STOCHASTIC POPULATION FORECASTS AND THEIR USES

Shripad Tuljapurkar

Paper number 0045

Department of Biological Sciences
Morrison Institute for Population and Resource Studies
Stanford University
Stanford, California 94305

1992

1 Introduction

The production of population forecasts (here also called projections) is often perceived as one of the most important tasks of demographers. Making a projection is a three-part job. We need initial conditions at the launch time, when the forecast is to begin; we must forecast the vital rates that determine population change, namely mortality, fertility and migration; and we have to use the vital rates to generate forecasts of population and various functionals of population. Initial conditions are often the easiest to establish if good Census data are available. Vital rates are difficult, since there is limited useful theory bearing on prediction. One approach to vital rates, traditional at the U.S. Census Bureau and elsewhere, is to rely on the expert opinion of informed historical analysts and generate "high-medium-low" scenarios for future rates. A different approach is to use historical data to construct stochastic models for the time-series of vital rates. This doesn't do away with expert opinion, which remains necessary for discriminating between models. But it does force the choice of an expicitly probabilistic model, a model that drives changes in vital rates over the course of the forecast period. Capturing historical variability in the rates in this way seems essential in the absence of an accurate theory that can predict the rates, and it also sets the stage for translating this variability to produce probabilistic forecasts.

This paper focuses on the third part of the forecaster's task, the propagation of uncertainty in population numbers. I first examine the theoretical underpinnings of forecasts that make use of stochastic models of vital rates. This paper takes an analytical view of the problem, concentrating on some basic but important features that are relevant to all such forecasts. Different but complementary aspects of stochastic projection are discussed by Cohen (1986), Alho and Spencer (1990) and Alho (1991). I will not have much to say about the modeling of vital rates themselves; for this, see for example Lee and Carter (1992) and Manton and Stallard (1984). Next I turn to a discussion of the ways in which probabilistic forecasts can be presented and used. I argue for embedding probabilistic forecasts in the framework of decision theory, and discuss ways in which this may be done.

I will begin with a discussion of recursion and conditioning in the simple context of a scalar population model. Several key features of the propagation of uncertainty are identified in this way, and I then show that these features also apply to the random matrix products that appear in demographic forecasts. I examine also the features of some population functionals that are commonly forecast. Then I turn to the representation of uncertainty in ways that are likely to be useful and intelligible to the users of forecasts.

2 How Uncertainty Propagates

The mathematics of population projection, as used by most demographers making age-structured forecasts, is in principle a simple recursion. However this simplicity is deceptive because of the multiplicative nature of the recursion, correlations over time, and the fact that we must project vectors. In this section I consider the effects of the first two of these factors by examining a simple scalar model.

Let M_t be the total number of some hypothetical population at a discrete time t; suppose that time is divided into suitable discrete units and that between t and $t+1$ the net rate of population growth is R_{t+1}, so that

$$M_{t+1} = R_{t+1} M_t. \tag{1}$$

In the class of models I consider here, the vital rates will always be independent of past population, excluding such things as an Easterlin (1980) feedback between population size and fertility. I don't exclude such effects because they are unimportant, but only to focus on the most widely used class of models. Therefore in equation (1), I assume that R_{t+1} is statistically independent of M_t. In such a model the dynamics of vital rates are described by a model for R_t in terms of its own past values. It is typical of such models, for example the time-series models of Lee and Carter (1992) and Bozik and Bell (1989), that the vital rates display serial autocorrelation over time, and this correlation has important consequences for the computatation of forecasts.

To see what autocorrelation does to recursion, take the sequence $R_t, t = 1, 2, \ldots$ to be serially correlated. Specifically let us suppose that

$$R_t = r + \epsilon H_t. \tag{2}$$

Here, H_t is a stationary zero-mean stochastic process with

$$\mathcal{E}(H_{t+m} H_t) = c(m) \neq 0. \tag{3}$$

The parameter ϵ in equation (2) is a smallness parameter, indicating that the stochastic variations in R_t are dominated by the deterministic rate r. It is not obvious that ϵ will be small in all conceivable cases, but it usually is when the conventional modeling strategy is followed, in which we attempt to capture most of the variability in a secular or deterministic model.

Let us begin the forecast at time $t = 0$ with a known M_0. The population at time t is given by

$$M_t = (R_t R_{t-1} \dots R_1) M_0.$$

It is crucial to recognize that we are now interested in calculating, for each $t > 0$, the moments $\mathcal{E}[M_t^j]$, $j = 1, 2, \dots$, conditional on the intial state M_0. In view of the serial dependence of equation (3), the conditional moments of M_t do not bear a simple relation to those of M_{t-1}.

