International Statistical Review (2004), **72,** 1, 19–36, Printed in The Netherlands

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

Nan Li[1], Ronald Lee[2] and Shripad Tuljapurkar[3]

[1]Max-Planck Institute for Demographic Research, 18057 Rostock, Germany. E-mail: NLi@demogr.mpg.de. [2]Demography and Economics, University of California, Berkeley, CA 94720, USA. E-mail:rlee@demog.berkeley.edu. [3]Department of Biological Sciences, Stanford University, Stanford CA 94305-5020, USA. E-mail: tulja@stanford.edu

Summary

The Lee–Carter method for modeling and forecasting mortality has been shown to work quite well given long time series of data. Here we consider how it can be used when there are few observations at uneven intervals. Assuming that the underlying model is correct and that the mortality index follows a random walk with drift, we find the method can be used with sparse data. The central forecast depends mainly on the first and last observation, and so can be generated with just two observations, preferably not too close in time. With three data points, uncertainty can also be estimated, although such estimates of uncertainty are themselves highly uncertain and improve with additional observations. We apply the methods to China and South Korea, which have 3 and 20 data points, respectively, at uneven intervals.

Key words: Mortality forecast; Limited data; Lee–Carter method.

1 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 & Miller, 2001). United Nations' projections for European and North American countries have also under-predicted life expectancy gains (Keilman, 1998; National Research Council, 2000). 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). 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.

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

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 & 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 been 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 & 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 & 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 & Miller, 2001). The LC method has also been used to forecast mortality for some Third World countries, for example Chile (Lee & 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 cannot 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, possibly spaced at varying intervals, as in the case of China, then in principle one can estimate all four items of information and produce full forecasts with no borrowing. The new methods for doing this are the central contribution of this paper. 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.

2 The LC Method Using Data at Single-year Intervals

Any use of statistical time series methods to model and forecast a variable relies on an explicit or implicit assumption that the process generating the series is similar over the historical period, and in the future which is to be forecasted. The usual assumption is that after an appropriate process of taking differences, the resulting process is covariance stationary. The LC method also requires this assumption, and it may be necessary to choose the relevant historical period to comply with this requirement. At the same time, one should not choose the historical period in such a way as to minimize the variance in the series. Some kinds of change in the structure of the process over the time can be handled through use of special models such as GARCH (e.g., Hamilton, 1994, p.665). The LC model uses time series methods to make forecasts over a much longer time horizon than is usually the case, and consequently questions can be raised about the appropriateness of assuming structural similarity of the process over such long time horizons. The LC model also makes assumptions about similarity of demographic structure of mortality over long time periods, by treating the relative proportional rates of change of mortality by age (the $b(x)$ coefficients to be defined below) as fixed over time. Applications of the model to industrial populations have found that in fact these relative rates of change were different in the first half of the 20th century than in the second half of the century, or alternatively that they are different at higher levels of life expectancy. Nonetheless, the basic LC model appears to fit the data for many populations surprisingly well.

Let the death rate for age x at time t be $m(x, t)$, for $t = 0, 1, 2, \ldots, T$, 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

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

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 $\sum_{t=0}^{T} \sum_{x=0}^{\infty} \varepsilon^2(x, t)$. Define the explanation ratio to be $R = 1 - \sum_{t=0}^{T} \sum_{x=0}^{\infty} \varepsilon^2(x, t) / \sum_{t=0}^{T} \sum_{x=0}^{\infty} \{\log[m(x, t)] - a(x)\}^2$. 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:

$$k(t) = k(t-1) + c + e(t)\sigma, \quad e(t)\tilde{\ }N(0,1), \quad E(e(s)e(t)) = 0. \tag{2}$$

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 $[k(t) - k(t-1)]$,

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

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

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

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 $\widehat{c}$. In other words, the $\widehat{c}$ in (3) is a statistic when $k(t)$ is a stochastic process, and a certain number when sample values of $k(t)$ is used. Let the expected value of $\widehat{c}$ be the average of all its sample values. The differences between the expected and sample values of $\widehat{c}$ can be defined as errors in estimating c. Using $\widehat{\sigma}$, the standard error in estimating c is expressed as

$$\sqrt{\mathrm{var}(\widehat{c})} = \sqrt{\frac{\sigma^2}{T}} \approx \frac{\widehat{\sigma}}{\sqrt{T}}. \tag{5}$$

Because the $e(t)s$ in (2) are normally distributed, so the distribution of statistic $\widehat{c}$ is also normal with mean c and variance $\mathrm{var}(\widehat{c})$, and therefore there is a standard-normal random variable η that makes

$$\widehat{c} = c + \sqrt{\mathrm{var}(\widehat{c})}\eta. \tag{6}$$

When $\widehat{c}$ is used as an estimated value in forecasting, how to deal with the η that is a certain but unknown value? We know the probability for random variable η to take any value, and thus in forecasting we can simulate the process in which η takes all possible values. In other words, although it is impossible to know the exact value of c, we know the probability for c to be in any range by (6), and simulating η utilizes this knowledge into forecasting. Randomly choosing a small range of

