International Journal of Forecasting 8 (1992) 315–327
North-Holland

Stochastic demographic forecasting

Ronald D. Lee

Department of Demography, University of California, 2232 Piedmont Avenue, Berkeley, CA 94720, USA

Abstract: This paper describes a particular approach to stochastic population forecasting, which is implemented for the USA through 2065. Statistical time series methods are combined with demographic models to produce plausible long run forecasts of vital rates, with probability distributions. The resulting mortality forecasts imply gains in future life expectancy that are roughly twice as large as those forecast by the Office of the Social Security Actuary. The method for forecasting fertility is useful mainly for providing estimates of the variance and autocorrelation in fertility forecast errors, and not for making stand-alone point forecasts. The forecasts of the probability distributions of the vital rates can then be used in a stochastic Leslie matrix to produce stochastic population forecasts. These provide probability distributions for the quantities forecast, reflecting lower bound estimates of uncertainty. Resulting stochastic forecasts of the elderly population, elderly dependency ratios, and payroll tax rates for health, education and pensions are presented.

Keywords: Population, Forecasting, Aging, Dependency, Mortality, Fertility, Projection, Social Security

1. Introduction

Demographic forecasts are used for many purposes. Sometimes the main interest is in the size of the future population, but more often it is in some component of the total, such as children or the elderly, or some function of the age distribution, such as a dependency ratio. Forecasts may have a powerful influence on policy. In the United States, forecasts of a rising elderly dependency ratio led to legislated increases in the pay roll tax rate for Social Security during the 1980s and continuing in the future, and scheduled postponements of the age of retirement. These demographically inspired increases have in turn shaped the level and incidence of federal taxes, permitting an offsetting reduction in income tax rates and a reduction in the progressivity of overall federal taxes.

Correspondence to: R.D. Lee, Department of Demography, University of California, 2232 Piedmont Avenue, Berkeley, CA 94720, USA.

The power and influence of long-run demographic forecasts derives from their reassuringly concrete aspect. A payroll tax policy based on the predictable retirement of the baby boom generations surely seems well founded and prudent. Yet even for this highly predictable event, much uncertainty remains. What proportion of the baby boomers will survive to retire, and how long will they live after retirement? How many children will be born in the coming decades to support them as workers in their retirement? Demographic forecasters attempt to indicate the uncertainty of their forecasts, but these attempts, leading to high, medium and low brackets on the forecasts, are often seen ex post to have been unsuccessful.

A number of approaches have been taken to the problem of systematically assessing the uncertainty of population forecasts. These include using time series methods to forecast the population growth rate [Cohen (1986)], stochastic simulations of population growth with randomly varying vital rates [Pflaumer (1988)], analysis of errors in past forecasts [Keyfitz (1981), Stoto (1983), US Bureau of the Census (1989)], and the development of stochastic models for the vital rates, which are then used in stochastic Leslie matrices to generate probability distributions for the future population [Sykes (1969), and Alho and Spencer (1985)].

This paper describes an approach taken to demographic forecasting in recent research by a team consisting of Carter, Lee, and Tuljapurkar. The research continues the development of the stochastic Leslie matrix approach mentioned above, combining statistical time series analysis with mathematical and statistical demography. There is particular emphasis on developing time series models for the vital rates which are suitable for long-run forecasting, with horizons of 75 years or so. The analysis of the propagation of error through the Leslie matrix, and the derivation of probability intervals for the forecasts are distinctive in not relying on linear approximations, and not using recursions, which are shown to be incorrect in the presence of autocorrelation [see Tuljapurkar (1992)].

An overview of the traditional approach to demographic forecasting and its limitations provides a convenient point of departure.

2. The traditional approach

Consider the way in which uncertainty is handled by traditional projections. Typically, high, medium and low trajectories are specified for each input (fertility, mortality and net migration). Many methods are employed to choose the medium trajectories (surveys of individual intentions, consultation of experts, statistical modeling, consideration of social and economic factors, and so on), but the high and low variants are usually chosen less systematically. Then an overall high projection of the population is made by taking together all the high trajectories, and a low projection by taking together all the low trajectories. While usually no probability interpretation is given to the resulting high–low interval, the user is certainty invited to believe that in some sense the intervals are likely to contain the actual future values.

Keyfitz (1981) and Stoto (1983) have analyzed the performance of past forecasts by the United Nations and the US Census Bureau, and concluded that the projected growth rate from the base year of the forecast up to T years later has a probability distribution which is largely independent of T.[1] On the basis of this work, it is possible to attach probability bounds to forecasts of population size, but not to other aspects of the forecast, such as age bands or ratios. This work, which treats the projection method as a black box, has proved very useful. The approach suggested here, however, takes the opposite tack of formally modeling the errors arising from some but not all sources, and evaluating their effects on the forecasted variables of interest.

The complexities of this more formal approach can be better appreciated if we reexamine the logic of the traditional high–low bounds, taking fertility as an example. Consider the number of births 40 years in the future. It will depend on the level of fertility in 40 years, and also on the number of reproductive-age women, which will in turn depend on the level of fertility for the first 25 years of the forecast. The high scenario assumes not only that fertility will be at the 'high' level in the fortieth year, which let us assume has probability p, but also that it was high *every year* in which the reproductive-age women were born. The probability of this composite outcome must be much less than p (it would be $p * p^{25}$ if fertility were independent from year to year), because annual variations in fertility tend to offset one another over time, resulting in a narrower interval covering probability p for composite outcomes such as births in 40 years. The degree to which the interval is narrower will depend on the extent of autocorrelation in the forecast errors. The confidence band on fertility by itself does not provide the information we need to assess the uncertainty of population projections; we also need to know the autocorrelation.

