December 2, 2002
date last saved: 12/01/02 3:51 PM
date last printed: 06/25/03 4:39 PM

Using the Lee-Carter Method to Forecast Mortality for Populations with Limited Data

Nan Li
Department of Sociology
University of Victoria
Victoria, Canada
linan@uvic.ca

Ronald Lee
Demography and Economics
University of California
2232 Piedmont Ave.
Berkeley, CA 94720
rlee@demog.berkeley.edu

Shripad Tuljapurkar
Department of Biological Sciences
Stanford University
Stanford CA 94305-5020
tulja@stanford.edu

Research for this paper was funded by a grant from NIA, R37-AG11761.

Introduction

Mortality forecasts are traditionally based on forecasters' subjective judgments, in light of historical data and expert opinions. This traditional method has been widely used for official mortality forecasts, and by international agencies. A range of uncertainty is indicated by high and low scenarios, which are also constructed through subjective judgements.

In the hands of a skilled and knowledgeable forecaster, the traditional method has the advantage of drawing on the full range of relevant knowledge for the middle forecast and the high-low range. However, it also has certain difficulties. First, official mortality projections in low mortality countries have been found to under-predict mortality declines and gains in life expectancy when compared to the subsequent outcomes (Keilman, 1997; National Research Council, 2000; Lee and Miller, 2001). United Nations' projections for European and North American countries have also under-predicted life expectancy gains (National Research Council, 2000:132). These errors have led to under-prediction of the elderly population, and particularly the oldest old. For Third World countries, the UN projections have come close on average for countries in Asia and Latin America (with only a small negative bias), but have seriously under-predicted gains in the Mideast/North Africa (National Research Council, 2000:132). For sub-Saharan Africa the projected gains have been much too great, due to the effects of the HIV/AIDS epidemic which could not have been anticipated. From this review of past performance, it appears that there may be a systematic downward bias in the traditional method, at least as it has been applied in this particularly historical period.

A second difficulty is that it is not clear how to interpret a variable's high-low range unless a corresponding probability for the range is stated. The traditional method, unfortunately, cannot provide such a probabilistic interpretation. Nor is it clear whether the range is supposed to refer to annual variations or to some sort of general trend or long run average. Third, it is not clear how to combine the uncertainty indicated by the high-low range with other uncertainties. How is the uncertainty of a forecast for a region, such as Asia, to combine the uncertainties of the forecasts of the individual countries in the region? Do we expect some cancellation of errors across the countries? Similarly, how are we to use the high-low range in assessing the overall uncertainty of a population projection that also involves high-low ranges for fertility and perhaps migration?

Recently, Lee and Carter (1992) developed a method (henceforth LC) that uses standard methods for forecasting a stochastic time series, together with a simple model for the age-time surface of the log of mortality, to model and forecast mortality. A forecast is produced for the probability distribution of each future age specific death. The method reduces the role of subjective judgment, since standard diagnostic and modeling procedures for statistical time series analysis are followed. Nonetheless, decisions must be made about how far back in history to begin, exactly what model to use, and how to treat historically specific shocks such as wars or intense epidemics.

The method has used to forecast mortality in a number of OECD countries. For the G7 countries, the LC method forecasted life expectancies that are significantly higher than official projections (Tuljapurkar, Li and Boe, 2000). Tests were performed for the US, in which projections were formulated at earlier dates based on data available before that date, and hypothetical tests were compared to the subsequent mortality (Lee and Miller, 2001). The resulting forecasts had a negative bias, but substantially less than the bias in the official projections of the time. The probability intervals were reasonably accurate. The 95% probability interval covered 97% of the subsequently observed life expectancies. Less complete performance tests for Canada, France, Sweden and Japan were also encouraging (Lee and Miller, 2001). The LC method has also been used to forecast mortality for some Third World countries, for example Chile (Lee and Rofman, 1994).

Like all time series analysis, the LC method extrapolates historical data. Applications to the US and other G7 countries were able to draw on mortality time series extending back at least a half-century, and often more. This was also true of the application to Chile. For most Third World countries, however, mortality data are very limited. For example for China, age-specific death rates at the national level are available only for the years 1974, 1981 and 1990. It has often been suggested that the LC approach can not be used widely for Third World countries because its data demands are too great, relative to what is typically available.

This paper discusses ways in which the LC method can be used for countries with limited mortality data.. To produce a LC forecast, four items of information are required, where the LC notation, to be introduced later, is given in parentheses: 1) a baseline age schedule of mortality (a_x); 2) the relative pace of change by age (b_x); 3) the overall rate of change (drift in the random walk model for k_t); 4) variability about the trend in mortality decline (the variance of the innovation term in the random walk model). Sometimes these items may be estimated for a particular country with severely limited data, using methods developed in this paper. In addition, it may sometimes be possible and desirable to borrow information from one or more other countries that are believed to be similar in relevant respects.

If age specific death rates are available for only a single year, then they can provide the baseline mortality schedule, and the other three items must be borrowed from another country. If age specific death rates are available for two years, then they can provide estimates of the baseline pattern, the pattern of change, and the rate of drift. One would need to borrow the variance from another country, but might also consider using another country for the drift and pattern of change as well, since these would be imprecisely estimated. If age specific death rates are available for three years, as in the case of China, then in principle one can estimate all four items of information and produce full forecasts with no borrowing. In practice, estimates may be too imprecise and one might want to borrow information, but that would not be a necessity.

In this paper, we will not develop the borrowing strategy, although it would also appear to be promising. Instead, we will consider single-country methods for dealing with

incomplete data, when we have age specific death rates for at least three periods, ideally separated by a number of years.

In order to apply the LC method to countries with limited mortality data, at least two questions need to be answered. The first is how to apply the LC method to mortality data collected at unequal intervals a number of years apart. The second question is what quality of results can we expect to derive through the LC method, when the historical data are only available for a small number of time points, as in the case of China. We answer these two questions in this paper.