η according to the corresponding probability, and a set of sample values of $e(s)$ that are independent with η for $s = (T + 1)$ to t, a trajectory of forecasted $k(t)$ for $t > T$ is obtained according to (2) and (6) as,

$$k(t) = k(T) + [\widehat{c} - \sqrt{\mathrm{var}(\widehat{c})}\eta](t - T) + \widehat{\sigma} \sum_{s=T+1}^{t} e(s). \tag{7}$$

Note that this particular trajectory for future $k(t)$ will depend partly on the estimated drift, $\widehat{c}$; partly on a randomly-simulated difference between the true c and $\widehat{c}$; 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 (7), the reason for η to be independent from $e(s)$ is, as pointed out by Lee & Carter (1992), that the η 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

$$\log[m(x, t)] = \log[m(x, T)] + b(x)[k(t) - k(T)], \tag{8}$$

and a large number of trajectories compose the stochastic forecasts of $m(x, t)$. Note that in (8), the most recently observed age-specific death rates, $m(x, T)$, are used as the baseline mortality rather than $a(x)$ 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).

3 The LC Method Using Data at Unequal Intervals

Here we begin our discussion of applying the LC method in the case of limited data, possibly with as few as three observations separated by unequal intervals. Obviously, standard statistical time series analysis cannot be used in this case. However, if we make the strong assumption that $k(t)$ follows a random walk with drift, then we can use as few as three observed $m(x, t)$ schedules to find all the parameters of the LC model. The trick is that because of the assumption we do not need to figure out the appropriate model, which would require much more data. Since the condition for the $k(t)$ in the LC model to be a random walk with drift is that mortality declines stably, which has already been observed for many countries, both developed and Third World, the strong assumption is defensible.

Now let mortality data be collected at times $u(0), u(1), \ldots, u(T)$. 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 $k(u(0)), k(u(1)), \ldots, k(u(T))$ are obtained.

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

$$k(u(t)) - k(u(t - 1)) = c[u(t) - u(t - 1)] + \sigma[e(u(t - 1) + 1) + \cdots + e(u(t))]. \tag{9}$$

Thus, for different t, $[k(u(t)) - k(u(t - 1))]$ are no longer identically distributed. Consequently, estimating c and σ from (9) cannot be as simple as that for i.i.d. variables.

Because the means of the second term in the right-hand side of (9) are still zero, the unbiased estimate of c is obtained as:

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

Since the variances of the second term in the right-hand side of (9) are no longer the same for different t, the derivation of the standard error of $e(u(t))$, $\widehat{\sigma}$, becomes somewhat complicated, and is derived in the appendix as

$$
\begin{aligned}
\widehat{\sigma}^2 &= \frac{\sum_{t=1}^{T}\{k(u(t)) - k(u(t-1)) - c[u(t) - u(t-1)]\}^2}{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}} \\
&\approx \frac{\sum_{t=1}^{T}\{k(u(t)) - k(u(t-1)) - \widehat{c}[u(t) - u(t-1)]\}^2}{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}}.
\end{aligned}
\tag{11}
$$

An illustration of (10) and (11) is given in Figure 1, with three observations of $k(t)$ shown as circles at times $u(0)$, $u(1)$ and $u(2)$. The slope of the straight line connecting the first and last values of $k(t)$ is the $\widehat{c}$ given by (9). The intermediate observation at time $u(1)$ is shown here as above this straight line. Its vertical displacement above the line is the sum of errors, i.e., the sum of realizations of $e(t)\sigma$ in the second term of the right hand side of (9), between times $u(0)$ and $u(1)$. The end of the upper dashed line shows the expected value of $k(t)$ at time $u(2)$, given the observed $k(t)$ at $u(1)$. The vertical distance between this expectation and the $k(t)$ observed at $u(2)$ is the sum of errors between times $u(1)$ and $u(2)$. Equation (11) describes how to calculate the strength of the errors, $\widehat{\sigma}$, according to these observed sum of errors.

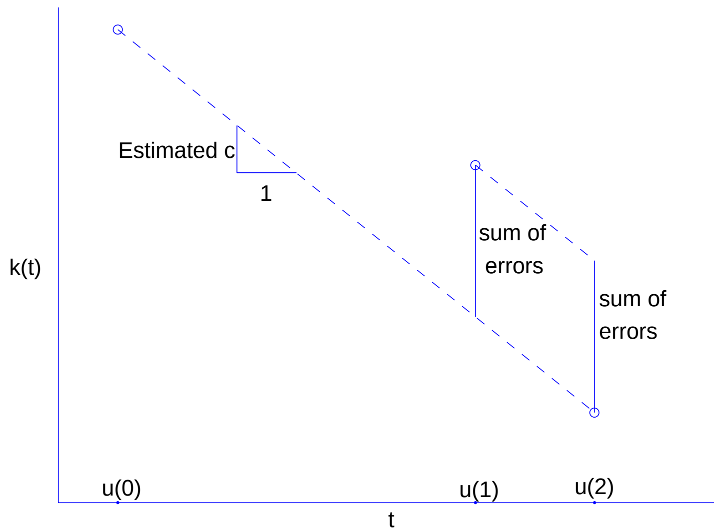

Figure 1. *Linear trend and errors.*

The standard error in estimating c, $\sqrt{\mathrm{var}(\widehat{c})}$, is obtained from (10) and (11) as

$$
\sqrt{\mathrm{var}(\widehat{c})} = \sqrt{\frac{\mathrm{var}\{\sum_{t=1}^{T}[e(u(t-1)+1) + \cdots + e(u(t))]\}}{[u(T) - u(0)]^2}} = \sqrt{\frac{\sigma^2}{u(T) - u(0)}} \approx \frac{\widehat{\sigma}}{\sqrt{u(T) - u(0)}}. \tag{12}
$$