In this scalar case we can write down exactly the moments of M_t. Note that

$$\prod_{i=1}^{t} R_i = r^t + \epsilon r^{t-1} \sum_i H_i + \epsilon^2 r^{t-2} \sum_{t \geq i > j \geq 0} H_i H_j + \cdots. \quad (4)$$

Assuming that we can compute the multitime moments of H_t, as we can for example in the case of a time-series model, we can use equation (4) to compute all moments of M_t conditional on M_0. It is instructive to write down the first three, as follows. The mean is

$$\mathcal{E}(M_t/M_0) = r^t + \epsilon^2 \, r^{t-2} \, t \, [\sum_{m=0}^{t-1} (1 - \frac{m}{t}) c(m)] + \cdots, \quad (5)$$

the variance is

$$var(M_t/M_0) = \epsilon^2 \, r^{2t-4} \, t \, [c(0) + 2 \sum_{m=0}^{t-1} (1 - \frac{m}{t}) c(m)] + \cdots, \quad (6)$$

and the third central moment is

$$\mathcal{E}[M_t - \mathcal{E} M_t]^3 = {M_0}^3 \, \epsilon^3 \, r^{3t-3} \sum_{t \geq i,j,k \geq 1} \mathcal{E}(H_i H_j H_k) + \cdots. \quad (7)$$

These equations are written to leading order in ϵ and they illustrate several key facts about the propagation of uncertainty.

1. Linearization. It is not uncommon to find analyses which assume that the mean changes simply at the deterministic rate r. This is clearly inadequate whenever the correlations in equation (5) do not disappear or cancel. They will disappear only if there is no autocorrelation, and they will cancel exactly only in unusual cases. Whenever there are serial correlations we must take them into account in computing the mean.

2. Moments and Smallness. We see that the central moment of order j is of leading order ϵ^j for $j \geq 2$. This suggests that for small ϵ it may be adequate to compute these using the leading orders as shown in the equations above. However we see also that as t increases and we go out further into the future, we will inevitably be making larger errors by such an approximation.

3. Closed Recursions Cannot Be Found. Examination of equation (5) and equation (6) shows that it is not possible to construct a recursion that describes the evolution of the first two moments as a closed system. That is, these moments at time t and the statistical properties of R_{t+1} are not sufficient information to compute these two moments at $t+1$. The actual moments at any time involve correlations over the entire time span of the forecast. This is true even in the lowest order approximation with respect to ϵ. It does not seem possible to construct a closed recursion in just two moments as claimed by Davis (1988).

4. Quadratic Approximation. I will refer to the terms explicitly shown in equations (5) to (7) as the quadratic approximation. This appears to be the minimal approximation that one can try to get away with. It is now tempting to think that we should compute $\mathcal{E}M_t$ and $\sigma_t = [var(M_t)]^{1/2}$ and present 95% confidence intervals as $[\mathcal{E}M_t \pm 2\sigma_t]$. This temptation is to be avoided for reasons given below. However the moments themselves are certainly useful and may be needed for the calculation of other demographically interesting quantities. The quadratic forecast will surely become inaccurate as time goes by, due to the growing number of correlations of order higher than two. It is relatively straightforward to extend this to a quartic forecast that includes all fourth-order moments of the stochastic terms H_t. The neccessity of such an extension dpends on the time-span of the forecast and remains a matter for empirical checks.

5. Distributional Inferences. The multiplicative character of equation (1) shows that $\log M_t$ will be approximately normally distributed as t increases. In the simple case considered here we can use standard probabilistic arguments to prove that a lognormal distribution actually emerges as $t \to \infty$ and to find its parameters. This is not as useful as one might suppose, since our focus is entirely on finite and often small

t. Even so, the lognormal distribution is a reasonable approximation for finite t although its moments will change with time. This distributional property of equation (1) is crucial to making correct statements about where the probability of M_t is concentrated. Instead of the method suggested above, we are better off using the properties of the lognormal distribution and the moments computed by the quadratic approximation to construct 95% confidence intervals for $\log M_t$ assuming that it is normally distributed. We can then transform these to get more accurate confidence intervals for M_t.

It is worth emphasizing the lognormal property of equation (1), since that property is often retained even if we have a nonstationary growth rate. For example, suppose that instead of equation (2), the growth rate has the form

$$R_t = r_t + \epsilon g_t H_t, \tag{8}$$

where r_t and g_t are time-varying. It is often the case that trends in fertility and/or mortality will lead to such behavior. As long as the series (g_t/r_t) satisfies some boundedness and summability conditions (cf. proofs of the standard limit theorems, Billingsley 1979), we will still find that

$$\frac{1}{\sqrt{t}}\left[\log M_t - \mathcal{E}\log M_t\right]$$

is approximately normally distributed with some time-dependent variance. This conclusion is easily checked by simulation in more complicated models, and makes a difference to statements about the probability distribution of M_t.

