



Perspectives on Mortality Forecasting

II. Probabilistic Models

Social Insurance Studies | No. 2

Editor
Nico Keilman

Swedish Social Insurance Agency

Series Editor
Edward Palmer

The Swedish Social Insurance Agency (*Försäkringskassan*) has a long standing commitment to promote research and evaluation of Swedish social insurance and social policy. The Social Insurance Agency meets this commitment by commissioning studies from scholars specializing in these areas. The purpose of the series Social Insurance Studies is to make studies and research focusing on important institutional and empirical issues in social insurance and social policy available to the international community of scholars and policy makers.



Published by: Försäkringskassan, Swedish Social Insurance Agency, 2005

Series Editor: Edward Palmer

Issue Editor: Nico Keilman

© The authors

The responsibility for the contents of the studies published in this series lies solely with the authors, and the views expressed in individual studies do not necessarily represent those of the leadership of the Swedish Social Insurance Agency.

Social Insurance Studies can be ordered through the homepage at www.forsakringskassan.se, by phone: +46 8 556 799 29 or by fax: +46 8 718 29 90, and cost SEK 120 excluding VAT and postage and packing.

Swedish Social Insurance Agency
Adolf Fredriks Kyrkogata 8
SE-103 51 STOCKHOLM, Sweden
Telephone: +46 8 786 90 00
E-mail: huvudkontoret@forsakringskassan.se

Graphic Production: Kristina Malm
Cover Design: Jenny Övrebö
Printed by Bodoni Tryck AB, Sundbyberg, 2005

ISBN 91-7500-326-0
ISSN 1651-680x

Preface

Mortality projections are an essential input for projections of the financial development of pension schemes. Governments and insurance companies all over the world rely on good mortality projections for efficient administration of their pension commitments. Ideally, the expected value of the difference between outcomes and projections would be close to zero. In practice, during recent decades, demographers have continually underestimated improvements in life expectancy for persons 60 and older. The demographic models used in projecting mortality are usually based on statistical modeling of historical data. The question is, is it possible to bring the results of mortality modeling closer to the ideal, and if so, what do demographers need to do to achieve this result?

This is the question that provided the impetus for forming the Stockholm Committee on Mortality Forecasting. The Swedish Social Insurance Agency (formerly National Social Insurance Board, RFV) is the national agency in Sweden responsible for providing a financial picture of Sweden's public pension system. The Swedish Social Insurance Agency has a long-standing interest in the development of modeling of pension schemes and participates actively in the international dialogue among experts in this area. The Stockholm Committee on Mortality Forecasting was created by RFV to bring together scholars from different disciplines working on issues in projecting mortality. The aim of the Committee is to survey the state of the art and to provide an impetus for the advancement of knowledge and better practice in forecasting mortality.

This is the second volume in a series presenting papers from workshops on mortality organized by the Stockholm Committee on Mortality Forecasting. The chapters focus on probabilistic (also labeled as stochastic) forecasts, in other words forecasts in which uncertainty has been quantified. Given a history of sizable forecasting errors, the first paper, by Nico Keilman, addresses the question of why demographic forecasts are uncertain. In the second paper Juha Alho outlines the statistical background of uncertain events and forecasts of these. In the third paper Maarten Alders and Joop de Beer sketch the approach taken by Statistics Netherlands in their stochastic forecast of mortality, while in the fourth paper Shripad Tuljapurkar presents a model for mortality analysis and forecasting that has proven to be feasible for probabilistic forecasts. He gives illustrations of US and Swedish mortality, and discusses also possible implications of uncertain mortality for future pension expenditures.

Probabilistic population forecasting is a relatively new development in demography and forecasting. The topic is of particular interest for the performance of pension systems in the future. As editor of *Social Insurance Studies*, it is my hope that the published proceedings of the Stockholm Committee on Mortality Forecasting will contribute to improve our understanding of the processes underlying increasing longevity.