When $[u(t) - u(t-1)] = 1$, (10), (11) and (12) reduce to (3), (4) and (5), respectively. Having the values of $\widehat{c}$ and $\widehat{\sigma}$, forecasting is carried out by (7) and (8), 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?

3.1 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, $\widehat{c}$ 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 $\widehat{c}$ 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 (10). 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 $\widehat{c}$, which is the slope of the line that connects the positions of $k(t)$ at 1974 and 1990. What happens in between, and how often it is observed, does not matter.

Second, (12) 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 (10), a given random disturbance in $k(u(T))$ or $k(u(0))$ will make smaller difference for $\widehat{c}$, when the denominator, $[u(T)-u(0)]$, is larger. In the example of China, if $\widehat{c}$ were not close enough to c, the reason would be that the period of 16 years is not long enough, not that the 3 time points are too few.

Turning to mean forecasts of $m(x,t)$, (8) shows that $a(x)$ can be omitted altogether in forecasting, and that $\text{mean}\{\log[m(x,t)] - \log[m(x,T)]\} = \text{mean}\{b(x)[k(t)-k(T)]\}$. 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.

3.2 The Probability Intervals for Forecasts Based on Data at Few Time Points—How Accurate Can They Be?

The probability intervals for $k(t)$, such as the 95% probability interval of $k(t)$ at different times, are based on $\widehat{\sigma}$ in (11), which captures historical random fluctuations in $k(t)$. To obtain positive $\widehat{\sigma}$ from (11), 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 $\widehat{\sigma}$ measures random deviation from the linear component of $k(t)$, its estimation error, measured by var($\widehat{\sigma}$), should depend also on the number of these fluctuations or the number of time points. Using the sampled value of $\widehat{\sigma}$, which is the unbiased estimate of σ, var($\widehat{\sigma}$) is described by

(14a) in appendix as

$$\sqrt{\operatorname{var}(\widehat{\sigma})} \approx \sqrt{\frac{1}{2\{u(T)-u(0)-\frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}\}}}\,\widehat{\sigma}. \tag{13}$$

Define the relative error in estimating σ, or the relative error of $\widehat{\sigma}$, as $\operatorname{re}(\widehat{\sigma}) = \sqrt{\operatorname{var}(\widehat{\sigma})}/\widehat{\sigma}$. Then for given the number of time points, the wider the span $(u(T) - u(0))$, the smaller the $\operatorname{re}(\widehat{\sigma})$. Fixing the span, $\operatorname{re}(\widehat{\sigma})$ decreases with increasing the number of time points. Given the span and the number of time points, $\operatorname{re}(\widehat{\sigma})$ reaches minimum when sampling in equal intervals. For the shortest span of three years, $\operatorname{re}(\widehat{\sigma})$ is as high as $1/\sqrt{2} \approx 0.707$.

Similar to the reason of (6), for statistic $\widehat{\sigma}$ there is a standard normal variable θ that makes

$$\sigma = \widehat{\sigma} - \sqrt{\operatorname{var}(\widehat{\sigma})}\theta = \widehat{\sigma}[1 - \operatorname{re}(\widehat{\sigma})\theta], \tag{14}$$

In forecasting, errors from estimating σ can be compensated, by using the σ in (14) to substitute $\widehat{\sigma}$ in (12) and (7) as

$$K(t,\theta) = k(T) + [\widehat{c} - \widehat{\sigma}\frac{1-\operatorname{re}(\widehat{\sigma})\theta}{\sqrt{u(T)-u(0)}}\eta](t-T) + \widehat{\sigma}[1-\operatorname{re}(\widehat{\sigma})\theta]\sum_{s=T+1}^{t} e(s). \tag{15}$$

How to include the full content of (15) in forecasting is an issue to be explored. In this paper our $K(t,\theta)$ is forecasted as an ordinary stochastic process, whose mean and variance must take certain value at any time. In (7), η may take different values in different forecast trajectories, since its effect is to make trajectories depart from the estimated center trend randomly, and hence can be described as a certain amount of uncertainty and then included in forecast uncertainty. In (15), however, θ cannot take different values in different forecast trajectories. This is because that allowing θ to take different values would make the variance of $K(t,\theta)$ uncertain, beyond the range of ordinary stochastic process. In other words, the effect of using different values of θ is to make forecast uncertainty stronger or weaker randomly, and such an effect cannot be described as a certain amount of additional uncertainty to be absorbed in forecast uncertainty. For this reason, we use $K(t,\theta)$ to distinguish (15) from (7), to indicate that $K(t,\theta)$ provides a family of statistic forecast, in which each member is a $k(t)$ that is produced by (15) using a specific value of θ with corresponding probability.

