Chapter 15

Forecasting Life Expectancy: A Statistical Look at Model Choice and Use of Auxiliary Series

Juha M. Alho

15.1 Why Forecast Life Expectancy?

Let $\mu(x, t)$ be the hazard (or force) of mortality in age x at time t . Define $p(x, t)$ as the probability of surviving to age x , under the hazards of time t , or

$$p(x, t) = \exp\left(-\int_0^x \mu(y, t) dy\right).$$

Then, the expectation of the remaining life time in age $x \geq 0$, equals

$$e_x(t) = \int_0^\infty p(x + y, t) dy / p(x, t).$$

These are synthetic period measures, i.e., they are intended to summarize the chances of survival at time t . Life expectancy at birth, $e_0(t)$, is the most frequently used summary measure. Despite their popularity life expectancies are not directly used in cohort-component population forecasting. Instead, proportions of type

$$p(x + 1)/p(x) = \exp(-\Lambda_x(t)),$$

where $\Lambda_x(t)$ is the increment of the cumulative hazard in age $[x, x + 1)$, are used for proportions of survivors from exact age x to exact age $x + 1$. Similarly, in the computation of present values of annuities, for example, a cohort perspective is

J. M. Alho (✉)

University of Joensuu, Joensuu, Finland

e-mail: juha.alho@helsinki.fi

© The Author(s) 2019

T. Bengtsson, N. Keilman (eds.), *Old and New Perspectives on Mortality Forecasting*, Demographic Research Monographs,
https://doi.org/10.1007/978-3-030-05075-7_15

185

necessary. In that case, the more relevant concept is the remaining life time of a person alive at exact age $x \geq 0$, at time t , which equals

$$c_x(t) = \int_0^{\infty} \exp\left(-\int_0^y \mu(x+u, t+u) du\right) dy.$$

Since mortality has typically declined, we expect that $e_x(t) \leq c_x(t)$. We note that even if life expectancies $e_x(t)$ have considerable descriptive value, they are of limited direct usefulness in population forecasting.

Taken together the values of $e_x(t)$ do determine the hazards $\mu(x, t)$ for a given t , but if only $e_0(t)$ is known, then infinitely many patterns $\mu(x, t)$'s would produce the same value $e_0(t)$. In special cases, such as a proportional hazards model ($\mu(x, t) = \mu(x)g(t)$ with $\mu(x)$ known) or a log-bilinear model of the Lee-Carter type ($\mu(x, t) = a(x) + b(x)g(t)$ with $a(x)$ and $b(x)$ known), a one-to-one correspondence exists (e.g., Alho 1989). In these cases forecasting $e_0(t)$ leads directly to estimates of age-specific mortality, but the assumption of known multipliers is strong. Given that the multipliers may change over time, it is not clear that this would, in practice, lead to a more accurate forecast of mortality hazards than forecasting the latter directly.

On the other hand, $e_0(t)$ might perform as an “auxiliary measure” if it behaves in a more time-invariant manner (e.g., Törnqvist 1949) than the age-specific series themselves. The recent finding of Oeppen and Vaupel (2002), in which the so-called best-practice life expectancy, i.e., the life expectancy of the country that is the highest at any given time, was shown to have evolved almost linearly for 160 years, points to this possibility. The first purpose of this paper is to establish the empirical relationship of the best-practice life expectancy to country-specific life expectancies in selected industrialized countries, during the latter part of the 1900's. Simple regression techniques will be used. The second purpose is to examine the statistical underpinnings of using best practice life expectancy as an auxiliary series for the prediction of the country-specific life expectancies.

15.2 Changes in Life Expectancy in 19 Industrialized Countries in 1950–2000

Oeppen and Vaupel (2002) show that the best practice life expectancy for females has followed remarkably well ($R^2 = 0.99$) the model:

$$\tilde{e}_0(t) = 45 + (t - 1840)/4,$$

for $t \geq 1840$. Could this “invariant” be used as an auxiliary series to improve accuracy?