3 Uncertainty and Age-Structure

The reader may wonder how well the lessons of the scalar model of the previous section carry over to the age-structured case that we are really interested in. The first task of this section is to show that the lessons all apply. The second task is to sketch out the way in which quartic and similar approximations can be formulated. The third task is to examine certain functionals of population and discuss their significance.

The age-structured forecasting model must deal with a vector N_t of population numbers by age at time t. It is possible that this vector is further disaggregated by race, socioeconomic variables, and sex. I am restricting myself here to linear models, in which these additional variables can be included in a generalized linear recursion: for example, if we include sex, we nevertheless project births by applying fertilities to the female population and then use the sex-ratio to sort out male from female births. As before we exclude nonlinear feedbacks between the demographic variables of interest. Given this assumption, I will confine discussion to the one-sex age-structured case. I also ignore immigration; most of the theoretical conclusions here carry over to the situation with immigration, so long as the level of immigration does not dominate population change.

Taking N_t to be the vector of female numbers by age, the vital rates for females are contained in a Leslie projection matrix which is written as X_t at time t, and the population dynamics are given by

$$N_{t+1} = X_{t+1} N_t. \tag{9}$$

This is just a multivariate version of equation (1). The problem now is to compute forecasts of N_t conditional on a known $N_0 = n_0$ at time $t = 0$. The vital rates in equation (9) are driven by some form of stochastic "engine": examples of this are time-series models that predict fertility and mortality, or models that predict mortality via an internal dynamic such as a diffusion in a space of risk-factors (Manton and Woodbury 1985). It is typical of these models that the vital rates, at least for human populations in this century, display serial correlation over time.

The problem of projecting the moments of N_t conditional on N_0 is a precise analog of the scalar problem discussed earlier. The recursion equation (9) can be solved formally in terms of a product of random matrices: this product is more complicated than that which solves equation (1), because the matrices involved do not commute. However there is a general theory of random matrix products in demography (Cohen 1979 , Tuljapurkar 1991) and it shows that the key conclusions of the scalar model extend to the matrix case:

1. Closed Recursions Cannot be Found. Just as in the scalar case the exact moments of N_t involve moments of X_t through the tth order multitime moments. In the presence of serial autocorrelation, this will preclude a closed form recursion involving just two moments. The same conclusion applies to approximations (see below).

2. Distributional Inferences. When the vital rates X_t form a stationary stochastic process, and under conditions satisfied by most demographic projection matrices, the stochastic theory of demography shows (Tuljapurkar and Orzack 1980) that at large t the distribution of the size of the population (or any subset of it) becomes lognormal. Simulations show that in practice the lognormality is achieved fairly rapidly.

   In the nonstationary case that we are most likely to confront, there will also be may conditions under which asymptotic lognormality applies. For finite times t, simulations with time-series models (Lee and Tuljapurkar 1991) show that lognormality is quickly achieved. At a minimum, therefore, it is necessary to assume that the distribution of population numbers is likely to be distinctly nonnormal. Computation of moments other than the first two is strongly advised to check on skewness and kurtosis. And, as a useful rule of thumb, a logarithmic transformation is likely to be the most appropriate normalizing transformation.

3. Approximations. We can mimic the approximation method of the scalar model in the matrix case. The key step is to decompose the vital rates as

   $$X_t = b_t + \epsilon H_t, \tag{10}$$

   where the deterministic b_t matrix contains trends and H_t is a matrix-valued zero-mean autocorrelated stochastic process. Now a multivariate analysis (Lee and Tuljapurkar 1991) can be carried out that generalizes the scalar formulae of the preceding section. The formulae that one obtains are more complicated, and require somewhat lengthy numerical evaluation, but are very similar in structure. There is a quadratic approximation that yields the first two moments to order ϵ^2: even in this approximation, it is not possible to find a closed recursion in the first two moments. We have found that in the matrix case, just as in the scalar case, the avaerage population vector is not accurately predicted by ignoring the stochastic terms in equation (10), even for small ϵ.