Edward Palmer

Members of the Stockholm Committee on Mortality Forecasting

Professor Edward Palmer, Committee Chairman
Uppsala University and Swedish Social Insurance Agency, Sweden

Professor Tommy Bengtsson, Committee Secretary
Lund University, Sweden

Professor Juha M. Alho
University of Joensuu, Finland

Professor Kaare Christensen
University of Southern Denmark, Odense

Professor Nico Keilman
University of Oslo, Norway

Professor James W. Vaupel
Director of Max Planck Institute for Demography, Rostock, Germany

Contents

Erroneous Population Forecasts

Nico Keilman 7

Remarks on the Use of Probabilities in Demography and Forecasting

Juha M. Alho 27

An Expert Knowledge Approach to Stochastic Mortality Forecasting in the Netherlands

Maarten Alders and Joop de Beer 39

Stochastic Forecasts of Mortality, Population and Pension Systems

Shripad Tuljapurkar 65

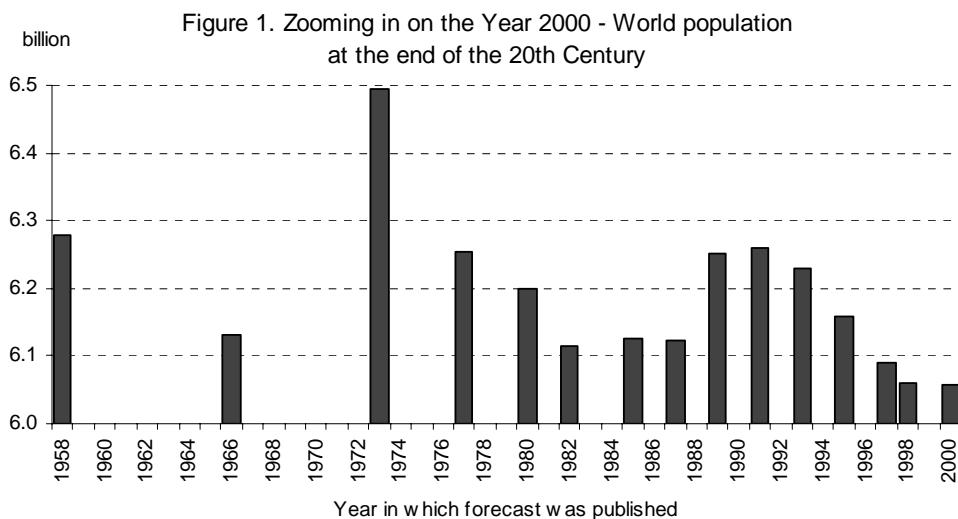
Erroneous Population Forecasts

Nico Keilman

Professor of Demography, Department of Economics at University of Oslo, Norway

1 Forecast accuracy

World population in the year 2000 was 6.09 billion, according to recent estimates by the United Nations (UN 2005). This number is almost 410 million lower than the year 2000-estimate that the UN expected in 1973. The UN has computed forecasts for the population of the world since the 1950s. Figure 1 shows that the calculations made in the 1980s were much closer to the current estimate than those published around 1990. Subsequent forecasts for the world population in 2000 show an irregular pattern: apparently, in 1973 and around 1990 it was rather difficult to predict world population size and much less so in the mid-1980s.



At first sight, the relative differences in Figure 1 appear small. The highest forecast came out in 1973. That forecast numbered 6.49 billion, only six per cent higher than the current estimate of 6.09 billion. However, the difference

is much larger in terms of population *growth*. The 1973 forecast covered the period 1965–2000. During those 35 years, a growth in world population by 3.20 billion was foreseen. According to the current estimate, the growth was 16 per cent lower: only 2.7 billion persons.

An important reason for lower population growth is that the world's birth rates fell stronger than previously thought. Thirty years ago, the UN expected a drop in total fertility by 1.4 children between the periods 1965–1970 and 1995–2000: from 4.7 to 3.3 children per woman on average. Recent estimates indicate that fertility initially was higher than previously thought, and that it fell steeper than expected in that thirty-year period, from 4.9 to 2.8.