Now consider the uncertainty in projecting the number of children under the age of 15 in 60 years. This involves a weighted sum of the number of births from 45 to 60 years in the future, with the weights equal to future survival probabilities. Ignoring the uncertainty in these weights, there should again be cancellation of errors in forecasting the births for each single year – 45 years ahead, 46, 47 and so on. The

[1] Analysis by the US Bureau of the Census (1989) questions this conclusion, however.

band with probability p of containing this age group size should be narrower still relative to the high and low range on fertility itself. But now we must take into account not only the autocorrelation in errors projecting fertility, but also the autocorrelation in errors in projecting births. Since there is very substantial overlap in the groups of reproductive-age women from one year to the next, a great deal of autocorrelation is automatically introduced by the mechanics of population renewal. Variation in mortality adds further complexity, as does migration. This discussion indicates the inconsistencies that arise when we try to give a probabilistic interpretation to the high–low intervals of the traditional method.

The problems become still more acute when we consider the uncertainty in forecasting an age group ratio such as the old age dependency ratio: the population age 65 and over divided by the population 20–64. The high forecast will combine a high denominator (resulting from high fertility) with a high numerator (resulting from low mortality), and the low forecast will pair a low denominator with a low numerator. The range generated in this way obviously will be much too narrow. In practice, the Social Security forecasts combine low fertility and low mortality in their 'pessimistic' scenario, and high–high in their 'optimistic' scenario, and the census projections suggest that a corresponding scenario be used when considering pension problems using their forecasts [Wade (1989)]. In addition to the difficulty in assigning probabilities to these scenarios, it is important to note that no consistently high or consistently low trajectory can cover the likelihood that fluctuations in vital rates will result in irregular age distributions.

There have been many attempts to model and forecast births or fertility as time series, and several that treat mortality in this way [see, for example, Lee (1974b), Saboia (1977), McDonald (1981), Alho and Spencer (1985, 1990), Bell and Monsell (1991), Carter and Lee (1986), Bozik and Bell (1987, 1989), Gomez (1989), McNown and Rogers (1989, 1992), Alho (1990), and others reviewed in Land (1986)]. For evaluations of this work, see Lee (1981), Land (1986), Lee (1991b) and Lee and Carter (1992). Suppose that an effort of this sort has been successfully carried out, resulting in a forecast of fertility, with a corresponding probability interval. How is this information to be used in preparing a population forecast? The medium forecast could be prepared on the basis of the expected value (point forecast) of fertility in each year. The result would *not* be the trajectory of expected values of population size, owing to the multiplicative way in which fertility enters the projection process [Tuljapurkar (1992)], but it would probably be reasonably close except for very long-run forecasts. The temptation would then be to use the upper and lower probability bound (whether taken at 95% coverage, or at two-thirds coverage, or whatever) in a traditional forecast. But this would be entirely mistaken, for at least four reasons. First, the probability coverage for the scenarios must be calculated jointly with mortality and migration. Second, the probability intervals are intended to cover annual variations in fertility, but as we have seen, this is not readily convertible into information about population size scenarios, since a good deal of cancellation of errors is to be expected. Third, the width of the probability interval for fertility or mortality is completely useless and uninterpretable for forecasting purposes without taking account of the autocovariance structure of the errors, which cannot be done using traditional methods. And fourth, all the problems of inconsistency remain undiminished in the probability intervals for different aspects of the projection, as discussed earlier. For these reasons, stochastic forecasts of fertility and mortality cannot readily be used to inform traditional demographic forecasts, or to evaluate the width of their high–low intervals. Lee (1991a) takes a small step towards bridging the gap by calculating probability intervals for the average level of fertility up to each forecast horizon, but such partial measures do not go nearly far enough.

Evidently, there is no hope of treating uncertainty in any way that is even roughly consistent, so long as we take the traditional approach. We turn, now, to a different approach, in which the vital rates are modeled as stochastic processes, and on the basis of these, a stochastic population projection is generated. Probability intervals can then be calculated for any quantity of interest. Unfortunately, this approach is very complex. Probability intervals cannot be calculated recursively, and for reasons indicated earlier, there is no way to combine the intervals for components to obtain an interval for the whole. Each must be calculated separately. Also, there are some sources of uncertainty that are not incorporated in this analysis. These include possible errors in the data for the jump-off population and in the data used for fitting the vital rate models; errors in specifying the models; errors in estimation of the model parameters; and future structural changes invalidating the estimated models.

3. Modeling vital rates

The first goal is to model the vital rates as stochastic processes. Fertility, mortality and migration vary strongly with age, so it is necessary to model and forecast age-specific rates. However, there is a strong tendency for the age-specific rates to move together, with respect both to trend and to fluctuation. There are several possible approaches. One, used in recent work on mortality by Rogers (1986) and McNown and Rogers (1989, 1992), is to develop an explicit analytic expression for the age profile, and then to model and forecast the time series of values of some or all of the parameters. They use the Heligman–Pollard (1980) mortality model, which uses eight parameters to describe the age profile of mortality. Similar approaches have sometimes been taken to forecast fertility. The advantage of these methods is a parsimonious expression for any particular age distribution. The drawbacks are first, that more than one variable must be forecast, and that covariances of the variables must be taken into account; second, in most published forecasts no probability intervals are provided, presumably owing to the problems with covariances just mentioned; third, in the case of mortality all published forecasts are for relatively short horizons of 10 or 15 years, so the long-run stability of such forecasts has not yet been established; and fourth, there is some indication that the time series of the parameter values are erratic, at least for the United States in recent years [see McNown and Rogers (1989) and Keyfitz (1990)].

The strategy pursued here is to generate a one-parameter family of age schedules, in the sense that variations in one parameter generate the entire range of schedules in the family. An actual description of a particular age schedule additionally requires twice as many age-varying coefficients as there are age groups. Different values of these coefficients define different families.[2]

Fertility, mortality and migration can all be handled in this way. Here, mortality will be used for illustration. Let $m_{x,t}$ be the central death rate for age x at time t. The model used for mortality is

$$\ln(m_{x,t}) = a_x + b_x k_t + e_{x,t} .$$

Here $\exp(a_x)$ describes the average shape of the age profile, and the b_x's describe the pattern of proportional deviations for this age profile when the parameter k varies. Parts of the Coale–Demeny (1983) life-table system are build on a model somewhat like this (where k is taken to equal e_{10}), and preliminary versions of the UN model life-table system also used an approach of this sort. Gomez (1990) found that out of the numerous models he considered using exploratory data analysis, this model gave the most satisfactory fit to historical Norwegian mortality data. The approach is related to the principal components methods developed by Bell and Monsell (1991) for mortality, and Bozick and Bell (1987, 1989) for fertility.

