of human populations, particularly among developed countries. In addition to representing a significant advance in forecasting methods when it was developed, its findings with respect to historical mortality trends are of significant substantive interest. Some twenty years after its initial development and widespread adoption among actuarial agencies, this manuscript revisits the model's predictive capability and the trends in mortality decline in the developed world since the end of the Cold War.

The Lee-Carter Model

In the Lee-Carter Model[1,2], the log of mortality at age x in year t is represented in this form:

$$\log(\boldsymbol{m_{x,t}}) = \boldsymbol{a_x} + \boldsymbol{b_x} * \boldsymbol{k_t} \ (1)$$

where $\mathbf{a_{1,x}}$ is a vector of average, across a given base set of years, log mortality rates by age, $\mathbf{b_{1,x}}$ is a vector that can be seen as representing the "slope" of log mortality rates by age over time, and $\mathbf{k_{t,1}}$ is a vector with each element representing the "level of mortality" in a given year and is explained further below.

The key methodological innovation behind this model was the application of singular value decomposition to a centered matrix of log age-specific mortality rates by year. This decomposition allows a mxn matrix **B** to be represented in the following fashion:

$$\boldsymbol{B} = \boldsymbol{UDV}^{\mathrm{T}} \ (2)$$

Here **U** and **V** are orthogonal matrices of dimensions mxm and nxn respectively and **D** is a mxn matrix of ordered singular values, decreasing in value, along the diagonal and 0 elsewhere. **B** can be reconstructed by, for each column i in 1:min(m,n), taking $\mathbf{U}_i(\mathbf{V}_i)^{\mathrm{T}}$, then taking the sum of the resulting mxn matrices weighted by their corresponding singular value.

The square of each singular value represents how much variance they explain across the columns of the original matrix **B**. For matrices of age-specific mortality rates, the lead singular value typically explains 80-90% of the variation in the data and thus represents the dominant time trend in mortality. Multiplying this by the first row of $\mathbf{V}^{\mathrm{T}}$ produces the **k**'s, which may be seen as an abstract, relative level

of mortality in each year, and by convention sum to 0. The first column of **U** becomes the vector **b**, which gives the age-specific response to the time trend, and by convention its elements sum to 1.

During the 20th century, the k's have been roughly linear over time in most developed countries[3]. As **a** and **b** are constant over time, a linear forecast of the **k**'s can be used to , in turn, forecast mortality rates and life expectancy and create meaningful prediction intervals for this forecast. This is also of great substantive interest as the forces that would produce a linear relationship over time – as opposed to a discontinuous form as new innovations emerge or a flattening curve as improvements in mortality diminish – remain unclear.

Extensions to this model have been made since its initial publication, allowing for "coherent" forecasts of subpopulations[4,5] and attempting to allow for changing age-specific responses to the main time trend[6,7]. However, the fundamental underpinnings of the model have remained 1) that there is a dominant single time trend that explains the overwhelming majority of the variation in mortality rates over time and 2) that this dominant time trend is linear. We investigate if these assumptions still hold and the implications for forecasts.

Trends in Mortality

Over the past century, there has been dramatic change in mortality rates and life expectancy in developed countries. In the United States, for example, estimated life expectancy rose from 47.3 years in 1900 to 76.8 in 2000[8]. The causes of death, likewise, have also changed from being largely infectious in origin to non-communicable, usually chronic diseases[9], in line with the ideas of the epidemiologic transition[10,11]. This theory offers a framework for how causes of death change in populations that are likewise undergoing a demographic transition: initially, life expectancy is in the 20s-30s, with deaths primarily from hunger, epidemic disease, and perinatal causes. As societies develop economically, mortality rates from these causes gradually decline and life expectancies rise into the 40s or 50s. Finally, as infections decline further, chronic diseases such as cancer and cardiovascular disease dominate the causes of death, although mortality rates continue to decline. At this point life expectancy is often in the 60s or 70s. Finally, a proposed fourth stage of so-called "delayed degenerative diseases" arises, as mortality from chronic diseases is pushed back to progressively older ages[11].

Under this model, increases in life expectancy lead to the emergence of diseases that are progressively harder to treat and whose mitigation leads to less gain in life expectancy, owing to their onset later in life. This would predict that improvements in mortality will gradually slow and the previously-described linear trend in a single dominant time signal may give way to non-linear trends or multiple relevant time signals. Thus, we also explore the Lee-Carter model's forecasts with actual trends in mortality since its inception.

Methods

Life tables and population size by age and sex for the G7 countries (Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States) were taken from the Human Mortality Database (HMD)[12]. Analysis of Germany was limited to the regions that comprised the former West Germany,

while national populations for the UK and France were used out of the available options. For population data during years in which a territorial change occurred, the populations listed following the territorial change were used.

Forecasts were made from 1990 to 2010 for each sex separately and combined, with two sets of base periods ending in 1989, beginning in 1950 (1956 for West Germany) or in 1960. The parameters of the Lee-Carter model were calculated using age-specific mortality rates across the base years. The values of *k* were calculated using both first and second-stage methods. First-stage *k*'s are calculated directly from the singular value decomposition as described in the introduction. Second-stage *k*'s are calculated as described elsewhere[2], with the aim of fitting them to observed total number of deaths *D* occurring in year *t* per the formula:

$$D_t = \sum_{x=0}^{n} N_{x,t} * \exp(a_x + b_x * k_t) \quad (3)$$

for all ages *x*, where *N* is the number of individuals of age *x* in year *t*, where **a** and **b** are the vectors described previously.

If the second-stage calculation is not undertaken, higher-order singular values can be used to investigate trends orthogonal to the lead singular value, producing $\mathbf{b}_j$ and $\mathbf{k}_j$ for the jth singular value.