The LC method using data at single-year intervals

Let the death rate for age x at time t be m(x,t), for t =0, 1, 2, …, T, and let the average over time of log(m(x,t)) be a(x). The LC method first applies the singular-value decomposition (SVD) on {log[m(x,t)]-a(x)} to obtain

$$\log[m(x,t)] = a(x) + b(x)k(t) + \varepsilon(x,t)\,. \qquad (1)$$

The purpose of using SVD is to transfer the task of forecasting an age-specific vector log[m(x,t)] into forecasting a scalar k(t), with small errors $\varepsilon(x,t)$. Notice that b(x)k(t) is an age (row) by time (column) matrix and the columns are proportional. The condition for $|\varepsilon(x,t)|$ to be small is that the columns of {log[m(x,t)]-a(x)} be close to proportional. This condition for $|\varepsilon(x,t)|$ to be small appears to hold not only for the G7 countries, but more generally, except for war and other unusual times. The SVD is a technique to maximally utilize the over-time similarity in the age pattern of {log[m(x,t)]-a(x)}, by finding b(x) and k(t) to minimize $\sum_{t=0}^{T}\sum_{x=0}^{\infty}\varepsilon(x,t)^2$. Define the explanation ratio to be $R = 1 - \sum_{t=0}^{T}\sum_{x=0}^{\infty}\varepsilon(x,t)^2 \Big/ \sum_{t=0}^{T}\sum_{x=0}^{\infty}\{\log[m(x,t)] - a(x)\}^2$. Actual values of R for the G7 countries over the period of 1950—1994 are greater than 0.94 in (Tuljapurkar, et al, 2000). In other words, more than 94% of age-specific mortality change in G7 countries between 1950 and 1994 was accounted for by change in k(t).

Ignoring the small errors $\varepsilon(x,t)$, the second stage of the LC method is to adjust k(t) to fit the reported values of life expectancy at time t. This stage leads to perfect description of life expectancy in history, and hence to better forecasts of future life expectancy in the future. (The original LC method fit the observed total number of deaths in the second stage, but fitting life expectancy is much simpler and works just as well.)

The adjusted k(t) is then modeled using standard time series methods. In most applications to date, it has been found that a random walk with drift fits very well, although it is not always the best model overall. Unless some other time series model is found to be substantially better, it is advisable to use the random walk with drift because of its simplicity and straightforward interpretation. The random walk with drift is expressed as follows:

$$k(t) = k(t-1) + c + e(t)\sigma, \quad e(t) \sim N(0,1), \quad E(e(s)e(t)) = 0. \qquad (2)$$

In (2), the drift term c, which is usually negative, represents the linear trend component in the change of k(t), while $e(t)\sigma$ represents deviations from this linear change as random fluctuations. A linear component exists in any change, and is generally more significant in shorter periods. According to (1), the linear component of k(t) corresponds to a constant rate of decline for m(x,t), reflecting a stable reduction in mortality. The linear component of k(t) has persisted through the second half of the 20th century and earlier for G7 countries (Tuljapurkar, et al, 2000). It should exist for other countries, so long as their mortality declines in a stable manner. This linear decline is the basis for the LC method to forecast mean mortality. Deviations from the linear change in k(t) are regarded as random fluctuations, modeled as $e(t)\sigma$, and then simulated to produce uncertainty for the forecasts.

For different t, [k(t)-k(t-1)] are assumed to be independently and identically distributed (i.i.d.) variables with mean c and standard deviation σ. Parameter c is estimated as the average across all observed t and t-1 of [k(t)-k(t-1)],

$$c = \frac{1}{T}\sum_{t=1}^{T}[k(t) - k(t-1)] = \frac{k(T) - k(0)}{T}. \qquad (3)$$

Using the estimated c, the standard error of $e(t)\sigma$ is estimated as

$$see = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}[k(t) - k(t-1) - c]^2}. \qquad (4)$$

The values of k(T) and k(0) in (3) are obtained only from one sample or realization of the matrix m(x,t). In other hypothetical realizations of history, yielding different samples, we would derive different sample values of c. Let the expected value of c be the average of all its sample values. The difference between the expected and sample values of c can be defined as the estimating errors of c. Using the estimated *see*, the standard error in estimating c is expressed as

$$\text{sec} = \sqrt{\frac{\sigma^2}{T}} \approx \frac{see}{\sqrt{T}}. \qquad (5)$$

Using the estimated values of c and *see*, and a set of sampled values of e(T) and e(s) for s=(T+1) to t, a trajectory of forecasted k(t) for t>T is obtained from (2) as,

$$k(t) = k(T) + [c + \text{sec}\cdot e(T)](t - T) + see\sum_{s=T+1}^{t} e(s). \qquad (6)$$

Note that this particular trajectory for future k(t) will depend partly on the estimated drift, c; partly on a random difference between the estimate of c and the true c; and partly on the random innovations.

A large number of such stochastically simulated trajectories for future k(t), 1000 in this paper, provides the basis for the stochastic forecast. The frequency distribution of these simulated trajectories provides an estimate of the probability distributions or confidence intervals for the forecast items of interest. In (6), the reason for e(T) to be independent from e(s) is, as pointed out by Lee and Carter (1992), that the e(T) describes random changes in the historical period while e(s) for s>T are in the future.

One trajectory of forecasted k(t) yields a corresponding trajectory of forecasted m(x,t) from (1) as

$$\log[m(x,t)] = \log[m(x,T)] + b(x)[k(t) - k(T)], \qquad (7)$$

and a large number of trajectories compose the stochastic forecasts of m(x,t). Note that in (7), the most recently observed age-specific death rates, m(x,T), are used as the baseline mortality rather than ax as in the original LC. This approach here seems preferable because it ensures that the forecasts begin from the most recently observed mortality schedule (Bell, 1997).

The LC method using data at unequal intervals

Now let mortality data be collected at times u(0), u(1), …, u(T). In the case of China, u(0)=1974, u(1)=1981, and u(2)=1990. Parameters a(x) are calculated as $\sum_{t=0}^{T} \log[m(x,u(t))]/T$. Applying SVD on [log[m(x,u(t))]-a(x)], b(x) and k(u(0)), k(u(1)), …, k(u(T)) are obtained.

For k(u(t)), however, (2) becomes

$$k(u(t)) - k(u(t-1)) = c[u(t) - u(t-1)] + \sigma[e(u(t-1)+1) + ... + e(u(t))]. \qquad (8)$$

Thus, for different t, [k(u(t))-k(u(t-1))] are no longer identically distributed. Consequently, estimating c and σ from (8) cannot be as simple as that for i.i.d. variables.