To examine this question empirically we have collected data on female life expectancies for 14 European countries, Australia, Canada, Japan, New Zealand,

Fig. 15.1 Life expectancies in 19 countries (Japan with a circle), and the best practice life expectancy (solid)

and the United States, for the periods 1950–55, 1955–1960, . . . , 1995–2000 (United Nations 2000). For ease of exposition, we denote the 5 year periods as $t = 1953, 1958, \dots, 1998$. Denoting life expectancy at birth in country $i = 1, 2, \dots, 19$ by $e_{0,i}(t)$ we define the variables of interest as:

$$\begin{aligned} \text{early life expectancy } LE53(i) &= e_{0,i}(1953); \\ \text{later life expectancy } LE78(i) &= e_{0,i}(1978); \\ \text{deviance } Dev(i) &= \tilde{e}_0(t) - e_{0,i}(t), \\ \text{early annual improvement } Early(i) &= (e_{0,i}(1978) - e_{0,i}(1953))/25, \\ \text{later annual improvement } Later(i) &= (e_{0,i}(1998) - e_{0,i}(1978))/20. \end{aligned}$$

Figure 15.1 shows the life expectancies of the 19 countries together with the best practice line. Two facts stand out. First, Japan has behaved in a radically different manner from the rest of the countries. A formal test using Mahalanobis’ distance (e.g., Afifi and Azen 1979, 282) also suggests that Japan is an outlier with a P-value < 0.001 . Second, all other countries appear to gradually veer off below the line. It is this set of 18 countries that we will be primarily concerned with in this paper.

To quantify the latter effect the following descriptive statistics were calculated for the 18 countries (Japan omitted):

Variable	N	Mean	Median	StDev
Dev53	18	−1.956	−1.550	1.942
Dev98	18	−3.928	−3.800	1.050

Fig. 15.2 Deviances in 1953 and 1998

Thus, the 18 countries that were an average of 2 years behind the best country in the early 1950’s (the best country being a member of the set of 18!), have fallen 2 years further behind in approximately 45 years. We also see that the spread among the 18 countries has decreased by a half.

For reference later, we note that *had* one forecasted life expectancy 45 years ahead in the first part of the 1950’s, by assuming that life expectancy will increase at the same rate as best practice life expectancy, then the average error in the 18 countries would have been 2 years.

Figure 15.2, which includes Japan, illustrates how different Japan is. However, it also reveals other interesting changes. For example, Denmark that was just under the best-practice line in the early 1950’s has fallen a full 6 years behind. The neighboring countries of Iceland, Norway and Sweden also fell behind, but by “three years only”. Thus, Denmark has, during a half a century, gradually distanced itself from the neighbors.

To examine country-specific changes more closely, we regressed the early improvement (Early) on life expectancy in the early 1950’s (LE53), among the 18 countries. The estimated coefficients are:

Predictor	Coef	SE Coef	T	P
Constant	1.4708	0.3257	4.52	0.000
LE53	−0.017432	0.004537	−3.82	0.002

with $R^2 = 47.7\%$. Regressing later improvement (Later) on life expectancy in the late 1970’s (LE78) yielded:

Predictor	Coef	SE Coef	T	P
Constant	2.5127	0.6091	4.13	0.001
LE78	−0.030312	0.007909	−3.83	0.001

Fig. 15.3 Early annual improvements as a function of life expectancy in 1953

Fig. 15.4 Later annual improvements as a function of life expectancy in 1978

with $R^2 = 47.9\%$. Figures 15.3 and 15.4 illustrate the same phenomenon. We find that in both cases the countries that had high life expectancy grew, on average, slower than those with low life expectancy. The well-known phenomenon of “regression to the mean” explains part of the changes, but we cannot ignore the possibility that there would be a tendency of having a lower rate of improvement when starting from a high value.