To construct survival probabilities and life expectancies, mortality in the first year of life was estimated using formulae for males, females, and both sexes combined as appropriate[13,14]. Observed life expectancy at birth, e_0, was generated in two fashions: one set was taken directly from HMD life tables, the other set was generated by summing the probability of mortality at each age *x*, q_x, across all ages *x*. The q_x values were taken from the HMD life tables or, for the first year of life, calculated using formulae 5-7 described below. These two approaches yield identical life expectancies.

To calculate the probability of death from birth to age one, ${}_1q_0$, given the infant mortality rate m_0, we use the formulae previously described by Preston[13,14] for males, females, and combined sexes as appropriate. To get this probability of death, the formula is:

$$q_0 = \frac{{}_1m_0}{1 + {}_1m_0(1 - {}_1a_0)} \quad (4)$$

Here ${}_1a_0$ (with ${}_1m_0 < 0.107$), the average number of years lived by those who died before age 1, is estimated by:

$$combined\ sexes:\ {}_1a_0 = 0.07 + 1.7 * {}_1m_0 \quad (5)$$

$$males:\ {}_1a_0 = 0.045 + 2.684 * {}_1m_0 \quad (6)$$

$$females:\ {}_1a_0 = 0.053 + 2.800 * {}_1m_0 \quad (7)$$

To investigate male-female differences, two approaches were taken. First was to construct coherent forecasts using the Lee-Li method[4]. This can be implemented in a simple manner by applying the standard Lee-Carter methods to a matrix containing both male and female age-specific mortality rates over time. This allows for coherence by having male and female mortality rates follow the same time-trend, with both age- and sex-specific responses to this trend. Second was to look at the effects of sex composition by age on life expectancy. This analysis was based on changes in variance of age of death over time[15,16], which has been increasing over time, suggesting mortality decline is driven by declining rates at older ages, which are predominantly female.

Analysis was conducted using R version 2.15.2 in the RStudio IDE version 0.97 with forecasting run using the *forecast* package[17–19].

Results

Mortality Trends

Plots of trends in life expectancy at birth for both sexes combined are shown in figure 1. Overall, the observed trends in life expectancy remain linear over the period studied. Forecasts using different base years do not differ substantially, although those using first and second stage *k*'s differ somewhat in the magnitude of difference between forecast and observed life expectancy, though not in the direction of the difference. Notably, for the four European countries and Canada, mean forecast life expectancy is substantially lower than what is observed, with the difference increasing over time – this is particularly striking for Italy, Germany, and the UK. The opposite is true of Japan, where forecast life expectancy for both sexes is greater than observed, and holds roughly constant over time. In the United States, forecasts are approximately congruent with observations.

Sex-specific results for each country are presented in figures 2 and 3 for males and females respectively. Among the results for females, most observed life expectancies were consistent with or somewhat above forecasts. The notable exceptions were the US and Canada, where observed life expectancy for women remained below forecasts throughout the period considered. For men in Canada, France, Germany, Italy, the UK, and the US observed life expectancy rose faster than was forecast, crossing the 95% forecast bound made with first-stage *k*'s in all cases except France. In Japan, by contrast, the trend was consistently below the mean forecast Coherent forecasts made using the Lee-Li method did not improve forecasts of male mortality.

"Observed" values of *k* for the forecast period were created by projecting actual log mortality rates onto *b;* chi-squared *p*-values for the observed *k*'s compared to the forecasts are shown in table 1, and the life expectancies calculated using these *k*'s are plotted in figures 1-3. Male life expectancies calculated using these observed *k*'s and the previously determined **a** and **b** were lower than observed life expectancies. Lee-Li coherent forecasts showed no improvement in forecasting male life expectancy, which continued to be underestimated.

As a result of the failure of observed *k*'s to adequately capture male mortality, we considered the possibility that the time trends in mortality associated with other singular values may be important. Using same method, we calculated a **b2** associated with the second singular value and **b3** associated with the third and projected male mortality rates onto them to produce observed *k2*'s and *k3*'s. These are plotted with the observed *k*'s, which show a small but growing contribution from *k2* for some countries, but with *k3* values staying near zero for all countries (figure 4). However, calculating male life expectancy using *k* and *k2* did not completely recover the difference with observed life expectancy.

Male-Female Compositional Effects

To resolve the apparent paradox of failed forecasts of male mortality with successful forecasts for combined sexes, we attempted to determine the effects age-specific sex composition. While life expectancy increases, variance in age of death allows for assessment of which ages are contributing most to these increases. Mortality declines below a particular threshold age tend to decrease variance, whereas declines above the threshold will tend to increase variance. Current work has shown that the threshold age has typically been around 65 years in developed countries and that, since the 1980s, variance has been increasing in these countries. These suggest that mortality improvements over the past thirty years – including the years considered in the above forecasts – have been occurring at ages above 65 years. These ages are composed of a greater proportion of women than men, and this compositional effect suggests an explanation for why forecasts for combined sexes, like those for women, were accurate even as those for men were not. For example, figure 5 shows observed life expectancy from 1989 to the most recent year with available data for the G7 countries, along with life expectancy calculated using the average of male and female age-specific mortality rates (effectively assuming an equal sex ratio at all ages). The life expectancies calculated using the averaged mortality rates are consistently lower than the observed life expectancies.

To further test the importance of sex composition, male and female age-specific mortality rates were averaged and used to produce forecasts of life expectancy. The mean forecasts are shown in figure 6 plotted with combined-sex mean forecast and observed life expectancies. The forecasts with sex-averaged mortality rates tend to be lower than combined-sex forecasts, diverging further over time. However, the difference is slight over the time frame considered.

Discussion

The observed trends in life expectancy since 1990 show that life expectancy in the developed world continues to rise in a monotonic, linear fashion, suggesting that the underlying structure of mortality decline has not changed. Further, this rise in life expectancy does not appear to have slowed compared to the historic trend as captured by the forecasts. We also find that male life expectancy has risen more rapidly than would be predicted using the standard Lee-Carter model, a feature that has been noted by others investigating the decline in the male-female difference in mortality[20,21]. The changes we observe in the separate-sex models cannot be explained by incorporating higher singular values into the model.