Because the means of the second term in the right-hand side of (8) are still zero, the unbiased estimate of c is obtained as:

$$c = \frac{\sum_{t=1}^{T}[k(u(t)) - k(u(t-1))]}{\sum_{t=1}^{T}[u(t) - u(t-1)]} = \frac{k(u(T)) - k(u(0))}{u(T) - u(0)}. \qquad (9)$$

Since the variances of the second term in the right-hand side of (8) are no longer the same for different t, the derivation of the standard error of e(u(t)), *see*, becomes somewhat complicated, and is derived in the appendix as

$$see^2 = \frac{\sum_{t=1}^{T}[(k(u(t)) - k(u(t-1)) - c[u(t) - u(t-1)]]^2}{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t) - u(t-1)]^2}{u(T) - u(0)}}. \quad (10)$$

The standard error in estimating c, *sec*, is obtained from (9) and (10) as

$$\sec^2 = \frac{\mathrm{var}\{\sum_{t=1}^{T}[e(u(t-1)+1) + ... + e(u(t))]\}}{[u(T) - u(0)]^2} = \frac{\sigma^2}{u(T) - u(0)} \approx \frac{see^2}{u(T) - u(0)}. \quad (11)$$

When [u(t)-u(t-1)]=1, (9), (10) and (11) reduce to (3), (4) and (5), respectively. Having the values of c and *see*, forecasting is carried out by (6) and (7), regardless of whether we are using data with single-year-intervals or unequal-intervals.

The equations presented above give the answer to the first question posed, how to apply the LC method to data observed at unequal-intervals. We now turn to the second question: when the historical data are available only at a few time points, what results can we realistically expect the LC method to provide?

The mean forecasts based on data at few time points

A special feature of the LC method is that it converts the task of forecasting an age-specific vector log[m(x,t)] into that of forecasting a scalar k(t). We will start by discussing how data limitations affect the forecast of k(t).

First, c is the average rate of decline in k(t), both for forecasting and for describing history. Just as the average speed of linear movement depends only on the initial and terminal positions and their times, so c is determined only by the first and last values of k(u(t)) and u(t), and is independent of other values of k(u(t)), as can be seen in (9). Thus, the mean forecasts of k(t) depend mainly on the death rates at starting and ending points of the historical period, and mortality data at years between the two points do not matter much. This property implies that the mean forecasts generated by applying the LC method to countries with limited data could be just as accurate as those for the G7 countries, if the formers' historical data span a long enough time period. In the example of China, the mean forecasts are determined by death rates in 1974 and 1990. What happens in between, and how often it is observed, does not matter.

Second, (11) indicates that the error in estimating c declines with the length of the historical period [u(T)-u(0)], not with the number of time points (T+1) at which mortality data are available. This conclusion can be explained intuitively. According to (9), a given random disturbance in k(u(T)) or k(u(0)) will make smaller difference for the estimated c, when the denominator, [u(T)-u(0)], is larger. In the example of China, if the estimated c were not close enough to its expected value, the reason would be that the period of 17 years is not long enough, not that the 3 time points are too few.

Turning to mean forecasts of m(x,t), (7) shows that a(x) can be omitted altogether in forecasting, and that $mean\{\log[m(x,t)]-\log(m(x,T)]\}=mean\{b(x)[k(t)-k(T)]\}$. We show in the appendix that b(x) is estimated without bias, and the errors in estimating b(x) are independent of k(t), so that mean forecasts of k(t) can be used to derive mean forecasts of m(x,t). The answer to a part of the second question, therefore, is that the LC method can provide accurate mean mortality forecasts for countries with historical data at only a few time points, if the earliest and latest points are sufficiently far apart in time.

The probability intervals for forecasts based on data at few time points

The probability intervals for k(t), such as the 95% probability interval of k(t) at different times, are based on *see* in (10), which captures historical random changes in k(t). To obtain positive *see* from (10), the number of time points must be larger than 2. In other words, for only two years of data, the LC method cannot provide uncertainty forecasts, since there is no deviation from the linear change of k(t).

Because *see* measures random deviation from the linear component of k(t), its estimation error, measured by var(*see*), should depend on the number of these fluctuations or the number of time points, not the length of the historical period. Since *see* usually is larger for faster changing k(t), we discuss *see*/|c|, which is the standard error per one year of linear change in k(t). For situation of unequal-intervals, (9) and (10) can be used to show how var(*see*/|c|) changes numerically with the number of time points.

We have drawn on the data used for the Tuljapurkar et al (2000) mortality forecasts for the G7 countries to study this point. Using mortality data for the extreme years of the G7 study, 1950 and 1994, and for one of the 43 intermediate years, we calculated 43 different values of *see*/|c|. The sample value of var(*see*/|c|) for 3 time points was obtained from the 43 resulting values of *see*/|c|, for each of the seven countries. To make a similar estimate for 4 time points, we can use all 903 possible ways of selecting two from the 43 years. To ease the computation, we can randomly choose 100 out of the 903 different values. Similarly, based on 100 different choices of 5 out of the 43 time points, we can again estimate *see*/|c| for 5 time points. Similarly, but starting from the other end, we can take away one of the 43 points in 43 different ways, to estimate *see*/|c| for using 44 time points, and so on. Figure 1, based on data from the G7 countries, plots the results.

It can be seen that reducing the number of time points raises var(*see*/|c|) for every country, as we would expect. However, we also see that there are large differences across

countries in var(*see*/|c|), particularly for smaller numbers of time points. It may be that the higher order moments of e(u(t)) differ for these countries.

To be conservative, we could choose the highest value of var(*see*/|c|) from Figure 1 to reflect the error in estimating *see*. Alternatively, we could select the average value of var(*see*/|c|) in Figure 1. Whichever value we choose for var(see/|c|)), the uncertainty of the forecast of k(t) is given as

$$k(t) = k(T) + [c + \frac{see + \sqrt{\text{var}(see)}e(T-1)}{\sqrt{u(T) - u(0)}}e(T)](t-T) + [see + \sqrt{\text{var}(see)}e(T-1)]\sum_{s=T+1}^{t} e(s). \quad (12)$$

When $\sqrt{\text{var}(see)} = 0$ and u(T)-u(0)=T, (12) reduces to (6). The reason why the random error in estimating *see*, described by $\sqrt{\text{var}(see)}e(T-1)$, is independent of e(T) is that the c is estimated without bias, so the mean$\{[c + \frac{see + \sqrt{\text{var}(see)}e(T-1)}{\sqrt{T}}e(T)]\}$=c. Because $\sum_{s=T+1}^{t} e(s)$ describes random changes in the future, while $\sqrt{\text{var}(see)}e(T-1)$ reflects estimating errors in using historical data, they should also be independent.