Accuracy statistics of the type given here are important indicators when judging the quality of population forecasts. Other aspects, such as the information content (for instance, does the forecast predict only total population, or also age groups?) and the usefulness for policy purposes (for instance, does the predicted trend imply immediate policy measures?) are relevant as well. Nevertheless, the degree to which the forecast reflects real trends is a key factor in assessing its quality, in particular when the forecast is used for planning purposes. For example, imagine a forecast, for which the odds are one against two that it will cover actual trends. This forecast should be handled much more cautiously than one that can be expected to be in error only one out of five times.

The purpose of this chapter is to give a broad review of the notions of population forecast errors and forecast accuracy. Why are population forecasts inaccurate? How large are the errors involved, when we analyse historical forecasts of fertility, mortality, and the age structure? Moreover, how can we compute expected errors in recent forecasts? We shall see that *probabilistic* population forecasts are necessary to assess the expected accuracy of a forecast, and that such probabilistic forecasts quantify expected accuracy and expected forecast errors much better than traditional deterministic forecasts do. The chapter concludes with some challenges in the field of probabilistic population forecasting.

The focus in this chapter is on population forecasts at the *national* level, computed by means of the *cohort component method*. I have largely restricted myself to national forecasts, because most of the empirical literature on forecast errors and forecast accuracy deals with forecasts at that level. Notable exceptions, to be discussed below, are analyses for major world regions by Lutz *et al.* (1996, 2001), and for all countries in the world by the US National Research Council (NRC 2000). The empirical accuracy of sub-national population forecasts has been evaluated since the 1950s (Smith *et al.*

2001), but the *expected* accuracy of such forecasts is largely uncharted terrain, *cf.* the concluding section. I focus on the cohort component method of population forecasting, because this method is the standard approach for population forecasting at the national level (Keilman and Crujisen 1992). Most of the empirical evidence stems from industrialized countries, although findings for less-developed countries will be mentioned occasionally.

Various terms are in use to express accuracy, and lack thereof. I shall use inaccuracy and uncertainty as equivalent notions. When a forecast is accurate, its errors are small. Forecast errors are a means of quantifying forecast accuracy and forecast uncertainty. *Empirical* errors may be computed based on a historical forecast, when its results are compared with actual population data observed some years after the forecast was computed. For a recent forecast, this is not possible. In that case, one may compute *expected* errors, by means of a statistical model.

2 Why population forecasts are inaccurate

Population forecasts are inaccurate because our understanding of demographic behaviour is imperfect. Keyfitz (1982) assessed various established and rudimentary demographic theories: demographic transition, effects of development, Caldwell's theory concerning education and fertility, urbanization, income distribution, Malthus' writings on population, human capital, the Easterlin effect, opportunity costs, prosperity and fertility, and childbearing intentions. He tried to discover whether these theories had improved demographic forecasting, but his conclusion was negative. Although many of the theories are extensively tested, they have limited predictive validity in space and time, are strongly conditional, or cannot be applied without the difficult prediction of non-demographic factors. Keyfitz' conclusion agrees with Ernest Nagel's opinion from 1961, that "... (un)like the laws of physics and chemistry, generalizations in the social sciences ... have at best only a severely restricted scope, limited to social phenomena occurring during a relatively brief historical epoch with special institutional settings." Similarly, Raymond Boudon (1986) concluded that general social science theories do not exist – they are all partial and local, and Louis Henry (1987) supports that view for the case of demography. Applied to demographic forecasting, this view implies that uncertainty is inherent, and not merely the result of our ignorance. Individuals make unpredictable choices regarding partnership and childbearing, health behaviour, and migration. Note that the views expressed by Nagel and Boudon are radically different from Laplace's view on chance and uncertainty: "Imagine ... an intelligence which could comprehend all the forces by which nature is animated ... To it nothing would be uncertain, and

the future, as the past, would be present to its eyes. “ (Laplace 1812–1820). This view suggests that our ignorance is temporary, and good research into human behaviour will increase our understanding and help formulating accurate forecasts.

