IIASA

How to Deal with Uncertainty in Population Forecasting?

Wolfgang Lutz and Joshua R. Goldstein
Guest Editors

RR-04-009
October 2004

REPRINT
RESEARCH
REPORT

How to Deal with Uncertainty in Population Forecasting?

Wolfgang Lutz
International Institute for Applied Systems Analysis, Laxenburg, Austria

Joshua R. Goldstein
Princeton University, Wallace Hall, Princeton, NJ, USA

Guest Editors

RR-04-009
October 2004

Reprinted from *International Statistical Review*, **72**(1&2):1–106, 157–208 (2004).

International Institute for Applied Systems Analysis • Schlossplatz 1 • A-2361 Laxenburg • Austria
Tel: (+43 2236) 807 • Fax: (+43 2236) 71313 • E-mail: publications@iiasa.ac.at • Web: www.iiasa.ac.at

Reprinted with permission from *International Statistical Review*, **72**(1&2):1–106, 157–208 (2004).

Table of Contents

Introduction: How to Deal with Uncertainty in Population Forecasting? 1
Wolfgang Lutz and Joshua R. Goldstein

Time Series Based Errors and Empirical Errors in Fertility Forecasts in the Nordic Countries 5
Nico Keilman and Dinh Quang Pham

Using the Lee–Carter Method to Forecast Mortality for Populations with Limited Data 19
Nan Li, Ronald Lee and Shripad Tuljapurkar

Mortality Forecasting and Trend Shifts: an Application of the Lee–Carter Model to Swedish Mortality Data 37
Hans Lundström and Jan Qvist

Toward a New Model for Probabilistic Household Forecasts 51
Jiang Leiwen and Brian C. O'Neill

Assumptions on Fertility in Stochastic Population Forecasts 65
Maarten Alders and Joop de Beer

Probabilistic Population Projections for India with Explicit Consideration of the Education-Fertility Link 81
Wolfgang Lutz and Sergei Scherbov

Simpler Probabilistic Population Forecasts: Making Scenarios Work 93
Joshua R. Goldstein

Conditional Probabilistic Population Forecasting 157
Warren C. Sanderson, Sergei Scherbov, Brian C. O'Neill and Wolfgang Lutz

Conditional Probabilistic Population Projections: An Application to Climate Change 167
Brian C. O'Neill

Random Scenario Forecasts Versus Stochastic Forecasts 185
Shripad Tuljapurkar, Ronald D. Lee and Qi Li

Developing Official Stochastic Population Forecasts at the US Census Bureau 201
John F. Long and Frederick W. Hollmann

International Statistical Review (2004), **72**, 1, 1–4, Printed in The Netherlands

Introduction: How to Deal with Uncertainty in Population Forecasting?

Wolfgang Lutz[1] and Joshua R. Goldstein[2]

[1]International Institute for Applied Systems Analysis, Laxenburg, Austria. E-mail: lutz@iiasa.ac.at
[2]Princeton University, Wallace Hall, Princeton, NJ, USA. E-mail: josh@princeton.edu

> *Demographers can no more be held responsible for inaccuracy in forecasting population 20 years ahead than geologists, meteorologists, or economists when they fail to announce earthquakes, cold winters, or depressions 20 years ahead. What we can be held responsible for is warning one another and our public what the error of our estimates is likely to be.—Nathan Keyfitz (1981)*

In the pursuit of our daily life, be it at the individual level or as a society, we are constantly guided by expectations about the future. Typically, these expectations are based on the assumption that the future is going to be more or less the same as what we currently experience or that there are clearly predictable regularities such as day and night and the change of seasons.

In some instances, however, we know with near certainty that things will change fundamentally. In the field of demography, it is virtually certain that European societies will get older and the proportion of the population above age 60 will increase significantly and the proportion below age 20 will shrink. This change that we expect over the coming decades is already embedded in the current age structure of the population. Only extremely unlikely events such as a new disease killing large proportions of the elderly population while leaving the younger unaffected could change this aging trend in the near to mid-term future. This nearly certain population aging will bring fundamental changes to the functioning of our societies and will have significant impacts on individual life course planning, family networks, pension systems and macro-economic development alike. In terms of pensions in a pay-as-you go system younger people today cannot count on the same contributions/benefits ratio as they observe with people retiring today.

Even in case of this highly predictable trend towards population aging the exact extent of aging is rather uncertain, especially in the longer run. Will the share of the population over age 60 in Western Europe increase from currently 20 percent to 29 percent or 43 percent by 2050 (the 80 percent prediction interval according to Lutz *et al.*, 2001). This is a very significant difference and the answer will largely depend on the still uncertain degree of future increases in life expectancy but also on the even more uncertain future trends in fertility and migration.

There is uncertainty in all three components of demographic change (fertility, mortality and migration). How should a forecaster deal with this uncertainty in a statistically consistent manner that is both informative to the users and feasible for the producing agencies? This is the topic of this set of eleven papers that are published in two blocks in this and the next issue of the *International Statistical Review*.

Over the past years a rapidly increasing body of literature has dealt with the issue of uncertainty in population forecasting. The *International Journal of Forecasting* published a special issue in 1992 (Ahlburg & Land); *Population and Development Review* published a special supplement in 1999 (Lutz *et al.*); and most recently the National Research Council (2000) has dealt extensively with uncertainty in its volume *Beyond Six Billion*. This is because population forecasts are important for a large community of users, forming the basis of social, economic, and environmental planning and policy making. The medium variant projections, typically considered to be the most likely forecasts, that have been produced by national and international agencies have played a useful role and have generally achieved impressive accuracy. Yet, it is increasingly recognized that the way these projections deal with the issue of uncertainty is unsatisfactory.

The current practice of providing "high" and "low" variants to communicate uncertainty around the medium projection suffer from several drawbacks. The most important are: (a) In many cases, variants only address fertility uncertainty, ignoring mortality and migration uncertainty; (b) The variants approach is unspecific about the probability range covered by the "high" and "low" variants; (c) The variants are probabilistically inconsistent when aggregating over countries or regions because the chances of extreme outcomes in many countries or regions at once are portrayed as being the same as an extreme outcome in a single country or region; and (d) The variants typically do not allow for temporal fluctuations such as baby booms and busts that can produce bulges in age structure.

Considerable scientific progress has been made in the field of probabilistic population forecasting, including the analysis of past projection errors, the use of expert knowledge and substantive arguments, and the development of stochastic models of fluctuating demographic rates. Several national statistical agencies, such as Norway, the Netherlands, Finland and Austria, have already published fully probabilistic forecasts and more, such as the United States, are planning to do so.

This set of eleven papers covers many of key issues currently discussed in this field. Without discussing the individual papers in this introduction we will only briefly mention some of the key questions the papers address. We end each item with citations to the relevant papers in this collection.

1. **Past projections errors:** The fact that population projections have been carried out for a long time can be used to compare the past projections to actual trends (*ex post* error analysis). This can give an important piece of information as to what can go wrong in population projections. For the future this information can serve as a yard stick of what one would assume to be the minimum error if one wants to be on the safe side, while it is of course problematic to assume that the future errors will be exactly a replication of the past errors. A more difficult challenge is to address the deeper reasons of why forecasters in the past have made certain erroneous assumptions and how we can learn from this for our new forecasts. (Keilman & Pham).
2. **Trends in the components:** Demographic forecasts typically treat uncertainty in each component of population change—fertility, mortality, and migration—separately. An active area of research has been to develop models for projecting each of these components. In mortality, the dominant approach in forecasting has been to rely on formal models extrapolating historical trends. In fertility and migration, where the patterns of determination have been more complex, the role of judgment has been greater. The challenge remains two-fold: on the one hand to incorporate substantive knowledge into formal models, and at the other extreme to formalize the use of substantive knowledge. This also holds for projections that go beyond age and sex through considering, for instance, household size. (Leiwen & O'Neill; Lundström & Qvist; Li *et al.*; and Alders & De Beer).
3. **The role of experts in defining uncertainty:** Experts play a key role in all population forecasting but their tasks depend on the chosen approach. In some instances the role of the expert is limited to choosing the model, including some of its key parameters and the reference data from which the estimates should be drawn. In other instances experts do also make assumptions about the likely future level of the demographic components (such as the average future fertility level) or limit the range of future values (minimum or maximum fertility) while using time series data to define the fluctuations within this range. In still other cases the experts also define the variance. Experts can also go further and include into their considerations some of the structural drivers of future fertility trends such as changes in the educational composition

of the population. (Lutz & Scherbov).

4. **Definition of the temporal process:** Traditional population projections have assumed smooth, monotonic paths for future trends in all three components but most importantly in fertility. This is also mostly the case for high and low fertility variants which assume (piecewise) linear trends to different target fertility levels. Fully stochastic projections on the other hand incorporate annual fluctuations in all three components which seem to better reflect the pattern seen in the real world. An interesting research question is to what degree piece-wise linear scenarios—following the tradition of most statistical agencies—can be seen as a stylized approximation to the results of fully stochastic models with annual fluctuations. There are currently two schools of thought: on the one hand, the two methods cannot produce perfectly identical results and when the variances of the underlying processes differ can produce vastly different results. On the other hand, it does appear that for some purposes the scenario approach can produce results that are close to the fully stochastic approach. Further work on this issue will be important as statistical agencies try to improve their description of uncertainty in forecasts and will influence whether they take a fully stochastic approach or transition incrementally from scenarios. (Tuljapurkar *et al.*; Goldstein).
5. **Conditional probabilistic forecasts:** While probabilistic population projections give a comprehensive description of the uncertainty range they do not provide users with information about the consequences of specific alternative fertility, mortality or migration trends. Especially, in the case of policy analysis, decision makers often want to know what will be the consequences of altering a demographic component by some degree. For this reason, some producers of probabilistic forecasts have also separately produced if-then scenario projections. The concept of conditional probabilistic projections makes this separation unnecessary because it allows the analyst to look at the range of uncertainty in the results conditioning on some specific subset of the full range of stochastic trends. (Sanderson *et al.*; O'Neill).
6. **Implementation by Statistical Agencies:** Official agencies must not only produce forecasts that are scientifically defensible but must also assure that their descriptions of uncertainty can be understood by the broad and heterogeneous set of users. As probabilistic forecasts become more common, agencies can begin to learn from each other's experience, but for now each producer is left to innovate on their own. The failures and successes of these endeavors if communicated appropriately, are likely to provide useful lessons for other countries around the world. (Long & Hollmann).

Not all demographers and statistical agencies are enthusiastically embracing probabilistic population projections for a number of substantive and institutional concerns. On the substantive side one concern raised is that the statement of precise uncertainty ranges (e.g. 95 percent intervals) conveys to the users a misleading sense of precision, as if one would have more detailed information than one actually has. In this context it is important to clearly tell the users that the stated uncertainty ranges should not be seen as precise objective probabilities but rather as indicative ranges depending on the specific model and parameter assumptions made according to the best judgement of the producers. A related substantive concern expressed (Lesthaeghe, 2002) is that probabilistic population forecasts tend to be "too mechanistic" and disregard most substantive scientific knowledge about the determinants of future mortality, fertility, and migration. This reservation does not concern all approaches to probabilistic projections equally. It applies more to models that are exclusively based on the extrapolation of past time series and the application of errors observed with passed projections to estimate future errors than it does to approaches based on substantive argumentation of experts about future trends and the relevant sources of uncertainty. Since this set of papers includes contributions choosing these different approaches the reader can make his/her own judgement about the relevance of this criticism for the approaches concerned.

There are also two reservations of a more institutional nature. Some people claim that probabilistic forecasts are too difficult and too complex to be understood and they point to the fact that most users of forecasts do not even consider the usual high and low variants and would be less likely to use full distributions. In response one can say that clearly for a large group of users there is for good and understandable reasons only interest in one forecast that is considered the most likely one. For the smaller group of users that in their applications do have to worry about the possibility of deviations from this most likely path, however, one can also make the point that they often do not use the high and low variants because they do not really understand what these variants stand for. And because of the above described problems and inconsistencies of the high and low variants there is no way to understand what they stand for. Instead, it seems to us that people interested in uncertainty should either use completely probability-free scenarios for sensitivity testing or fully probabilistic projections for a more comprehensive view of the uncertainty involved.

A final institutional concern is that probabilistic population projections are too difficult to be implemented by statistical agencies that do have a lack of skilled manpower in this field. This is a genuine concern for many countries and international agencies alike. The solution to this problem can come from two sides: on the one hand demographers can try to develop simpler approaches to probabilistic forecasting that as much as possible build on established expertise and procedures in statistical agencies and on the other hand those agencies can involve in continued training in order to be able to apply these increasingly powerful and useful new ways of dealing with uncertainty in population forecasting.

The papers published in these two issues of the *International Statistical Review* have originally been presented at a seminar held in Vienna in December 2002 (sponsored by the Vienna Institute of Demography and the International Institute for Applied Systems Analysis, IIASA), and have been extensively refereed. The participants in this meeting also worked on a consensus statement from which some of the above mentioned points are taken and which concludes with the sentence: "We believe that the quantification of uncertainty will enhance the usefulness of population projections and make the work of forecasting agencies an even more valuable product for planners, policy-makers, scientists, and the public around the world."

References

Ahlburg, D.A. & Land, K.C. (Guest Eds.) (1992). *International Journal of Forecasting*, **8**(3), November 1992.

Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review*, **7**(4), 579–593.

Lutz, W., Sanderson, W. & Scherbov, S. (2001). The end of world population growth. *Nature*, **412**, 543–545.

Lutz, W., Vaupel, J. & Ahlburg, D. (Eds.) (1999). *Frontiers of Population Forecasting*, supplement to volume **24** (1998) of *Population and Development Review*.

Lesthaeghe, R. (2002). Personal communication.

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

International Statistical Review (2004), **72**, 1, 5–18, Printed in The Netherlands

Time Series Based Errors and Empirical Errors in Fertility Forecasts in the Nordic Countries*

Nico Keilman[1,2] and Dinh Quang Pham[2]

[1] *Department of Economics, University of Oslo, Norway. E-mail: nico.keilman@econ.uio.no*
[2] *Statistics Norway, Oslo, Norway*

Summary

We use ARCH time series models to derive model based prediction intervals for the Total Fertility Rate (TFR) in Norway, Sweden, Finland, and Denmark up to 2050. For the short term (5–10 yrs), expected TFR-errors are compared with empirical forecast errors observed in historical population forecasts prepared by the statistical agencies in these countries since 1969. Medium-term and long-term (up to 50 years) errors are compared with error patterns based on so-called naïve forecasts, i.e. forecasts that assume that recently observed TFR-levels also apply for the future.

Key words: Time series; ARCH model; Stochastic population forecast; Total Fertility Rate; Empirical forecast errors; Naïve forecast; Nordic countries.

1 Introduction

Long-term population forecasts, covering a period of fifty years or more, are useful in a number of fields, two of which are analyses of the impact of population trends on contributions and expenditures for old-age pensions, and studies of demographically induced resource use and climate change. Such long-term forecasts are necessarily uncertain: for a given country, one may imagine many possible demographic futures, but some of these population developments are more probable than others. This calls for stochastic population forecasts, i.e. forecasts in terms of prediction intervals. Such prediction intervals quantify uncertainty—they express the expected probability that the future population (or age group, or number of births) falls within a certain range.

A number of recent stochastic population forecasts have used some form of time series analysis for one or more key indicators, in order to assess the expected accuracy of predicted values for these indicators. The most commonly used summary indicator for the level of fertility in a certain year t is the Total Fertility Rate (TFR), defined as the number of children a woman is expected to have over her lifetime, if the age-specific fertility of year t would hold through the woman's childbearing ages, and she would survive to the end of childbearing ages. Time series models were used to predict the TFR in stochastic forecasts prepared for the US (Lee & Tuljapurkar, 1994), Finland (Alho, 1998), the Netherlands (De Beer & Alders, 1999), and Norway (Keilman *et al.*, 2001). One attractive property of time series models is that they not only give a prediction of future values of the variable in question, but also allow us to compute prediction intervals.

*Revised version of a paper presented at the seminar "How to deal with uncertainty in population forecasting?" Vienna, 12–14 December 2002. Comments made by seminar participants are gratefully acknowledged. Research was supported by grant nr. SERD-2000-00172 under the Fifth Framework of the European Commission.

A common finding with TFR-time series in industrialized countries is that these are non-stationary. As a consequence, long run prediction intervals, when unchecked, may become extremely wide. Therefore, adjustments are necessary. For instance, Lee & Tuljapurkar (1994) introduced upper and lower bounds to the TFR by a generalized logit-transformation. This way they constrained TFR-predictions to between 0 and 4 children per woman on average. Alho (1998) found that time-series based TFR-prediction intervals 50 years ahead were 15 per cent wider than those obtained based on the volatility in the historical TFR-observations, and he decided to rely on the latter type of intervals. De Beer & Alders (1999) initially found a 95-per cent prediction interval for the TFR in 2050 equal to [0.6–2.8] based on time series models. Next, an analysis of fertility by birth order led them to suggest that an interval of [1.1–2.3] would be more appropriate. Keilman *et al.* (2001) simulated predicted TFR-values, and rejected TFR-simulations that would fall outside the interval [0.5–4] in any year up to 2050.

Checks of this kind for time series predictions involve judgement. Therefore, additional information will be useful when judging whether the subjective decisions and the adjustments are reasonable. One source of additional information is the accuracy of old TFR-predictions. Many statistical agencies have published population forecasts in the past, including TFR-predictions. When compared with actual TFR-values observed for the years after the forecast was made, one can check the observed TFR-accuracy against predicted TFR-accuracy, obtained by time series models.

In practice however, analyses of this type are restricted to forecasts up to 10 years ahead, seldom longer. This is because few statistical agencies have documented their population forecasts in sufficient detail before the 1960s, and hence the series of historical errors is rather short. However, fertility assumptions in official forecasts, in particular those on the long term, are often close to a recently observed level, which is held constant for the future. Hence, starting from a time series of observed fertility levels, one could compute so-called naïve errors, i.e. errors that result from the naïve assumption that future fertility will be the same as the current one. For a few countries, for example the Nordic countries, long time series for the TFR exist. This allows us to construct rather large data sets with naïve TFR-errors.

The purpose of this paper is to compare the results of Total Fertility Rate predictions made by time series methods with those obtained by analysing both observed forecast errors and naïve errors. The focus is on the expected accuracy of the predictions, i.e. on the width of the prediction intervals. We shall use time series models to derive model-based prediction intervals for the TFR in Norway, Denmark, Sweden, and Finland up to 2050. For the short term and medium term (up to 10 yrs), expected TFR-errors are compared with empirical forecast errors observed in historical population forecasts prepared by the statistical agencies in these countries since 1969. Long-term TFR-errors obtained from the time series models are compared with naïve errors in the TFR. The longest TFR-series we have is that for Finland, which starts in 1776. The Norwegian and Swedish series start in 1845 and 1855, respectively, while observations for Denmark are available from 1911.

2 Time Series Models

Figure 1 plots the Total Fertility Rate for the four countries. The data sources are listed in the Appendix.

The four countries show a similar pattern in the TFR, which reflects the demographic transition, followed by the effects of the economic recession in the 1930s and the baby boom in the 1950s and 1960s. Plots of first differences in the TFR (not shown here) revealed less similarity. Finland exhibits the strongest fluctuations, in particular before 1950. In the 20th century, all four countries show a tendency towards lower variability in the TFR, although this tendency is much stronger for Finland than for the other three countries, see Table 1. Major events, such as the two world wars, and the occurrence of the Spanish Influenza in 1918/1919 are clearly reflected in the series for all four countries.

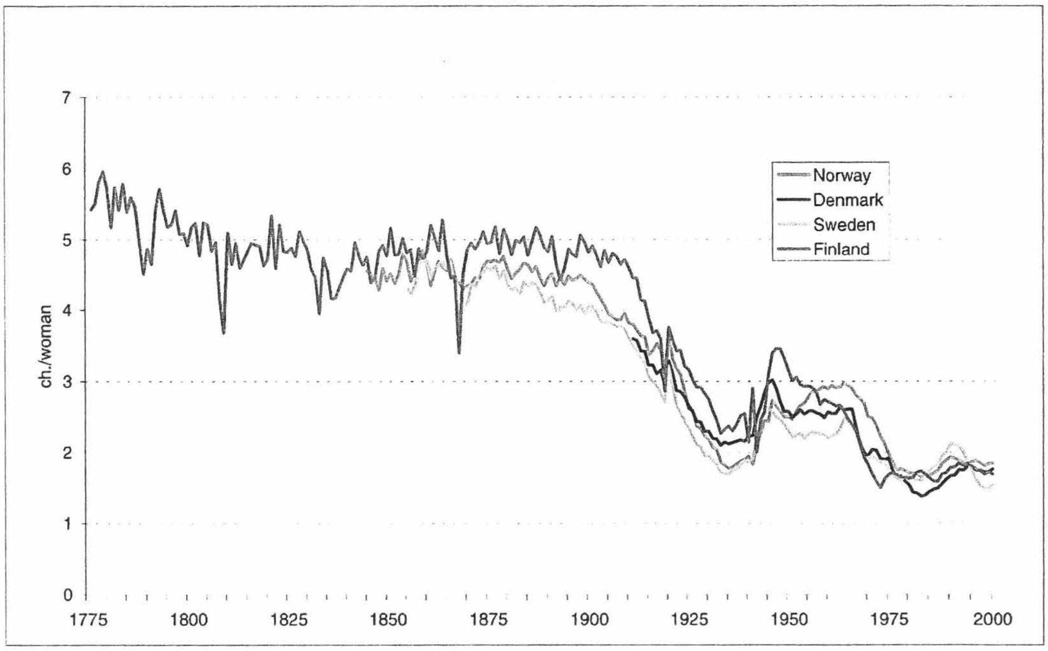


Figure 1. *Total fertility rate.*

Table 1

Standard deviations of the first differences in the TFR for various sub-periods.

	Norway	Denmark	Sweden	Finland
1776–1850	–	–	–	0.346
1851–1900	0.135	–	0.142	0.319
1901–1950	0.138	0.123	0.126	0.288
1951–2000	0.065	0.074	0.073	0.071

An important question is how much of the data should be used in the modelling. Several issues are at stake here. First, Box & Jenkins (1970, p.18) suggest at least 50 observations for ARIMA-type of time series models, although annual models (in contrast to monthly time series) probably need somewhat shorter series. Second, the quality of the data is better for the 20th century than for earlier years. This is particularly true for the denominators of the fertility rates, i.e. the annual numbers of women by single years of age. Third, one may question the relevance of data as long back as the mid-1800s. Current childbearing behaviour is very different from that of women in the 19th century. Fourth, our ultimate goal is to compute long-term predictions of some 50 years ahead, which necessitates a long series.

The ultimate choice is necessarily a subjective one, which includes a good deal of judgement and arbitrariness. We believe that we strike a reasonable balance between conflicting goals by selecting the 20th century as the basis for our models. An analysis based on the last 50 years, say, would be unfortunate: it would include the baby boom of the 1950s and early 1960s, but not the low fertility of the 1930s, to which the boom was a reaction, at least partly. A base period stretching back into the 19th century would be hampered by problems of data quality, and it would also unrealistically assume that the demographic behaviour over such a long period could be captured by one and the same model. In a sensitivity analysis we also experimented with base period 1945–2000. For Norway and Finland we found 95 per cent prediction intervals that were smaller (by 1.4 and 0.5 children per woman on average, respectively) than those that we have accepted for further analysis (see Table 4). For Denmark and Sweden they were larger (by 0.8 and 1.2 children per woman, respectively). To increase comparability across countries, we used a TFR time series starting in 1900 for Denmark as well. Danish TFR-values for the years 1900–1910 were estimated on the basis of observed Crude Birth rate values; see the Appendix.

Traditional time series models of the ARIMA type assume homoscedasticity, i.e. constant residual variance. Given the tendency towards less variability in the TFR in recent decades, such traditional models could not be used. The Autoregressive Conditional Heteroscedastic (ARCH) model introduced in Engle (1982) combines time-varying variance levels with an autoregressive process. This model, and its generalizations (generalized, integrated, and exponential ARCH models, to name a few) have gained popularity in recent years (Bollerslev, 1986). The model has already proven useful in analysing economic phenomena such as inflation rates, volatility in macroeconomic variables, and foreign exchange markets; see Bollerslev (1986) for a review. Application to demographic time series is less widespread. Yet, given the varying levels of volatility in the TFR during the 20th century, an ARCH-type of model is an obvious candidate.

Let Z_t be the logarithm of the TFR in year t. Then the model is

$$\begin{aligned} Z_t &= C + \phi Z_{t-1} + \nu_t + \eta_1 U_{1,t} + \eta_2 U_{2,t} + \eta_3 U_{3,t} + \eta_4 U_{4,t} \\ \nu_t &= \psi_1 \nu_{t-1} + \psi_2 \nu_{t-2} + \cdots + \psi_m \nu_{t-m} + \varepsilon_t \\ \varepsilon_t &= \left(\sqrt{h_t}\right) e_t \\ h_t &= \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \end{aligned} \tag{1}$$

where $e_t \sim N(0, 1)$. This is the AR(m)–ARCH(q) model. The outliers caused by the two world wars and by the Spanish Influenza are handled by between two (Denmark and Sweden) and four (Norway, Finland) dummy variables $U_{i,t}$. In addition we have $\omega > 0$ and $\alpha_i \geq 0$.

The maximum number of terms m included in the autoregressive expression of ν_t was set equal to 10, but few of the ψ-estimates turned out to be significantly different from zero. Similarly, estimates for α_i suggested that the order (q) of the CH-part of the model could be one or two, not higher. This led to the specifications for the AR(m)–ARCH(q) model with dummy variables for the four countries as given in Table 2.

Table 2

ARCH models with dummy variables for the TFR in Norway, Denmark, Sweden, and Finland, 1900–2000.

	Dummy variables	AR:ν_t =	CH:h_t =
Norway	1915, 1919, 1920, 1946	$\psi_2 \nu_{t-2} + \varepsilon_t$	$\omega + \alpha_1 \varepsilon_{t-1}^2$
Denmark	1920, 1942	$\psi_1 \nu_{t-1} + \varepsilon_t$	$\omega + \alpha_1 \varepsilon_{t-1}^2$
Sweden	1920, [1942–45][1]	$\psi_1 \nu_{t-1} + \varepsilon_t$	$\omega + \alpha_1 \varepsilon_{t-1}^2$
Finland	1919, 1920, 1940, 1941	$\psi_2 \nu_{t-2} + \varepsilon_t$	$\omega + \alpha_1 \varepsilon_{t-1}^2$

Note 1: one common dummy variable for each year in the period 1942–45.

Table 3 lists the estimation results, with t-values in parentheses. None of the C-estimates were significantly different from zero. Yet the constant was retained in the model: trial calculations omitting the constant resulted in implausibly low point predictions for the TFR in 2050, ranging from a high 1.38 children per woman in Norway, to a low 1.21 children per woman in Sweden. The predictions for 2050 based on the model with constant are in the interval from 1.71 (Denmark) to 1.46 (Finland).

Note the high α_1-estimate for Finland. It reflects the large variability in the Finnish data. The Swedish α_1-estimate is close to zero—however, omitting the ARCH(1) part from the model (which essentially boils down to assuming constant variance) would lead to rejecting the null hypothesis of constant variance in the residuals, as trial calculations showed.

Table 3

Parameter estimation results for the models in Table 2; t-values in parentheses.

	η_1	η_2	η_3	η_4	C	ϕ	ψ_1	ψ_2	ω	α_1
Norway	-0.0680	-0.0531	0.1830	0.1016	0.0114	0.9754		0.4771	0.0005	0.3626
	(-2.34)	(-2.41)	(8.78)	(2.28)	(0.61)	(48.07)		(5.41)	(5.18)	(2.45)
Denmark	0.1493	0.0901			0.0157	0.9701		0.3249	0.00009	0.2721
	(2.06)	(2.76)			(0.86)	(48.23)		(3.35)	(7.18)	(1.95)
Sweden	0.2258	0.0753			0.0215	0.9588	0.5788		0.0007	0.0000
	(10.40)	(4.51)			(1.15)	(48.69)	(6.53)		(6.01)	(0.00)
Finland	-0.2085	0.2909	-0.1095	0.1563	0.0053	0.9812		0.2135	0.0007	0.7080
	(-2.14)	(1.72)	(-6.43)	(7.46)	(0.47)	(98.03)		(2.38)	(4.24)	(3.48)

The overall impression is that model (1) is a useful device to capture the TFR-trends in the four Nordic countries during the past century. We used the model to compute prediction intervals for the future TFR up to 2050. Since we cannot be certain that the estimated coefficients are equal to the real ones, we used simulation to obtain these intervals. In each of the 5000 simulation runs, parameter values were drawn from a multivariate normal distribution, with expectation equal to the parameter estimates in Table 3, and with corresponding covariance matrix as estimated earlier. The possibility that a pandemic as bad as the Spanish Flu, or a war with consequences as catastrophic as WWI or WWII could occur during the prediction period, was included in these simulations. For each of the two dummy variables, we first drew a random number from the binomial distribution with a probability of "catastrophe" equal to 1/101. Next, the starting year for the catastrophe was determined on the basis of a random draw from the uniform distribution on the interval [2001, 2050]. Finally, the appropriate η-value was drawn from its estimated distribution.

The estimates for the AR-coefficient ϕ are all between 0.96 and 0.98. Although those estimates were extremely sharp (cfr. the corresponding t-values), the simulations resulted in a few ϕ-values equal to or larger than one. These were rejected, and ϕ was redrawn until 5000 admissible values had been obtained for each country.

Table 4 lists point forecasts (i.e. expected values) and prediction intervals for selected years up to 2050. Long-range 95 per cent prediction intervals are roughly two children per woman wide. Ten years ahead 95 per cent intervals are 1.1 to 1.2 children per woman wide.

The time-series model predicts an expected TFR-value up to 2050 around 1.7 children per woman, except for Finland, where the prediction is a low 1.5. Although this is not an impossible value, it is much lower than that in official population forecasts for Finland, for instance Statistics Finland (1.75, see Council of Europe, 2001) or the United Nations (1.85, see `http : //www.un.org/esa/population/publications/wpp2002` /wpp2002annextables.PDF). In this paper, we focus primarily on the width of the predictive distribution, much less on its central tendency. In an actual fertility prediction, all long-term point predictions would have to be examined critically, for example by inspecting the mean number of children born to women in successive birth generations.

Are the long-term prediction intervals in Table 4 reasonable? One may object that the interval

Table 4

Simulated TFR-point forecasts and prediction intervals.

	forecast	67% interval lower bound	67% interval upper bound	67% interval interval width	95% interval lower bound	95% interval upper bound	95% interval interval width
				Norway			
2000	1.85						
2005	1.86	1.72	2.02	0.30	1.55	2.21	0.76
2010	1.83	1.61	2.10	0.49	1.37	2.48	1.11
2015	1.80	1.53	2.16	0.63	1.23	2.63	1.40
2035	1.72	1.33	2.23	0.90	0.90	2.84	1.94
2050	1.68	1.24	2.25	1.01	0.77	2.96	2.19
				Denmark			
2000	1.77						
2005	1.78	1.63	1.97	0.34	1.45	2.18	0.73
2010	1.77	1.54	2.06	0.52	1.32	2.40	1.08
2015	1.76	1.48	2.10	0.62	1.21	2.53	1.32
2035	1.73	1.34	2.19	0.85	0.94	2.73	1.79
2050	1.71	1.27	2.21	0.94	0.82	2.82	2.00
				Sweden			
2000	1.54						
2005	1.61	1.44	1.79	0.35	1.29	2.03	0.74
2010	1.62	1.37	1.92	0.55	1.12	2.30	1.18
2015	1.64	1.32	2.00	0.68	1.02	2.49	1.47
2035	1.66	1.19	2.16	0.97	0.76	2.80	2.04
2050	1.67	1.16	2.18	1.12	0.69	2.83	2.14
				Finland			
2000	1.70						
2005	1.67	1.53	1.83	0.30	1.34	2.07	0.73
2010	1.63	1.42	1.89	0.47	1.17	2.28	1.11
2015	1.60	1.33	1.93	0.60	1.03	2.45	1.42
2035	1.51	1.12	2.02	0.90	0.73	2.82	2.09
2050	1.46	1.02	2.03	1.01	0.60	2.83	2.23

bounds are rather wide. The Norwegian model predicts a 16 per cent chance of a TFR in 2050 of at least 2.2 children per woman, which is much higher than the European average of 1.4, but close to the current fertility level of developing countries such as Brazil, Lebanon, or Tunisia. The lower 95 per cent bound in Finland in 2050 is 0.6 children per woman. Both upper and lower bounds are incompatible with current demographic theories for industrialized countries, one may argue. However, these objections are unconvincing, in our view, for three reasons. First, fertility theories have poor predictive performance. Keyfitz (1982) assessed various theories, and concluded that they have limited predictive validity in space and time, are strongly conditional, or cannot be applied without the difficult prediction of non-demographic factors. Fertility theories developed in more recent years do not fare much better (Lee, 1997, p.52; Van de Kaa, 1996, p.390). This situation is not specific for demography. Any generalizations about human behaviour are bound to be narrowly restricted to specific institutional settings or particular epochs (Nagel, 1961; Boudon, 1986; Henry, 1987). Second, research on forecasting in other fields shows that subject matter experts often are too confident, in the sense that they tend to give prediction intervals that are too narrow (Armstrong, 1985). We have no reasons to believe that the situation in demographic forecasting is very different. Third, the interval *bounds* should not be taken as *probable* fertility levels. On the contrary, the 95 per cent bounds merely indicate the *outliers* of predicted fertility paths. Even the 67 per cent bounds reflect TFR-intervals with chances equal to only one in six.

3 Other Time Series Models

In order to assess the robustness of the prediction intervals obtained in Section 2, we have experimented with several other time series models for Z_t:

- a pure AR(m)-model
- an AR(m)–CH(1) model
- an AR(m)-model with dummy variables.

The results can be summarized as follows.

Fitting an AR(m)-model or an AR(m)–CH(1) model

A purely autoregressive model for Z_t (with maximum lags equal to 5 for Norway, 2 for Denmark, 1 for Sweden, and 2 for Finland) indeed indicated non-constant variance: using a Portmanteau Q-test and a Lagrange Multiplier (LM-) test, a hypothesis of homoscedastic residuals had to be rejected at the five per cent level for Norway, Sweden, and Finland. For Denmark, such a hypothesis was not rejected at the ten per cent level. When we introduced a CH(1)-part to the model in order to account for heteroscedastic residuals, the situation improved considerably for Denmark and Sweden, but for Norway and Finland, there were still some signs of heteroscedasticity at lag 1. A Kolmogorov–Smirnov test hypothesis for normality could only be accepted for Norway and Denmark (at the five per cent level), not for the other two countries.

Fitting an AR(m)-model with dummy variables

Can the non-constant residual variance be captured by introducing dummy variables to the AR(m)-model? This is the case for three countries: Finland was the exception. For the other three countries, dummy variables for the periods around 1918, 1944, and 1970 were introduced. A hypothesis of homoscedastic residuals at all lags could not be rejected (5%) for all three countries. The 95% prediction intervals in 2050 turned out to be 2.7 (Norway), 3.0 (Denmark), and 3.5 (Sweden) children per woman wide—much wider than the intervals in Table 4.

Based on these sensitivity tests we conclude that the ARCH-model in expression (1) gives a useful and reliable description of the development in the TFR in the four countries in the previous century. Next it remains to be seen whether the prediction intervals of this model are reasonable.

4 Errors in Historical TFR Forecasts

Prediction intervals determine the *expected* errors in the *current* forecast. Investigating *observed* errors in *historical* forecasts could provide an independent check of the expected errors. We have analysed the errors in historical TFR-forecasts prepared by statistical agencies in the four countries. Bibliographic details are given in the appendix. The help of Timo Nikander and Ossi Honkanen in Helsinki, Anna Qvist in Copenhagen, Jan Qvist in Stockholm, and Åke Nilsson in Ørebro is gratefully acknowledged.

For Norway, we have TFR-forecasts starting in 1969, for Denmark the series starts in 1974, and for Sweden and Finland the starting year is 1972. Assumed TFR-values for each forecast from the jump-off year until 2000 were compared with observed values (Council of Europe, 2001). Most forecasts had more than one fertility variant, often two, or three. In that case, we included all variants in the data, because very few of the forecast reports contained a clear advice as to which of the variants the statistical agency considered as the most probable one at the time of publication. Hence, it was left to the user to select one of the variants. We may assume that all variants have been used, although the middle one probably more often than the high or the low one (in case there were three variants). Below we shall use the standard deviation of the observed TFR-error. For Norway, the error patterns in this indicator based on all forecast variants were very close to those based on main variants only. For Sweden, the all-variants standard deviations were approximately 10 per cent higher than those based on main variants. We have 31 series of TFR-forecasts for Norway, 51 for Denmark, 23 for Sweden, and 34 for Finland. Each series was ordered by forecast duration, with the jump-off year defined as duration 0.

Since TFR-forecasts may be higher or lower than observed TFR-values, we have used the *signed error* of the TFR-forecasts, defined as the assumed minus the observed value. Therefore, a positive (negative) error indicates a value that is too high (too low). The purpose of the current analysis is to check the uncertainty in model-based TFR-predictions, in other words, the width of the model-based prediction intervals. Thus, for each forecast duration, we computed the standard deviation of the signed TFR-errors, as a measure of uncertainty. The empirical standard deviation reflects uncertainty in the future TFR appropriately, provided that the expected value of the TFR is predicted correctly. The latter assumption may be relaxed by inspecting the Root Mean Squared Error (RMSE) of the TFR, which adds a bias component to the standard deviation (Maddala, 1977). The empirical RMSE's for the four countries diverged only slowly from the standard deviations, and hence for the cases of Norway, Denmark, and Sweden they declined for long durations, too (although the decline set in a few years later than that in the standard deviation). Figure 2 plots these standard deviations.

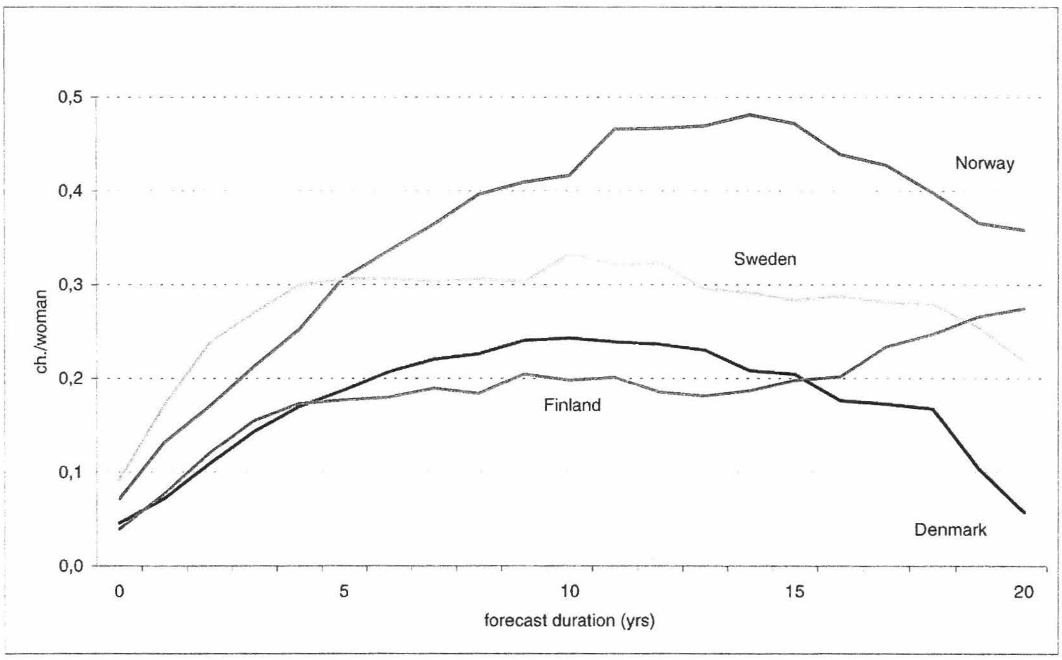


Figure 2. *Standard deviations of signed TFR-errors.*

Figure 2 shows that for the first years of the forecast, the standard deviations increase regularly, from less than 0.1 children per woman in the first forecast year (duration 0), to 0.2–0.3 children per woman five years ahead. Next, the patterns tend to stabilize (quite soon for Sweden, but at 10–15 years ahead also for Denmark and Norway). This stabilization, however, should not be interpreted as an indication that the uncertainty in TFR-predictions prepared since 1969 was roughly constant for durations longer than five years ahead. The stable pattern for durations beyond five years ahead in Figure 2 is a consequence of the fact that the data set is right-censored. The year 2000 was the last year for which TFR-errors were computed. Hence, there are very few errors for long forecast durations. For instance, 15 years ahead we have only 18 observations for Norway, 13 for Denmark, 11 for Sweden, and 16 for Finland. Therefore, the long-term standard deviations in Figure 2 are not very reliable. At the same time, the errors at long durations all apply to the relatively stable period of the last ten years—see Figure 1.

Assuming that the errors for a given duration and a particular country are approximately normally distributed, we can construct a confidence interval for the estimated standard deviation of that distribution. Table 5 gives these intervals, which were obtained on the basis of a χ^2-distribution with degrees of freedom equal to the number of observations minus 1.

Table 5

Estimated standard deviations in TFR-errors for forecast durations of 5 and 10 years, with corresponding 95-per cent confidence intervals in parentheses.

	5 years ahead	10 years ahead
Norway	0.308 (0.240–0.428)	0.417 (0.320–0.595)
Denmark	0.188 (0.152–0.245)	0.243 (0.187–0.348)
Sweden	0.307 (0.237–0.434)	0.333 (0.248–0.507)
Finland	0.177 (0.142–0.236)	0.198 (0.150–0.293)

We note that for forecasts five years ahead, errors for Norway and Sweden have considerably higher standard deviations than those for Denmark and Finland, even when estimation uncertainty is taken into account. This is probably due to the steep fall in the Norwegian TFR in the early 1970s, and the strong fluctuations in the Swedish TFR around 1990, see Figure 1. These factors made it rather difficult to prepare an accurate TFR-forecast, at least more difficult than for Denmark and Finland. The estimates and confidence intervals 10 years ahead still reflect this to a certain extent, although the (unknown) theoretical standard deviations of Denmark and Sweden are probably not very different from each other.

5 Errors in Naïve TFR-forecasts

Inspection of historical official population forecasts shows that in many cases, forecasters in effect have assumed that the TFR would remain relatively close to its current level, even in the long run; see, for example, Lee (1974) for the US, Keilman (1990) for the Netherlands, and Texmon (1992) for Norway. This observation led Alho (1990) to propose inferring prediction intervals for the TFR from the errors of what he called naïve forecasts, i.e. forecasts based on the naïve assumption that current levels will persist over a long time. Clearly, in times of falling fertility, demographers will extrapolate the decline for at least a few years, but the examples mentioned above demonstrate that a constant level is assumed quite soon.

We have computed errors in the naïve TFR forecasts for Norway, Sweden, and Finland for the years 1900–2000, and for Denmark for the period 1911–2000. Hence the naïve TFR forecasts for the first three countries are based on observed TFR levels for the years 1899–1999 (duration 0), 1888–1988 (duration 1), 1887–1987 (duration 2), etc. This means that there are 101 errors for forecast durations 0–50 years for Norway and Finland, and for durations 0–46 years for Sweden. For durations 47, 48, 49, and 50 years there are 100, 99, 98, and 97 naïve errors for the latter country. For Denmark, the error series are much shorter, in particular for longer forecast durations: the length declines regularly

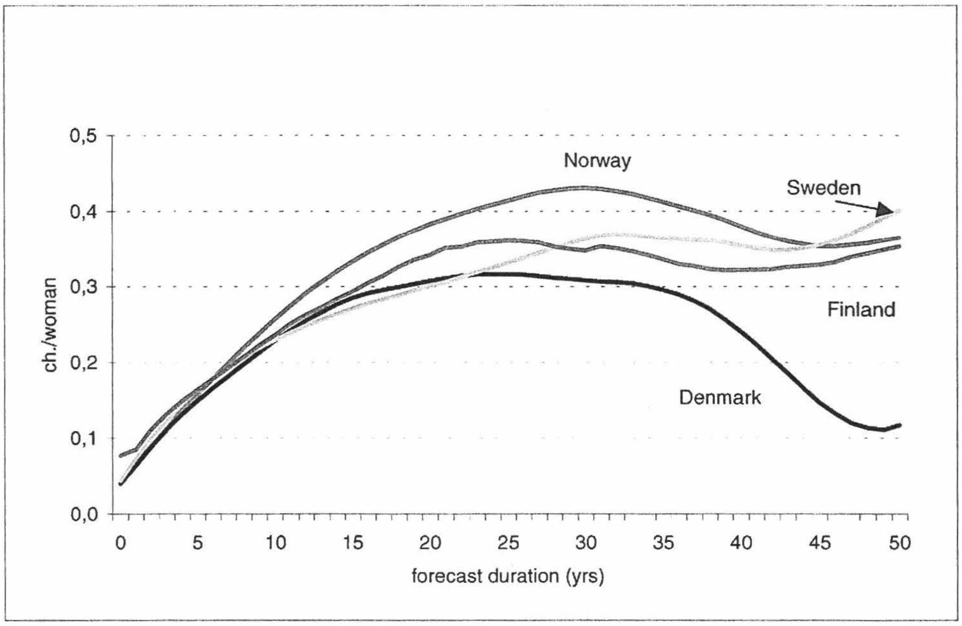


Figure 3. *Standard deviation of signed naïve TFR-errors.*

from 89 errors for duration 0 to 39 for duration 50. Figure 3 plots the standard deviations of these errors. We computed the errors in the log scale, but transformed the results back to the original scale.

The Danish standard deviation falls for forecast durations longer than 20 years, because the long-term errors apply to the period after World War II. Although this period included the baby boom and bust, the volatility in the TFR measured this way was less than that for the whole 20th century. The other three countries show tendencies of the same pattern, but much less extreme than the curve for Denmark. The four curves coincide quite well up to 20 years ahead. As to the long-term behaviour, Figure 3 suggests (ignoring the curve for Denmark) that standard deviations 50 years into the future are around 0.4 children per woman.

We tested an assumption of normally distributed naïve TFR-errors at forecast durations 35 and 50 years ahead for Norway, Sweden and Finland. This assumption had to be rejected (5%) for Norway at a duration of 50 years, but not for the other five cases. Thus for the latter cases we can infer the width of the 67 per cent prediction interval as twice the standard deviation. For the case of Norway 50 years ahead, we have to be more cautious. Not knowing the distribution, Chebyshev's inequality tells us that at least a share equal to $(1 - 1/k^2)$ of probability mass is covered by a $k\sigma$-interval (centered around the mean). Hence the width of the 67-percent prediction interval in this case is estimated as *at least* $\sqrt{3}$ times the standard deviation.

6 Comparison of Model-based and Empirical Prediction Intervals

Figure 4 facilitates a comparison of the expected width of prediction intervals from the three sources: time-series models, errors in historical forecasts, and naïve errors. We have omitted long-term naïve errors for Denmark for reasons mentioned in Section 5.

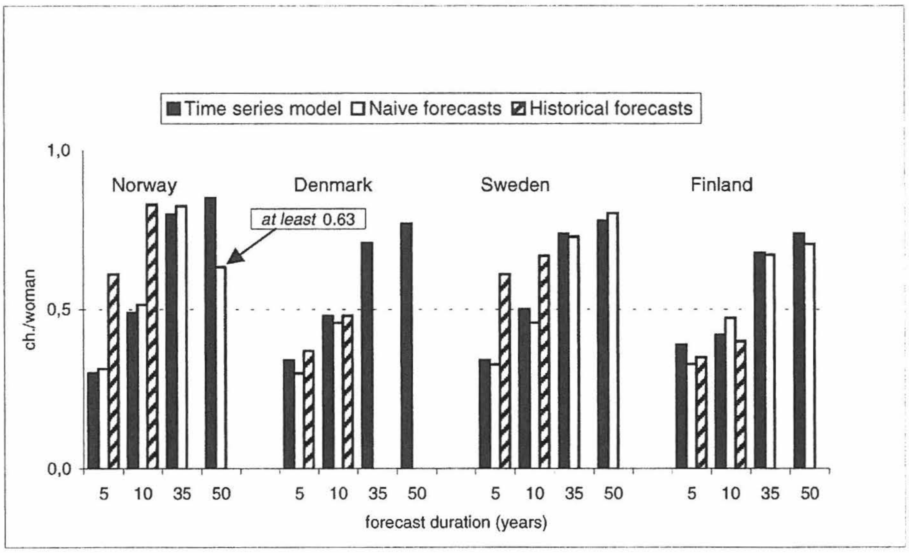


Figure 4. *Width of 67 per cent prediction intervals for the TFR.*

Short term

At a forecast duration of 5 years ahead, the model-based 67 per cent prediction intervals for Denmark and Finland agree quite well with those based on historical errors: in both countries the intervals are 0.3–0.4 children per woman wide. For Norway and Sweden the historical errors are much larger than those stemming from time-series models. One reason may be that the historical errors are not normally distributed. We have relatively few observations (23 for Sweden and 25 for Norway), and therefore the normality assumption could not be tested. In any case, one has to be cautious. The naïve intervals are very close to the model-based ones.

When we consider ten-year ahead forecasts, the number of empirically observed TFR-errors becomes even less (ranging from 17 to 22 for the four countries). With due caution one may conclude that the intervals are roughly 0.5 children per woman wide, primarily based on the time-series models and the naïve errors, but (for the cases of Denmark and Finland) supported by the historical errors.

Medium term

At the medium term, at 35 years into the future, historical errors cannot be used. We note that there is rather good agreement between model-based errors and naïve errors for Norway, Sweden, and Finland. The model-based intervals are approximately 0.7–0.8 children per woman wide for these three countries.

Long term

At 50 years into the future, the agreement between model-based intervals and naïve intervals is rather good for the three countries for which we could do the comparison. Figure 4 suggests 67 per cent intervals that are approximately 0.75–0.85 children per woman wide for Norway, Sweden, and Finland.

7 Conclusion

In this paper we have shown how time series methods may be combined with information from historical and naïve forecasts when one wants to construct long-term prediction intervals for the Total Fertility Rate (TFR). We compared the expected accuracy of three types of TFR-predictions for Norway, Denmark, Sweden, and Finland. Time-series models were fitted to data for the period 1900–2000. These resulted in model-based prediction intervals for the TFR up to 2050. For the short term (5–10 yrs), we analysed empirical forecast errors observed in historical population forecasts prepared by the statistical agencies in these countries after 1969. We also analysed medium-term and long-term (up to 50 years) error patterns based on so-called naïve forecasts, i.e. forecasts that assume that recently observed TFR-levels apply for the future.

For the short term, model-based intervals and those derived from historical errors tend to be of the same order of magnitude, although we have to be cautious with historical errors, because our data set is rather limited. Naïve errors provide useful information both for the short and the long run. Indeed, model-based intervals 50 years ahead agree quite well with naïve errors-based intervals, except for Denmark. For the latter country, the data set did not allow us to compute reliable naïve error patterns for forecast periods beyond 20 years. In general, one may conclude that historical TFR-errors and naïve errors do not indicate that model-based prediction intervals are excessively wide. We found 67 per cent intervals that are approximately 0.5 children per woman wide for a forecast horizon of 10 years, and roughly 0.85 children per woman wide 50 years ahead.

The three types of data sets show us that we have to be modest when we try to predict the accuracy of our TFR-forecasts, at least when we assume

- the same variability in the future TFR as that in the past 100 years (ARCH-model)
- the same short-term forecastability as in the past 30 years (historical errors)
- the same long-term forecastability as in the past 90–150 years (naïve errors).

Some scholars may react that the model-based intervals are too wide and thus reflect too much uncertainty in the TFR-predictions. Our analysis does not give statistical evidence for such a conclusion. If, nonetheless, one believes (for instance, on judgemental grounds) that the expected accuracy of TFR-predictions is better than what time series models, historical forecasts, and naïve forecasts show, the results presented in this paper can serve as a useful benchmark in the discussion.

References

Alho, J. (1990). Stochastic methods in population forecasting. *Internat. J. Forecasting*, **6**, 521–530.
Alho, J. (1998). *A stochastic forecast of the population of Finland.* Reviews 1998/4. Helsinki: Statistics Finland.
Armstrong, J. (1985). *Long-range forecasting: From crystal ball to computer.* New York: Wiley (2nd ed.).
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, **31**, 307–327.
Boudon, R. (1986). *Theories of social change: A critical appraisal.* Cambridge: Polity Press.
Box, G.E.P. & Jenkins, G.M. (1970). *Time series analysis: Forecasting and control.* San Francisco: Holden Day.
Brunborg, H. & Mamelund, S.-E. (1994). *Kohort- og periodefruktbarhet i Norge 1820–1999* (Cohort and Period Fertility for Norway 1820–1993). Report no. 94/27. Oslo: Statistics Norway.
Chesnais, J.-C. (1992). *The Demographic transition: Stages, patterns, and economic implications.* Oxford: Clarendon Press.
Council of Europe (2001). Recent demographic developments in Europe 2001. Strasbourg: Council of Europe Publishing.
De Beer, J. & Alders, M. (1999). Probabilistic population and household forecasts for the Netherlands. Paper ECE-Eurostat Work Session on Demographic Projections, Perugia, Italy, 3–7 May 1999. Internet `www.unece.org/stats/documents/1999.05.projections.htm`
Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987–1008.
Henry, L. (1987). Perspectives et prévision. In *Les projections démographiques. Actes du VIIIe Colloque National de Démographie.* Tome 1. Paris: Presses Universitaires de France (Travaux et Documents Cahier no. 116).
Keilman, N. (1990). *Uncertainty in national population forecasting: Issues, backgrounds, analyses, recommendations.* Amsterdam: Swets and Zeitlinger.
Keilman, N., Pham, D.Q. & Hetland, A. (2001). *Norway's uncertain demographic future.* Social and Economic Studies 105. Oslo: Statistics Norway. Internet `www.ssb.no/english/subjects/02/03/sos105_en`.

Keyfitz, N. (1982). Can knowledge improve forecasts? *Population and Development Review*, **8**, 729–751.
Lee, R. (1974). Forecasting births in post-transition populations: Stochastic renewal with serially correlated fertility. *J. Amer. Statist. Assoc.*, **69**, 607–617.
Lee, R. (1997). History of demography in the U.S. since 1945. In *Les contours de la démographie au seuil du XXIe siècle.*, Eds. J.-C. Chasteland and L. Roussel, pp. 31–56. Paris: Institut National d'Etudes Démographiques, Presses Universitaires de France.
Lee, R. & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *J. Amer. Statist. Assoc.*, **89**, 1175–1189.
Maddala, G. (1977). *Econometrics*. Auckland etc., McGraw-Hill International Book Company.
Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanations.* New York: Harcourt, Brace and World.
Statistics Norway (1994). *Framskriving av folkemengden 1993–2050* (Population Projections 1993–2050). Oslo/Kongsvinger: Statistics Norway.
Statistics Norway (1997). *Framskriving av folkemengden 1996–2050* (Population Projections 1996–2050). Oslo/Kongsvinger: Statistics Norway.
Statistics Norway (1999). Fortsatt befolkningsvekst. Befolkningsframskrivinger. Nasjonal og regionale tall, 1999–2050 (Population growth continues. Population forecasts: national and regional results 1999–2050). *Dagens statistikk*, 17.11.1999. Internet `http : //www.ssb.no/folkfram`.
Texmon, I. (1992). Norske befolkningsframskrivinger 1969–1990 (Norwegian population projections 1969–1990). In *Mennesker og Modeller*, Eds. O. Ljones, B. Moen, and L. Østby, pp. 285–311. Oslo/Kongsvinger: Statistics Norway.
Turpeinen, O. (1979). Fertility and mortality in Finland since 1750. *Population Studies*, **33**(1), 101–114.
United Nations. Internet `http : //www.un.org/esa/population/publications/wpp2002/`wpp2002annextables.PDF.
Van de Kaa, D. (1996). Anchored narratives: The story and findings of half a century of research into the determinants of fertility. *Population Studies*, **50**(3), 389–432.

Résumé

Nous avons construit un modèle de séries temporelles du type ARCH pour calculer des intervalles de prédiction pour l'Indice Synthétique de Fécondité (ISF) pour la Norvège, la Suède, la Finlande, et le Danemark jusqu'à l'année 2050. Pour le court terme (5–10 ans dans le futur), on compare les erreurs attendues pour l'ISF avec les erreurs calculées dans des prévisions démographiques historiques, préparées par des bureaux de statistique dans ces pays depuis 1969. Les erreurs à moyen terme et long terme (jusqu'à 50 ans dans le futur), sont comparées avec des structures d'erreur fondée sur des prévisions dites "naïves"— c'est-à-dire, des prévisions qui supposent que le niveau d'ISF observé pour une période récente est valable aussi pour le futur.

À court terme, nous trouvons que les intervalles de prédiction calculés par le modèle de séries temporelles et ceux dérivés des erreurs historiques sont du même ordre d'amplitude. Cependant, il faut être prudent, car la collecte des données de base historiques est limitée. Les erreurs "naïves" fournissent de l'information utile pour le court terme et le long terme. En effet, des intervalles de prédiction fondés sur des erreurs naïves à 50 ans dans le futur se comparent très bien avec des intervalles fondés sur le modèle de séries temporelles, sauf pour le Danemark. Pour ce pays, les données de base ne nous permettent pas de calculer des intervalles "naïfs" pour des périodes de prévision au-delà de 20 ans. En général, on peut conclure que les erreurs historiques et les erreurs naïves ne montrent pas que les intervalles de prédiction fondés sur des modèles de séries temporelles du type ARCH sont excessivement larges. Nous avons constaté que les intervalles à 67 pour cent de l'ISF ont une amplitude d'environ 0.5 enfants par femme à l'horizon de 10 ans, et approximativement 0.85 enfants par femme à 50 ans.

8 Appendix: Data Sources

Observed TFR

The Council of Europe (2001) publishes annual TFR-values for Norway, Denmark, Sweden, and Finland for the years 1960–2000.
Chesnais (1992, Appendix 2) tabulates annual observed TFR-values for Sweden since 1855 and for Denmark starting in 1911.
Brunborg & Mamelund (1994, Table 1a) give annual TFR-values starting in 1845.
Turpeinen (1979, p.112) gives Finnish TFR-values for each year in the period 1776–1925. Those for later years are taken from the Statistical Yearbook of Finland.

For Denmark, Chesnais tabulates annual TFR-values starting in 1911, and isolated data points for 1903 and 1908. On the other hand, his Table A1.5 lists pre-1911 values for the Danish Crude Birth Rate (CBR). We have estimated TFR-values for the years 1900–1910 based on a linear regression between TFR and CBR. A plot of TFR and CBR showed a near linear relationship for the years 1911–1940. We have used this relationship and "backcasted" the Danish TFR for the years 1900–

1910. This way, we obtained the values and 95 per cent prediction intervals as given in Table A1. The TFR in Figure 1 and its first difference do not reveal any apparent anomalies in the predictions for the years 1900–1910.