The preceding discussion focuses on moments of the population vector N_t and those conclusions also apply to the simplest functional of this vector, namely the size of a subset of the population. However demographers are also interested in more complicated functionals of the population vector, and different considerations can arise, as I now illustrate. Denoting the scalar product of two vectors a and b by (a, b), the size of a subset of population is a quantity of the form (c, N_t), where the vector c has ones where one wants to count age-classes and zeros elsewhere. With this notation, we can define a set of dependency ratios: let c_Y be a vector that has ones for ages under 20 and zeros else, let c_W be a vector that has ones for ages 20-64 and zeros else, and c_O be a vector that has ones for ages over 64 and zeros else. Then the dependency ratios usually of interest are: that for young people, that for old people, and a total,

$$\begin{aligned} D_{Yt} &= \frac{(c_Y, N_t)}{(c_W, N_t)}, \\ D_{Ot} &= \frac{(c_O, N_t)}{(c_W, N_t)}, \\ D_t &= D_{Yt} + D_{Ot}. \end{aligned} \tag{11}$$

Suppose that we have followed the mathods described above and have in hand a stochastic forecast of N_t. To make accurate forecasts of the dependency ratios we must note two important things:

a. Ratios are nonlinear functions of N_t and therefore their average values depend on the variance and possibly higher moments of N_t. It is very misleading to use as predictors of these ratios the values found by substituting $\mathcal{E}N_t$ for N_t on the right side of equation (11).

b. These ratios are actually functions of the population's age-structure, i.e., of the vector of proportions by age rather than of numbers by age.

Letting P_t be the total population number, the age-structure vector is

$$W_t = \frac{N_t}{P_t}. \tag{12}$$

Little is known about the distributional properties of W_t from the theory. The lognormality results discussed earlier apply to P_t and components of N_t, but the proportions in W_t certainly need not be lognormally distributed. In the cases that we have studied, it is perfectly adequate to compute $\mathcal{E}W_t$ and the variance of W_t and then construct confidence intervals in the usual way.

More complicated functionals are also of interest and I will consider an example in the next section.

4 How Does It Help To Know The Variance?

A common response to a stochastic forecast is: "the end result is a mean forecast together with, say, 95% confidence intervals. But a consumer of forecasts takes the mean to be the best guess, and it is not clear what he is to do with the intervals. How do you make estimators of uncertainty meaningful to the users of a forecast?" I believe this question is a serious one. If the maker of a stochastic forecast aims to provide better information, not merely to display technical virtuosity, much more attention must be paid to translating uncertainty estimates into measures meaningful to the user.

A technically trained person might say that he knows what a variance is, and the information doesn't require further translation. This is narrowly true, and one response is that users of forecasts may not be suitably trained and therefore require more assistance. But the problem goes deeper, as may be seen by considering the analogous question of choosing where to invest one's money. Suppose one is given a foreacasted rate of return for a possible investment; it is only rarely that one is given the corresponding variance, but let us suppose that we have the variance as well. What are we to make of it? The answer lies in the revelations of portfolio theory (Markovitz 1959 , and in a practical vein, Malkiel 1990), which demonstrates that variance is a measure of risk, and that there is a negative correlation between average reward and the varaince of reward. The same theory allows one to combine average return, variance of return, and level of risk aversion, in a way that allows individuals to compare investment alternatives. This theory also allows one to compare the rationale of decision-making by different individuals. Absent this framework, even the technically trained person is not sure how variance should influence investment choice.

I believe this kind of framework is needed for users of stochastic demographic forecasts, and it lies ready to hand in the theory of decision analysis (Raiffa 1970 , Howard 1971). I will illustrate with an example, a simplified model of the balance in the Social Security Trust Fund. Let this balance at time t be B_t, and let r_i be the interest rate earned on this balance (assumed fixed and known). Further suppose that in periods $t > 0$ (the known initial point is at time $t = 0$) the wage rate (constant dollars per person per year) for age-group a is $w_t(a)$; the social security tax-rate (percentage of earnings) for age-group a is $\tau_t(a)$; and, the rate at which benefits are paid out (constant dollars per person per year) for age-group a is $\beta_t(a)$. We can model the dynamics of B_t by the equations:

$$
\begin{aligned}
I_t &= \sum_a \tau_t(a)\, w_t(a)\, N_t(a), \\
E_t &= \sum_a \beta_t(a)\, N_t(a), \\
B_{t+1} &= I_t - E_t + (1 + r_i)B_t.
\end{aligned}
\tag{13}
$$

If we solve this system formally we find that B_t is a functional of the entire history of population, i.e., it depends on the set of vectors $N_1, N_2, \ldots, N_t$. The demographic uncertainty in these population vectors therefore determines the uncertainty in B_t over time.