The cause of the divergence between forecast and observed male life expectancy is uncertain. However, one of the major features of health improvement over the late 20th century was decline in

cardiovascular disease (CVD) mortality. In general, CVD mortality among men has declined at a rate faster than in women[22–24]. This lends itself to two hypotheses:

1. Improvements in the treatment of CVD and its risk factors since the 1980s, including use of ACE inhibitors, beta blockers, and statins as well as advances in the care of myocardial infarction, have generally benefitted men more than women. There has been some evidence for differential benefit, primarily in terms of utilization of new technologies in women versus men[25,26].
2. Changes in health behavior, particularly tobacco use, have changed more rapidly and dramatically among men than women, and account for this differential. This has recent support from work by Preston and Wang incorporating tobacco use into Lee-Carter forecasting[27].

It is also notable that, while forecasts have diverged substantially from observed life expectancy for males, they remain within prediction limits for combined sexes. While the evidence presented is not conclusive, it suggests that sex-composition effects may explain this particular observation.

Further study should focus on sex differences in mortality. This includes decomposition of the sex- and age-specific contributions to combined-sex life expectancy. Additionally, this should clarify whether differences in CVD mortality decline between men and women can explain the observed difference in the accuracy of the Lee-Carter forecasts and, if this is the case, whether this is driven by differences in health behaviors or in benefit from new health technology.

References

1. Lee, R. D. & Carter, L. R. Modeling and Forecasting U.S. Mortality. *J. Am. Stat. Assoc.* **87,** 659–671 (1992).

2. Lee, R. The Lee -Carter Method for Forecasting Mortality, with Various Extensions and Applications. *North Am. Actuar. J.* **4,** 80–93 (1999).

3. Tuljapurkar, S., Li, N. & Boe, C. A universal pattern of mortality decline in the G7 countries. *Nature* **405,** 789–92 (2000).

4. Li, N. & Lee, R. D. Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. *Demography* **42,** 575–594 (2005).

5. Hyndman, R. J., Booth, H. & Yasmeen, F. Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography* **50,** 261–83 (2013).

6. Li, N. & Gerland, P. Modifying the Lee-Carter method to project mortality changes up to 2100. in Paper presented at the Population Association of A (2012). at <http://esa.un.org/unpd/ppp/pdf/Li-Gerland_2011_Modifying the Lee-Carter method to project mortality changes up to 2100.pdf>

7. Li, N., Lee, R. & Gerland, P. Extending the Lee-carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography* **50,** 2037–51 (2013).

8. Arias, E. United States Life Tables, 2008. *Natl. Vital Stat. Reports* **61,** (2012).

9. Armstrong, G. L., Conn, L. A. & Pinner, R. W. Trends in infectious disease mortality in the United States during the 20th century. *JAMA* **281,** 61–6 (1999).

10. Omran, A. R. The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Mem. Fund Q.* **49,** 509–38 (1971).

11. Olshansky, S. J. & Ault, A. B. The fourth stage of the epidemiologic transition: the age of delayed degenerative diseases. *Milbank Q.* **64,** 355–91 (1986).

12. University of California Berkeley (USA) & Max Planck Institute for Demographic Research (Germany). Human Mortality Database. at <www.mortality.org>

13. Preston, S. H., Keyfitz, N. & Schoen, R. *Causes of death: life tables for national population*. 787 (Seminar Press, 1972). at <http://books.google.com/books?id=EuwbAAAAIAAJ&pgis=1>

14. Preston, S., Heuveline, P. & Guillot, M. *Demography: Measuring and Modeling Population Processes*. 291 (Wiley, 2001). at <http://books.google.com/books?id=xlv2H6LpNAwC&pgis=1>

15. Edwards, R. D. & Tuljapurkar, S. Inequality in Life Spans and a New Perspective on Mortality Convergence Across Industrialized Countries. *Popul. Dev. Rev.* **31,** 645–674 (2005).

16. Tuljapurkar, S. & Edwards, R. D. Variance in death and its implications for modeling and forecasting mortality. *Demogr. Res.* **24,** 497–526 (2011).

17. R Core Team. R: A Language and Environment for Statistical Computing. (2012). at <http://www.r-project.org/>

18. Hyndman, R. J. & Khandakar, Y. Automatic Time Series Forecasting: The forecast Package for R. *J. Stat. Softw.* **27,** 1–22 (2008).

19. RStudio. RStudio : Integrated development environment for R. (2012). at <http://www.rstudio.org/>

20. Wang, H. & Preston, S. H. Forecasting United States mortality using cohort smoking histories. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 393–8 (2009).

21. Technical Panel on Assumptions and Methods (2011). *Report to the Social Security Advisory Board*. (2011). at <http://www.ssab.gov/Reports/2011_TPAM_Final_Report.pdf>

22. Roger, V. L. *et al.* Trends in incidence, severity, and outcome of hospitalized myocardial infarction. *Circulation* **121,** 863–9 (2010).

23. Parikh, N. I. *et al.* Long-term trends in myocardial infarction incidence and case fatality in the National Heart, Lung, and Blood Institute's Framingham Heart study. *Circulation* **119,** 1203–10 (2009).

24. Levi, F. Trends in mortality from cardiovascular and cerebrovascular diseases in Europe and other areas of the world. *Heart* **88,** 119–124 (2002).

25. Vaccarino, V. *et al.* Sex and Racial Differences in the Management of Acute Myocardial Infarction, 1994 through 2002. *N. Engl. J. Med.* **353,** 671–682 (2005).

26. Gouni-Berthold, I., Berthold, H. K., Mantzoros, C. S., Böhm, M. & Krone, W. Sex disparities in the treatment and control of cardiovascular risk factors in type 2 diabetes. *Diabetes Care* **31,** 1389–91 (2008).

27. Preston, S. H. & Wang, H. Sex Mortality Differences in the United States: The Role of Cohort Smoking Patterns. *Demography* **43,** 631–646 (2006).