We then examined the persistence of improvement among the 18 countries. Correlations (with P-values for the hypothesis of zero correlation in parenthesis) between Later, LE78, and Early were (Japan omitted):

	Later	LE78
LE78	−0.692 (0.001)	
Early	0.342 (0.165)	−0.081 (0.748)

This suggests that there may be some persistence. However, when Later is regressed on LE78 and Early, the coefficients are

Predictor	Coef	SE Coef	T	P
Constant	2.3514	0.5855	4.02	0.001
LE78	−0.029288	0.007525	−3.89	0.001
Early	0.3617	0.2163	1.67	0.115

with $R^2 = 50.2\%$ (adjusted for the number of explanatory variables). While the regression is marginally better than the one not including Early (with $R^2 = 47.9\%$), the effect of Early is small and not significant. The regression is compatible with the notion that current level rather than past improvement has had a systematic association with the later development.

Descriptive statistics on early and later improvement among the 18 countries are as follows (Japan omitted):

Variable	N	Mean	Median	StDev
Early	18	0.2280	0.2140	0.0490
Later	18	0.1789	0.1950	0.0618

Had these statistics been used to forecast life expectancy in the late 1970's for the late 1990's, the average error would have been $20 (0.2280 - 0.1789) = 0.982$, as opposed to the average error of $20 \times (0.25 - 0.1789) = 1.422$ years that would have resulted from the use of the best practice line. I.e., the error of the latter forecast would have been about 50% higher.

We conclude that during 1950–2000, as life expectancy has increased, its annual improvement has gradually decreased. Based on Figs. 15.3 and 15.4 this holds for Japan, as well. The 18 countries have also come closer together, and they have fallen further behind Japan.

15.3 Conditions on the Usefulness of an Auxiliary Series

The model for the best-practice life expectancy says that (female) life expectancy at birth increases by 0.25 years every calendar year, but the 18 countries have fallen from 1.5 years behind in the 1950's to nearly 4 years behind in the late 1990's, on average. The deviance for the average of the 18 countries is a roughly linear function of time ($R^2 = 86.1\%$), and we estimate that the deviance has increased by about 0.05 years each calendar year. In 50 years time the best-practice line would imply an increase of 12.5 years, but if the average of the 18 countries continues to fall behind,

the increase would be less, or $12.5 - 0.05 \times 50 = 10.0$ years. In general, we might wish to establish an empirical relationship between the best practice line and the measure of interest, which we take here to be the average of the 18 countries.

Suppose there are some functions $f_j(t)$, $j = 0, 1, 2, \dots$, such that an invariant $g(t)$ is of the form

$$g(t) = \sum_{j=0}^m \alpha_j f_j(t).$$

Suppose the series of interest, say $e(t)$, is related to the invariant via

$$e(t) - g(t) = \sum_{j=0}^n \beta_j f_j(t) + \epsilon(t)$$

Where $\epsilon(t)$ is random with expectation $E[\epsilon(t)] = 0$. If $n \geq m$, then the *same* (e.g., generalized least squares) forecast for $e(t)$ is obtained by (a) modeling the difference $e(t) - g(t)$ and adding the result to $g(t)$ that is assumed to be known, or (b) by modeling $e(t)$ directly with the same explanatory variables $f_j(t)$, $j = 1, \dots, n$, but with modified coefficients $\gamma_j = \beta_j + \alpha_j$ (take $\alpha_j = 0$ for $j > m$). This follows from the fact that if the result of (a) is known, then the result of (b) can be deduced, and *vice versa*. Thus, in this case the knowledge of the invariant provides no help.

On the other hand, suppose $m > n$, or the invariant $g(t)$ behaves in a more complex manner than the deviance $e(t) - g(t)$. In this case, if the future values of the invariant can be assumed to be known for all t , we can reduce the dimensionality of the problem to m explanatory variables by modeling the deviance from the invariant. This can be of important practical use, especially if the future values of some of the functions $f_j(t)$, $j = n + 1, \dots, m$, are unknown. From this perspective having a linear invariant (with $m = 2$ only) is, paradoxically, the least helpful!