Table A1

Backcasts and 95% prediction interval bounds for the TFR of Denmark, 1900–1910.

	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910
Prediction	4.07	4.07	3.99	3.91	3.94	3.86	3.88	3.83	3.89	3.83	3.71
L95	4.01	4.01	3.93	3.85	3.88	3.80	3.81	3.77	3.83	3.77	3.65
U95	4.12	4.12	4.04	3.96	3.99	3.91	3.93	3.88	3.94	3.88	3.77

Chesnais (1992, p.545) reports a TFR in 1908 equal to 3.84, and in 1903 equal to 4.04. Our estimate for 1908 is close to Chesnais' value, while the 1903-estimate is somewhat lower.

TFR Predictions in Official Population Forecasts

Norway

We have updated the data originally assembled by Texmon (1992), who collected, among others, TFR errors for the forecasts of 1969–1987 during the years 1969–1989. Updates were made based on Statistics Norway (1994, 1997, 1999).

Denmark

Various issues of "Befolkning og valg" published by Danmarks Statistik, viz.1983:15, 1984:20, 1985:12, 1986:16, 1988:13, 1989:12, 1990:17, 1992:2, 1992:16, 1993:13, 1994:12, 1995:15, 1996:14, 1997:12, 1998:15, 1999:11, 2000:11, 2001:11. For earlier forecasts we consulted "Statistiske Efterretninger" issues 1978 A39, 1980 A7, 1981 A40, and 1982 A33, and "Statistiske Undersøgelser" nr. 33 published in 1975.

Sweden

Jan Qvist and Åke Nilsson (personal communications) of Statistics Sweden provided us with assumed TFR-values in forecasts prepared by Statistics Sweden starting in 1972. Most of these values can also be found in the following reports published by Statistics Sweden:
"Information i prognosfrågor" issues 1972:5, 1973:6, 1974:7, 1975:6, 1976:3, 1978:5, 1983:2. Later forecasts have been published in "Demografiska rapporter" 1986, 1989:1, 1991:1, 1994:3, and 2000:1. Forecasts with base years 1997, 1998, 1999 have been published at Statistics Sweden's home page, and in Statistical Yearbooks of Sweden issues 1998, 1999, 2000.

Finland

From Ossi Honkanen of Statistics Finland (personal communication) we received assumed TFR-values in official Finnish population forecasts starting in 1972. Most of these values were annual. In some cases, we interpolated linearly between values given for distinct calendar years. Most of these values can also be found in the following reports published by Statistics Finland:
"Statistisk rapport" issues VÄ 1972:7, VÄ 1973:6, VÄ 1975:12, VÄ 1978:17, VÄ 1982:5, VÄ 1982:5, and VÄ 1985:10; "Statistiska meddelanden" 49:1972; "Statistical surveys" issues 52:1974, 64:1979, 70:1983; "Muistio" 91:1984; "SVT/OSF Population" issues 1989:3, 1992:6, 1993:10, 1995:9, and 1998:6.

[*Received December 2002, accepted November 2003*]

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

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

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 var$(\widehat{c})$, and therefore there is a standard-normal random variable η that makes

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

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,

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

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.

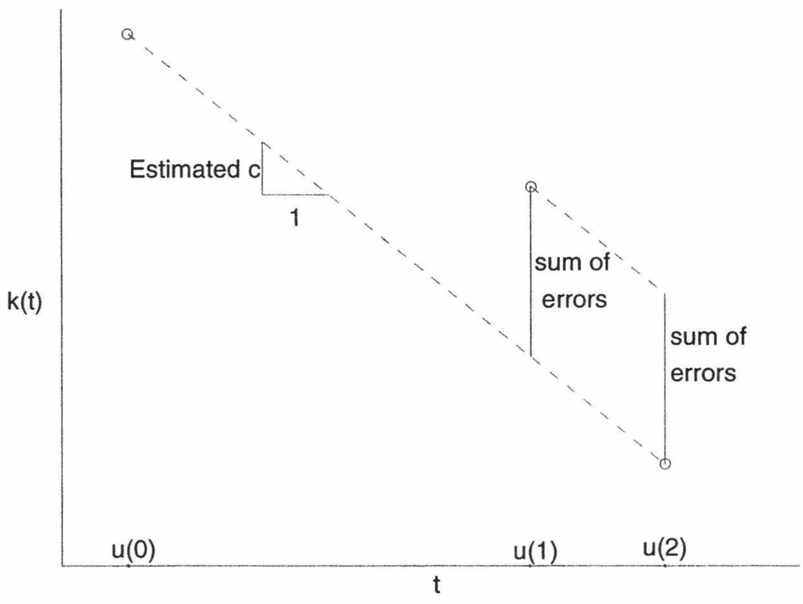


Figure 1. *Linear trend and errors.*

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

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

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 $\text{var}(\widehat{\sigma})$, 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, $\text{var}(\widehat{\sigma})$ is described by

(14a) in appendix as

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

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

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

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

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

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$[\widehat{c} - \widehat{\sigma}\frac{1-\mathrm{re}(\widehat{\sigma})\theta}{\sqrt{u(T)-u(0)}}\eta] = c$. 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.

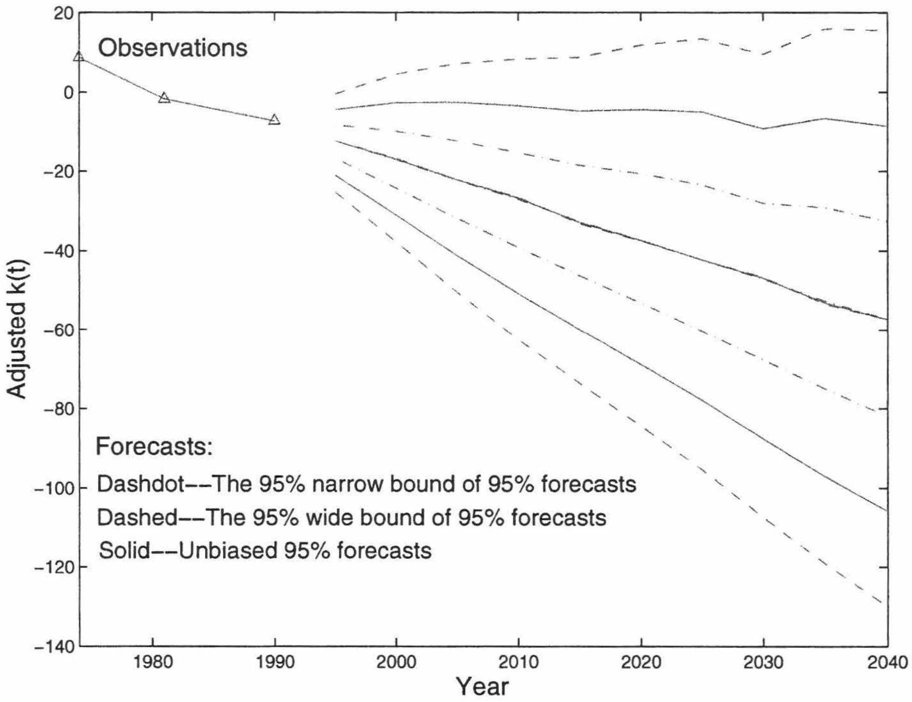


Figure 2. *Observations and 95% forecasts of $k(t)$ of China.*

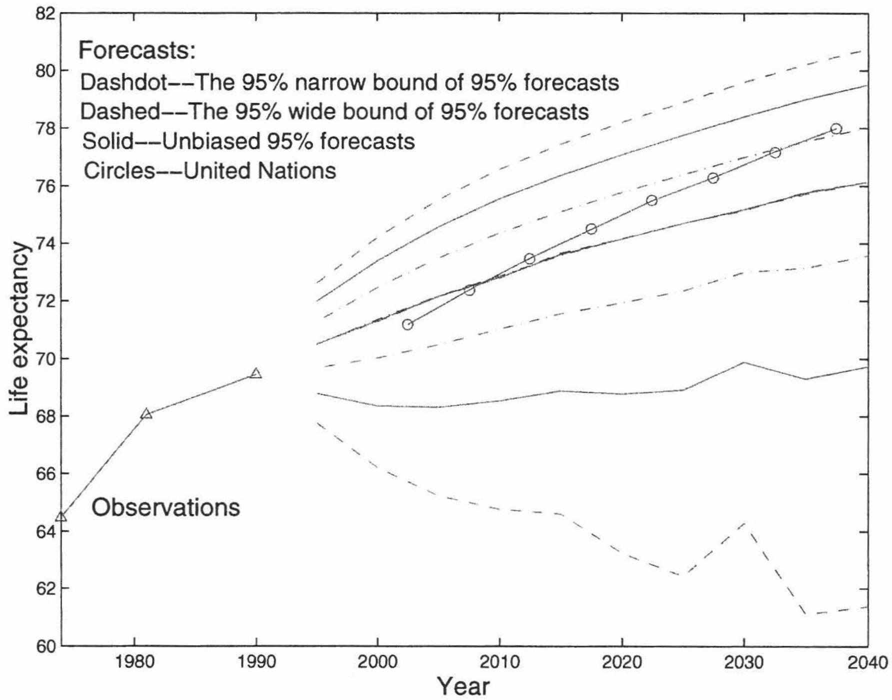


Figure 3. *Observations and 95% forecasts of life expectancy of China.*

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 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

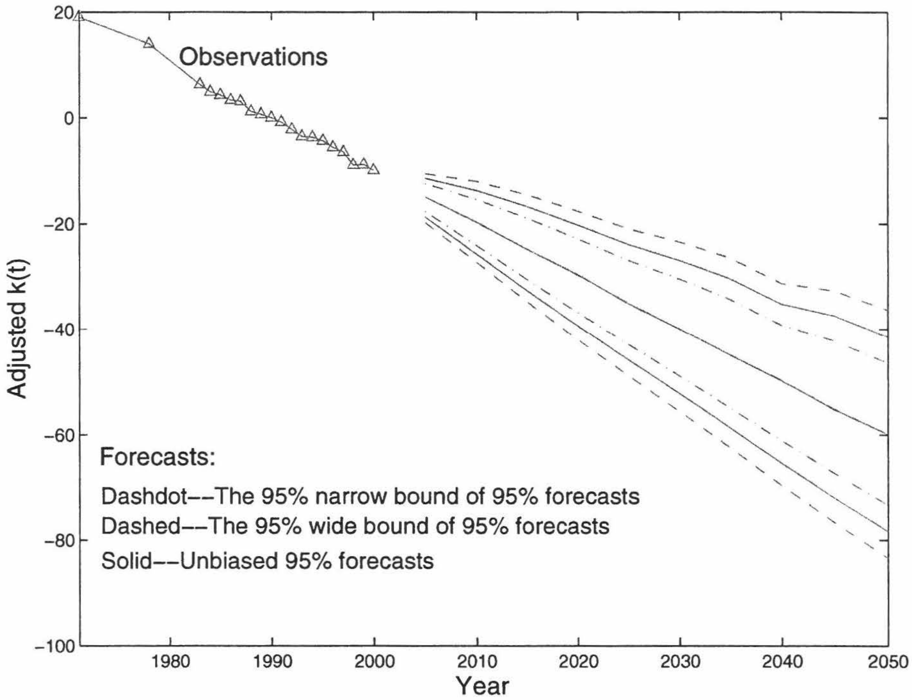


Figure 4. *Observations and 95% forecasts of $k(t)$ of South Korea.*

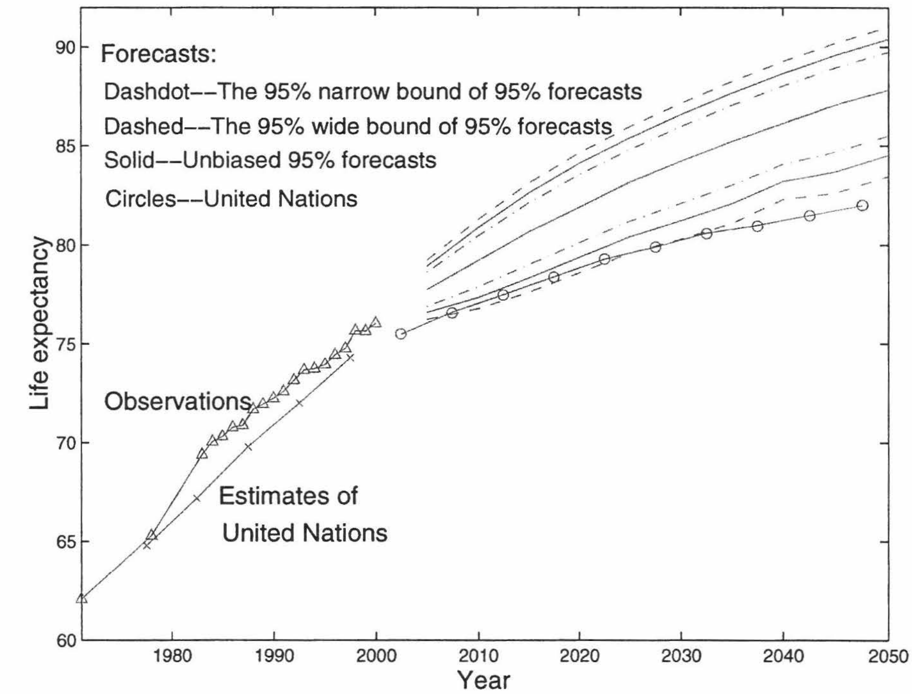


Figure 5. *Observations and 95% forecasts of life expectancy of South Korea.*

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

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

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 $[1 - \frac{u(t)-u(t-1)}{u(T)-u(0)}][u(t) - u(t-1)]$:

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

Therefore,

$$\sum_{t=1}^{T}\frac{v(t)}{\sigma^2} \sim \chi^2(\sum_{t=1}^{T}[1 - \frac{u(t) - u(t-1)}{u(T) - u(0)}][u(t) - u(t-1)]) =$$

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

and

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

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

$$\text{var}(\widehat{\sigma}^2) = \text{var}[\frac{\sigma^2}{u(T) - u(0) - \frac{\sum_{t=1}^{T}[u(t)-u(t-1)]^2}{u(T)-u(0)}}\sum_{t=1}^{T}\frac{v(t)}{\sigma^2}] =$$

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

Thus

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

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

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

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

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

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

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

Because (13a) describes the distribution of a positive variable ($\hat{\sigma}$) 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 $(k(t) - k(t-1) - \hat{c})$ to simulate 1000 values of $\hat{\sigma}$ using (4), for each case of using data in 2 to 100 single-year intervals. Meanwhile, we analytically calculate the distribution of $\hat{\sigma}$ using (12a) where $\sqrt{\mathrm{var}(\hat{\sigma})}$ is given by (14a). Comparing values $\hat{\sigma}$ from simulating and analytical calculation, Figure 6 indicates that (13a) is accurate. Substituting σ by $\hat{\sigma}$ will make (14a) further approximate, but derives the relative error of $\hat{\sigma}$, $\mathrm{re}(\hat{\sigma}) = \sqrt{\mathrm{var}(\hat{\sigma})}/\hat{\sigma}$. Comparing to simulated values of $\mathrm{re}(\hat{\sigma})$ that are obtained from simulated $\hat{\sigma}$, Figure 7 shows the analytical calculated $\mathrm{re}(\hat{\sigma})$ given by (14a) where σ is substituted by $\hat{\sigma}$ is also accurate.

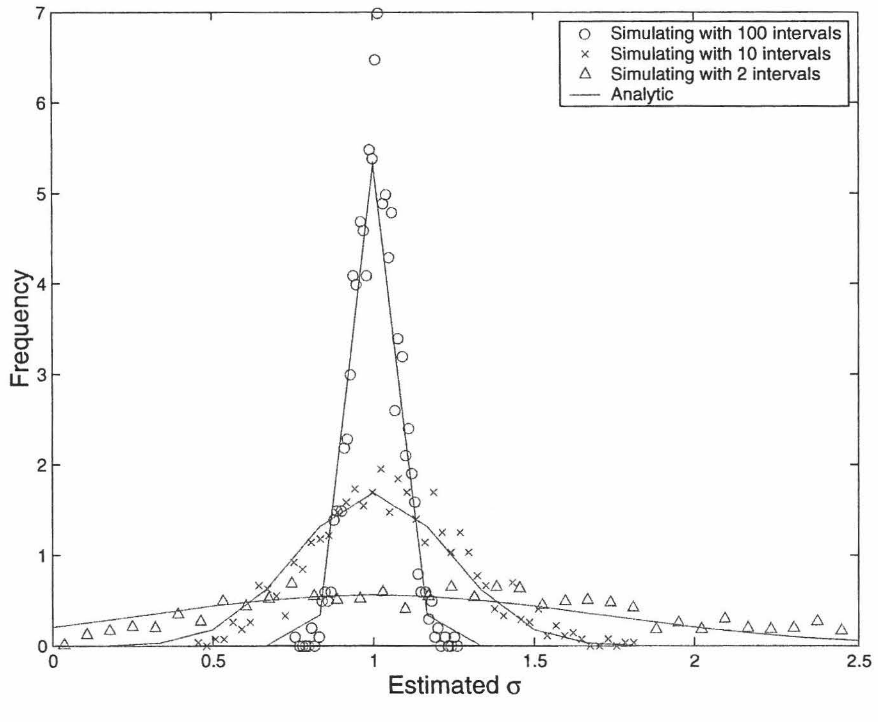


Figure 6. *Simulated and analytical distributions of estimated σ.*

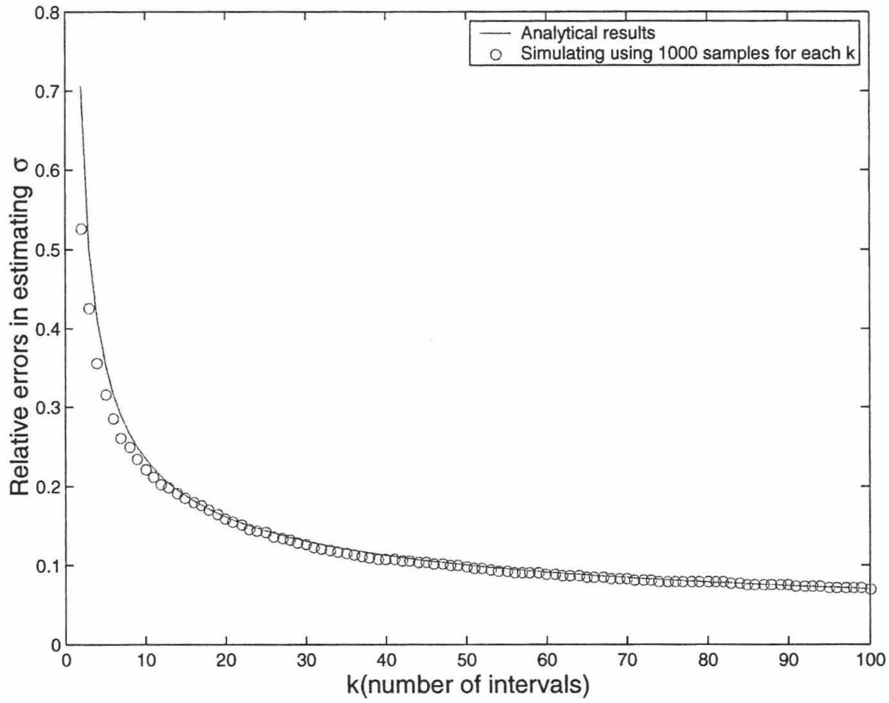


Figure 7. *Relative errors in estimating σ.*

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,

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

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

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

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

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

Equations (16a)–(18a) show that $\text{var}(a(x))$ and $\text{var}(b(x))$ 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*]

International Statistical Review (2004), **72**, 1, 37–50, Printed in The Netherlands

Mortality Forecasting and Trend Shifts: an Application of the Lee–Carter Model to Swedish Mortality Data*

Hans Lundström and Jan Qvist

Statistics Sweden, Stockholm

Summary

In this paper we examine how the Lee–Carter model fares with Swedish data for the period 1901–2001 and for segments of this period. We have choosen to censor ages less than age 40 as those ages only are of marginal interest to the forecast. At age 40 some 98 to 99 percent of the birth cohorts are survivors. In the study we only consider the unweighted k_t estimates. The Lee–Carter model provides very good fits to the data. When splitting up the base period there seems to be an interaction beween the age and time components of the model. In order to deal with the different phases of falling mortality for males and females possibly one should choose the past 25 years as a base in the model. Selecting the base period is however a judgmental issue depending on the main focus of the forecast. Is it long-term, short-term or, as in Sweden, a combination of both?

Key words: Mortality; Forecasting; Lee–Carter model; Trend shifts.

1 Introduction

In 2001, the life expectancy in Sweden reached 82.1 and 77.6 years for females and males, respectively. In a temporal perspective this means that life expectancies increased by about 25 years since 1900. This historically unprecedented advance in longevity has bestowed upon us longer lives and better opportunities for realizing our full potential as human beings. At the same time, however, this development also imposes its challenges. Over time, falling fertility and declining mortality have resulted in an aging population; the proportion of old people is now steadily increasing in the industrialized world. This is a new social and economic reality to which society is adapting itself. As a result, projecting the old age population is a more important endeavor than it used to be and mortality forecasting has come more into focus.

In the context of making population forecasts, Statistics Sweden faces the dual task of making both long and short term projections. For this reason it is necessary for us to focus on both short and long term developments in mortality.

Existing methods for making population projections naturally fall into three categories: (i) extrapolation of mortality trends without considering causes of death, (ii) cause of death specific extrapolations, and (iii) extrapolations that build on factors such as changing life styles and advances in medicine among other things. Most institutions working with population projections undoubtedly work with these three aspects of mortality change albeit the third one can have more a character of speculation.

*Working Paper presented at the seminar on "How to deal with uncertainty in population forecasting?" in Vienna, December 12–14, 2002

In addition to several relatively simple time series models for projecting age-specific mortality, a new ingenious time series model has been suggested by Lee & Carter (1992). This model involves partly a time invariant mortality schedule, partly age-specific and time-specific parameters which provide a temporal adjustment to the mortality schedule and hence provides a temporal description of age-specific mortality. This model also accommodates the application of a random-walk model for portrayal of the temporal stochastic effects of mortality. It is generally believed that the best approach to estimating the model is to make use of long time series of age-specific mortality.

Wilmoth (1993) has proposed a weighted estimation procedure for the Lee–Carter model. In the recent past the interaction between age-bound and time-bound changes have been discussed in the literature (Carter & Prskawetz, 2001; Booth, Maindonald & Smith, 2002).

In this paper we examine how the Lee–Carter model operates with Swedish data for the period 1901–2001. The first part of the paper is dedicated to a description of how age-specific mortality by sex has changed over time. To this end we have made use of calendar year experiences for the above mentioned period. In the second part of the paper we apply the Lee–Carter model to the data. This application involves that we make use partly of the complete time series of data, partly to segments of the time series. This gives us an opportunity to discuss how the estimated parameters depend on the choice of time period. Although the application is limited to Swedish data, we believe that it has relevance in the context of other mortality experiences as well.

The last part of the paper addresses issues concerning application of the Lee–Carter model to age and cause specific mortality.

2 Data and Method

The analysis of mortality and trend shifts is based on yearly central death rates by sex for the period 1901–2001. The central death rates have been estimated as

$$m_x^t = \frac{d_x^t}{(P_{x-1}^{t-1} + P_x^t)/2}$$

where d_x^t is the number of deaths at age x (end of year) in calendar year t, P_x^t is the population aged x December 31 in year t.

For the description of age-specific mortality, the Lee–Carter method (Lee & Carter, 1992) is

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

a_x age-specific constants describing the general pattern of mortality for the whole base period
k_t index of the level of mortality capturing the main time trend in death rates
b_x age-specific constants describing the relative speed of change in mortality at each age
$e_{x,t}$ is an error term.

To obtain unique parameter estimates, a_x is determined as the mean of $\ln(m_{x,t})$ over time for each x. The sum of b_x for all x is set to 1, and the sum of k_t for all t to 0. To find a least squares solution, the singular value decomposition (SVD) method was applied to the normalized matrix of death rates (Good, 1969). The first column vectors of the returned decomposition matrices and leading value were used. In calculating the parameter estimates the SVD subroutine in the SAS (IML) software package was applied. In this study, we only consider unweighted k_t.

In this model, the time factor k_t is intrinsically viewed as a stochastic process, and is found to be accurately modeled by a simple time series model

$$k_t = k_{t-1} - z + e_t$$

where z is a constant average rate of decline, and e_t is a random term whose statistical properties are estimated from the data. This particular aspect of the model is not discussed nor used in the paper.

3 Mortality Decline in Sweden 1900 to 2001

In 1900, the life expectancy for females was 53.7 years. In the year 2001 it had increased to 82.1 years. For men the similar increase was from 50.8 to 77.5 years. During this period the average increase in life expectancy per calendar year was 0.28 for females and 0.26 for men. In reality, however, the increase in life expectancy was more pronounced before 1950 than after (table 1). Before 1950 the average increase in life expectancy per calendar year was 0.38 year for both males and females, with a steady sex differential of about 3 years.

Table 1

Life expectancy at birth and at age 65 and average annual increase by sex.

	Life expectancy at birth		Life expectancy at age 65	
Year	Males	Females	Males	Females
1900	50.8	53.7	12.1	13.0
1950	69.8	72.4	13.5	14.3
1980	72.8	78.9	14.3	18.0
2001	77.6	82.1	16.9	20.1

Average yearly increase in remaining life expectancy for periods between 1900 and 2001.

	at birth		at age 65	
Period	Males	Females	Males	Females
1900–2001	0.26	0.28	0.05	0.07
1900–1950	0.38	0.38	0.03	0.03
1951–2001	0.15	0.19	0.06	0.11
1951–1980	0.10	0.21	0.02	0.12
1981–2001	0.23	0.15	0.12	0.10

From 1950 and onwards the yearly increase in life expectancy slowed down (figure 1). For males it almost came to a standstill during the 1960s. Beginning about 1980 we saw a new trend shift once again involving a rapidly increasing life expectancy for males. In 1980, the difference in life expectancy between males and females had grown to six years. After 1980 the male–female differential in life expectancy abated and in 2001 had declined to 4.5 years in favor of females.

The remaining life expectancy at age 65 shows clear trend shifts in 1950 and 1980. After 1950 the increase in life e_{65} slowed down for males. From 1980 and onwards, the increase in e_{65} for males gained some momentum.

Figures 2 and 3 give an overall portrayal of a century of mortality decline in Sweden. The data is by single-year ages for ages 10–99 years. In spite of the massive amount of detail and random fluctuation, the overall picture is one of a smooth temporal development. The Spanish Flu left a noticeable footprint in 1918. With few exceptions, the mortality curves for the different ages follow each other with but few crossovers; especially for ages between 50 and 90. At higher ages, random fluctuations make it difficult to identify the different curves from each other.

Up to 1950 the mortality decline was essentially the same for males and females with rapid declines at younger ages and a slower decline among the elderly. Falling rates of infectious diseases and diseases of the respiratory organs account for this improvement. Among the elderly the mortality decline was slow, especially for males.

From 1950 and onwards, we see a clear trend shift. For males the mortality decline stopped and

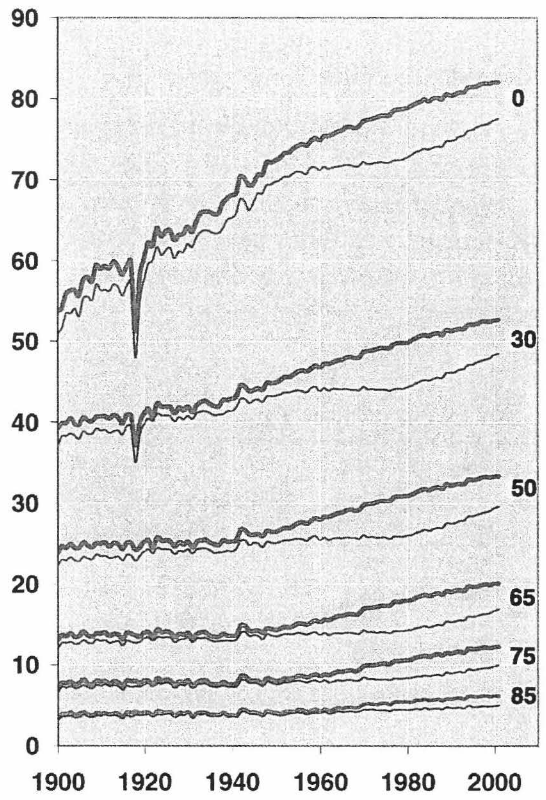


Figure 1. *Life expectancy at birth and at ages 30, 50, 65, 75 and 85: Sweden 1900–2001.*

turned into an increase at ages between 30 and 55. For females a more pronounced mortality decline started at all ages in 1945. For middle-aged males the increase in mortality was mainly due to increases in heart and vascular diseases.

From the end of the 1970s there was another trend shift. Mortality rates for males started to decline once more, possibly due to health related information and a change toward a more healthy life style with fewer smokers, a reduction in alcohol consumption, improved dietary habits, and general improvements in the health sector. For middle aged females, the rapid mortality decline originating during the mid 1930s slowed down around 1960.

The overall picture is that the age pattern of mortality has changed over time. On the whole, a linear decline in age specific mortality on the logarithmic scale is only visible for shorter periods. For females however a linear trend in mortality decline seems approximately true since 1960. For males, a linear trend seems to have prevailed only during the past 20–25 years.

Fitting the Lee–Carter Model

The above-mentioned descriptive analysis of mortality changes in Sweden between 1901 and 2001 is now reviewed by means of recourse to the Lee–Carter model. To this end, we study how the unweighted estimates of $k(t)$ vary over time. In passing, it must be noted that estimates of $b(x)$ and $k(t)$ are complimentary, that is, it is the product of the two that brings forth the time changes in

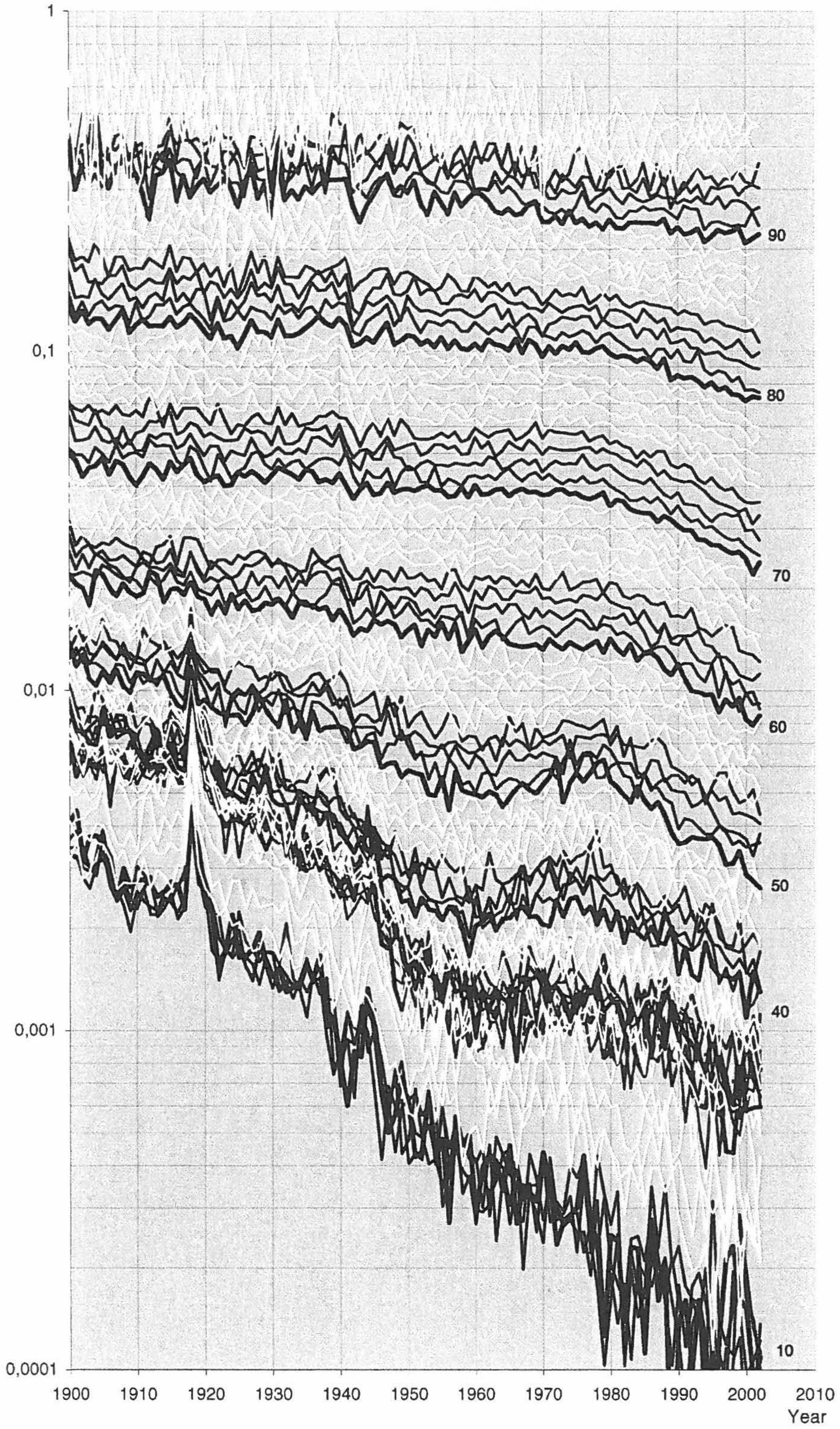


Figure 2. *Central death rates for males 10 to 99 years of age. Sweden 1900–2001 (Log scale).*

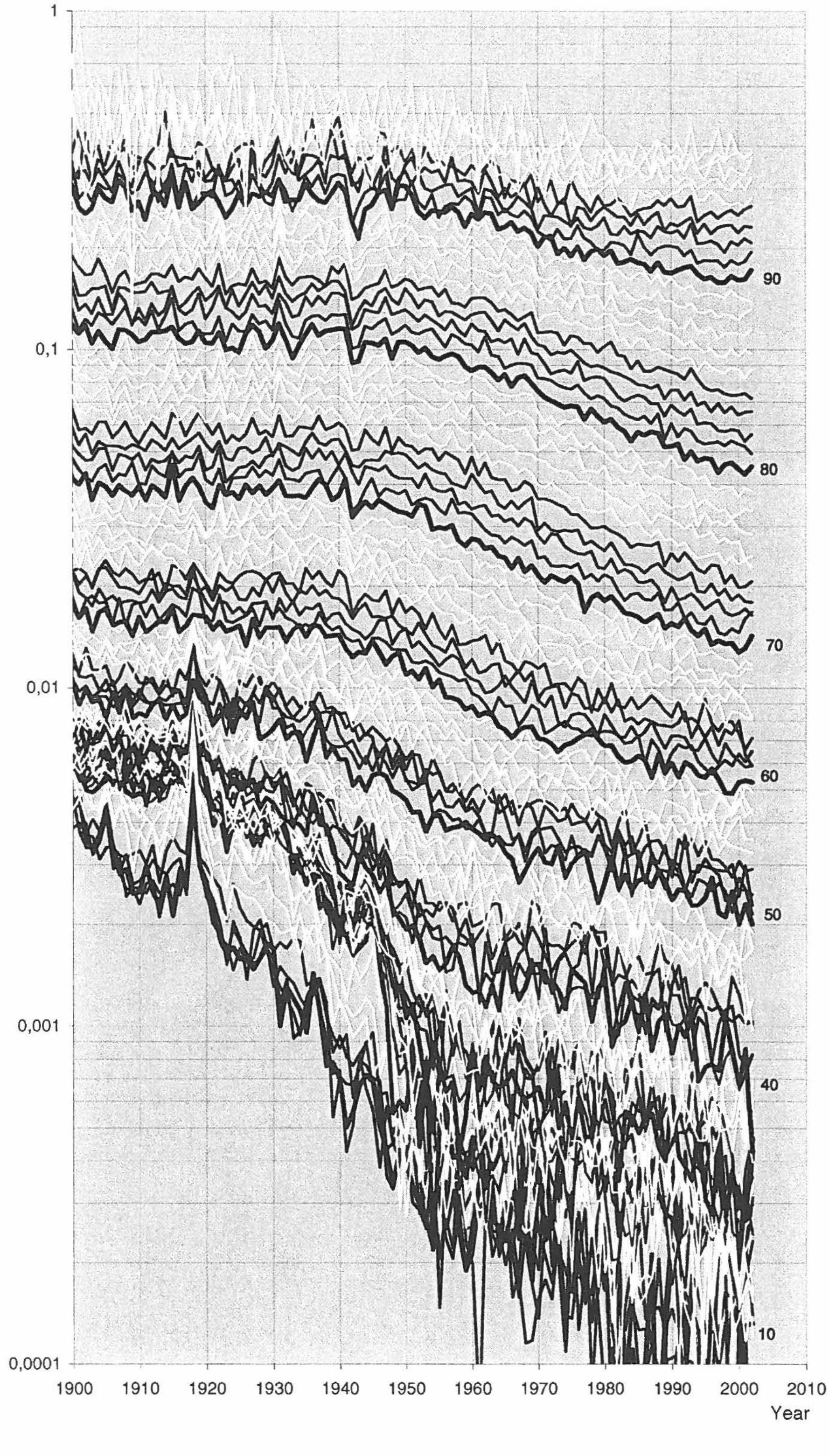


Figure 3. *Central death rates for females 10 to 99 years of age. Sweden 1900–2001 (Log scale).*

age-specific mortality. To begin with, we study the estimates with respect to the period 1901–2001. This set of estimates of $b(x)$ and $k(t)$ forms the basis for various comparisons with corresponding estimates for other time periods.

Figure 4.1 shows $k(t)$ for calendar years 1901–2001 (all ages involved). It will be appreciated that $k(t)$ for both males and females displays an obvious linear trend. In consequence, it may be concluded that age-specific mortality (in the sense of single-year age central death rates) has fallen nearly exponentially over time. The broken linearity around 1918 is attributable to the well-known Spanish flu. A close inspection of figure 4.1 also reveals that mortality fell the fastest around the middle of the century and that, on the whole, the drop in mortality was faster for females than for males. These results tie in very well with the descriptive analysis.

Figure 4.2 shows estimates of the age-specific component $b(x)$. The figure shows, as indeed is well known, that the mortality decline was more pronounced for younger than for older people. Clearly, during the 20th century, the drop in mortality for a life aged 1 was several times that of a life aged 70, for example.

Figures 5.1 and 5.2 show observed central death rates as well as rates fitted from the model. The goodness of fit for the unweighted estimation process brings forth some discrepancies. The progression of $k(t)$ provides a better fit to younger than to older ages. The most visible discrepancy is at old age for men as well as for women. Projecting the old age population as correctly as possible is important to us. A weighted estimation process may make up for the lack of fit relative to the unweighted process (Wilmoth, 1993). In the following, we have instead choosen to censor ages less than age 40. The reason for this is that the spectrum of diseases is more stable or homogeneous at ages 40 and over; in fact, during the past decades, mortality changes at ages less than 40 have had a decreasing impact on the projected population. Indeed, in the Swedish life table for the year 2000, at age 40 some 98 or 99 percent of the birth cohorts are survivors.

When disregarding mortality below age 40, the fit provided by the model was somewhat improved. Nevertheless, several departures from a "good fit" remained (figures 4.3, 4.4, 5.3 and 5.4).

Changes during the first half of the 20th century, 1901–1951

During the first half of the 20th century, the mortality index $k(t)$ for ages 40 and over decreases much the same way for males and females. The exception is for 1918 on account of the Spanish flu. The index accelerates somewhat toward the middle of the century (figures 4.5 and 4.6).

The age-specific modulation $b(x)$ is nearly the same for males and females. The reduction in mortality is almost exclusively at ages between 40 and 65. At ages above 65, the reduction is more or less insignificant. Stated otherwise, at ages where chronic diseases establish themselves, the reduction in mortality is but slight. It is noticeable that the reduction in mortality, partly due to effective treatment of infectious diseases, mainly is a phenomenon among those aged 40–60.

Figures 5.5 and 5.6 display the small differences between observed central death rates and those fitted from the model. Generally, the unfolding of $k(t)$ is the same at all ages in the chosen age interval. Hence, the unweighted Lee–Carter model provides an unusually close fit at ages 40–90.

The first half of the 20th century involves changes that sometimes are described as the epidemiological transition which involves that mortality due to infectious diseases and diseases of the respiratory organs fall precipitously. During this time period the life expectancy for men increased from 51.6 to 70.0 years and for females from 54.1 to 72.7 years.

Changes 1951–2001

During the second half of the century, $k(t)$ follows a linear and smooth development for females (figure 4.7). The fall in mortality for women aged 40–90 is steeper than during the previous fifty-year

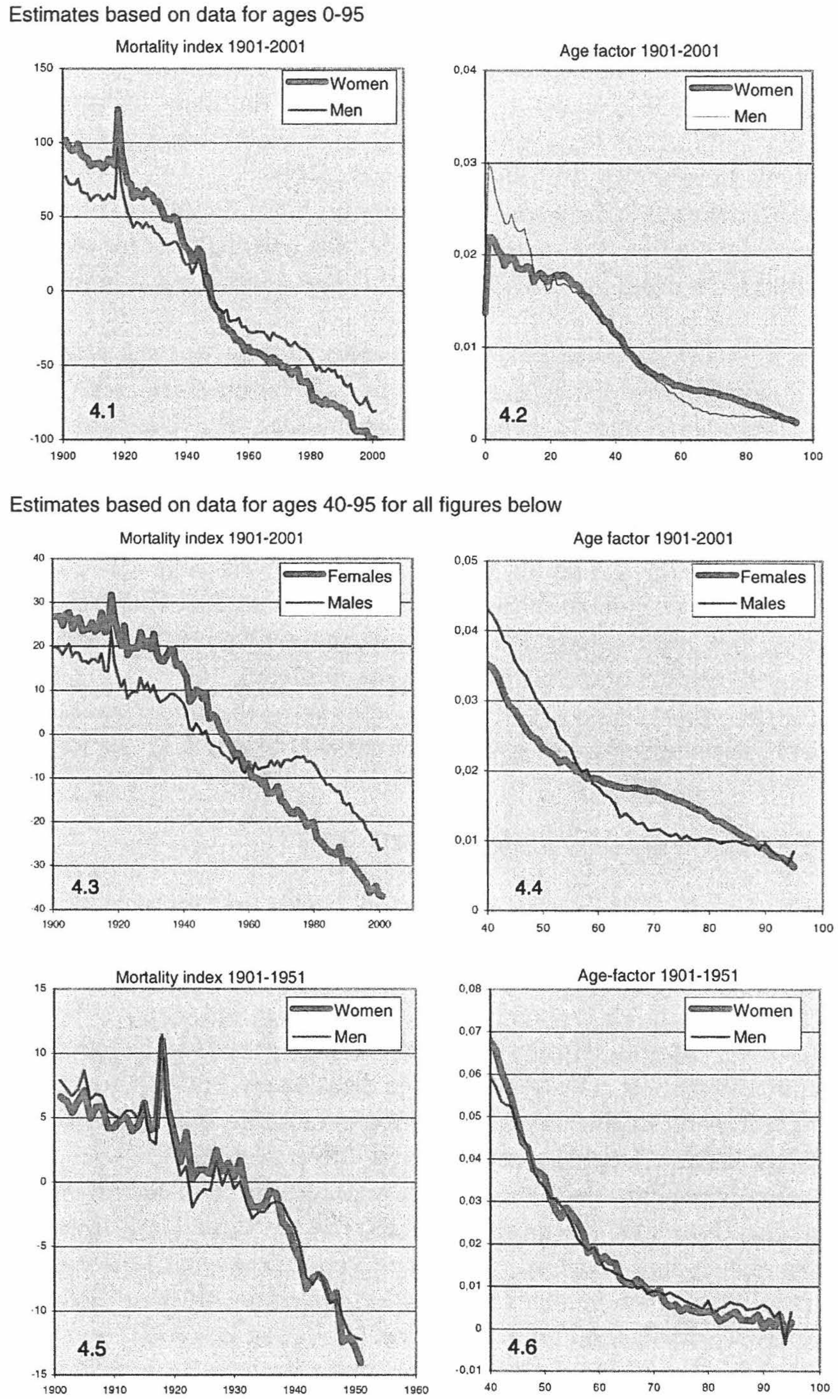


Figure 4.1–4.6. *Mortality index k_t and age factor b_x in L–C model for different periods.*

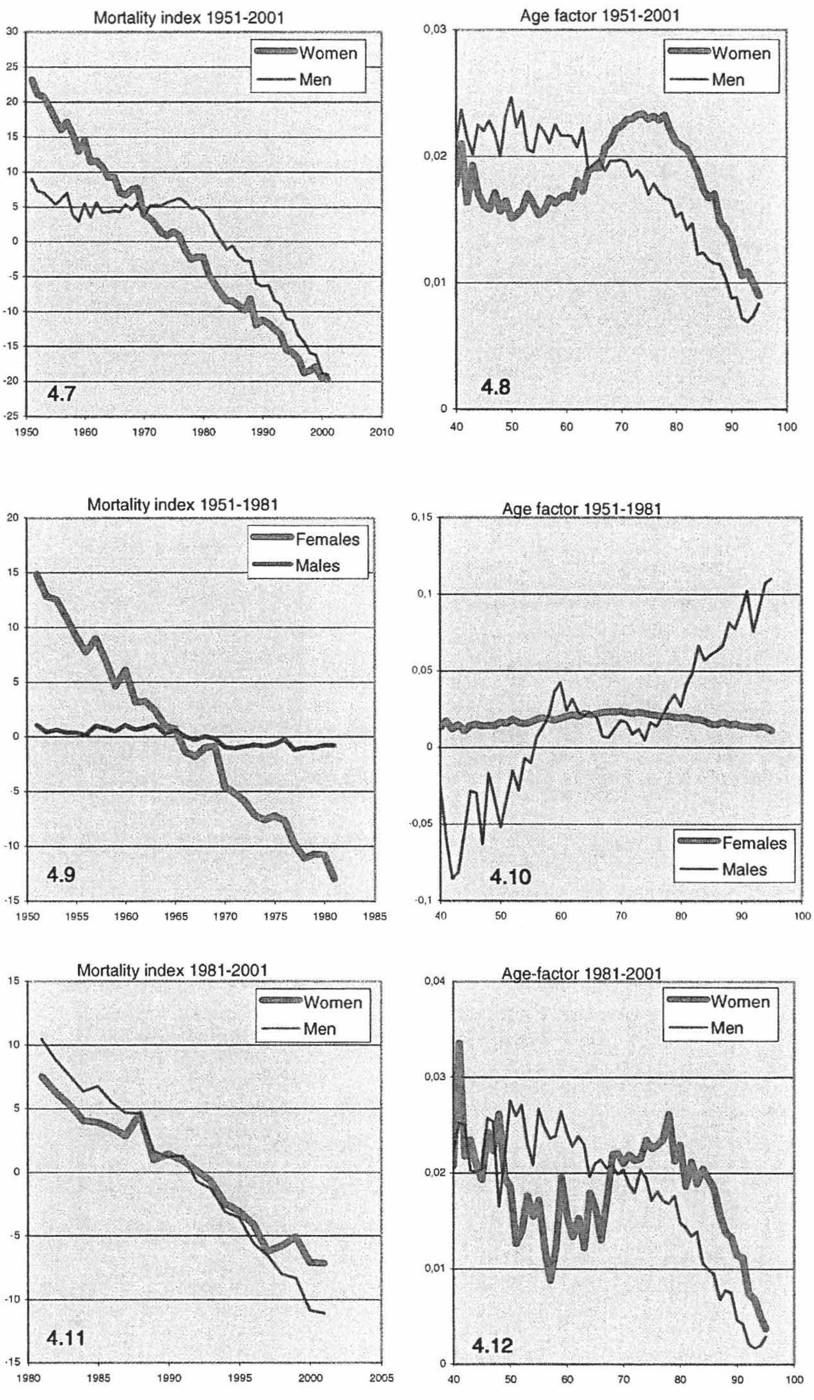


Figure 4.7–4.12. *Mortality index k_t and age factor b_x in L–C model for different periods.*

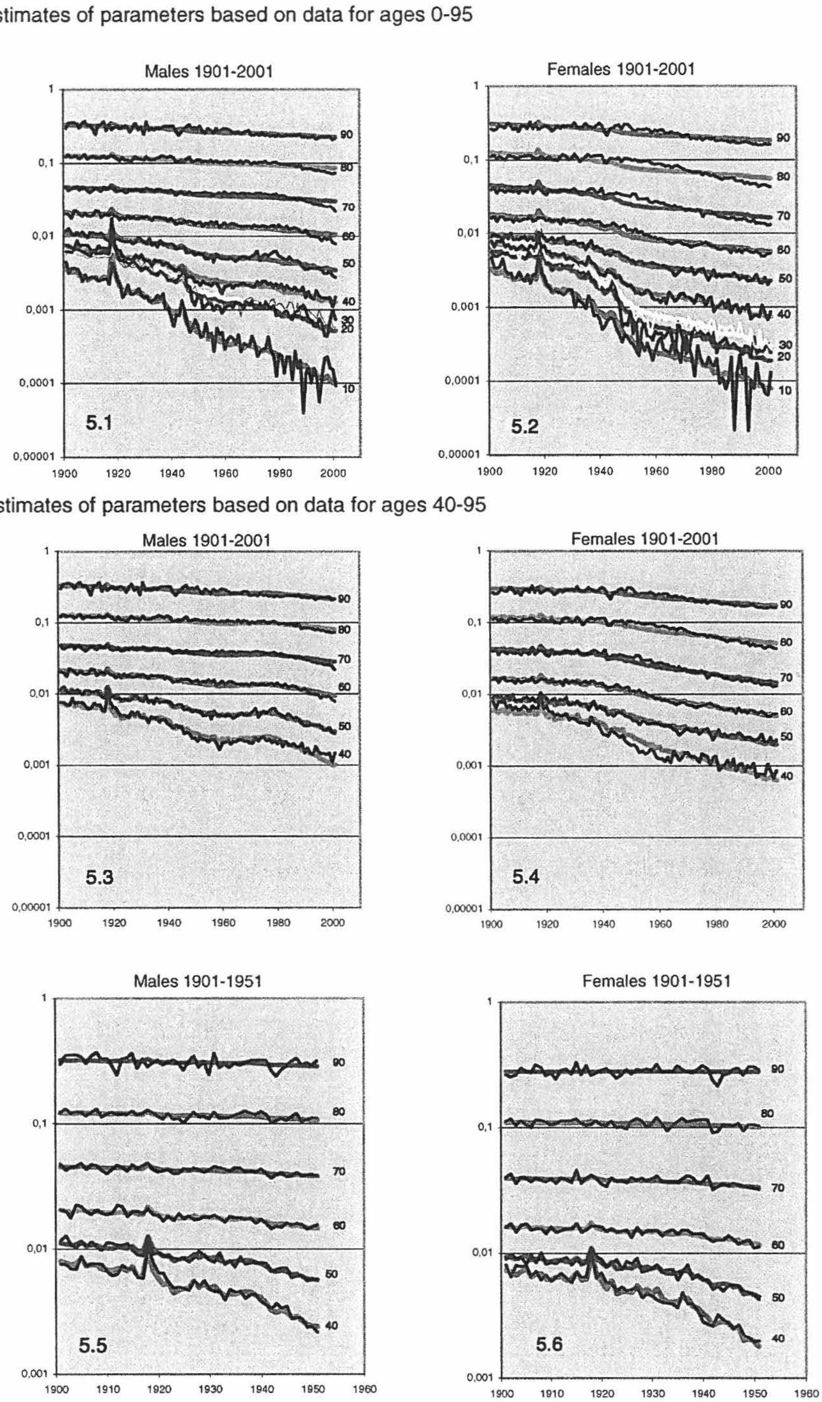


Figure 5.1–5.6. *Observed and fitted death rates from L-C model for different periods (selected ages).*

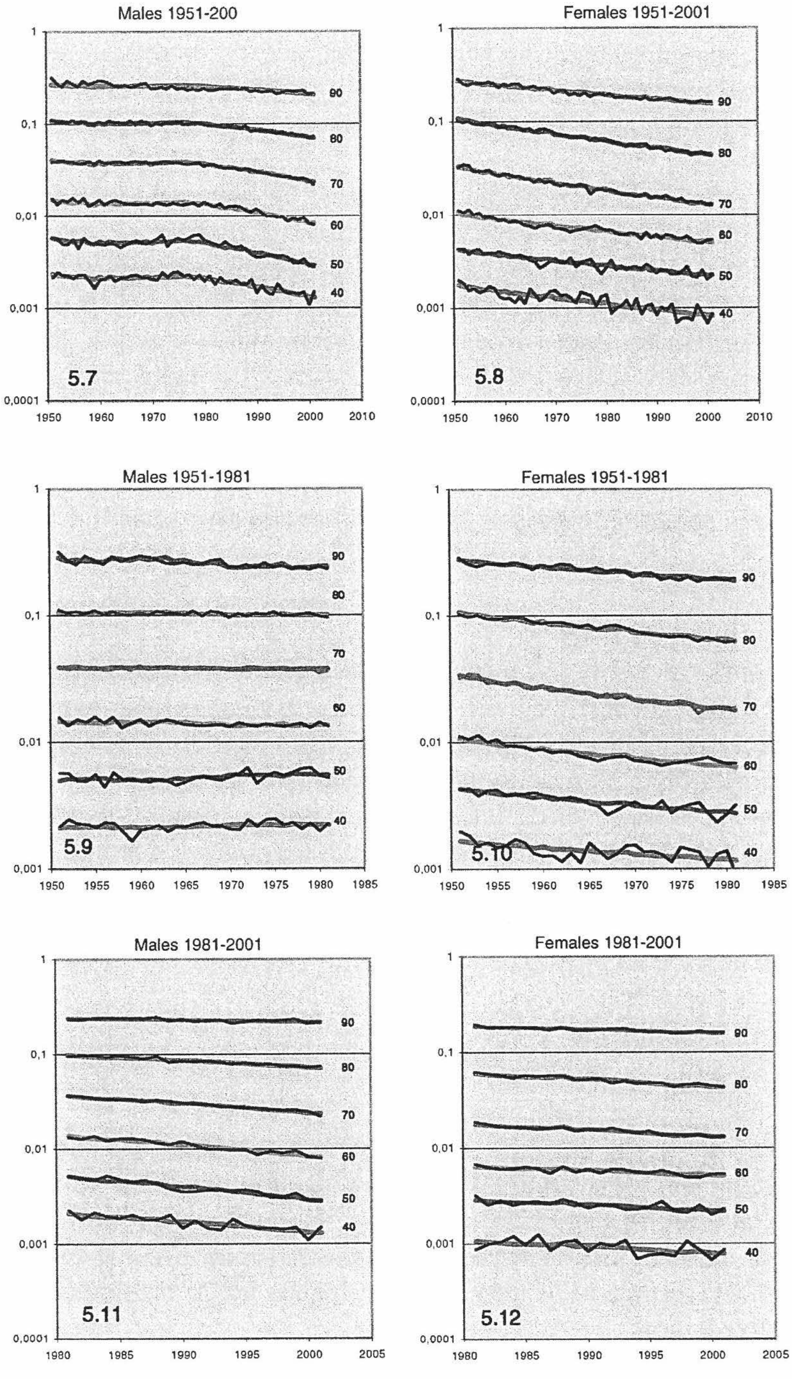


Figure 5.7–5.12. *Observed and fitted death rates from L–C model for different periods (selected ages).*

period. For males the time changes are considerably different. During most of the period there is no fall in mortality for men. However thirty years into the period the index drops markedly. The changes in cause of death structure for both men and women are, in the first place, due to falling incidences of cardio-vascular diseases. This characterizes the latest unfolding of the epidemiological transition.

The age effects for this period stand out as very different from those of the first half of the 20th century. A peak of pronounced age effects is visible at steadily increasing ages. For females the transition toward lower mortality has its foundation at ages 70–80 years. For males this foundation is centered among the middle-aged. Nevertheless, pronounced age effects are also visible for men at higher ages.

The fit afforded by the Lee–Carter model is very good for both males and females. This despite the fact that the unfolding of $k(t)$ for males varied a great deal (figures 5.7 and 5.8).

Fundamentally, during the second part of the century the life expectancy for females increased from 72.7 to 82.1 years while for males the increase was from 70.0 to 77.6 years. In other words, the life expectancy for females increased a great deal more than for males. The differential in life expectancy between the sexes is now about 4.5 years. Some 20 years ago it was as much as 6 years.

In order to arrive at a clearer picture of mortality trends during the latter part of the 20th century, we have divided this period into two segments; namely the periods 1951–1981 and 1981–2001. What demands this division of time is mainly the mortality changes affecting males. For women, there is not much difference in parameter estimates between those two time periods.

1951–1981

During this period the mortality index $k(t)$ falls very slowly for males (figure 4.9). The fall for females however is significant and maintains strong linearity. On the other hand, the age factor $b(x)$ is now altogether reversed relative to estimates for previous periods. In as much as $b(x)$ now takes on negative values, signifying increasing mortality, the scale on which $b(x)$ is measured is much different from diagrams where $k(t)$ declines. The increase in mortality is mainly for those aged 40–55 years. Around age 70, estimates approach 0 from the positive side indicative of a very slight fall in mortality. At most ages estimates of $b(x)$ are relatively high, something that is balanced by decreasing estimates of $k(t)$. The fall in mortality at the highest ages however is on par with what has previously been observed. For women the age factor effects are smoothly distributed, centering at ages around 70.

1981–2001

Once again, in recent years the mortality index has taken on a linear and decreasing form for both males and females. The fall is less obvious for females than for men. A tampering effect is found for females during the most recent years (figure 4.11).

The age factor for males (figure 4.12) shows a relatively strong effect at ages up to 70–80 years. For females such effects seem to be located at ages as high as 80. At higher middle-ages, age effects for women are relatively weak.

4 Summary and Commentary

In Sweden there was a substantial drop in mortality at all ages during the 20th century. Throughout this period it was mortality among the youngest that fell the most. Largely, males and females share this development. Sweden has changed from being a nation with high to one of low mortality. Similar developments are found for the majority of industrialized countries.

As a result of changes in mortality during this century, life expectancy at birth has increased by some 25 to 30 years. Increases in life expectancy were more pronounced during the first half of the century than during the latter. In particular, females have come to enjoy much longer life expectancies than males. This change however has mostly taken place during the latter half of the century. This effect is largely due to falling mortality at higher ages, something that materializes when studying changes in the remaining life expectancy at adult ages.

The determinants behind this improvement primarily fall into two different categories. In the first place comes a reduction of infectious diseases during the first part of the century, something that mostly benefited younger persons. Improved and more nutritious diets along with improved sanitary conditions and housing, supplemented by advances in the treatment of infectious diseases underlie this development. Nevertheless, some diseases such as pneumonia and certain specific infectious diseases remained responsible for sustained mortality.