This model cannot be fit by simple regression, because there is no observed variable on the right-hand side. Nonetheless, a least-squares solution exists, and can be found using the first element of the Singular Value Decomposition [or SVD; see Wilmoth (1990) or Lee and Carter (1992) for details]. On inspection we can see that the solution cannot possibly be unique, however. To distinguish a unique solution, I will impose the further conditions that the sum of the b_x's equal 1.0, and that the sum of the k_t's equal zero. Under these assumptions, the a_x's are simply the average values over time of the $\ln(m_{x,t})$ for each x.

The model was fit to US data from 1933 to 1987, and to Chilean data from 1952 to 1987, both for sexes combined. [For a treatment of male and female mortality differences, see Carter and Lee (1992).] Exhibit 1 shows the estimated values of a_x and b_x for the two populations. The a_x's, as noted, are just the average

[2] In effect, this is the way we *define* a family of age schedules, since it is of course true that variation in other dimensions can be induced by varying other parameters. Our usage here is consistent with standard terminology in demography, as employed with respect to the Coale–Demeny (1983) model life tables, for example.

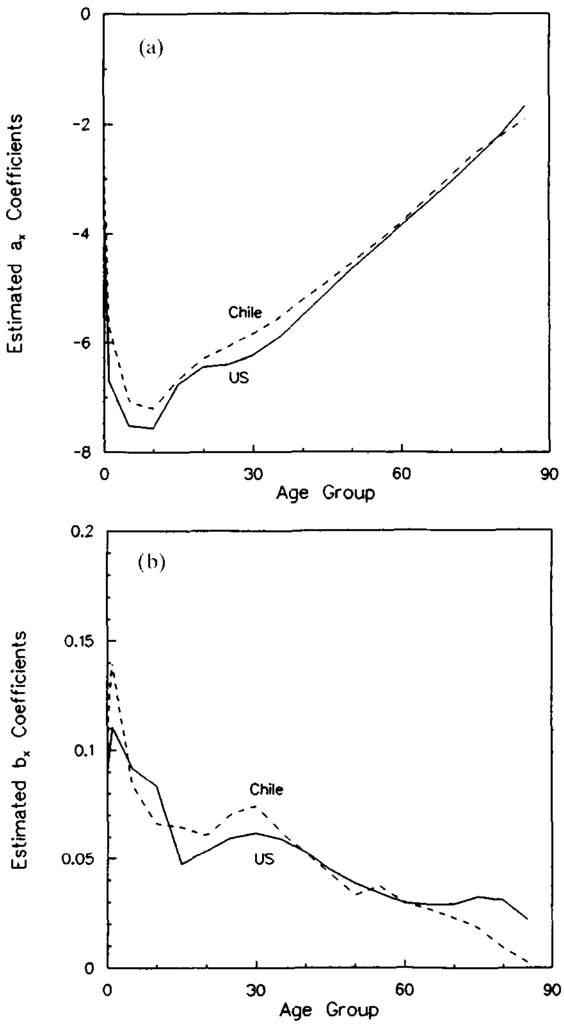

Exhibit 1. (a) Comparison of estimated a_x coefficients for the United States, 1933–1987, and Chile, 1952–1987. (b) Comparison of estimated b_x coefficients for the United States, 1933–1987, and Chile, 1952–1987. *Note*: The age groups are 0, 1–4, 5–9, . . . , 80–84, 85+.

values of the logs of the death rates. Not surprisingly, the Chilean coefficients lie above those of the United States at all ages except the highest, reflecting the fact that mortality was higher, on average, in Chile from 1952 to 1987, than in the United States from 1933 to 1987.[3] The b_x's describe the relative sensitivity of death rates to variation in k. It is also not surprising that sets of coefficients for the United States and Chile look quite similar. Because of the normalization, their absolute levels have no particular meaning. With N age groups, if $b_x = b_y = 1/N$ for all x and y, then all the rates would move up and down proportionately, maintaining constant ratios to one another. However, it can be seen that in fact some ages are much more sensitive than others. Generally speaking, the younger the age, the greater its sensitivity to variation in k. The exponential rate of change of an age group's mortality is proportional to the b_x value: $\mathrm{d}\ln(m_{x,t})/\mathrm{d}t = (\mathrm{d}k_t/\mathrm{d}t)b_x$. If k declines linearly with time, then $\mathrm{d}k_t/\mathrm{d}t$ will be constant and each m_x will decline at its own constant exponential rate.

[3] The Chilean data treat 65+ as the open age group. The Coale–Guo (1989) method was used to construct age-specific mortality estimates at older ages. For the United States, the open age interval is 85+. Lee and Carter used the Coale–Guo method to estimate mortality up to 105–109, but the figures shown in Exhibit 1 are not affected. The quality of age reporting above age 65 is probably poor in the United States and in many other countries as well.

4. Second stage estimation of k

The estimation process just described generates values for a_x, b_x and k_t, and we could proceed directly to the next step of modeling k_t as a time series process. Instead, we make a second-stage estimate of k by finding that value of k_t which, for a given population age distribution and the previously estimated parameters a_x and b_x, implies exactly the observed number of total deaths for the year in question. That is, we search for k_t such that

$$D_t = \sum \exp(a_x + b_x k_t) N_{x,t} ,$$

where D_t is total deaths in year t, and $N_{x,t}$ is the population age x in year t.

There are several advantages to making a second-stage estimate of k in this way. First, this guarantees that the life tables fitted over the sample years will be consistent with the data. Because the first-stage estimation was based on logs of death rates rather than the death rates themselves, sizable discrepancies can occur between predicted and actual deaths. Second, in this way the empirical time series of k can be extended to include years for which age-specific data on mortality. are not available, since the second-stage estimate of k yields an indirect estimate of mortality. For the United States, this allows the base year of the forecast to be brought forward by two or three years, because of the long time lag in publishing age-specific rates. For Chile, it allows us to fill in several gaps in the middle of the time series of mortality.