Whichever view is correct, demographic behaviour is not well explained as of today. When explanation is problematic, forecasting is even more difficult. Therefore, in addition to whatever fragmentary insight demographers obtain from behavioural sciences, they rely heavily on current real trends in vital processes, and they extrapolate those trends into the future. Hence, they face a problem when the indicators show unexpected changes in level or slope. It will not be clear whether these are caused by random fluctuations, or whether there is a structural change in the underlying trends. A trend shift that is perceived as random will first lead to large forecast errors. This effect is known in forecasting literature as assumption drag (Ascher 1978). Later, when the new trend is acknowledged, it will be included in the forecast updates and the errors will diminish. On the other hand, random fluctuations that are perceived as a trend shift will cause forecast errors, which will have a fluctuating effect on subsequent forecasts.

3 Empirical evidence from historical forecasts

There is a large literature, in which historical population forecasts are evaluated against observed statistics (Preston 1974; Calot and Chesnais 1978; Inoue and Yu 1979; Keyfitz 1981; Stoto 1983; 1987; Pflaumer 1988; Keilman 1997, 1998, 2000, 2001; Keilman and Pham 2004; National Research Council 2000). These studies have shown, among others, that forecast accuracy is better for short than for long forecast durations, and that it is better for large than for small populations. They also learned us that forecasts of the old and the young tend to be less accurate than those of intermediate age groups, and that there are considerable differences in accuracy between regions and components. Finally, poor data quality tends to go together with poor forecast performance. This relationship is stronger for mortality than for fertility, and stronger for short-term than for long-term forecasts. Selected examples of these general findings will be given below.

3.1 Forecasts are more accurate for short than for long forecast durations

Duration dependence of forecast accuracy is explained by the fact that the more years a forecast covers, the greater is the chance that unforeseen developments will produce unexpected changes in fertility, mortality, or migration.

The US National Research Council (NRC) evaluated the accuracy of nine total population size forecasts for countries of the world. Four of these were published by the United Nations (between 1973 and 1994), four by the World Bank (between 1972 and 1990), and one by the US Census Bureau (1987). The absolute percentage error, that is the forecast error irrespective of sign, increased from 5 per cent on average for five-year ahead forecasts, to 9 per cent 15 years ahead, and to 14 per cent 25 years ahead (NRC 2000). The average was computed over all countries and all forecasts. Developed countries had errors that were lower, and increased slower by forecast duration: from 2 (5 years ahead) to 4–5 (25 years ahead) per cent. A striking feature of these errors is that, even at duration zero, *i.e.*, in the forecast's base year, the errors are not negligible. Hence, forecasts start off with an incorrect base line population. For countries in Africa and the Middle East this base line error was highest: five per cent. Base line errors reflect poor data quality: when the forecasts were made, demographers worked with the best data that were available, but in retrospect, those data were revised.

Total fertility showed average errors from 0.4 children per woman after five years, to 0.6 and 0.8 children per woman after 15 and 25 years, with higher than average errors for European countries. In an evaluation of ten TFR-forecasts made by the UN since 1965, I found that for Europe as a whole, TFR errors were lower, and increased slower: from 0.2 children per woman after five years, to 0.5 after 15 years (Keilman 2001). An analysis of the errors observed in TFR forecasts in 14 European countries made since the 1960s shows that TFR-predictions have been wrong by 0.3 children per woman for forecasts 15 years ahead, and 0.4 children per woman 25 years ahead (Keilman and Pham 2004). Life expectancy was wrong by 2.3 (five years ahead), 3.5 (15 years ahead) and 4.3 (25 years ahead) years on average in the NRC evaluation. In 14 European countries, life expectancy forecasts tended to be too low by 1.0–1.3 and 3.2–3.4 years at forecast horizons of 10 and 20 years ahead, respectively.

3.2 Forecasts are more accurate for large than for small populations

A size effect in empirical errors at the sub national level was established already fifty years ago (White 1954), and reconfirmed repeatedly (see Smith *et al.* 2001 for an overview). Schéele (1981) found that the absolute error in small area forecasts within the Stockholm area was approximately proportional to the square root of population size, *i.e.*, a power of 0.5 (see also Bandel Bäckman and Schéele 1995). Later, Tayman *et al.* (1998) confirmed such a power law for small area forecasts in San Diego County, California, when

they found that the mean absolute percentage forecast error was proportional to population size raised to the power 0.4.