In the second place it is observed that the reduction in chronic diseases during the second part of the century to an increasing extent involves the aged population. This drop is mainly due to reduced illness from cardio-vascular diseases. Fewer smokers, more healthy diets and increased physical activity together with advances in medicine are thought to underpin this development.

The Lee–Carter model is widely applied for the purpose of projecting mortality. The model involves a three-stage process of application. These steps build on a base model, weighting of parameters and extrapolation. In this paper we have limited ourselves to working with the unweighted model and performed an explorative analysis. We believe that working with unweighted parameters facilitates a better understanding of discrepancies that arise from working with different time periods. In this perspective, we have primarily focused on mortality at ages above 40 since $k(t)$ is more consistent for such ages.

The unweighted Lee–Carter model provides very good fits to Swedish mortality over extended time periods. Nevertheless, it also has its weak sides.

When estimating the parameters in the model using the whole 20th century as a base period, the structure of the constant $b(x)$ estimates reflect a general age structure for the whole period. Splitting up the base period, the drop in mortality changes its main age location from younger ages during the early part of the century to older ages toward the latter. This suggests an interaction between $b(x)$ and $k(t)$. Possibly this effect can be lessened by application of the estimation procedure suggested by Wilmoth (1993).

Another problem concerning estimation of the model is that the reduction in cardio-vascular diseases has been somewhat irregular. While, at the beginning of the 1950s, the risks of cardio-vascular diseases among females declined fast over time, such a development did not take place for males. Instead, such male mortality risks increased for some decades. Possibly improved economic conditions during this period involved a range of negative health consequences. Increased use of tobacco and alcohol paired with employment demanding less physical activity than before is thought to have left its marks on the mortality development.

Health agencies and news media began to inform the public about the risk factors underlying cardio-vascular diseases as a result of extensive epidemiological and medical studies. However, it was not until the end of the 1970s that clearly visible changes in the health behavior of males began to surface. Today the proportion of smokers among Swedish men is among the lowest in the world, about 20 percent. In terms of estimating future mortality trends it is necessary to consider the different phases of mortality transition for males and females; the distance between these phases is about 25 years.

The Lee–Carter Model as a Means of Projecting Mortality

Our analysis of Swedish mortality suggests that using the past fifty years as a base period for the Lee–Carter model is preferable to making use of an entire century of mortality change. This choice is motivated due to the fact that main drops in mortality come from two different sources, namely a reduction in infectious diseases and a reduction in chronic disease. Furthermore, in order to deal with the different phases of falling mortality for males and females, one should possibly limit the focus to the past 25 years. Besides, males who began a downward mortality transition later than females with respect to chronic diseases have recently experienced deeper drops than females. This we believe is especially called for when making mortality projections for Sweden in the short and medium term. On the other hand, we recognize the need for the length of the base period to be reasonably on par with the length of the projection period which extends to 2050.

The Lee–Carter model is sometimes used for making long-term extrapolations of life expectancies (periods that extend more than 50 years into the future). The base for making such extrapolations is also lengthy and often covers the entire 20th century. When little consideration is given to the time-bound changes in the distribution of causes of death, perhaps it is justified to say that what underlies the extrapolation is an assumptive improvement in health behavior and living styles. What base period one should choose depends on the objective of the forecast. If it is primarily a long term forecast of life expectancy, irregular short term trends are not so important. On the other hand, in a short term forecast they can be crucial for calculations such as the tendency of stagnation in mortality decline for females in the late 1990s.

References

Booth, H., Maindonald, J. & Smith, L. (2002). Age-time interactions in mortality projection: Applying Lee–Carter to Australia. Working Papers in Demography. Revised 2 August 2002, Demography and Sociology Program, Research School of Social Sciences, The Australian National University.

Carter, L. & Prskawetz, A. (2001). Examining Structural Shifts in Mortality Using the Lee–Carter Method. MPIDR Working Paper WP 2001–007, March 2001. Max Planck Institute for Demographic Research, Rostock.

Good, I.J. (1969). Some Applications of the Singular Decomposition of a Matrix. *Technometrics*, **11**(4), 823–831.

Lee, R.D. & Carter, L.R. (1992). Modeling and Forecasting U.S. Mortality. *J. Amer. Statist. Assoc.*, **87**(419), 659–671.

Wilmoth, J. (1993). Computational Methods for Fitting and Extrapolating the Lee–Carter Model of Mortality Change. Technical Report, Department of Demography, University of California, Berkeley.

Résumé

Dans cet article nous examinons comment le modèle Lee–Carter se comporte avec les données suédoises pour la période 1991–2001 et pour des parties de cette période. Nous avons choisi d'exclure les âges en dessous de 40 ans car ces âges ont un intérêt marginal pour la projection. A 40 ans, de 98 à 99% des membres d'une cohorte sont encore vivants. Dans l'étude nous considérons uniquement les estimateurs non pondérés. Le modèle Lee–Carter fournit de très bons ajustements pour les données. En divisant la période de base il semble qu'il y ait une interaction entre les composantes âge et temps du modèle. Afin de tenir compte des différentes phases de chute de la mortalité pour les hommes et les femmes, on devrait choisir les 25 années passées comme période de base dans le modèle. Le choix de la période de base est cependant une question de jugement selon l'objectif principal de la projection. Est-ce à long terme, à court terme ou comme en Suède une combinaison des deux?

[Received May 2003, accepted October 2003]

International Statistical Review (2004), **72**, 1, 51–64, Printed in The Netherlands

Toward a New Model for Probabilistic Household Forecasts

Jiang Leiwen[1] and Brian C. O'Neill[2]

[1]*Watson Institute for International Studies, Brown University, Providence, RI, 02912, USA. E-mail: Leiwen_Jiang@brown.edu* [2]*International Institute for Applied System Analysis, Laxenburg, Austria*

Summary

Household projections are key components of analyses of several issues of social concern, including the welfare of the elderly, housing, and environmentally significant consumption patterns. Researchers or policy makers that use such projections need appropriate representations of uncertainty in order to inform their analyses. However, the weaknesses of the traditional approach of providing alternative variants to single "best guess" projection are magnified in household projections, which have many output variables of interest, and many input variables beyond fertility, mortality, and migration. We review current methods of household projections and the potential for using them to produce probabilistic projections, which would address many of these weaknesses. We then propose a new framework for a household projection method of intermediate complexity that we believe is a good candidate for providing a basis for further development of probabilistic household projections. An extension of the traditional headship rate approach, this method is based on modelling changes in headship rates decomposed by household size as a function of variables describing demographic events such as parity specific fertility, union formation and dissolution, and leaving home. It has moderate data requirements, manageable complexity, allows for direct specification of demographic events, and produces output that includes the most important household characteristics for many applications. An illustration of how such a model might be constructed, using data on the U.S. and China over the past several decades, demonstrates the viability of the approach.

Key words: Model; Probabilistic forecast; Household.

1 Introduction

Anticipating changes in the number, size, and composition of households is an important element of many issues of social concern. For example, the living arrangements of the elderly are a key determinant of their needs for socio-economic, physical and psychic assistance, and household projections are therefore critical to understanding challenges in this area (Gonnot *et al.*, 1995; Bongaarts & Zimmer, 2001). Similarly, several studies stress the importance of the household as a unit of consumption. Household members share living space, energy, water, and other goods, leading to potentially important economies of scale. Household composition, particularly by age, also affects consumption patterns (King 1999; Select Committee on Environment, Transport and Regional Affairs, 1998; Van Diepen, 1994, 1995; Martin, 1999; Muller *et al.*, 1999). Changes in consumption due to shifts in distributions of households by type can induce significant environmental pressure (MacKellar *et al.*, 1994; Jiang, 1999; O'Neill *et al.*, 2001; Prskawetz *et al.*, 2001; Liu *et al.*, 2003; Keilman, 2003).

No international organization produces household projections for countries or regions of the world, as is done for population by the United Nations, IIASA, the World Bank, and the U.S. Census Bureau. Rather, statistical offices of some individual countries produce their own projections to meet needs regarding housing demand, household services, or elderly support (King, 1999; Hollmann *et al.*, 2000).

Most existing household forecasts are based on a headship rate approach, in which an independent population projection is combined with a forecast of probabilities, by certain population subgroups, of heading households of different types. Household projections therefore must cope with uncertainty not only in the population projection, but also in the projection of headship rates. The most common approach is either to produce only a single best guess projection, or a small number of alternative variants (King, 1999; Hollmann *et al.*, 2000). However, the variants approach is not well suited for household forecasts. First, it is difficult to define variants for headship rates, because the link between demographic events—which provide the theoretical basis for alternative variants—and the headship rates is unclear. Second, even if the household projections are performed with a more sophisticated method based directly on demographic events, the variant approach is unwieldy. There are many potentially relevant input variables that could be varied in alternative variants; in addition to total fertility (parity specific), mortality, and migration, other important factors are age at leaving home, and union formation and dissolution rates. It is not even clear which output variable should be used as the metric for the typical high/medium/low variants often produced in this approach. Should these be defined in terms of numbers of households, number of elderly households, age composition, multi-generation households, sex of household head, or what? Users have a range of different needs, and there is no single outcome that is of primary importance in all applications. Even if one were to select a single outcome as a basis for defining variants, the projection would be "probabilistically inconsistent" (Lee, 1999) in that the highest variant in terms of that outcome would not be the highest variant in terms of others. In summary, the weaknesses of the traditional variants approach in population forecasting (Keilman, 2003; Lee, 1999; Alho, 1998) are compounded for household forecasts.

Probabilistic household projections could address some of these problems. A single probabilistic projection accounting for uncertainty in the various demographic rates important to household dynamics would produce a distribution of outcomes that could be analysed with respect to a number of different metrics. Just as a probabilistic population projection can produce consistent uncertainty distributions for population size and for age structure, a probabilistic household projection could produce consistent uncertainty distributions for household numbers, sizes, and age composition, all of which would also be consistent with the uncertainty distributions for population size and age structure. Probabilistic household projections would also be valuable to users who employ them in applied analyses that support decision-making. Quantification of the likelihood of various outcomes is valuable to formal decision analysis, as well as input to policy discussions (O'Neill, 2004).

Probabilistic household projections would, however, be a formidable task. In addition to the methodological issues confronted by probabilistic population projections, they would also have to address new issues related to the additional variables and the demographic events and social processes they represent. We first briefly review current methods of household projections and the potential for using them to produce probabilistic projections. We also discuss the one existing probabilistic household projection in the literature (de Beer & Alders, 1999; Alders, 2001). We then propose a new household projection method of intermediate complexity that we believe is a good candidate for providing a basis for further development of probabilistic household projections. It has moderate data requirements, manageable complexity, allows for direct specification of demographic events, and produces output that includes the most important household characteristics for many applications.

2 Household Projection Models

Several different types of household projection models have been developed over the past few decades (Kuijsten & Vossen, 1988). Econometric models (both static and dynamic) model the relationships between household characteristics and socio-economic, cultural, and environmental determinants. Examples include the Cornell model in the U.S.A. (Caldwell *et al.*, 1979), the UPDATE model for households in small areas in England (Duley *et al.*, 1988), and the NEDYMAS model in the Netherlands (Nelissen & Vossen, 1989). However, these models are much less developed in terms of household structures than purely demographic models. Our interest here is in models capable of producing at least a modest set of household characteristics as output, so we focus our discussion on three types of purely demographic models: headship rate models (static macro-demographic models), micro-simulation models, and dynamic macro-demographic models.

2.1 Headship Rate Model

The headship rate method involves extrapolating proportions of household heads in population categories defined by certain combinations of age, sex, and possibly marital status. The headship rate projections are combined with an independent projection of the population by age and sex (and marital status) to produce a projection of households broken down by demographic characteristics of the head of the household.

Three different methodologies for projecting headship rates have been used: (1) linearly or exponentially extrapolating these changes into the future; (2) using regression models of the relationships between headship rates and major socio-economic factors combined with projections of those factors; and (3) developing scenarios based on information on government policies, such as housing policy for the elderly or for university students.

Projections of headship rates are intended to represent the results of dynamic demographic processes; however, the demographic processes themselves remain black boxes. Thus it is very difficult to incorporate demographic assumptions about future changes in demographic rates such as fertility, marriage, divorce, and mortality into the headship rate projections. Nevertheless the headship rate method is often used, because it is easy to apply and its data demands are modest. Over the past few decades, household projection models have been predominantly of the headship rate type.

A headship rate method is an attractive option as a basis for probabilistic projections given its limited data demands, relative simplicity, and manageable number of variables. Headship rate models could be made probabilistic by combining probabilistic population projections with probabilistic projections of the headship rates, perhaps defined using some combination of time series models, ex-post error analysis, and expert judgement. The task would be challenging, given the large number of rates that would need to be projected. Each age- and sex-specific headship rate would need its own forecast, particularly since changes in rates can vary strongly across groups. For example, headship rates changed the most in the older age group, particularly among women, between 1940 and 1980 in the U.S. (Kobrin, 1973; Carliner, 1975; Sweet, 1984).

Moreover, assumptions about the covariance of changes in the components and headship rates, across age and sex groups, and between the forecast years would need to be made. While statistical analysis of relationships between the components and headship rates would be possible, it would be more direct to analyze relationships between the components and the household-related demographic rates (union formation and dissolution, leaving home) themselves, and the results would be easier to assess. For example, the correlation between headship rates and fertility is not straightforward. Low fertility is often associated with the postponement of marriage. If the age at leaving home is also postponed, headship rates would decline; if not, they would increase due to increasing numbers of young single adults. Thus it would be preferable to explicitly include in a household projection assumptions regarding changes in the age at leaving home and age at marriage, rather than to directly model the relationship between fertility and headship rates.

2.2 *Microsimulation*

Microsimulation household models are virtually probabilistic (Hammel *et al.*, 1976; Wachter, 1987; Smith, 1987; Galler, 1988; Nelissen, 1991; Oskamp, 1997). These models take individuals as the unit of analysis, simulate each demographic event, and determine the changes in individual status usually using Monte Carlo methods combined with probability distributions for transition rates (Wachter, 1998). The model has major advantages in studying the variability of individuals and their distributions across households. For example, this method is especially well suited for studying kin networks, in which available kin may or may not live together in a household. Goldstein & Wachter (2001) recently applied a new method of using contemporary survey data as a basis for projecting spouse and sibling ties differentiated by educational level and race. Limitations of the method include the very large data requirements, and the fact that the simulation itself often must include a very large "population" of individuals to reduce sampling error.

2.3 *Dynamic Macro-demographic Models*

In dynamic macro-demographic models, the unit of analysis is the group, e.g. an age and parity class or all households of a given type. Transition matrices are used to update the composition of the population by group over time (Keilman *et al.*, 1988). Prominent examples include the LIPRO Model developed by Van Imhof & Keilman (1991). LIPRO projects a large number of household types and accounts for many types of demographic events; however, data on transition probabilities (and projections of their changes) among 69 types of events that individuals may experience are necessary. These data have to be collected in a special survey because they are not available in the conventional demographic data sources.

The ProFamy model developed by Zeng *et al.* (1997) takes a somewhat different approach that extends Bongaarts' nuclear status life table model. ProFamy only requires data that can usually be obtained from conventional demographic data sources. The model uses the individual as the unit of forecast; all individuals are classified according to eight dimensions of demographic status. Projections are performed based on status transition rates, and then distributions of households by size and type are derived based on characteristics of reference persons (or household 'markers') in a manner that produces consistent projections of households and individuals. Since 1997, Profamy has been used to produce household forecasts for Germany, the U.S. (Zeng *et al.*, 1999), China (Jiang, 1999), and Austria (Prskawetz *et al.*, 2001).

Probabilistic projections based on dynamic macro-dynamic household models would be possible by defining uncertainty distributions for the transition rates. The main challenge would be the very large state space such projections would need to cover. For example, the transition rates rely on standard schedules defined by age-, sex-, (sometimes parity-, and marital-status-) specific rates for mortality, first marriage, parity specific fertility, migration, leaving parental home, divorce, remarriage by the divorced, and remarriage by the widowed. Deriving probability distributions for all of these events, given limited data on many events, would be problematic, as would account for possible correlations among events.

Nonetheless, the macro-demographic approach is attractive in that it allows the application of available data, and of expert opinion, directly to demographic events themselves rather than to indirect measures such as headship rates. One possibility for managing the large state space is to simplify the model by only retaining the demographic events that the household projection in a given setting is most sensitive to. For example, fertility and leaving home are the most important for household formation in developing countries like China, while marriage and union dissolution are important for some developed countries.

2.4 An Existing Probabilistic Household Forecast

The only existing probabilistic household forecast was made by De Beer and Alders at Statistics Netherlands (De Beer & Alders, 1999; Alders, 2001) for the Netherlands up to year 2050. The Statistics Netherlands household forecasting model in general is an approach somewhat between a headship rate model and a macro dynamic model. It consists of four steps:

1. Project the population by age, sex and marital status;
2. Assume future trends in household position rates distinguished by age, sex and marital status, for six household positions: living at the parental home, living alone, living with a partner, being a single parent, living in an institution, and other;
3. Based on the results of steps 1 and 2, calculate the number of persons by age, sex and household position;
4. Calculate the number of households by assuming that persons living alone and lone parents count for one household, persons living with a partner count for a half household, and children living with their parents and people living in an institution count for no households.

Uncertainty in the future number of households is derived by defining uncertainty distributions for household position rates in the year 2050, and interpolating between the resulting population distributions by household position and current conditions. Obviously, the different position rates are not independent: if people within a given population group are more likely to occupy a certain position in a household, then they must be less likely to occupy others. The authors approach this problem by defining uncertainty distributions for each position rate in a step-wise fashion.

- First, an assumption is made about the probability distribution for living in an institution. Conditional on that assumption, and the distribution of population by age and sex, the uncertainty distribution for rates of (and total population) living in private households by age and sex are calculated.
- Second, an uncertainty distribution for the rates of leaving parental home is assumed; combined with the distribution of population living in private households, this yields an uncertainty distribution for the population leaving parental home.
- Third, given the uncertainty distribution for the population leaving parental home, the uncertainty distribution for the probability of living alone is assumed. The rest of the people are mainly persons living with a partner.

For the year of 2050, the uncertainty distributions for the percentage of the population in the different household positions are derived by repeated draws from these underlying uncertainty distributions. Perfect correlation across ages and sex is assumed. Each draw yields the number of persons by age, sex and household position, and then the number of households. The percentages for years between the baseline and 2050 are interpolated, implying perfect autocorrelation.

Results show that in 2050 the 67% uncertainty interval ranges from 7.5 million to 9 million. The width of this interval is smaller than that between deterministic high and low variants from Statistics Netherlands. While De Beer & Alders (1999) presents only the uncertainty in total numbers of households, uncertainty intervals for average household size and for population (and proportions) by five household positions (living alone and lone parent are combined) are also available from this model (M. Alders, personal communication).

This approach has the advantage of combining assumptions about demographic events with household formation and dissolution, while avoiding the problem of requiring large amounts of data unavailable from conventional demographic sources. However, the principal limitation is that while it provides abundant information on the population in different household positions, it produces limited information about household types and size. For many users, the most important information from a household forecast is the number of households, the distribution of households by size, and the distribution of households by age of the householders, especially for application to analysis of consumption (MacKellar *et al.*, 1995; Jiang, 1999; O'Neill & Chen, 2002; Prskawetz *et al.*, 2001; Karl, 2000).

3 A New Approach—the Size/Age-Specific Headship Rates Model

We propose a new method of making household forecasts that extends the traditional headship rate method. Instead of applying total age-specific headship rates to the future population, the new method applies age-specific headship rates decomposed by household size. In addition, changes in these age- and size-specific headship rates are modeled as functions of the demographic processes to which they are most sensitive. This method has the advantages that (1) it produces output on future changes in household numbers, size distributions, and distributions by the age of the householder, which have been identified as the most useful characteristics in applied problems; (2) it explicitly models the relationships between demographic events and headship rates, rather than treating them as a "black box"; and (3) it is simple enough that the resulting model could be used as the basis for more computationally intensive applications, such as probabilistic projections, or integrated assessments of interdisciplinary issues.

We illustrate the new approach with the cases of U.S. and China. Using census data, we analyze the changes in age-specific headship rate by household size for the period 1960–1990 in the U.S. and 1982–2000 in China (Figures 1 and 2). We describe observed changes in rates, and propose qualitative explanations for these changes in terms of trends in specific demographic events. We argue that these explanations provide a strong basis for specification of a model of size-specific headship rates and describe a model framework in terms of key variables likely to explain most of the changes in these rates. This framework could be used as the basis for a new household projection model and estimated from data and from the results of more complex projection models.

Figures 1a and 2a show that the *total* age-specific headship rates for the U.S. and, especially, for China have been quite stable over the past decades. In the U.S., headship rates have increased somewhat over this period, but the general age pattern has not changed much. There are clear differences across countries: China has lower headship rates beyond about age 55, and before age 35. The lower headship rate in early adulthood is driven primarily by later age at leaving parental home compared to most Western countries. The low rates at older age groups are a result of the preference for living with an adult child (mostly a son). After retirement, Chinese elderly usually pass on the household head title to their adult children, resulting in a significant decrease of headship rate in older age groups. In contrast, most American elderly continue to head a household so that headship rates increase until age 75.

The relative stability of *total* headship rates in both countries masks the fact that there have been substantial changes in the propensity of household formation and household dissolution over the past few decades. Figures 1 and 2 show that *size-specific* headship rates in U.S. and China have changed considerably, especially in China. These changes have been driven by trends in demographic rates such as fertility, mortality, union formation and dissolution, and leaving home. We argue that changes in headship rates for each household size are traceable to trends in particular kinds of demographic rates, and thus make a theoretically grounded basis for a projection model.

Before considering each household size in turn, note that in general, the headship rates for different sizes display clear overall age patterns in both the U.S. and China. For one- and two-person households, the headship rate is high for the elderly above age 60 and for young adults below age 30. For households with 3 or more members, the headship rates are low in young adult age groups, increase and reach a peak for the age group 30–45, then decrease and become very small for the oldest old.

For households of size 1, the shape of the age-specific headship rate curve at younger ages is

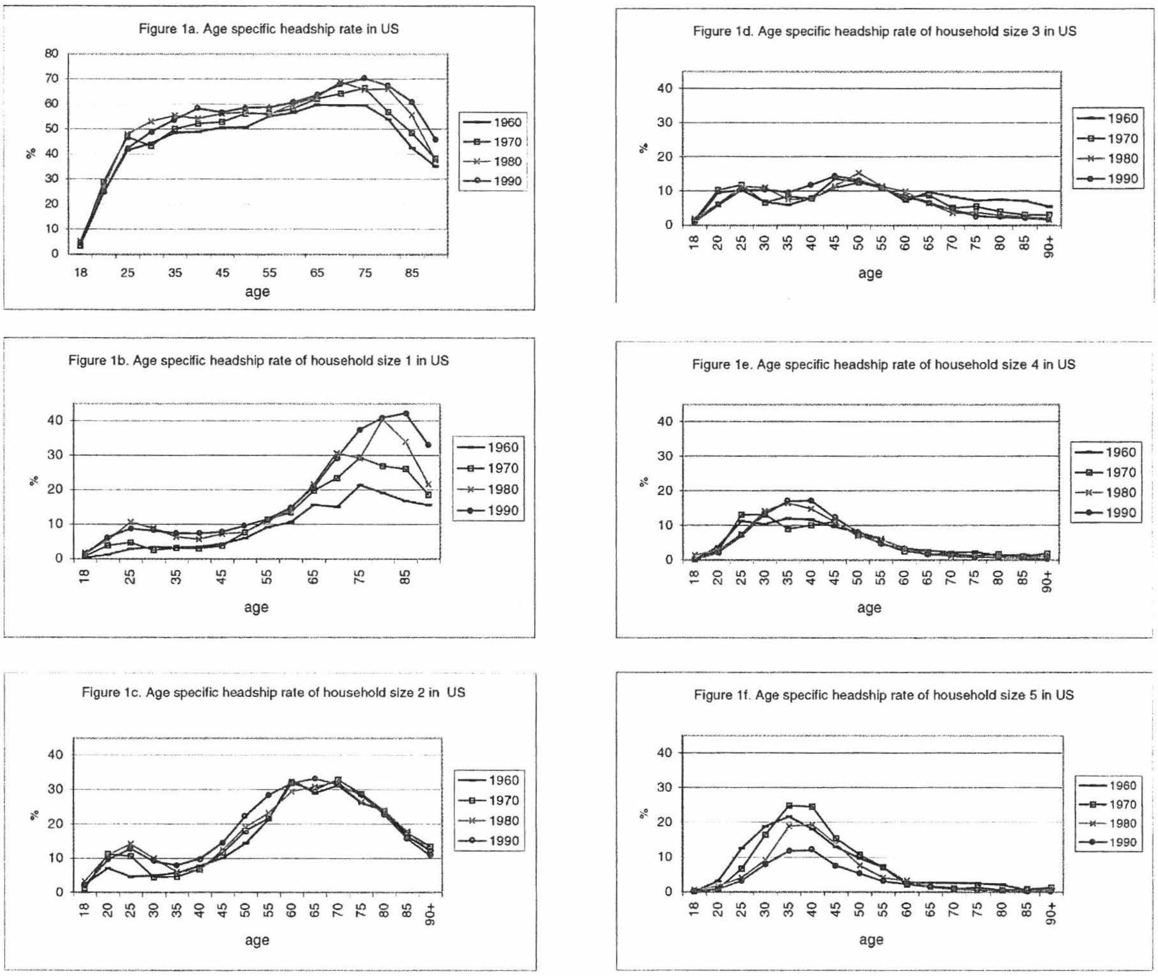


Figure 1. *Age-specific headship rate by household size in U.S., 1960–1990.*

determined primarily by the age-specific rates at which children leave parental homes, and the age-specific rates of marriage and divorce. At older ages, changes in life expectancy and the changes in the gender disparity of life expectancy are probably the primary determinants of the shape of the curve, with smaller contributions from changes in the divorce/remarriage rate and in preferences for living independently from children.

Figure 1 indicates that headship rates in the U.S. increased the most for households of size 1 over the period of 1960–1990. The increase is especially significant among the elderly. We hypothesize that changes in life expectancy have been the primary reason. In the U.S., life expectancy at birth increased from 69.9 to 75.4 between 1960 and 1990. By itself, this improvement would be expected to be associated with a greater likelihood of living with a spouse at older ages; all else equal, this would actually decrease headship rates of one-person households. However, at the same time life expectancy was increasing, the difference in life expectancy between men and women increased as well, from 6.6 to 7.1 years over the same period. The difference in remaining life expectancy at age 65 increased even more, from 2.8 to 3.9 years (Federal Interagency Forum on Aging-Related Statistics, 2002). The trend toward larger gender differences would be expected to lead to increased numbers of single person households. Rates of union dissolution have increased over this period and have

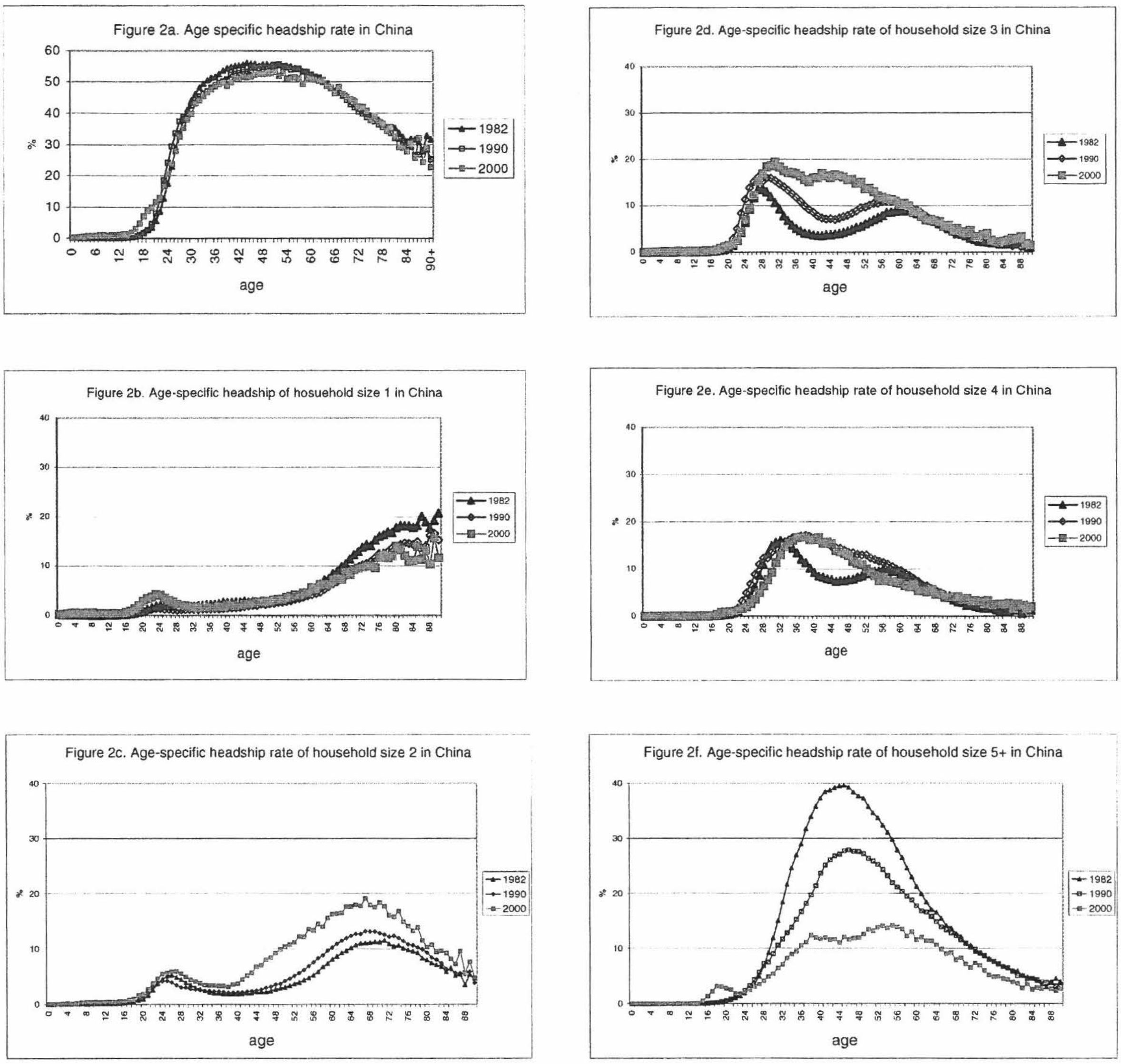


Figure 2. *Age-specific headship rate by household size of China, 1982–2000.*

likely contributed to the change in headship rates as well. The divorce rate (number of divorces per 100 existing marriages) increased from 9% in 1960 to 21% in 1987 (Popenoe, 1993); the proportion of a given marriage cohort that voluntarily ends their union was around 33% in the middle of the century, but increased to more than 50% in the end of 1980s (Furstenberg, 1990). Increasing union dissolution would lead to larger number of people entering the older age groups without a spouse, and thus more likely to be living alone. Lastly, improvements in the health and financial status of older Americans may have led to a continuation of a trend toward preferences for living independently at older ages. At the beginning of the 20th century, more than 70 percent of people aged 65 or older resided with kin. By 1980, only 23 percent of elderly lived with relatives (Bianchi & Casper, 2000). Older Americans now are more likely to spend their later years with spouse or living alone than with adult children.

Changes in the one-person household headship rate at younger ages have likely been driven by changes in the age at leaving home (an event which tends to create one-person households), and the age of union formation (an event which tends to reduce the number of one-person households).

Before 1950, most young Americans left their parental home for or after marriage. Since then, more and more young people leave home for education, employment or simply for increased independence. At the same time, Americans were significantly postponing marriage. In 1960, only 10 percent of women ages 25 to 29 had never married. By the end of the century, more than 40% of women in that age group had not been married, and the majority (52%) of men were still unmarried at these ages (Bianchi & Casper, 2000). The combined effect of these trends led to increases in the one-person headship rate between 1960–1980. During the 1980s, there was a slight *decrease* in the headship rate for young adults. This occurrence actually strengthens the case for ascribing changes to changes in leaving home and union formation/dissolution. During that decade, the average age of leaving parental home increased and the proportion returning to parental home increased, because of the increase in housing prices and improved standards of living in the parental home (Goldscheider *et al.*, 1999). Similarly, the divorce rate has declined since the early 1980s (Bianchi & Casper, 2000); given that most divorces happen among young couples below aged 30, this trend counteracts the effects of later marriage.

Trends in China for one-person headship rates at older ages were opposite to the general trend in the U.S., showing a decrease in the period 1982 to 2000 (Figure 2). This trend is consistent with the observed increase of life expectancy over this period (from 67.8 to 70.0), coupled with no change in the gender difference in life expectancy. Since divorce rates have not changed substantially, and neither have preferences (the elderly still prefer to live with children, particularly in rural China), these two factors did not influence one-person headship rates as they probably did in the U.S. The observed increase in the headship rate in younger age groups most likely resulted from a trend toward more and more young people leaving the parental home for the purpose of education, jobs, or for independence.

For two-person households, changes in the rate and mean age of marriage (for single people) and of first birth and divorce (for married couples) are the primary determinants the shape of headship rate in the early adulthood group. For the late-middle aged group, changes in the propensity of leaving parental home may affect the headship rate when there is only one child remaining in a household. At older ages, mortality is the prime determinant.

In the United States, although the two-person headship rate did not change much for the elderly, it increased somewhat for other age groups, most likely due to the lengthening of the interval between marriage and first birth (Morgan, 1996). Another reason may be the increase in the divorce rate, which one would expect to create a larger number of single parent households with one child. An increase in non-marital fertility may also have played a role (Popenoe, 1993).

In China, there was a striking increase in the two-person headship rate above age 40, especially between 1990 and 2000. This trend is likely a reflection of the increasing number of children establishing their own nuclear households, leaving parents behind in an "empty nest". At older ages, this reflects the improvement in mortality, which prolongs the period during which both spouses are living together.

For households of size 3, the shape of the headship rate curve in the childbearing age groups changes with fertility of parity 1 and 2, both of which can affect the headship rate in either direction depending on the number of parents in the household. Union dissolution rates will also affect headship rates for this size household, as will changes in the propensity of leaving parental home.

In the U.S., the three-person headship rate decreased for older age groups due to the increasing propensity of children to leave the parental home and of the elderly to live independently from their children. Changes at younger ages are subtler. One identifiable trend is the shift in the first peak in headship rates from the mid-20s age group to the early 30s age group (a second, slightly larger peak follows in the mid to late 40s). This shift is likely due to the increase in the mean age of first births. The probability of completing first-birth by women aged 35 decreased from more than 45% in the 1950s to less than 30% in the 1970s (Morgan, 1996). The slight dip in the headship rate after this initial peak is due largely to second births, and the second peak after age 45 is due to the first children leaving home.

Changes in the headship rate for size 3 in China have been quite different. The most important change has been the large increase in the headship rate for the middle age groups, due to an increasing proportion of couples with only one child since the early 1980s. A second change has been first an increase (during the 1980s) and then a decrease (during the 1990s) of the headship rate in the early 20s age group. This pattern likely reflects the trend toward younger marriage and earlier first births in the 1980s, as the family planning policy of "Later marriage, longer birth interval, and fewer children" implemented in the 1970s was relaxed, and then a reversal toward later marriages and first births during the 1990s due to socio-economic and cultural changes supporting these trends.

For four-person households, changes in fertility at parity 2 and 3 are the main determinant of changes in the pattern of headship rate. For example, in the U.S., the peak in the headship rate curve increased in magnitude and shifted from occurring in the mid to late 20s in the 1960s and 1970s, to the mid to late 30s in the 1980s and 1990s. This change was driven by a shift to later second births and a dramatic decrease in the probability of giving third births (Morgan, 1996).

In China the headship rate increased in the middle age groups due to a decline in the probability of third births, especially in rural areas during the 1980s. The headship rate declined during the 1990s for people aged 25–35 because young couples increasingly had only one child. Headship rates also declined in the 1990s for the late middle age and early old age groups, most likely due to the growing number of first children moving out for job opportunities and education.

The headship rate pattern for size 5+ is driven largely by fertility of parity 3+ and the propensity of leaving parental home. For example, the U.S. generally experienced a large decrease in headship rates, and a shift in the peak rate to older ages. This pattern reflects later childbearing, along with a sharp decline in fertility since the end of the 1960s. Increasing union dissolution and early ages at leaving home have also contributed.

In China, there have been very large decreases in the 5+-person headship rate over the past 20 years, due to rapid fertility declines. However, there is a small increase of headship rate for ages 16–24, a result of an increase in the number of young people leaving parental home for jobs and education, and sharing living space in order to reduce the expense of housing.

In summary, the evolution of age patterns of size-specific headship rates has a clear connection to the changes of demographic events. The opposite approach, tracing the impacts across household sizes of given demographic events, can also give insight into the functional relationships between events and headship rates.

For example, changes in parity-specific fertility affect headship rates for different sized households in a number of ways, as we have discussed. The increase in fertility of first births will change the headship rate for the childbearing age group (15–49) as follows: decrease headship rates of size 1 (as single parents have a child and become two person households), increase headship rates of size 2 for single parent households (the consequence of the first effect), decrease headship rates of size 2 for two parent households, and increase headship rates of size 3 of two parent households. Effects of fertility by parity on headship rates for all household sizes are summarized in Table 1.

Similarly, the impacts of increasing divorce rates on the age patterns of size-specific headship rates are summarized in Table 2. Empirical studies indicate that divorce rate is the highest among those aged 25–35. Young heads with no children tend to have less stable marriages than those who have children. Therefore, an increase in the divorce rate will increase headship rates of size 1 for the young age group, decrease headship rates of size 2 for young couples without children, increase headship rates of size 2 for single parents with one child, and decrease headship rates of size 3 for couples with one child.

Table 1

Impacts of increase in the total fertility rate on headship rates, while all else is equal.

Headship rate	Parity		
	1	2	3
Household size 1	- (15-49, 1 parent)	No effect	No effect
Household size 2	+ (15-49, 1 parent) - (15-49, 2 parents)	- (15-49, 1 parent)	No effect
Household size 3	+ (15-49, 2 parents)	+ (15-49, 1 parent)	No effect
Household size 4	No effect	+ (15-49, 2 parents)	+ (15-49, 1 parent) - (15-49, 2 parents)
Household size 5+	No effect	No effect	+ (15-49, 2 parents)

Table 2

Impact of the increase in divorce rate on headship rates while else is equal.

Household Size	Effect on headship rate in young age group
1	+
2	- (couple with no child) + (couple with one child)
3	- (couple with one child)
...	

4 Conclusion: Toward a New Model

These examples strongly suggest that, based on current understanding and empirical data, changes in the age patterns of size-specific headship rates can be modelled by parameterizing the changes for different age groups as a function of the dynamics of important demographic events. We propose that age- and size-specific headship rates are best modelled as the sum of three functions, one describing each of three broad age categories that serve as proxies for different life cycle stages: young adulthood, middle age, and elderly. Thus, the starting point for our basic model, based on our preliminary analysis for the U.S. and China, is:

$$h(i,x) = g_1(i,x) + g_2(i,x) + g_3(i,x) + \varepsilon(i,x)$$

where $h(i,x)$ is the headship rate for household size i at age x, $g(i,x)$ is a function capturing effects acting mainly on one of three broad age categories (1 = young adult; 2 = middle age; 3 = elderly), and ε is an error term. In turn,

$$g(i,x) = f(r_1, r_2, r_3, \ldots)$$

where r_n is a variable describing a particular demographic rate. The demographic rates that we expect to influence each of our three "g" functions are given in Table 3, which summarizes the discussion in the previous section.

Existing empirical data will likely be insufficient to estimate all model parameters, given that the household formation and dissolution involves many demographic events, and data for some events are unavailable in conventional population data sources. Thus, it would likely be necessary to use "synthetic data" generated from a more complex household projection model, such as existing dynamic household models. For example, one strategy would be to generate a number of scenarios that systematically vary different demographic rates, and observe the effect on size-specific headship rates. Model parameters could then be estimated from this "data". The end result would be a reduced-form model with parameters estimated such that it reproduced the behavior of the more complex model, and could be tested against empirical data for specific countries.

Table 3

Demographic events and parameters of importance to a model of size-specific headship rates.

Household Size	Demographic events and parameters		
	Early adulthood group	Middle age group	Senior age group
1	Propensity of leaving home, age and rate of marriage, divorce rate	Marriage, remarriage and divorce rate, propensity of leaving home	Mortality, propensity of leaving home
2	Age and rate of marriage, age and rate of first birth, divorce rate	Propensity of leaving home, divorce rate, mortality	Mortality, propensity of leaving parental home
3	Fertility rate of first and second birth, divorce rate	Propensity of leaving home, divorce rate	Mortality, propensity of leaving home
4	Fertility rate of second and third birth, divorce rate	Fertility of second and third birth, propensity of leaving home	Mortality, propensity of leaving home
...	...	...	...

Our extension of the headship rate approach is amenable for use in probabilistic forecasting because it reduces the complexity of the model (compared to full macro-demographic household models) while retaining demographic processes as important inputs to which expert opinion and other forms of information can be applied in order to develop forecasts. For example, in this model one might specify probability distributions not only for fertility, mortality, and migration, but also for variables such as the mean age at leaving home, mean ages of parity-specific births, and rates and mean ages of marriage and divorce. Judgment, supported by empirical analysis, would also need to be applied to the correlation matrix. For example, in many countries changes in age at marriage are likely to be correlated with changes in age at first birth, and fertility may be correlated with mean age at leaving home. Careful attention would need to be paid to these possibilities.

Here we have presented the basic model framework, and used data to support our argument regarding its viability. Clearly, further work is required to develop a first illustrative model for testing. However, our analysis demonstrates one plausible strategy for achieving the goal of probabilistic household projections.

Acknowledgements

Jiang Leiwen's work was supported by a grant from the U.S. Department of Energy program on Integrated Assessment of Global Climate Change Research.
We thank Wolfgang Lutz for helpful conversations in formulating ideas for this paper.

References

Alders, M. (2001). Huishoudensprognose 2000–2050: veronderstellingen over onzekerheidsmarges. *Maandstatistiek van de bevolking*, Jaargang **49** (August) (in Dutch).
Alho, J.M. (1998). *A Stochastic Forecast of the Population of Finland*. Helsinki: Statistics Finland.
Bianchi, S.M. & Casper, L.M. (2000). American families. *Population Bulletin*, **55**(4).
Bongaarts, J. & Zimmer, Z. (2001). Living arrangements of older adults in the developing world: Analysis of DHS household surveys. Policy Research Division Working Paper no. 148. New York: Population Council.
Caldwell, S.B., Greene, W., Mount, T., Saltzman, S. & Broyd, R. (1979). Forecasting regional energy demand with linked macro/micro models. *Papers of the Regional Science Association*, **43**, 99–113.
Carliner, G. (1975). Determinants of household headship. *Journal of Marriage and the Family*, **37**(1), 28–38.
De Beer, J. & Alders, M. (1999). Probabilistic population and household forecasts for the Netherlands. Paper for the European Population Conference, August 30–September 3, 1999, The Hague, The Netherlands.
Duley, C., Rees, P. & Clarke, M. (1988). A microsimulation model for updating households in small areas between census. Paper presented at the NIDI/RUU workshop "Multi-state demography: measurement, analysis, forecasting", Oct. 31–Nov. 4, 1988, Zeist, The Netherlands.

Federal Interagency Forum on Aging-Related Statistics. (2002). http://www.aoa.dhhs.gov/agingstats.
Furstenberg, F.F., Jr. (1990). Divorce and the American family. *Annual Review of Sociology*, **16**.
Galler, H. (1988). Microsimulation of household formation and dissolution. In *Modelling Household Formation and Dissolution*, Eds. N. Keilman, A. Kuijsten and A. Vossen, pp. 139–159. Oxford, UK: Clarendon Press.
Goldscheider, F., Goldscheider, C., St. Clair, P. & Hodges, J. (1999). Changes in returning home in the United States, 1925–1985. *Social Forces*, **78**(2).
Goldstein, J. & Wachter, K. (2001). Survey-based stochastic kinship forecasting. Paper presented at the General Conference of IUSSP, August 2001, Salvador, Brazil.
Gonnot, J.-P., Keilman, N. & Prinz, C. (Eds.) (1995). *Social Security, Household and Family Dynamics in Aging Societies*. Boston: Kluwer Academic Publishers.
Hammel, E., Hutchinson, D., Wachter, K., Lundy, R. & Deuel, R. (1976). The SOCSIM demographic-sociological microsimulation program. Research Series No. 27. Berkeley, CA: Institute of International Studies, University of California.
Hollmann, F.W., Mulder, T.J. & Kallan, J.E. (2000). Methodology and assumptions for the population projection of the United States: 1999 to 2100. Working Paper No. 38. Washington, DC: Population Division, U.S. Census Bureau.
Jiang, L. (1999). *Population and Sustainable Development in China—Population and Household Scenarios for Two Regions*. Amsterdam: Thelasis.
Karl, G. (2000). *Human Settlement Statistics*. United Nations Center for Human Settlements (UNCHS).
Keilman, N. (2003). The threat of small households. *Nature*, **421**(6922), 489–490.
Keilman, N., Kuijsten, A. & Vossen, A. (Eds.) (1988). *Modelling Household Formation and Dissolution*. Oxford, UK: Clarendon Press.
King, D. (1999). Official household projections in England: Methodology, usage and sensitivity tests. Paper presented at "Joint ECE-EUROSTAT Work Session on Demographic Projections", May 3–7, 1999, Perugia, Italy.
Kobrin, F.E. (1973). Household headship and its changes in the United States, 1940–1960, 1970. *Journal of the American Statistical Association*, **68**(344), 793–800.
Kuijsten, A. & Vossen, A. (1988). Introduction. In *Modelling Household Formation and Dissolution*, Eds. N. Keilman, A. Kuijsten and A. Vossen, pp. 3–12. Oxford, UK: Clarendon Press.
Lee, R.D. (1999). Probabilistic approaches to population forecasting. In *Frontiers of Population Forecasting*, Supplement to Vol. **24**, 1998, Eds. W. Lutz, J.W. Vaupel and D.A. Ahlburg. New York: Population and Development Council.
Liu, J., Daily, G., Ehrlich, P. & Luck, G. (2003). Effects of household dynamics on resource consumption and biodiversity. *Nature*, **421**(6922), 530–533.
MacKellar, F.L., Lutz, W., Prinz, C. & Goujon, A. (1995). Population, households and CO_2 emissions. *Population and Development Review*, **21**(4), 849–865.
Martin, N. (1999). Population, households and domestic water use in countries of the Mediterranean Middle East. Interim Report IR-99-032. Laxenburg, Austria: IIASA.
Morgan, P. (1996). Characteristic feature of modern American fertility. In the Supplement "Fertility in the United States: New Patterns, New Theories", pp. 19–63. *Population and Development Review*, Volume **22**. New York: Population Council.
Muller, C., Gnanasekaran, G.S. & Knapp, K. (1999). *Housing and Living Arrangements of the Elderly, An International Comparison Study*. New York: International Longevity Center–USA, Ltd. Almanac Phase 4.
Nelissen, J. (1991). Household and education projections by means of a microsimulation model. *Economic Modelling*, **8**(4), 480–511.
Nelissen, J. & Vossen, A. (1989). Projecting household dynamics: a scenario-based microsimulation approach. *European Journal of Population*, **5**(3), 253–279.
O'Neill, B.C. (2004). Conditional probabilistic projections: an application to climate change. *Internat. Statist. Rev.*, **72**(2), forthcoming.
O'Neill, B.C. & Chen, B.S. (2002). Demographic determinants of household energy use in the United States. In *Population and Environment: Methods and Analysis*, Eds. W. Lutz, A. Prskawetz and W.C. Sanderson, pp. 53–88, Supplement to Vol. **28**, 2002, *Population and Development Review*, New York: Population Council.
O'Neill, B.C., MacKellar, F.L. & Lutz, W. (2001). *Population and Climate Change*. Cambridge, UK: Cambridge University Press.
Oskamp, A. (1997). *Local Housing Market Simulation: A Micro Approach*. Amsterdam: Thesis Publishers.
Popenoe, D. (1993). American decline, 1960–1990: A review and appraisal. *Journal of Marriage and the Family*, **55**(3), 527–542.
Prskawetz, A., Jiang, L. & O'Neill, B. (2001). Demographic composition and projections of car use in Austria. Paper presented at the General Conference of IUSSP, August 2001, Salvador, Brazil.
Select Committee on Environment, Transport and Regional Affairs. (1998). *Tenth Report*. London, UK: The United Kingdom Parliament.
Smith, J.E. (1987). The computer simulation of kin sets and kin counts. In *Family Demography: Methods and their Application*, Eds. J. Bongaarts, T.K. Burch and K.W. Wachter, pp. 249–266. Oxford, UK: Clarendon Press.
Sweet, J.A. (1984). Components of changes in the number of households: 1970–1980. *Demography*, **21**(2), 129–140.
Van Diepen, A. (1994). Demografische ontwikkelingen en milieugebruik. In *Bevolkingsvraagstukken in Nederland Anno 1994: Demografische Ontwikkelingen in Maatschappelijk Perspectief*, Eds. N. van Nimwegen and G. Beets, pp. 357–374. NIDI Report No. 35. The Hague, the Netherlands: NIDI (in Dutch).
Van Diepen, A. (1995). Population, land use and housing trends in the Netherlands since 1950. Working Paper WP-95-63. Laxenburg, Austria: IIASA.
Van Imhoff, E. & Keilman, N. (1991). *Lipro 2.0: An Application of a Dynamic Demographic Projection Model to Household Structure in The Netherlands*. Amsterdam/Lisse and Berwyn, PA: Swets & Zeitlinger Inc.

Wachter, K. (1987). Microsimulation of household cycles. In *Family Demography: Methods and their Application*, Eds. J. Bongaarts, T.K. Burch and K.W. Wachter, pp. 215–227. Oxford, UK: Clarendon Press.
Wachter, K. (1998). SOCSIM: Description of the Program, `http : //demog.berkeley.edu`.
Zeng, Y., Vaupel, J.W. & Wang, Z. (1997). A multidimensional model for projecting family households—with an illustrative numerical application. *Mathematical Population Studies*, **15**(3), 187–216.
Zeng, Y., Vaupel, J.W. & Wang, Z. (1999). Household projection using conventional demographic data. In *Frontiers of Population Forecasting*, Eds. W. Lutz, J.W. Vaupel and Dennis A. Ahlburg, Supplement to Vol. **24**, 1998, *Population and Development Review*. New York: Population Council.

Résumé

Les projections de ménages sont des éléments clés des analyses de problèmes sociaux tels que la situation des personnes âgées, le logement, les habitudes de consommation et leurs conséquences sur l'environnement. Les chercheurs ou responsables politiques qui utilisent de telles projections ont besoin de représentations correctes de l'incertitude pour étayer leurs analyses. Cependant, les faiblesses de l'approche traditionnelle consistant à fournir des variantes alternatives à la projection fondée sur la "meilleure hypothèse" sont amplifiées dans les projections de ménages, qui produisent en sortie de nombreuses variables intéressantes et utilisent en entrée nombre de variables autres que la fécondité, la mortalité et les migrations. Nous examinons les méthodes actuelles de projections de ménages et leur potentiel d'utilisation pour produire des projections probabilistes palliant toutes ces faiblesses. Nous proposons ensuite un nouveau cadre pour une méthode d'une complexité moyenne qui à nos yeux pourra fournir une bonne base pour un développement ultérieur des projections probabilistes des ménages. Extension de l'approche traditionnelle des taux directeurs, cette méthode est fondée sur la modélisation des changements dans ces taux, décomposés par taille des ménages, en fonction de variables décrivant les évènements démographiques tels que: fécondité spécifique de parité, formation et dissolution d'une union, départ du domicile. Elle a peu d'exigences de données, une complexité gérable, permet la spécification directe des évènements démographiques et produit en sortie les caractéristiques les plus importantes des ménages pour de nombreuses applications. Une illustration de la façon dont le modèle pourrait être construit, utilisant des données des Etats-Unis et de la Chine sur les dix dernières années, démontre la viabilité de cette approche.

[*Received May, 2003, accepted December, 2003*]

International Statistical Review (2004), **72**, 1, 65–79, Printed in The Netherlands

Assumptions on Fertility in Stochastic Population Forecasts*

Maarten Alders and Joop de Beer

Statistics Netherlands, Department of Statistical Analysis, Voorburg, The Netherlands. E-mail: mals@cbs.nl; jber@cbs.nl

Summary

In recent years Statistics Netherlands has published several stochastic population forecasts. The degree of uncertainty of the future population is assessed on the basis of assumptions about the probability distribution of future fertility, mortality and migration. The assumptions on fertility are based on an analysis of historic forecasts of the total fertility rate (TFR), on time-series models of observations of the TFR, and on expert knowledge. This latter argument-based approach refers to the TFR distinguished by birth order. In the most recent Dutch forecast the 95% forecast interval of the total fertility rate in 2050 is assumed to range from 1.2 to 2.3 children per woman.

Key words: Stochastic population forecasts; Uncertainty; Expert knowledge; Fertility.

1 Introduction

The Dutch population forecasts published by Statistics Netherlands every other year project the future size and age structure of the population of the Netherlands up to 2050. The forecasts are based on assumptions about future changes in fertility, mortality, and international migration. Obviously, the validity of assumptions on changes in the long run is uncertain. It is important that users of forecasts are aware of the degree of uncertainty. In order to give information about the degree of uncertainty in population forecasts, Statistics Netherlands produces stochastic population and household forecasts (De Beer & Alders, 1999). Instead of publishing two alternative deterministic (low and high) variants in addition to the medium variant, as was the practice up to a few years ago, forecast intervals are made. These intervals are calculated by means of Monte Carlo simulations. The simulations are based on assumptions about the probability distributions of future fertility, mortality, and international migration (see e.g. Alders & De Beer, 2002).

In the Dutch population forecasts the assumptions on the most likely future changes in fertility are specified by making assumptions on the future values of the total cohort fertility rate (CFR) distinguished by birth order. In the most recent Dutch forecasts, assumptions underlying the medium variant, describing the most likely development, are based on a detailed analysis of age-specific fertility rates, both from a cohort and a period perspective. Moreover, evidence on the desired number of children from the Netherlands Family and Fertility Survey is used. These analyses have lead to the assumption in the 2000 Population forecast that future generations of women will have 1.75 children per woman on average. In the long run, age-specific fertility rates are assumed to remain constant. As a consequence, the total fertility rate in 2050 equals the average number of children for women born

*Paper presented at the seminar "How to deal with uncertainty in population forecasting?", 12–14 December 2002, Vienna, Austria.

around 2020. It should be noted that the description 'medium variant' refers to the former practice when several deterministic variants were published. Since no variants are published anymore, it is no longer appropriate to speak of a medium variant. However, abandoning this terminology would make users think that the medium variant is different from the most likely forecast. For this reason we will still use the term 'medium variant' when we are referring to the most likely forecast.

This paper examines how assumptions on the uncertainty of future changes in fertility in the long run can be specified. The use of expert knowledge plays an important role in the assumption making process. Section 2 describes the general methodology that is used for stochastic population forecasting. In section 3 the methods are applied to the Dutch fertility forecasts.

2 Methods for Stochastic Forecasting

Population forecasts are calculated by means of the cohort component model. The input of the model is based on assumptions about future changes in fertility, mortality, and migration. In the Dutch population forecasts, assumptions on fertility refer to age-specific rates distinguished by parity, mortality assumptions refer to age- and sex-specific mortality rates, assumptions about immigration refer to absolute numbers distinguished by age, sex and country of birth, and assumptions on emigration are based on a distinction of emigration rates by age, sex and country of birth.

Forecast intervals can be derived from simulations. On the basis of an assessment of the probability of the bandwidth of future values of fertility, mortality, and migration, the probability distribution of the future population size and age structure can be calculated by means of Monte Carlo simulations. For each year in the forecast period, values of the total fertility rate, life expectancy at birth of men and women, numbers of immigrants and emigration rates are drawn from the probability distributions. Subsequently, age-sex-specific fertility, mortality and emigration rates, and immigration numbers are specified. Each draw results in a population by age and sex at the end of each year. Thus the simulations provide a distribution of the population by age and sex in each forecast year.

To make the simulations, several assumptions have to be made. First, the type of probability distribution has to be specified. Subsequently, assumptions about the parameter values have to be made. The assumption about the mean or median value can be derived from the medium variant. Next, assumptions about the value of the standard deviation have to be assessed. In the case of asymmetric probability distributions, additional parameters have to be specified. Finally, assumptions about the covariances between the forecast errors across age, between the forecast years, and between the components have to be specified (see e.g. Lee, 1996).

The main assumptions underlying the probability distribution of the future population relate to the variance of the distributions of future fertility, mortality, and migration. The values of the variance can be assessed by using one of the following three methods:

a. an analysis of errors of past forecasts
b. model-based estimates of forecast errors
c. expert knowledge or judgement.

These methods do not exclude each other; rather they may complement each other. For example, even if the estimate of the variance is based on past errors or on a time-series model, judgement plays an important role. In publications, however, the role of judgement is not always made explicit.

2.1 An Analysis of Errors of Past Forecasts

The probability of a forecast interval can be assessed on the basis of an analysis of the errors of forecasts published in the past. On the assumption that the errors are approximately normally distributed—or can be modelled by some other distribution—and that the future distribution of the errors is the same as the past distribution, these errors can be used to calculate the probability of forecast intervals of new forecasts. Keilman (1990) examines the errors of forecasts of fertility, mortality, and migration of Dutch population forecasts published between 1950 and 1980. He finds considerable differences between the errors of the three components. For example, errors in life expectancy grow considerably more slowly than errors in the total fertility rate. Furthermore, he examines to what extent errors vary between periods and whether errors of recent forecasts are smaller than those of older forecasts, taking into account the effect of differences in the length of the forecast period.

In order to assess to what extent errors of past forecasts provide useful information on the degree of uncertainty of new forecasts it has to be examined whether new forecasts are better than older ones. This comparison of errors of forecasts made in different periods requires that we know the reasons why the forecaster chose a specific method for a certain forecast period. This information enables us to conclude whether a certain forecast was accurate, because the forecaster chose the right method for the right period, or whether the forecaster was just more lucky in one period than in another. Thus simply comparing the forecast errors of successive forecasts does not tell us whether recent forecasts are better than preceding forecasts.

Another question related to the analysis of errors of past forecasts is how likely it is that similar developments will occur again. The 1965-based forecasts in the Netherlands were rather poor because forecasters did not anticipate the sharp decline in fertility between 1965 and 1975. If it is assumed that it is very unlikely that such developments will occur again, one may conclude that errors of new forecasts are likely to be smaller than those of the 1965-based forecasts. For that reason, the degree of uncertainty of new forecasts can be based on errors of forecasts that were made after 1965.

The decision as to which past forecasts are to be included is a matter of judgement. Obviously, one may argue that an 'objective' method would be to include all forecasts made in the past. However, this implies that the results depend on the number of forecasts that were made in different periods. Since more forecasts were made after 1985 than in earlier periods, the errors of more recent forecasts weigh more heavily in calculating the average size of errors. On the other hand, for long-term forecasts, one major problem in using errors of past forecasts for assessing the degree of uncertainty of new forecasts is that the sample of past forecasts tends to be biased towards the older ones. As for recent forecasts the accuracy cannot yet be checked except for the short term (Lutz, Goldstein & Prinz, 1996). Forecast errors for the very long term result from forecasts made a long time ago.