For populations with seriously incomplete data on fertility or mortality, the second-step estimation offers many opportunities for indirect estimation. For example in China there are national age-specific mortality data for only a few years, and population age distributions for four scattered census years over a 40-year period, although total births and deaths are available annually. a_x and b_x can be estimated from as few as two life tables. Given these, one can begin with the earliest census, and find the value of k_t and the corresponding life table (for single years of age) for the following year. This can be used to survive forward the population for one year, and likewise to survive births. This procedure yields an estimate of the population age distribution one year later. The process can be replicated recursively up to the present. In this way the population and its mortality history can be reconstructed annually, and a long time series of k_t can be created as a basis for forecasting. This method is a version of inverse projection [see Lee (1974a)]. Similar methods can be used to reconstruct a time series of fertility as a basis for forecasting.

5. Vital rates as stochastic processes

The next step is to model k as a stochastic time series process. This is done using standard Box–Jenkins procedures. In most mortality applications k_t is quite well modeled as a random walk with drift $k_t = c + k_{t-1} + u_t$. In this case, the forecast of k changes linearly and each forecasted death rate changes at a constant exponential rate. However, sometimes a model of this general form but with an added moving average term or autoregressive term is superior. In this case the pattern of change is somewhat different.

Note that each of the m_x's is now itself modeled as a stochastic process driven by the process k. Note also that if we ignore the error term $e_{x,t}$, the variations in the $\ln(m_x)$'s will be perfectly correlated with one another, since all are linear functions of the same time-varying parameter k. This is a very convenient feature of the model, for it means that we can calculate the probability bounds on all (period) life-table functions directly from the probability bounds on the forecasts of k, without having to worry about cancellation of errors. For a discussion of errors arising from $e_{x,t}$ and from the estimation of a_x and b_x, which are here ignored, see the appendix to Lee and Carter (1992).

We can now use the fitted time series model for k to forecast it over the desired horizon. Exhibit 2 shows past values of k for the United States, 1900–1989, and their forecasts from 1990 to 2065. Note that the estimated values of k over the base period change quite linearly. In fact, the change in k over the first half of the period almost exactly equals its change in the second half. This contrasts with our usual understanding that the pace of mortality decline has decelarated over time, an understanding based on the trends in life expectancy at birth, which we will examine later. The approximate linearity of k in the base period is a great advantage from the point of view of forecasting. Long-term extrapolation is always a hazardous undertaking, but it is less so when supported in this way by the regularity of change in a 90-year empirical series. Also note that, aside from the influenza epidemic of 1918, the variability of the series is similar throughout the period. This also is a desirable feature for forecasting purposes.

Exhibit 2 shows the point forecasts, which are

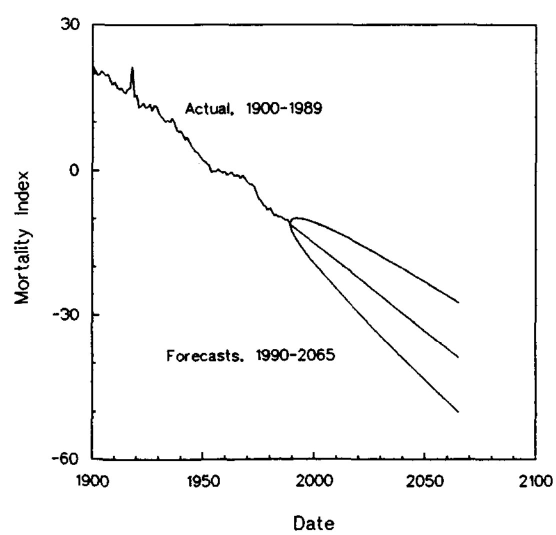

Exhibit 2. Estimated and forecast values of the mortality parameter k for US data, 1900–1989 and 1990–2065, with 95% probability intervals. *Note*: The 95% probability interval is shown. It reflects only innovation error, not uncertainty from estimation of drift or other sources. These are second-stage estimates of k. The forecasts are based on a random walk with drift, with a dummy variable for the 1918 influenza epidemic.

essentially a linear extrapolation of the base period series. The analysis of Chilean data showed a similar result, as did Gomez's (1990) analysis of Norwegian mortality data which took a slightly different approach. Ninety-five percent probability intervals are also shown for the forecast of k. The figure shows intervals that reflect only the error term in the random walk, which expresses the innovation error. However, these could be augmented to include uncertainty about the rate of decline in k, which is itself estimated [see Lee and Carter (1992)].

The next step is to convert the forecasts of k into forecasts of life-table functions. This is done using the equation which gives $\ln(m_{x,t})$ as a function of k, given the previously estimated age-specific coefficients a_x and b_x. Once the implied forecasts of $m_{x,t}$ have been recovered in this way, any desired life-table function can be calculated. For period life-table functions, the probability intervals can be found directly from the intervals on k. To find probability intervals for cohort life-table functions, we would have to take into account the autocovariance structure of errors in k, which is less straightforward.

Exhibit 3 displays actual (fitted) base period values of life expectancy at birth along with the forecasts derived from the forecasts of k in Exhibit 2. The figure also shows the high, medium and low forecasts by the Census Bureau (1989) and the Social Security Actuary [Wade, (1989)]. There are several points to note. First, we (Lee and Carter) forecast an increase in life expectancy of 10.6 years (sexes combined), from 75.5 to 86.1, between 1989 and 2065. This is more than twice as great as the increase projected by the Census Bureau and the Social Security Actuary. The difference translates into substantially higher forecasts of the elderly population, to be discussed later. Second, although we projected a linear trend in k, the forecast for life expectancy is for growth at a slowing pace. The difference is due to the decreasing entropy of the survival curve [see Keyfitz (1977)]. When mortality is higher, general mortality decline saves many young lives which contribute more to gains in life expectancy; when mortality is lower, general mortality decline saves largely old lives which contribute less to gains in life expectancy. Third, the probability interval is fairly tight, even though the plot includes uncertainty in the estimated drift term in addition to the innovation error. Previous use of statistical time series methods to forecast vital rate series had generally found very wide probability intervals, so wide that there seemed to be little information in the forecasts after a couple of decades. We see here that this need not be the case; indeed, these intervals have been criticized as being implausibly narrow. Note, however, the intervals on the forecasts of k were not particularly narrow. We believe that the narrowness of the forecasts of life expectancy arises from the low entropy of the survival curve when life expectancy reaches high levels: even sizable variations in k have little effect on the level of life expectancy, because mortality has become so concentrated at the older ages. This concentration does not necessarily represent a Fries-like 'compression' of mortality [Fries (1980)], but occurs even if, as here, mortality declines at all ages at historical rates. An increasingly small proportion of deaths occur before age 65, for example.