Tables

Table 1

	Canada	France	GBR	Germany	Italy	Japan	USA
Comb. Sexes	0.598	0.97	0.831	0.371	0.422	0.061	0.91
Males	0.052	0.903	0.777	0.090	0.002	0.044	0.794
Females	0.957	0.979	0.835	0.689	0.886	0.012	0.915

Chi-squared *p*-values for the difference between observed and forecast *k*'s

Figures

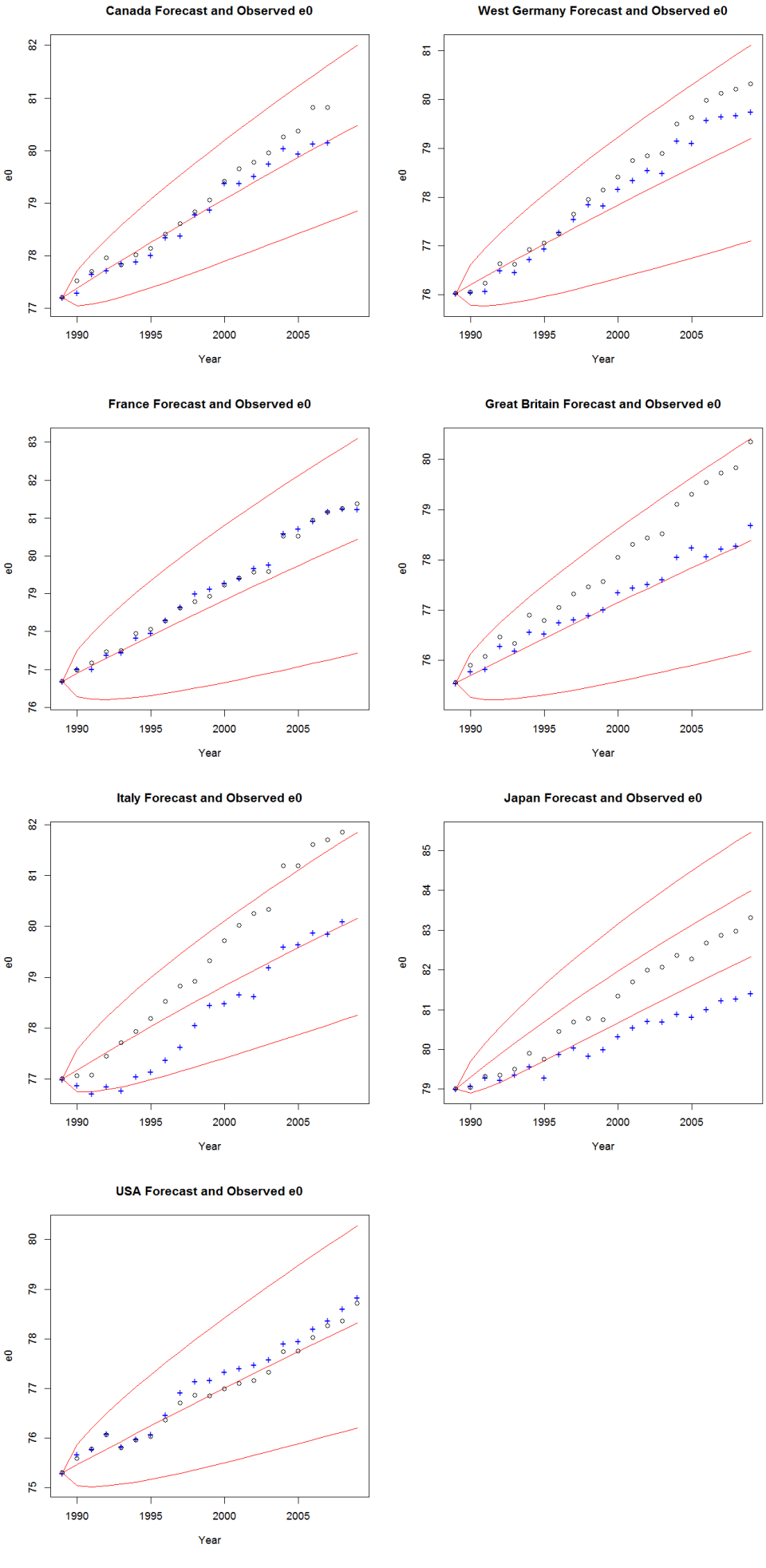


Figure 1: Combined sexes forecast, observed, and observed k_t-derived e0. Red lines: Forecast with 95% prediction interval. Black circles: Observed. Blue crosses: Derived from observed k_t