Just as the estimating errors of *c* raise the variance of k(t), so do the errors in estimating *see*, as shown by (9a) in the appendix. In other words, when the data are available at only a small number of years, the uncertainty forecasts that the LC method provides include additional variances, due to errors in estimating historical uncertainty.

Turning to the uncertainty of forecasts of m(x,t), we show in the appendix that the errors in estimating b(x) are negligible, when the explanation ratio of SVD is high and the number of time points is small. Thus, the uncertainty of forecasts of m(x,t) derives exclusively from uncertainty in the forecasts of k(t).

The answer to the other part of the second question is, therefore, is that the LC method applied to countries with only a few years of data can estimate the uncertainty of the forecasts. However, with only a few years of data, there will be additional variances due to errors in estimating the historical uncertainty. When the number of time points increases from 3, var(*see*) declines fast in Figure 1, so that the additional variances would decrease quickly.

Application to China

To provide an example of the worst situation for the LC method to estimate the uncertainty of its forecasts, we will apply it to the case of China. We use China's two-sex

combined mortality data for the years 1974, 1981 and 1990[1]. These data are in 5-year age groups and the open age interval covers 85 years and older. The Coale-Guo (1989) approach is used to extend death rates up to the group aged 105 to109 years, so that ages 110 and older form the open age interval. Applying SVD to these data, the explanation ratio is 0.96. In general, SVD tends to result in an higher explanation ratio when there are fewer years of data because then the number of parameters is relatively greater compared to the number of observations. In China, the year 1974 represented the time when both rural and urban populations were covered by essential but efficient health-care systems, and in the years 1981 and 1990 the rural health-care system collapsed due to the reform launched in 1978. Given the major change in the health-care system, 0.96 is a high value for the explanation ratio.

The mean forecasts would reflect longer trend of mortality change, if there were mortality data before 1974 or after 1990; but they do not require data at years between 1974 and 1990. Figure 3 compares our mean forecast of life expectancy for China to the United Nations middle projection (2001). The two forecasts are quite close overall, but our forecasts are initially higher and subsequently lower than those of the United Nations. Considering the impact on the health-care system from the urban reform in the 1990s, a life expectancy lower than our forecasts might well be observed, say from the 2000 census. Assuming a quick reinstatement of the healthcare system at the national level, our longer-term forecasts could turn out to be too low. These possibilities, however, are based more on subjective judgments than on recorded trends.

The random change in mortality, measured by $see/|c| = 1.74$, is stronger than in any of the G7 countries in the period 1950 to 1994. Considering the recent changes in the health-care system of China, such a high value of $see/|c|$ is not surprising. Without considering errors in estimating *see*, the estimates of forecast uncertainty expressed as 95% probability intervals for k(t) and life expectancy, are shown by the dashed curves in Figures 2 and 3, respectively. The value of *see*, however, is estimated from data at only three time points and hence may not be close enough to its expected value. To evaluate effects on the estimates of forecast uncertainty from the potentially inaccurate estimate of *see*, we take $var(see/|c|)=0.7$, which is larger than the largest in Figure 1. Taking this unlikely-high estimation error into account, the resulting 95% confidence intervals for k(t) and life expectancy are plotted by the solid curves in Figures 2 and 3. To different readers, the confidence intervals in Figure 3 may or may not be too wide, but these intervals are better than the high-low ranges which have no probabilistic interpretation.

Application to South Korea

Between the mortality data situation of China and the G7 countries, there are many Third World nations in transition from having limited mortality data to collecting death reports annually. All Third World countries will move through this transition sooner or later. For these countries, age-specific death rates are available annually in

[1] Data of years 1981 and 1990 are from census of 1982 and 1990. The 1974 data are from the China Death Cause Survey of 1973—1975,Yearbook of Chinese Population, 1985.

recent periods. However, such periods are often not long enough for the LC method to provide accurate forecasts. For these countries, the LC method can be used to forecast mortality by combining the recent annual data with earlier data available at unequal time intervals. The formulas developed in this paper apply directly to these countries, because whether or not the recent data are collected annually does not matter. To provide an example for using the LC method to these countries, we choose the case of South Korea.

The sex-combined age-specific death rates of South Korea are available for the years 1972, 1978, and then annually for 1983 through 2000[2]. The period that contains annual data lasts for17 years. Although it is hard to determine whether such a period is long enough to apply the LC method, adding data at the two earlier years improves the situation in any case. These data are also in 5-year age group and the open age interval covers 80 years and older for most of the years. The Gompertz formula is used to estimate the death date for the age group 80-84. The Coale-Guo approach is then used to extend death rates to the age group 105-109 years, and ages 110 and older form the open age interval. The explanation ratio of the fitted SVD model is only 0.84, implying that the changes in the age pattern of mortality have been stronger and less regular in South Korea than in China and the G7 countries.

The LC method uses a drift term in the random walk model to describe the linear change in k(t), and treats deviations of k(t) from this linear change as random fluctuations. When there are only a few years of data, these deviations are assumed to be random fluctuations, although it is not possible to rule out the presence of a nonlinear trend. In the case of South Korea, with 20 time points over a period of 28 years, we are on firmer ground. Figure 4 shows clearly that the k(t) did indeed change linearly with random fluctuations about the trend.

If there were no random fluctuations, the linear trend in the historical change of k(t) would suggest forecasting future changes of k(t) along such a linear trend, as is done for the mean forecasts of k(t) for 2002 through 2050 plotted in Figure 4. In history, however, k(t) did not change exactly along the linear trend, but fluctuated around it randomly. The standard error of these random fluctuations, estimated as *see*, measures the amount of uncertainty around the linear historical trajectory. Figure 1indicates that estimates of uncertainty should be quite accurate when based on the see for 20 years of data. The random walk model derives uncertainty forecasts for k(t), as described by the 95% confidence intervals in Figure 4, assuming that the random disturbances in the future will resemble those in the past. The forecasts of k(t) simply extrapolate the historical mean trend and uncertainty into the future, without subjective judgments.