At the international level, the NRC analysis referred to earlier showed that the absolute percentage error in forecasts of total population size was 5.5 per cent on average, the average being taken over all countries and all nine forecast rounds. However, for countries with less than one million inhabitants, the average was 3 percentage points *higher*; for countries with a population of at least one million, the error was 0.7 percentage points *lower* (controlling, among others, for forecast length, year forecasted, forecast round, and whether or not the country had had a recent census; see NRC 2000, Appendix Table B7).

There are three reasons for the size effect in forecast accuracy. First, at the international scale, forecasters tend to pay less attention to the smallest countries, and take special care with the largest ones (NRC 2000). Second, both at the international and the local scale, small countries and areas are stronger affected by random fluctuations than large ones. In fact, many errors at the lower regional level cancel after aggregation. This explains irregular patterns and randomness in historical series of vital statistics at the lower level, leading to unexpected real developments after the forecast was produced. Third, for small areas the impact of migration on total population is strong compared to fertility and mortality, while, at the same time, migration is the least predictable of the three components.

3.3 Forecasts of the old and the young tend to be less accurate than those of intermediate age groups

In medium sized and large countries and regions, international migration has much less effect on the age structure than fertility or mortality. Therefore, a typical age pattern is often observed for accuracy. For many developed countries, a plot of relative forecast errors against age reveals large and positive errors (*i.e.*, too high forecasts) for young age groups, and large negative errors (too low forecasts) for the elderly. Errors for intermediate age groups are small. This age effect in forecast accuracy has been established for Europe, Northern America, and Latin America, and for countries such as Canada, Denmark, the Netherlands, Norway, and the United Kingdom (Keilman 1997, 1998). The fall in birth rates in the 1970s came fully unexpected for many demographers, which led to too high forecasts for young age groups. At the same time, mortality forecasts were often too pessimistic, in particular for women – hence the forecasts predicted too few elderly. The relative errors for the oldest old are often of the same order of magnitude as those for the

youngest age groups: plus or minus 15 per cent or more for forecasts 15 years into the future.

3.4 Accuracy differs between components and regions

In an analysis of the accuracy of 16 sets of population projections that the UN published between 1951 and 1998, I found considerable variation among ten large countries and seven major regions (Keilman 2001). Problems are largest in pre-transition countries, in particular in Asia. The quality of UN data for total fertility and the life expectancy has been problematic in the past for China, Pakistan, and Bangladesh. The poor data quality for these countries went together with large errors in projected total fertility and life expectancy. For Africa as a whole, data on total population and age structure have been revised substantially in the past, and this is a likely reason for the poor performance of the projections in that region. Nigeria, the only African country in my analysis, underwent major revisions in its data in connection with the Census of 1991. In turn, historical estimates of fertility and mortality indicators had to be adjusted, and this explains large projection errors in the age structure, in total fertility and in the life expectancy for this country. The problematic data situation for the former USSR is well known, in particular that for mortality data. The result was that, on average, life expectancy projections were too high by 2.9 years, which in turn caused large errors in projected age structures for the elderly. For Europe and Northern America, data quality is generally good. Yet, as noted in Section 3.3, the two regions have large errors in long-range projections of their age structures, caused by unforeseen trend shifts in fertility and mortality in the 1960s and 1970s.

The analysis of the statistical distribution of observed forecast errors for 14 European countries showed that a normal distribution fitted well for errors in life expectancies (Keilman and Pham 2004): TFR-errors, on the other hand, were exponentially distributed. This indicates that the probability for extremely large error values was greater for the TFR than for the life expectancy. Extreme errors for net migration are even more likely.

4 The expected accuracy of current forecasts

Forecast users should be informed about the expected accuracy of the numbers they work with. It focuses their attention on alternative population futures that may have different implications, and it requires them to decide what forecast horizon to take seriously. Just because a forecast covers 100 years does not mean that one should necessarily use that long a forecast (NRC 2000). In that sense, empirical errors observed in a series of historical

forecasts for a certain country can give strong indications of the accuracy of the nation's current forecast. However, these historical errors are just one realization of a statistical process, which applied to the past. Expected errors for the *current* forecast can only be assessed when the population forecast is couched in probabilistic form.