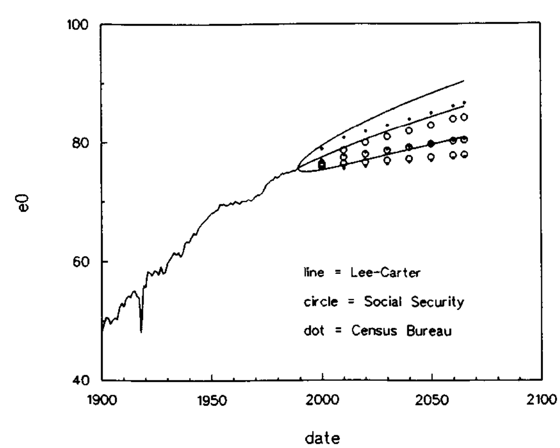

Exhibit 3. Actual and forecast values of life expectancy at birth for the United States, 1900–1989 and 1990–2065: Lee–Carter vs. official projections. *Note*: The Lee–Carter e_0 forecasts are derived from the forecasts of k in Exhibit 2, but the 95% probability intervals shown include uncertainty from the estimated drift term as well as innovation error [see Lee and Carter, (1992) for details]. High, medium and low projections from the US Bureau of the Census (1989) and the Social Security Actuary [Wade (1989)] are also shown.

6. Fertility

Procedures for fertility are in some respects similar to those for mortality, but in other respects quite different. Differences arise in part because fertility series in industrial populations are typically much less regular than mortality series. They fluctuate widely, and their pasts appear to contain less useful information about their futures. The proposed strategy is to depart from standard Box–Jenkins procedures by introducing prior constraints on the fitted models. There are several points to note. First, the basic model described above for the age distribution of mortality can be used, but it is much more satisfactory for unlogged fertility rates.[4] As for mortality, a second-stage estimate of the fertility parameter, say f_t, was made. Second, we know that average fertility is bounded below by zero and above by a bio-social upper limit of about 16 children per woman. Point forecasts of fertility may violate these bounds, but in any event the usual assumption of normally distributed errors certainly allocates positive probability to levels outside these bounds. To get around this difficulty, an inverse logistic transform of f_t may be forecast rather than f_t itself. Although developed independently in this work, an identical procedure was first used by Alho (1990). The inverse logistic transform can be chosen such that the derived forecast of f_t itself must fall between preassigned bounds. For the United States after 1900, I have taken these bounds to be 0 and 4 children per woman. Finally, the ultimate fertility level implied by unconstrained time series models, even using the logistic transform, appears quite capricious, and depends on aspects of the model that are not well estimated.[5] In the illustrative application below, the long-run equilibrium level of the total fertility rate was constrained to be 2.0.

Let $f_{x,t}$ be the birth rate for women age x at time t. Then the basic model of age variation is

$$f_{x,t} = a_x + b_x f_t \,.$$

The goal is to forecast f_t with suitable constraints. With upper and lower limits of U and L, and an equilibrium level of F^*, the transformed fertility index g_t is calculated as

$$g_t = \ln\{(f_t + A - L)/((U - f_t - A)\} \,,$$

where A is the sum of the a_x coefficients. Then the series g_t-G^* is modeled, with G^* chosen such that when $g = G^*$, then $f = F^*$ [see Lee (1991a) for details].

Obviously, the point of estimating a model constrained in these ways cannot be to learn something new about the likely future trend of fertility, since the ultimate future level is assumed a priori rather than estimated. Rather, the point is to learn something about the degree of uncertainty in the forecast, and more particularly about the autocovariance structure of the errors. As discussed earlier, this autocovariance structure is an absolutely essential ingredient for evaluating the uncertainty of the implied population forecasts. While the point forecasts of fertility are quite sensitive to the prespecified equilibrium level, the width of the 95% probability band is relatively independent of prespecified range, as is the autocovariance structure [see Lee (1991a)]. The imposed upper and lower bounds evidently mainly affect the upper and lower 2.5% of the distribution of the fertility forecast, rather than the central 95%.

Exhibit 4 plots the resulting forecast for the total Fertility Rate, along with 95% probability intervals. The intervals appear to be very wide, but recall that these should contain 95% of the annual variations in fertility. We are accustomed to seeing intervals for traditional forecasts, intervals that are designed to generate plausible high–low bounds for the total population. Therefore the high–low fertility assumptions must be much narrower than would be necessary to cover 95% of the annual variations, most of which would cancel. For a further discussion, see Lee (1991a).

Modeling and forecasting net migration proceeds in a similar way, although because reliable data are scarce for the United States, the age distribution of net migrants is simply assumed to remain fixed at that assumed in the census projections [US Bureau of the Census (1989)]. Net migration can be positive or negative, and has no clear upper bound; consequently, bound constraints were not specified. However, an equilibrium level consistent with the long-run assumption of the census projections was prespecified.

[4] In the logged version, the estimate of the fertility parameter f (corresponding to k in the mortality model) is unduly influenced by the secular trend in fertility at the older ages.

[5] In particular, the forecast behaves very differently depending on whether or not a constant term is included, and whether or not the original series is differenced before modeling. Often, there is little purely statistical basis for making decisions on these aspects of the model.