The θ should be independent with η, because $\widehat{c}$ is estimated without bias so that mean$[\widehat{c} - \widehat{\sigma}\frac{1-\operatorname{re}(\widehat{\sigma})\theta}{\sqrt{u(T)-u(0)}}\eta] = c$. Since $\sum_{s=T+1}^{t} e(s)$ describes random changes in the future, while θ reflects estimating errors in using historical data, they should also be independent.

Because the mean of $K(t,\theta)$ with respect to η and $\sum_{s=T+1}^{t} e(s)$ is $k(T) + c(t-T)$, independent of θ, indicating that errors in estimating uncertainty does not affect the mean forecast of $k(t)$.

Taking $\theta = 0$, at which the probability density function of θ reaches maximum, (15) reduces to (7) and produces the most possible uncertainty forecasts of $k(t)$. Since $\widehat{\sigma}$ is the unbiased estimate of σ, the most possible uncertainty forecast is also unbiased. To reach the most possible and unbiased forecast, the forecasted 95% probability intervals of $k(t)$ should be given by (15) using $\theta = 0$, or simply by (7).

Errors in estimating uncertainty would lead to uncertainty of uncertainty, which can be described by the 95% wide and narrow bounds of the forecasted 95% probability intervals. Using $\theta = 1.96$, the 95% probability intervals yielded by (15) can be called wide bounds of the forecasted 95% probability intervals, because they are wider than that of using $\theta = 0$. Similarly, taking $\theta = -1.96$, (15) produces narrow bounds of the forecasted 95% probability intervals. Because the chance of having a 95% probability interval that is wider than the wide bound or narrower than the narrow bound is 5%, the range between wide and narrow bounds covers 95% of all possible 95% probability intervals. For this reason, the wide and narrow bounds obtained from using $\theta = \pm 1.96$ in (15) can be called the 95% wide and narrow bounds of the forecasted 95% probability intervals.

Turning to the uncertainty forecasts of $m(x, t)$, we show in the appendix C 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 $k(t)$. Using the 95% wide and narrow bounds of $k(t)$, (8) provides 95% wide and narrow bounds of $m(x, t)$ and of life expectancies.

The answer to the other part of the second question is, therefore, that the LC method can provide uncertainty forecasts for countries with limited data. However, with only a few years of data, the uncertainty forecasts would be remarkably uncertain. Of course, when plentiful data are available, the uncertainty of uncertainty would be negligible, because re($\widehat{\sigma}$) approaches zero when $u(T) - u(0) = T$ and T increases.

4 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. 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. 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 a 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 health-care 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.

Without considering errors in estimating $\widehat{\sigma}$, the unbiased uncertainty forecasts, expressed as 95% probability intervals for $k(t)$ and life expectancy, are shown by the solid curves in Figures 2 and 3, respectively. Because the 95% probability interval for life expectancy covers more than 10 years at 2040, the uncertainty is strong. Considering the recent changes in the health-care system of China, this high uncertainty is not surprising. The value of $\widehat{\sigma}$, however, is estimated from data at only three time points and hence may not be close enough to its expected value. The relative estimating error, re($\widehat{\sigma}$), is about 0.252, which is quite high. Taking this estimation error into account, the resulting 95% wide and narrow bounds of 95% probability intervals for $k(t)$ and life expectancy are plotted in Figures 2 and 3, by dashed and dash-dot curves respectively. To different readers, this may or may not be too uncertain, but these intervals are better than the high-low ranges, which have no probabilistic interpretation.