The corresponding forecasts of life expectancy, derived from the forecasts of k(t) shown in Figure 4, are shown in Figure 5. It can be seen that the mean forecasts from using the LC method are significantly higher than those of the United Nations. Most of

[2] Data for years 1983 through 2000 were obtained from the Korea National Statistical Office (http://www.nso.go.kr/eng/). For 1972 and 1978 data were obtained from the United Nations (through personal communication with Thomas Buettner).

the difference can be attributed to the lower United Nations estimates of South Korea's life expectancy for 1980 to 1995. However, the United Nations forecasts would still be lower than the LC forecasts, even if the data used were the same.

For China, the 50-year LC forecast is for life expectancy of 76 in 2040, a gain of about 7 years over the level observed in 1990. The projected pace of increase is modest, at 1.4 years per decade. For South Korea, the 50-year LC forecast is for life expectancy of 88 in 2050, a gain of 12 years over the level observed in 2000. The forecasted pace of increase in South Korea is 2.4 years per decade, the rate of increase found by Oeppen and Vaupel (2002) for the record (or leader) national life expectancy from 1840 to 2000. Despite the historical precedent, this seems to be a very fast rate. The 2050 life expectancy forecast for South Korea is ahead of all LC forecasts for the G7 except that of Japan (Tuljapurkar et al, 2000). Is this reasonable and plausible? Or would we expect the pace of improvement in South Korean mortality to decelerate as it approached the life expectancy levels of the leader countries?

This question raises the general issue of whether mortality forecasts should be done not country by country, but rather for collections of countries in some coordinated way. One possibility is to model mortality change in individual nations as a process of convergence toward a trending target. That target could be tied to international trends, but reflect individual features of each country. The process of convergence would be subject to disturbance, as would the evolution of the international trend. Lee (2002) has developed a preliminary analysis of this sort. However, it is important to note that in these LC forecasts, Japan remains in the leader position, well ahead of South Korea. Therefore, the case for deceleration would have to be based solely on the plausibility that South Korea could overtake the leading European countries by 2050, which it is now trailing by 2 to 4 years.

Discussion

The methods developed here extend the LC approach to situations in which mortality data are available at only a few points in time, and at unevenly spaced intervals, situations often encountered in statistics for Third World countries. We have shown that useful forecasts can still be derived, both for the mean and for the probability interval about the mean forecast. Other modifications of the approach, not developed here, would include borrowing missing information from similar countries, and forecasting mortality change as a process of convergence within an international system.

Appendix

A. Estimating variance for independently distributed variable e(u(t))

Similar to the single-year-interval situation, we start from describing $E\{[k(u(t))-k(u(t-1))-c[u(t)-u(t-1)]]^2\}$, using the c estimated from (9). Since [k(u(t))-k(u(t-1))] are independently distributed, so that which one to be the first does not matter and we may focus on t=1. Suppose that for t=1 the second term of the right-hand side of (8) is [e(1)+ e(2)+ …+ e(m)], and the e(t) included in the whole historical period are e(1), e(2), …, e(n), there is

$$E\{[k(u(1))-k(u(0))-\frac{\sum_{t=1}^{T}[k(u(t))-k(u(t-1))]}{\sum_{t=1}^{T}[u(t)-u(t-1)]}[u(1)-u(0)]]^2\}$$

$$=E\{[[k(u(1))-k(u(0))]-c[u(1)-u(0)]-(\frac{\sum_{t=1}^{T}[k(u(t))-k(u(t-1))]}{\sum_{t=1}^{T}[u(t)-u(t-1)]}-c)[u(1)-u(0)]]^2\} \quad (1a)$$

$$=E\{[\sigma\sum_{i=1}^{m}e(i)-[\frac{\sum_{t=1}^{T}[k(u(t))-k(u(t-1))]}{\sum_{t=1}^{T}[u(t)-u(t-1)]}-c][u(1)-u(0)]]^2\}$$

$$=E\{[\sigma\sum_{i=1}^{m}e(i)-\frac{\sum_{t=1}^{T}[k(u(t))-k(u(t-1))]-c[u(t)-u(t-1)]}{\sum_{t=1}^{T}[u(t)-u(t-1)]}[u(1)-u(0)]]^2\}$$

$$=E\{[\sigma\sum_{i=1}^{m}e(i)-\frac{m\sigma\sum_{i=1}^{T}e(i)}{n}]^2\}$$

$$=\frac{\sigma^2}{n^2}E\{[(n-m)(e(1)+...+e(m))-m(e(m+1)+...+e(n))]^2\}.$$

Notice that all e(i) in the last row of the right-hand side of (1a) are i.i.d. variables and are different each other with respect to i, all cross terms, e(s)e(t), shall disappear. Therefore

$$\{[k(u(1))-k(u(0))-c[u(1)-u(0)]]^2\}$$

$$=\frac{\sigma^2}{n^2}E\{[(n-m)^2[e(1)^2+e(2)^2+...+e(m)^2]+m^2[e(m+1)^2+...+e(n)^2]\}$$

$$=(\frac{n-m}{n})^2 m\sigma^2+(\frac{m}{n})^2(n-m)\sigma^2 \qquad (2a)$$

$$=(\frac{n-m}{n})m\sigma^2$$

$$=[1-\frac{u(1)-u(0)}{u(T+1)-u(1)}][u(1)-u(0)]\sigma^2.$$

Because which [k(u(t))-k(u(t-1))] to be used as the first does not matter, (2a) applies to any t:

$$E\{[k(u(t))-k(u(t-1))-c[u(t)-u(t-1)]]^2\}=[1-\frac{u(t)-u(t-1)}{u(T+1)-u(1)}][u(t)-u(t-1)]\sigma^2. \quad (3a)$$

Sum (3a) through all t and divide the coefficient of σ^2 on both sides, there is

$$\sigma^2=E\{\frac{\sum_{t=1}^{T}[k(u(t))-k(u(t-1))-C[u(t)-u(t-1)]]^2}{[u(T)-u(0)-\frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}]}\}. \qquad (4a)$$

Therefore,

$$see^2=\frac{\sum_{t=1}^{T}[k(u(t))-k(u(t-1))-C[u(t)-u(t-1)]]^2}{[u(T)-u(0)-\frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}]}, \qquad (5a)$$

is the unbiased estimate of σ^2.