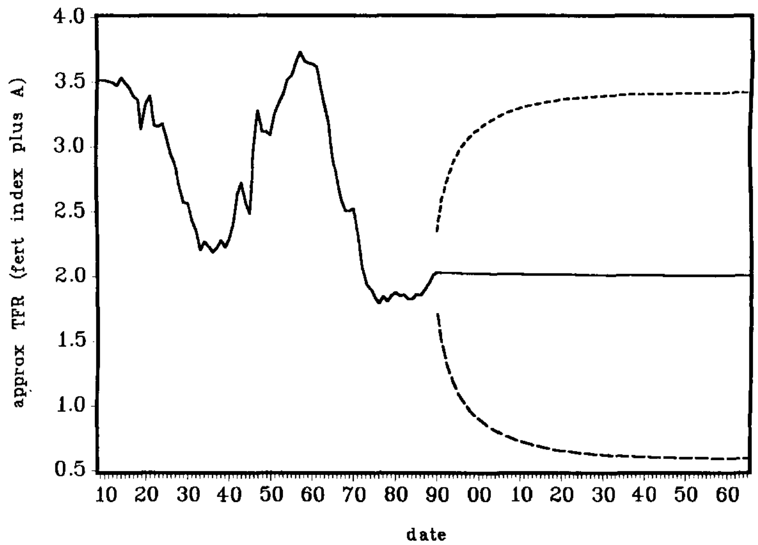

Exhibit 4. Actual and forecast values of the total fertility rate for the United States, 1909–1989 and 1990–2065, with 95% probability interval. *Note*: Forecasts are based on a logistic-transformed index, constraining fertility to lie between 0 and 4 children per woman. The ultimate level of the forecast is constrained to be 2.0. See Lee (1991) for details.

7. Stochastic population projections

Having developed stochastic models of age specific fertility, mortality and migration, the next step is to construct stochastic forecasts of the population itself and its age distribution. This approach was apparently first implemented by Sykes (1969), and in more detail by Alho and Spencer (1985). The Sykes (1969) paper pioneered in focusing on time series variability as the most important source of uncertainty in population forecasts, but the modeling of the vital rates was rather crude and the forecast, with 15-year age groups, was only illustrative. Alho and Spencer (1985) pioneered in a number of ways, including considering a broader range of sources of error, and incorporating expert opinion. However, they restrict their forecasts to a 15-year horizon in which it is not necessary to deal with compounding second generation uncertainty (uncertain fertility of uncertain numbers of reproductive-age women), and they do not consider probability intervals of functions of the age distribution such as ratios of age group sizes. The projections described here have a long horizon, covering the period from 1990 to 2065. The uncertainty of various functions of the age distributions is considered, including age group size ratios, and implied tax rates for transfers. Results reported here pertain to female dominant single sex projections.

The propagation of error in the Leslie matrix is analyzed using procedures described in Tuljapurkar (1990, 1992) and in Lee and Tuljapurkar (1991). Although the details of these procedures will not be discussed in this paper, there are several points to note, illustrating the complexity of the problem. First, the stochastic point forecasts of population variables are not derived from the point forecasts of the individual rates. Instead, they must take into account the entire probability distribution for each rate, because of the non-linear nature of population renewal. The following example, due to Tuljapurkar, shows why: Alternating annual population growth rates of zero and 0.02 average out to a growth rate of 0.01, which would be the point forecast for the growth rate. But the population subject to these growth rates will grow at a lesser average rate equal to $\{\mathrm{sqrt}(1.00*1.02)-1\}$ per year, or 0.00995. This example also shows why it may be important to screen out implausible extreme values from the forecasts of the vital rates by specifying upper and lower bounds. [For a more formal statement, see Tuljapurkar (1992)]. Second, the stochastic forecast is not recursive when errors are autocorrelated. To move ahead one step, we need more information than is contained in the population age distribution at the previous step. The moments of the population distribution, even the first, depend on non-linear functions of the autocovariances of the errors which preclude recursive calculation [Tuljapurkar (1992)].

Detailed (but preliminary) results of the stochastic population projections can be found in Lee and Tuljapurkar (1991). The discussion here will be limited to some of the more important implications of the results, particularly those having to do with population aging and its implications. First consider the elderly population. Because we forecast mortality to decline more rapidly than do the Census Bureau and the Social Security Actuary, we also forecast a large elderly population. Specifically, our point forecast of the population 65 and over for 2065 is more than 18% larger than the figures for the US Bureau of the Census (1989) and Social Security [Wade (1989)]. For the population 85 and over, we expect 28% more than the Census Bureau and over 40% more than Social Security. These differences in point forecasts have major implications for the challenge posed by population aging. The low forecasts of the Census Bureau and Social Security are very similar to our lower 95% probability bound. However, our upper bound has 5.6 million more people age 85 and over than does the Census, and 11.4 million (57%) more than does Social Security! From these comparisons, it appears that the Social Security and Census forecasts may be understating both the expected level and the upside range of the future proportions of elderly. These aspects of their forecasts follow from their rather lower forecast of life expectancy gains as discussed earlier.

For some policy purposes, we are more interested in the elderly dependency ratios than we are in the absolute numbers of elderly. One of the major shortcomings of the traditional forecast methods is their inability to convey a sense of the uncertainty of dependency ratios. Because the high–low scenarios are assumed to hold fixed over the full range of the forecast, the extreme age group ratios are those that result from permanently high fertility and low mortality, or permanently low fertility and high mortality. In some variations, permanently high fertility and high mortality, or low fertility and low mortality, may be assumed to capture 'optimistic' and 'pessimistic' outcomes from the point of view of the Social Security system [Wade (1989)]. None of these cases reflects the possibilities that fertility might be initially very low, creating a small cohort of those who will be old at the end of the forecast period, and that fertility might then rise, creating a large number of working-age people to support them. Nor can they reflect the possibility that fertility might start high for a decade, then go very low for fifty years, and then again be very high at the end of the forecast period, creating a severe life cycle squeeze, with a small number of working-age people with large numbers of both children and elderly to support at the same time. To capture such possibilities, and to assign them appropriate probabilities, it is necessary to allow vital rates to vary stochastically, which traditional methods cannot.