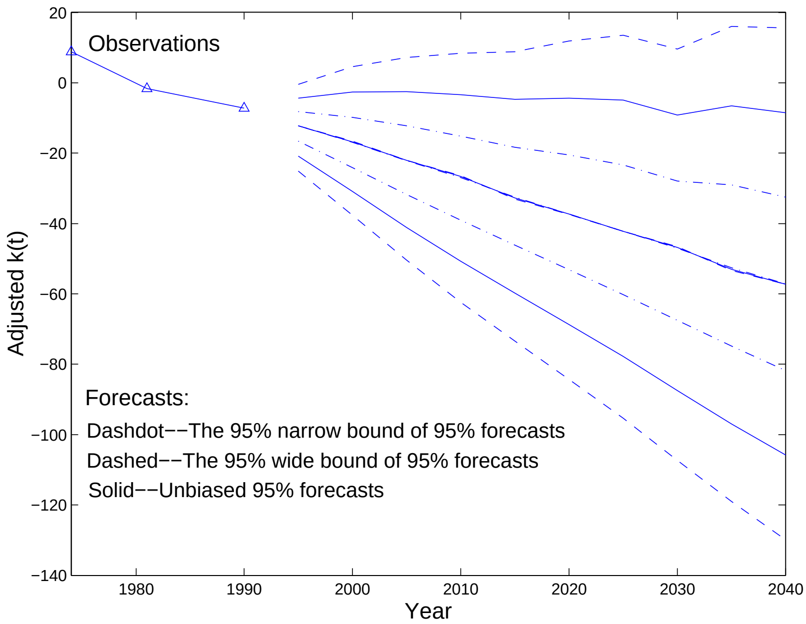

Figure 2. *Observations and 95% forecasts of $k(t)$ of China.*

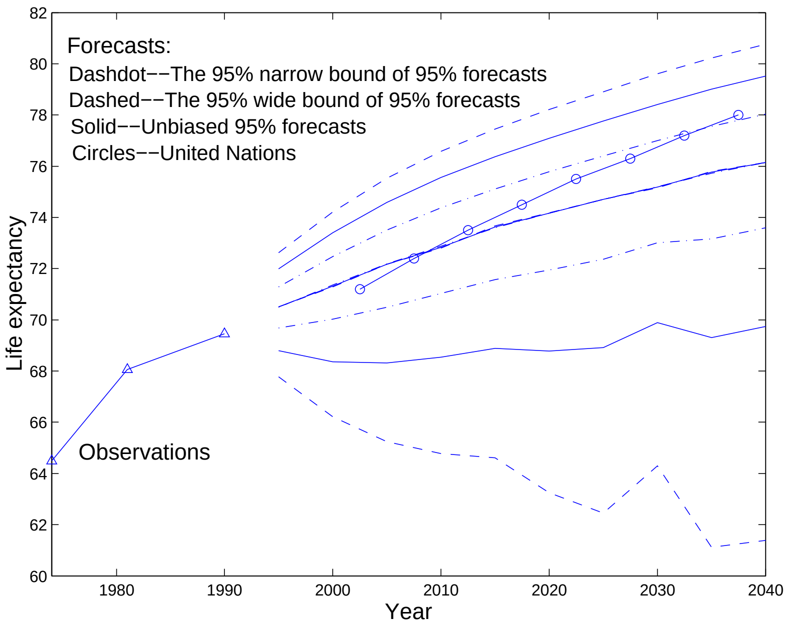

Figure 3. *Observations and 95% forecasts of life expectancy of China.*

5 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 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. 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 period that contains annual data lasts for18 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 (1989) 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 $\widehat{\sigma}$, measures the amount of uncertainty around the linear historical trajectory. Assuming that the random disturbances in the future will resemble those in the past, the random walk model derives the unbiased uncertainty forecasts for $k(t)$, as described by the 95% probability intervals and plotted in solid curves in Figure 4. The forecasts of $k(t)$ simply extrapolate the historical mean trend and uncertainty into the future, without subjective judgments. Because mortality data are available at more time points and in longer period than that of China, the $\text{re}(\widehat{\sigma})$ is about 0.139, much smaller than that of China. As a result, the uncertainty of uncertainty forecast is much less than that of China, as can be seen in Figure 4.

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 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