One way of assessing forecast errors in the long term is to extrapolate forecast errors by means of a time-series model (De Beer, 1997). The size of forecast errors for the long term can be projected on the basis of forecast errors of recent forecasts for the short and medium term. Thus, estimates of ex ante forecast errors can be based on an extrapolation of ex post errors. This method combines the analysis of past errors with the use of model-based estimates of errors.

2.2 *Model-based Estimates of Forecast Errors*

Instead of assuming that future forecast errors will be similar to errors of past forecasts, one may attempt to estimate the size of future forecast errors on the basis of the assumptions underlying the methods used in making new forecasts. If the forecasts are based on an extrapolation of observed trends, ex ante forecast uncertainty can be assessed on the basis of the time-series model used for producing the extrapolations. If the forecasts are based on a stochastic time-series model, the model produces not only the point forecast, but also the probability distribution. For example, ARIMA (Autoregressive Integrated Moving Average)-models are stochastic univariate time-series models that can be used for calculating the probability distribution of a forecast (Box & Jenkins, 1970). Alternatively, a structural time-series model can be used for this purpose (Harvey, 1989). The latter model is based on a Bayesian approach: the probability distribution may change as new observations become available. The Kalman filter is used for updating the estimates of the parameters.

One problem in using stochastic models for assessing the probability of a forecast is that the probability depends on the assumption that the model is correct. Obviously, the validity of this assumption is uncertain, particularly in the long run. If the point forecast of the time-series model does not correspond with the medium variant, the forecaster apparently does not regard the time-series model as correct. Moreover, time-series forecasting models were developed for short term horizons, and they are not generally suitable for long term forecasts (Lee, 1996). Usually, stochastic time-series models are identified on the basis of autocorrelations for short time intervals only. Alternatively, the form of the time-series model can be based on a judgement to constrain the long-run behaviour of the point forecasts such that they are in line with the medium variant of the official forecast (Tuljapurkar, 1996). However, one should be careful in using such a model for calculating the variance of ex ante forecast errors, because of the uncertainty of the validity of the constraint imposed on the model. In assessing the degree of uncertainty of the projections of the model one should take into account the uncertainty of the constraint, which is based on judgement.

Rather than identifying an appropriate time-series model and analytically deriving forecast intervals from the selected model, empirical forecast errors can be assessed by means of calculating the forecast errors of simple baseline projections. Alho (1998) notes that the point forecasts of the official Finnish population forecasts are similar to projections of simple baseline projections, such as assuming a constant rate of change of age-specific mortality rates. If these baseline projections are applied to past observations, forecast errors can be calculated. The relationship between these forecast errors and the length of the forecast period can be used to assess forecast intervals for new forecasts.

2.3 Expert Knowledge or Judgement

In assessing the probability of forecast intervals on the basis of either an analysis of errors of past forecasts or an estimate of the size of model-based errors, it is assumed that the future will be like the past. Instead, the probability of forecasts can be assessed on the basis of experts' opinions about the possibility of changes in trends. For example, fertility forecasts are usually based on the assumption that trends like the increase of female labour force participation will continue. Even though a reversal of this trend may not be assumed to occur in the most likely variant, an assessment of the probability of such an event is needed to determine the uncertainty of the forecast. More generally, an assessment of ex ante uncertainty requires assumptions about the probability that the future will be different from the past. If a forecast is based on an extrapolation of past trends, the assessment of the probability of structural changes which may cause a reversal of trends cannot be derived directly from an analysis of historical data and therefore requires the judgement of the forecaster. Lutz, Goldstein & Prinz (1996) argue that subjective distributions are to be preferred to a time-series approach, because "structural changes and unexpected events are likely to happen". Lutz, Sanderson, Scherbov & Goujon (1996) assess the probability of forecasts on the basis of opinions of a group of experts. The experts are asked to indicate the upper and lower boundaries of 90 percent forecast intervals for the total fertility rate, life expectancy, and net migration up to the year 2030. Subjective probability distributions of a number of experts are combined in order to diminish the danger of individual bias.

In the Dutch population forecasts, an assessment of the degree of uncertainty of fertility forecasts is primarily based on expert judgement, taking into account errors of past forecasts and model-based estimates of the forecast errors.

3 Assumptions on Fertility

In the Dutch population forecasts the assumptions about future changes in fertility are based on a distinction by parity. Accordingly, the assumptions about the standard deviation of forecast errors are based on a distinction by parity. The assumptions relate to the upper and lower limits of a 95% interval. This implies that it is assumed that it is very unlikely that a forecast that is higher than the upper limit or lower than the lower limits will come true.

The assumptions underlying the medium variant refer to cohort fertility. In the long run, age-specific fertility rates are assumed to remain constant. As a consequence, the total fertility rate in 2050 equals the average number of children for women born around 2020. The 2000 Dutch population forecast assumes a level of 1.75 children per woman for the TFR in 2050 (De Jong, 2001).

3.1 An Analysis of Errors of Past Forecasts of Fertility

The first population forecast that was published by Statistics Netherlands dates from 1950. In total 23 forecasts of the TFR are published. Between the mid-eighties and mid-nineties the population forecasts were revised every year. Since 1996 the forecasts for the long run are revised every other year, and for the short run (5 years ahead) in the intermediate years. Figure 1 shows some of the historic forecasts of the TFR. As is clear from figure 1, the 1965-based forecast was rather poor since it did not anticipate the sharp decline in fertility during the late sixties and early seventies.

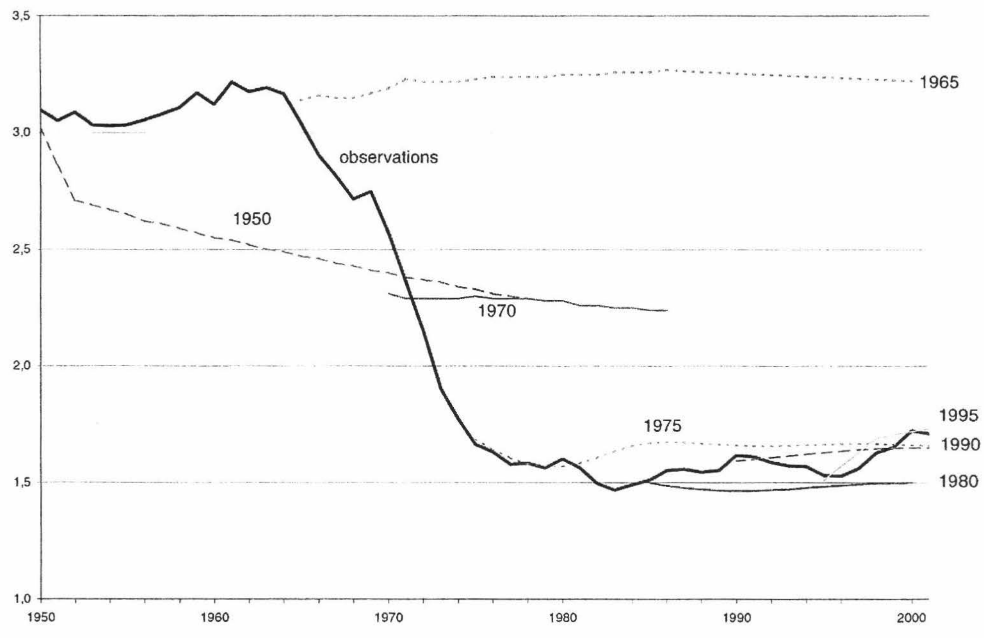


Figure 1. *Historic forecasts and observations of the TFR in the Netherlands.*

To assess the accuracy of the (medium variants of the) historic forecasts, errors are calculated by subtracting the observed values from the forecasted values. An empirical 67% interval was then obtained specifying the lower and upper boundaries such that one sixth of the errors is lower than the lower boundary and one sixth is higher than the upper boundary. Similarly, a 95% interval is calculated. Figure 2 shows the duration specific intervals, together with the mean of the errors. It should be noted that the number of forecasts on which the intervals are based, decreases with the forecast lead time. The intervals for the duration of 3 years are based on 20 forecasts, and those for the duration of 20 years on only 4 forecasts.

What strikes one's attention from figure 2 is that the intervals are rather asymmetric around zero. The mean is significantly positive. This is to a large extent due to the 1965-based forecast. If the same analysis is applied to all forecasts *except* the 1965-based forecast, the mean is close to zero (figure 3). Moreover, the 95% interval is much smaller and more symmetric. As was stated in section 2.1 the question is to what extent this analysis provides useful information about the degree of uncertainty of new forecasts. If it is assumed that it is unlikely that developments that took place in the sixties will occur again, one may assume that errors of new forecasts will be smaller than those of the 1965-based forecast.

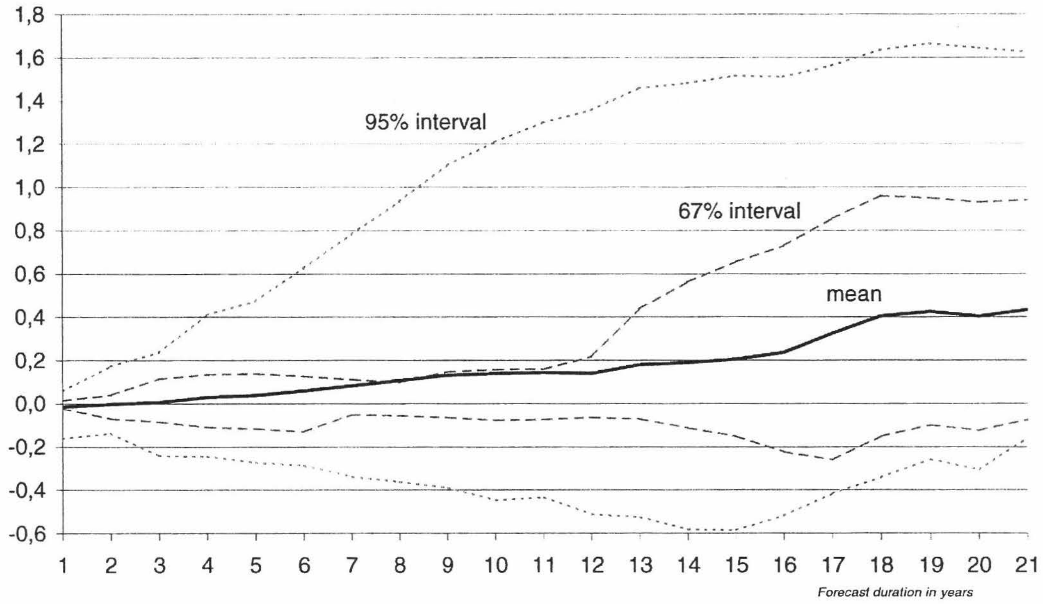


Figure 2. *Mean and 67% and 95% interval of errors of past forecasts of the TFR.*

The question as to which historic forecast to use can only be based on judgement. Figure 4 compares the analysis in which all forecasts are used with analyses in which the 1965-based forecast and the forecasts made before 1975 are excluded respectively. The duration is restricted to 10 years. For this duration at least 14 historic forecasts are available if all forecasts are included.

Two conclusions can be derived from figure 4. First, the differences between the 67% interval and the 95% interval indicate that the errors are not normally distributed. Second, the way in which the errors increase over the forecast lead time is somewhat different than in the latest forecast where it was assumed that the errors increase proportional to the square root of the forecast lead time (like in a random walk model).

One problem in using forecast errors from historic forecasts for assessing intervals of long-term forecasts is that there are no forecast errors for a period of 50 years. Therefore the standard deviation of forecast errors for the long run is assessed on the basis of a projection of forecast errors for shorter time horizons. A random walk model is used for this purpose.

Table 1 shows the 95% interval for the TFR in 2050. If the model is applied to all forecasts since

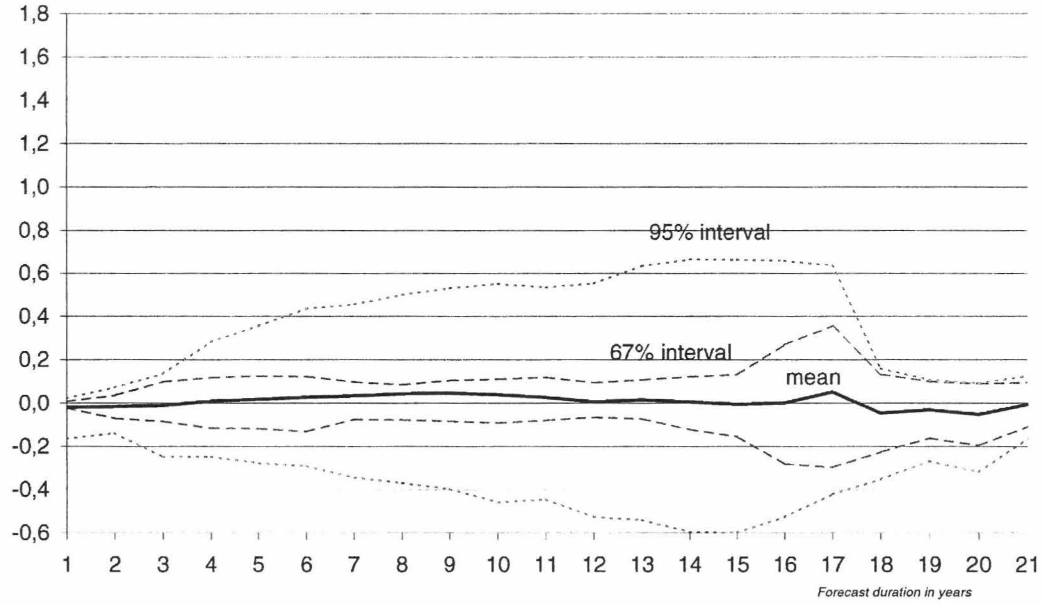


Figure 3. *Mean and 67% and 95% interval of errors of past forecasts of the TFR, excluding 1965-based forecast.*

Table 1

Random walk model of forecasts errors of the TFR.

	95% interval of TFR in 2050	
	Lower bound	Upper bound
All forecasts since 1950	0.89	2.61
All forecasts excl. 1965 Forecast	0.97	2.53
All forecasts since 1975	1.23	2.27
2000 Population forecast	1.20	2.30

1950 the 95% interval ranges from 0.9 to 2.6 children per woman. If only forecasts made since 1975 are included, the interval is again much smaller: 1.2–2.3 children per woman. The interval assumed in the latest Dutch forecast equals that of the forecasts made since 1975. The decision as to which forecasts one is to include cannot be made on purely statistical grounds. Judgement plays a role. This choice is based on an analysis of fertility rates by parity. The large forecast errors in the 1960s and 1970s occurred when fertility was at a considerably higher level than at present. Fertility rates for third and fourth births were much larger. Similar forecast errors in the same direction seem very unlikely in the future. Fertility rates of parity 3 and higher can hardly decline to the same extent as in the past, since they are already rather low. This would imply that fertility rates of parity 1 and 2 would have to decline much more than they did in the past. It would imply, for example, that a vast majority of women would remain childless. The Netherlands Fertility and Family Surveys (FFS), however, indicate that only a small minority of women from the younger generations wish to remain childless. Thus, very low fertility levels would imply that a large number of women could not realise their expectations. Similar large errors in an upward direction do not seem likely either, because of the changed position of women. The strong increase in the female labour force participation rates does not make it very likely that the percentage of women having three or four children will increase

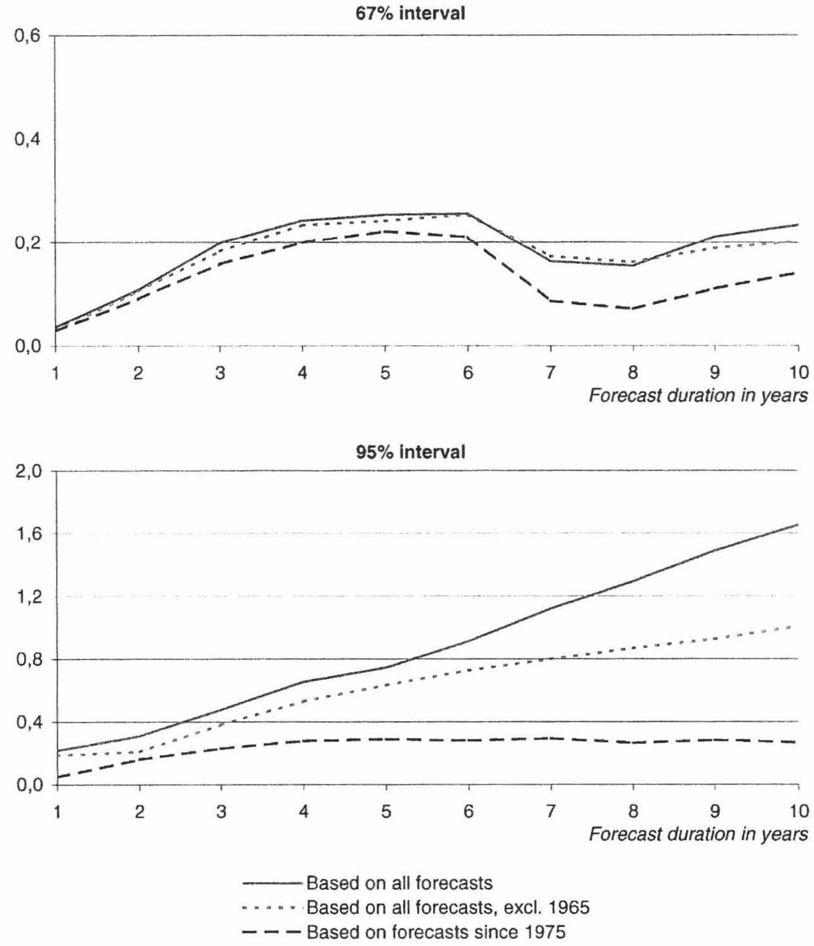


Figure 4. *Width of 67% and 95% interval of errors of past forecasts of the TFR.*

strongly. For this reason it is assumed that the interval of new forecasts is smaller than that based on the errors of the forecasts made before 1975.

It should be noted that the 95% interval in 2050 as modelled by the random walk model of the errors of all forecasts (0.89–2.61 children per woman) is small compared to the corresponding empirical 95% interval as shown in figure 4. Since only a few historic forecasts are available, the latter interval is determined to a large extent by outliers (e.g. the 1965-based forecast). Although such outliers cannot be excluded, the probability that such errors will occur again is very small as was explained before.

3.2 Model-based Estimates of Forecast Errors of Fertility

The assessment of the standard deviation of the TFR can also be based on time-series models. A random walk model (without drift) seems an appropriate choice since the forecasts that were made in the past resemble random walk forecasts. Except for the 1950-based forecast all forecasts assume more or less a continuation of the current level of the TFR (see figure 1). Moreover, whereas other models (with more parameters) may provide a better fit of the data, they may lead to an underestimation of the degree of uncertainty since it is not sure that such a model is also valid for the future TFR.

The random walk model is applied twice. First, it is applied to observations of the TFR since 1950, and second to observations since 1975. Figure 5 shows the results of this model when applied to the TFR.

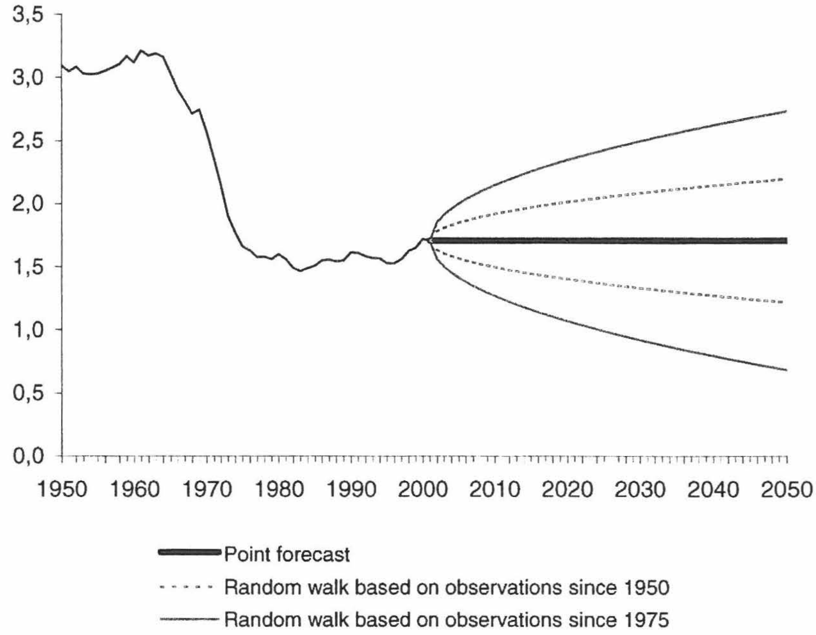


Figure 5. *Point forecast and 95% forecast interval of TFR based on a random walk model.*

Not surprisingly it makes a big difference whether or not observations before 1975 are included. If the observations before 1975 are excluded the 95% interval of the TFR in 2050 ranges from 1.2 to 2.2 children per woman. Otherwise, the interval ranges from 0.7 to 2.7 children per woman. These differences are largely due to the sharp decline in the fertility rates in the late sixties, in particular fertility rates for parity 3, 4 and higher. If the random walk model is applied to the TFR distinguished by parity, the differences are rather large for parity 3 and higher and rather small for parity 1 and 2.

In section 3.1 we noted that it is a matter of judgement whether or not to include old forecast errors in assessing forecast intervals of new forecasts. The same applies to the model-based estimates of forecast errors.

3.3 An Expert Knowledge Approach to Fertility Forecasts

The expert knowledge approach to specify assumptions about the forecast interval of the average number of children is based on the distribution by parity. Note that if there is no perfect correlation between the forecast errors for the separate parities, the 95% forecast interval for the total number of children does not equal the sum of the 95% intervals for the separate parities. However, if the correlation is strong, albeit not perfect, the difference with the 95% interval is only small. Since the assumptions refer to the order of magnitude of the probabilities rather than the exact level, the difference may be ignored in specifying the assumptions. Thus it is assumed that the forecast interval of the total number of children equals the sum of the assumed forecast intervals for the separate parity-specific fertility rates.

First children

In the Netherlands about 90% percent of women born in 1945 had at least one child (see table 2). For younger generations the percentage of women who had their first child before age 30 has decreased considerably. According to the medium variant of the 2000 Dutch population forecasts, the percentage of young women having their first child before age 30 will be only half that of the generation born in 1950. However, above age 30 the differences between successive cohorts will become much smaller. According to the medium variant, 80% of future generations of women will have at least one child. For the upper limit of the 95% forecast interval, 90% seems a reasonable assumption. There are several reasons for this assumption. First, the percentage cannot be much higher because of infecundity. It is estimated that about 5% of couples cannot have children because of infecundity (De Jong & Steenhof, 2000). Moreover, not all women will find a partner. Furthermore, dissolution of a partnership may lead to childlessness. Consequently, it is assumed to be unlikely that the percentage of women born in 2020 that will remain childless will be lower than 10. This percentage is high compared with historic values. Assuming that 90% of women would have at least one child would imply that either fertility above age 30 would increase very strongly or that there would be an increase of fertility at young ages, in contrast with the trend of the last decades. Thus, 10% childlessness seems an appropriate choice for the upper limit of the 95% interval of first children (table 2).

Table 2

Cohort fertility rates by birth order.

	Cohort 1945	Cohort 2020		
Number of children	Observation	Lower limit of 95% interval	Medium variant	Upper limit of 95% interval
1	0.88	0.60	0.80	0.90
2	0.75	0.50	0.65	0.80
3	0.25	0.05	0.20	0.40
4 or more	0.12	0.05	0.10	0.20
Total	2.00	1.20	1.75	2.30

A reasonable assumption for the lower limit of the 95% interval seems a value of 60%. It seems unlikely that childlessness would be over 40%. One of the reasons for this assumption is that successive fertility surveys, held by Statistics Netherlands, indicate that the vast majority of women wants to have children. For example, in the 1998 Fertility and Family Survey less than 10 percent of women with a low level of educational attainment and less than 20 percent of highly educated women who were born in the second half of the 1960s expect to remain childless (De Graaf & Steenhof, 1999). A 40% level of childlessness would imply that either a large percentage of women would not realise their expectations, for example, as women would face increasing problems in combining children and a paid job or their attitudes towards having children would change dramatically, as younger women would give higher priority to their career than to raising children or, for many women, postponement would result in ultimate childlessness. Of course, such a development is not impossible but it does not seem very likely either. It would imply that the percentage of women having children would decrease even more than it did during the last decades. The value of 60% is lower than the current period total fertility rate for first births. This level will only be reached by young cohorts if a sudden end to the increasing trend in the percentage of women having their first child above age 30 takes place.

Second children

About 75% of women born in 1945 had two children or more. According to the medium variant of the 2000 Dutch population forecasts, 65% of future generations of women will have at least two children. This implies that 15% will have only one child. For the upper limit of the 95% interval, 80% of women having at least two children seems a reasonable assumption. This is equal to the level for women born in the second half of the 1930s. This percentage implies that a very high proportion of women that has children will have at least two children. A higher percentage does not seem likely since there will always be women who will not have any children at all. Moreover, some women will have only one child, due to the dissolution of their relationship or because they have their first child at a relatively high age.

The lower limit of the 95% interval is assumed to be 50% (see table 2). In view of the strong preference for a family size of two children which appears from the Dutch FFS, a lower percentage does not seem likely. Moreover, recent data on fertility show that women no longer postpone having their first child. This enlarges the likelihood that a second child will be born. If the average age at the birth of a first child would continue to increase, the probability that a second child will be born will become smaller, due to infecundity or because women believe that they are too old for another child.

Third children

The percentage of women having at least three children declined sharply between cohorts 1935 and 1945 from about 45% to 25%. According to the medium variant of the Dutch forecast, 20% of future generations of women will have at least three children. This implies that about one third of women having a second child will also have a third child. In comparison to the other parities, uncertainty for parity three is relatively large, because having a third child in the Netherlands is more or less a free choice. It is much less bounded by norms and values than having a first or second child. This explains why the fertility rates for third children do not show a monotonous development. While the other parities have shown very stable patterns, the figures for parity three fluctuate: a decrease for women born in the thirties and the first half of the forties, followed by an increase for women born in the fifties and a decrease for those born in the sixties.

The lower limit of the 95% interval is assumed to equal 5%. Obviously the percentage cannot be much lower. This is considerably lower than the percentages observed in the past. The upper limit is assumed to be 40%. Even in Sweden, where the number of births of third children increased sharply in the early 1990s, the period fertility rate for parity three did not reach this level. Moreover, this peak was only temporary. In the mid 1990s fertility rates in Sweden dropped sharply.

Four or more children

The fertility rate for fourth and higher order births equals 0.10 for women born after 1945. Note that since this figure refers to an addition of birth orders 4, 5, etc. this cannot be interpreted as the fraction of women having at least four children. According to the medium variant the fertility rate for birth order 4+ equals 0.10. The lower limit of the 95% interval is assumed to equal 0.05. Obviously the value cannot be much lower, taking into account that this is the addition of birth orders 4 and higher. The upper limit of the 95% interval is assumed to equal 0.20. This does not seem to be very high compared with past developments, but it does not seem very likely that the high level of fertility of generations born before the Second World War will be reached again. In combination with a high level of labour force participation of women it is not very likely that a high percentage of women will have four or more children. One possible cause of a high level of fertility could be the increase in the foreign population, but that effect will probably be small. True, the percentage of foreign women in the childbearing ages is increasing, but the level of fertility of foreign-born women has been decreasing strongly.

Total fertility

Women born in 1945 had two children on average. According to the medium variant of the 2000 Dutch population forecasts the average number of children in future generations will equal 1.75. If the 95% forecast intervals for the separate birth orders are added up, the resulting interval for total fertility ranges from 1.2 to 2.3 children per woman. A higher value than 2.3 would imply that the level of fertility of women born before the Second World War would be approached (for example, women born in 1935 had on average 2.5 children). Due to the changed situation of women (higher level of educational attainment, higher labour force participation, changed sex roles within the family) this seems rather unlikely. A lower value than 1.2 does not seem very likely either. It would imply that either women will by no means have the number of children they want or that the attitudes towards having children would have to change drastically. True, developments during the last decades have shown that the level of fertility can change sharply. However, the reduction concerned mainly a decrease in the number of third and following children. A comparable decrease starting from the current level would imply a much more fundamental change. It would not imply that women would have less children, but rather that considerably more women will not have children at all (even a strong increase in the percentage of women having only one child would not be sufficient).

3.4 Specification of the Assumptions

The preceding sections describe three different methods to assess the standard deviation of the TFR. The results of the methods that use historic forecasts or time-series models depend to a large extent on which forecasts or base period they pertain to. It is up to the forecaster to select the most appropriate one. Table 3 summarises the main results obtained so far.

Table 3
Comparison of methods.

	95% interval of TFR in 2050	
	Lower bound	Upper bound
Extrapolation of errors of historic forecasts		
Based on all forecasts since 1950	0.89	2.61
Based on all forecasts excl. 1965 forecast	0.97	2.53
Based on all forecasts since 1975	1.23	2.27
Random walk model of TFR		
Based on observations since 1950	0.69	2.73
Based on observations since 1975	1.22	2.20
Expert judgement (parity-specific)	1.20	2.30

As mentioned in section 2, both an analysis of historic forecast errors and model-based estimates of forecast variances can be combined with expert judgement. One problem in using information on historic forecasts to assess the uncertainty of new long-run forecasts is that there are hardly any data on forecast errors for the long term. Alternatively, forecast errors for the long term can be projected on the basis of forecast errors for the short and medium term. Time-series of historic forecast errors can be modelled as a random walk model (without drift). On the basis of this model, applied to all forecasts made since 1950, the standard error of forecast errors 50 years ahead is estimated at 0.44. This implies that the 95% forecast interval for the year 2050 equals 0.9–2.6 children per woman. If only forecasts made since 1975 are incorporated the interval is much narrower: 1.2–2.3 children per woman.

Alternatively, the standard error of forecast errors can be projected on the basis of a time-series model describing the development of the TFR. If all observations since 1950 are included, the lower and upper bound of the 95% interval in 2050 are 0.7 and 2.7 children per woman respectively. This seems an unlikely wide margin. The lower limit would imply that only a minority of women would have children. The upper limit would imply that a vast majority of women would have three or more children. The probability of both developments is extremely small. Even for a 95% interval this margin seems to be much too wide. However, if only observations after 1975 are included the interval is again much smaller.

So, next to the question as to which model or method is preferred the question is which observation period (of forecast errors or of observed values of the TFR) should be selected. This question is answered by the expert judgement approach. Arguments are given as to why it is unlikely that the developments of the late sixties will occur again, or analogously, that forecast errors made in the 1965-based forecast will be made again. So, it is not surprising that the results of the argument-based approach are very similar to the results of the model based on forecast errors made since 1975 and to the results of the model based on the observations of the TFR after 1975.

To conclude, in the Dutch population forecasts the 95% forecast interval of the TFR in 2050 was assumed to equal 1.2–2.3 children per woman. Moreover, it was assumed that the forecast errors increase over time similar to the random walk model (errors increase with the square root of the forecast horizon).

The judgemental assumptions made in section 3.3 refer to cohort fertility rather than period fertility. However, the forecast interval for the TFR equals the interval for cohort total fertility. To obtain this result simulations of age-specific fertility rates arranged by calendar year are transposed into an arrangement by year of birth of the mother. This transposed set of simulations is used to calculate intervals for the completed cohort fertility rate (CFR). To compare the intervals for the TFR and the CFR the interval for a specific cohort is compared with the interval of the TFR 30 years later. The reason for this shift of 30 years is that women are on average 30 years old at childbearing. Figure 6 shows that the intervals of the TFR and the CFR are indeed almost equal. Keilman & Hetland (1999) draw the same conclusion on the basis of Norwegian data.

The width of the 95% interval of 1.2–2.3 children per woman in 2050 in the Dutch population forecast corresponds reasonably well with that of the 90% forecast interval for Germany, ranging from 1.0 to 1.9 in 2030, which is assumed by Lutz & Scherbov (1998). Keilman, Pham & Hetland (2002) assume a much wider interval for Norway. On the basis of an ARIMA model they calculate a 95% interval of the TFR in 2050 ranging from 0.7 to 3.5 children per woman. The result was obtained by restricting the TFR to the interval 0.5–4 children per woman. The authors interpret this latter interval as a "100 per cent prediction interval". The width of the 95% interval depends heavily on the assumption about the maximum value of the TFR. If the authors would have assumed the 100 per cent interval to range between, say, 1 and 3, their 95% interval would have been much more close to that for the Netherlands. Thus, even though a time-series model is used, the forecast interval is to a large extent determined by judgement.

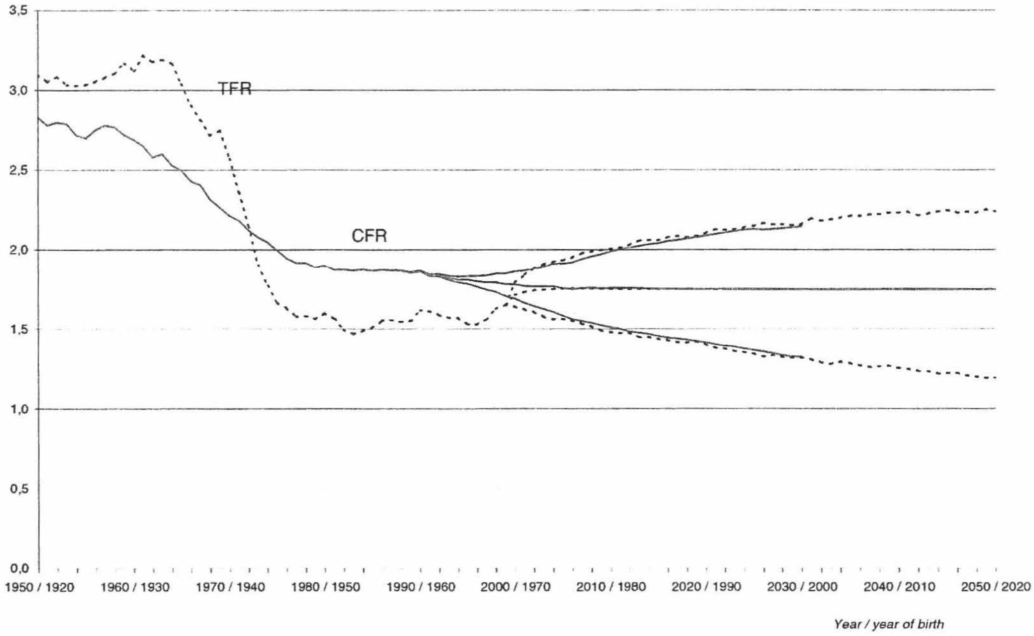


Figure 6. *Comparison of the 95% interval of period fertility (TFR) and cohort fertility (CFR).*

4 Discussion

Long-term developments in fertility are uncertain. To assess the degree of uncertainty of future developments in fertility and other demographic events several methods may be used: an analysis of errors of past forecasts, a statistical (time-series) model and expert knowledge or judgement. These methods do not exclude each other; rather they may complement each other. For example, even if the assessment of the degree of uncertainty is based on past errors or on a time-series model, judgement plays an important role. However, in publications the role of judgement is not always made explicit.

Expert knowledge or judgement usually plays a significant role in population forecasting. The choice of the model explaining or describing past developments cannot be made on purely objective, statistical criteria. Moreover, the application of a model requires assumptions about the way parameters and explanatory variables may change. Thus, forecasts of the future cannot be derived unambiguously from observations of the past. Judgement plays a decisive role in both the choice of the method and the way it is applied. "There can never be a population projection without personal judgement. Even models largely based on past time-series are subject to a serious judgemental issue of whether to assume structural continuity or any alternative structure" (Lutz, Goldstein & Prinz, 1996).

One main question in (stochastic) forecasting is which observation period should be the basis for the projections. An extrapolation of changes observed in the last 20 or 30 years may result in quite different projections than an extrapolation of changes in the last 50 years or more. Another important question is the choice of the extrapolation procedure: linear or non-linear. This question is difficult to answer on empirical grounds: different mathematical functions may describe observed developments equally well, but may lead to quite different projections in the long term. In conclusion, judgement plays an important role in fertility extrapolations.

Instead of an extrapolation of trends forecasts of fertility may be based on an explanatory approach. In making population forecasts usually a qualitative approach is followed. On the basis of an overview of the main determinants of fertility (e.g. changes in life style, labour force participation of women, educational level, etc.) and of assumptions about both the impact of these determinants on the development of fertility and future changes in the determinants, it is assessed in which direction fertility may change. Clearly, if no quantitative model is specified, the assumptions about the future change in fertility are largely based on judgement. However, even if a quantitative model would be available, judgement would still play an important role, since assumptions would have to be made about the future development in explanatory variables.

The methodology described has been applied to assess the uncertainty of future fertility in the Netherlands. In comparison to some other European countries fertility in the Netherlands has been rather stable in the past decades. Therefore, an assessment of the uncertainty of fertility in these countries may lead to broader forecast intervals.

References

Alders, M. & De Beer, J. (2002). An expert knowledge approach to stochastic mortality forecasting in the Netherlands. Paper for the Meeting on stochastic models for forecasting mortality, National Social Insurance Board, 29–30 January 2002, Stockholm, Sweden.

Alho, J.M. (1998). A stochastic forecast of the population of Finland. Reviews 1998/4, Statistics Finland, Helsinki.

Box, G.E.P. & Jenkins, J.M. (1970). *Time series analysis. Forecasting and control.* San Francisco: Holden-Day.

De Beer, J. (1997). The effect of uncertainty of migration on national population forecasts: the case of the Netherlands. *Journal of Official Statistics*, **13**, 227–243.

De Beer, J. & Alders, M. (1999). Probabilistic population and household forecasts for the Netherlands. Paper presented at the Joint ECE-Eurostat Work Session on Demographic Projections, Perugia, 3–7 May 1999.

De Graaf, A. & Steenhof, L. (1999). Relatie- en gezinsvorming van generaties 1945–1979: uitkomsten van het Onderzoek Gezinsvorming 1998. *Maandstatistiek van de bevolking*, **47**(12), 21–36.

De Jong, A. (2001). Bevolkingsprognose 2000–2050: veronderstellingen en methodiek. *Maandstatistiek van de bevolking*, **49**(1), 17–21.

De Jong, A.H. & Steenhof, L. (2000). Infecundity: a result of postponed childbearing? Paper presented at the BSPS-NVD-URUR Conference, 21 August–1 September 2000, Utrecht, the Netherlands. Published at www.cbs.nl

Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter.* Cambridge: Cambridge University Press.

Keilman, N.W. (1990). *Uncertainty in population forecasting: issues, backgrounds, analyses, recommendations.* Amsterdam: Swets & Zeitlinger.

Keilman, N. & Hetland, A. (1999). Simulated confidence intervals for future period and cohort fertility in Norway. Paper for the Joint ECE-Eurostat Work Session on Demographic Projections, Perugia, 3–7 May 1999.

Keilman, N., Pham, D.Q. & Hetland, A. (2002). Why population forecast should be probabilistic—illustrated by the case of Norway. *Demographic Research*, Vol. **6**, Art. 16, published 28 May 2002 on www.demographic – research.org

Lee, R. (1996). Probability approaches to population forecasting. Paper presented at IIASA Meeting on Population Forecasting, June 6–8, Laxenburg.

Lutz, W., Goldstein, J.R. & Prinz, C. (1996). Alternative approaches to population projection. In *The future population of the world. What can we assume today?*, Ed. W. Lutz, pp. 14–44. London: Earthscan.

Lutz, W., Sanderson, W., Scherbov, S. & Goujon, A. (1996). World population scenarios for the 21st century. In *The future population of the world. What can we assume today?*, Ed. W. Lutz, pp. 361–396. London: Earthscan.

Lutz, W. & Scherbov, S. (1998). Probabilistische Bevölkerungsprognosen für Deutschland. *Zeitschrift für Bevölkerungswissenschaft*, **23**, 83–109.

Tuljapurkar, S. (1996). Uncertainty in demographic projections: methods and meanings. Paper presented at the Sixth Annual Conference on Applied and Business Demography, September 19–21, Bowling Green.

Résumé

Au cours de ces dernières annes, l'Office statistique néerlandais (CBS) a publié un certain nombre de prévisions stochastiques sur la population. Le degré d'incertitude de la future population est évalué à partir d'hypothèses sur la distribution probable de la fécondité, de la mortalité et de la migration. Les hypothèses relatives à la fécondité se basent sur une analyse des prévisions historiques de l'indicateur conjoncturel de fécondité (ICF), sur des modèles d'observation de séries chronologiques de l'ICF et sur l'avis des experts. Cette dernière approche basée sur les arguments s'applique à l'ICF distribué par rang de naissance. Dans les prévisions néerlandaises les plus récentes, on considère que l'intervalle de prévision de 95% de l'indicateur conjoncturel de fécondité en 2050 varie entre 1,2 et 2,3 enfants par femme.

[Received December 2002, accepted November 2003]

International Statistical Review (2004), **72**, 1, 81–92, Printed in The Netherlands

Probabilistic Population Projections for India with Explicit Consideration of the Education-Fertility Link

Wolfgang Lutz[1] and Sergei Scherbov[2]

[1]*International Institute for Applied Systems Analysis, Laxenburg, Austria. E-mail: lutz@iiasa.ac.at*
[2]*Vienna Institute of Demography, Austrian Academy of Sciences, Vienna, Austria. E-mail: sergei.scherbov@oeaw.ac.at*

Summary

Among the different sources of uncertainty in population forecasting, uncertain changes in the structure of heterogeneous populations have received little attention so far, although they can have significant impacts. Here we focus on the effect of changes in the educational composition of the population on the overall fertility of the population in the presence of strong fertility differentials by education. With data from India we show that alternative paths of future female enrolment in education result in significantly different total fertility rates (TFR) for the country over the coming decades, even assuming identical fertility trends within each education group. These results from multi-state population projections by education are then translated into a fully probabilistic population projection for India in which the results of alternative education scenarios are assumed to expand the uncertainty range of the future TFR in the total population.

This first attempt to endogenize structural change with respect to education—which is the greatest measurable source of fertility heterogeneity in Asia—has resulted from a larger exercise of the Asian MetaCentre for Population and Sustainable Development Analysis to collect empirical information, scientific arguments as well as informed opinions about likely future population trends in Asia from a large number of population experts in the region. In this process, future changes in the educational composition of the population have been identified as a key driver of future fertility.

The actual probabilistic population projections for India show that with high certainty, the Indian population will continue to grow to about 1.3 billion over the next quarter of a century. After that the uncertainty will get much wider, ranging from a continued strong increase to the beginning of a population decline in India.

Key words: India; Population; Education; Projections; Uncertainty.

1 Background

Asia's population, which comprises more than half of the world's population, is currently going through significant structural changes. While the population of Asia is still expected to grow from currently 3.5 billion to around 4.8 billion in 2050, for the second half of the century we expect the beginning of a population decline (see Lutz *et al.*, 2003). Simultaneously, it will experience significant population aging. Asia is also in the midst of a major structural transition with respect to its human capital, i.e., the number of people by level of education. While the great majority of older cohorts of Asians have little to no education, the younger cohorts tend to be much better educated.

Since the better-educated women have lower fertility, this has direct demographic consequences. The improvements have been particularly impressive in China, where it is estimated that within two decades China will have more working age people with secondary and tertiary education than Europe and North America taken together (Lutz & Goujon, 2001).

But trends in Asia are very heterogeneous. China, which is currently the world's most populous country, has fertility well below the replacement level and is soon expected to be surpassed in size by India, which has higher fertility and a much younger population. For most of the 21st century, India will be the world's most populous country, with a projected 17 percent of the world's population in 2050. Hence, the future of the world's population will be significantly influenced by the future trends in India's population.

These future demographic trends in India are, however, highly uncertain. A lot will depend on the still unknown speed of fertility decline and the post-transition fertility level as well as the future spread of the AIDS epidemic or other health problems. The projections cited above only refer to the best guess from today's perspective, or more precisely the median of an uncertainty distribution. It is hard to say precisely how social, economic and even environmental changes will play together in determining the future trends in fertility, mortality and migration in India. On the other hand, we are not completely ignorant about the future size or structure of the population. Many of the people who will be alive in 2030 have already been born and we know their cohort sizes. Also, we can assume with high probability that a country in the midst of its fertility transition will continue with its fertility decline until a low level is reached. How should we communicate this to policy makers, who want to get the best synthesis of what we can say about the future trends to base their planning on these forecasts? To tell them that we simply do not know will not serve them well, especially if we think we know more than nothing. It is a bit more informative to say, here is one scenario and here is another, but we do not know how likely they are. This still does not help them much in their planning. Giving them just one projection, however, and telling them that this is the way the future will look, is probably highly welcomed by the planners but dangerous, if there are any costs associated with a projection error. A false feeling of certainty can be harmful and we should not create it, especially when we know better. Hence, the challenge for the population forecaster is to try to be specific of what we think is more certain and what is less certain, and to what degree uncertainties differ. The most comprehensive and efficient way to do so is to produce probabilistic forecasts.

In the context of developing countries with very limited data availability, methods of probabilistic population projections that are based on time series analysis are often not applicable because there are no time series. Also, methods that are based on the *ex post* error analysis of past population projections are of very limited applicability in a country that is in the midst of a demographic transition. Because of this one still expects significant structural change—we know that the future will be different from the demographic changes that occurred over the past decades. This essentially leaves us with the third approach to probabilistic forecasting, which is based on the evaluation of expert arguments. The strengths and weaknesses of expert-based approaches are extensively discussed elsewhere (Lutz *et al.*, 2004; Lutz *et al.*, 1999). In short, the key to avoid most personal biases in assessing the uncertainty ranges of future demographic trends is to focus on the evaluation of substantive arguments that experts put on the table and hence try to inform the subjective probability distribution by as much objective and scientific information as possible.

In the Asian context this task was taken on by the Asian MetaCentre for Population and Sustainable Development Analysis. The Asian MetaCentre is a group of population research institutes throughout Asia with headquarters at the National University of Singapore and a training branch at Chulalongkorn University in Bangkok (for more information, see `www.populationasia.org`). A series of two workshops in 2001 and 2002 was dedicated to an exercise to identify the main drivers of demographic change in Asia and to translate these insights into probabilistic projections. This included the more traditional analysis of past time series when available, the study of the errors of past projections—

where such could be found—and individual expert statements about the assumptions of the most likely trends in vital rates together with subjective quantitative assessments of the uncertainty ranges. This exercise also included new methods, such as a rather large number of in-depth interviews of experts in which they were extensively asked about their reasons for making certain assumptions rather than others, and during which an attempt was made to assess with which assumptions the experts felt more confident or less confident. These interviews were conducted by an expert in cognitive science, who is an expert on experts, but not on population. Based on these findings the exercise then went one step further to try to explicitly model the one structural determinant that has been singled out as the most important one, namely female education, and make the future fertility uncertainty dependent on future education in the context of probabilistic projections. To our knowledge such a structural approach to probabilistic projections has not been applied before.

This approach of expert argument-based projections distinguishes between resource experts who provide the arguments and meta-experts who guide the process and operationalize the assumptions. In this exercise carried out by the Asian MetaCentre, the authors of this paper served as the meta-experts with more than 50 Asian national population experts serving as resource experts. This is explained in more detail in Lutz & Scherbov (2003). In this paper we only focus on discussing the assumptions and projections for India.

2 Fertility Assumptions

We chose to single out the one structural argument about the determinants of the future course in fertility that would feature most prominently in most of the expert interviews. The choice was not difficult, because in addition to the rather diffuse reference to all kinds of government policies, female education clearly stood out as the single most important factor mentioned. Almost all of the national experts from the different countries in Asia involved in the exercise consistently mentioned that they thought that the improving level of female education has been and will be the main driver of fertility decline.

The line of argumentation in these interviews seems clear: The combination of great educational fertility differentials in Asia (more highly-educated women have significantly lower fertility) with the fact that younger women are and will be more educated than older women, greatly contributes to fertility decline. These educational fertility differentials are pervasive all over Asia and in countries with very different overall levels of fertility.

Table 1 gives the recent educational fertility differentials for women in India and China. It shows the total fertility rate (TFR) for four categories of women: those without any formal education and those with some primary, secondary and tertiary education. The precise definition and discussion of these categories and data sources are given in Lutz & Goujon (2001). The table clearly and impressively shows how in these two major countries, higher education is associated with lower fertility. The total TFR gives the fertility of the total population with weights of the different educational groups corresponding to the current educational composition in India and China.

Table 1

Education-specific differentials and total TFR for women in India and China.

	India 2000	China 2000
No education	3.78	2.43
Some primary education	2.89	2.14
Some secondary education	2.36	1.63
Some tertiary education	1.96	1.08
Total Fertility Rate	3.30	1.80

Figure 1 gives the age-, sex- and education-specific population pyramid for India as estimated for the year 2000. The pyramid clearly shows that younger cohorts of women in India are better educated than older ones, although the gender gap is still significant.

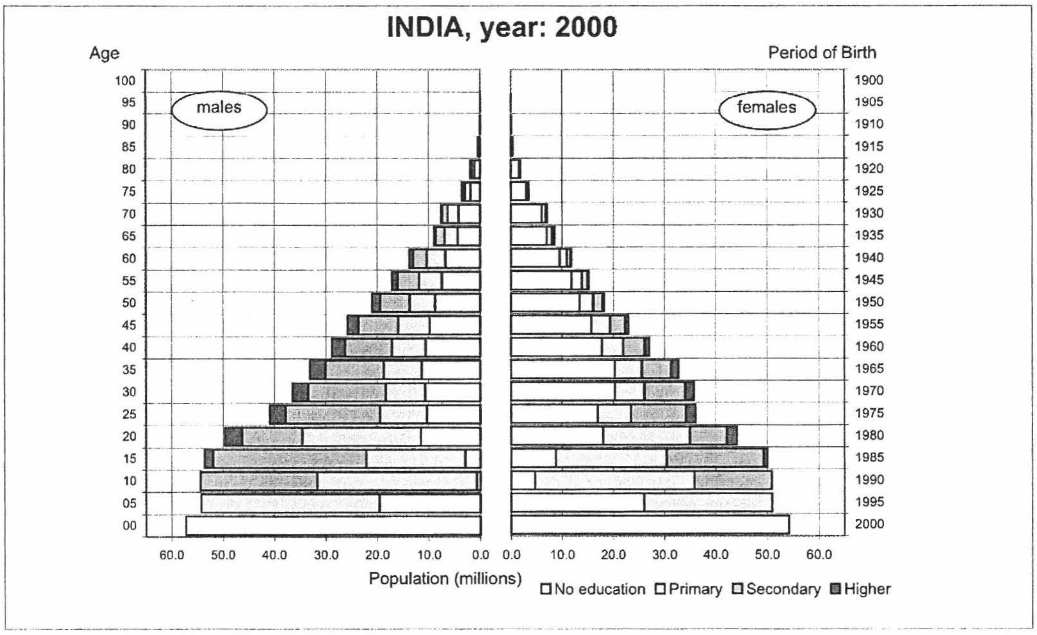


Figure 1. *Population pyramid for India, 2000.*

When we think about the future of the national-level fertility in India, we must differentiate between two different effects: (a) the change in the educational composition of the population, and (b) the fertility trends within each educational group. We will first discuss these factors separately and then (c) study their joint effect.

(a) Alternative trends in education

As can be inferred from Figure 1 future improvement in the educational composition of the population is a near certainty. It is already pre-programmed into the age structure. The younger, better-educated cohorts will inevitably become older and replace the older, less-educated cohorts. Disregarding the unlikely case of massive adult education programs only the future education of the young cohorts is uncertain at this point, depending on future school enrolment rates at different levels. Policies can make a difference here. We capture this difference through two alternative and rather extreme scenarios on girls' and boys' education in India, one (pessimistic) scenario in which all enrolment rates stay constant in the future [scenario A: constant enrolment] and another in which India manages to implement the ambitious education goals as defined in the 1994 Cairo World Population Conference (ICPD) [scenario B: ICPD]. This highly optimistic scenario assumes the elimination of the gender gap in primary and secondary education by 2005–10, 90 percent net primary enrolment by 2010–15 and secondary enrolment of 75 percent by 2025–30, as well as an increase in transition to tertiary education by 5 percentage points until 2025–30. Trends between

2000 and the target year are based on linear interpolation. The results from these two scenarios are depicted in Figure 2 for the year 2030.

Figure 2 clearly shows that the two scenarios are virtually identical in 2030 for men and women above age 55. For younger cohorts the constant enrolment rates scenario shows an essentially frozen educational composition with a significant gender gap in education being maintained. Under the ICPD scenario the educational composition improves significantly for the younger cohorts and the gender gap essentially disappears below age 25.

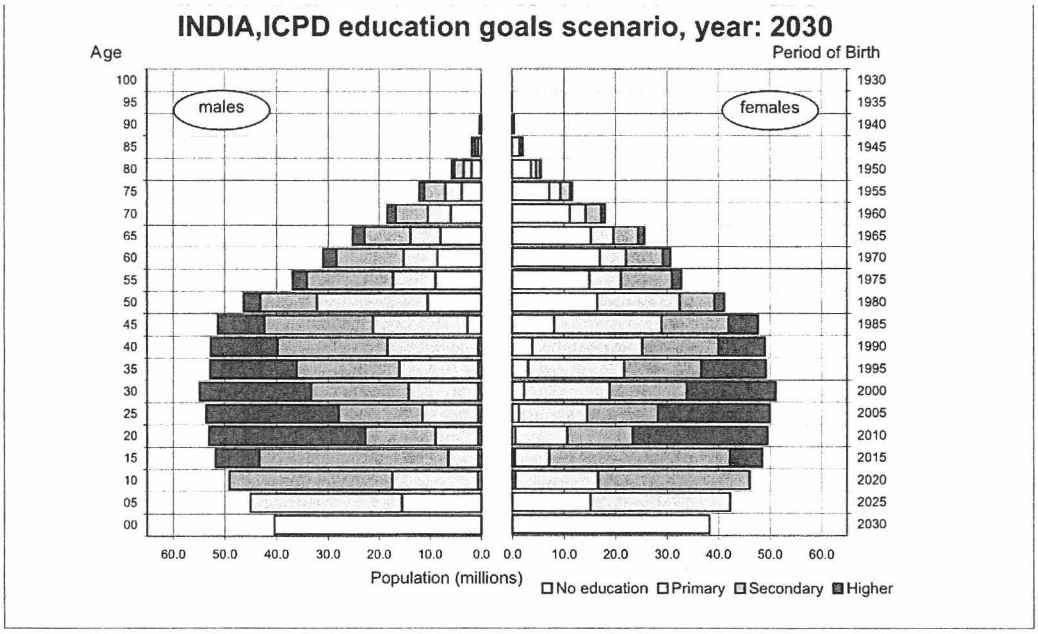
(b) Trends in education-specific fertility

During the course of the Asian MetaCentre exercise, there were quite some discussions about what should be assumed in terms of future education-specific fertility trends. We considered three different options for dealing with this issues: (1) assuming proportional fertility changes in all educational categories, (2) assuming convergence of all educational fertility trends to one target level, or (3) assuming that education-specific fertilities will move to the observed levels of another country that is already further advanced in the process of fertility decline. Option (1) is not meaningful as a general rule because in some countries the fertility of university graduates is already so low that no further declines will be expected, despite declines in average fertility. Also, for countries for which time series exist, the declines do not seem to go in parallel. Option (2) is not meaningful in this context, because if we have complete convergence, by definition the changes in the educational composition do not affect aggregate fertility. And substantively all of the Asian countries studied maintained significant differentials, even at very low aggregate fertility levels. For these reasons we chose Option (3), which is consistent with the frequent demographic practice to think in terms of analogies, as is more generally done in the context of the demographic transition. Specifically, for the case of India presented here, we assumed that in 2030, India will have education-specific fertility rates comparable to those of China today.

What is the substantive justification for this particular assumed analogy between China today and India tomorrow? First, the aggregate TFR in India in 2030 is assumed by most projections to go to a level similar to that in China in 2000. Also, India and China are both huge and greatly heterogeneous countries. In this sense the national-level fertility differentials by education do not just reflect some volatile patterns as they may appear in small populations. Finally, the data given in Table 1 show declines in each educational group of between 26 and 45 percent or on average one-third. A sensitivity analysis that assumes that each of the fertility rates in India in 2000 declines linearly by one-third until 2030 yields virtually the same results as the assumed move to the Chinese pattern of 2000.

(c) Combined scenarios

Table 2 defines 12 scenarios that result from the cross-classification of different future trends in the educational structure of the population, and different trends in education-specific fertility rates. The first four scenarios that keep the educational structure of the population frozen are of a purely hypothetical nature because, as has been mentioned above, it is already embedded in the age structure that the younger, more educated age groups over time will move the age scale and improve the average education of the female population of reproductive age. But it is still important to talk about this hypothetical case of the educational composition by age remaining constant in its current form, because it serves as a point of reference in the minds of experts thinking about this issue. For this reason Figure 3 shows the aggregate level TFR resulting from scenarios 1 and 2 in 2050 as black dots to the right of the figure. If the educational composition remains frozen and the education-specific fertility remains constant (scenario 1), then clearly the aggregate TFR



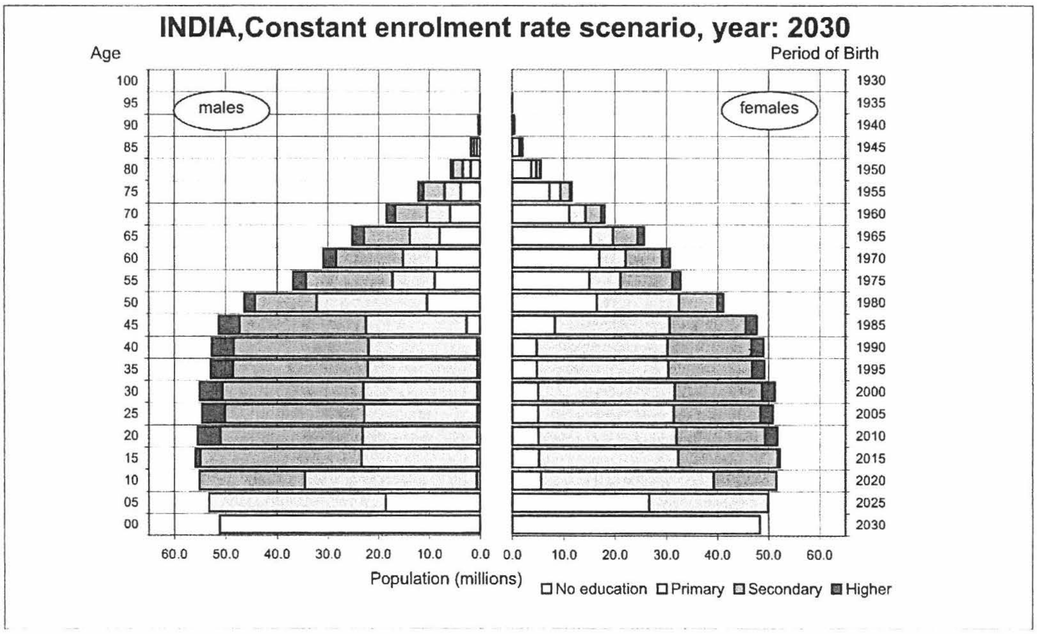


Figure 2. *Education-specific age pyramids for India under the two alternative education scenarios in 2030 as described in the text.*

remains constant over time. Scenario 2 gives the case in which the educational composition remains frozen, but education-specific fertility rates decline as discussed below. In this case the aggregate TFR declines by about one child, which is the fertility trend effect that is completely free of the effect of the changing educational structure.

