Comment

JUHA M. ALHO*

Lee and Carter are to be congratulated for their elegant contribution to the methods of demographic forecasting. They first study the quality of the age-specific mortality data from 1900–1989 and then use essentially the first principal component of the logarithmically transformed data to model past changes. This reduces the task of forecasting into a problem of forecasting a univariate mortality index $\mathbf{k}_t$. The authors carefully apply standard time series techniques to forecast $\mathbf{k}_t$ and age-specific mortality up to the year 2065. Interestingly, their extrapolations lead to similar decreases in future mortality as the low mortality variant of the judgmental forecasts of the Social Security Administration's Office of the Actuary. This suggests that the high–low prediction interval of the Office of the Actuary may not cover all reasonable contingencies.

Because Lee and Carter intend their forecasts as direct competitors of the official forecasts, I will comment on some of the limitations of their extrapolative technique and the credibility of the interval forecasts it produces. I will conclude with a note on terminology.

1. EXTRAPOLATION MODEL

I am less optimistic than the authors about the realism of the family of the life tables that is generated by the past average mortality and its first principal component. By construction, such life tables produce reasonable (in this case very good) fits to data from which they are estimated, but it is not at all clear that life tables corresponding to the values of $\mathbf{k}_t$ that are outside the observed range continue to be "reasonable." A case in point is Figure 4. The observed lack of fit in age group 15–20 is accentuated by the extrapolation to produce an implausibly sharp increase in mortality in the year 2065. Moreover, I expect that the technique will get into trouble if applied to time periods with stagnating mortality, such as the years 1954–1968 in the U.S., when male mortality *increased* slightly. During such periods mortality may increase in some ages and decrease in others. Extrapolation can rapidly magnify such effects out of proportion.

Of course, such issues will be problems with any purely extrapolative technique. Because purely judgmental forecasts that rely on *expert judgment* appear not to be optimal either, it might be valuable to try techniques that combine the two approaches. From a statistical point of view expert judgment can be interpreted as representing such extraneous information that has not been reflected in the past of the time series under study. As discussed in Alho (1992), Bayesian techniques, mixed forecasting, or the techniques of combining forecasts provide tools for such an undertaking. Combining the extrapolation forecasts of Lee and Carter with those of the Office of the Actuary would produce intermediate values depending on the weight given to each. With equal weights, a point forecast of the joint life expectancy of males and females would be about 83.5 years, for example.

2. INTERVAL FORECASTS

A nice feature of Lee and Carter's technique is that the only variable that needs to be forecast, $\mathbf{k}_t$, happens to have behaved fairly linearly during this century. In consequence, a random walk model with a drift fits the data well—perhaps too well! The forecasts produced by this model are essentially straight lines through the first and last observation. Experimentation with alternative observation periods and a transparent ruler in Figure 2 suggests that the 95% prediction intervals displayed may be too narrow.

One reason for the narrowness of the intervals is that the authors used a dummy for the flu epidemic of 1918. Although we would not expect flu to have a similar impact in the future, there may well be other dramatic changes in mortality. I don't think AIDS will be such a factor, but something equally unpredictable, perhaps caused by war or environmental catastrophe, could dramatically alter mortality schedules. Because the authors do not make any allowance for possible modeling error in their interval estimates, I think the exclusion of the flu epidemic from error estimation is hard to defend. The authors note in passing that including the epidemic in error estimation would make the prediction interval of $\mathbf{k}_t$ 57% wider in the year 2065. It would have been interesting to see the effect in the intermediate years and on life expectancy, especially because we see from Appendix B that all error terms are not used in calculating the prediction intervals for life expectancy. I believe that the wider intervals would have been more credible than the ones presented.

To illustrate, suppose that inflating the error estimate of $\mathbf{k}_t$ would increase the uncertainty of the forecasts of life expectancy by one-half. Then a 95% prediction interval around the point forecast that combines the Lee–Carter extrapolations with the judgment of the Office of the Actuary would be from approximately 75 to 89.5 years.

3. TERMINOLOGY

Although the lack of ideas frequently goes hand in hand with an interest in semantics, I will risk the appearances and comment on the use of the term "confidence interval" in connection with forecasting. I think it might be clearer to

* Juha M. Alho is Professor of Statistics, University of Joensuu, SF-80101 Joensuu, Finland.