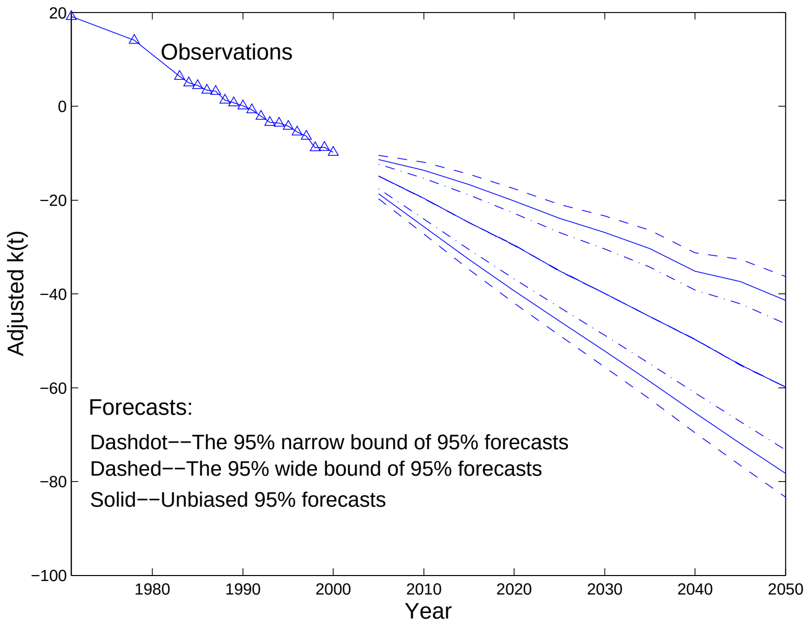

Figure 4. *Observations and 95% forecasts of $k(t)$ of South Korea.*

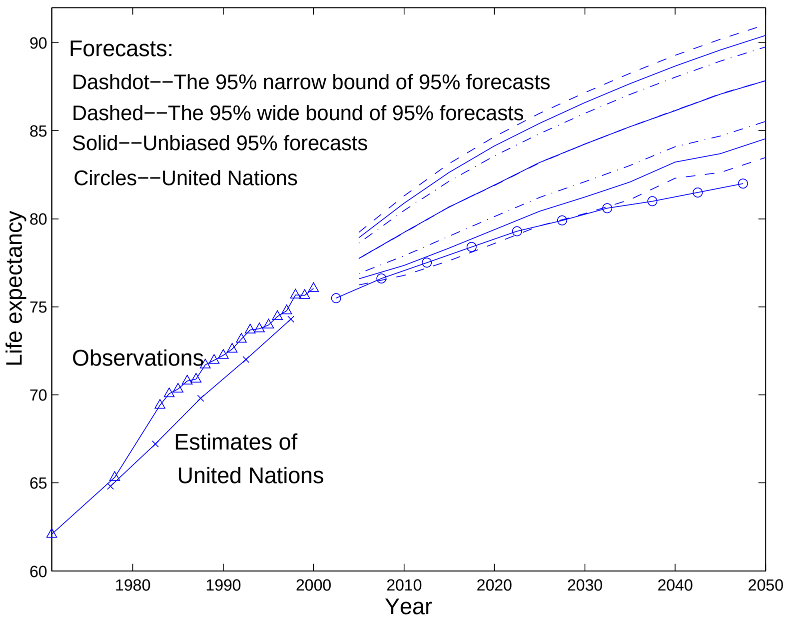

Figure 5. *Observations and 95% forecasts of life expectancy of South Korea.*

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 & 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.

6 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.

References

Agresti, A. & Finlay, B. (1997). *Statistical Methods for the Social Sciences*. New Jersey: Prentice Hall, Inc.

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

Coale, A. & Guo, G. (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*. London: Sage Publications.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.

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

Keilman, N. (1998). How Accurate Are the United Nations World Population Projections? *Population and Development Review*, **24**, 15–41.

Korea National Statistical Office (2002). http : //www.nso.kr/eng/

Lee, R. (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. & Rofman, R. (1994). Modelacion y Proyeccion de la Mortalidad en Chile. In *NOTAS de Poblacion*, **XXII**, No. 59 (Junio), pp.183–213.

Lee, R.D. & Carter, L. (1992). Modeling and Forecasting the Time Series of U.S. Mortality. *J. Amer. Statist. Assoc.*, **87**, 659–671.

Lee, R.D. & Miller, T. (2001). Evaluating the Performance of the Lee–Carter Method for Forecasting Mortality. *Demography*, **38**, 537–549.

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

Oeppen, J. & Vaupel, J.W. (2002). Broken limits to life expectancy. *Science*, **296**, 1029–1031.

Tuljapurkar, S., Li, N. & Boe, C. (2000). A Universal Pattern of Mortality Change in the G7 Countries. *Nature*, **405**, 789–792.

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

Yearbook of Chinese Population (1985). Edited by the Statistics Bureau of China, Chinese Statistics Publisher.

Résumé

La méthode Lee–Carter de modélisation et de prévision de la mortalité a prouvé son bon fonctionnement avec des séries de données existant sur une longue période. Nous envisageons ici son utilisation lorsqu'on ne dispose que de quelques observations à intervalles irréguliers. En supposant que le modèle sous-jacent est correct et que l'indice de mortalité suit une marche aléatoire avec dérive, nous trouvons que cette méthode peut être utilisée avec des données éparses. La prévision centrale dépend alors principalement de la première et de la dernière observation. Elle peut donc être générée à partir de deux observations seulement, de préférence pas trop proches dans le temps. Avec trois points, on peut aussi estimer l'aléa, bienqu'un tel estimateur de l'aléa soit lui-même très aléatoire. Il s'améliore cependant lorsqu'on dispose d'observations supplémentaires. Nous appliquons notre méthode à la Chine et à la Corée du Sud, pour lesquelles nous avons respectivement 3 et 20 points à intervalles irréguliers.

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 (10). 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 (9) is $[e(1) + e(2) + \cdots + e(m)]$, and the η that covers in the whole historical period includes $e(1), e(2), \ldots, e(n)$, there is

$$
\begin{aligned}
&E\{[k(u(1)) - k(u(0)) - \frac{\sum_{t=1}^{T}[k(u(t)) - k(u(t-1))]}{\sum_{t=1}^{T}[u(t) - u(t-1)]}[u(1) - u(0)]]^2\} \\
&= E\{[k(u(1)) - k(u(0)) - c[u(1) - u(0)] - \\
&\qquad\qquad (\frac{\sum_{t=1}^{T}[k(u(t)) - k(u(t-1))]}{\sum_{t=1}^{T}[u(t) - u(t-1)]} - c)[u(1) - u(0)]]^2\} \\
&= E\{[\sigma \sum_{i=1}^{m} e(i) - [\frac{\sum_{t=1}^{T}[k(u(t)) - k(u(t-1))]}{\sum_{t=1}^{T}[u(t) - u(t-1)]} - c][u(1) - u(0)]]^2\} \qquad (1a) \\
&= E\{[\sigma \sum_{i=1}^{m} e(i) - \frac{\sum_{t=1}^{T}[k(u(t)) - k(u(t-1))] - c[u(t) - u(t-1)]}{\sum_{t=1}^{T}[u(t) - u(t-1)]}[u(1) - u(0)]]^2\} \\
&= E\{[\sigma \sum_{i=1}^{m} e(i) - \frac{m\sigma \sum_{i=1}^{n} e(i)}{n}]^2\} \\
&= \frac{\sigma^2}{n^2} E\{[(n-m)(e(1) + \cdots + e(m)) - m(e(m+1) + \cdots + e(n))]^2\}.
\end{aligned}
$$