Table 2

Definition of 12 scenarios combining different possible trends in the educational structure with different assumptions about education-specific fertility trends.

	Educational Structure of Population		
Fertility trends for educational categories	Structure constant (purely hypothetical)	Enrolment rates constant	ICPD goals for enrolment
All fertility rates constant	1	5 (c.enr-c.fert)	9 (ICPD-c.fert)
Rates reach Chinese level by 2030	2	6 (c.enr-medium)	10 (ICPD-medium)
China +0.5 by 2030	3	7 (c.enr-high)	11 (ICPD-high)
China −0.5 by 2030	4	8 (c.enr-low)	12 (ICPD-low)

Figure 3 gives four lines for the cross-classification of the two assumptions for education-specific fertility (constant versus linear change to the Chinese pattern) with the two education assumptions (constant enrolment versus the ICPD scenario). When comparing scenarios 5 and 9 and 6 and 10, respectively, we see that the changing educational structure makes a significant difference, even with identical education-specific fertility trends. By 2050 this difference accounts for more than half a child, i.e., more than one-third of the level of fertility as given by scenario 10. If we also consider the hypothetical case of a frozen education structure (scenario 2), the difference due to differential educational structures becomes almost one child. This clearly illustrates that the experts have made a valid and quantitatively important point when they suggested that the changing educational structure would be a major force towards lower fertility.

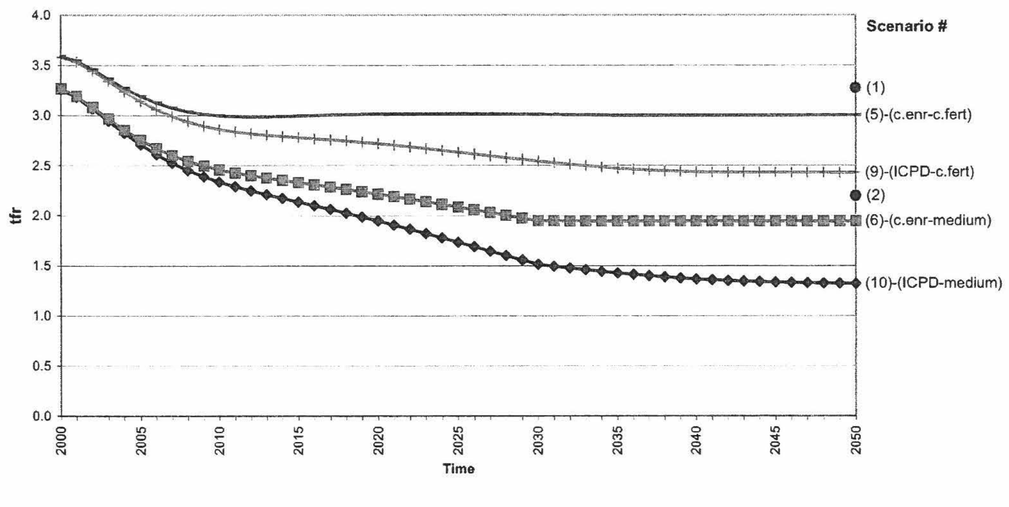


Figure 3. *Selected scenarios for combining education assumptions with education-specific fertility trends.*

3 Probabilistic Population Projections

How should these insights be translated into probabilistic fertility assumptions? The first thing that is unclear in this context is whether experts, when they stressed the effect of the changing educational composition, had one of the two extremes—the constant enrolment scenario or the ICPD scenario—or something in between in mind. A qualitative discussion with some of the experts at the second seminar indicated that this uncertainty about future school enrolment should be assumed to be part of the total fertility uncertainty. As to the uncertainty of education-specific fertility trends, it was assumed that the uncertainty range considered here would be half a child up and down, as compared to the mean trend (scenarios 6 and 10), which is the linear move from the Indian 2000 rates to the Chinese 2000 rates by 2030. This is consistent with what was assumed in Lutz *et al.* (2001). There the 80 percent range was assumed to be plus/minus one child if the TFR was above 3.0, and plus/minus 0.5 if it was below 2.0, with linear interpolation in between. Since this was supposed to include the education uncertainty in addition to the education-specific fertility uncertainty, these assumptions are roughly consistent.

Figure 4 gives the four scenarios that combine the plus/minus 0.5 children in education-specific fertility assumptions with the two extreme education scenarios. Following the logic outlined above the appearing range between scenarios 7 and 12 is then taken to represent the 90 percent uncertainty interval of a normal distribution representing India's fertility uncertainty in any given year between 2000 and 2050.

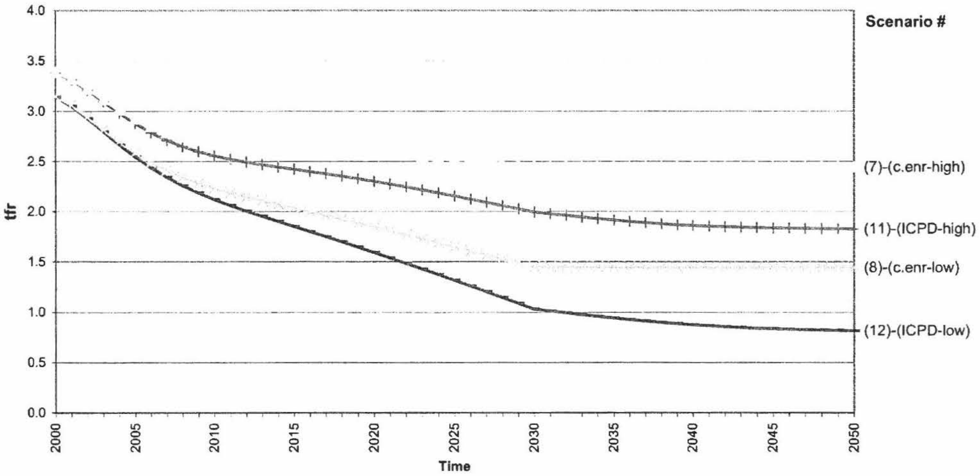


Figure 4. *Selected scenarios combining education-specific fertility trends, which are 0.5 children higher and lower than the mean with different future education trends.*

A final step was to translate these fertility uncertainty ranges, which include the education uncertainty in addition to the uncertainty of education-specific fertility trends, into full probabilistic population projections. While the fertility distribution is based on the range between scenarios 7 and 12 as discussed above, the stochastic mortality assumptions are taken from the projections given in Lutz *et al.* (2001). It is assumed that life expectancy increases by two years per decade with an 80 percent uncertainty range from zero to four years of improvement per decade. For simplicity a closed population is assumed here. The stochastic model chosen assumes annual fluctuations in birth and death rates following the model that is described in Lutz *et al.* (2004).

Figure 5 presents the fractiles of the distribution of India's future population size resulting from 1,000 simulations. The white line gives the median of the uncertainty distribution with half of the simulated population futures above that line and half below. The black area shows the inner 20 percent of the uncertainty distribution, the dark gray the inner 60 percent, and the light gray the 95 percent interval. Five percent of the simulated paths lie either above or below the outer lines. As with all such projections the uncertainty range opens up over time. The graph also shows that for the coming two decades the uncertainty range is very narrow, which implies that with very high probability, India's population will increase from presently 1 billion to more than 1.2 billion by 2020. The median shows a stabilization of India's population size after 2035 at a level of roughly 1.4 billion. But the uncertainty range becomes very broad beyond 2030. This is due to the fact that after around 30 years the uncertainty about the number of potential mothers is added to that of the number of births per woman. While in more than 20 percent of the simulations, India's population starts to decline after 2030, in another 20 percent of the cases, it shows very significant continued growth, with the upper end of the 95 percent interval reaching 1.7 billion by 2050.

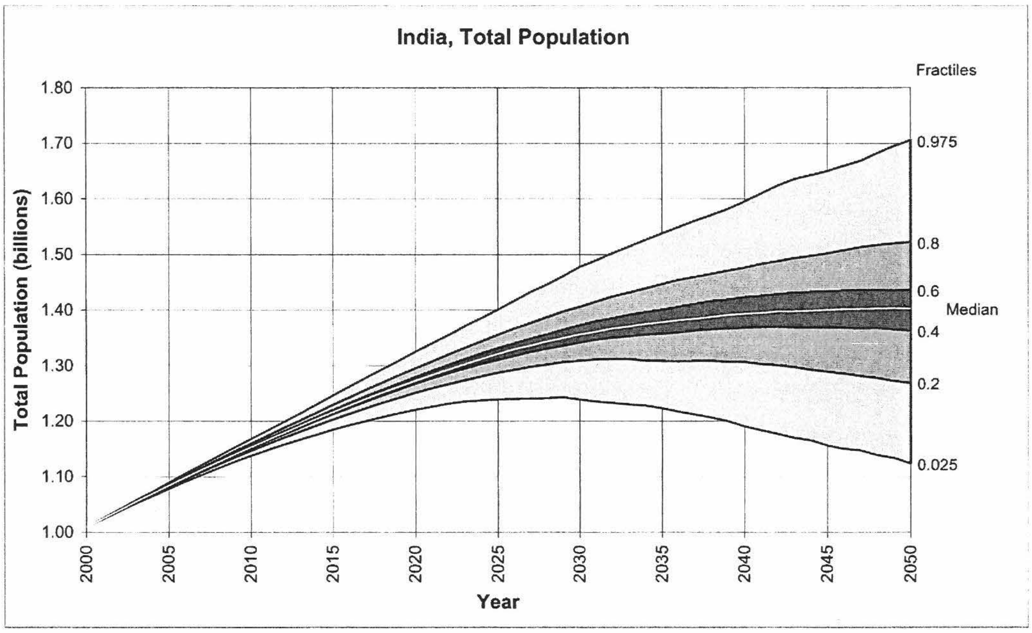


Figure 5. *Resulting distribution of total population size in India.*

As discussed above, the future fertility uncertainty is greatly influenced by the uncertainty in the future of education trends. This becomes apparent when comparing the total population numbers in 2050 that result from scenarios 6 and 10 as discussed above, i.e., the central fertility decline assumption combined with the two different educational scenarios. The difference is very significant and in the order of 0.2 billion. In other words, these calculations imply that with otherwise identical assumptions, an India that will follow the ICPD education goals will have 200 million people less than an India with constant school enrolment ratios.

Table 3 gives the numerical results of these new probabilistic population projections for India. The values given refer to the median, with the 80 percent intervals given in parentheses. This shows that despite an expected further population growth of around 40 percent over the coming five decades, India's population will also become significantly older. The proportion above age 65 will increase from currently only 5 percent to around 14 percent with the 80 percent uncertainty range going from 0.12 to 0.16. The proportion of children below age 15 will likely decrease to about half its level, from currently 0.35 to only 0.17. For the children, the 80 percent uncertainty interval is much larger, ranging from 0.11 to 0.21.

Table 3

Medians and 80 percent ranges (in parentheses) for selected projection output parameters in India.

	2000	2025	2050
India, Total Population	1.009 (1.009–1.009)	1.321 (1.268–1.374)	1.403 (1.208–1.603)
India, Proportion below age 15	0.335 (0.335–0.335)	0.238 (0.212–0.260)	0.167 (0.113–0.213)
India, Proportion 15–65	0.615 (0.615–0.615)	0.691 (0.671–0.714)	0.697 (0.660–0.733)
India, Proportion above age 65	0.050 (0.050–0.050)	0.071 (0.068–0.075)	0.138 (0.117–0.161)
India, Old-Age Dependency Ratio (65+ / 15–65)	0.081 (0.081–0.081)	0.103 (0.099–0.107)	0.198 (0.172–0.228)
India, Support Ratio	12.395 (12.395–12.395)	9.690 (9.332–10.053)	5.060 (4.388–5.811)

The fact that the uncertainty ranges differ greatly by age is most clearly shown by the probabilistic age pyramid in Figure 6. Unlike such pyramids for industrialized countries, where there is visible uncertainty at a very old age due to the uncertainty about the path of future old age mortality, this is not yet visible in India because the population is still much younger and life expectancy is still much lower. Hence, the giant share of the uncertainty of India's future population is due to fertility uncertainty and to a smaller extent to uncertainty about future child and young adult mortality.

These results for India show that the world's most populous country of the 21st century is likely to simultaneously experience significant further population growth and population aging. The proportion above age 60 will almost triple over the coming half century. This will pose formidable challenges to old age social and economic support systems. At the same time the total population will continue to grow significantly by some 40 percent, bringing many of the infrastructural problems associated with rapid population growth. The good news is that over the coming decades the proportion of the total population that is of working age will also increase substantially, thus providing a "demographic window" or "demographic bonus" for more rapid economic development. If this opportunity is used for heavy investments in education and human capital formation (along the lines of the IPCD scenario) this would significantly strengthen India's economic and political standing in the world and improve the quality of life of its citizens.

4 Discussion

The paramount importance of the specific path of the future fertility trends in India on the total population size is an *ex post* justification for spending a considerable amount of time and effort to capture some of the structural determinants of future fertility trends in India. Whether there will be 200 million more Indians in 2050 is not a trivial issue. As we have shown, this will depend significantly on the future educational efforts in India.

The analysis presented here has only been a first step in the direction of trying to incorporate this very important structural dimension of population uncertainty into population projections. Much more attention should be given to this and other important structural drivers of fertility decline in

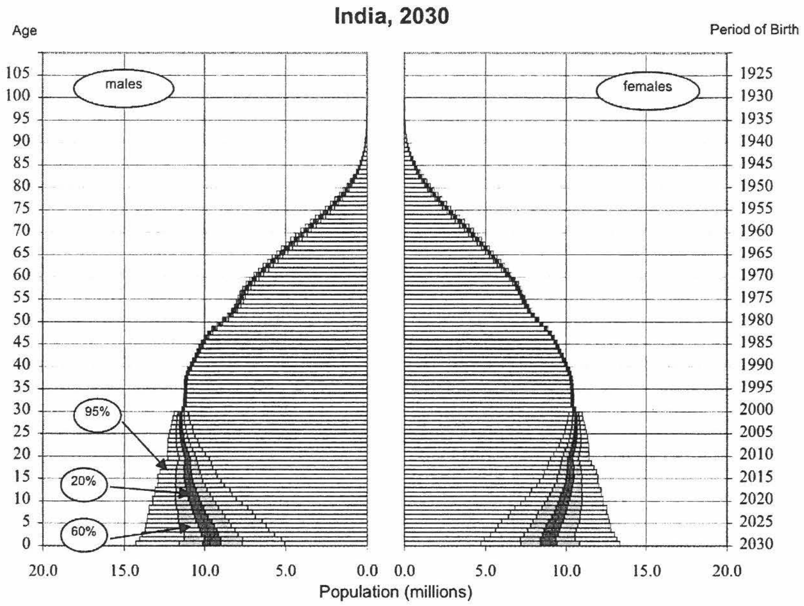


Figure 6. *Probabilistic population pyramid for India in 2030.*

developing countries in the future. Here we focused on female education as probably the most important observable source of fertility heterogeneity. Urbanization would be the next logical candidate for such analysis. But it should probably be cross-classified with education (Cao & Lutz, 2004).

The great sensitivity of the future level of aggregate fertility to the change in the educational composition of the population raises disturbing questions about the way fertility assumptions are usually derived for developing countries that are in the midst of structural changes of the sort described here for India. When future paths of aggregate-level fertility are being assumed in regular population projections, then these should reflect all possible forces that influence aggregate fertility. It is not clear, however, that without an explicit formal model of the sort presented in this paper, the extent of the effect of these compositional changes can be adequately assessed in a more or less intuitive manner. In principle, this reflects the general problem of describing the dynamics of heterogeneous populations, which can show unpredicted behavior if the heterogeneity is not accounted for. The best strategy thus should be to make sure that at least the observable and most significant measurable sources of heterogeneity are explicitly incorporated in the analysis of population dynamics. In most developing countries education is such a source of heterogeneity that should be made explicit.

In conclusion we can say that the explicit consideration of some of the most important drivers of changes in demographic rates can make a major difference in the way we see the population evolve in the future. This is particularly true in the context of probabilistic population projections, where the uncertainty about the evolution of the structure of heterogeneous populations is added to the uncertainty about demographic trends within each sub-population. This aspect is quantitatively more important in heterogeneous populations, such as India, than in the more homogeneous ones.

How to deal with this issue in the context of specific populations to be forecasted, cannot be determined purely by statistical models. It requires deep substantive analysis and inevitably a degree of expert judgment. But in any case we should be careful to base our assumptions on explicit arguments that are open to the usual instruments of scientific review and evaluation.

Acknowledgements

The authors wish to thank Kannan Navanetham for his most valuable input to this paper. We also want to thank all of the population experts from Southeast Asia that participated in this expert inquiry, which was organized by the Asian MetaCentre for Population and Sustainable Development Analysis.

References

Cao, G.-y. & Lutz, W. (2004). China's future urban and rural population by level of education. In *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*, Eds. W. Lutz and W.C. Sanderson. London: Earthscan.

Lutz, W. & Goujon, A. (2001). The world's changing human capital stock: Multi-state population projections by educational attainment. *Population and Development Review*, **27**(2), 323–339.

Lutz, W., Sanderson, W.C. & Scherbov, S. (1999). Expert-based probabilistic population projections. In *Frontiers of Population Forecasting*. A Supplement to Vol. **24**, 1998, Population and Development Review, Eds. W. Lutz, J.W. Vaupel and D.A. Ahlburg, pp. 139–155. New York: The Population Council.

Lutz, W., Sanderson, W.C. & Scherbov, S. (2001). The end of world population growth. *Nature*, **412**, 543–545.

Lutz, W., Sanderson, W.C. & Scherbov, S. (2003). The end of population growth in Asia. *Journal of Population Research*, **20**(1), 125–141.

Lutz, W., Sanderson, W.C. & Scherbov, S. (2004). The end of world population growth. In *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*, Eds. W. Lutz and W.C. Sanderson. London: Earthscan.

Lutz, W. & Scherbov, S. (2003). Toward Structural and Argument-Based Probabilistic Population Projections in Asia: Endogenizing the Education-Fertility Links. Interim Report IR-03-014. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Résumé

Parmi les différentes sources d'incertitude dans les projections de population, les changements dans la structure de populations hétérogènes ont reçu peu d'attention jusqu'ici, en dépit de l'impact significatif qu'ils peuvent avoir. Nous nous concentrons ici sur les effets de changements dans la composition éducative de la population sur la fertilité globale de la population, en présence de forts différentiels de fertilité par niveau d'éducation. A partir de données de l'Inde, nous montrons que des parcours éducatifs différents des femmes conduisent à des taux globaux de fertilité significativement différents pour le pays dans les décennies à venir, même avec l'hypothèse que les tendances de fertilité restent identiques à l'intérieur de chaque groupe éducatif. Ces résultats de projections de population par niveau éducatif sont ensuite traduits en une projection de population entièrement probabiliste pour l'Inde, dans laquelle les résultats de scénarios éducatifs alternatifs sont supposés accroître le degré d'incertitude du futur taux global de fertilité dans la population totale. Cette première tentative d'endogénéiser le changement structurel en fonction de l'éducation—qui est la source la plus mesurable d'hétérogénéité de la fertilité en Asie—est issue d'un exercice plus large conduit par le Métacentre Asiatique pour l'analyse de la population et du développement durable dans le but de collecter de l'information empirique, des arguments scientifiques ainsi que l'opinion d'un grand nombre d'experts de la population de la région sur les futures tendances probables de la population en Asie. Dans ce processus, les changements futurs dans la structure de la population par niveau d'éducation ont été identifiés comme un facteur clé de la fertilité future. Les projections de population probabilistes pour l'Inde montrent avec un degré élevé de certitude que la population va continuer à croître jusqu'à environ 1.3 milliard dans le prochain quart de siècle. Au delà l'incertitude devient beaucoup plus grande, entre le prolongement d'un fort accroissement et le début d'une baisse de la population indienne.

[Received February, 2003, accepted November, 2003]

International Statistical Review (2004), **72**, 1, 93–106, Printed in The Netherlands

Simpler Probabilistic Population Forecasts: Making Scenarios Work

Joshua R. Goldstein

Office of Population Research, Princeton University, Wallace Hall, Princeton, NJ, USA. E-mail: josh@princeton.edu

Summary

The traditional high-low-medium scenario approach to quantifying uncertainty in population forecasts has been criticized as lacking probabilistic meaning and consistency. This paper shows, under certain assumptions, how appropriately calibrated scenarios can be used to approximate the uncertainty intervals on future population size and age structure obtained with fully stochastic forecasts. As many forecasting organizations already produce scenarios and because dealing with them is familiar territory, the methods presented here offer an attractive intermediate position between probabilistically inconsistent scenario analysis and fully stochastic forecasts.

Key words: Age structure; Population forecasting; Population size; Scenarios; Stochastic; Uncertainty.

1 Introduction

Traditionally, population forecasts have included high, medium, and low variants, with the medium being the most likely outcome. The high and low scenarios are said to be "plausible" but lack any probabilistic interpretation. High and low scenarios tend to produce extremes in either population size or in population age structure, but not in both at once. Scenarios typically have smooth time trajectories, assigning zero probability to fluctuations in fertility and mortality that occur in real populations. A recent evaluation concludes that "no consistent probabilistic interpretation *can* be given to the high-low scenarios in population forecasts". (National Research Council, 2000).

Stochastic population forecasts have been developed that try to address these difficulties. A number of approaches have been tried. Approaches based on stochastic renewal (Lee & Tuljapurkar, 1994; Keilman, Pham & Hetland, 2002; Alho & Spencer, 1991) have used time series analysis to forecast future mortality and fertility rates which are allowed to fluctuate randomly over time. (Typically, migration remains deterministic.) Monte Carlo methods are then used to produce a large number of future population trajectories. The quantiles of these simulated trajectories give the prediction intervals for measurements of population size, and age-structure. An analytic theory of stochastic renewal has also been developed (Cohen, 1986; Lee & Tuljapurkar, 1994; Alho & Spencer, 1991), but modern computing power makes the Monte Carlo approach simpler in most cases. The world forecasts by Lutz, Sanderson & Scherbov (2001) have used a hybrid of methods including expert opinion, scenarios, and auto-correlated time series to produce probabilistic projections.

Fully stochastic forecasts quantify the uncertainty in population size, age groups, and age structure in a probabilistically consistent way. This internal consistency allows the incorporation of mortality and fertility uncertainty, simultaneous prediction intervals for several age groups and functions of age groups, and population outcomes to be integrated over time. This last feature is of particular advantage when assessing the viability of national pension systems for which the age structure of a population over many decades is of interest (Lee & Tuljapurkar, 1998).

Despite these advantages, population projectors have been slow to switch over to fully stochastic forecasts. Their reluctance is not surprising. Stochastic forecasts are data-intensive, requiring detailed historical information to estimate the variability of vital rates. They also assume that variability in the future will be similar to that of the past. Finally, from a technical point of view, stochastic forecasts are difficult to produce—whereas scenario projections require relatively little training and can be produced with existing software, stochastic forecasts have required research scientists knowledgeable about time series estimation and customized programming.

Moreover, scenarios are easy to explain to users and easy for users to interpret. The middle scenario is interpreted as the most likely outcome (Keyfitz, 1972). "High" and "low" may lack assigned probabilities but do have simple conditional interpretations. *If* fertility and mortality follow the hypothesized paths *then* the impact for future population will be as shown. This if-then approach translates vital rate trajectories into population outcomes, a rough sensitivity analysis of how population size and age-structure depend on levels of fertility, mortality, and migration

This paper addresses the reluctance of forecasters to make the leap from using scenarios to fully stochastic forecasts. We do this by describing the conditions under which the results of scenario-based forecasts closely approximate those that could have been obtained using a fully stochastic approach.

The procedure presented here is a compromise. It does not produce exactly correct results, but it yields surprisingly accurate figures with little additional effort on the part of the forecasters. Those who use this approach have the benefits of both using a methodology with which they are familiar and obtaining the desirable features of probabilistically consistent uncertainty intervals. To do this, we first give scenarios a specific probabilistic interpretation; then we combine the results for different age-segments in a probabilistically sensitive manner.

Past comparisons between scenarios and stochastic forecasts have compared forecasts that differ in many dimensions including both the expected level of fertility and mortality and the range of fertility and mortality (Lee & Tuljapurkar, 1994; Keilman *et al.*, 2002). The approach here is to calibrate the two approaches so that a comparison can reveal the differences and similarities intrinsic to the two methods. Lutz & Scherbov (1998) used a similar approach but a different method of calibration.

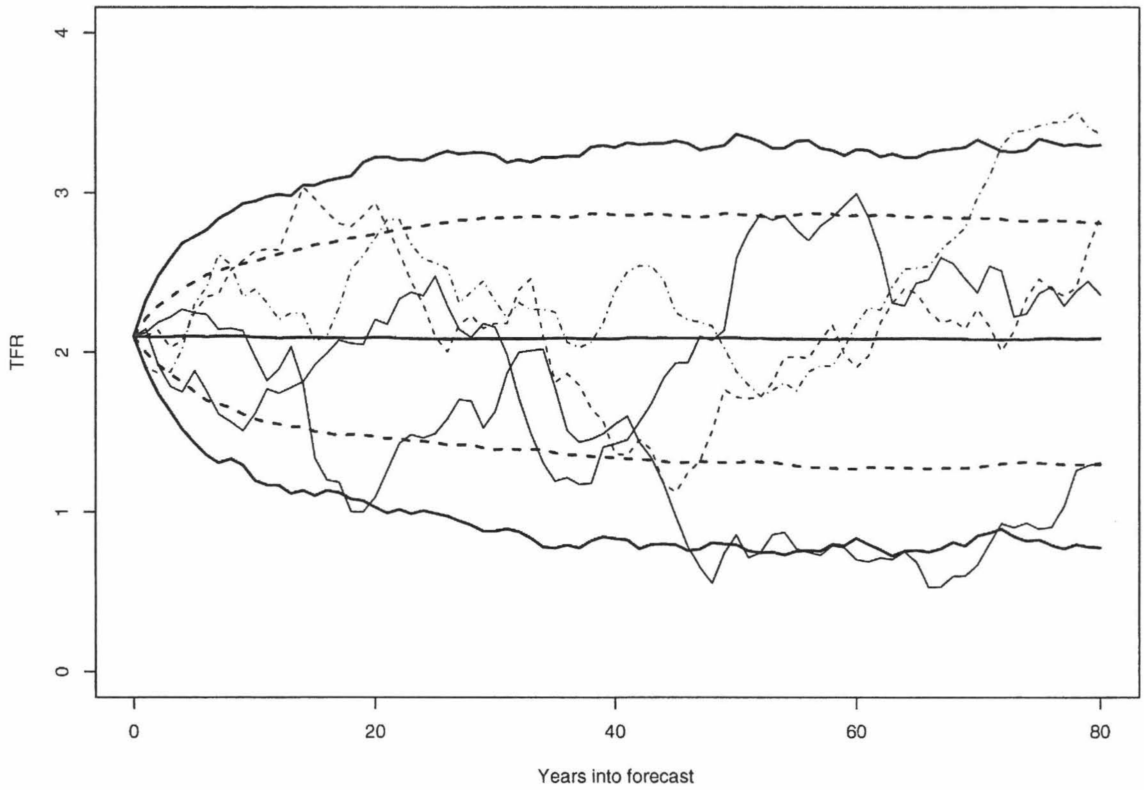
The paper begins with a discussion of how to give of scenarios, particularly fertility scenarios, a specific probabilistic interpretation. Next we consider how to combine the results of forecasts for different age-segments in a probabilistically sensitive manner. Throughout, we compare the results of this modified scenario approach with published fully stochastic forecasts by Lee & Tuljapurkar (1994).

2 The Modified Scenario Approach

Our first task is to assign a probabilistic interpretation to the vital rates scenarios. Figures 1 and 2 show some sample paths from stochastic forecasts of fertility and life expectancy. The distance between the bold solid lines gives the 95 percent uncertainty intervals for fertility and life expectancy in any given year. The vertical span between the heavy dashed lines in the fertility figure gives the 95 percent interval of the long-term average of fertility, which will be described below. The intervals for for the average are narrower than the intervals for a given year because because the ups and downs in fertility cancels out to some extent.

A key difference between scenario-based forecasts and stochastic forecasts is that the scenarios incorporate nearly perfect temporal autocorrelation. If, for example, fertility is high in the first 5 years of the scenario forecast, it is assumed to remain high for the entire forecast. Stochastic forecasts, on the other hand, use time series models to incorporate an empirically estimated amount

Figure 1. *Sample paths and forecast intervals of stochastic forecast of total fertility rate.*



Stochastic fertility model is a mean-constrained ARMA(1,1) model with coefficients given in Lee & Tuljapurkar (1994). The figure shows 4 sample paths generated with Monte Carlo simulation; these can be identified by their large ups and downs. The heavy solid U-shaped outer bounds give the 95 percent prediction interval for the TFR in any year, with the heavy middle line giving the median TFR. The heavy dashed U-shaped inner bounds give the 95 percent prediction interval for the cumulative average of the TFR up to that date. Both prediction intervals were estimated from the quantiles of 1000 simulated runs. Note that the bounds on the cumulative average cover roughly the same interval from years 20 through 80.

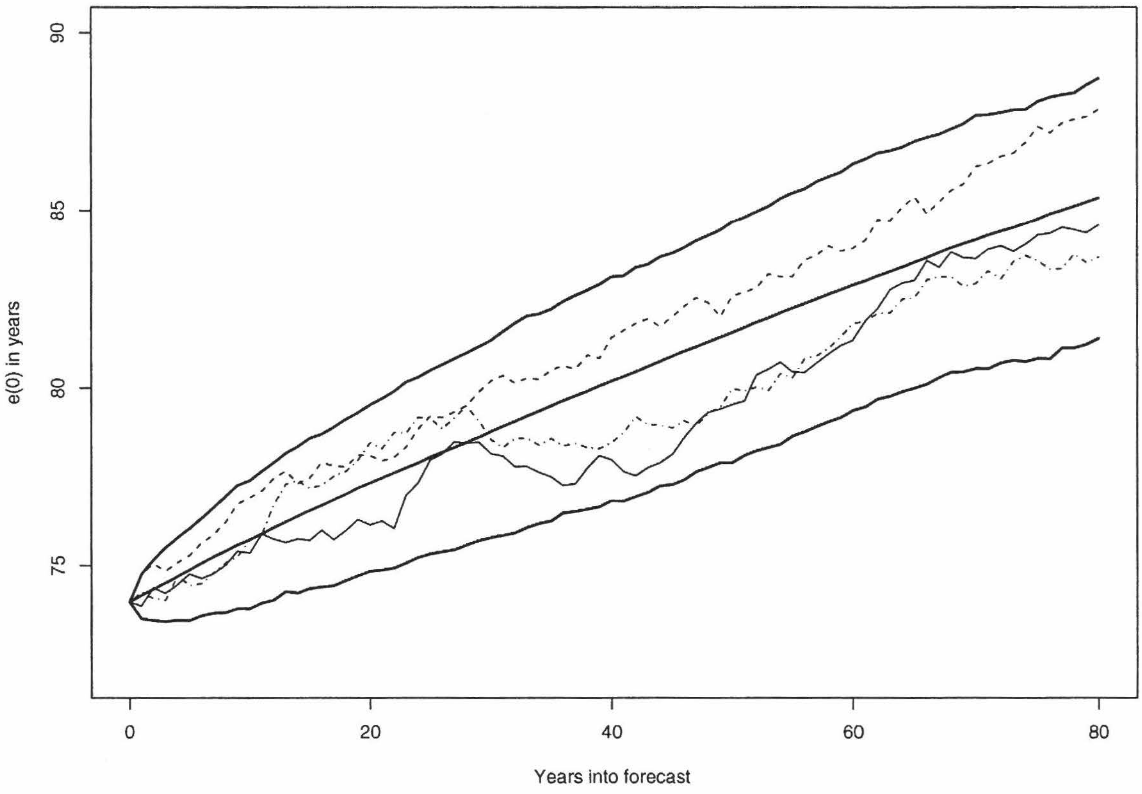
of autocorrelation over time. In the illustration, the fertility model is a constrained mean ARMA(1,1) model (Lee, 1993) and the mortality model is a random walk with drift (Lee & Carter, 1992a).

In an earlier paper, Lutz & Scherbov (1998) used the prediction interval on fertility in a given year to construct scenarios. They found that this produced larger estimates of uncertainty than the fully stochastic approach because—as Lee (1999) points out the stochastic forecasts incorporate some "cancellation of errors", whereby a path that has high fertility in one year is likely not to be the same path that has high fertility in the next year.

An alternative approach, adopted here, is to calibrate scenarios to a smoothed version of the stochastic realizations, averaging out the highs and lows. Specifically, we construct scenarios based on the uncertainty in the cumulative average of fertility from the beginning of the forecast to the date of interest. The cumulative average $\bar{X}(t)$ is defined as $\sum_0^t X(i)/t$.

An interesting feature of the cumulative average is that its variance is nearly constant after about year 20 of the forecast. In the absence of serial correlation, the variance of the cumulative average would shrink over time. However, the observed temporal autocorrelation is sufficient to keep the uncertainty interval nearly stable from year 20 through year 80 of the forecast, and shrink significantly

Figure 2. *Sample paths and forecast interval of stochastic forecast of life expectancy.*



Stochastic mortality model is a random walk with drift (Lee & Carter, 1992a) with specification and coefficients given by Lee & Tuljapurkar (1994); life expectancy is not forecasted directly but rather is a function of the mortality index. The figure shows 3 sample paths generated with Monte Carlo simulation; these can be identified by their large ups and downs. The heavy solid U-shaped outer bounds give the 95 percent prediction interval for life expectancy in any year, with the heavy middle line giving the median life expectancy, as estimated from the quantiles of 1000 simulated runs.

only later on. This property is convenient because it means that the bounds on the calibrated fertility scenarios can be approximated by a moving moving average with a window of several decades, and that the choice of window width will have a small effect on the interval coverage. The stability of the uncertainty interval of the cumulative average is a robust finding. It also is produced by other autoregressive specifications (e.g., AR(1)) and in other populations in the industrialized world (e.g. France). However, it is possible to construct models that have less stable uncertainty intervals for the cumulative average of forecast fertility (Congressional Budget Office, 2001).

For fertility the use of the variation in long-term averages seems justifiable on several counts. First, forecasters are generally not interested in the number of births in a particular year, but rather in the number of births over broad time intervals that determine the size of future broad age groups. Second, the wide span of reproductive years for any cohort means that over the course of generations, the number of potential parents in a broad age interval rather than the numbers of potential parents in a single year of age are what—in combination with period fertility—determine the number of births in a given year.

Analytically, it is difficult to show that the cumulative average is the right kind of average to use in age-structured populations experiencing stochastic renewal. However, a simple non-overlapping generation model in which one generation reproduces the next can be used to show that calibrating scenarios to the cumulative average will give approximately the same variance as the fully stochastic approach. This is demonstrated in appendix B.

Stochastic forecasters have used the cumulative average to compare the variability of their fertility forecasts with scenarios (Lee, 1993; Congressional Budget Office, 2001). The modified scenario approach set out here simply takes this comparison one step further, interpreting the prediction interval bounds of the cumulative average as high and low scenarios.

The second step in improving scenario-based forecasts is to combine the uncertainty of different age groups in a probabilistically sensible manner. The temporal correlation of the scenarios for fertility assure that a large number of births in the first years of a forecast will always be accompanied by a larger than expected number of births later on. Similarly, high fertility and low mortality will create larger than expected population sizes across all ages, while low fertility and high mortality will do the opposite. The resulting estimates of variability in age structure will be distorted because of the perfect correlations that are produced using the scenarios. A way to correct this distortion is to take account of the magnitude of correlation that exists between broad age-segments.

Table 1

Age segment correlations and coefficients of variation from stochastic forecasts.

Forecast Year	Correlation O, W	Correlation Y, W	Correlation O, Y	Coef. of Variation Y	Coef. of Variation W	Coef. of Variation O
10	1.00	0.10	0.09	0.06	0.00	0.01
20	1.00	0.06	0.05	0.15	0.00	0.02
30	0.23	0.63	0.02	0.22	0.03	0.04
40	0.12	0.78	0.01	0.30	0.07	0.06
50	0.08	0.82	-0.01	0.39	0.12	0.08
60	0.04	0.83	-0.04	0.46	0.20	0.09
70	0.12	0.86	0.00	0.52	0.28	0.10
80	0.47	0.88	0.31	0.57	0.36	0.16

Estimated from simulation of stochastic forecasts as given in Lee & Tuljapurkar (1994) with 400 independent runs. Y is youth population aged 0–19, W is working-age population aged 20–64, and O is old-age population aged 65+. Coefficient of variation is ratio of the observed standard deviation to the observed mean.

Table 1 shows the correlations estimated between broad age groups over time for the fully stochastic forecast developed by Lee & Tuljapurkar (1994). The correlations between the working age and old age segments are close to zero for years 30 through 70, due to the independence of fertility and mortality assumed in the stochastic forecasts. The high correlations between the working and old age segments in the first two decades is due to the mortality model being used which assumes correlated effects across all ages in a given period (Lee & Carter, 1992a). This high degree of correlation can be neglected since there is virtually no variation in the size of the working age population. The working age and elderly age segments become positively correlated at the very end of the forecast when fertility uncertainty reaches the oldest age segment. Generational reproduction then introduces a positive correlation between the size of the elderly generation and those born afterwards. With these two caveats, there is little correlation between the size of the working and old-age population segments.

Generational reproduction also assures a high degree of correlation between generations for working and young age segments for forecast years 30 through 80. Before about year 30, the slight uncertainty in the number of working age is determined by mortality. After about year 30, the size of the first wave of births determines in large part the size of the second wave, producing the high correlations we see in Table 1. Although uncertainty in fertility rates also has an effect, it is of second order, particularly when there is positive autocorrelation of fertility rates over time.

The set of correlations and seen in Table 1 are based on my replication of the Lee and Tuljapurkar stochastic forecasts of the U.S. population. In general, a forecaster will not have a fully stochastic forecast in hand—that is why he or she is interested in using scenarios. Accordingly, rather than using the correlations produced by the U.S. forecasts, the approach we use to approximate the uncertainty in dependency ratios assumes independence ($\rho = 0$) between the elderly and the rest of the population and perfect correlation ($\rho = 1$) between the size of the young and working population.

Assuming zero correlation between the elderly and the rest of the population also allows one to treat fertility and mortality as independent factors for the time period we are considering, since the number of elderly is unaffected by future fertility for the first 60 or 70 years of the forecast.

3 Results

We now investigate how close the modified scenario approach comes to reproducing the uncertainty estimates from fully stochastic forecasts, assumed to give the true uncertainty of future population estimates. It should be kept in mind, however, that even fully stochastic forecasts omit many sources of uncertainty including estimation of the baseline population, specification error in time series models, and the probability of significant events that did not occur in the past, such as nuclear war (Lee & Carter, 1992b).

The stochastic forecasts of the United States population by Lee & Tuljapurkar (1994) constitute our benchmark case. (For details see appendix A.) For simplicity, we use the estimated Lee and Tuljapurkar time series models for fertility and mortality but omit migration and project only the female population.

Fertility scenarios are calibrated to the same 95 percent prediction bounds of *cumulative average* fertility produced by the stochastic forecasts—the heavy dashed lines in Figure 1. The life expectancy scenarios are calibrated to the 95 percent prediction interval of the forecast shown in Figure 2. Using this broader interval for life expectancy, rather than some kind of cumulative average, turns out to produce reasonable uncertainty ranges.

The observed 1990 U.S. female population is then projected according to three variants, which we label "large", "small", and "medium". The large variant is for high fertility and low mortality. The medium variant is for medium fertility and mortality. The small variant is for low fertility and high mortality.

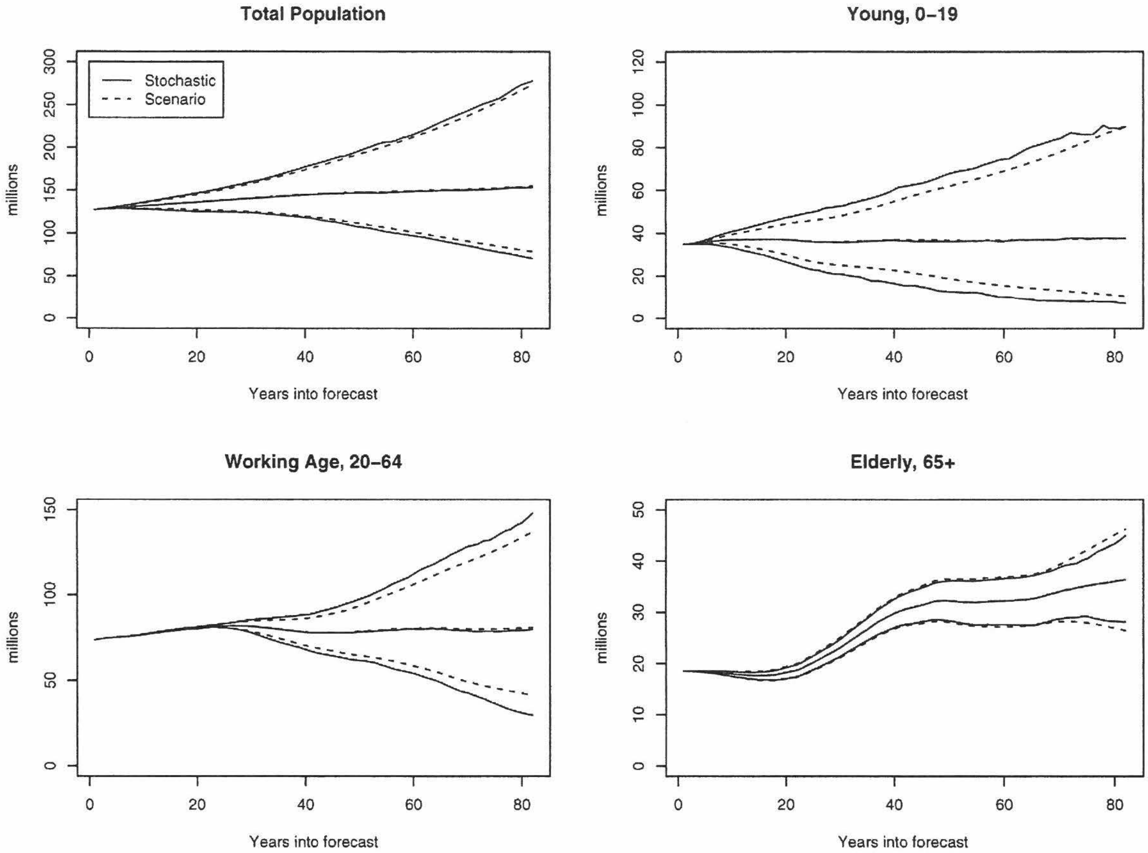
This choice of variants assures the greatest variation in age-segment sizes. As will be seen, a further round of calculation is needed to take into account the proper level of uncertainty on age structure.

3.1 Uncertainty in Total Population Size and the Size of Broad Age Segments

Figure 3 presents the total population size and the number of young, working and elderly estimated over time by the three variants. Fertility is the overwhelming source of uncertainty in population size (Alho, 1992), and we can see that the calibration of fertility scenarios to the prediction interval on the cumulative average of stochastic fertility has indeed created a very close match between the uncertainty intervals from the two approaches. This is a new finding. Past comparisons have shown that stochastic forecasts and scenarios have different ranges of uncertainty around age-segment sizes and total population (Lee & Tuljapurkar, 1994; Lutz & Scherbov, 1998). But no earlier comparisons have calibrated fertility so that its effective variation is similar in the two approaches.

The extremely close match for total population size between the two approaches results from two offsetting differences between the stochastic and scenario forecasts. The scenarios couple high

Figure 3. *Population size according to stochastic and scenario forecasts of U.S. females from 1990.*



In the scenario approach uncertainty for component age groups sums to the uncertainty in the population. In the stochastic approach, the uncertainty in the total population is less than the sum of the uncertainty of the components because of offsetting variability. Scenarios are calibrated to the uncertainty ranges in mortality and cumulative fertility.

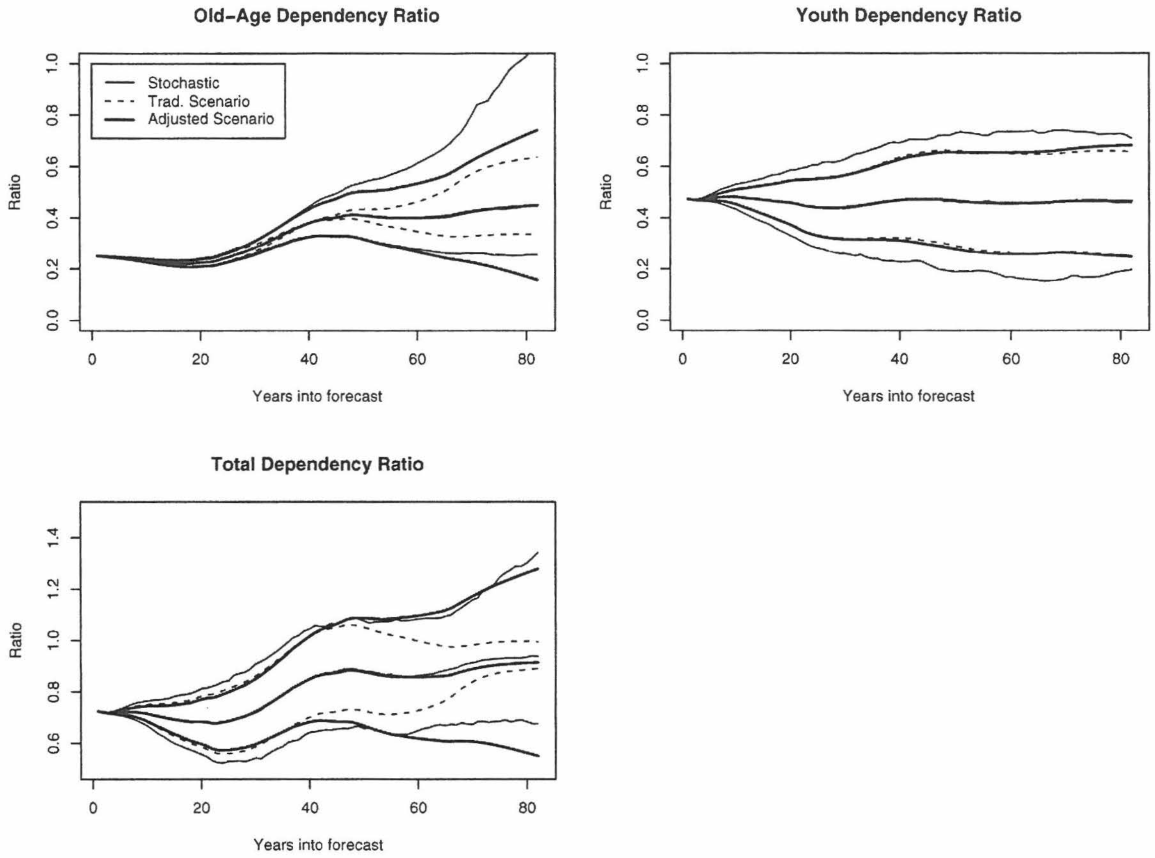
fertility with low mortality and vice-versa, inflating slightly the bounds on total population size compared to the independence between factors assumed in the stochastic forecast. On the other hand, we can see that the scenarios underestimate slightly the width of the interval covering each of the age segments, presumably because they do not allow the same degree of short-term variation in fertility rates that produces extremes in age-group sizes. The net result is that the scenarios produce almost exactly the same uncertainty ranges on total population size as the fully stochastic forecasts. An important qualitative feature is that the scenarios pick up the asymetry in the prediction interval seen in the stochastic forecasts that is due to the log-normality of population size predicted by stochastic renewal theory (Tuljapurkar, 1992).

3.2 Uncertainty in Age Structure Measures

To produce uncertainty estimates on age-structural measures from scenarios, we combine the uncertainty of different age groups in a probabilistically sensitive manner, taking into account the general magnitude of the correlations between age segments observed in Table 1.

Demographers define dependency ratios as the ratio of dependent population age groups to the working age population. Following convention, the working age population W consists of those aged

Figure 4. *Stochastic forecast of total fertility rate 95% forecast interval for each year and 95% forecast interval for cumulative average.*



The traditional scenario approach uses within-scenario age group sizes. The adjusted approach takes account of the possibility that all age group sizes may not be "large", or "small", at the same time.

20 through 64, the young population Y of those aged 0 through 19, and the elderly population O of those aged 65 and above. Then, the old-age dependency ratio (ODR) is O/W; the youth dependency ratio (YDR) is Y/W; and the total dependency ratio (TDR) is $(Y+O)/W$, the sum of the youth and old-age dependency ratios.

Our strategy is to estimate the uncertainty in dependency ratios using the standard approximation for the variance of correlated random variables (e.g. Rice (1995)),

$$\sigma^2_{X/Y} \approx \left(\frac{\mu_X}{\mu_Y}\right)^2 \left(c_X^2 + c_Y^2 - 2\rho_{XY}c_Xc_Y\right), \tag{1}$$

where μ_X and μ_Y are the means of X and Y, c_X and c_Y are the coefficients of variation (the ratio of the standard deviation to the mean), and ρ_{XY} is the correlation between X and Y.

Using the exact values of the correlation observed between age groups would produce the best estimate of uncertainty. In general, however, these will not be known, and so we assume correlations of 1 between the young and working age populations and 0 between the elderly and non-elderly populations. We estimate the coefficients of variation, assuming normality, as one-fourth the ratio of the width of the prediction interval to the central forecast.

Figure 4 shows the dependency ratios of the stochastic approach and those produced by the modified scenarios. The figure also shows the dependency ratio ranges produced by taking the range of dependency ratios directly from the scenario output.

In the case of the old-age dependency ratio, the modified scenario approach produces an uncertainty interval that tracks the stochastic intervals nearly exactly for the first 40 to 50 years of the forecast, and does so much better than the direct use of scenarios. After this, the width of the modified scenario intervals are comparable to the stochastic intervals but the use of the normal approximation does not capture the asymmetry of the stochastic forecast intervals. Still, allowing fertility and mortality to vary independently produces much more accurate uncertainty intervals than does the traditional use of scenarios.

In the case of the youth dependency ratio, our assumption of perfect correlation between the young and working age populations is equivalent to the traditional use of scenarios. Both scenario approaches produce somewhat narrower uncertainty bounds than they should. For example, using the observed correlation between the young and the working age segments given in Table 1 would increase the uncertainty by a factor of almost a third in the 70th year of the forecast. Still, our general finding is that the bounds on the youth dependency ratio, even assuming perfect correlation, are of the same general magnitude as those produced by the stochastic forecast.

Finally, we turn to the total dependency ratio, which requires an additional assumption to be made about the covariance between the youth and elderly dependency ratios, since $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y)$. The fully stochastic results (not shown) reveal that in these forecasts the old-age and youth dependency ratios are *negatively* correlated, although it is not clear that this is a general property across populations. Our estimates of uncertainty in the old-age and youth dependency ratios are both on the low side, so assuming no covariance between the two ratios offsets this bias. The result is a remarkably close match in the uncertainty intervals from the stochastic and the modified scenario forecasts. The improvement relative to the traditional use of scenarios is dramatic.

4 Adjusting Dependency Ratios from Traditional Scenarios: A Worked Example

In applications of the method proposed here, one would not generally have estimated versions of the fertility and mortality time series—if one did, one could just continue with the fully stochastic forecast. The more likely real-world situation is that one has the output from scenarios and wants to present these in a probabilistically sensible manner. To do this, one would take the output for each age group from the scenarios, use the bounds generated for each age group and total population directly from the scenarios, but recomputing the uncertainty in the total dependency ratios using the methods proposed here.

As an example, Table 2 presents the projected populations in year 50 of the forecast (year 2040) based on the scenarios used in this paper. The coefficients of variation estimated as (Large–Small)/(4 × Medium), based on assuming normality with the Medium scenario as mean. We then apply equation 1 assuming $\rho = 0$ for the ODR and $\rho = 1$ for the YDR. The TDR is calculated by assuming independence between the ODR and YDR, making the variances additive.

The table shows the original and adjusted dependency ratios along with the ranges calculated using equation 1. We see that the YDR is unaffected by the adjustment, since the scenarios implicitly assume correlations of 1 between the young and working age groups. The uncertainty range in the old-age dependency ratio, on the other hand, increases substantially, some 5-fold, due to our adjustment. This is because the original scenarios always combine large working populations with large elderly populations and small working populations with small working populations, whereas in fact independence—or something very close to it—is warranted. The adjusted uncertainty in the total dependency ratio is also larger than the uncertainty implied by the original scenarios. Here the adjustment increases the estimate of uncertainty by about a factor of 1/4. The adjusted TDR differs

Table 2

Illustrative example of calculation of adjusted dependency ratios.

(a) Age-group results of year 50 of scenario forecasts

	Millions of females			
Age group	Small	Medium	Large	Coef. of Var.
0–19	36.7	18.4	62.5	0.30
20–64	78.9	64.2	94.8	0.10
65+	32.2	27.9	36.5	0.07

(b) Original and adjusted dependency ratios

	ODR		YDR		TDR	
	original	adjusted	original	adjusted	original	adjusted
Low	0.39	0.31	0.29	0.29	0.72	0.66
Medium	0.41	0.41	0.47	0.47	0.87	0.87
High	0.43	0.51	0.66	0.66	1.04	1.08

Coefficients of variation estimated using normal approximation as one-fourth the difference between Large and Small divided by Medium. "Original" dependency ratio uncertainty is directly from scenario outputs. "Adjusted" uncertainty range between "Low" and "High" is the 95% forecast interval based on equation 1 and normal approximation. Adjusted TDR assumes independence of YDR and ODR.

from the original TDR by larger amounts later in the forecast (See Figure 4).

This example illustrates how to use the scenario results, conditional on having scenarios that themselves have some probabilistic interpretation. In real applications, one would need to specify the scenarios. Alternatives to stochastic time series could involve the use of judgement and expert opinion (Lutz *et al.*, 2001) and/or *ex post* evaluation of past scenarios. These approaches are discussed further below.

5 Discussion

In this section, we first discuss the need for accurate estimates of uncertainty and then consider ways to implement the modified scenario approach using stochastic time series methods, expert opinion, and existing scenario-based forecasts.

5.1 How Accurate do Uncertainty Measures Need to Be?

The modified scenario approach produces uncertainty intervals that are broadly similar to those produced by the fully stochastic forecasts. But differences emerge, particularly for forecast horizons beyond about 60 years when fertility uncertainty begins to influence the number of elderly.

How accurate do estimates of uncertainty need to be? It is sometimes argued that estimates of uncertainty from the fully stochastic forecasts are misleadingly precise. Exact intervals are produced but these are conditional on the correct specification of the time series model and more importantly on the assumption that past variability is a good estimate of future uncertainty. In such a situation, approximations of stochastic forecasts such as those produced here may arguably give a somewhat more appropriate, even if less precise, depiction of the uncertainty in uncertainty estimates.

At a practical level, the degree of uncertainty at longer time horizons is often of less interest to contemporary planners because advance warning will come if unexpected changes take place. For example, a baby boom in the decade of the 2020s would surprise most contemporary observers. But because it will take some years for these newborns to enter school, even more years for them to enter the workforce, and many decades for them to enter the retirement system, there will be advance notice, which new forecasts can then take into account. Even when a long planning horizon is legally mandated, as is the 75 years for actuarial reports of the U.S. social security system, policy debates usually revolve around changes expected in the next several decades. A final factor giving short horizons greater importance is that policy makers recognize that the distant future is more uncertain than the near future and are thus less likely to lock in inflexible policies with extremely long time horizons. For all of these reasons it is important to emphasize the success of the modified scenarios during the first five decades of the forecast.

5.2 Calibrating Scenarios

Here, I have calibrated scenarios to the prediction intervals generated by stochastic time-series models. When time-series models for vital rates are available, it will usually be worth continuing with fully stochastic forecasts of the population, with calibrated scenarios best used as a tool for checking stochastic forecasts.

The modified scenario approach is most useful when time series estimates are either unavailable or, for whatever reason, unbelievable. In these instances, one would want to calibrate the scenarios to some other standard. Two approaches seem reasonable. The first is a minor modification of existing scenario methods whereby experts either within or outside of forecasting agencies specify high and low scenarios. In this case, the results presented here suggest that there be an explicit statement by those designing fertility scenarios that the scenarios apply to long-term averages, not to the uncertainty in any given year. The stability of the uncertainty interval of the averages over a broad range of cumulative averages from 20 to 80 years reduces the need for experts to worry about the precise form of the average they are considering, as long as it is made clear that it is at least several decades.

A second approach would use the effective uncertainty range covered by past scenarios and apply it to the current production of scenarios. Stoto has found that forecasts in the industrial world tend to cover about a two-thirds prediction interval in population size in any given year (Stoto, 1983). The present analysis suggests doing an *ex post* analysis exercize but with a focus on the cumulative average of fertility. The quantile range covered by past scenarios could then be used to give a probabilistic interpretation to the results from new scenario-based projections, extrapolating these *ex post* errors into the future.

Scenario forecasts could also be made without specifying the probability interval covered by "high" and "low" but by using the methods of combining age segments and calculating the uncertainty in dependency ratios. This would assure greater consistency in the uncertainty range of population age group sizes, total population, and dependency ratios, so that all would cover approximately the same uncertainty interval, even if it remained undefined.

6 Conclusion

We have found that scenarios, when used with some care, can come close to reproducing the uncertainty estimates produced by fully stochastic population forecasts. The key elements are the specification of a probability interval covered by the scenarios, the calibration of fertility scenarios to variability in long-term average fertility, and the probabilistic combination of uncertainty from distinct age segments.

Several caveats are in order. The first is that the uncertainty estimates from our use of scenarios came closest to the estimates from stochastic scenarios for the first 60 years or so. For long term forecasting, full incorporation of stochastic renewal becomes more important. Second, our results apply to populations in which future fertility and mortality can be assumed to be independent and where there is little mortality at younger ages. The probabilistically sensitive use of scenarios in populations in which mortality effects the size of the young and working age populations needs further research. Third, the methods used here apply only to broad age segments. They are inappropriate for assessing the relationships between the joint uncertainty of functions of narrower age groups. For uncertainty estimates for detailed age-distributions, the fully stochastic approach is needed. Finally, further research is needed to assess uncertainty in time-integrated functions of population structure, such as the viability of pay-as-you-go pension systems. Unlike the scenarios presented here, the fully stochastic approach can be integrated over time without concerns about probabilistic consistency.