We have forecast the elderly dependency ratio – that is, the ratio of the population 65 and over to the population 20–64 [see Lee and Tuljapurkar (1991)]. This depends on the uncertainty in fertility as well as the uncertainty in mortality. For 2065, our point forecast of the dependency rate is 14% higher than that of census, and 22% higher than that of Social Security. The 95% probability intervals that we generate for 2065 are *twice* as wide those of Social Security, and eight times as wide as those implied by the Census Bureau's high–low brackets! However, the Census Bureau explicitly recommends using a specially constructed forecast combining low mortality and low fertility to get a sense of the high range for elderly dependency. This special scenario is very similar to the Social Security pessimistic projection, although our upper bound is 33% higher than both of these. Probably the main reason for our much wider intervals is that our stochastic fertility model expresses more uncertainty about future fertility, and incorporates the possibility of long swings in fertility. We believe that this is an important and realistic feature of our approach.

The total dependency ratio is the ratio of the population under age 20 and over 65, to the population aged 20–64. It allows for offsetting variations in the dependency of children and the elderly, and therefore it is not surprising that its forecast is far less uncertain than is the elderly dependency ratio.

Dependency ratio calculations give each age group a weight of unity. If we are interested in the implications of our forecasts for public expenditures, however, we might want to weight the age groups by the amount they receive in transfers or subsidies to create a numerator for the ratio, and weight the age groups by the amount of their labor earnings to form the denominator. Then the ratio is a kind of implied average payroll tax rate necessary to support health, education, pensions, and other social welfare programs, on the assumptions that all revenues for these programs come from taxes on labor earnings. Holding these tax and transfer based weights constant over the forecast horizon, we can calculate point forecasts and confidence intervals for ratios of this sort [see Lee and Tuljapurkar (1991)]. Our preliminary forecast is that the payroll tax for the elderly will more than double by 2065, with 95% bounds containing possibilities ranging from a return to the low levels of today to more than tripling – a range from 0.12 to 0.44 in 2065. This is a very wide interval, but earlier, from 2015 to 2030, when the baby boom retires, the interval is much narrower, primarily reflecting uncertainty about mortality, not fertility. A substantial rise in payroll taxes seems virtually certain, even though the projection fan widens to contain lower tax rates later in the century.

Just as the forecasts of the total dependency ratio were less uncertain than those of the elderly dependency ratio, so the forecast of the payroll tax rates necessary to fund all public sector transfers, to children, younger adults and the elderly, are less uncertain than are those for transfers to the elderly alone. The point forecast is for a 50% increase in taxes in 2065, with a range from the current level up to a doubling of the current level. Once again, an increase around 2030 appears nearly certain, while the longer run picture is far less so.

These calculations are based on the simple assumption that, age for age, the costs of publicly provided health care, education, and pension programs remain the same in relation to age-specific earnings as they are today. This assumption is relatively robust to variations in inflation, but pension costs will be sensitive to the rate of growth of productivity, which is additionally uncertain. Health care costs are likely to continue to rise in real terms for some time to come, contributing further uncertainty. The calculations also assume, contrary to fact, that these public transfer expenditures must be funded exclusively from labor income. Despite these shortcomings, these calculated tax rates provide a suggestive interpretation of the age distribution changes we forecast, and illustrate the power of the stochastic forecast.

8. Problems and possibilities

This paper has described an approach to stochastic population forecasting. It has shown how statistical time series methods can be melded with demographic models to produce plausible long-run forecasts of vital rates, with probability distributions. The method used to forecast mortality has some promise for improving both the point forecasts of mortality and measures of their uncertainty. The method for forecasting fertility, by contrast, is useful mainly for providing estimates of the variance and autocorrelation in fertility, items that are centrally important for the next step. The vital rate forecasts can then be combined in a stochastic Leslie matrix, and used to produce stochastic population forecasts. These provide probability distributions of the quantities forecast. The resulting measures of uncertainty must be viewed as lower bounds on uncertainty, since a number of the potential sources of uncertainty have not been incorporated in the analysis.

Estimates of uncertainty should, in principle, be crucial for policy formation based on population forecasts. However, there are serious barriers to the more widespread use of these or similar methods. First, while the methods described here for modeling the vital rates are not particularly difficult to employ, the calculation of the stochastic population forecast itself is extremely complex to implement. We hope to develop some simple approximations or rules of thumb for deriving rough probability intervals from the models of the vital rates. Alho (1991) presents some interesting ideas for making the results of stochastic forecasts accessible to users. Second, planners are not accustomed to working with forecasts of probability distributions. In theory, it is straightforward: the planner uses a loss function to weight each possible outcome, and then integrates over the weighted probabilities. This yields a number which may be quite different from the expected value of the forecasts, which results from an equal weighting of all outcomes. In practice, however, loss functions are seldom if ever used by the consumers of population forecasts. Therefore methods for using stochastic population forecasts must be developed hand-in-hand with methods for carrying them out. The calculation of probability distributions for payroll tax rates is a small step in that direction. In future research, we intend to calculate forecasts of the size of the Social Security trust fund, and flows between Treasury and Social Security. Dynamic programming will be used to find optimal trajectories for payroll tax rates, conditional on the loss function. While the research described here represents an advance, it is clear that much remains to be done.

Acknowledgments

Research for this paper was supported by a grant from NICHD, RO1-HD24982. Much of the work described was conducted jointly with Lawrence Carter and with Shripad Tuljapurkar.

References

Alho, J.M., 1990, "Stochastic methods in population forecasting", *International Journal of Forecasting*, 6, 521–530.

Alho, J.M. and B.D. Spencer, 1991 "A population forecast as a database: Implementing the stochastic propagation of error", *Journal of Official Statistics*, 7, 295–310.

Alho, J.M. and B.D. Spencer, 1985, "Uncertain population forecasting", *Journal of the American Statistical Association*, 80 no. 390, 306–314.