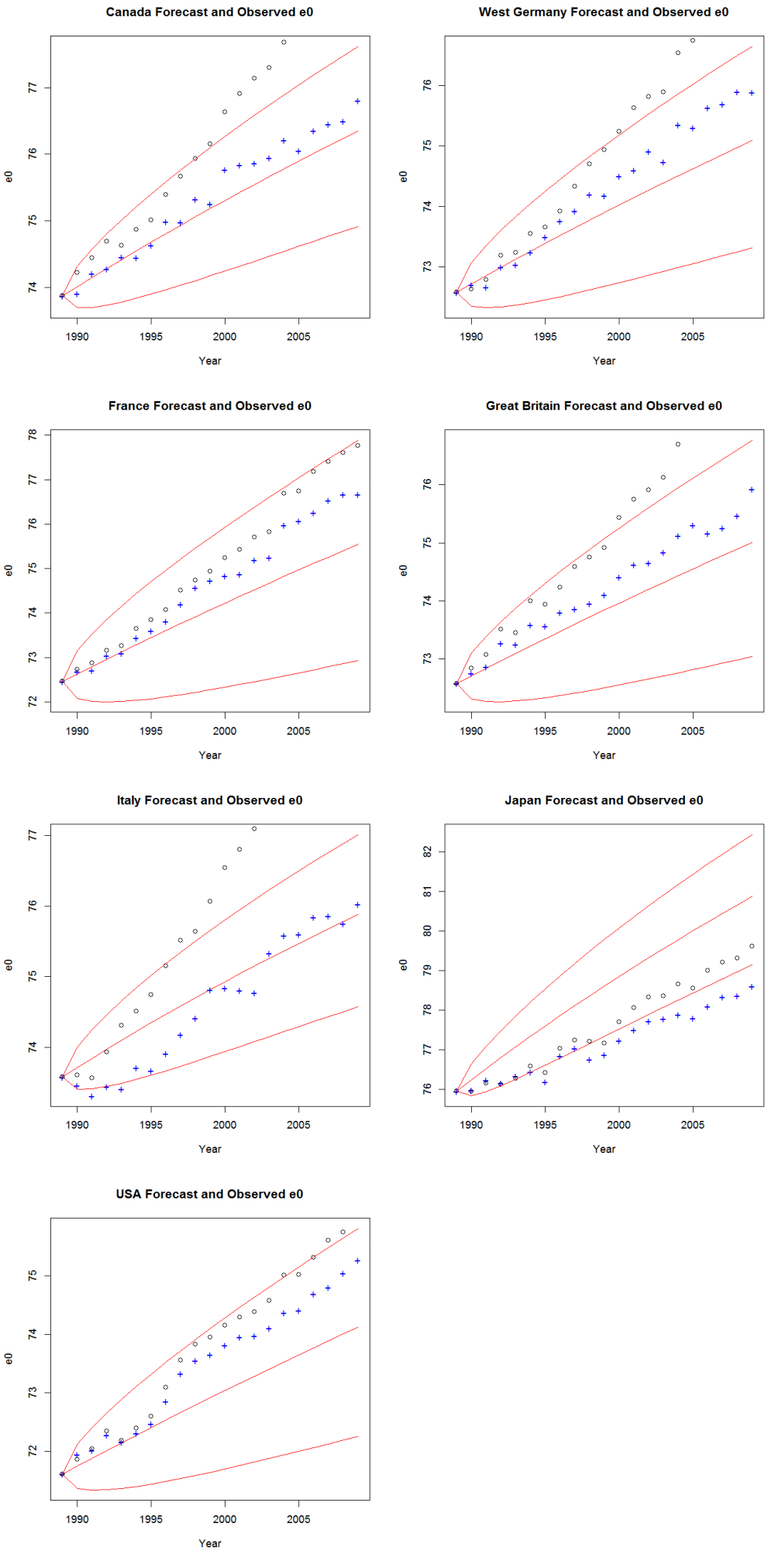
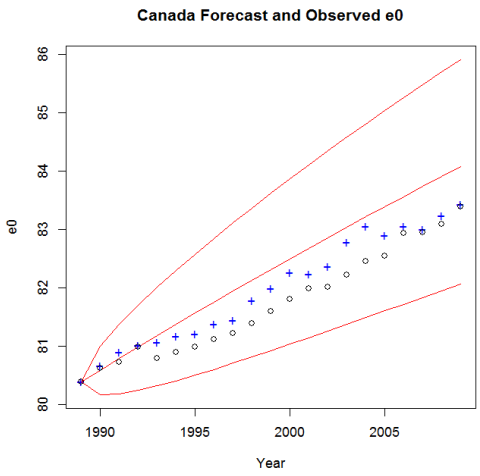
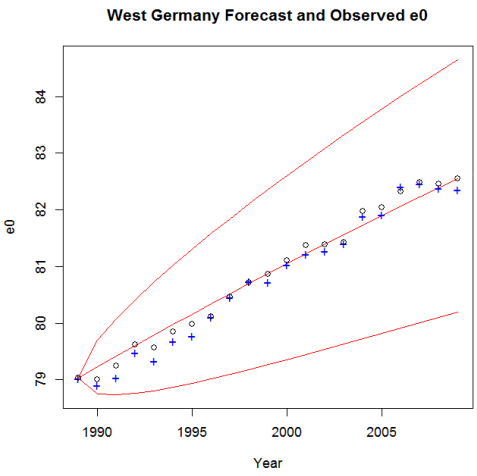
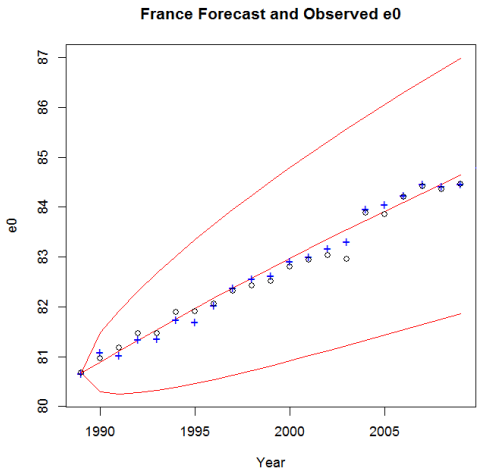
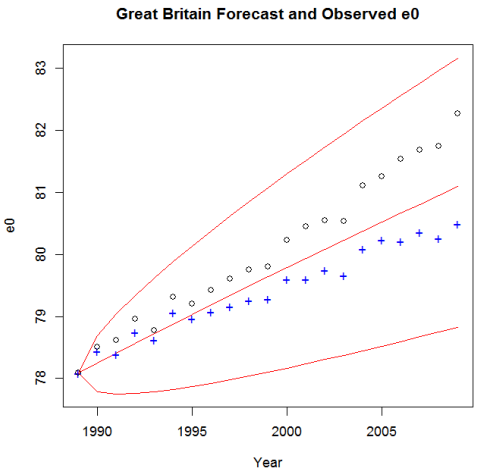