Overall, our results suggest some interesting future applications of probabilistically sensible scenario forecasts. It may be possible to translate existing scenario forecasts into approximate probabilistic forecasts fairly easily. It also suggests a method for translating expert-based forecasts into probability intervals. Finally, the approach offers a useful first step for those agencies considering the production of probabilistic forecasts using the fully stochastic time-series approach.

Acknowledgement

I thank Gyanendra Badgaiyan for research assistance and am grateful for comments on an earlier version of this paper to Juha Alho, Ronald Lee, Tim Miller, Guy Stecklov, Shripad Tuljapurkar, Warren Sanderson, and participants of the Stochastic Forecasting Workshop held in December 2002 at the Vienna Institute for Demography.

References

Alho, J.M. & Spencer, B.D. (1991). A population forecast as a data base: implementing the stochastic propagation of error. *Journal of Official Statistics*, **7**, 295–310.

Alho, J.M. (1992). The magnitude of error due to different vital processes in population forecasts. *International Journal of Forecasting*, **8**, 301–314.

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

Congressional Budget Office (2001). Uncertainty in social security's long-term finances: a stochastic analysis. CBO Paper, December 2001, Congress of the United States.

Keilman, N., Pham, D.Q. & Hetland, A. (2002). Why population forecasts should be probabilistic—illustrated by the case of Norway. *Demographic Research*, **6**, 410–453, Article 15.

Keyfitz, N. (1972). On future population. *J. Amer. Statist. Assoc.*, **67**(338), 347–363.

Lee, R.D. (1993). Modeling and forecasting the time series of U.S. fertility: age distribution, range, and ultimate level. *International Journal of Forecasting*, **9**, 187–202.

Lee, R.D. (1999). Probabilistic approaches to population forecasting. In *Frontiers of Population Forecasting*, Vol. **24** of *Supplement to Population and Development Review*, Eds. W. Lutz, J. Vaupel and D. Ahlburg, pp. 156–190. New York: Population Council.

Lee, R.D. & Carter, L. (1992a). Modeling and forecasting the time series of U.S. mortality. *J. Amer. Statist. Assoc.*, **87**, 187–202.

Lee, R.D. & Carter, L.R. (1992b). Modeling and forecasting U.S. mortality: rejoinder. *J. Amer. Statist. Assoc.*, **87**(419), 674–675.

Lee, R.D. & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: beyond high, medium, and low. *J. Amer. Statist. Assoc.*, **89**(428), 1175-1189.

Lee, R.D. & Tuljapurkar, S. (1998). *Stochastic forecasts for social security*. In *Frontiers in the Economics of Aging*, Ed. D. Wise, pp. 393–420. Chicago: University of Chicago Press.

Lutz, W. & Scherbov, S. (1998). An expert-based framework for probabilistic national population projections: the example of Austria. *European Journal of Population*, **14**, 1–17.

Lutz, W., Sanderson, W. & Scherbov, S. (2001). The end of world population growth. *Nature*, **412**, 543–545.

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

Rice, J.A. (1995). *Mathematical Statistics and Data Analysis*, second edn. Belmont, California: Duxbury.

Stoto, M. (1983). The accuracy of population projections. *J. Amer. Statist. Assoc.*, **78**, 13–20.

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

Résumé

L'approche traditionnelle à trois scénarios (supérieur-moyen-inférieur) pour quantifier l'incertitude dans les projections de population a été critiquée pour son manque de signification et de cohérence probabilistes. Cet article montre, sous certaines hypothèses, comment on peut utiliser à bon escient des scénarios calibrés pour estimer les intervalles d'incertitude sur la taille et la structure par âge de la population totale, obtenues avec des projections entièrement stochastiques. De nombreuses organisations produisent déjà des scénarios de projections que l'on sait traiter. Aussi les méthodes présentées ici apportent un point de vue intermédiaire intéressant entre l'analyse d'un scénario incohérent au sens probabiliste et les projections entièrement stochastiques.

Appendices

A Stochastic Processes for Vital Rates

All of the estimates in this paper are based on the estimates in Lee & Tuljapurkar (1994) and Lee (1993). The fertility model used is a constrained autoregressive moving average model, ARMA(1,1), given by

$$F_t = cF_{t-1} + F(1-c)u_t + du_{t-1}, \tag{2}$$

with $c = 0.9676$ and $d = 0.47978$, the ultimate mean level equal to $F = 2.1$ and with $u_t \sim \mathcal{N}(0, \sigma_u)$, $\sigma_u = 0.110663$. This index of fertility is translated to age specific fertility rates according to the schedules in Lee (1993).

The mortality model is a random walk with drift, given by

$$K_t = K_t - z + \epsilon_t, \tag{3}$$

with $z = 0.365$ and $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$, $\sigma_\epsilon = 0.651$. This mortality factor is translated to age-specific mortality rates, namely

$$\log m_{x,t} = \alpha_x + \beta_x K_t.$$

The schedules α_x and β_x are given in Lee & Carter (1992a).

Both formulations assume perfect autocorrelation between age-groups from year to year. Estimates of uncertainty omit uncertainty in the estimation of parameters, following Lee & Tuljapurkar (1994).

B A Generational Model of Stochastic Renewal

To represent the stochastic forecast, let X_t be the net reproductive rate of generation t such that

$$G_t = (X_t \cdot X_{t-1} \cdot \ldots \cdot X_1)G_0.$$

The X_t are serially correlated random variables such that

$$X_t = M + \epsilon_t,$$

where M is the expected net reproduction rate and ϵ is a random variable with zero mean, variance σ_t^2 and $\text{cov}(\epsilon_t, \epsilon_{t-1}) \geq 0$.

As an alternative, generation size in the scenario based forecast can be defined as

$$G_t^s = Y^t G_0,$$

where now Y is a single random variable that when multiplied by itself t times will give the change in G_0 that gives G_t.

Letting $Z_t^s = Y^t$ and $Z_t = \prod^t X_i$, we wish to find $E(Y)$ and $\text{var}(Y)$ such that $E(Z_t) = E(Z_t^s)$ and $\text{var}(Z_t) = \text{var}(Z_t^s)$. Using the delta method to first order for the expectation (Rice, 1995), $E(Z_t) \approx E(X_1) \ldots E(X_t) = M^t$ and $E(Z_t^s) \approx E(Y)^t$. Thus $E(Y) \approx M$. The same method, after some algebra, gives a first order approximation of the variance of the generational stochastic model of

$$\mathrm{var}(Z_t) \approx M^{2(t-1)}\mathrm{var}(X_1 + \ldots + X_t).$$

Substituting $Y = X_i$ for all i, the delta approximation of the generational scenario model gives

$$\mathrm{var}(Z_t^s) \approx M^{2(t-1)}\mathrm{var}(tY).$$

We can now find the variance of Y by setting $\mathrm{var}(Z_t) = \mathrm{var}(Z_t^s)$. This gives

$$\mathrm{var}(Y) = \mathrm{var}\left(\frac{X_1 + \ldots + X_t}{t}\right),$$

which shows that calibrating Y to the variance of the cumulative average of X will equalize the variances of generation size.

The generational model differs from that used in applications in several respects. First, in applications the projections are age-structured, with age groups of 1 or 5 years rather than generational. Second, the scenario calibration ends up giving the scenarios Y a different variance for each year of the forecast, although as we saw this stabilizes after about year 20 because of autocorrelation in the process for annual fertility fluctuations. Finally, the application considers both fertility and mortality independently, rather than net fertility, although the distinction in low mortality populations is not substantial.

The simulation results shown in the body of the paper show that these differences do not, in practice, change the approximate equivalence of the two approaches.

[*Received December 2002, accepted November 2003*]

International Statistical Review (2004), **72**, 2, 157–166, Printed in The Netherlands

Conditional Probabilistic Population Forecasting

Warren C. Sanderson[1], Sergei Scherbov[2], Brian C. O'Neill[3] and Wolfgang Lutz[4]

[1]*State University of New York, Stony Brook, NY, USA. E-mail: wsanderson@notes.cc.sunysb.edu*
[2]*Vienna Institute of Demography, Austrian Academy of Sciences and the International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. E-mail: sergei.scherbov@assoc.oeaw.ac.at*
[3]*International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria and Watson Institute for International Studies, Brown University, Providence, RI, USA*
E-mail: oneill@iiasa.ac.at
[4]*International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria and Vienna Institute of Demography of the Austrian Academy of Sciences. E-mail: lutz@iiasa.ac.at*

Summary

Since policy-makers often prefer to think in terms of alternative scenarios, the question has arisen as to whether it is possible to make conditional population forecasts in a probabilistic context. This paper shows that it is both possible and useful to make these forecasts. We do this with two different kinds of examples. The first is the probabilistic analog of deterministic scenario analysis. Conditional probabilistic scenario analysis is essential for policy-makers because it allows them to answer "what if" type questions properly when outcomes are uncertain. The second is a new category that we call "future jump-off date forecasts". Future jump-off date forecasts are valuable because they show policy-makers the likelihood that crucial features of today's forecasts will also be present in forecasts made in the future.

Key words: Forecasting; Population forecasting; Probabilistic forecasting; Scenarios; Scenario analysis.

1 Introduction

The last decade and a half has witnessed rapid development in the area of probabilistic population forecasting (see Alho, 1990, 1997; Alho & Spencer, 1985; Keilman *et al.*, 2002; Lee, 1999; Lee & Tuljapurkar, 1994; Lutz *et al.*, 1996, 1997, 2001; Lutz & Scherbov, 1998; and Pflaumer, 1988, among others). A probabilistic forecast goes beyond a traditional deterministic one by providing an integrated estimate of the forecast's uncertainty, often a crucial quantity for decision-makers. These forecasts give distributions of outcomes rather than single numbers resulting from alternative scenarios. Since policy-makers often prefer to think in terms of alternative scenarios (for example, outcomes with and without a certain policy), the question has arisen as to whether it is possible to make conditional forecasts in a probabilistic context.

This paper answers that question by demonstrating how to obtain conditional probabilistic population forecasts. We do this with two different kinds of examples. The first is the probabilistic analog of deterministic scenarios and the second is a new category that we call "future jump-off date forecasts". Both are important for policy analysis.

Scenario analysis is essential for policy-makers because it allows them to answer "what if" type questions. For example, they may want to know *what* the age structure of their country would be in fifty years *if* fertility were lower than in the official projections. Future jump-off date forecasts are valuable because they help in answering questions about the value of waiting to learn about how the future is unfolding. For example, a country may be deciding on whether to build up a retirement fund for its citizens. The decision could be made to raise taxes now or to wait ten years to improve its projections of future population aging. Future jump-off date forecasts allow us to assess how much uncertainty about the future is likely to be resolved by waiting.

In Section 2, we briefly discuss the probabilistic forecasting methodology used in Lutz *et al.* (2001). It is the basis for the quantitative examples in the next two sections. In Section 3, we discuss the probabilistic counterpart of traditional scenario analysis. Section 4 presents a first look at future jump-off date forecasts. Section 5 contains some concluding thoughts.

2 An Introduction to the Methodology

Creating population forecasts from an initial distribution of the population by age and sex and forecasts of total fertility rates (TFRs), life expectancies at birth, and net migration rates is a widely accepted procedure. Probabilistic population forecasts differ from deterministic forecasts in that they quantify the uncertainty of the course of future rates and therefore must specify future total fertility rates, life expectancies, and net migration rates as distributions and not as points. Distributions can also be used to quantify other uncertainties such as those relating to the base population size.

In order to generate the required distributions, Lutz *et al.* (2001) let ν be the total fertility rate, the change in life expectancy at birth, or net migration to be forecasted for periods 1 through T and ν_t be its forecasted value at time t. The forecasted value, ν_t, can be expressed as the sum of two terms, its trend (mean) at time t, $\overline{\nu}_t$, and its deviation from the mean at time t, ε_t. In other words, $\nu_t = \overline{\nu}_t + \varepsilon_t$, where ε_t is the idiosyncratic noise. The $\overline{\nu}_t$ were chosen based on the arguments given in Lutz *et al.* (1994, 1996) and updated based on subsequent information. The ε_t term is assumed to be a normally distributed random variable with mean zero and standard deviation $\sigma(\varepsilon_t)$. The $\sigma(\varepsilon_t)$ are also based on arguments from the same sources.

Due to the persistence of the factors represented by the ε_t, we would generally expect them to be autocorrelated. One of the most commonly used methods of specifying how the ε_t term evolves over time is the simple autoregressive formation (AR(1)), where $\varepsilon_t = \alpha \cdot \varepsilon_{t-1} + u_t$, where u_t is an independently distributed normal random variable with mean zero and standard deviation $\sigma(u)$. Another commonly used method is the moving average formation of order q, MA(q) where q is the number of lagged terms in the moving average. We use the following moving average specification:

$$\varepsilon_t = \sum_{i=0}^{q} \alpha_i \cdot u_{t-i},$$

where u_{t-i} are independently distributed standard normal random variables. To ensure that the standard deviation of ε_t is equal to its prespecified value,

$$\alpha_i = \frac{\sigma(\varepsilon_t)}{\sqrt{q+1}}.$$

The choice between AR(1) and MA(q) does not have to do with estimation, but rather with representation. Data do not exist that would allow the estimation of the parameters of either specification at the regional level used in Lutz *et al.* (2001). Neither is more theoretically correct than the other. Both are just approximations to a far more complex reality. When comparably parameterized, they produce very similar distributions of ε_t.

The choice between the two, therefore, rests on which more accurately reflects arguments concerning the future. From our perspective, the moving average approach has the advantage that the $\sigma(\varepsilon_t)$

terms appear explicitly making it easier to translate ideas about the future into that specification.

The future levels of vital rates can be correlated in different ways. Most important are (a) the correlations between deviations from assumed average trends in fertility and mortality rates, (b) the autocorrelation of deviations within each series of vital rates and (c) the correlations among the deviations from the average vital rate trends in different world regions. The forecasts of the world's population used in this paper assume: (1) a zero correlation between fertility and mortality deviations from their trends within regions, (2) a 31 term moving average specification separately for fertility and mortality deviations, which implies an autocorrelation between deviations one year apart of around 0.96, and (3) cross-regional correlations of fertility and mortality deviations within each year of 0.7 and 0.9 respectively. This methodology is considerably different from the one used in Lutz *et al.* (1996, 1997), where piecewise linear paths for future vital rates were used.

Due to temporal and regional correlations, vital rates paths for all regions are determined simultaneously and then used to make population forecasts, which were aggregated to the world total. This process was repeated 2,000 times, generating a distribution of world population sizes for each year from 2001 to 2100.

3 Conditional Probabilistic Forecasting and Scenario Analysis

One important audience for probabilistic forecasts is the user community. Often when demographers want to communicate the importance of particular variables in their forecasts to members of this community, they use scenarios. In population forecasting, scenarios are typically clear "if ... then" statements in which the implications of a certain set of assumptions on fertility, mortality and migration are being demonstrated (Lutz, 1995). Such scenarios can illustrate the laws of population dynamics but do not give the user any information about the likelihood of the described path. For instance, an immediate replacement fertility scenario merely shows what would happen if fertility immediately jumped to the replacement level without saying that this is a likely or even plausible path. For policy makers who want to know what would be the long-term consequences of alternative fertility trends resulting from alternative policies, for example, such scenarios can nonetheless be useful guides. Conditional probabilistic forecasting is a way of posing and answering the same type of question within a probabilistic framework.

The first discussion of conditional probabilistic population forecasts, of which we are aware, appears in Alho (1997). Alho first turned the deterministic world population forecasts in Lutz *et al.* (1994) into a probabilistic one and computed the probability of the world's population falling between the high and low scenarios. Next, Alho considered the case where the UN's world population forecasts for 2025 could be regarded as a Lutz *et al.* (1994) forecast conditional on the success of family planning programs. Alho showed that if the probability of being between the UN's high and low variants was 75 percent, then those programs would have to reduce the variance of the probabilistic version of the Lutz *et al.* (1994) forecasts by at least 42 percent. Alho regards this as "much too high to be credible" in light "of the past record of ineffectiveness of government interventions concerning fertility in the industrialized countries" (p. 83). He showed that if the reduction in the variance were less than 42 percent, then the probability content of the interval between the UN high and low variants must be less than 75 percent.

Alho (1997) is an example of taking known unconditional and conditional distributions and learning about the nature of the conditional distribution by studying the plausibility of the conditions needed to obtain it from the unconditional one. Here, an example that is at the other end of the continuum, is presented. Starting with unconditional distributions and conditions that are of interest to policy-makers, the example demonstrates how probabilistic forecasting can produce conditional distributions that are useful in scenario analysis. There are many possible intermediate cases as well, where information about some aspects of conditional distributions and some features of the conditions themselves are combined in order to investigate particular questions. An example of this can be found in O'Neill in this issue (O'Neill, 2004).

The approach used here was developed in Sanderson *et al.* (2004). An application on whether immigration can compensate for Europe's low fertility appears in Lutz & Scherbov (2002). The example begins with Figure 1, which shows the distribution of the world's population in 2050 conditional on average fertility and mortality levels for the world over the period 2000–2050. The x-axis is divided into three ranges labeled "low fertility", "medium fertility" and "high fertility". Low fertility includes all of the 2,000 simulated futures where the average total fertility rate in 2000–2050 was below 1.6. Medium fertility includes those paths where the average total fertility rate was between 1.6 and 1.8; and high fertility includes paths in which the average total fertility rate (over the whole projection period) was above 1.8.

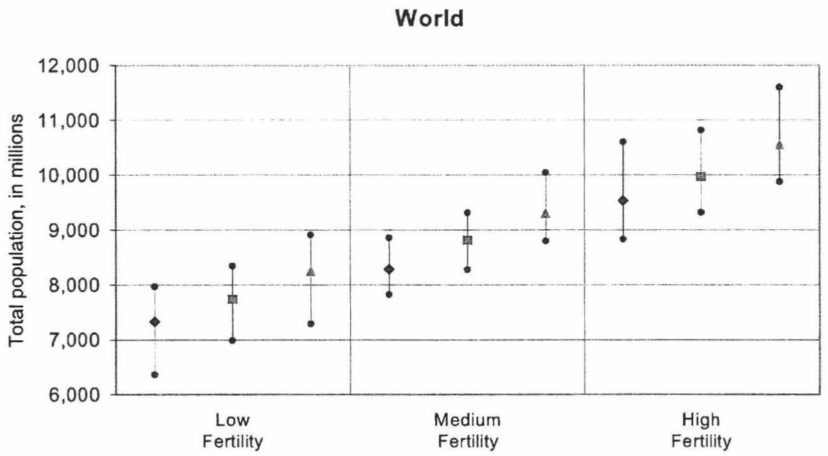


Figure 1. *Median and interdecile ranges for the world population, conditional on three alternative fertility and mortality levels. The three lines within each category refer to the low (left), central (middle) and high (right) groups of life expectancy. Source: Authors' calculations.*

Within each of the three panels there are three lines that have different symbols near their centers. The lines with the diamonds near their centers refer to paths where the global average life expectancy at birth was lower than 68 years. The lines with the dented squares refer to paths where the average life expectancy was between 68 and 71 and the lines with triangle shapes to paths with average life expectancies over 71. The aggregations of the total fertility rates and life expectancies at birth were chosen so that one-third of our paths was in each group. The symbols are placed at the medians of the distributions. The circles at the endpoints of the lines indicate the 80 percent prediction intervals.

Now we are in the position to answer some "what-if"-type questions. For example, what would be the effect on world population size in 2050 of high fertility trends versus low fertility trends over the coming decades combined with the medium range of uncertainty for future mortality? We can immediately read the answer off the figure. In the middle group, the median population of the world in 2050, if we experienced low fertility, would be around 7.7 billion people with the 80 percent prediction interval covering the range 7.0 to 8.3 billion people. If we experienced a high fertility world, the median population would be considerably higher, around 10.0 billion people, with a prediction interval between 9.2 and 10.9 billion people. The difference between the medians is 2.3 billion people, which is quite large considering that the median of the unconditional population distribution is 8.8 billion people. Clearly, the difference in fertility is very significant.

We can also read the figure to tell us about the influence of differences in life expectancies on future population size. We can do this easily by looking at the middle panel, labeled "medium fertility". When life expectancies are in the low group, the median population size is 8.3 billion. When they are in the high group, the median population is 9.2 billion. Therefore, in 2050 the effect on population size of moving from low to high fertility, keeping life expectancy constant, is much larger than the effect of moving from low to high life expectancy, keeping fertility constant.

Figure 2 is similar to Figure 1, except that it deals with the proportion 60 years and above. As fertility increases, the proportion 60 and above decreases, but as life expectancy increases, the proportion gets larger. Let us consider the difference in the proportion due to having high fertility as opposed to low fertility, again assuming medium life expectancy. The median proportion is 25 percent when fertility is low and around 19 percent when it is high. Assuming medium fertility and varying mortality, we see that when mortality is low the proportion is below 20 percent, compared to 24 percent when mortality is high. Thus, the effects of fertility and mortality are more similar in determining the proportion 60 and above than they are in determining population size.

The two examples in this section show that in making the transition from deterministic to probabilistic forecasting, we do not have to give up on answering the kinds of "what-if" questions that users and policy-makers so often pose.

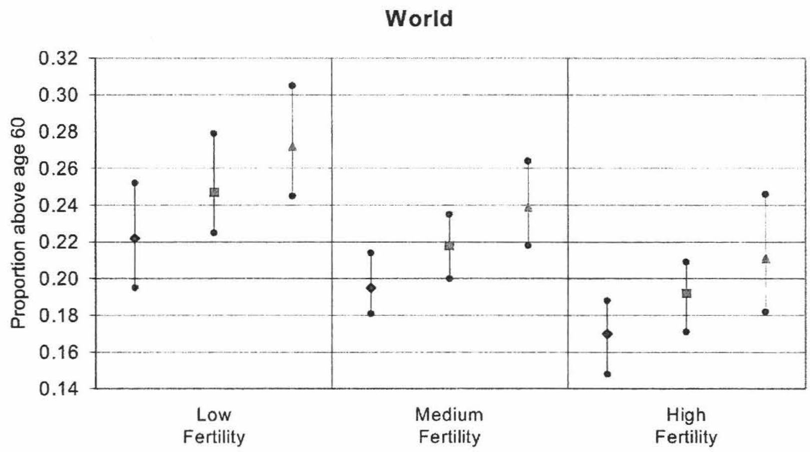


Figure 2. *Median and interdecile ranges for the global proportion above age 60, conditional on three alternative fertility and mortality levels. The three lines within each category refer to the low (left), central (middle) and high (right) groups of life expectancy. Source: Authors' calculations.*

4 Conditional Probabilistic Forecasts with Future Jump-Off Dates

In many policy areas, we come across the question: Should we act now or should we wait until we learn more? Waiting has a cost because it can foreclose certain policy options or make them more expensive. On the other hand, by waiting policy-makers could possibly acquire important and relevant information, and avoid potentially unnecessary policy interventions. Since population is an important driver of many processes, it is valuable to know how much the demographic outlook might change if we wait.

For example, the question of whether to act now or wait to learn more is central to the debate over climate change policy. The climate change issue is characterized by both long timescales—today's emissions of greenhouse gases will affect the climate for decades to centuries—and substantial uncertainties in climate impacts on society and costs of emissions reductions. Many argue that it would be beneficial to wait to learn more (and reduce uncertainties) before deciding whether, and how much, to reduce emissions. This strategy would avoid investments in emissions reductions that may turn out to be unnecessary. Others argue that reductions should begin now, because if climate change turns out to be serious, we would later regret not acting early. The question of how the potential for learning about various aspects of the problem affects today's optimal decision remains unresolved (Webster, 2002). It is possible that learning about the outlook for future population growth could impact such decisions. Population is one factor affecting the outlook for future greenhouse gas emissions. If, by waiting a decade or two, we learn that population is likely to be much lower in the future than we currently expect, our outlook for future emissions will also likely be revised downward, reducing the urgency of emissions reductions. If we learn that population is likely to be much higher, our outlook for emissions will also be higher, justifying more aggressive action to reduce emissions.

Probabilistic forecasts with future jump-off dates are constructed to help us learn about the value of waiting for more information. These forecasts are, of course, conditional on what happens between the beginning of the current forecast period and the future jump-off date. For example, imagine that it is the year 2000 and forecasts are made of the distribution of the size of the world's population in 2050. How different would the forecasted distribution of population sizes be in 2050 if the forecast were made in 2010 instead of 2000? We do not have to wait to 2010 to answer this question. The technique of making probabilistic forecasts with future jump-off dates allows us to think about this question now.

Projections in 2010 may differ from projections in 2000 because something is learned between now and then. At a minimum, the values of demographic variables like population size, fertility, mortality, and migration in that ten-year period will be observed. Other factors such as new policies, economic trends, or social conditions that are relevant to the outlook for future demographic rates will also be observed. It is possible as well that demographic theory will be improved through research, that new breakthroughs in health (or new epidemics of disease) will occur, or that new contraceptive technology will be developed. All of these types of learning could change the outlook for the future. Learning based on these other factors is not considered here. In the example below, learning is only based on the observation of demographic variables. While learning by observation is only one type of learning, it is likely to be an important one in population projections.

In this section, we take some small first steps toward understanding how this passive learning process takes place, so that users of forecasts are not surprised when forecasts change and so that policy makers can use probabilistic forecasts in the design of adaptive policies.

Let us imagine that it is now 2010 and all the relevant population information has been compiled and is available. Certainly it would be appropriate to make new forecasts, even if the methodology and assumptions that were originally used were completely correct. The forecasts based in 2010 would take into account what actually happened between 2000 and 2010. Without actually making new forecasts, the projections made in 2000 could be used to anticipate what new projections would look like.

In order to make this inquiry practical, a very simple approach will be used here. Instead of observing exact population characteristics, the assumption is made that only whether or not global population size is above or below the median of its distribution can be observed. There is nothing theoretically attractive in dividing the observations into only two groups in 2010, but it makes this introduction to passive demographic learning as simple as possible.

Table 1 consists of two panels. Panel A provides the distributions of future world population size expressed in intervals below 6 billion, 6 to 7 billion, 7 to 8 billion, and so on with the uppermost interval being above 12 billion. The numbers in the cells are the percentages of our 2,000 simulated future population paths. Median population sizes are in column 9. The tenth column contains an uncertainty measure, the relative interdecile range (RIDR) defined as the difference between the ninth decile and the first decile of the distribution divided by the median.

Panel B is based on a division of the 2010 distribution into population paths that were above the median in that year and those that were below it. There are 1,000 observations in each of these subgroups. There are two rows in Panel B for each decade following 2010, one labeled with an "L" and another with an "H". The "L" rows are the population distributions at the indicated date for the observations that were below the median in 2010 and the "H" rows are from the paths that were above the median in 2010.

Table 1

Forecasted distributions of the world's population size beginning in 2000 and beginning in 2010. Source: Authors' calculations.

World	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: 2000 Jump-Off Date										
	Below 6	6 to 7	7 to 8	8 to 9	9 to 10	10 to 11	11 to 12	Above 12	Median	RIDR*
2000	0	100	0	0	0	0	0	0	6055	0
2010	0	85.05	14.95	0	0	0	0	0	6828	0.062
2020	0	8.05	81.45	10.5	0	0	0	0	7538	0.129
2030	0.1	3.05	41.85	47.35	7.5	0.15	0	0	8085	0.195
2040	0.15	3.75	24.85	40.25	25.15	5.2	0.65	0	8525	0.27
2050	0.5	5.05	18.95	30.95	26.8	13.45	3.3	1	8796	0.352
2060	1.55	7	17.45	25.75	22.35	16.25	6.45	3.2	8935	0.427
2070	4	8.35	16.9	21.2	19.9	15.05	8.5	6.1	8974	0.52
2080	6.8	10.1	16	18.45	18.45	12.55	9	8.65	8890	0.606
2090	10	12.75	14.85	17.9	15	12.25	6.45	10.8	8678	0.702
2100	14.25	14.05	14.45	16.5	12.9	10.45	6.85	10.55	8413	0.779

* Relative Interdecile Range (RIDR) is measured as the difference between the ninth decile and the first decile divided by the median.

World	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel B: 2010 Jump-Off Date										
	Below 6	6 to 7	7 to 8	8 to 9	9 to 10	10 to 11	11 to 12	Above 12	Median	RIDR*
2020/L	0	16.1	83.9	0	0	0	0	0	7268	0.088
2020/H	0	0	79	21	0	0	0	0	7787	0.081
2030/L	0.2	6.1	70.3	23.4	0	0	0	0	7704	0.147
2030/H	0	0	13.4	71.3	15	0.3	0	0	8486	0.139
2040/L	0.3	7.5	42.3	42.4	7.4	0.1	0	0	7996	0.224
2040/H	0	0	7.4	38.1	42.9	10.3	1.3	0	9083	0.216
2050/L	1	9.9	31.1	36.6	17.9	3.2	0.3	0	8152	0.294
2050/H	0	0.2	6.8	25.3	35.7	23.7	6.3	2	9521	0.279
2060/L	3.1	12.8	26.9	31.3	17.3	6.8	1.6	0.2	8256	0.389
2060/H	0	1.2	8	20.2	27.4	25.7	11.3	6.2	9760	0.352
2070/L	7.8	13.6	23.9	25.2	17.4	8.7	1.9	1.5	8213	0.485
2070/H	0.2	3.1	9.9	17.2	22.4	21.4	15.1	10.7	9891	0.438
2080/L	12.5	14.7	22.2	19.8	16.9	7.9	4.1	1.9	8045	0.568
2080/H	1.1	5.5	9.8	17.1	20	17.2	13.9	15.4	9816	0.536
2090/L	16.1	18	18.5	18.3	13.3	9.4	3.6	2.8	7888	0.641
2090/H	3.9	7.5	11.2	17.5	16.7	15.1	9.3	18.8	9638	0.647
2100/L	22	17.7	16.6	17.6	9.9	9	3.6	3.6	7652	0.716
2100/H	6.5	10.4	12.3	15.4	15.9	11.9	10.1	17.5	9328	0.734

* Relative Interdecile Range (RIDR) is measured as the difference between the ninth decile and the first decile divided by the median.

One disadvantage of this very simplified example is that the forecasts with jump-off dates in 2000 and 2010 are not exactly comparable. The vital rate paths used in the 2000 forecasts all start at their observed values, while the paths in forecasts that have the 2010 jump-off date have a distribution of starting values. One way of testing the plausibility of this example is to consider the uncertainty of forecasts of various durations based on a jump-off date of 2000 and a jump-off date of 2010. Holding duration constant, the example would be questionable if the uncertainties of N year ahead forecasts were very different depending on whether they were made in 2000 or 2010. When the jump-off date is the year 2000 and a forecast is made for 10 years into the future, the uncertainty measure in 2010 is 0.062, which can be read off the row in Panel A labeled 2010. In the case of a forecast made 10 years ahead based on being below the median in 2010, the uncertainty measure in 2020 is 0.088. This can be read off the row in Panel B labeled 2020/L.

The uncertainty measures for 10- through 90-year ahead forecasts based on 2000 and the two sub-samples from 2010 are shown in Figure 3. The results from the two 2010 groups track those from 2000 quite well, but are always slightly higher than the uncertainty measures based on 2000. Figure 3 is what is expected given the construction of the example and it suggests that it is plausible to proceed.

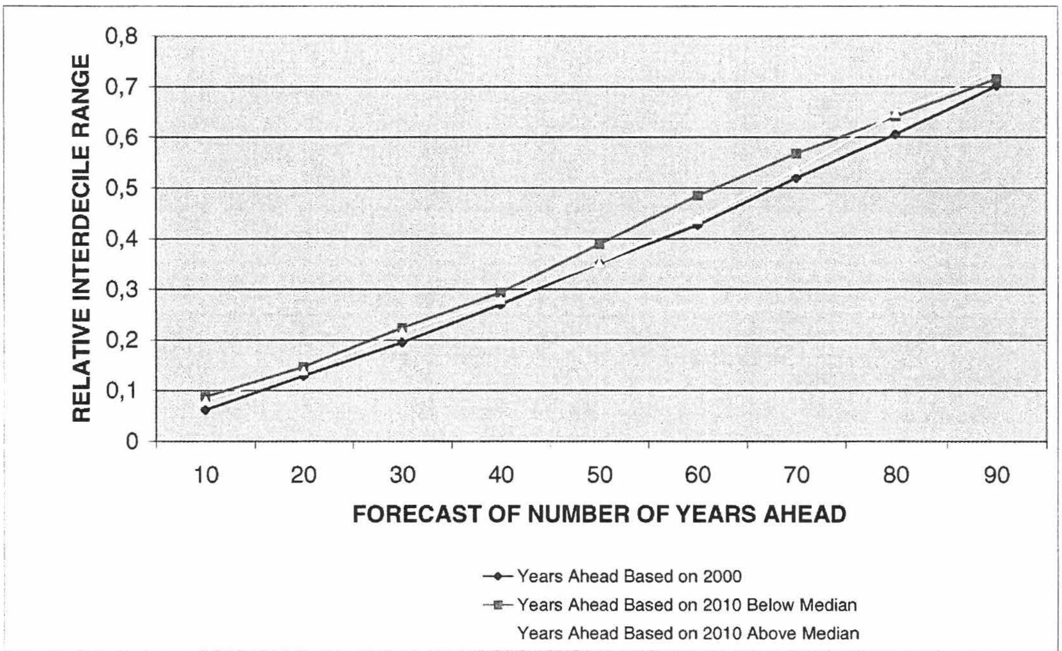


Figure 3. *Comparison of the Relative Interdecile Ranges (RIDR) for forecasts made for 10 through 90 years ahead starting from 2000 and starting from an observation either above the median or below the median in 2010. Source: Authors' calculations.*

The median population forecast for 2100 based on information up to 2000 is 8.414 billion people. After a 10 year wait, the median forecast for 2100, based on being above or below the median in 2010, would either be 7.652 billion or 9.328 billion. It would seem that if anyone were to predict 914 million more people in the world in 2100 from the perspective of 2010, the forecaster must have made a big mistake in 2000. Yet, this could well happen even if the methodology is probabilistically correct. A prediction of a 2100 population size in 2010 that is 762 million smaller than the one that was predicted in 2000 is also easily possible. These are substantial differences. Clearly, forecasts of the future will be different in 2010 than they are today. One interesting feature of probabilistic forecasting is that it can give us some idea about how much different future forecasts could be from current ones, and with what likelihood.

Population size in 2010 has such a persistent effect because of a number of factors. First, past population size influences future population size. Paths that yield large populations in 2010 will also yield large populations in 2100, even if population growth rates after 2010 are the same. Second, some populations are large in 2010 because they had high fertility rates. These high fertility rates alter the age structure of the population making it younger. Younger populations tend to grow more, other things being equal, a process that demographers call "population momentum". Fertility and mortality themselves have persistence built into them. The persistence of fertility and mortality means that on paths where fertility was high and mortality was low, leading to relatively large populations in 2010, they are likely to remain high and low respectively for a while. The effects of the persistence of fertility and mortality over time are compounded by the relatively high interregional correlations of fertility and mortality, by the persistence caused by population momentum and by the size effect itself.

It is crucially important that attention be given not only to the effects of the passage of time on the median forecast, but to the entire distribution of forecasted population sizes. Most of the differences in the distributions based on the paths above and below the median in 2010 are in the extremes (tails) of the distributions. For example, 6.5 percent of the paths that were above the median in 2010 resulted in populations of less than 6 billion in 2100, compared to 22.0 percent of the paths that were below the median in 2010. The difference at the high end of the distribution is even more striking. Over 17 percent of the paths that were above the median in 2010 ended the century with 12 billion people or more. In contrast only 3.6 percent of the paths that were below the median in 2010 did so.

This has been a very short and simplified presentation of the basic concepts of conditional probabilistic forecasting with future jump-off dates. It is meant only to be suggestive. This analysis of learning with the passage of time has illustrated how sensitive the long-term population outlook is to near-term trends. It can also help to understand why projections of population size in 2100 have changed so significantly over the past 10 years. We have simply learned a great deal over the past decade. During this decade population growth has been lower than originally expected and this has significantly decreased our new long-term expectations.

5 Concluding Thoughts

Conditional probabilistic projections represent a way to combine the benefits of probabilistic projections, particularly the quantification of uncertainty, with the benefits of alternative scenarios, which give clear indications of the sensitivity of results to underlying assumptions. We have shown that the same kinds of conclusions about, for example, the relative importance of fertility and mortality trends to population size outcomes can be drawn using conditional probabilistic forecasts as can be drawn using alternative deterministic scenarios. An added benefit is that the conditional probabilistic forecasts provide an estimate of the likelihood of the underlying demographic conditions, as well as an estimate of their effect on outcomes. These projections can be extremely useful to both the research and policy communities. For instance, many analyses of the potential for long-term environmental change are based on the approach of considering a set of alternative future scenarios conditional on different sets of assumptions about future development. The scenario approach dominates as a response to deep uncertainty about the many socio-economic, technological, and environmental factors that must be included in such analyses. Conditional probabilistic projections present a possible means of retaining some of the advantages of the probabilistic approach without discarding the benefits of conditional scenarios (see O'Neill in this issue for an example (O'Neill, 2004)).

Probabilistic projections with future jump-off dates, which are conditional on how population characteristics evolve between now and the future jump-off date, present a new way to address an important set of research questions that are also policy relevant. They provide a means of anticipating how our forecasts might change in the future and how likely those changes appear to be at the moment. These kinds of analyses cannot be done deterministically. As there can be costs and benefits to changes in the outlook for the future, these projections could have interesting new applications. For example in the climate change issue, the prospects of learning about technological costs, or about physical aspects of the climate system, have been incorporated into analyses of whether it is better to act now or to wait to learn more. However as far as we are aware, no such analysis—for climate change or any other issue—has been performed taking into account the prospects for learning about the outlook for population.

References

Alho, J. (1990). Stochastic methods in population forecasting. *International Journal of Forecasting*, **6**(4), 521–530.

Alho, J. (1997). Scenarios, uncertainty and conditional forecasts of the world population. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, **160**(1), 71–85.

Alho, J. & Spencer, B. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, **80**(390), 306–314.

Keilman, N., Pham, D. & Hetland, A. (2002). Why population forecasts should be probabilistic—illustrated by the case of Norway. *Demographic Research*, **6**(15), 408–454. http://www.demographic-research.org/

Lee, R.D. (1999). Probabilistic approaches to population forecasting. In *Frontiers of Population Forecasting*, A supplement to vol. **24**, 1998, *Population and Development Review*, Eds. W. Lutz, J.W. Vaupel and D.A. Ahlburg, pp. 156–190. New York: The Population Council.

Lee, R. & Tuljapurkar, S. (1994). Stochastic population projections for the United States: Beyond high, medium and low. *Journal of the American Statistical Association*, **89**(428), 1175–1189.

Lutz, W. (1995). Scenario Analysis in Population Projection. Working Paper WP-95-57. Laxenburg, Austria: IIASA.

Lutz, W., Prinz, C. & Langgassner, J. (1994). The IIASA world population scenarios to 2030. In *The Future Population of the World: What Can We Assume Today?*, Ed. W. Lutz, pp. 391–422. London: Earthscan.

Lutz, W., Sanderson, W. & Scherbov, S. (1996). Probabilistic population projections based on expert opinion. In *The Future Population of the World: What Can We Assume Today?*, Ed. W. Lutz, pp. 397–428, Revised edition. London: Earthscan.

Lutz, W., Sanderson, W. & Scherbov, S. (1997). Doubling of world population unlikely. *Nature*, **387**(6635), 803–805.

Lutz, W., Sanderson, W. & Scherbov, S. (2001). The end of world population growth. *Nature*, **412**, 543–545.

Lutz, W. & Scherbov, S. (1998). An expert-based framework for probabilistic national population projections: The example of Austria. *European Journal of Population*, **14**, 1–14.

Lutz, W. & Scherbov, S. (2002). Can immigration compensate for Europe's low fertility. Interim Report, IR-02-052. Laxenburg, Austria: IIASA.

O'Neill, B.C. (2004). Conditional population projections: An application to climate change. *International Statistical Review*, **72**, 167–184.

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

Sanderson, W.C., Scherbov, S., Lutz, W. & O'Neill, B.C. (2004). Applications of probabilistic population forecasting. In *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*, Eds. W. Lutz and W.C. Sanderson. London: Earthscan.

Webster, M. (2002). The curious role of "learning" in climate policy: Should we wait for more data? *The Energy Journal*, **23**(2). 97–119.

Résumé

Puisque les décideurs politiques préfèrent souvent penser en terme de scénarios alternatifs, la question se pose de savoir s'il est possible de faire des prévisions démographiques conditionnelles dans un contexte probabiliste. Cet article montre à la fois le bien-fondé et l'utilité de telles prévisions, et cela à travers deux types d'exemples. Le premier est l'analogue probabiliste de l'analyse par scénarios déterministes. L'analyse conditionnelle de scénarios probabilistes est essentielle pour les décideurs politiques parce qu'elle leur permet de répondre convenablement aux questions de type "que se passera t'il si" en cas de résultats incertains. Le second exemple est une nouvelle catégorie que nous appelons "future jump-off date forecasts" (prévisions démographiques dans laquelle l'année de base est déjà dans le futur). Ces dernières sont essentielles parce qu'elles révèlent aux décideurs politiques avec quelle probabilité les caractéristiques cruciales des prévisions démographiques d'aujourd'hui seront aussi actuelles pour les prévisions faites à l'avenir.

[*Received May 2003, accepted January 2004*]

International Statistical Review (2004), **72**, 2, 167–184, Printed in The Netherlands

Conditional Probabilistic Population Projections: An Application to Climate Change

Brian C. O'Neill

International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria and Watson Institute for International Studies, Brown University, Providence, RI, 02912, USA
E-mail: oneill@iiasa.ac.at

Summary

Future changes in population size, composition, and spatial distribution are key factors in the analysis of climate change, and their future evolution is highly uncertain. In climate change analyses, population uncertainty has traditionally been accounted for by using alternative scenarios spanning a range of outcomes. This paper illustrates how conditional probabilistic projections offer a means of combining probabilistic approaches with the scenario-based approach typically employed in the development of greenhouse gas emissions projections. The illustration combines a set of emissions scenarios developed by the Intergovernmental Panel on Climate Change (IPCC) with existing probabilistic population projections from IIASA. Results demonstrate that conditional probabilistic projections have the potential to account more fully for uncertainty in emissions within conditional storylines about future development patterns, to provide a context for judging the consistency of individual scenarios with a given storyline, and to provide insight into relative likelihoods across storylines, at least from a demographic perspective. They may also serve as a step toward more comprehensive quantification of uncertainty in emissions projections.

Key words: Population; Projection; Uncertainty; Scenario; Climate change.

1 Introduction

The threat of human-induced climate change, popularly known as global warming, presents a difficult challenge to society over the coming decades (Watson *et al.*, 2001). The production of so-called "greenhouse gases" as a result of human activity, mainly due to energy production through the burning of fossil fuels such as coal, oil, and natural gas, is expected to lead to a generalized warming of the Earth's surface, rising sea levels, and changes in precipitation patterns. The potential impacts of these changes are many and varied—more frequent and intense heat waves, changes in the frequency of droughts and floods, increased coastal flooding, and more damaging storm surges—all with attendant consequences for human health, agriculture, economic activity, biodiversity, and ecosystem functioning.

Future changes in population size, composition, and spatial distribution are key factors in the analysis of climate change (O'Neill *et al.*, 2001). Demographics enter the problem in a number of ways: as one determinant (among many others) of emissions of greenhouse gases, and therefore of how much climate may change in the future; as a determinant of the impacts of climate change, and therefore how serious any given level of climate change might be; and potentially as a factor in political arrangements for addressing climate change (as, for example, in emissions reduction schemes based on per capita emissions levels). It is not just population size that matters: age structure, urbanization, and composition by household size and type are all of potential importance to these different aspects of the climate issue, as are population characteristics that are less demographic in the traditional sense but are often analyzed by demographers: educational status, health status, and poverty.

As demographers are well aware, there is great uncertainty in projecting any of these factors into the long-term future. However, the population component of the climate problem is in no way unusual in this respect. Uncertainties can be even larger in other factors, such as future labor productivity, rates of technological progress and of per capita GDP growth, consumption patterns, the response of the climate system to greenhouse gas emissions, and impacts of climate change on ecosystems and societies.

Recently, growing attention has been focused on methods of accounting for uncertainty in various components of the climate problem (see, e.g., Webster *et al.*, 2002; Wigley & Raper, 2001; Moss & Schneider, 2000; Lempert *et al.*, 2000; Nordhaus & Popp, 1997). However, no study has devoted any specific attention to quantification of population uncertainty and its relevance to integrated assessments of climate change, despite the existence of fully probabilistic, global population projections (Lutz *et al.*, 2001).

This paper proposes a new approach to accounting for demographic uncertainty within projections of greenhouse gas emissions. Accounting for uncertainty in emissions is important for at least two reasons. First, emissions projections play a key role in estimating what the consequences of climate change might be if no action were taken in response to it. Such estimates essentially define climate change as a problem to the extent that serious consequences seem possible. Emissions are the first step in this "cascade of uncertainty" (Moss & Schneider, 2000) from emissions, to climate change, to impacts, and therefore play a key role. Second, analyses of the costs of achieving given emissions reduction targets are sensitive to the so-called reference emissions path: the assumed level of emissions over time in the absence of policy. Reaching a given target in an otherwise high-emissions world would be much more difficult than in a world in which emissions would turn out to be quite low anyway.

To date, there have been two main approaches to accounting for uncertainty in emissions: alternative scenarios, and fully probabilistic projections. Here approach is proposed that combines elements of both: conditional probabilistic projections. Section 2 describes current approaches in more detail, and the rationale for the conditional probabilistic approach. Following that (section 3) the methodology is discussed and applied to a well-known set of emissions scenarios. Results are presented (section 4), and the paper concludes with a short discussion (section 5).

2 Uncertainty and Greenhouse Gas Emissions

Alternative Scenarios

Currently, the dominant approach to accounting for uncertainty in future greenhouse gas emissions is the development and analysis of alternative scenarios. The scenario approach to uncertainty has a history of at least 50 years, beginning with post-World War II military planning, extending to business strategy development for major corporations, and more recently to planning for sustainable development (Schwartz, 1991). Scenarios are stories about the way the future might unfold. They can be qualitative, quantitative, or both, and they can be used in a variety of ways, including for educating participants in the scenario development process about the issues at hand, or for communicating key insights to the intended audience. Scenarios tend to be used as an approach to uncertainty when problems are complex and uncertainties are very large, precluding meaningful estimates of the likelihood of various future outcomes. In decision analysis, scenarios are often used as a basis for searching for "robust strategies", i.e. options for responding to a problem that are likely to work reasonably well regardless of how the future unfolds (e.g., Lempert *et al.*, 2000).

The Intergovernmental Panel on Climate Change (IPCC) used the scenario approach to develop a plausible range of climate change consequences assuming no additional climate policy measures are taken in the future. Based on a set of 40 different emissions scenarios (detailed in the IPCC Special Report on Emissions Scenarios (SRES), Nakicenovic *et al.*, 2000), the IPCC estimates that the increase in global average temperature over the next 100 years could range from 1.4°C to 5.8°C (Cubasch *et al.*, 2001). The SRES emissions scenarios are not probabilistic. Instead, the rationale was to explore a wide range of plausible futures, each "equally sound" (Nakicenovic *et al.*, 2000) but with no judgment made as to their relative likelihood.

The SRES scenario development process consisted of three main steps. First, four different qualitative "storylines" were developed, describing broad socio-economic and technological development patterns that could unfold over the 21st century. While the SRES storylines are multidimensional, they can usefully be distinguished along two dimensions, "globalization" and "sustainability". Globalization refers to "the extent of economic convergence and social and cultural interactions across the regions" and sustainability to "the balance between economic objectives and environmental and equity objectives" (Nakicenovic *et al.*, 2000, Section 4.2.1). The storylines were called A1, A2, B1, and B2, with no significance given to this ordering; 1 vs. 2 distinguishes the degree of globalization vs. regionalization, and the A vs. B distinction is based on the degree of economic vs. environmental orientation. Brief summaries are included in Table 1 to make the discussion more concrete.

Table 1

Brief summaries of the SRES storylines, from Nakicenovic et al. *(2000, Section 4.2.1).*

Storyline	Summary
A1	"The A1 storyline and scenario family describes a future world of very rapid economic growth, low population growth, and the rapid introduction of new and more efficient technologies. Major underlying themes are convergence among regions, capacity building, and increased cultural and social interactions, with a substantial reduction in regional differences in per capita income. The A1 scenario family develops into four groups that describe alternative directions of technological change in the energy system."
A2	"The A2 storyline and scenario family describes a very heterogeneous world. The underlying theme is self-reliance and preservation of local identities. Fertility patterns across regions converge very slowly, which results in high population growth. Economic development is primarily regionally oriented and per capita economic growth and technological change are more fragmented and slower than in other storylines."
B1	"The B1 storyline and scenario family describes a convergent world with the same low population growth as in the A1 storyline, but with rapid changes in economic structures toward a service and information economy, with reductions in material intensity, and the introduction of clean and resource-efficient technologies. The emphasis is on global solutions to economic, social, and environmental sustainability, including improved equity, but without additional climate initiatives."
B2	"The B2 storyline and scenario family describes a world in which the emphasis is on local solutions to economic, social, and environmental sustainability. It is a world with moderate population growth, intermediate levels of economic development, and less rapid and more diverse technological change than in the B1 and A1 storylines. While the scenario is also oriented toward environmental protection and social equity, it focuses on local and regional levels."

Next, particular quantitative paths for fundamental driving forces of emissions, including population and gross domestic product (GDP), were selected that were judged to be consistent with each storyline. Finally, six different modeling teams produced quantitative interpretations of the storylines, using the quantitative paths for driving forces as inputs, resulting in 40 different scenarios for energy use, land use, and associated greenhouse gas emissions over the next 100 years. Figure 1 shows the range of CO_2 emissions resulting from the SRES scenarios, categorized by storyline. The range is very wide, covering a possible increase in emissions of CO_2 by a factor of 5 over the next 100 years, as well as a decline to emissions below today's level.

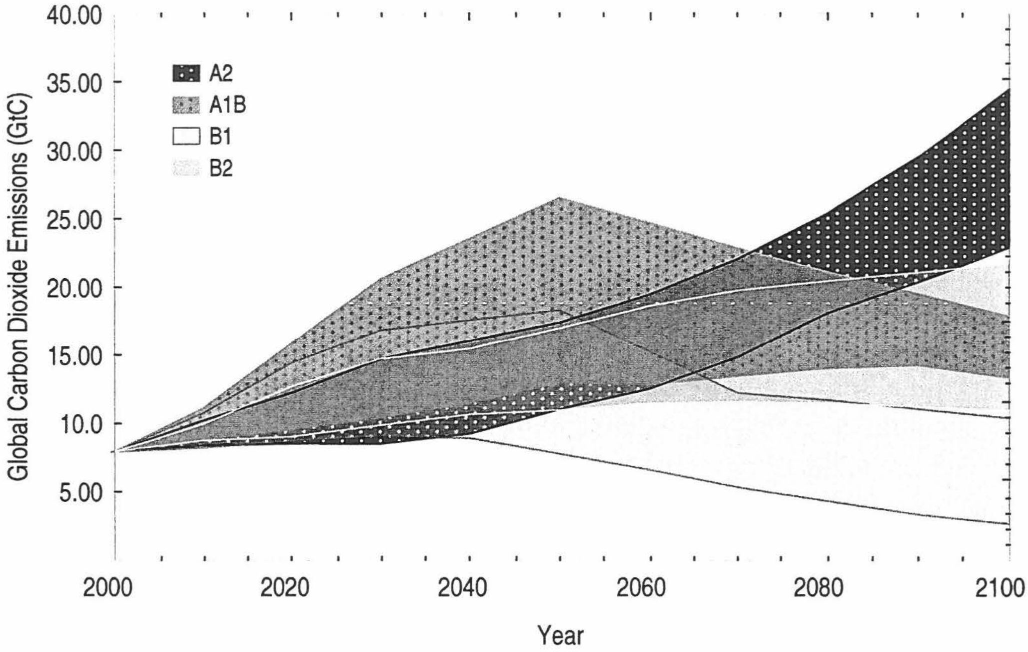


Figure 1. *Range of global CO_2 emissions covered by each SRES storyline. Several variants of the A1 storyline were produced; only one (A1B) is shown.*
Source: Based on data in Nakicenovic et al., *2000.*

Probabilistic Emissions Projections

Several studies have produced fully probabilistic emissions projections. Early work by Nordhaus & Yohe (1983) and Edmonds *et al.* (1986) developed probabilistic versions of energy-economic growth models to calculate uncertainty distributions for optimal emissions, given uncertainty in a long list of parameters, including those affecting population growth, labor productivity, the energy system (e.g., elasticities of substitution between fuels), and the carbon cycle. These projections were the first to systematically explore uncertainty in long-term emissions projections, and were useful in estimating how wide the plausible range of outcomes actually was (Keepin, 1986).

More recently, Wigley & Raper (2001) made their own assumptions about the relative likelihood of the SRES scenarios. They assign each SRES scenario an equal probability as part of an effort to produce probabilistic projections of global average temperature change (see New & Hulme (2000) for a similar approach and Dessai & Hulme (2001) for an approach that treats storylines independently). In contrast, Webster *et al.* (2002) take an approach independent of SRES. They use an expert opinion-based estimate of uncertainty distributions for key parameters of an energy-economic growth model, which they then use to generate an uncertainty distribution for future emissions and climate change.

There are a number of points of view on the desirability, and credibility, of assigning subjective relative probabilities to alternative scenarios of future emissions. Schneider (2001, 2002) argues for developing subjective probability estimates based on expert opinion, since the judgment of experts is preferable, in his view, to the plethora of judgments that different users and policy makers will make in the absence of guidance on this point. Webster *et al.* (2002) amplify this view, describing the question as not one of *whether* probabilities will be assigned to scenarios, but *when* and *by whom*. Since methods have been developed to control known biases in expert judgment, it is better that probabilities be assigned by experts in decision analytic techniques than by users. In contrast, Grübler & Nakicenovic (2001) argue that assigning subjective probabilities to emissions scenarios would be inappropriate and infeasible. Probability distributions for conditions such as population, economic growth, technological development, agricultural practices, diets, etc. (all determinants of emissions) would themselves be insufficient; probabilities for ways in which these conditions would interact would also be necessary, producing an unmanageable explosion of uncertainty, in their judgment. Attempts to assign subjective probabilities in the face of such complications risks "dismissal of uncertainty in favour of spuriously constructed 'expert' opinion".

Conditional Probabilistic Emissions Projections: Rationale

One route to managing the complexity of the problem is to preserve the scenario approach to uncertainty when considering possible broad patterns of development in the future (i.e., storylines), but use subjective probabilities in quantifying uncertainty in the consequences of those development patterns for emissions. The basic rationale for such conditional probabilistic projections is that more meaningful judgments can be made about the likelihood of future trends given a particular development path, than can be made about the relative likelihood of different development paths actually occurring. Stated in terms of the SRES scenarios, one may be able to assign more meaningful probabilities to emissions outcomes conditional on, say, the A2 world coming to pass, than one could to the relative likelihood of the A1, A2, B1 and B2 storylines occurring in the first place.

Intuitively, one expects this proposition would hold, because by reducing the state space that must be considered, the conditional approach makes uncertainties more manageable. For example, it may be judged difficult, if not impossible, to assign meaningful probabilities to the rate of per capita economic growth over the coming decades. One reason may be that it is difficult to judge how globalization is likely to proceed—whether trade barriers will become more or less restrictive, for example—which will have important consequences for economic growth. However, if one restricts the set of possible futures that must be considered to only those in which globalization proceeds rapidly and trade barriers are reduced, confidence in judgments of the probability distribution of future economic growth rates, conditional on that storyline, may increase. This type of reasoning can be extended to many different factors that are included in storylines of the future. In fact, it would seem that as long as the conditions placed on the future (through storylines) are relevant to the outcomes of interest, then it should be possible to make better judgments of conditional probabilities of outcomes than of unconditional ones.

The following sections illustrate the possibility of developing conditional probabilistic emissions scenarios, using the IPCC SRES scenarios as a basis. First probabilistic population projections conditional on the storylines used in the SRES scenarios are developed. Next, they are combined through simple linear scaling with per capita emissions rates derived from the SRES scenarios. While this latter step is an oversimplification of the relationships between population and other factors affecting emissions, it serves to illustrate the kind of emissions results one might obtain from a fuller accounting for population uncertainty.

The results of developing conditional probabilistic projections both for population and for emissions can be useful for several reasons:

1. They can be used as a means of judging the likelihood of a single existing scenario (of population or of emissions) given a particular storyline;
2. They can be used to develop a single new scenario that has a probabilistic interpretation for use in scenario analysis (e.g., O'Neill *et al.*, 2003);
3. They can provide a fuller accounting for uncertainty (in population or in emissions) within storylines;
4. They might provide some insight into the relative likelihood of outcomes across storylines (discussed further below);
5. They may serve as a step toward more comprehensive probabilistic projections in the future.

3 Methodology

Conditional probabilistic population forecasts were first discussed by Alho (1997), who began with known unconditional and conditional projections, and evaluated the plausibility that the conditional projections could be derived from the unconditional ones. Sanderson *et al.* (2004, see this issue) take a different approach, beginning with unconditional projections and deriving conditional forecasts that demonstrate, within a probabilistic framework, the sensitivity of outcomes to key variables. Lutz & Scherbov (2002) use the conditional probabilistic approach in a similar manner to compare the sensitivity of the European population to changes in immigration and fertility. These studies demonstrate that sensitivity analysis, traditionally carried out within a deterministic framework, is also possible within a probabilistic framework. This paper builds on that approach by applying the concept of conditional probabilistic projections to emissions scenario development. The necessary steps are:

1. Specify qualitative descriptions of future trends in demographic rates (fertility, mortality, migration) judged to be consistent with a given storyline;
2. Define quantitative probability distributions of demographic rates consistent with the qualitative descriptions;
3. Generate a probability distribution of population outcomes conditional on the distributions of demographic rates;
4. Generate a probability distribution of emissions based on the distribution of population outcomes.