An alternative point of view is that *if* there is information about the difference $e(t) - g(t)$ that has not been reflected in the past values of the series $e(t)$, then such information can be introduced *via judgment* into forecasting. In the example at hand, suppose one believes that there is a feedback mechanism in operation such that if the life expectancy of a country falls sufficiently far behind the best-practice life expectancy, then corrective action will be taken by the society to reduce the deviance, in the future. This is a reasonable hypothesis, and presumably such an effect could manifest itself in the future. For example, even though Denmark has distanced itself from its neighbors for a half a century, perhaps later it will recoup some of the loss. More generally, if the 18 countries that have fallen behind Japan transform their life style in such a way that it resembles more that of Japan in terms of nutrition, job-security, attitude to leisure etc., then maybe they will begin to catch up. However, as this is a strong judgmental assumption that has to be defended by means other than statistical analysis, we will next pursue a number of alternatives that a statistical analyst might consider.

15.4 Model Choice

Figure 15.5 shows, in accordance with the earlier analyses, that the average improvement was higher in the early part of the observation period than in the later part. If the intention is to forecast until, say, 2050, the observation period is rather short, and alternative ways of viewing the trend are plausible. (a) Disregarding the first appearance, if we assume that the series is actually stationary, then the mean (*) is approximately the best predictor after a few years. (b) If we think that the series is a random walk, then the last observation (·) is the best predictor. (c) If we think that there is an exponentially linearly declining trend in the series, then the best prediction also declines exponentially (×). (d) If we think there is a linear trend, then the best predictor is the estimated linear line (+).

Forecasting as far as 2050, a choice between (a) – (d) can make a tremendous difference (this was pointed out in a more general context by Whelpton et al. 1947, already):

- using the historical average we expect to gain $50 \times 0.2062 = 10.3$ years;
- using the latest value we expect to gain $50 \times 0.15 = 7.5$ years;
- using the exponential trend we expect to gain 5.9 years;
- using linear trend we expect to gain 4.0 years.

All values are below the expected gain of 12.5 years derived from the linear model for the best practice life expectancy.

To distinguish between the models we can first examine the estimated variance of the residuals under models (a) – (d) and the best practice line model that assumes a

Fig. 15.5 Average annual improvement in average life expectancy during five-year periods, of the 18 countries (Japan excluded), in 1950–2000, and four forecasts based on historical average (*), last observed value (·), exponential trend (×) and linear trend (+)

constant rate of increase of 0.25 years per calendar year. The number of data points is $n = 10$ (from ten 5-year periods), and the number of estimates of annual increase is $n - 1 = 9$. The residual degrees of freedom in models (a) – (d) are 8, 8, 7, and 7, respectively. The best practice line model has 9 degrees of freedom, because it has no estimated parameters. Compared in this manner we find that the estimated variances of the residuals in the five models are 0.0041, 0.0042, 0.0031, 0.0031 and 0.0056. In view of Fig. 15.5, it is not surprising that the two regression models lead to the best fit. Similarly, it is not surprising that the last model with a rate coming from the outside of the data set fits the worst. The fact that the random walk model is not among the best is informative. Although the regression models fit the best, we recognize that the data period is short and one cannot take results of this type as decisive.

Another possibility is to try to find supporting evidence based on alternative approaches to the same problem. Here the “rates-to-life expectancy” comparison is available. The life expectancy of the Finnish women in 2000 was 81.0 years, or essentially the same as the average of 80.6 for the late 1990’s, of the 18 countries. A stochastic forecast (Alho 2002) that assumed the decline in age-specific mortality to continue in each age at the rate of the most recent 15 years lead to a median of female life expectancy in 2050 of 86.7, indicating a gain of 6 years. This agrees with the assumption of an exponential decline model (c). We will examine this model further.