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

$$
\begin{aligned}
&E\{[k(u(1)) - k(u(0)) - c[u(1) - u(0)]]^2\} \\
&= \frac{\sigma^2}{n^2} E\{[(n-m)^2(e^2(1) + \cdots + e^2(m)) + m^2(e^2(m+1) + \cdots + e^2(n))]\} \\
&= (\frac{n-m}{n})^2 m\sigma^2 + (\frac{m}{n})^2 (n-m)\sigma^2 \qquad (2a) \\
&= (\frac{n-m}{n}) m\sigma^2 \\
&= [1 - \frac{u(1) - u(0)}{u(T) - u(0)}][u(1) - u(0)]\sigma^2.
\end{aligned}
$$

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

$$E\{[k(u(t)) - k(u(t-1)) - c[u(t) - u(t-1)]]^2\} = [1 - \frac{u(t) - u(t-1)}{u(T) - u(0)}][u(t) - u(t-1)]\sigma^2. \quad (3a)$$

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

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

Therefore

$$\widehat{\sigma}^2 = \frac{\sum_{t=1}^{T}\{k(u(t)) - k(u(t-1)) - c[u(t) - u(t-1)]\}^2}{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}}, \quad (5a)$$

is the unbiased estimate of σ^2.

B. Errors in estimating $\widehat{\sigma}$

Let

$$v(t) = \{k(u(t)) - k(u(t-1)) - \widehat{c}[u(t) - u(t-1)]\}^2. \quad (6a)$$

From (3a), $v(t)$ includes $[1 - \frac{u(t)-u(t-1)}{u(T)-u(0)}][u(t) - u(t-1)]$ squared i.i.d. variables that are assumed normal with mean zero and variance σ^2. Thus, $v(t)/\sigma^2$ obeys the Chi-square distribution with the degree of freedom $[1 - \frac{u(t)-u(t-1)}{u(T)-u(0)}][u(t) - u(t-1)]$:

$$\frac{v(t)}{\sigma^2} \tilde{} \chi^2([1 - \frac{u(t) - u(t-1)}{u(T) - u(0)}][u(t) - u(t-1)]). \quad (7a)$$

Therefore,

$$\sum_{t=1}^{T}\frac{v(t)}{\sigma^2} \tilde{} \chi^2(\sum_{t=1}^{T}[1 - \frac{u(t) - u(t-1)}{u(T) - u(0)}][u(t) - u(t-1)]) = \chi^2(u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t) - u(t-1)]^2}{u(T) - u(0)}), \quad (8a)$$

and

$$\mathrm{var}[\sum_{t=1}^{T}\frac{v(t)}{\sigma^2}] = 2\{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t) - u(t-1)]^2}{u(T) - u(0)}\}. \quad (9a)$$

According to (5a) and (9a), the variance of statistic $\widehat{\sigma}^2$ is obtained as

$$\mathrm{var}(\widehat{\sigma}^2) = \mathrm{var}[\frac{\sigma^2}{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}}\sum_{t=1}^{T}\frac{v(t)}{\sigma^2}] = \frac{2\sigma^4}{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}}. \quad (10a)$$

Thus

$$\sqrt{\mathrm{var}(\widehat{\sigma}^2)} = \sqrt{\frac{2}{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}}}\sigma^2. \quad (11a)$$

On the other hand, statistic $\widehat{\sigma}$ can be assumed normally distributed approximately

$$\widehat{\sigma} \approx \sigma + \sqrt{\operatorname{var}(\widehat{\sigma})}e, e\tilde{\ }N(0,1). \tag{12a}$$

Because that $\widehat{\sigma}$ can also be written as

$$\widehat{\sigma} = \sqrt{\widehat{\sigma}^2} \approx [\sigma^2 + \sqrt{\operatorname{var}(\widehat{\sigma}^2)}e] \approx \sigma + \frac{\sqrt{\operatorname{var}(\widehat{\sigma}^2)}}{2\sigma}e, \tag{13a}$$

the standard error of $\widehat{\sigma}$ is obtained from (11a)–(13a)

$$\sqrt{\operatorname{var}(\widehat{\sigma})} = \sqrt{\frac{1}{2[u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}]}}\sigma. \tag{14a}$$

Because (13a) describes the distribution of a positive variable $(\widehat{\sigma})$ as normal, it is not exact and will make (14a) approximate. To examine the accuracy of (13a), we choose $\sigma = 1$ and randomly select 1000 values of $e(t)$ as the sample values of $(k(t) - k(t-1) - \widehat{c})$ to simulate 1000 values of $\widehat{\sigma}$ using (4), for each case of using data in 2 to 100 single-year intervals. Meanwhile, we analytically calculate the distribution of $\widehat{\sigma}$ using (12a) where $\sqrt{\operatorname{var}(\widehat{\sigma})}$ is given by (14a). Comparing values $\widehat{\sigma}$ from simulating and analytical calculation, Figure 6 indicates that (13a) is accurate. Substituting σ by $\widehat{\sigma}$ will make (14a) further approximate, but derives the relative error of $\widehat{\sigma}$, $\operatorname{re}(\widehat{\sigma}) = \sqrt{\operatorname{var}(\widehat{\sigma})}/\widehat{\sigma}$. Comparing to simulated values of $\operatorname{re}(\widehat{\sigma})$ that are obtained from simulated $\widehat{\sigma}$, Figure 7 shows the analytical calculated $\operatorname{re}(\widehat{\sigma})$ given by (14a) where σ is substituted by $\widehat{\sigma}$ is also accurate.