B. Errors in estimating *see* raise variance of k(t)

Let

$$A = \frac{see(T-1)}{\sqrt{u(T)-u(0)}}, \quad B = see, \quad C = \sqrt{\text{var}(see)}\,/\,see,$$
$$X = e(T), \quad Y = \sum_{s=T+1}^{t} e(s), Z = e(T-1), \tag{6a}$$

(12) is written as

$$k(t) = k(T) + c(t-T) + (AX + BY)(1 + CZ). \tag{7a}$$

Denote the probability density function of (X, Y, Z) by F(x,y,z), and notice that the means of X, Y, and Z are zero, the variance of k(t) is

$$\text{var}[k(t)] = \sum_x \sum_y \sum_z (Ax + By)^2 (1 + Cz)^2 F(x, y, z). \tag{8a}$$

Notice that X, Y, and Z are independent each other and denote their probability density functions as Fx(x), Fy(y) and Fz(z) respectively, there is

$$\begin{aligned}\text{var}[k(t)] &= [A^2 \sum_x x^2 Fx(x) + 2AB \sum_x xFx(x) \sum_y yf(y) + B^2 \sum_y y^2 Fy(y)][\sum_z (1 + Cz)^2 Fz(z)] \\ &= [A^2\ \text{var}(X) + B^2\ \text{var}(Y)][\sum_z Fz(z) + 2C \sum_z zFz(z) + C^2 \sum_z z^2 Fz(z)] \\ &= [A^2\ \text{var}(X) + B^2\ \text{var}(Y)][1 + C^2\ \text{var}(Z)] \geq [A^2\ \text{var}(X) + B^2\ \text{var}(Y)].\end{aligned} \tag{9a}$$

Because $[A^2\ \text{var}(X) + B^2\ \text{var}(Y)]$ is the variance of k(t) when $C = \sqrt{see}\,/\,see = 0$, errors in estimating *see* raise the variance of k(t).

C. Errors in estimating a(x) and b(x)

In order to discuss errors in estimating a(x) and b(x), their expected values of must be defined. Viewing the values of m(x,t) as of from one sample, the corresponding values of a(x), b(x), k(t) and $\varepsilon(x,t)$ in (1) are also from this sample. The values of m(x,t) would be different in other samples, so that (1) would produce different sample values of a(x), b(x), k(t) and $\varepsilon(x,t)$ in other samples. Let expected values of a(x) and b(x) be corresponding averages of all sample values, the errors in estimating a(x) and b(x) can be defined as the differences between sample and expected values.

Without enough sample values of a(x) and b(x), their expected values cannot be obtained and therefore assumptions have to be introduced. For example, in assessing the errors of estimating c, e(t) in (2) are assumed as i.i.d. variables. In order to assess errors in estimating a(x) and b(x), $\varepsilon(x,t)$ in (1) have to be assumed as i.i.d. variables over time t, and independent across age x. In fact, these assumptions have already been used in applying SVD, because SVD minimizes $\sum_{t=0}^{T} \sum_{x=0}^{\infty} \varepsilon^2(x,t)$, and terms $\varepsilon(x,t)\varepsilon(y,s)$ are ignored.

Noticing that the LC method uses k(t) to explain m(x,t) in history and forecast m(x,t) in the future, k(t) and m(x,t) can be regarded as independent and dependent variables respectively. Observable variable values may be common, but not necessary. In structural equation models, for example (Agresti and Finlay, 1997: 634—638), independent and dependent variables are unobservable but measured using factor analysis on other observable variables. In terms of structural equation model, k(t) is a latent variable that describes the underlying force of mortality change, and SVD is used to measure the values of k(t) from observed m(x,t). From this point of view, although values of a(x) and b(x) are estimated by SVD, they can be re-estimated using ordinary least square (OLS) on the unequal-interval version of (1) for each x separately,

$$\log[m(x,u(t))] = a(x) + b(x)k(u(t)) + \varepsilon(x,u(t)) . \qquad (10a)$$

In (6a), values of log[m(x,u(t))] are observed, of k(u(t)) are measured by SVD, and $\varepsilon(x,u(t))$ are assumed i.i.d variables. The reason of using OLS is that its estimates of a(x) and b(x) are identical to that of SVD, since otherwise one of the SVD or OLS does not minimize its target function. There are three reasons of doing the re-estimation. The first one is that it interprets a(x) and b(x) as unbiased estimates in terms of OLS. The second reason is this re-estimation points out that the errors in estimating a(x) and b(x) can be assumed independent from k(t), because these errors come from $\varepsilon(x,u(t))$ that are orthogonal to k(t) according to SVD. The third reason is that the re-estimation assesses errors in estimating a(x) and b(x) (e.g., Fox ,1997: 115) as

$$\text{var}(a(x)) = \frac{\sigma_\varepsilon^2(x)}{u(T) - u(0)}, \qquad (11a)$$

$$\text{var}(b(x)) = \frac{\sigma_\varepsilon^2(x)}{\sum_{t=0}^{T} k^2(u(t))}, \qquad (12a)$$

$$\sigma_\varepsilon^2(x) \approx \frac{\sum_{t=0}^{T} \varepsilon(x,u(t))^2}{T} . \qquad (13a)$$

Equations (11a)—(13a) show that var(a(x)) and var(b(x)) come from the SVD errors $\varepsilon(x,u(t))$. Involving estimating errors in a(x) and b(x), therefore, is to take the SVD errors into account. By doing so, potential improvements, in explaining historical change of m(x,t), would be to reduce the (1-R) unexplained fraction left by SVD to some extend, which may not be necessary when the R is close to 1. To do so, $\sigma_\varepsilon^2(x)$ needs to be precisely estimated, which is impossible for using data at small number of time points. Therefore, involving estimating errors in a(x) and b(x) is an issue that is sophisticated

when the SVD explanation ratio is high, and difficult when the number time points is small.

References

Agresti, A. and B. Finlay, 1997. Statistical Methods for the Social Sciences. Prentice Hall, Inc. New Jersey.

Bell, William R. (1997) "Comparing and Assessing Time Series Methods for Forecasting Age Specific Demographic Rates" *Journal of Official Statistics* 13:279-303.