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Applications & Case Studies

reserve this term for interval estimators of parameters. Interval estimators of random variables (which is what we are discussing here) might be better called "prediction intervals" or "forecast intervals" as, say, in Kendall and Buckland (1971).

ADDITIONAL REFERENCES

Alho, J. M. (1992), "Estimating the Strength of Expert Judgment: The Case of U.S. Mortality Rates," *Journal of Forecasting*, 11, 157-167.

Kendall, M. G., and Buckland, W. R. (1976), *A Dictionary of Statistical Terms*, London: Longman Group.

Rejoinder

RONALD D. LEE and LAWRENCE R. CARTER

We are grateful for the thoughtful comments on our article provided by two scholars who have themselves made pioneering contributions to the problems we discuss. Because we disagree with some of McNown's points, we will devote the most space to them.

McNown asserts that our method is "equivalent" to "directly projecting each age-specific mortality rate at its own historical rate of exponential decline" . . . "despite their statements to the contrary." This is an important point, because our method is rather complicated, whereas straight extrapolation is very simple. We have addressed this point in our article and also will respond at some length here: First, in our model each death rate declines at its own exponential rate only when **k** declines linearly. This is not an assumption of the model, and in other applications **k** might follow some other sort of process. Second, if each age-specific rate is forecast separately, then deriving confidence intervals for forecasts of period life table functions such as life expectancy, that depend on many death rates, requires taking into account the covariance matrix of errors.

Third, in response to this comment, we have tried two versions of directly extrapolating individual age specific rates. We forecasted to 2065 using the endpoint-to-endpoint exponential rates of decline from 1933–1987 to extrapolate to 2065. The resulting rate forecasts were lower than ours for ages below 10, higher from 10–45, lower from 45–75, and higher thereafter. The percentage differences were often appreciable, ranging from plus 65 to minus 11. We also forecasted using regressions of the logs of the death rates on a constant and time. Such forecasts did indeed often come close to ours for 2065, although individual age group differences are as large as 25%. For example, for age group 30–34 our own forecast is .000180. An endpoint-to-endpoint extrapolation yields .00298, and the regression-based extrapolation yields .000225. Furthermore, comparing the regression estimates of rates of decline to our $\mathbf{b}_x$s shows that they differ by up to 8% after equivalent normalization. Comparison of the 95% probability interval for this age group in 2065 shows wider discrepancies: Our range is .00009 to .00036; the regression interval is .00016–.00031, or about half as wide (these figures do not reflect parameter uncertainty). In sum, the methods we tried for directly forecasting the individual rates led to point forecasts which, although somewhat similar to ours, differed in both level and age pattern, contrary to McNown's assertion that they would be "identical." They also led to very different confidence intervals.

Fourth, our method incorporates procedures for indirect estimation of mortality in periods when age-specific mortality data are unavailable. In our article we extended the time series back from 1933 to 1900 in this way, and forward from 1987 to 1989. This aspect of the method is helpful in applications for populations of developed countries and will be absolutely essential in many applications for populations of less developed countries such as China, where estimates of age-specific mortality may be available for only one or two years.

McNown also suggests that we actually have a 24-parameter model of mortality change, consisting of the 23$\mathbf{b}_x$s plus **k**. Perhaps our difference on this point is just semantic. The $\mathbf{b}_x$s are fixed by age and so do not change over time. Only **k** changes over time, and so only a single parameter, **k**, needs to be forecast. In the common language of demography we have a one-parameter family of life tables in exactly the same sense that the Coale–Demeny model life tables for a given region and sex are one-parameter life tables, even though the construction of the Coale–Demeny life tables involved two regression coefficients at each age, corresponding to our $\mathbf{a}_x$ and $\mathbf{b}_x$ coefficients. In our case a value of **k** allows us to identify uniquely a corresponding life table from the family. Of course one could vary the underlying coefficients ($\mathbf{a}_x$ and $\mathbf{b}_x$), but then one would be providing a basis for a new family of life tables. To forecast from a two-parameter family of life tables, such as those of Ledermann or Brass, one would have to forecast two parameters. The model used by McNown and Rogers can describe a single life table very efficiently, using only nine parameters versus the 47 required for our model. But in their 1990 forecasting application, even though six of these parameters are held constant over the forecast range, it still is necessary to forecast three of them. With six parameters held constant, this could be identified as a three-parameter life table system.

The two commentators suggest that our out-of-sample forecasts of the age pattern of mortality may not be "rea-

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Applications & Case Studies