Take the point of view of a particular consumer of forecasts, say Congress or the Social Security Administration. The above model is more directly relevant to such users than the population forecasts. Furthermore, their interest is in a particular objective, say the solvency of the Trust Fund, over a particular time horizon, and the key issue about population uncertainty is: how does it affect this objective? One may formulate this concern as a particular stochastic problem: given the model (13), intial conditions N_0, B_0, and a time-course of all relevant rates in the future, what is the probability of the Fund becoming insolvent? Formally, we would compute the probability

$$
\text{Prob}[B_t < b*], \text{ for } t = 1, \ldots, T, \tag{14}
$$

where $b*$ is a "solvency threshold" and T is the time horizon. Now such probabilities would, I believe, be much more useful than a stochastic forecast.

Going one essential step further, one could embed this model into a stochastic decision-making context and provide even richer information to a decision maker. Suppose that we regard the tax-rate as adjustable over time, and that our management objective is to minimize the prospect of insolvency. Then the set $w_t, t = 1, \ldots, T$ is the space of decision variables, with appropriate constraints deriving from legislative and political considerations. One way of providing insight into the way in which alternative policies affect the future is to show how the probabilities of equation (14) vary when adopt alternative decisions. This can be done with a relatively straightforward sensitivity analysis. A more complex exercise would be to use stochastic dynamic programming. For example, define an "optimally safe" decision set as the collection of tax-rates which solves the minimization problem

$$\min_{w_t, t=1,\ldots,T} \text{Prob}[(\min_t B_t) < b*].$$

Solving this problem would provide a reference point in "policy" space against which alternatives could be evaluated. A combination of the preceding calculations would show how sharply defined the optimum is, and thus how much it matters if a suboptimal decision is taken.

It is clear that one can widen the set of objective functions to suit the objectives of the user, and incorporate risk-aversion or more general constraints. The particular analyses to be done can be tailored to specific users. Such an approach will be technically more demanding than the demographic forecasting task, but should be a goal at which at least some demographers direct their efforts. The rewards in terms of the reception, utilization, and value of demographic forecasts promise to be great.

5 Acknowledgements

My work owes much to a collaborative project with Ronald Lee and Larry Carter on forecasting. Jim Vaupel, David Lane, Denis Ahlburg and the Aging group at Minnesota provided useful comments; Marc Feldman provided encouragement and support at Stanford. Support for this work comes from NICHD grant HD-16640 to the author, and from the Morrison Institute for Population and Resource Studies at Stanford.

REFERENCES

Alho, J.M. 1991. Stochastic methods in population forecasting. Int. J. Forecasting. (In Press).

Alho, J.M. and Spencer, B.D. 1990. Error models for official mortality forecasts. J. Am. Stat. Assoc. 85:609-616.

Billingsley, P. 1979. *Probability and measure.* John Wiley: New York.

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

Cohen, J.E. 1986. Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. Demography. 23:105-126.

Cohen, J.E. 1979. Ergodic theorems of demography. Bull. Amer. Math. Soc. N.S. 1:275-295.

Davis, W.W. 1988. Calculation of the variance of population forecasts. United Staates Bureau of the Census, Statistical Division Research Reports, CENSUS/SRD/RR-88/20.

Easterlin, R.A. 1980. *Birth and fortune.* Basic Books.

Howard, R.A. 1971. *Dynamic probabilistic systems, V.2: Semi-Markov and decision processes.* Wiley.

Lee, R.D. and Carter, L. 1992. Modeling and forecasting the time series of U.S. mortality. J. Am. Stat. Assoc. (In Press).

Lee, R.D. and Tuljapurkar, S.D. 1991. Stochastic population projections for the U.S.: Beyond high, medium and low. Paper presented at the 1990 meeting of the Population Association of America.

Malkiel, B.G. 1990. *A random walk down Wall Street.* W.W. Norton: New York.

Manton, K. and Stallard. E. 1984. *Recent tends in mortality analysis.* Academic Press: Orlando.

Manton, K. and Woodbury, M.A. 1985. A continuous-time multivariate

Gaussian stochastic process model of change in discrete and continuous state variables. Pages 277-315 in *Sociological methodology*, N. Tuma, Ed. Jossey-Bass.

Markowitz, H. *Portfolio selection: efficient diversification of investments.* Wiley.

Riffa, H. 1970. *Decision analysis.* Addison-Wesley.

Tuljapurkar, S.D. 1990. *Population dynamics in variable environments.* Springer-Verlag.

Tuljapurkar, S.D. and Orzack, S.H. 1980. Population dynamics in variable environments I. Long-run growth rates and extinction. Theor. Pop. Biol. 18:314-342.