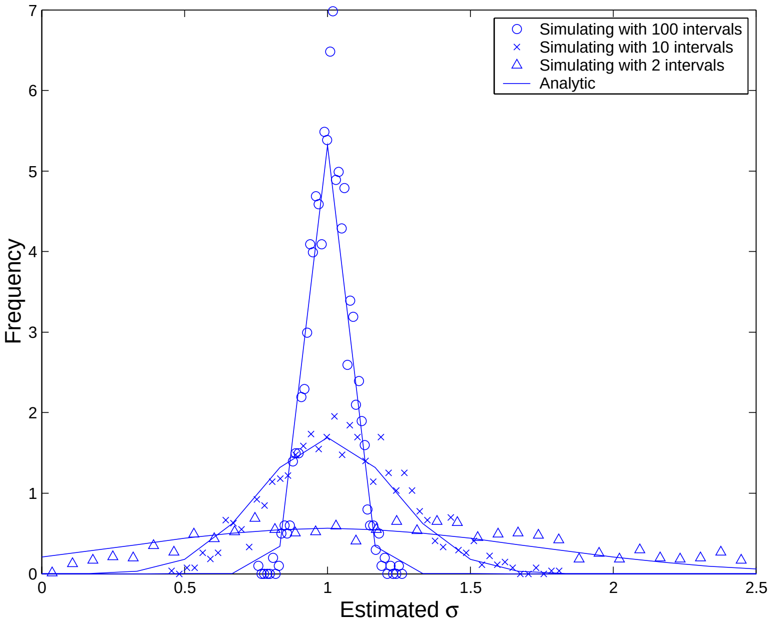

Figure 6. *Simulated and analytical distributions of estimated σ.*

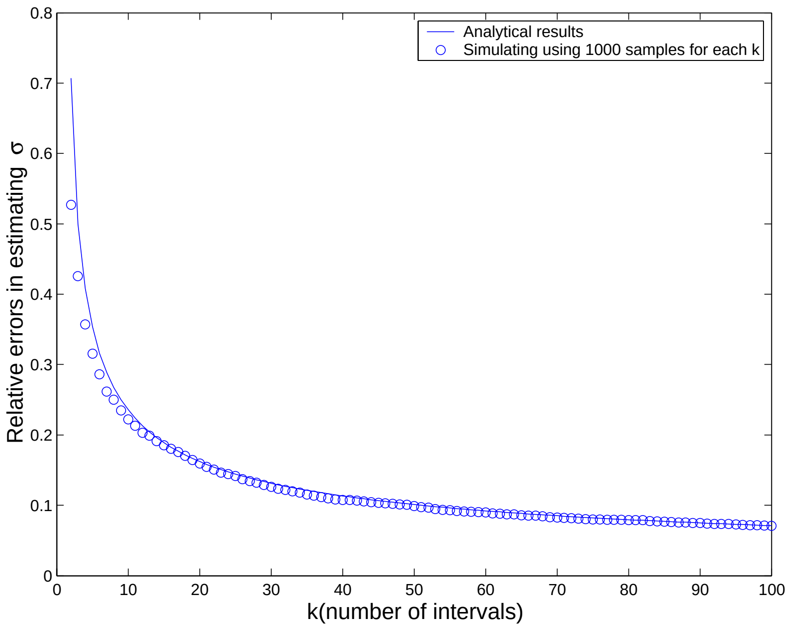

Figure 7. *Relative errors in estimating σ.*

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

In order to discuss errors in estimating $a(x)$ and $b(x)$, their expected values must be defined. Viewing the values of $m(x,t)$ as 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, η 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 $\sum_{t=1}^{T}\sum_{x=0}^{\infty}\varepsilon^2(x,t)$, 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 & Finlay, 1997, pp.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,

$$\log[m(x,u(t))] = a(x) + b(x)k(u(t)) + \varepsilon(x,u(t)). \tag{15a}$$

In (15a), 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, p.115) as

$$\mathrm{var}(a(x)) = \frac{\sigma_\varepsilon^2(x)}{u(T) - u(0)}, \tag{16a}$$

$$\mathrm{var}(b(x)) = \frac{\sigma_\varepsilon^2(x)}{\sum_{t=0}^{T} k^2(u(t))}, \tag{17a}$$

$$\sigma_\varepsilon^2(x) \approx \frac{\sum_{t=0}^{T} \varepsilon^2(u(t))}{T}. \tag{18a}$$

Equations (16a)–(18a) show that $\mathrm{var}(a(x))$ and $\mathrm{var}(b(x))$ 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 a 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.

[*Received December 2002, accepted November 2003*]