Alho, J.M. and B.D. Spencer, 1990, "Error models for official mortality forecasts", *Journal of the American Statistical Association*, 85, 609–616.

Bell, W.R. and B.C. Monsell 1991, "Using principal components in time series modeling and forecasting of age-specific mortality rates", paper presented at the 1991 Annual Meetings of the Population Association of America, Washington, DC, March.

Bozik, J. and W. Bell 1987, "Forecasting age-specific fertility using principal components", SRD Report Series, RR-87/19, US Bureau of the Census.

Bozik, J. and W. Bell, 1989, "Time series, modeling for the principal components approach to forecasting age-specific fertility", paper presented at the 1989 Meetings of the Population Association of America.

Carter, L. and R.D. Lee, 1986, "Joint forecasts of U.S. marital fertility, nuptiality, births and marriages using time series models", *Journal of the American Statistical Association*, 81, no. 396, 902–911.

Carter, L. and R.D. Lee, 1992 "Modeling and forecasting US mortality: Differentials in life expectancy by sex", *International Journal of Forecasting*, 8, 393–411 (this issue).

Coale, A.J. and P. Demeny, 1983 *Regional Model Life Tables and Stable Populations*, 2nd edn. (Princeton University Press, Princeton, N.J.)

Coale, A. and G. Guo, 1989, "Revised regional model life tables at very low levels of mortality", *Population Index*, 55 no. 4, 613–643.

Cohen, J. 1986, "Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches", *Demography*, 23, no. 1, 105–126.

Fries, J.F., 1980, "Aging, natural death and the compression of morbidity", *New England Journal of Medicine*, 303, 130–135.

Gomez de Leon, J. 1990, "Empirical DEA models to fit and project time series of age-specific mortality rates", unpublished manuscript.

Heligman, L. and J.H. Pollard 1980, "The age pattern of mortality", *Journal of the Institute of Actuaries*, 107, 49–80.

Keyfitz, N., 1977, *Applied Mathematical Demography* Wiley, New York).

Keyfitz, N., 1981, "The limits of population forecasting", *Population and Development Review*, 7 no. 4, 579–593.

Keyfitz, N., 1990, "Future mortality and its uncertainty", paper presented at the Annual Meetings of the Population Association of America.

Land, K., 1986, "Methods for national population forecasts: A review", *Journal of the American Statistical Association*, 81, no. 396, 888–901.

Lee, R., 1974a, "Estimating series of vital rates and age structures from baptisms and burials: a new technique", *Population Studies*, 28, no. 3, 495–512.

Lee, R., 1974b, "Forecasting births in post-transitional populations: Stochastic renewal with serially correlated fertility", *Journal of the American Statistical Association*, 69, no. 247, 607–617.

Lee, R., 1981, "Comment on 'McDonald: Modelling demographic relationships: An Analysis of Forecast Functions for Australian Live Births'", *Journal of the American Statistical Association*, 76, no. 376, 793–795.

Lee, R.D. 1991a, "Modeling and forecasting the time series of U.S. fertility: Age distribution, range and ultimate level", Unpublished Manuscript, Demography, University of California at Berkeley.

Lee, R.D. 1991b, "Modeling and forecasting mortality in Chile: Application of a new method", Unpublished Manuscript, Demography, University of California at Berkeley.

Lee, R.D. and L. Carter, 1992, "Modeling and forecasting the time series of U.S. mortality", *Journal of the American Statistical Association*, 87, no. 419, 659–671.

Lee, R.D. and S. Tuljapurkar 1991 "Stochastic population projections for the U.S.: Beyond high, medium and low", paper presented at the Annual Meetings of the Population Association of America, Washington, DC.

McDonald, J. 1981, "Modling demographic relationships: An analysis of forecast functions for Australian births", *Journal of the American Statistical Association*, 76, 782–792.

McNown, R. and A. Rogers, 1989, "Forecasting mortality: A parameterized time series approach", *Demography*, 26, no. 4. 645–660.

McNown, R. and A. Rogers, 1992, "Forecasting cause-specific mortality using time series methods, *International Journal of Forecasting*, 8, 413–432 (this issue).

Pflaumer, P., 1988, "Confidence intervals for population projections based on Monte Carol methods", *International Journal of Forecasting*, 4, 135–142.

Saboia, J.L.M., 1977, "Autoregressive integrated moving average (ARIMA) models for birth forecasting", *Journal of the American Statistical Association*, 72, 264–270.

Rogers, A., 1986, "Parameterized multistate population dynamics and projections", *Journal of the American Statistical Association*, 81, no. 393, 48–61.

Stoto, M., 1983, "The accuracy of population projections", *Journal of the American Statistical Association*, 78, no. 381. 13–20.

Sykes, Z.M., 1969, "Some stochastic versions of the matrix model for population dynamics", *Journal of the American Statistical Association*, 44, 111–130.

Tuljapurkar, S., 1990, *Population Dynamics in Variable Environments, Lecture Notes in Biomathematics* 85. (Springer-Verlag, Berlin).

Tuljarpurkar, S., 1992, "Stochastic population forecasts and their uses", *International Journal of Forecasting* 8, 385–391 (this issue).

US Bureau of the Census 1989, Current Population Reports, Series P-25, No. 1018, Projections for the Population of the United States, by Age, Sex, and Race: 1988 to 2080, by Gregory Spencer (US Government Printing Office, Washington, DC).

Wade, V., 1980, Social Security Area Population Projections: 1989, Actuarial Study No. 105 (Office of the Actuary, Social Security Administration, Washington, DC).

Wilmoth, J.R., 1990, "Variation in vital rates by age, period, and cohort", in: C. Clogg, ed., *Sociological Methodology* vol. 20 (Blackwell, Oxford) 295–335.

Biography: Ronald D. LEE taught economics and demography at the University of Michigan from 1971 to 1979, and since then has been at the University of California at Berkeley in Demography and Economics. His primary fields of interest, besides demographic forecasting, are the interactions of population with economy in preindustrial times, and the influence of age structure and intergenerational transfers in economic growth. He is a member of the U.S. National Academy of Sciences, and a Past President of the Population Association of America.