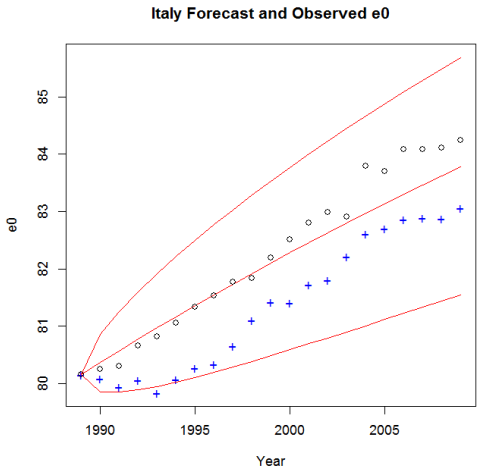
Figure 2: Male forecast, observed, and observed k_t-derived e0. Red lines: Forecast with 95% prediction interval. Black circles: Observed. Blue crosses: Derived from observed k_t

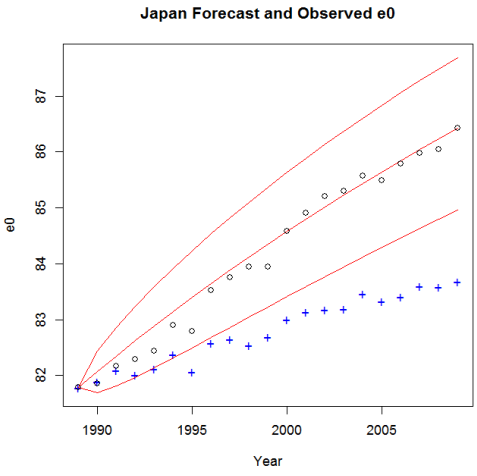












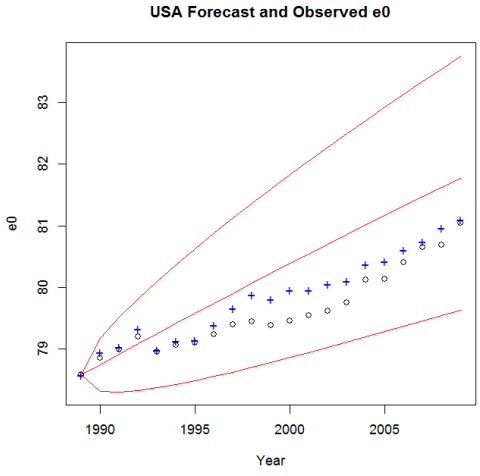
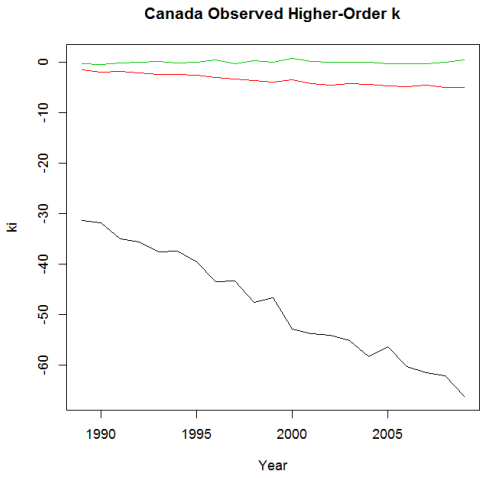
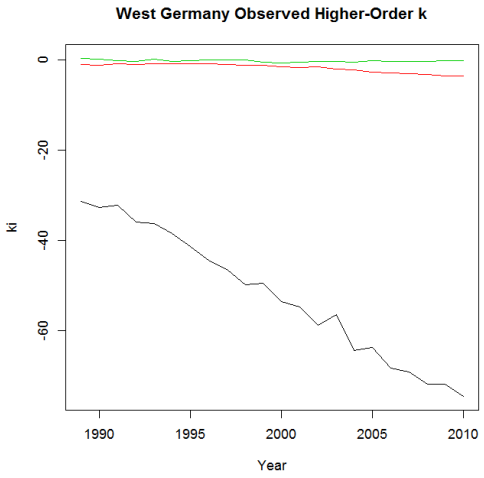
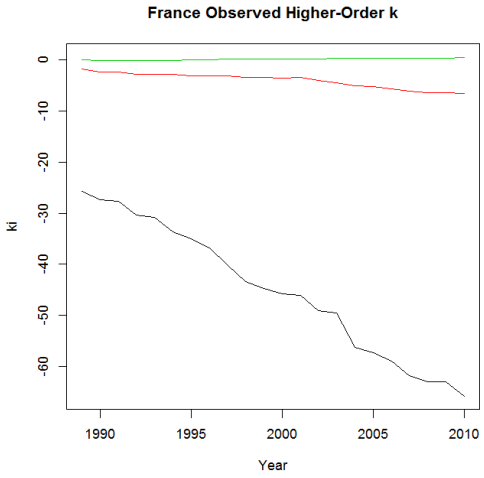
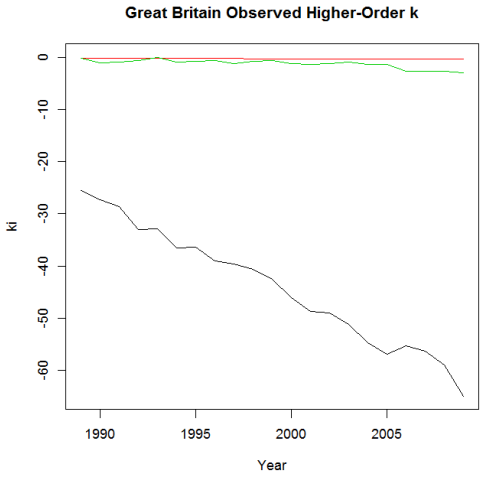