Consider a function $e(t)$ such that $e(0) = A$ and $e'(t) = e^{\alpha - \beta t}$, $\beta > 0$, for $t \geq 0$. It follows that

$$e(t) = A + B(1 - e^{-\alpha t}),$$

where $B = e^{\alpha}/\beta$. Taking $t = 0$ to correspond to the late 1990’s, we have $A = 80.6$ and our empirical estimates can be translated to values $\alpha = -1.8472$ and $\beta = 0.01151$, which imply that $B = 13.7$. Under this model the average life expectancy of the 18 countries would never exceed $80.6 + 13.7 = 94.3$ years. For the year 2050 we would get the value $80.6 + 6.2 = 86.8$, for example. (The increase here is slightly larger than the 5.9 years given above, because the starting period is earlier.)

To complement the above point estimates we note that by using the so-called delta-method (e.g., Rao 1973, 385–6) we can compute a standard error for the estimate of B , as 9.4 years. Thus a 95% confidence interval for the additional improvement is quite wide, approximately 13.7 ± 18.4 years. From this, the 95% upper limit for the average life expectancy of the 18 countries would be about $94.3 + 18.4 = 112.7$ years. Of course, even under this model, individual people can live much longer.

Figure 15.6 has a graph of the past data together with a point forecast until the late 2040’s. Visually, the slight concavity smoothly continues from the past data to the point forecast.

Fig. 15.6 Average life expectancy of the 18 countries in 1950–2000 continued with a forecast based on an exponential trend in annual improvements for 2001–2050

15.5 Concluding Remarks

We have investigated statistically the possible use of the best-practice life expectancy as an aid in forecasting the life expectancy of industrialized countries. The evidence shows that during the past 50 years this would have been overly optimistic. The results do not preclude the possibility that in the longer term a comparison to the best practice line might prove to be useful, but beliefs concerning this cannot be based on statistical analyses of the type we have conducted. Instead, arguments concerning processes, whose effects have not manifested themselves yet, are required.

Better fits would have been provided by models that incorporate the slowing down of improvement in life expectancy, among the countries studied. A model that assumed a geometric slowing down leads to an absolute upper bound for life expectancy, but estimates about this upper bound are statistically quite uncertain. The validity of such a model cannot be ascertained based on the short data period we consider.

Independently of whether life expectancy turns out to be approximately linear or concave (or convex!) in the long run, there may well be other periods besides the latter part of the twentieth century, in which groups of countries veer off the trend for decades. From the perspective of individual countries this possibility would have to be allowed in the construction of prediction intervals.

In case one is not willing to choose an appropriate model at all, one can try to assign probabilities to each model, and do model averaging (Draper 1995). This approach has the advantage of leading to more honest prediction intervals, as it does

not condition on a particular choice, but the disadvantage is that it requires the assignment of probabilities. It may be difficult to achieve a consensus on the latter.

References

- Afifi, A. A., & Azen, S. P. (1979). *Statistical analysis* (2nd ed.). New York: Academic.
- Alho, J. M. (1989). Relating changes in life expectancy to changes in mortality. *Demography*, 26, 705–709.
- Alho, J. M. (2002). *The population of Finland in 2050 and beyond* (Discussion Paper 826). Helsinki: The Research Institute of Finnish Economy.
- Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, B*, 45–97.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570), 1029–1031.
- Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.
- Törnqvist, L. (1949). Näkökohdat, jotka ovat määränneet primääristen prognoosiolettamusten valinnan. In J. Hyppölä, A. Tunkelo & L. Törnqvist (Eds.), *Suomen väestöä, sen uusiutumista ja tulevaa kehitystä koskevia laskelmia* (Tilastollisia tiedontantoja 38). Helsinki: Tilastokeskus.
- United Nations. (2000). *World population prospects. The 2000 revision: Volume I: Comprehensive tables*. New York: United Nations.
- Whelpton, P. K., Eldridge, H. T., & Siegel, J. S. (1947). *Forecasts of the population of the United States 1945–1975*. Washington, DC: U.S. Bureau of the Census.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.