Coale, A. and G. Guo, 1989. Revised Regional Model Life Tables at Very Low Levels of Mortality. *Population Index* 55: 613—643.

Fox , J., 1997. Applied Regression Analysis, Linear Models, and Related Methods. Sage Publications, London.

Keilman, Nico (1997) "Ex-post errors in official population forecasts in industrialized countries." *Journal of Official Statistics* (Statistics Sweden) 13(3): 245-277.

Keilman, N., 1998. How Accurate Are the United Nations World Population Projections? Population and Development Review 24: 15—41.

Ronald Lee (2002) "Mortality Forecasts and Linear Life Expectancy Trends," paper prepared for a meeting on mortality forecasts, for the Swedish National Insurance Board, Bund, Sweden, September 4, 2002.

Lee, R. D. and L. Carter, 1992. Modeling and Forecasting the Time Series of U.S. Mortality. *Journal of the American Statistical Association* 87: 659—71.

Lee, Ronald and Rafael Rofman (1994) "Modelacion y Proyeccion de la Mortalidad en Chile" in *NOTAS de Poblacio*n, v.XXII, No. 59 (Junio), pp.183-213.

Lee, R. D. and T. Miller, 2001. Evaluating the Performance of the Lee-Carter method for Forecasting Mortality. Demography 38: 537—49.

National Research Council, 2000. Beyond Six Billion: Forecasting the World's Population, edited by J. Bongaarts and R. A. Bulatao. Washington, DC, National Academy Press.

Oeppen, J. and J. W. Vaupel, 2002. Broken limits to life expectancy. Science 296: 1029—31.

Tuljapurkar, S., N. Li and C. Boe, 2000. A Universal Pattern of Mortality change in the G7 Countries. Nature 405:789—92.

United Nations, 2001. World Population Prospects. The 2000 Revision. New York.

Figure 1. Sampled values of var(see/|c|)