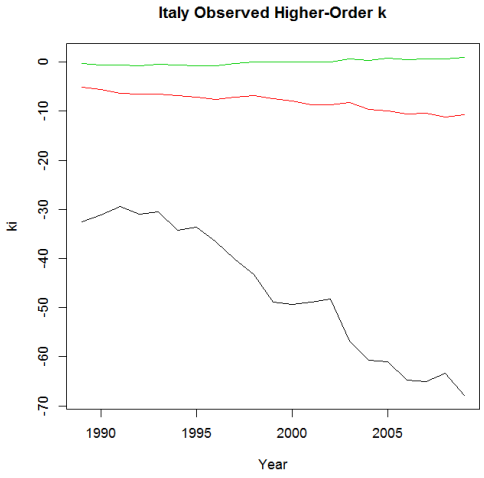
Figure 3: Female forecast, observed, and observed k_t-derived e0. Red lines: Forecast with 95% prediction interval. Black circles: Observed. Blue crosses: Derived from observed k_t

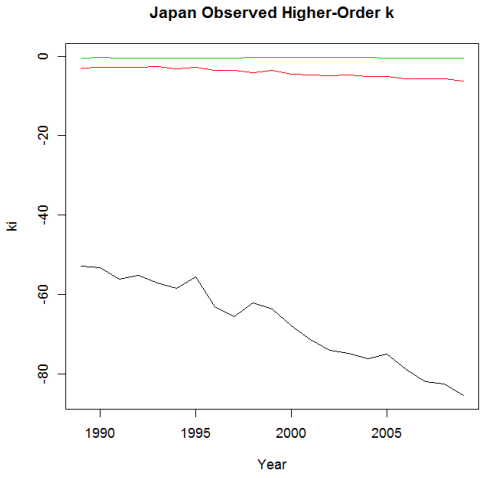












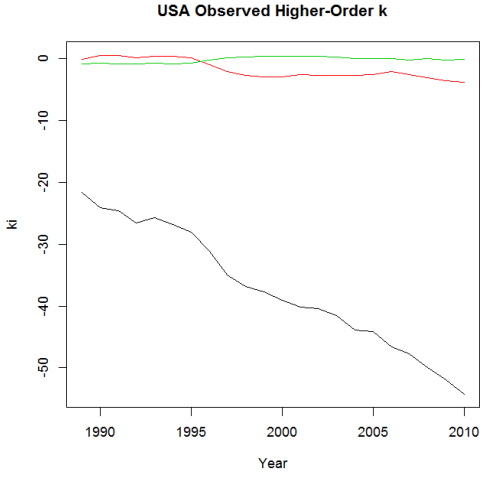
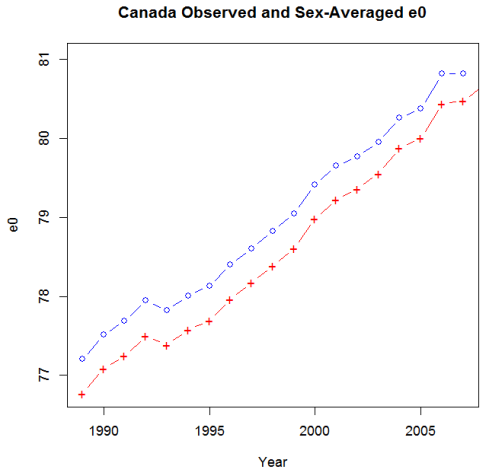
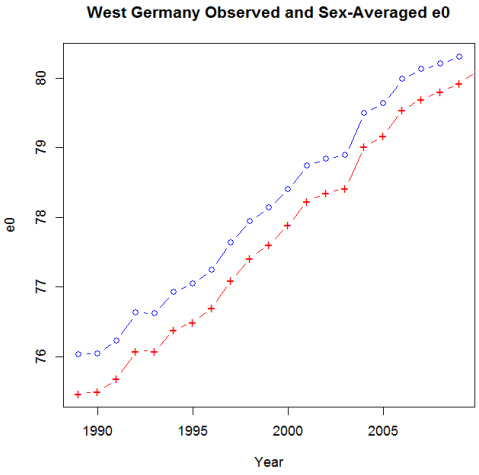
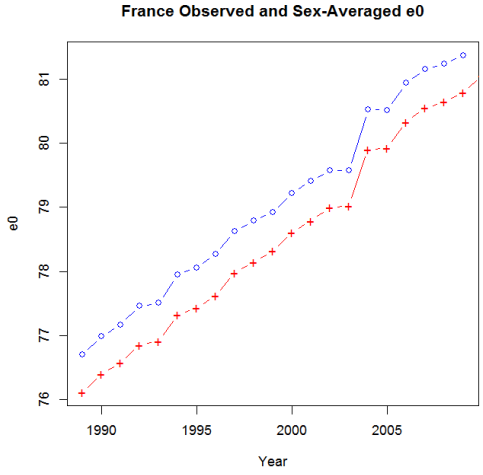
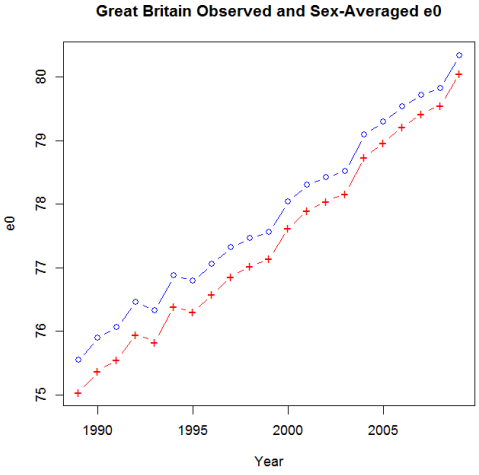


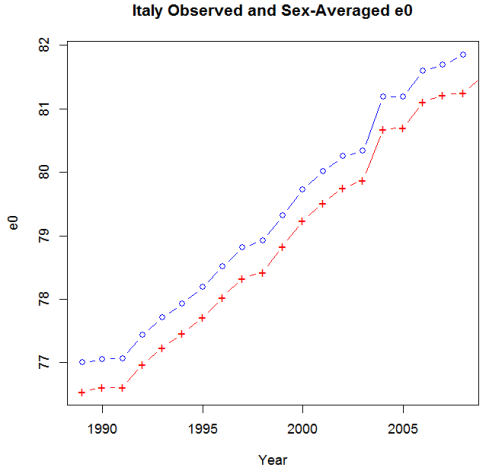
Figure 4: Observed higher-order *k*'s for males. Associated with 1st (black), 2nd (red), and 3rd (green) singular values from SVD of centered log mortality rates.