Table 2 provides a summary of the qualitative fertility, mortality and migration assumptions made by the SRES authors for each storyline for the industrialized country regions (IND) and the developing country regions (DEV). The "high", "medium", and "low" descriptions are interpreted as relative to the overall outlook within each region (i.e., high fertility in the IND region means the high end of the plausible range for that region, but may in fact be lower than low fertility paths for the DEV region, which occupy the low end of the plausible range for the DEV region). Based on descriptions in the SRES report (primarily in sections 4.3 and 4.4), a summary of the reasoning for these choices is as follows (it would be possible to offer alternative interpretations of the demographic rates most likely to be associated with each storyline, but it is considered outside the scope of this paper):

- In the A1 storyline, rapid economic development, associated with improved education and reduced income disparities, is assumed to drive a relatively rapid fertility decline in the high fertility regions. Fertility is generally below replacement level in the long run. Fertility in industrialized regions is assumed to follow a medium path at least in part so that, relative to the developing regions, the scenario "is consistent with the assumption of convergence of social and economic development" (p. 193); i.e., the assumption is that social and economic convergence will lead to demographic convergence as well. For mortality, it is assumed that the conditions leading to low fertility are also consistent with relatively low mortality, so mortality is assumed to be low in all regions. No explicit discussion of migration is provided, although the projection eventually adopted assumes medium migration levels.
- In the A2 storyline, the regional orientation and slower rate of economic growth, limited flow of people and ideas across regions, and orientation toward family and community values was judged to be consistent with a relatively high fertility in all world regions. Mortality was assumed to be high as well, based on the assumption that conditions leading to high fertility would also lead to relatively high mortality in all regions. Although the storyline describes a limited flow of people across regions, the medium migration flows were chosen, as in all other storylines.

- The B1 storyline shares the same population projection as the A1 storyline, although for somewhat different reasons. Rapid social development, particularly for women, and an emphasis on education drives a relatively rapid decline in fertility in developing country regions (as opposed to the A1 storyline, in which economic development is seen as the main driver). Reasoning for fertility in industrialized countries, and for mortality and migration assumptions, are the same as in A1.
- In the B2 storyline, economic development is moderate, particularly in the developing country regions. However education and welfare programs are pursued widely and local inequity is reduced through strong community support networks. The mix of moderate economic development, and strong but heterogeneous social development results in an assumption of medium fertility and mortality paths. Migration is again assumed to be medium, with no explicit discussion of this choice.

For each storyline, the SRES authors selected one population projection from the literature that was judged to be consistent with these assumptions (see Table 2). A similar approach (selection or creation of a single scenario within storylines) was taken for quantifying GDP and rates of technological progress within storylines.

The next step in developing conditional probabilistic projections (which will substitute for the single scenarios used in SRES) is to express the qualitative demographic assumptions associated with each storyline as quantitative uncertainty distributions. These conditional distributions are derived from the global probabilistic projection from IIASA (Lutz *et al.*, 2001). The methodology for generating the IIASA projection consists of a combination of expert- and argument-based judgment on future fertility, mortality, and migration; a time series approach to generating individual trajectories for each rate; and consideration of the magnitude of errors in past projections to guard against underestimation of uncertainty. The final product is a distribution of 2000 equally likely simulations at the level of 13 world regions. The conditional probabilistic projections are generated by selecting from this complete (unconditional) distribution a subset of simulations consistent with the SRES assumptions as expressed in Table 2.

Table 2

Demographic assumptions in SRES storylines (IND=industrialized country regions; DEV=developing country regions).

Storyline	Fertility	Mortality	Migration	Projection Source
A1/B1	IND: medium DEV: low	IND: low DEV: low	IND: medium DEV: medium	Lutz *et al.*, 1996
A2	IND: high DEV: high	IND: high DEV: high	IND: medium DEV: medium	Lutz *et al.*, 1996
B2	IND: medium DEV: medium	IND: medium DEV: medium	IND: medium DEV: medium	UN, 1998

In principle, the "high", "medium", and "low" categories for demographic rates would be best defined by the scenario-builders themselves, because they may have had specific ideas in mind that are not clearly stated in the storyline descriptions. For example, was "high" fertility in the A2 world intended to imply a rather extreme case (perhaps near the 95th percentile of possible outcomes), or simply a fertility rate that is moderately higher than an unconditional medium or "best guess" outcome? Should fertility always be high in this scenario, or should it be high on average, with periods of medium or low fertility also possible? Should it be high in all regions simultaneously, or high on average across all regions? In the absence of specific guidance in the SRES report, this paper takes as a starting point that high, medium, and low should represent ranges of demographic rates that are equally likely and are mutually exclusive. That is, in each region and for each variable representing one of the demographic components (total fertility rate, life expectancy, and net migration) high, medium, and low should be defined as those ranges covering the top, middle, and bottom third, respectively, of all outcomes in the IIASA unconditional probabilistic projections.

To carry out the selection process, an index of each demographic rate must be adopted, and an appropriate level of regional aggregation chosen. As an index, the average values of each variable over the period 2000–2050 were used, since several essential elements of the reasoning behind the choice of population scenario in SRES apply to the currently high fertility regions (e.g., relationships between fertility and economic and social conditions, and the correlation between fertility and mortality), and these regions will have above replacement fertility mainly in the first half of the century. For regional disaggregation, average values for two world regions, Industrialized (IND) and Developing (DEV) were considered, which matches the level of detail for which qualitative storyline elements relevant to demographics are provided in SRES. Averages are calculated from underlying rates in the IIASA projections. Fertility, life expectancy, and net migration for the IND and DEV regions are specified as the population-weighted average of these variables across the subregions. For fertility, the population aged 15–65 is used, the most relevant age group available. For net migration, the absolute values of the subregional levels are used.

The selection process therefore consists of producing an average fertility, mortality, and migration rate for each of 1,000 simulations in the unconditional projection, with averaging performed over the period 2000–2050 and over two groups of regions. Simulations are then ranked according to each demographic rate in each region, and assigned to high, medium, and low categories by dividing each ranking into three sets of outcomes containing the first 333, next 334, and last 333 simulations. Those simulations matching the assumptions in Table 2 for all three rates for each SRES storyline become part of the conditional probabilistic projection for that storyline. This approach produces what might be called input-constrained conditional probabilistic projections, since the conditions (storylines) impose constraints on the demographic rates that serve as inputs to the projection. (It is also possible to develop output-constrained projections, in which constraints on population size or age structure are used to select individual simulations, a type of inverse approach.)

Before examining results, it is important to stress that alternative approaches to the selection process are possible. For example, the definition of the high, medium, and low categories could be based on intervals of absolute values of demographic rates rather than on the assumption that each category should be equally likely. This strategy is not pursued here because there is no obvious basis for defining the intervals of absolute values that would clearly recommend it. Additionally, category definitions need not be derived by simply dividing the unconditional distributions into mutually exclusive terciles. This division is clearly a simplification that gives all simulations within a tercile a weight of 1, and all outside the tercile a weight of zero. One could consider a definition in which a more smoothly varying distribution of weights was applied to the unconditional distribution. Furthermore, one could choose a different metric than the 50-year average value used here. For example, one might choose to extend the averaging period to 100 years, or to use consecutive or overlapping 50-year periods to extend the constraints over the whole century (O'Neill, submitted). These choices have implications for trends in the medians and for the variance of the conditional distributions of the underlying rates. Finally, one could choose different levels of regional aggregation to which the criteria are applied. Applying criteria to all 13 IIASA regions simultaneously, versus applying them to averages across several regions, also affects the variance of the conditional distributions of demographic rates. The methodological choices made in this paper are intended to match the level of specificity provided in the SRES storylines, and to remain as simple as possible.

4 Results

First we discuss the results for the population outcomes, and then for emissions.

Conditional Probabilistic Population Projections

Figures 2a–c show the results of the process of selection of population distributions, for the world and two regions for the A2 storyline. Figure 2a shows the individual simulations underlying the conditional projection for the world, along with the median and 60% and 95% uncertainty intervals. For comparison, the single population path used in the SRES A2 storyline is shown as a thick gray line. Two main conclusions emerge. First, the probabilistic projections indicate substantial scope for uncertainty in population size conditional on the A2 storyline. The 60% uncertainty interval for global population in 2100 is 8.5 to 12 billion, and the 95% uncertainty interval is 7 to 15 billion. Second, the SRES population assumption is high relative to the conditional probabilistic projection. Early in the century it falls at about the upper limit of the 60% uncertainty interval, and by the end of the century is at 15 billion, the upper limit of the 95% uncertainty interval and about 4 billion higher than the median.

The systematic downward bias in the probabilistic projections relative to the SRES assumptions for A2 is due primarily to changes in the outlook for fertility and mortality in different regions of the world. The population projections used in SRES were developed in 1996, while the conditional distributions are based on projections made in 2001. Figures 2b and 2c show that the DEV region mirrors the global result, with the SRES population size reaching the upper bound of the 95% uncertainty interval by the end of the century, while the IND region assumption is high but more moderately so, falling just above the 60% uncertainty interval. More detailed regional analysis shows that difference between the SRES population assumptions and the conditional projections vary across smaller regions. Regions showing the largest differences include the Reforming Economies region (mainly the Former Soviet Union) and the Centrally Planned Asia region (mainly China) by the end of the century.

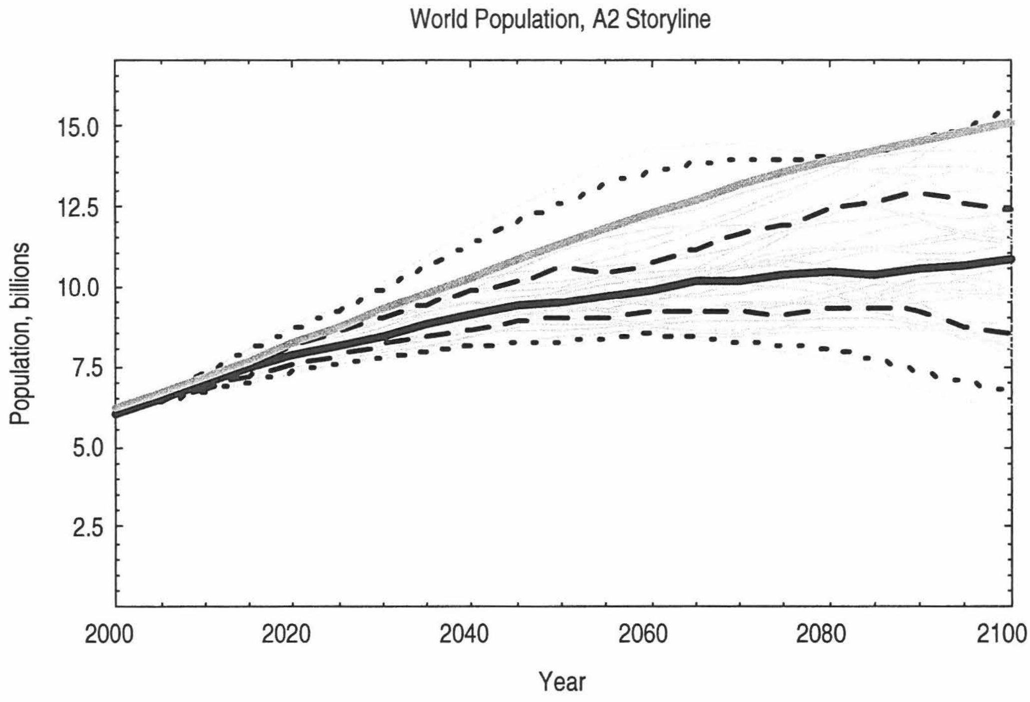
Figure 3 summarizes the results for all storylines for the years 2050 and 2100 for the world and the two regions. It shows that the uncertainty range in the conditional projections is largest in the A2 storyline (when fertility is assumed to be high), and the smallest in the A1/B1 storylines (when fertility is assumed to be low), although in all storylines the conditional projections show a 95% uncertainty interval of at least +/-20% in terms of absolute population size relative to the median. In the year 2050, the SRES population assumptions are relatively close to the median of the conditional projections for the A1/B1 and the B2 storylines. However by the end of the century, they appear high not only in the A2 storyline described above, but also in the DEV region in the B2 storyline (falling well above the 95% uncertainty interval), and in the IND region in the A1/B1 storyline (falling at about the upper end of the 95% uncertainty interval).

The quantitative differences between the SRES population assumptions and the conditional projections are sensitive to the definitions of high, medium, and low, as well as to the time periods and regional aggregations to which the criteria are applied, although the general patterns appear to be robust.

Relative Likelihoods Across Storylines

One intriguing possibility offered by conditional probabilistic projections is that, if they are drawn from a full unconditional projection, they should provide a basis for judging the relative likelihood across two or more conditional projections, not just the relative likelihood of outcomes within a given conditional projection. For example, if a conditional projection based on storyline X consists of twice as many (equally likely) simulations drawn from the unconditional distribution as a conditional projection based on storyline Y, then the demographic consequences of storyline X should be twice as likely as the demographic consequences of storyline Y. Thus, comparing the conditional projections developed to be consistent with the four SRES storylines should give an indication of the relative likelihoods of the SRES storylines, at least from a demographic perspective. Such demographically-

(2a)



(2b)

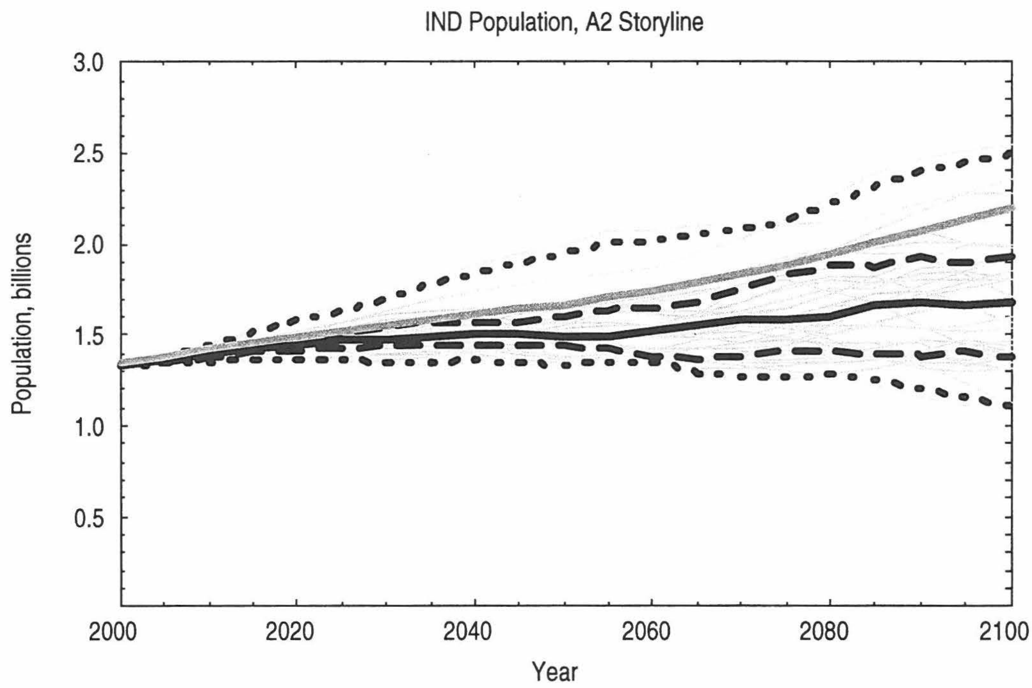


Figure 2a–2b. *Probabilistic population projections conditional on the A2 storyline. Thin gray lines: individual simulations from IIASA 2001 making up the conditional probabilistic projections. Black lines: median (solid), 60% uncertainty interval (dashed), and 95% uncertainty interval (dotted) for the IIASA 2001 conditional projections. Thick gray line: SRES A2 population assumption. Figures a–c show results for the World, industrialized country region (IND), and developing country region (DEV).*

(2c)

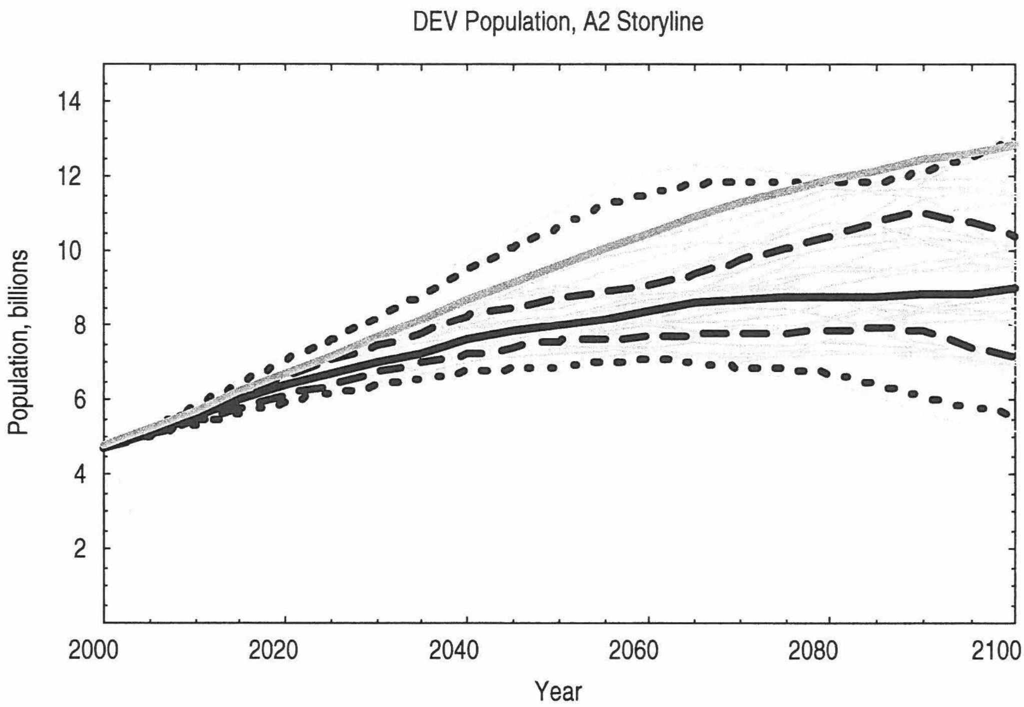


Figure 2c. *Continued.*

based relative likelihoods could be combined with perspectives on other aspects of the scenarios to inform a larger process of considering relative likelihoods of alternative emissions scenarios.

Table 3 shows, for each storyline, the number of simulations that match assumptions regarding each of the demographic rates individually, as well as combinations of rates. For example, it shows that, based on the definitions of high, medium, and low established above, there are 7 simulations (out of 1,000 total that were available for this analysis) consistent with the A1 and B1 storylines in terms of fertility, mortality, and migration at the level of two world regions. There are 32 simulations consistent with the A2 storyline, and 14 simulations consistent with the B2 storyline. Thus, given the IIASA population projections, the SRES storylines, and the category definitions used here, the demographic assumptions consistent with the B2 and especially the A1/B1 storylines appear to be substantially less likely than the assumptions underlying the A2 storyline.

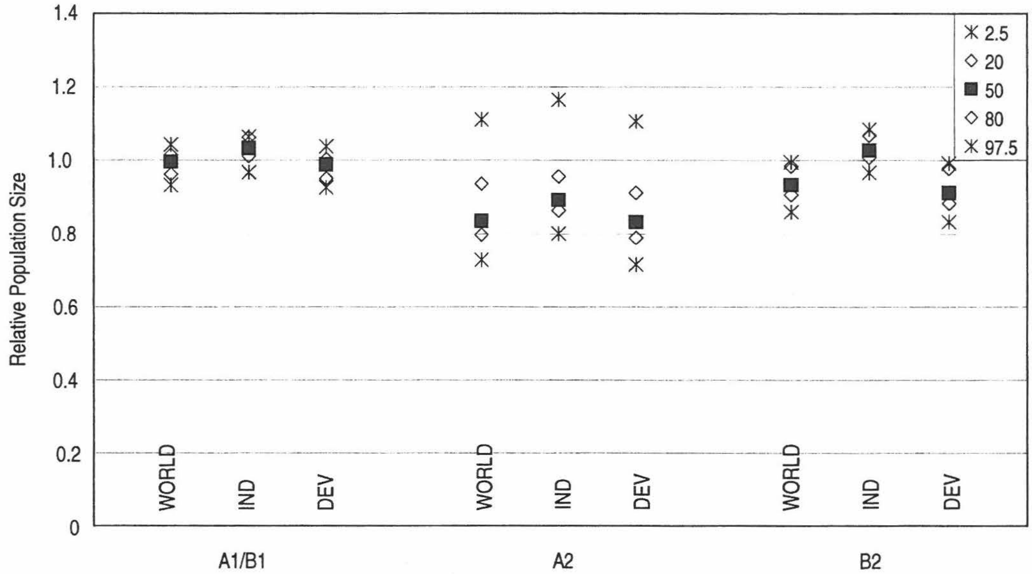
Table 3

Number of simulations meeting fertility (F), life expectancy (LE), and migration (M) criteria, as well as combinations of them, associated with the SRES storylines.

Storyline	F	LE	M	F+LE	F+LE+M
A1/B1	62	301	334	16	7
A2	268	297	334	80	32
B2	207	266	334	49	14

There are two separate reasons for this result. The A1/B1 projection is relatively unlikely primarily because it mixes medium fertility rates in the IND region with low fertility rates in the DEV region. Because the IIASA projections assume a relatively strong correlation in fertility rates across regions,

(3a)



(3b)

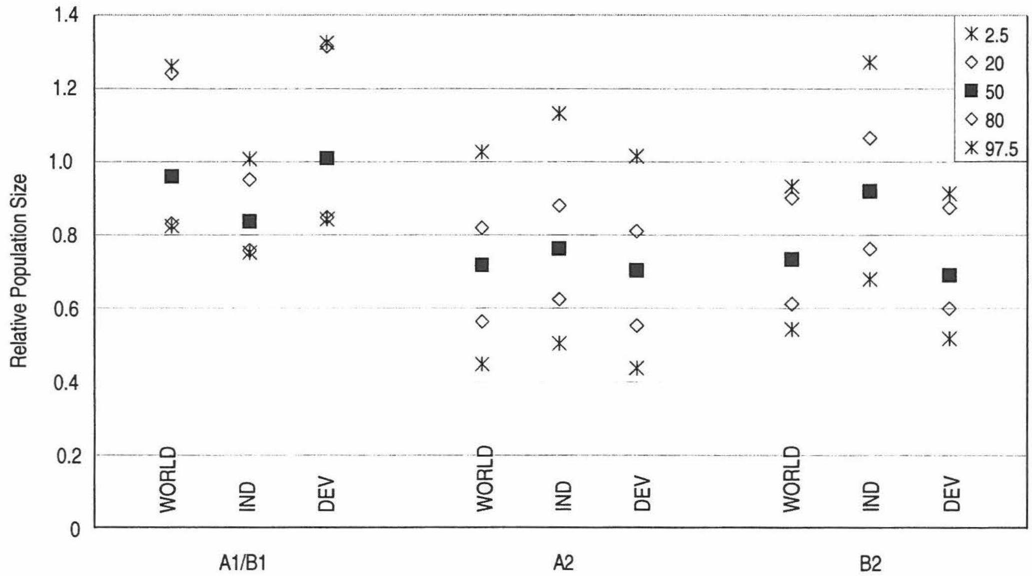


Figure 3a–b. *Population size of the median and the bounds of the 60% and 95% uncertainty intervals in conditional probabilistic projections, relative to population size assumed in SRES, for the year 2050 (a) and 2100 (b).*

this combination is less likely than the A2 assumption of high fertility in both regions. As shown in Table 3, there are only 62 simulations that meet the A1/B1 fertility assumptions, while there are 268 simulations that meet the A2 fertility assumptions.

The reason that the B2 projection is relatively unlikely is less obvious. A counterintuitive consequence of the regional correlations in demographic rates is that, even if it is equally likely that a single variable in a single region will fall in the high, medium, or low category at any given time, it is *not* equally likely that several regions will simultaneously experience high, medium, or low rates within one simulation. In general, it is substantially more likely that several regions will experience simultaneously low or high rates, than simultaneous medium rates. For example, Table 3 shows that while there are 268 simulations in which fertility is in the high category in each of the two world regions (i.e., in the A2 storyline), there are only 207 in which it is in the medium category in both regions (i.e., in the B2 storyline). This pattern holds for mortality assumptions as well: there are more simulations (297) in which mortality is high in both regions than simulations (266) in which it is medium in both regions. (Results follow the same general pattern when demographic rates in particular years, rather than averaged over time, are used, and when based on alternative regions or groups of regions.)

This result is likely due to a combination of factors. One factor is a consequence of the nature of the standard means of generating correlated variables. This process weights extreme values more heavily (through the squared terms in the definition of the correlation coefficient) than values close to the means, and therefore the generation process produces values that are more strongly correlated when they are relatively high or low, and more weakly correlated near the means. Another factor is that, even if the correlation were identical across categories, imperfect correlation produces a relative deficit in pairs of medium values because the "medium" category is bounded both from above and from below, while the "high" category has only a lower boundary and the "low" category has only an upper boundary. As a result, when imperfect correlation spreads the location of variable pairs across the state space, more are "lost" from the medium category than from the high or low categories. The net result of both factors is that scenarios consistent with medium values for vital rates in all regions (such as B2) are less likely than scenarios consistent with high values for fertility and mortality in all regions (such as A2).

It is possible that this pattern could be consistent with expectations for the real world. It may be that extreme values of fertility or mortality in one region would be expected to exert a strong influence on rates in other regions, but that moderate values would have a weaker influence. However, the empirical basis appears to be too thin to strongly support any particular claim at this point. Thus no great weight can be placed on conclusions regarding relative likelihoods of different storylines.

Conditional Probabilistic Emissions Projections

In principle, deriving conditional probabilistic emissions projections would require quantifying uncertainty within storylines not only in population, but also in each of the other elements driving emissions, as well as in the relationships between them. In this paper, the matter is greatly simplified by assuming that all SRES emissions scenarios within each storyline are equally likely, converting each to a per capita CO_2 emissions path (excluding land use emissions), and combining all per capita emissions paths with all population simulations underlying the conditional population projections for each storyline. The result is a set of equally likely emissions paths for each storyline, which can then be interpreted as a conditional probabilistic emissions projections and compared with the scenarios developed in SRES.

Neither the assumption of equal likelihood across the SRES per capita emissions scenarios within storylines, nor the assumption of a linear relationship between population and per capita emissions, is well founded. As discussed above (section 2), the SRES authors made no judgments about relative likelihoods across scenarios, and attempts to make such judgments have been controversial. Here, the assumption of equal likelihood of emissions scenarios *within* storylines probably has more to recommend it than judgments *across* storylines, but still must be treated as speculative. Whether population is linearly related to per capita CO_2 emissions or not is also questionable. There is some evidence that historically this has been the case (Dietz & Rosa, 1997), and typical optimal growth models used in climate analysis also tend to demonstrate an essentially linear relationship (Gaffin & O'Neill, 1997; O'Neill *et al.*, 2001). However, the dependence of emissions on population has not been analyzed in more detailed models.

Figure 4 shows the conditional probabilistic emissions projection for the A2 storyline, compared with the four A2 emissions scenarios from SRES. (Two SRES scenarios were dropped from this storyline: A2 IMAGE, results of which are not provided in SRES for all time periods, and A2–A1 MINICAM which models a combined storyline scenario that switches to A1 assumptions partway through the simulation period.)
The same basic conclusions drawn for the conditional population projections hold here as well. First, the conditional probabilistic emissions projections indicate a substantially wider range of uncertainty in emissions than do the original four SRES scenarios. The 60% uncertainty interval is 18–27 GtC/yr by 2100, and the 95% uncertainty interval is 14–32 GtC/yr. In comparison, the SRES A2 scenarios span a much smaller range of 28–33 GtC/yr. Second, the probabilistic projections are generally substantially lower than the SRES A2 emissions scenarios at the global level. While one of the four SRES scenarios is below the median over the first 30 years, by the end of the century all four are above the 60% uncertainty interval, and one of them is above the 95% uncertainty interval.

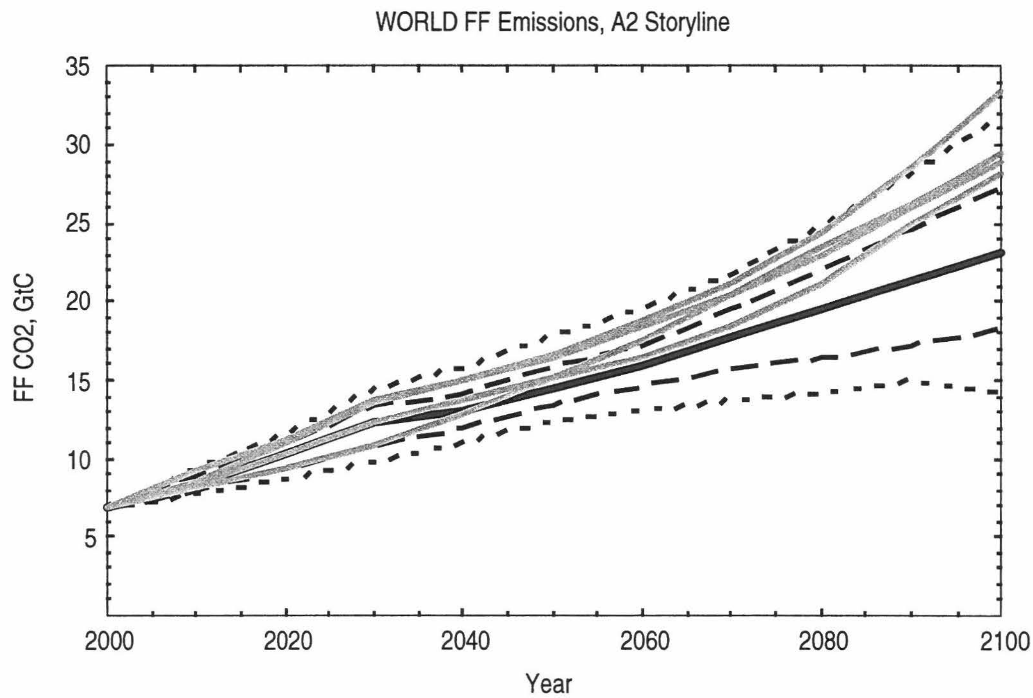


Figure 4. *Probabilistic emissions projections for the world conditional on the A2 storyline. Black lines: median (solid), 60% uncertainty interval (dashed), and 95% uncertainty interval (dotted) for emissions based on the IISA 2001 conditional projections. Gray line: SRES A2 emission scenarios.*

Figure 5 presents results for all storylines for the world and the two regions, reported in terms of cumulative CO_2 emissions over the period 2000–2100 (cumulative emissions are a better indicator of impact on climate than emissions at a point in time). It shows that differences between the conditional projections and the SRES scenarios are the largest for the A2 storyline. In the B2 storyline, the SRES scenarios cover about half of the 95% uncertainty interval in the conditional projection, although they are still relatively high, with nearly all scenarios falling above the median of the conditional projections, particularly in the developing country region. For the A1 and B1 storylines the SRES scenarios essentially span the full 95% uncertainty interval of the conditional projections, both at the world level and within the two regions. This occurs even though the conditional probabilistic population projection for these storylines covers a wide range by 2100, and is substantially lower than the single SRES population assumption in the industrialized country region (Figure 3b). The reason the SRES emissions scenarios compare differently to the conditional projections is that the SRES scenarios include a fairly wide range of possible per capita emissions trends in the A1 and B1 storylines (particularly in A1). Thus in terms of total emissions, uncertainty in per capita emissions can to some extent make up for a lack of uncertainty in population. This is not in general a good way to account for potential uncertainty in outcomes, however, because there are often other aspects of an analysis, besides total emissions, that may also depend on the population assumptions—land use change or climate change impacts are two examples. Uncertainty in per capita emissions will not make up for lack of uncertainty in population assumptions when addressing these other components of the problem.

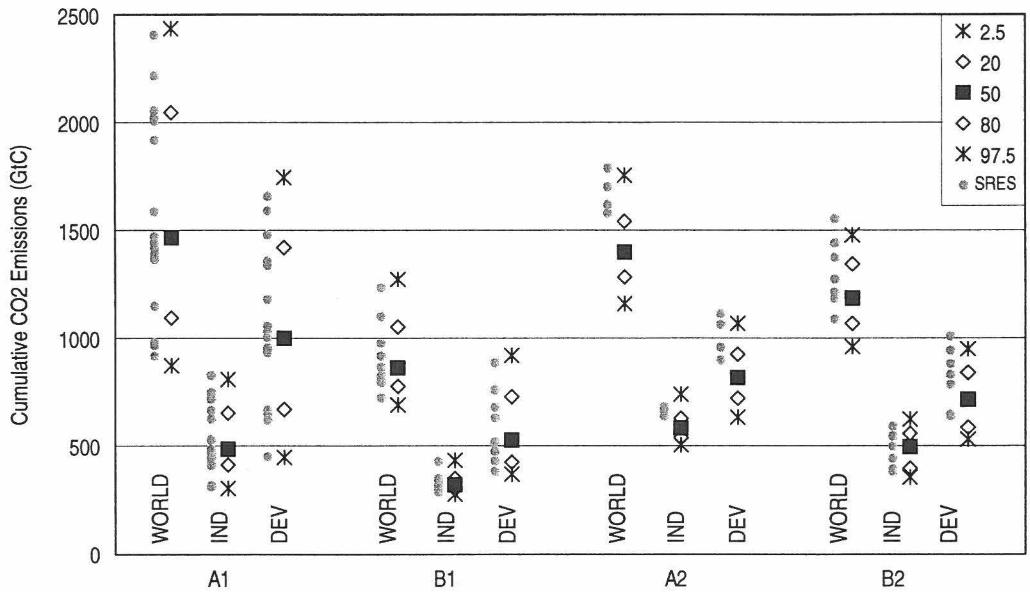


Figure 5. *Cumulative CO_2 emissions, 2000–2100, for individual SRES scenarios (gray dots), and percentiles (median and the bounds of the 60% and 95% uncertainty intervals) of conditional probabilistic projections, for each storyline.*

5 Discussion and Conclusions

Uncertainty in future greenhouse gas emissions has typically been approached in one of two ways: through the development and analysis of alternative scenarios, and through fully probabilistic projections based largely on expert opinion. Here I have proposed a new option: the use of conditional probabilistic projections. This approach essentially combines the other two. Alternative sets of conditions are developed using scenario development methods, while uncertainty given a particular set of conditions is then defined probabilistically. The example presented here illustrated the concept using the SRES storylines as constraints in developing a set of four conditional probabilistic population and emissions projections.

Results demonstrate that the approach can provide a fuller accounting for uncertainty in population and emissions within storylines. They show, for example, that the uncertainty range for population within each storyline could well be quite wide, in contrast to the single population scenario assumed in SRES. In terms of emissions, the conditional probabilistic projections indicate an uncertainty distribution in the A2 and, to a lesser extent, the B2 storylines that is substantially wider than the range covered by the SRES scenarios. Furthermore, the conditional probabilistic approach can give a sense of whether existing emissions scenarios are biased in a particular direction. The B2 and, particularly, the A2 scenarios used in SRES fall at the upper end of the conditional distribution of emissions, suggesting that a fuller accounting for population uncertainty, using the most recent population projections, would produce generally lower emissions within these two storylines. These conclusions must remain tentative since they make important assumptions about relative likelihoods of per capita emissions paths and about relationships between variables. They are also dependent on the particular choices made in defining the criteria for the conditional projections. Nonetheless they suggest that substantial changes in the outlook for emissions (within storylines) are possible using the conditional probabilistic approach.

This exercise also raised several points regarding the population projections on which it is based. For example, the set of 1000 simulations available for the IIASA (Lutz *et al.*, 2001) projections contained relatively few simulations that matched a rather broad set of criteria for three variables and two world regions over one time period. Adding additional criteria, more regions, or additional time periods as constraints quickly makes the task impossible. This implies that when producing unconditional population projections, it would be useful to generate a much larger number of simulations in order to support applications calling for the conditional probabilistic approach. In addition, this application to the SRES scenarios demonstrated that the correlations assumed in the unconditional population projections have important implications for results. The IIASA projections are based on a single assumption about the strength of correlations between variables, over time, and across regions. This substantially reduces the number of simulations matching storylines that assume a different correlation. One effect is simply practical: it may be hard to find enough simulations in this case to characterize fully the uncertainty distribution. However it also means that the judgment of the relative likelihood across storylines, which should in principle be possible using the conditional probabilistic approach, is extremely sensitive to the specification of correlations. Conclusions regarding relative likelihoods must await further work on the empirical and theoretical basis for these correlations.

Finally, while the aim of this paper has been to explore how probabilistic approaches to uncertainty in population might benefit emissions projections, it is also worth considering the benefits of the storyline approach to the treatment of uncertainty in population projections. Storylines are a device for structuring thinking about a future with deep uncertainty. They are also a means of making projections more useful to users. Assumptions regarding the reasoning behind the choice of driving forces, parameter values, and modeling approaches are made more explicit. More explicit discussion of the reasoning motivating the input assumptions that underlie probabilistic population projections would make them more useful to the interdisciplinary research community. Demographers producing such projections might consider storylines as one way of communicating this information. Users from other fields need to know the conditions under which demographers think that, for example, fertility may turn out to be high in South Asia, or very low in Western Europe, or what might lead trends across regions to be strongly or weakly correlated. For a user searching for ways to join projections from one field with projections from another, it is this underlying reasoning that provides the link. Making this reasoning as explicit as possible would greatly facilitate interdisciplinary work.

Acknowledgements

The author thanks Wolfgang Lutz and Sergei Scherbov for many valuable discussions, particularly regarding the influence on results of correlations in demographic rates across regions.

References

Alho, J. (1997). Scenarios, uncertainty and conditional forecasts of the world population. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, **160**(1), 71–85.

Cubasch *et al.* (2001). *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, UK: Cambridge University Press.

Dessai, S. & Hulme, M. (2001). Climatic implications of revised IPCC emissions scenarios, the Kyoto Protocol and quantification of uncertainties. *Integrated Assessment*, **2**, 159–170.

Dietz, T. & Rosa, E.A. (1997). Effects of population and affluence on CO_2 emissions. *Proceedings of the National Academy of Sciences USA*, **94**, 175–179.

Edmonds, J., Reilly, J., Gardner, R. & Brenkert, A. (1986). *Uncertainty in Future Global Energy Use and Fossil Fuel CO_2 Emissions 1975 to 2075* (with appendices). Washington, DC: United States Department of Energy.

Gaffin, S.R. & O'Neill, B.C. (1997). Population and global warming with and without CO_2 targets. *Population and Environment*, **18**(4), 389–413.

Grübler, A. & Nakicenovic, N. (2001). Identifying dangers in an uncertain climate. *Nature*, **412**(6842), 15.

Keepin, B. (1986). Review of global energy and carbon dioxide projections. *Annual Review of Energy*, **11**, 357–392.

Lempert, R.J., Schlesinger, M.E., Banks, S.C. & Andronova, N.G. (2000). The impacts of climate variability on near-term policy choices and the value of information. *Climatic Change*, **45**, 129–161.

Lutz, W., Sanderson, W., Scherbov, S. & Goujon, A. (1996). World population scenarios for the 21st century. In *The Future Population of the World. What Can We Assume Today?* (Revised Edition), Ed. W. Lutz, pp. 361–396. London: Earthscan.

Lutz, W., Sanderson, W. & Scherbov, S. (2001). The end of world population growth. *Nature*, **412**, 543–545.

Lutz, W. & Scherbov, S. (2002). Can immigration compensate for Europe's low fertility? Interim Report, IR-02-052. Laxenburg, Austria: IIASA.

Moss, R.H. & Schneider, S.H. (2000). Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting. In *Cross Cutting Issues Guidance Papers*. Geneva, Switzerland: Intergovernmental Panel on Climate Change.

Nakicenovic, N. *et al.* (2000). *Special Report on Emissions Scenarios*. Cambridge, UK: Cambridge University Press for the Intergovernmental Panel on Climate Change.

New, M. & Hulme, M. (2000). Representing uncertainty in climate change scenarios: a Monte Carlo approach. *Integrated Assessment*, **1**, 203–213.

Nordhaus, W.D. & Popp, D. (1997). What is the value of scientific knowledge? An application to global warming using the PRICE model. *The Energy Journal*, **318**, 1–45.

Nordhaus, W.D. & Yohe, G. (1983). Future carbon dioxide emissions from fossil fuels, Cowles Foundation Paper No. 580. New Haven, CT: Yale University.

O'Neill, B.C., MacKellar, F.L. & Lutz, W. (2001). *Population and Climate Change*. Cambridge, UK: Cambrige University Press.

O'Neill, B.C. (submitted). Population scenarios based on probabilistic projections: An application for the Millennium Ecosystem Assessment. Submitted to *Population and Environment*.

Sanderson, W.C., Scherbov, S., O'Neill, B.C. & Lutz, W. (2004). Conditional probabilistic population forecasting. *International Statistical Review*, **72**, 157–166.

Schneider, S.H. (2001). What is 'dangerous' climate change? *Nature*, **411**, 17–19.

Schneider, S.H. (2002). Can we estimate the likelihood of climatic changes at 2100? *Climatic Change*, **52**, 441–451.

Schwartz, P. (1991). *The Art of the Long View*. New York: Doubleday.

UN. (1998). *World Population Projections to 2150*. New York: United Nations.

Watson, R.T. & the Core Writing Team (35 authors). (2001). *Climate Change 2001: Synthesis Report*. Cambridge, UK: Cambridge University Press for the Intergovernmental Panel on Climate Change.

Webster, M.D., Babiker, M., Mayer, M., Reilly, J.M., Harnisch, J., Sarofim, M.C. & Wang, C. (2002). Uncertainty in emissions projections for climate models. *Atmospheric Environment*, **36**(22), 3659–3670.

Wigley, T.M.L. & Raper, S.C.B. (2001). Interpretation of high projections for global-mean warming. *Science*, **293**, 451–454.

Résumé

Les changements futurs dans la taille, la composition et la distribution spatiale de la population sont des facteurs clés dans l'analyse du changement climatique, et leur évolution future est très incertaine. Dans les analyses du changement climatique, on tient traditionnellement compte de l'incertitude sur la population en utilisant des scénarios alternatifs couvrant un éventail de résultats. Cet article illustre comment des projections à probabilité conditionnelle permettent de combiner les approches probabilistes avec l'approche basée sur des scénarios, typiquement employée dans les travaux de projections d'émissions de gaz à effet de serre. La présentation combine un ensemble de scénarios d'émissions développé par le Panel

Intergouvernemental sur le changement climatique (IPCC) avec des projections de population probabilistes existantes de l'IIASA. Les résultats démontrent que les projections à probabilité conditionnelle peuvent expliquer plus complètement l'incertitude sur les émissions dans le cadre de scénarios conditionnels des modèles de développement futurs, qu'elles peuvent permettre de juger de la cohérence de scénarios individuels avec un scénario donné, et de fournir une idée des vraisemblances relatives dans les scénarios, au moins d'un point de vue démographique. Ils peuvent aussi servir d'étape vers une quantification plus précise de l'incertitude dans les projections d'émissions.

[*Received May 2003, accepted January 2004*]

International Statistical Review (2004), **72**, 2, 185–199, Printed in The Netherlands

Random Scenario Forecasts Versus Stochastic Forecasts

Shripad Tuljapurkar[1], Ronald D. Lee[2] and Qi Li[3]

[1]*Biological Sciences, Stanford University, Stanford, CA 94305, USA. E-mail: tulja@stanford.edu*
[2]*Demography and Economics, University of California, Berkeley, CA 94305, USA*
[3]*Management Sciences and Engineering, Stanford University, Stanford, CA 94305, USA*

Summary

Probabilistic population forecasts are useful because they describe uncertainty in a quantitatively useful way. One approach (that we call LT) uses historical data to estimate stochastic models (e.g., a time series model) of vital rates, and then makes forecasts. Another (we call it RS) began as a kind of randomized scenario: we consider its simplest variant, in which expert opinion is used to make probability distributions for terminal vital rates, and smooth trajectories are followed over time. We use analysis and examples to show several key differences between these methods: serial correlations in the forecast are much smaller in LT; the variance in LT models of vital rates (especially fertility) is much higher than in RS models that are based on official expert scenarios; trajectories in LT are much more irregular than in RS; probability intervals in LT tend to widen faster over forecast time. Newer versions of RS have been developed that reduce or eliminate some of these differences.

Key words: Probabilistic forecast; Population forecast; Trajectory; Vital rates; Scenario; Random scenario; Dependency ratio.

1 Introduction

Population forecasts are widely used by governments, and agencies such as the United Nations, as an essential part of policy analysis. A population forecast requires launch data—the size and composition of the initial population, its vital rates—and a procedure for forecasting the population's vital rates (mortality, fertility, immigration) over the period covered by (also called the span of) the forecast. Forecasters recognize that the demographic basis for forecasting vital rates over time is inadequate, in the sense that any single forecast is likely to be off the mark. In consequence, demographers attempt to incorporate uncertainty into forecasts, usually by generating a range of forecasts along with some indication of their relative likelihood. A widely used method to indicate uncertainty is to prepare three forecasts, called High–Medium–Low, in which the Medium is most likely, and the others indicate the range of less likely outcomes. This is what the United Nations and the United States Census Bureau, for example, do in official projections. In practice there may be many more than three underlying forecasts but they are combined to present three outcomes for any quantity of interest—we call these *scenario* forecasts. The shortcomings of such forecasts are discussed by, for example, Lee & Tuljapurkar (2000).

There are three general ways of improving on scenario forecasts. One method, due to Keyfitz (1981), Stoto (1983) and Keilman (1998), starts with a historical analysis of errors in past forecasts. The statistical properties of errors are assumed to be the same in future forecasts as in the past, so that the historical data may be used to generate prediction errors. A second approach is a *stochastic* forecast as described by Lee & Tuljapurkar (1994)—a statistical analysis of historical data is used to estimate dynamic stochastic models (e.g., a time series model) of the variability in vital rates. A stochastic forecast is then made using these models—here we call it the LT forecast—that yields probabilities for different ranges of any outcome of interest. In effect, the LT method relies upon a probability distribution for the trajectories of the process that is generated dynamically. The main point of the LT method is that a forecast should reflect as directly as possible the uncertainty displayed by historical data. The philosophy of the LT method is based on and shared by other dynamic approaches to generating future uncertainty, notably the work of Alho & Spencer (1985), Cohen (1986), and Alho (1990) who uses in addition a form of *post hoc* error analysis to project predictive probability distributions.

A third approach due to Pflaumer (1988) and Lutz *et al.* (1996, 1997) uses a *random scenario* method that we refer to here as an RS forecast. In its original versions, an RS forecast uses historical and other analysis, and expert opinion, to assert that vital rates at a particular future target date (or set of dates) will lie in a specified range and have a particular probability distribution. Expert opinion may be solicited formally: see Lutz *et al.* (1996) for a discussion and example. To generate an RS forecast, one target value of vital rates at the target date is selected from the assumed distribution, and it is assumed that vital rates will follow some specified smooth trajectory from launch value to target value. A forecast is made using vital rates on this trajectory; a new target is selected, a new trajectory of rates is generated, a new forecast is made; and so on until one has a large set of forecasts. These forecasts have a probability distribution that reflects the forecaster's assumptions about target values, their distribution, and the trajectory of vital rates. This probability distribution is now used as a basis for making probabilistic forecasts for quantities of interest.

The original RS and LT approaches are very different in their approach to formulating the dynamics that underlie their respective forecasts. The objective of this paper is to compare the two approaches in terms of the properties of the forecasts that they generate. We use two examples to compare the original RS and LT approaches—first, a stylized model of a population without age structure, and second, a complete projection model for the US population with a launch date of 1997. The stylized model highlights the differences between the dynamics of the stochastic processes in RS and LT, and the complex model shows how the content of the forecasts differs in various substantive ways. Our practical aim is to educate makers and users of forecasts to the key differences that arise between the two types of forecast, especially in terms of probability distributions of outcomes and dynamic correlations between important variables over time.

Our comparison deliberately employs stylized versions of RS and LT in which any differences between the two approaches are likely to be large. In practice, as pointed out by many forecasters, virtually all forecasts may require elements of both approaches. For example, the LT approach exploits long runs of accurate historical data when they are available, whereas the RS approach is a way of filling in gaps in the data or in the mechanisms of change. Thus it is natural and sometimes essential to consider intermediate modelling methods. One notable effort in this direction is the work of Lutz *et al.* (2001) who extended the original RS method by assuming that the vital rate trajectories have an additional random component. Their new method is a substantial advance over the original RS, but we will not discuss it here. Our goal is not to make a comprehensive comparison between actual forecast outcomes using the latest methods, but only to show how different forecast methodologies can generate persistent differences in the forecasts that they yield. Armed with the understanding that we provide here of how these differences arise in simple cases, the reader will be prepared to explore how the newer methods of Lutz *et al.* (2001) narrow these differences to good effect.

The next two sections compare a simple stylized RS model and an LT model for two cases: first, a simple scalar population dynamics, and second, the more realistic case of projecting the US population. The final section discusses our findings and conclusions.

2 Scalar Forecasts

We begin with the simplest setting in which to examine how one makes a forecast. Consider a population which has total size $N(t)$ at the end of year t—in this section we ignore age structure and other details. During the subsequent year the population is assumed to grow by a ratio $R(t + 1)$ to a number

$$N(t + 1) = R(t + 1)N(t). \tag{1}$$

To make a forecast starting with a known $N(0)$ at year $t = 0$ until the end of a forecast span of T years, we need only specify the growth factors $R(1), R(2), \ldots, R(T)$. How do we specify and compare forecasts in the simplest versions of the RS and LT approaches? We present below simple but representative models for the dynamics as used in RS and LT, and examine their outcomes; analytical results are given but the focus is on graphical comparisons using a representative numerical example.

A Simple RS Model and Forecast

A simple stylized RS forecast is described by assumptions about the final distribution of the growth rates, and the intermediate trajectories.

1. There is a known initial growth factor $R(0)$ and initial population $N(0)$.
2. The growth factor in the last year, $R(T)$, has a most likely value R_M. With high probability (say 90%) the value of $R(T)$ will lie in the range from $R_L < R_M$ to $R_H > R_M$. In practice, these values can be based on a survey of expert opinion.
3. The value of $R(T)$ has either a uniform or a normal random distribution with mean R_M and a variance σ_R that reflects the range assumed above.
4. Given any particular value of $R(T)$, the growth factors for all forecast periods are obtained by interpolation between $R(0)$ and R_1,

$$R(t) = (1 - h(t))R(0) + h(t)R(T), \tag{2}$$

where $h(0) = 0, 0 \leq h(t) \leq 1$, and $h(T) = 1$. We assume a specific interpolation function, e.g., $h(t) = (t/T)$.

To make the RS forecast we simply select a value of $R(T)$ according to the assumptions, generate a trajectory of $N(t)$ between $t = 1$ and T, and repeat. Our numerical illustrations are based on 500 such trajectories.

A Simple LT Model and Forecast

A simple LT model for this setting is described by assumptions about a stochastic process that generates trajectories of growth rates over time.

1. There is a known initial growth factor $R(0)$ and initial population $N(0)$.
2. The growth factor $R(t)$ is a serially correlated random process with an expected mean of R_M and follows the equation

$$R(t) - R_M = \rho(R(t - 1) - R_M) + e(t), \tag{3}$$

where $e(t)$ is an iid random process, independent of $R(t)$, with zero mean and variance σ_e.
3. We specify R_M, based on either evidence about trends or expert opinion.
4. We specify ρ and σ_e. In actual application, we would estimate these by fitting the model equation (3) to historical data— this is a fit conditional on the assumed value R_M.

To make an LT forecast we generate a stochastic sequence of $R(t)$ for times $t = 1$ to T using equation (3), forecast the corresponding sequence of $N(t)$, and repeat. Here too, our numerical illustrations are based on 500 such trajectories.

A Numerical Example

In both the RS and LT cases we obtain many (here, 500) trajectories and then use these samples of the forecast trajectory to generate probability forecasts. For example, if $T = 50$, we can describe our forecast of $N(25)$ for $t = 25$ in several ways:

1. Plot the 500 trajectories of $N(t)$ through $t = 25$.
2. Use a histogram (or some better statistical density estimator) to estimate the probability distribution of $N(25)$.
3. Display a predictive distribution that shows the estimated quantiles of the forecast distribution of $N(25)$.
4. Estimate the mean and standard deviation of $N(25)$.

In our numerical example, for both RS and LT we set initial values

$$R(0) = 1.028, \ N(0) = 6.$$

The target for the RS example at $T = 50$ is chosen to be symmetric with

$$R_L = 0.01, \ R_M = 0.02, \ R_H = 0.03.$$

For a normally distributed target we set the RS final variance to be $\sigma_R = (R_M - R_L)/1.6448$. For a uniformly distributed target we set $\sigma_R = (R_M - R_L)/0.45$.

In the RS example we pick some $R(T)$ from the target distribution and then connect $R(0)$ to this final value by an interpolation as in equation (2). The baseline case we use is the **linear** case with $h(t) = (t/T)$. We compare this with a **slower** than linear approach to the target with a *quadratic* interpolation $h(t) = (t/T)^2$. We also compare the baseline with a **faster** than linear approach to the target with a *square-root* interpolation $h(t) = (t/T)^{1/2}$.

The steady state variance of the LT growth rates is

$$\sigma_{LT}^2 = \sigma_e^2/(1 - \rho^2). \tag{4}$$

We use values of ρ from the set $\{-0.9, -0.5, 0, 0.5, 0.9\}$. The last value is what we expect for a highly correlated series, as is typical of human fertility rates.

Making a Comparison

How should we make a sensible comparison between our simple stylized RS and LT models? In a practical case, such as the US forecast in the next section, the two methods use distinct approaches to arrive at the statistical dynamics of growth rates. As a result, the means and variances of the vital rates over time may well be different and in turn will produce differences in the forecasts—we must then keep track of such differences in making any comparison. But in the present simple case we are free to set the parameters in a way that highlights only the essential differences in the dynamics of simple RS and LT models. To keep the models as comparable as possible, we set the LT mean value of growth rate to be R_M, the same as the RS value, and also make the asymptotic variances equal by setting $\sigma_{LT} = \sigma_R$ which means that we set

$$\sigma_e = \sqrt{1 - \rho^2}\, \sigma_R. \tag{5}$$

Given these parameter settings, any differences we find are due solely to the different ways in which the two approaches model the vital rates.

Comparing Forecast Properties

Trajectories

Figure 1 shows 25 trajectories from an RS simulation using a normally distributed target and linear interpolation; also shown are the trajectories that go to the High, Mean and Low target growth rates. Note the smooth trajectories do not cross, and the High–Mean–Low range contains almost all the results.

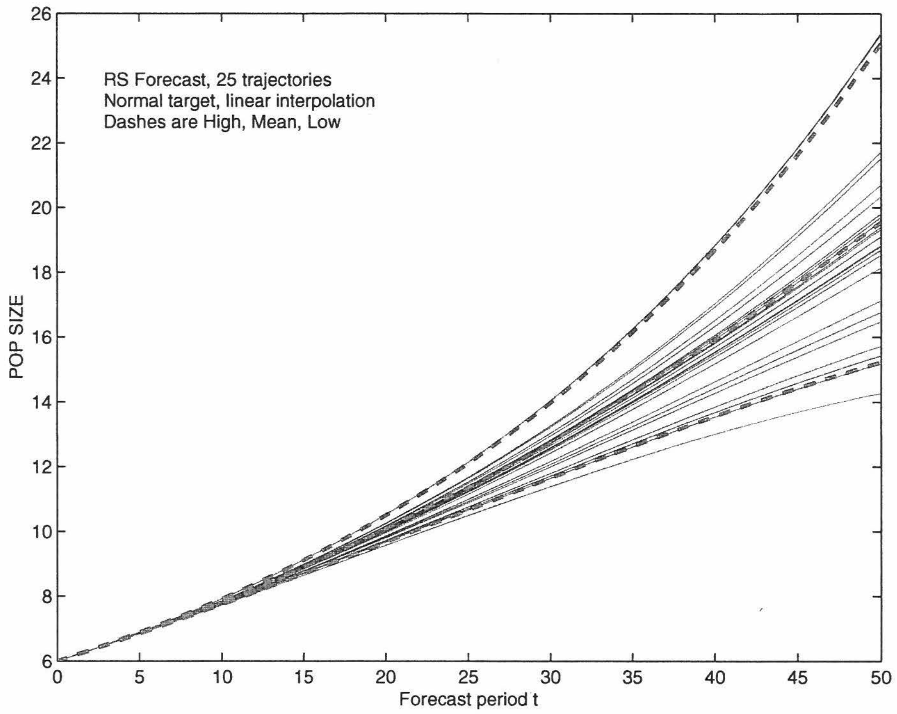


Figure 1. *A subset of trajectories (solid lines) of the RS version of a scalar projection, also shown (dashed lines) are trajectories based on the high, medium, low targets.*

Figure 2 shows 25 trajectories from an LT simulation with $\rho = +0.9$. Note that trajectories do cross, there is a much wider range of population sizes and a "looser" distribution. Instead of a High-Mean-Low we show three trajectories that span the outcomes at the end of the projection, i.e., the trajectories whose terminal values are the median and the upper and lower 95% at the end of the forecast.

Probability Distributions of Forecast Population

A general feature of RS is that the shape of the target distribution of $R(T)$, which is set by assumption, largely determines the shape of the distribution of $N(T)$. To see why, note that in RS

$$N(T) = [(1 - h(1))R(0) + h(1)R(T)]\,[(1 - h(2))R(0) + h(2)R(T)]\,\ldots \qquad (6)$$
$$[(1 - h(1))R(0) + h(T - 1)R(T)]\,N(0).$$

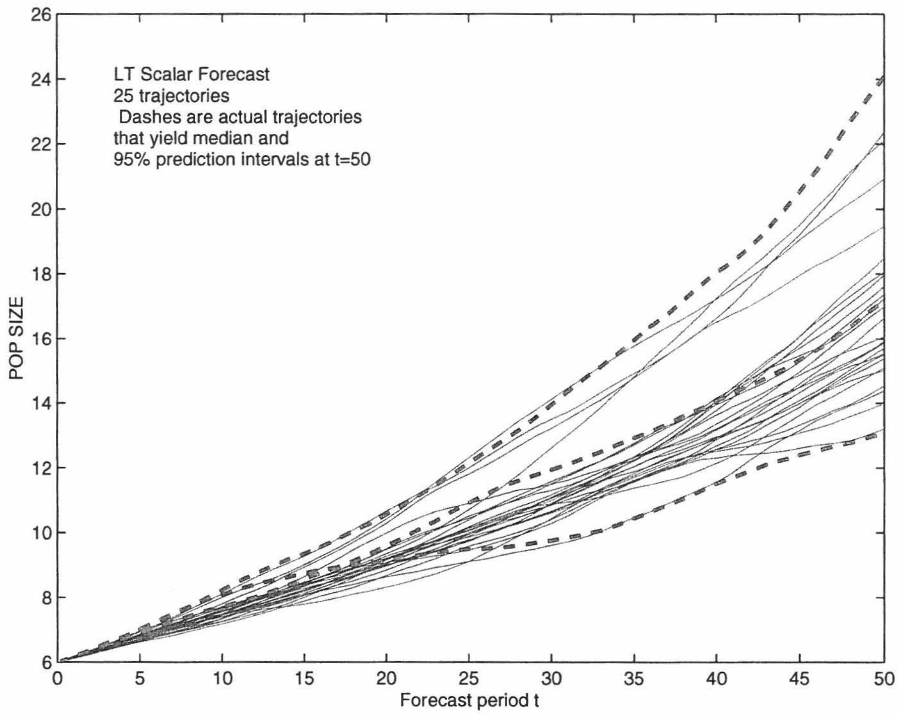


Figure 2. *A subset of trajectories (solid lines) of the LT version of a scalar projection, also shown (dashed lines) are 95 percent prediction intervals based on 500 simulated trajectories.*

Hence $\log N(T)$ has the same distribution as

$$\sum_i \log([(1 - h(i))R(0) + h(i)R(T)]).$$

The factors in this sum are perfectly correlated with each other and the distribution of each will be very similar to that of $R(T)$ – to obtain each factor we multiply $R(T)$ by a number and add a number. Therefore the distribution of $N(T)$ must be similar to the distribution of $R(T)$.