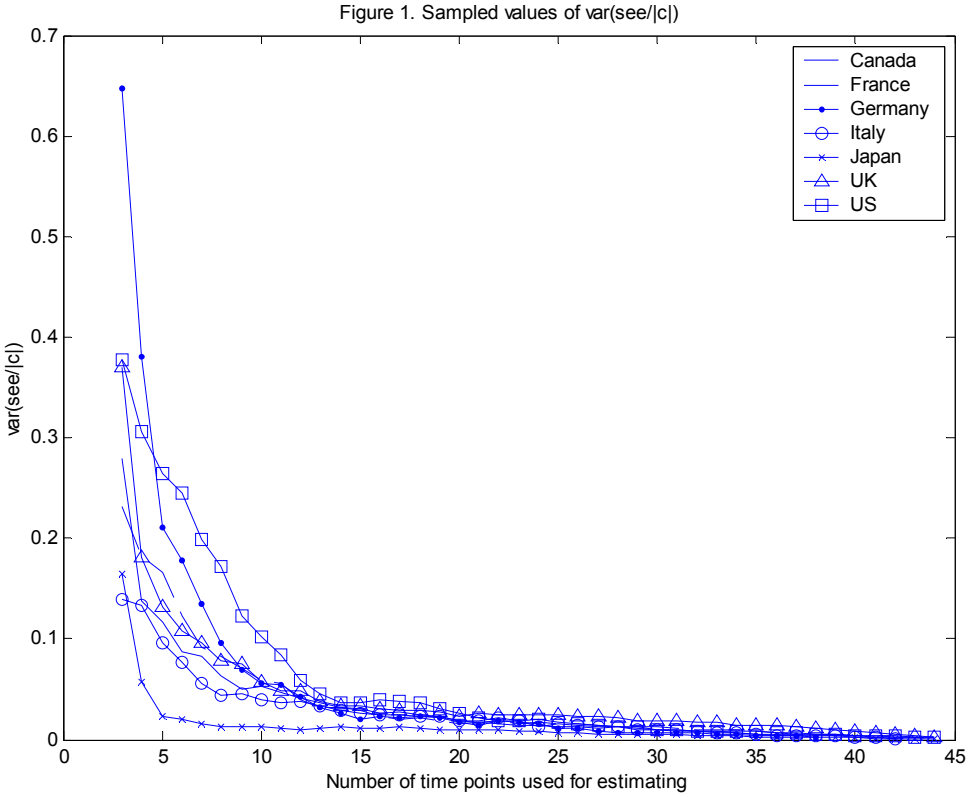

Figure 2. Historical values and 95% forecasted ranges of k(t) of China

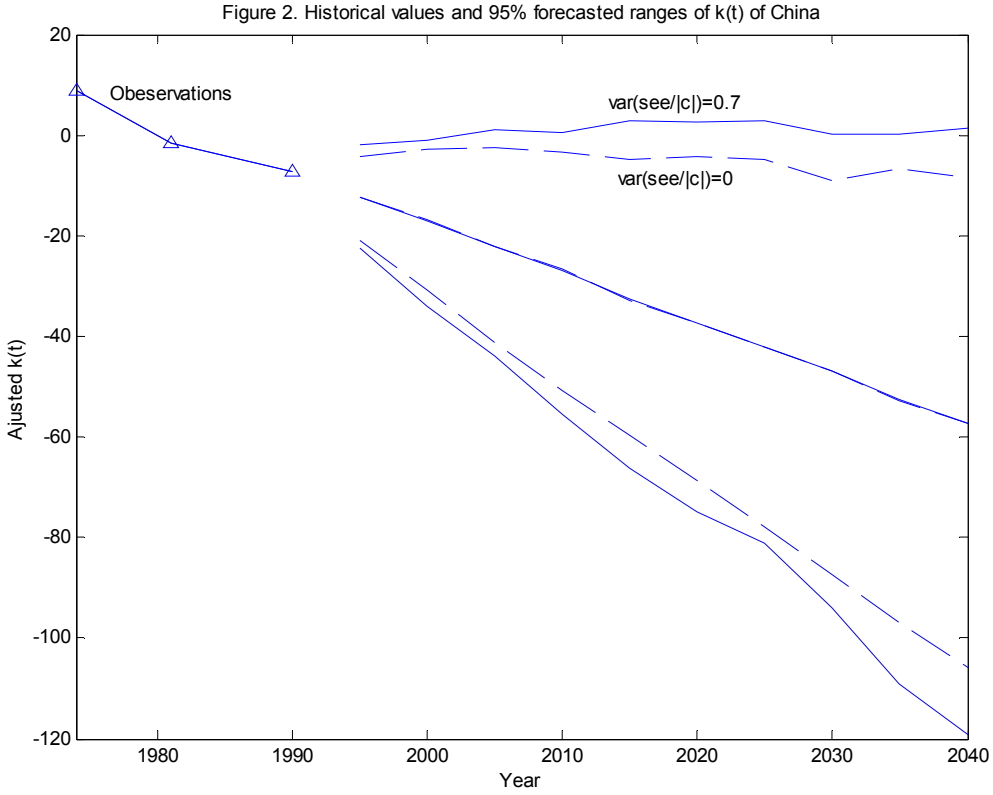

Figure 3. Historical values and 95% forecasted ranges of life expectancy of China

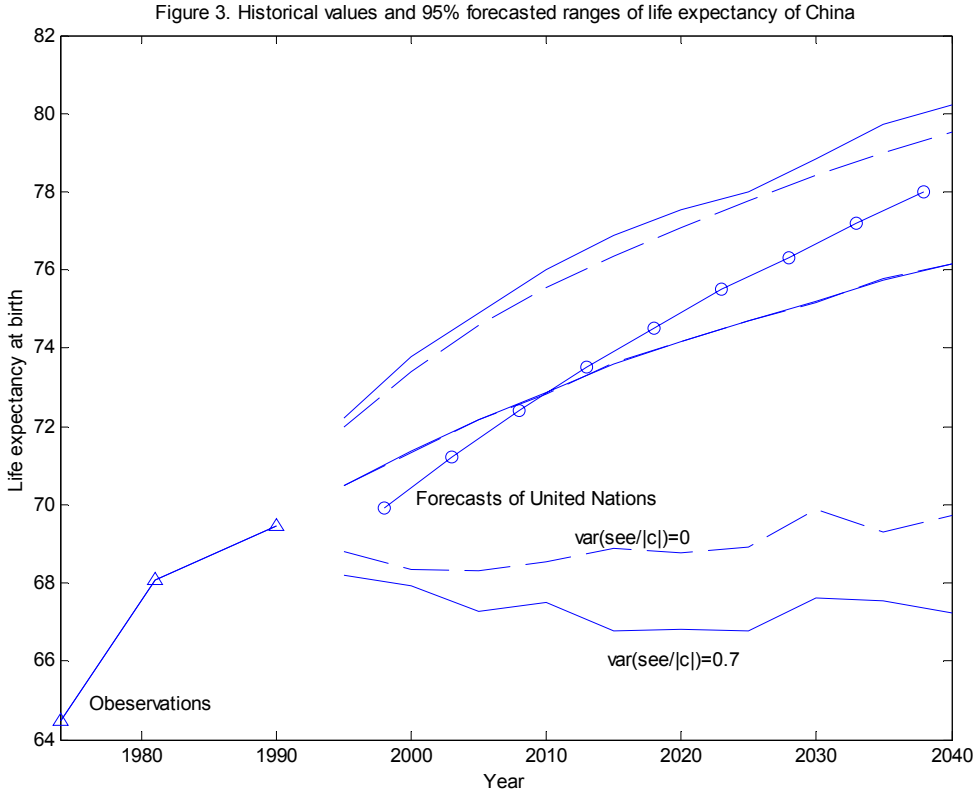

Figure 4. Historical and 95% forecasted ranges of k(t) of South Korea

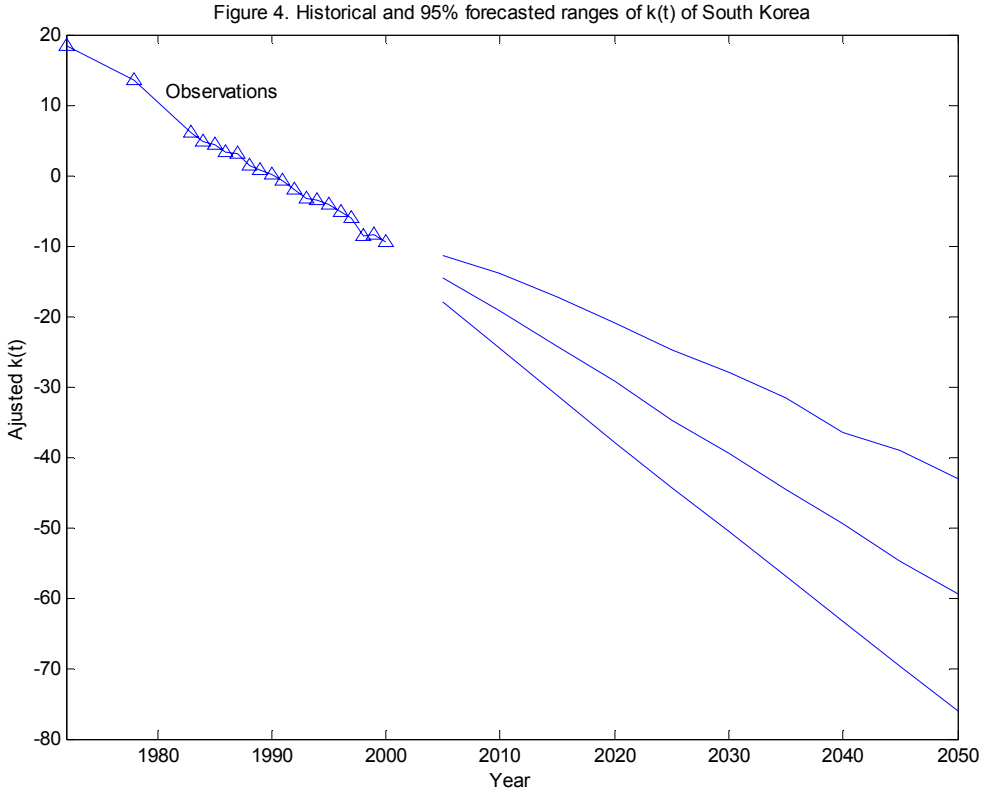

Figure 5. Historical values and 95% forecasted ranges of life expectancy of South Korea