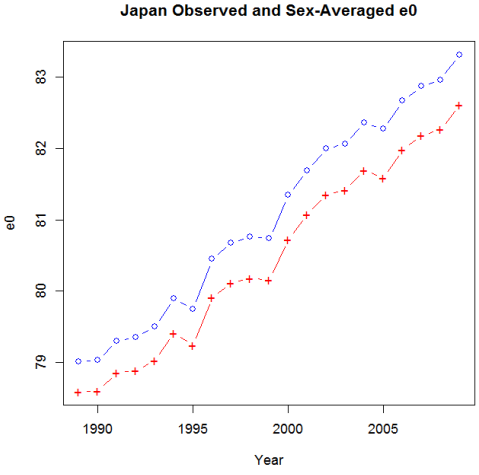












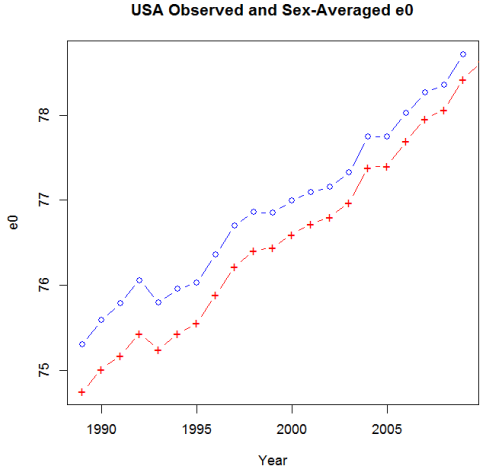


Figure 5: Life expectancy calculated for combined sexes (blue) and using averaged male and female age-specific mortality rates (red).

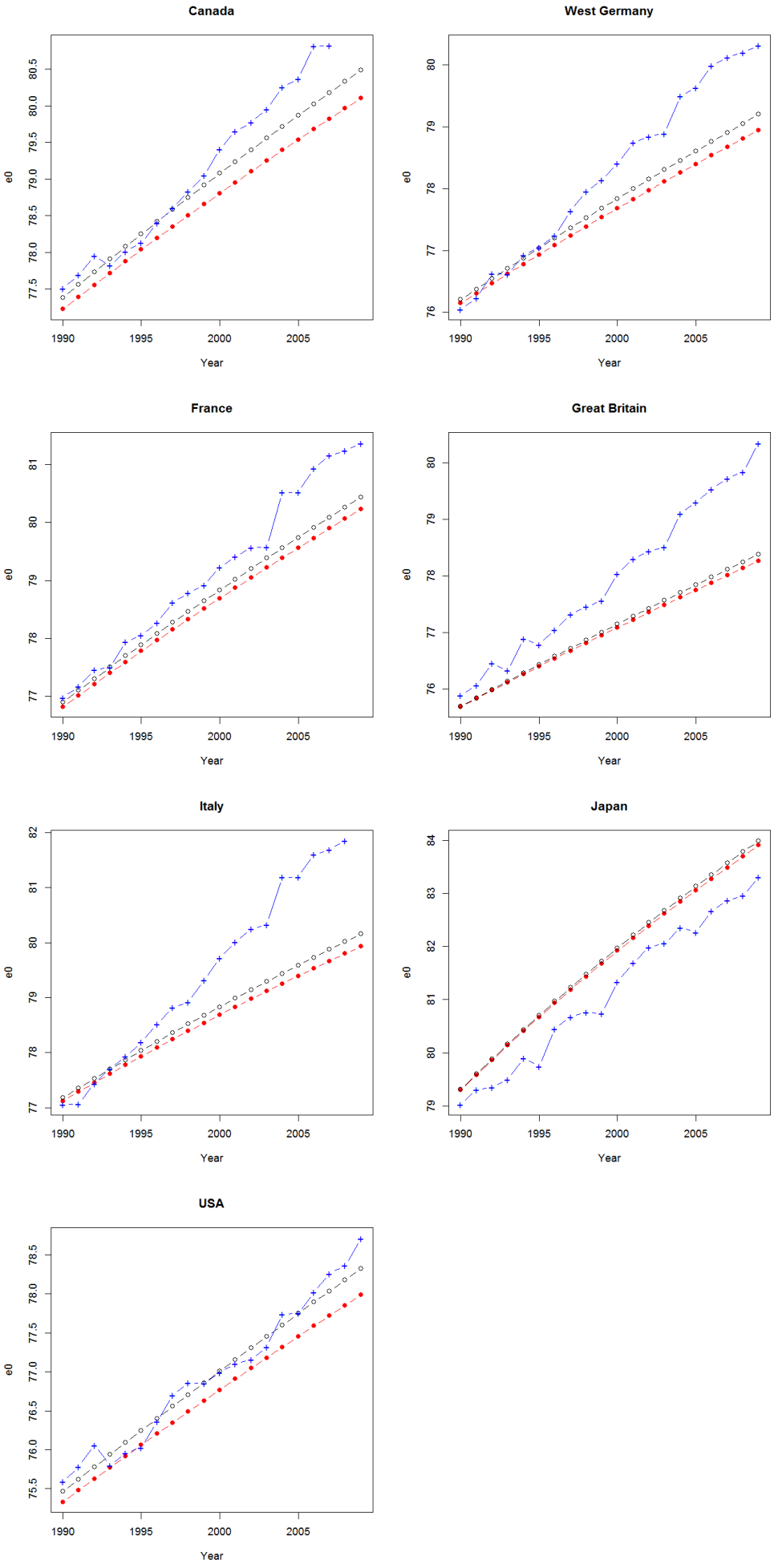


Figure 6: Life expectancy for combined sexes observed (blue) and forecast (black) and forecast with sex-averaged rates (red).