This distributional feature of RS is illustrated by Figure 3, in which the top panel is a histogram of the forecasted values of $N(50)$, the lower panel is a histogram of $\log N(50)$. This figure was generated using a uniformly distributed target, i.e., a uniform distribution for $R(50)$, and we see that the result is an approximately uniformly distribution of $N(50)$. Note that the logarithm of N is also approximately uniformly distributed. We note also that if we instead use a normally distributed RS target, we obtain an approximately normal distribution (not shown here) for both $N(50)$ and its logarithm.

In contrast to RS, a general feature of LT is that the distribution of $N(T)$ for large T (say $T > 10$ or so), is always approximately lognormal. To see why note that in LT we have

$$\log N(T) - \log N(0) = \sum_i \log R(i), \tag{7}$$

where the $R(i)$ are a serially correlated sequence. But the elements of the sequence are **not** perfectly correlated and in fact the correlation between $R(t)$ and $R(t+s)$ goes to zero at the rate ρ^s as the difference $s > 0$ increases. This implies that the sum on the right side of the above equation must obey the central limit theorem and be asymptotically normally distributed. In other words, $\log N(t)$ in LT will be roughly normally distributed and $N(t)$ will be roughly lognormally distributed. We do not illustrate this feature here.

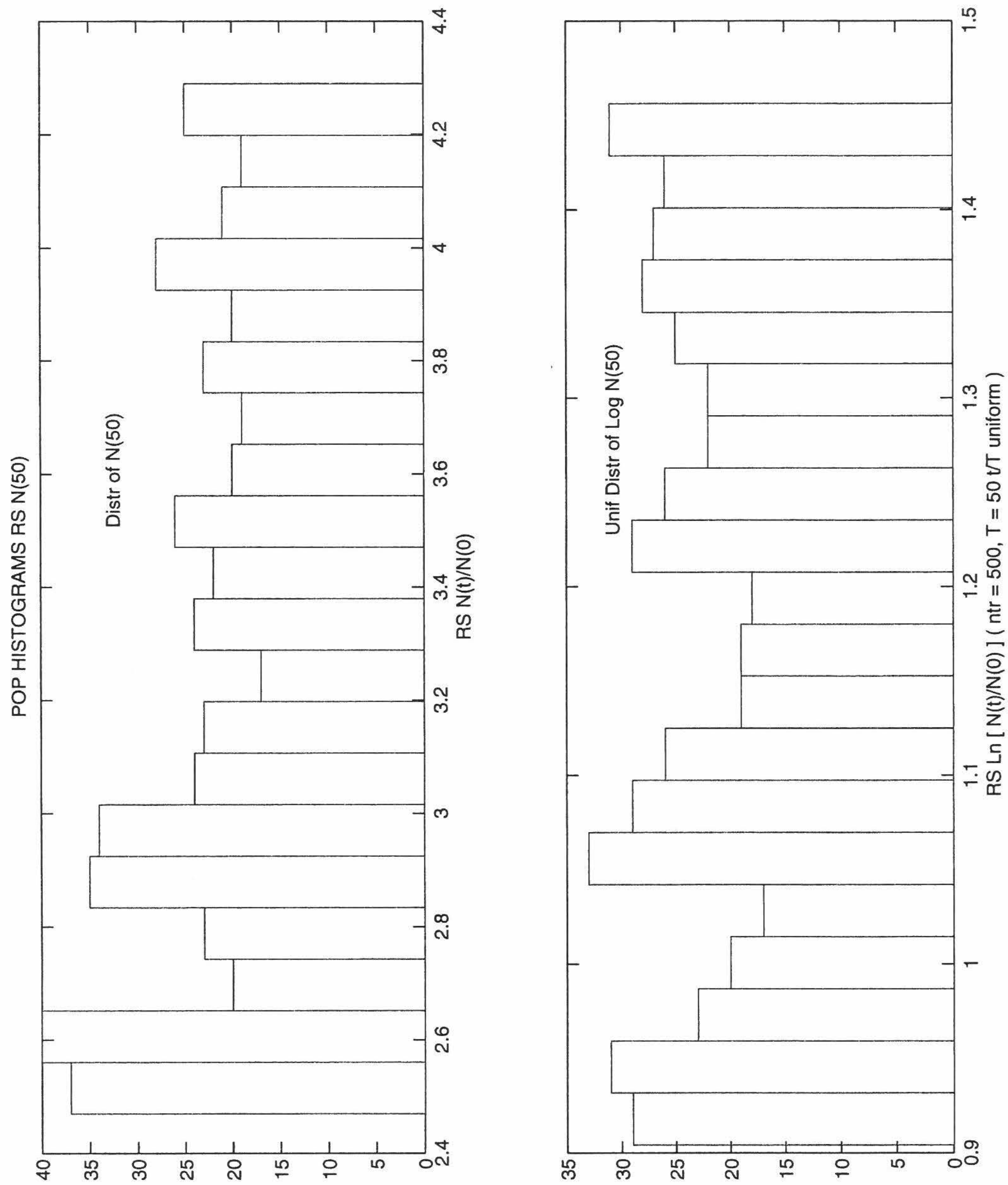


Figure 3. *Histograms of 500 forecast values of population in a scalar projection RS model that uses a uniform distribution of the target growth rates in the final year. Upper panel shows the population, lower panel the logarithm of population.*

Growth Rate of Variance

The growth rate of forecast variance in RS depends strongly on the trajectory along which vital rates converge to their target value (e.g., faster or slower than linear) and only weakly on the shape of the target distribution. This is illustrated in Figure 4. To focus on the growth rate of the variance in $N(t)$ we have scaled out the average growth rate, which in RS is the growth rate along the trajectory towards R_M. As we might expect, variance grows most rapidly when growth rates converge slowly to the target—in this case different RS trajectories start at the same growth rate but diverge rapidly. This also means that each trajectory is pretty close to its target rate after only a few years. Note that the variances in Figure 4 grow quadratically with time. This too is a feature of the RS projection, as may be seen from equation (7)—the variance of that sum of t perfectly correlated terms will grow as t^2.

In contrast consider equation (7) for LT—the variance of a sum of terms whose correlation dies out as the terms become more distant grows linearly as t. The contrast between forecast variances in RS and LT is shown in Figure 5. The three LT curves show how strongly serial correlation ρ affects the LT variance, but all three LT curves rise linearly with t at long times. The RS curve shows its characteristic quadratic rise. Note also that the small t variances change very differently—the LT variances rises quite rapidly with time at first, then drops to a linear change. The RS variance always grows quadratically. Thus early into the forecast the LT method will usually result in substantially greater forecast variance than will RS.

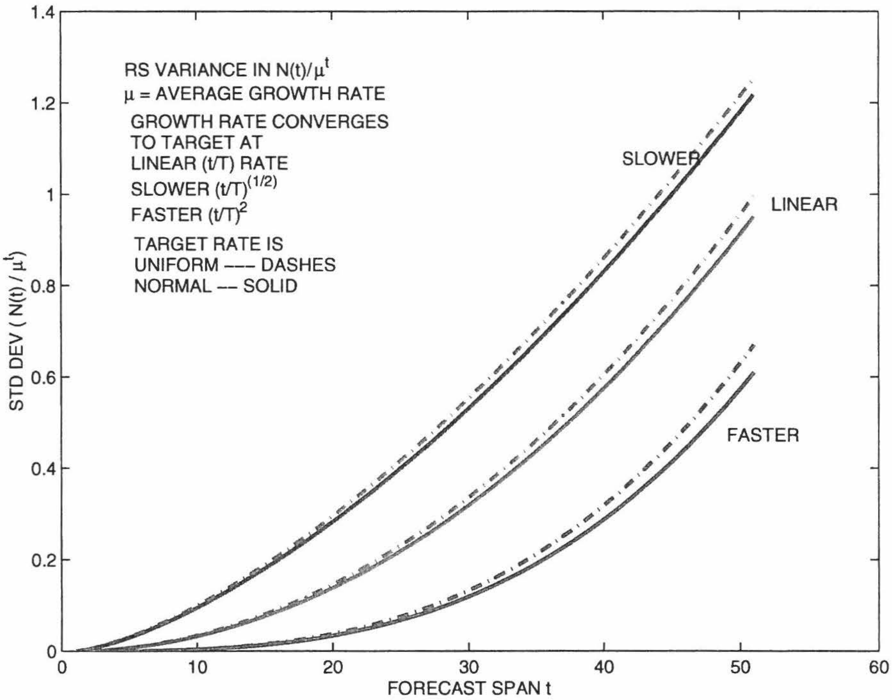


Figure 4. *Forecast variance as a function of forecast span in a scalar projection RS model. The three pairs of curves obtain for slow, linear, or fast approaches to the target. The choice of target distribution does not matter: solid lines for uniform target, dashed lines for a normal target.*

3 Age-structured Forecasts for the US

We must first consider the appropriate setting in which to compare forecasts. In the scalar model of the previous section, we were free to set the parameters to minimize the difference between forecast methods. In contrast, here we are describing an actual country and demographic history for which both the simple RS method and the LT method must be based on separate analyses that result in dynamic models of mortality and fertility change. For example, the LT model below fits a time series model for fertility but constrains the long-run average fertility level. We constrain this average to equal the ultimate medium fertility scenario of the US Social Security Administration (a similar value is used by the US Census Bureau). The LT time series model generates fertilities that vary substantially around the constrained average. These official agencies also provide a range of high (optimistic) and low (pessimistic) values for the ultimate fertility, and in our simple RS model we assume that they would cover the ultimate probability range (95 percent in case of a normally distributed target). The trajectories of the simple RS are then bounded by that range at all times, and one can only change the variance by changing the target range—this is certainly possible but for

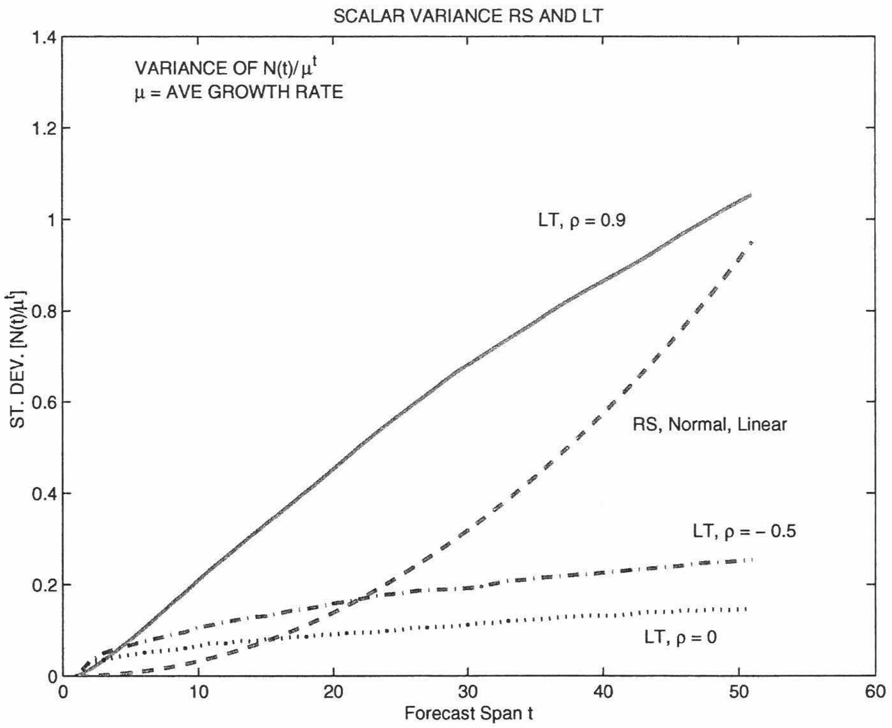


Figure 5. *Forecast variances compared for RS and LT in a scalar projection. Variance for LT is shown for three values of serial correlation ρ—solid line for $\rho = +0.9$, dotted line for $\rho = 0$, and dash-dot line for $\rho = -0.5$. Variance for RS is shown by the dashed line. See text for further discussion.*

our example we rely on the ranges used by official agencies. Thus here we cannot and do not force an equality in the long-run variances of the LT and RS methods. On the other hand, we do impose long-run similarities in the behavior of average fertility in the two approaches, because both require the assumption of a target. We discuss these issues further in the next section.

LT Forecast

The LT forecast for the US is the one used by Anderson, Tuljapurkar & Lee (2001) with launch data in 1996. Mortality follows a Lee–Carter model estimated using data from 1933 to 1996, fertility follows a time series model using data for the same period (for details consult the reference cited or Lee & Tuljapurkar, 1994), and the model trajectories for fertility and mortality are not correlated with each other. We note that fertility displays very high positive serial correlation and that the long-run variance in the total fertility rate projected by LT is high. Immigration is deterministic and follows the US Social Security Administration (1996) medium scenario, hereafter called the SSA medium scenario. The forecast span is 1997 to 2096.

RS Forecast

We first set assumptions for a stylized version of RS, based loosely on the conventional (not random) scenario methods used by the SSA. Our simple RS model is based on the early work of Pflaumer (1988) and Lutz *et al.* (1996) and is not representative of the newer RS methods in Lutz *et al.* (2001). The TFR in the target year 2096 is taken to have High, Medium, Low values of 2.2, 1.9, 1.6, respectively. The Medium to Low difference is used to set the variance of a normally distributed target TFR. The age distribution of fertility is taken to be fixed over time at the value in the launch year. We choose a target in 2096 at random from this distribution and use linear interpolation to generate TFR values for all the launch years: we do this for every choice of target TFR.

Mortality targets are set in terms of targets for the value in 2096 of the expected lifespan at birth, denoted by e_0. We start with SSA scenarios for 2075 and extend them by interpolation to get scenarios for 2096 (see table 1):

Table 1

e_0	Male 2075	Female 2075	Combined 2075	Combined 2096
High	83.1	88.2	85.7	88.3
Medium	79.3	84.2	81.8	83.3
Low	76.4	81.2	78.8	79.5

We use the difference between Medium and Low to set the variance of a normally distributed target e_0 for each sex. To make population projections we need the full age-specific death rates—we project these in four steps. One, choose a random $e_0(T)$ target for each sex. Two, assume that age-specific central death rates $m(x, t)$ at $t = 0$ and at $t = T$ are proportional, $m(x, t) = \lambda\, m(x, 0)$. Three, since we know $m(x, 0)$ we can use $e_0(T)$ to compute a proportionality factor $\lambda(T)$ such that the final death rates yield the desired expectation of life. Four, use linear interpolation between $m(x, 0)$ and $m(x, T)$ to find the death rates in any forecast year: do this for each sex along each trajectory.

Forecasts of Population

Any forecast of an age and sex structured population over a long span contains a huge amount of information. We choose only a few components of the forecast for our comparison, aiming to identify the key differences that obtain between LT and simple RS methods. Where relevant we indicate the substantive importance of the variables we have chosen to discuss.

We begin with the dramatic difference between LT and RS forecasts in terms of the width of their prediction intervals for total population over the entire forecast span, shown by the 95% prediction intervals in Figure 6. Note that the LT forecast is substantially more uncertain even very early into the forecast. We note that in this example the RS forecast of population size is roughly normal because the underlying vital rates have normally distributed targets.

Next, we consider (Figure 7) projections of one component of population, the age class 0 to 1 year—births. Births matter to numerous policy issues, ranging from expenditures on maternal care to education. The upper left panel compares the mean of the LT and RS forecasts. Note that the RS forecast shows a "classical" demographic convergence with damped waves that die out by 2050. The mean LT forecast shows substantial variation over time reflecting predicted future baby booms. Although the scale of the upper left panel makes the predicted means of the two methods look fairly close, there are differences of the order of several hundred thousand births over the entire forecast—see the logarithmic plot in the lower left panel. The upper right panel of Figure 7 shows the enormous difference in variance that we have mentioned before. The lower panels of Figure 7 make the same comparison between the logarithm of population in the age group 0 to 1 year. Considering that the distributions of RS and LT forecasts are very different (RS will be roughly normal, LT will be roughly lognormal) the logarithmic means show a larger difference than the arithmetic means.

Figure 8 shows that the variance difference propagates into other important quantities—here the old age dependency ratio(ODR), defined as the ratio of the numbers of people over 65 years of age to the number of people between ages 20 and 64. The ODR is key to the management of pension policy, in which context it is a plausible proxy for the tax rates needed to support the elderly. The first notable point in Figure 8 is the close agreement between the RS mean forecast and the LT median—this means that even though we did not force assumptions in the two methods to agree, the average values of their forecasts are often close, as is the case here for ODR and to a lesser extent for

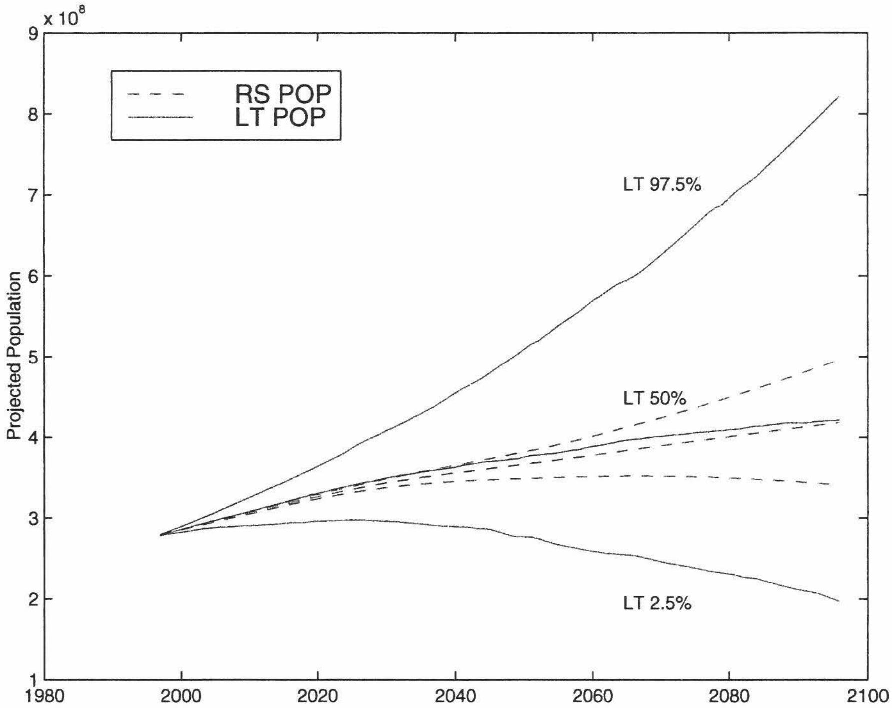


Figure 6. *Percentiles showing the 95 percent predictive distribution of total population in RS (dashed) and LT (solid) projections for the United States, using age and sex specific projection models.*

the number of births. The second point to note is the enormous difference between the uncertainty of the forecasts—a difference that has implications for any policy analysis. Finally, it is interesting that the RS scenario shows very tight bounds, similar to what the US Census Bureau was reporting about a decade ago.

The differences in forecast variance we have found above for total population, births, and ODR are found in every other projected variable that we examined, for example, the young dependency ratio (YDR). But the similarity between mean forecasts also persists across these variables.

Temporal Correlation in Forecasts

As pointed out earlier (Lee, 1999), variables in a scenario or simple RS forecast are typically highly correlated over time along trajectories, and this tends to lock the forecasted variables into highly correlated trajectories. We illustrate this observation with two examples.

Figure 9 shows the correlation between the values of e_0 (sexes combined) and TFR for both forecast methods in 2050. Note that the RS values fall in a tight grouping—indeed they are, as they should be, a bivariate normal distribution with larger variance along the fertility axis. The LT values are substantially more widely distributed: although TFR will be normally distributed in LT the distribution of e_0 is not normal.

Finally, Figure 10 shows estimates of the serial autocorrelation along trajectories of the ODR, estimated using standard time series estimators. Notice that even after 90 years the RS forecasts have a correlation of nearly 0.8 with their starting values. The serial correlation of the LT forecasts falls much more rapidly, declining to about 0.3 in about 25 years.

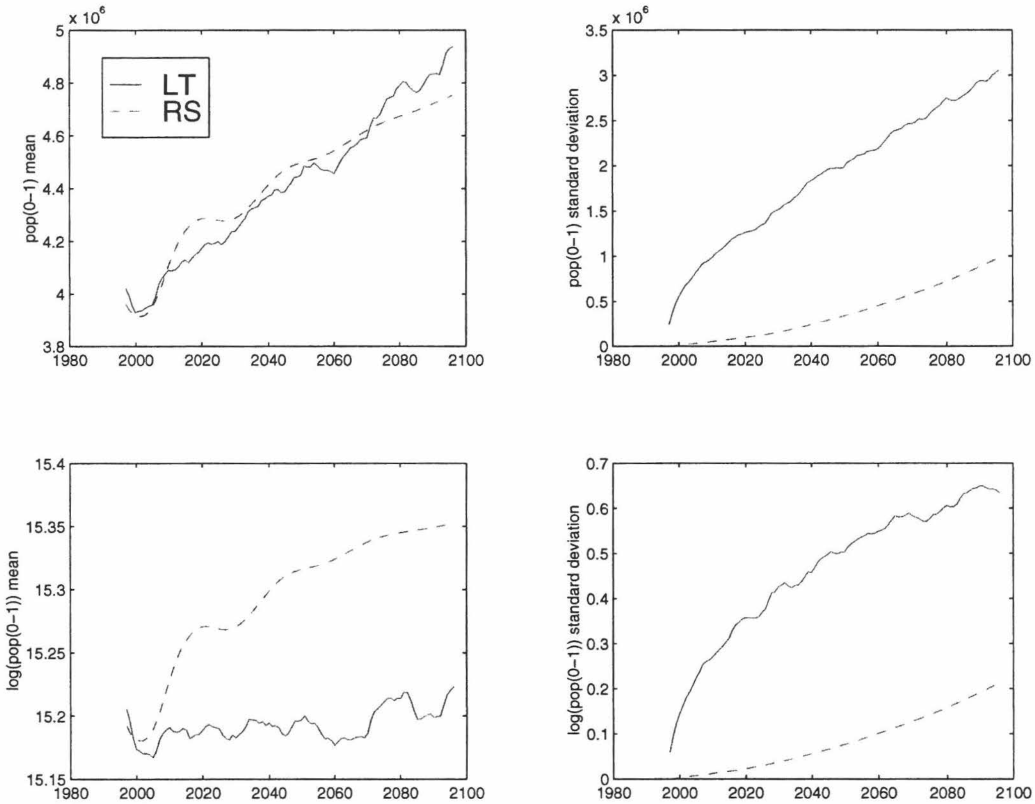


Figure 7. *Averages and variances of births forecast by RS (solid) and LT (dashed) projections for the United States. Upper panels show mean and variance for numbers of births, lower panels for the logarithm of the number of births.*

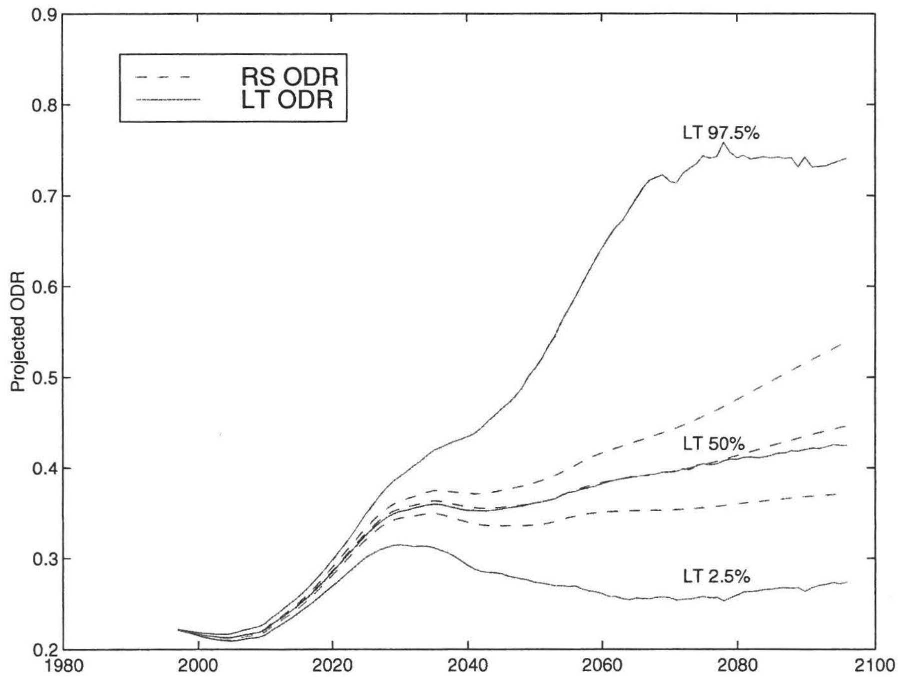


Figure 8. *Percentiles showing the 95 percent predictive distribution of the old age dependency ratio in RS (dashed) and LT (solid) projections for the United States, using age and sex specific projection models.*

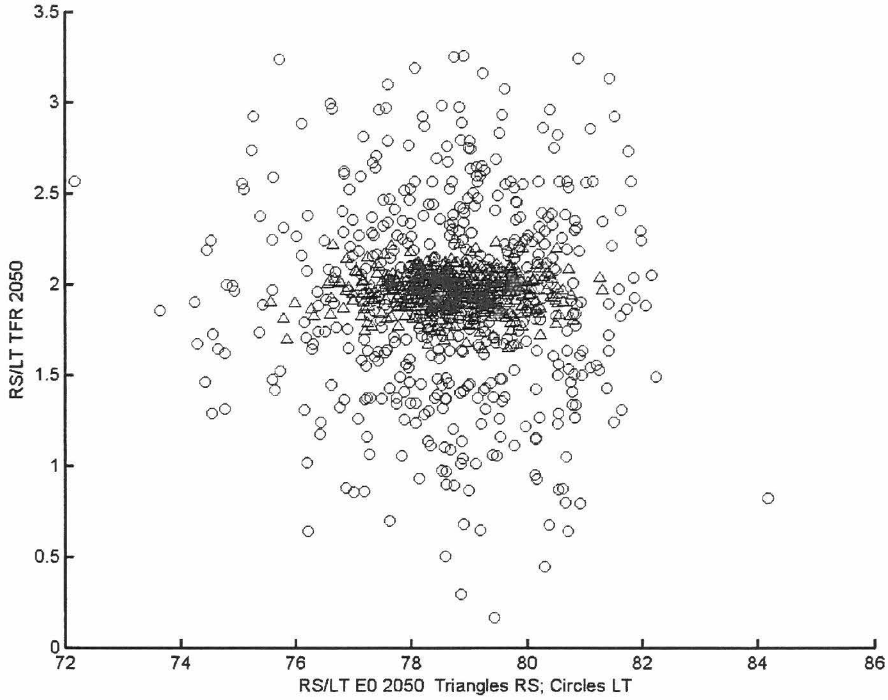


Figure 9. *Scatter plot showing simulated future values in 2050 of life expectancy e_0 and total fertility rate in RS (triangles) and LT (circles) models for the United States.*

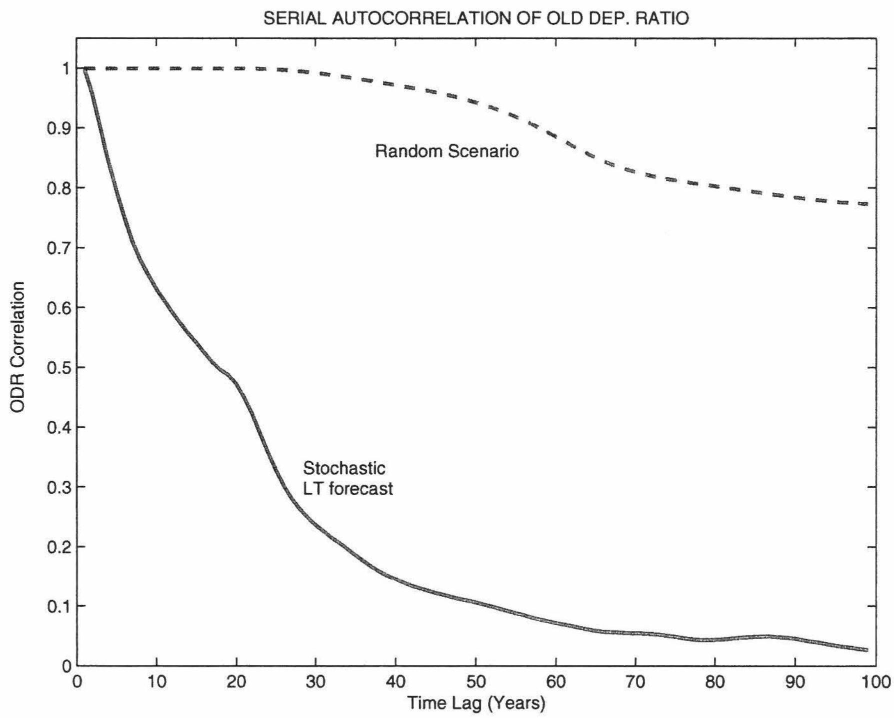


Figure 10. *Serial autocorrelation of the old-age dependency ratio as forecast by RS (dashes) and LT (solid) models for the United States.*

4 Discussion and Conclusions

We used two forecasting contexts, one simple and stylized, one complex, to compare the probabilistic forecasts that result from a simple random scenario (RS) approach and a stochastic modelling (LT) approach. Our goal has been to show what dynamic and substantive differences result from the differences in the formulation, estimation and dynamics of the models that generate the forecasts.

In our stylized scalar model, we enforced equality of both the long-run averages and long-run variances of the vital rate (the population growth rate). This means that the differences we observed are driven entirely by the different ways in which uncertainty propagates in the RS and LT formulations of forecast dynamics. We found some key differences. First, the simple RS method yields predictive distributions whose shape is determined by the forecaster's assumptions about the shape of the target distribution of the random scenarios. In contrast, LT usually generates lognormal distributions of population and its components. Second, the rate of growth of forecast variance in RS is generally quadratic with time at long forecast spans, whereas LT yields a variance that increases linearly with time. Third, at short times into the forecast, the forecast variance initially grows faster in LT than RS.

In our more complex example of the United States population, we construct a simple RS model and contrast it with an LT model that we previously developed. The long-run average vital rates in the two models are similar, but the variances in the long-run are different as dictated by the different models in the two approaches and their estimation. We noted one useful similarity and some important differences. The useful similarity is that the averages of many forecast quantities (e.g., old-age dependency ratio, total population) are quite close in the two approaches. Presumably this a reflection of the similarity in long-run average values of vital rates. One major difference we found was in the level of uncertainty as measured by the width of the predictive distributions of all quantities—the LT method yields greater uncertainty than RS over most of the forecast span with the difference increasing rapidly as we go more than say 25 years into the forecast. The driving force here appears to be the much larger variance of fertility in the LT model. The second major difference is driven by the high positive correlation along forecast trajectories in the RS model: this correlation persists over many decades, in contrast to LT in which the correlations damp out within a generation length.

We note that the Lutz *et al.* (1996) projections assume much larger variance in TFR than do we in our simple RS method; indeed, they have in 2010 for North America a range of 1.4 to 2.3 to cover 90 percent of a normal distribution. As noted, we relied on officially defined scenarios to obtain a range from an "optimistic" high to a "pessimistic" low. On the other hand, Lutz *et al.* (1996) use scenario arguments to arrive at the very different view that target fertility values lie in a much wider range. Indeed, we would argue that the use of a much more variable scenario by Lutz *et al.* supports a point we make—that simply "randomizing" a non-stochastic high-low scenario will not serve to capture the real variability of vital rates.

The RS methods we use here are a "pure" random scenario method that has since been extended by Lutz *et al.* (2001) and there are related developments by others (e.g., Pflaumer, 1992). The LT methods that we use here are not purely derived by estimating stochastic models; indeed, as we have said, they require the imposition of long-run constraints on the average of fertility. The direction of new developments is towards hybrid methods that combine elements of the simple RS methods we use here with the LT method and related stochastic models. Our goal has been to provide a stylized but sharp comparison of the consequences of the two approaches when they are not hybridized. Makers and users of forecasts should be aware of these differences and the degree to which they may influence the content and interpretation of any probabilistic forecast. We believe that hybrid methods are a logical and important direction in the business of making better and more useful forecasts.

Acknowledgements

This work was supported by grants from the National Institute of Aging and the Michigan Retirement Research Center to Tuljapurkar and to Lee, and by the Center for the Economics and Demography of Aging at the University of California at Berkeley. Li was also supported in part by the Morrison Institute of Population and Resource Studies at Stanford University.

References

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

Alho, J.M. & Spencer, B.D. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, **80**(390), 306–314.

Anderson, M., Tuljapurkar, S. & Lee, R. (1999). Chances are ... Stochastic forecasts of the social security trust fund and attempts to save it. Paper presented at the 1999 Conference on Retirement Research, Center for Retirement Research, Boston College. Available on the web site of the Michigan Retirement Research Center.

Cohen, J. (1986). Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. *Demography*, **23**(1), 105–126.

Keilman, N. (1998). How accurate are the United Nations world population projections? *Population and Development Review*, **24**(Supplement), 15–41.

Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review*, **7**(4), 579–593.

Lee, R. & Tuljapurkar, S. (1994). Stochastic population forecasts of the U.S.: Beyond high, medium, low. *Journal of the American Statistical Association*, **89**, 1175–1189.

Lee, R. & Tuljapurkar, S. (2000). Population forecasting for fiscal planning: Issues and innovations. In *Demography and Fiscal Policy*, Eds. Alan Auerbach and Ronald Lee, pp. 7–57. Cambridge University Press.

Lee, R. (1999). Probabilistic Approaches to Population Forecasting. In *Population Development Review, 1999 Rethinking Population Projections*, Eds. W. Lutz, J. Vaupel and D. Ahlburg, supplement to **24**, pp.156–190.

Lutz, W., Sanderson, W. & Scherbov, S. (1996). Probabilistic population projections based on expert opinion. In *The Future Population of the World. What Can We Assume Today?*, Ed. W. Lutz, pp. 397–428, Revised Edition. London: Earthscan.

Lutz, W., Sanderson, W. & Scherbov, S. (1997). Doubling of world population unlikely. *Nature*, **387**, 803–805.

Lutz, W., Sanderson, W. & Scherbov, S. (2001). The end of world population growth. *Nature*, **412**, 543–545.

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

Pflaumer, P. (1992). Forecasting US population totals with the US Box–Jenkins approach. *International Journal of Forecasting*, **8**(3), 329–338.

Stoto, M. (1983). The accuracy of population projections. *Journal of the American Statistical Association*, **78**(381), 13–20.

United States Social Security Administration. (1996). The 1996 OASDI Trustees Report. Government Printing Office.

Résumé

Les projections de population probabilistes sont utiles car leur façon de répresenter l'incertitude est utilisable quantitativement. Une approche possible (que nous appellerons LT) consiste à utiliser des données historiques pour estimer de modèles stochastiques (c'est à dire des modélisations en séries temporelles) des taux démographiques et, à partir de là, faire des prévisions. Une autre approche (que nous appellerons RS) trouve sa source dans l'utilisation de sorte de scénarios aléatoires. Nous considérons ici sa variante la plus simple, dans laquelle on se sert de l'opinion d'experts pour construire une distribution de probabilité des taux démographiques finaux, puis on projette une trajectoire lissée dans le temps. Au moyen d'arguments analytiques et d'exemples, nous montrons plusieurs différences importantes entre ces méthodes: les corrélations au sein des séries lors des prévisions sont bien plus faibles dans les modèles LT; la variance des taux démographiques (en particulier celle de la fertilité) est bien plus élevée dans la modélisations LT que dans les modélisations RS, qui se fondent sur les scénarios des experts publics; la trajectoire des modèles LT est bien plus irrégulière qu'avec LS; les intervalles de confiance dans les modèles LT ont tendance à augmenter plus vite au cours de la période de prévision. De nouvelles versions des modèles RS ont été développées et réduisent ou éliminent certaines des différences.

Mot-clés: Prévision probabiliste; Prévision démographique; Trajectoire; Taux démograpiques; Scénario; Scénarios aléatoires; Ratio de dépendence.

[Received May 2003, accepted January 2004]

International Statistical Review (2004), **72**, 2, 201–208, Printed in The Netherlands

Developing Official Stochastic Population Forecasts at the US Census Bureau

John F. Long and Frederick W. Hollmann

U.S. Census Bureau, Washington, USA
E-mail: John.F.Long@census.gov; Frederick.W.Hollmann@census.gov

Summary

The U.S. Census Bureau is approaching a critical decision regarding a major facet of its methodology for forecasting the United States population. In the past, it has relied on alternative scenarios, low, medium, and high, to reflect varying interpretations of current trends in fertility, mortality, and international migration to forecast population. This approach has limitations that have been noted in the recent literature on population forecasting. Census Bureau researchers are embarking on an attempt to incorporate probabilistic reasoning to forecast prediction intervals around point forecasts to future dates. The current literature offers a choice of approaches to this problem. We are opting to employ formal time series modeling of parameters of fertility, mortality, and international migration, with stochastic renewal processes. The endeavor is complicated by the administrative necessity to produce a large amount of racial and Hispanic origin detail in the population, as well as the ubiquitous cross-categories of age and sex. As official population forecasts must meet user demand, we are faced with the added challenge of presenting and supporting the resulting product in a way that is comprehensible to users, many of whom have little or no comprehension of the technical forecasting literature, and are accustomed to simple, deterministic answers. We may well find a need to modify our strategy, depending on the realities that may emerge from the limitations of data, the administrative requirements of the product, and the diverse needs of our user community.

Key words: Stochastic forecasting; Racial disaggregation; Hispanic origin; Composition effects; User demand.

1 Introduction

In the early 1980's, the U.S. Census Bureau's population projections program was in a state of flux. For the previous three decades, the Census Bureau projections had been chasing the increases and decreases in American fertility during the baby boom and subsequent bust. All of its efforts had gone into devising projections that could cope with the rapidly changing fertility—on the one hand collecting data on birth expectations and emphasizing cohort completion rates and on the other hand defending itself by making the argument that the results were merely projections (of current trends or alternative assumptions) and should not be taken as forecasts. The alternative series were labeled by numbers or letters rather than high, medium or low and on some occasions an even number of series were projected so that no one could be considered the most-likely series.

By 1980, fertility rates had entered a period of relative stability while major changes in the other components of population change, mortality and immigration had become more volatile. Both within and outside the Census Bureau, interest in the error structure of previous population projections raised the issue that official projections were indeed used as forecasts and needed to be treated as such (Keyfitz, 1981; Stoto, 1983). Moreover, even though most users wanted a single forecasted number, it was important to provide an explicit range around the projected numbers that represented the total variability of the population and not just variability associated with fertility uncertainty.

In an exchange of correspondence that year between Nathan Keyfitz at Harvard, William Kruskal at Chicago, and staff members of the U.S. Census Bureau, the importance of the error structure in population projections was discussed. Should official population projections be recast as forecasts with true error ranges that represented an historical empirically determined error range or should they take the more modest step of explicitly designating the alternative series as low and high, expanding the uncertainty analysis to mortality and immigration as well as fertility, and increasing the level of detail and sophistication of the projection methodology. Given the state of the theory, methods, data series, and computational power available at that time, the Census Bureau decided to take the latter course and embarked on a multi-year project to expand the detail and sophistication of the national population projections while retaining an essentially deterministic approach.

The 1988 set of national population projections consisted of 30 series, 27 of which represented the combinations of high, middle, and low assumptions around each of the three components of population change (U.S. Census Bureau, 1988). The resulting projections permitted sensitivity analysis of the independent effects of fertility, mortality, and immigration on the age structure of the population (Long, 1989). This method attempted to address the general question by defining "high" and "low" series of the major inputs, and defining "highest" and "lowest" series projections by jointly assuming the extreme values for each component. The projections were based on classic cohort-component methodology; hence (for example) the highest series was composed of the high-level assumptions of fertility and net migration to the United States, and the low-level assumption for mortality. These high and low assumptions were not probabilistic, but represented a qualitative judgment of the forecasters about "useful" upper and lower limits. In hindsight, the range often proved too small in the early years of the projection and perhaps too large for the later decades of the projection, as noted by Lee & Tuljapurkar (1994) and others. The range of high and low series offered a limited way to view the uncertainty of projected total population—but did not adequately represent the variability of internal ratios such as the proportion of the population in working ages.

Other expansions in the sophistication and detail of the projections continued—resulting in an improved handling of the dynamics of mortality, incorporation of short-range time-series ARIMA forecasts of fertility rates, disaggregation of net immigration to various types of in-migration flows and separate emigration rates for the native and foreign-born populations, expansion of race categories to 8 combinations of race and ethnicity, and the expansion of age detail to single years of age to 100 plus.

Today, we find ourselves at a crossroads again. Advancements in the theory and methods of population forecasting (Ahlburg & Lutz, 1998) combined with the expanded computing power make probabilistic population projections feasible. The remaining question is "Can they be produced and distributed to the government and to the public in such a way as to serve the uses to which official population projections have been used in the past?" The answer involves both technical and policy issues.

2 Choice of Probabilistic Methods

One of the first choices to be made is whether to pursue an expert-based probabilistic method (Lutz, Sanderson & Scherbov, 1998) or a stochastic forecasting approach (Lee, 1998). Although it might seem at first glance that the expert-based approach would be more consistent with the U.S. Census Bureau tradition of using assumptions to set alternative variants of population change (Long, 1984), there are several reasons to begin the foray into alternative probabilistic projections with a stochastic forecasting approach.

Most importantly, the Census Bureau projections have always depended heavily on the development of a disaggregated set of data about the current and historical data on age-sex-race specific rates of fertility, mortality, and (recently) immigration. The empirical data base supplies the key grounding for initial conditions and short-term trends and even provides the database on which to validate the plausibility of alternative assumptions. Stochastic forecasts (especially those based on statistical time series analysis) provide the best basis to continue to ground official forecasts in the current and historical data.

Secondly the Census Bureau has more experience in statistical time series analysis than it does in developing formal models of expert opinion. Several of the Census Bureau's projections in the 1990's used short-term forecasts of fertility rates based on statistical time-series as one aspect of the projection method. Although the Census Bureau has had scattered workshops on the future course of fertility, there is no formal expert panel or advisory committee for setting assumptions of future levels—much less a formal methodology for determining the uncertainty in those estimates.

Finally, stochastic forecasts with their ability to permit canceling of errors over time due to the less than perfect autocorrelation of errors appear to give the best chance of providing error ranges that are sufficiently wide in the short run without having excessively large error ranges in the very long run.

The approach adopted in our current attempt is that of a stochastic forecast, based on an *ex post* analysis of historical trends in fertility, mortality, and net international migration. The forecasts are to be developed using time series procedures derived from Box–Jenkins (ARIMA methods). The approach follows most closely the approach favored by Ronald Lee and his colleagues (e.g., Lee & Tuljapurkar, 1994), although we are attempting alternative approaches to defining the parameters of the components of change. We are treating fertility, mortality, and migration separately. We define parameters for each component, and employ a vector autoregression (ARIMA) to forecast them. Regardless of how parameters are defined, a fundamental requirement of the process is a substantial time series of data, with some demographic detail consistent over time.

3 Demographic Disaggregation

A fundamental aspect of the Census Bureau projections has been the role of disaggregated calculations for specific demographic subpopulations with different trajectories for each race and ethnic group. In fact, it is this disaggregation combined with the marked differential in fertility by race in previous decades and the current marked differential in fertility by Hispanic origin that has been one of the major factors in driving demographic change in the Census Bureau's recent projections.

Racial detail has been an attribute of United States population projections almost as long as they have been produced. Early projections were often simply disaggregated into a white/non-white racial dichotomy. Later projections expanded the racial categories to white, black, and other. In 1990, dichotomous Hispanic origin was introduced to the matrix. The number of racial and ethnic categories increased to eight (Hispanic origin crossed by four race groups: white, black, Asian and Pacific Islander, and American Indian or Alaskan Native) in our last projection series released in 2000. An important technical aspect of this practice is that the evolution of forecast population is subject to racial composition effects, meaning that forecast population change is influenced by the racial and ethnic composition of the population in the presence of differential rates of fertility and mortality for the various groups. Hence, the projection of racial and ethnic categories is inseparable from the projection of the population by age and sex.

The base population for the current forecasts will be the 2002 estimate of the U.S. national population (consistent with the 2000 census results) disaggregated by single years of age, sex, and three race/ethnic categories. To reduce complexity, we are confining our stochastic projections to three categories, 1) Hispanic origin, of any race, 2) non-Hispanic, reporting black alone, and 3) all others. By forecasting these three groups stochastically, we determine the total population by age and sex and its forecast uncertainty. More detailed racial categories can then be projected deterministically within the three large ones (see following section on expanding detail for by race and Hispanic origin). This three-category distribution has the following desirable properties, from the standpoint of population forecasting.

1. The categories are large, so there should be no difficulty with projecting single-year age distributions too "thin" to produce good results.
2. Differential fertility and mortality among the three categories can be documented from historical data, although there is some potential for bias in the comparisons. Hispanic origin was not directly affected by the change to "check all that apply" in race definition. The non-Hispanic Black population was affected by the change in race definition, but relatively marginally.
3. Differences among the categories with respect to demographic rates are substantial. The age pattern of childbearing and excess mortality (notably for infants) for the non-Hispanic Black population are distinctive. Fertility continues to be somewhat higher for the Hispanic population than for most non-Hispanic racial categories.

As the decennial census is seen as defining the population universe for the projections, we have allowed no uncertainty in the census population. Associated with this assumption is the proviso that all population forecasts, including any estimates of uncertainty, are viewed as representing what Census 2000 would have enumerated at a future date.

Parameters of the Components of Change

In this method, rates for each of the major components of population change—fertility, mortality, and international migration—will be forecasted independently, and we assume no time series correlation among the three major components. The independent variable of the vector autoregression in each case is a vector of parameters that summarize the demographic detail in the historical series and that can be used to reproduce the detail in the projected series. The large amount of detail in our planned projections places considerable stress on the need to summarize the inputs to the time series models through parameters.

In the case of fertility, we plan to base our parameters on the mean and variance of the age pattern of fertility intensity by parity, as developed by Kohler & Ortega (2001). In this model, fertility is measured by the age schedule of rates of fertility for women of a given parity. While the schedule is computed from cross-sectional (as opposed to cohort-based) information, the rates are adjusted for the effects of tempo changes in the fertility of cohorts of women that produce bias when synthetic cohort information is analyzed. The fertility intensity schedule is summarized by its mean age and the variance of age, for each parity category of women. In our application, we plan to calculate these parameters for the first four parities of women. From the previous section, we must carry this out for three racial and ethnic categories, yielding a total of 12 categories for which mean and variance must be estimated and forecasted, or 24 parameters of fertility. The determinants of secular fertility changes in the United States tend to be expressed across racial and ethnic lines, but not in "lock step", and there is undoubtedly some time series correlation among the behavior of women of different parity. It must therefore be assumed that much of the time series covariance among the 24 parameters is non-trivial.

This analysis is somewhat challenged by problems encountered in the data, which come from three sources. For numerators of rates, we rely on birth registration data supplied initially from reports of local administrators to state health departments. For the count of women by age, sex, and race, and Hispanic origin, we rely on administrative population estimates based on decennial census data. As these estimates do not provide women by parity, we measure the distribution of women by parity using data from the Current Population Survey (CPS), a monthly survey used to monitor the

U.S. labor force that has included an annual supplementary questionnaire on childbearing. The time series of historical data is limited to 20 years, from 1980 to 1999, largely because of the lack of population data on Hispanic origin prior to 1980. Even in the period since 1980, the number of U.S. states reporting Hispanic origin in birth registration data did not embrace the entire country until 1993. Hence, the estimation of fertility for the Hispanic origin population must be imputed from the experience of reporting states, resulting in some spurious discontinuities in the series. The fertility supplement to the CPS, used to estimate parity of women, is also lacking for a few of these years because of resource limitations, so the data must be interpolated from neighboring years. Finally, the limitation of the series to 20 years means that we are unable to calculate a full variance-covariance matrix of the 24 parameters, so that some of the covariance assumptions must be postulated.

For mortality, we plan to develop a system of parameters after the model developed by Lee & Carter (1992). There is a possibility that we will need to increase the number of parameters to account for perturbations in the age pattern of mortality that do not conform to the Lee–Carter model. Be this as it may, we do not anticipate that the number of parameters will pose as much of a problem with mortality as it did with fertility. As with fertility, we will rely on a data series consisting of 20 years of death registration data and population estimates. The limitations to the data on deaths by Hispanic origin are more severe than with fertility, with a substantial portion of the population in non-reporting states until the late 1990s.

For international migration, the development of parameters will be somewhat cruder, consisting of a level of net migration and an indicator of central tendency for the age distribution, but disaggregated by groups of country of origin. The necessity to view international migration by country of origin is important for two reasons. First, the trend in migration to and from the United States is very much governed by the source of immigration, so disaggregating by groups of countries is essential to the interpretation of historical trends. Second, country of birth is an effective indicator of the racial and ethnic composition of the net migration flow, which can be determined after the forecast trends have been produced. For a historical series, we rely on 30 years of legal immigration data from the U.S. Immigration and Naturalization Service. This series requires considerable interpretation, because legal immigration to the United States does not embrace all residential movement across U.S. frontiers. Indirect estimates of undocumented migration and emigration from the United States have been developed, but generally not as an annual series; hence, assumptions must be made regarding the impact of these components on the level and the degree of fluctuations in the annual series of net migration to the United States, which determine the degree of uncertainty in the forecast series.

Generating Models and Realizations

The actual implementation of this modeling strategy in producing population forecasts is comparatively routine, but very computer-intensive. Once the vector autoregressive models for the three major components are developed, the parameters of fertility, mortality, and international migration are forecast via a stochastic renewal process. Multiple forecast series, or "realizations" are produced for the three vectors of parameters, and they are combined to form the basis for projections of population. The full schedules of age-sex-specific fertility and mortality rates, and the age-sex schedule of net international migration are generated from the parameters. Each population realization is developed by the cohort-component method, and all the elements that will compose the ultimate population forecast product are retained for each realization. Central values and prediction intervals of total population, age and sex categories, and summary demographic indicators can be determined from the realizations. Due to the large number of parameters, it is anticipated that the number of realizations will be very large, so we must prepare to abandon the individual realizations once the product demand has been fully assessed.

Expanding the Race and Hispanic Detail

Even with the presence of race and Hispanic detail in the forecasting model, the results do not provide the level of output detail for race and ethnicity required of the Census Bureau's official population projections. As a result a post-hoc method must be designed to provide the required detail.

In the 2000 census the four racial categories were expanded to five, as the Asian and Pacific Islander category was divided into two categories, Asian, and Native Hawaiian and Other Pacific Islanders. The instruction requiring a single race response was modified to an instruction of "check all that apply". The possibility of responses covering more than one of the five major categories increased the number of mutually exclusive response categories to 31, which, when crossed with dichotomous Hispanic origin yields 62 racial and ethnic categories to project, compared to 8 in previous series.

Although Census Bureau forecasts do not need to show data for all 62 possible race and Hispanic origin combinations, they do need to show the following two summary classifications: a) "minimum" value for the five racial groups, composed of people who would report a single race only, with a sixth category defined by all those reporting two or more races, and b) "maximum" values for the five groups, composed of all who would report the given race either alone or in combination with another race. Both of these distributions can be shown for the Hispanic population and for the non-Hispanic population. In order to generate these two distributions from mutually exclusive categories, it is necessary to forecast the entire matrix of 62, each by age and sex.

In the past, our strategy for developing projections of the 8 groups has been to treat them as separate populations in a cohort component framework, with the total population being a simple aggregation of the eight groups. Of paramount importance to this approach is the working assumption that the 8 categories are closed with respect to childbearing. Fertility rates can thus be applied to projected cohorts of women to produce the population of the youngest age group within the racial and ethnic category. Under the new definition, this simplifying assumption is not plausible, since the multiple-report (2 or more races) categories owe their existence in part to racial differences between mother and child brought about by intermarriage.

This complication along with the lack of any historical administrative time series for any demographic component of this detailed racial distribution renders impossible a stochastic treatment of all 62 racial and ethnic categories. Consequently we settled on the simplified three-category race/ethnic distribution described in our methodology above.

Then the specific race/ethnic distribution of children is separately modeled from data on the racial and ethnic composition of family households from the 2000 census. As there is no historical information whatever on the trend in the relationship of race of child to race of parents, we make a simplifying assumption that ignores any uncertainty in the specification of a child's race by race of parent. With this extension to the basic stochastic forecast, we are able to obtain much of the demographic detail that users have come to expect in our population forecasts.

4 Policy Issues and Users Demands for Official Stochastic Forecasts

Once the Census Bureau produces a set of official stochastic forecasts, our next step will be to lead users and governmental officials to accept and use them. There are several hurdles in this step that will require careful attention. Official projections or forecasts are expected to have several characteristics.

First, many if not most users are interested in the single set of numbers that represents the most likely course of population change. If the methodology used to create stochastic forecasts does not also provide the best estimator of the future population then those users will be ill-served. Users will need to be cautioned about the change from a population variant approach. In prior projections, there was internal consistency within a series—if not across series. Thus a series with a high fertility rate would be associated with a given mean age, dependency ratio, etc. With the stochastic method, this would no longer be the case since the forecast is for the distribution and even the set of medial values does not represent a coherent scenario.

As we have already discussed, one of these characteristics is explicit detail about policy-relevant demographic groups. Racial, Hispanic origin, age, and sex detail are expected and will be provided by our methods. Some other detail that has been present in recent projections will be missing from this series. Separate projections of the native and foreign born populations that were a feature of our last projections will not be possible in this series given the lack of data for a stochastic method.

Population forecasts must be for a long time horizon, irrespective of the limitations of the baseline data series that are used to specify models of uncertainty. These forecasts will go to at least 2080 and possibly to 2100. The estimation of uncertainty in the population series must be comprehensive. For example, it is not useful to restrict our analysis of uncertainty to only those components of change, or those population groups for which it can be readily estimated. Uncertainty in the resulting population will be interpreted as though it were fully measured, even if it is not. On the other hand, wide ranges for the long-run forecasts may cast doubt on the utility of the projections in the short-run.

User demand tends to be shaped by past practice. Many of our users in other state and federal agencies, for example, employ U.S. Census Bureau projections of the national population to project related social and economic series. Their methodology is driven by past practice. This requires that all of the elements of our "preferred" middle series that were available with past deterministic projections be present, irrespective of our methodology.

Population forecasts must be comprehensible to a wide variety of user interests, hence there tends to be resistance to expanding the complexity of the product in ways that are not supported by mandate. There is complexity in the production of stochastic forecasts but the results can be presented relatively simply to an audience that is already accustomed to sample survey variance. The problem may in fact be that users will too easily make the analogy with survey variance and put more trust in the variance estimates than they deserve at this level of development.

5 Conclusions

Despite all these cautions, now would seem to be the time to try to produce a set of official stochastic population forecasts for the United States. The needed data sets are fairly complete and accurate; the major components do not seem to be undergoing substantial structural changes; and the methods and computer capability can handle the extensive calculations required. It must be done with caution and there must be alternative plans to modify results if the time series approach leads to unreasonable median forecasts. In fact, a fully deterministic fall-back position is required if the methodological development fails to be completed in time to produce the required set of projections by the end of 2003. Nonetheless, unlike the early 1980's, the Census Bureau is now attempting to produce official stochastic population forecasts.

References

Ahlburg, D.A. & Lutz, W. (1998). Introduction: The Need to Rethink Approaches to Population Forecasts. In *Frontiers of Population Forecasting*. Supplement to *Population and Development Review*, **24**, 1–14, Eds. W. Lutz, J.W. Vaupel and D.A. Ahlburg.

Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review*, **7**(4), 579–593.

Kohler, H.P. & Ortega, A. (2001). *Period Parity Progression Measures with Continued Fertility Postponement: A New Look at the Implications of Delayed Childbearing for Cohort Fertility*. Max Planck Institut für Demografische Forschung, Working Paper No. 2001-001 (January).

Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* **87**(419), 659–675.

Lee, R. & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low.

Journal of the American Statistical Association, **89**(428), 1175–1189.
Lee, R.D. (1998). Probabilistic approaches to population forecasting. *Population and Development Review*, **24** (supplement), 156–190.
Long, J.F. (1984). U.S. national population projection methods: A view from four forecasting traditions. *Insurance: Mathematics and Economics*, **3**, 231–239.
Long, J.F. (1989). The relative effects of fertility, mortality, and immigration on projected population age structure. In *Future Demographic Trends in Europe and North America*, Ed. W. Lutz. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Lutz, W., Sanderson, W.C. & Scherbov, S. (1998). Expert-based probabilistic population projections. *Population and Development Review*, supplement, **24**, 139–155.
Lutz, W., Vaupel, J.W. & Ahlburg, D.A. (Eds.) (1998). *Frontiers of Population Forecasting*. Supplement to *Population and Development Review*, **24**.
Stoto, M. (1983). The accuracy of population projections. *Journal of the American Statistical Association*, **78**, 13–20.
U.S. Bureau of the Census (1988). Projections of the Population of the United States by Age, Sex, and Race: 1983 to 2080. *Current Population Reports*, Series P-25, No. 1018.

Résumé

Le Bureau du Census aux Etats-Unis se prépare à une décision critique concernant un aspect primordial de sa méthodologie pour projeter la population du pays. Dans le passé, celle-ci s'est appuyée sur des scénarios alternatifs, inférieur, moyen et supérieur, pour refléter les interprétations variables des tendances courantes de fécondité, de mortalité et des migrations internationales afin de projeter la population. Cette approche a des limites qui ont été soulignées dans la littérature récente sur les projections de population. Les chercheurs du Bureau du Census se lancent dans une tentative d'incorporer le raisonnement probabiliste pour projeter des intervalles de prédiction autour de valeurs prévues à des dates futures. La littérature courante offre un choix d'approches à ce problème. Nous faisons l'option d'employer la modélisation formelle de séries temporelles de paramètres de fécondité, de mortalité et de migration internationale avec des processus de renouvellement stochastiques. Le travail est compliqué par la nécessité administrative de produire un grand nombre de détails sur l'origine raciale et hispanique de la population, ainsi que la répartition traditionnelle par âge et sexe. Les projections de population officielle devant répondre à la demande des utilisateurs, nous sommes confrontés au défi supplémentaire de présenter et d'appuyer les résultats d'une manière compréhensible pour les utilisateurs; beaucoup d'entre eux en effet ont une faible compréhension de la littérature technique de projection et sont habitués à des réponses simples et déterministes. Il se peut que nous soyons amenés à modifier notre stratégie en fonction des réalités qui peuvent ressortir des limites des données, des exigences administratives du produit et des besoins divers de notre communauté d'utilisateurs.

[Received February 2003, accepted November 2003]