A probabilistic population forecast of the cohort component type requires the joint statistical distribution of all of its input parameters. Because there are hundreds of input parameters, one simplifies the probabilistic model in two ways. First, one focuses on just a few key parameters (for instance, total fertility, life expectancy, net immigration).¹ Second, one ignores certain correlations, for instance those between components, and sometimes also those in the age patterns of fertility, mortality, or migration.²

In probabilistic forecasts, an important type of correlation is that across time (serial correlation). Levels of fertility and mortality change only slowly over time. Thus, when fertility or mortality is high one year, a high level the next year is also likely, but not 100 per cent certain. This implies a strong, but not perfect serial correlation for these two components. International migration is much more volatile, but economic, legal, political, and social conditions stretching over several years affect migration flows to a certain extent, and some degree of serial correlation should be expected. In the probabilistic forecasts for the United States (Lee and Tuljapurkar 1994), Finland (Alho 1998), the Netherlands (De Beer and Alders 1999), and Norway (Keilman *et al.* 2001, 2002) these correlation patterns were estimated based on time series models. For Austria (Hanika *et al.* 1997) and for large world regions (Lutz and Scherbov 1998a, 1998b) perfect autocorrelation was assumed for the summary parameters (total fertility, life expectancy, and net migration). This

¹ A cohort component forecast that has one-year age groups requires 35 fertility rates, 200 death rates, and some 140 parameters for net migration *for each forecast year*. With age groups and time intervals equal to five years, a forecast for a period of fifty years, say, still requires that one specify the joint statistical distribution of $(7+40+28)*10=750$ parameters.

² For Western countries, there is little or no reason to assume correlation between the components of fertility, mortality, and migration. Nor is there any empirical evidence of such correlation (Lee and Tuljapurkar 1994; Keilman 1997). In developing countries, disasters and catastrophes may have an impact both on mortality, fertility, and migration, and a correlation between the three components cannot be excluded. There may also be a positive correlation between the levels of immigration and childbearing in Western countries with extremely high immigration from developing countries.

assumption underestimates uncertainty (Lee 1999). In recent work for world regions, Lutz, Sanderson, and Scherbov relaxed the assumption of perfect autocorrelation (Lutz *et al.* 2001).

Three main methods are in use for computing probabilistic forecasts of the summary indicators: time series extrapolation, expert judgement, and extrapolation of historical forecast errors (Lee 1999; NRC 2000). The three approaches are complementary, and elements of all three are often combined. Time series methods and expert judgement result in the distribution of the parameter in question around its expected value. In contrast, an extrapolation of empirical errors gives the distribution centred around zero (assuming an expected error equal to zero), and the expected value of the population variable is taken from a deterministic forecast computed in the traditional manner.

Time series methods are based on the assumption that historical values of the variable of interest have been generated by means of a statistical model, which also holds for the future. A widely used method is that of Autoregressive Integrated Moving Average (ARIMA)-models. These time series models were developed for short horizons. When applied to long-run population forecasting, the point forecast and the prediction intervals may become unrealistic (Sanderson 1995). Judgmental methods (see below) can be applied to correct or constrain such unreasonable predictions (Lee 1993; Tuljapurkar 1996).

Expert judgement can be used when expected values and corresponding prediction intervals are hard to obtain by formal methods. In demographic forecasting, the method has been pioneered by Lutz and colleagues (Lutz *et al.* 1996; Hanika *et al.* 1997; Lutz and Scherbov 1998a, 1998b). A group of experts is asked to indicate the probability that a summary parameter, such as the TFR, falls within a certain pre-specified range for some target year, for instance the range determined by the high and the low variant of an independently prepared population forecast. The subjective probability distributions obtained this way from a number of experts are combined in order to reduce individual bias. A major weakness of this approach, at least based upon the experiences from other disciplines, is that experts often are too confident, *i.e.*, that they tend to attach a too high probability to a given interval (Armstrong 1985). A second problem is that an expert would have problems with sensibly guessing whether a certain interval corresponds to probability bounds with 90 per cent coverage versus 95 per cent or 99 per cent (Lee 1999).

Extrapolation of empirical errors requires observed errors from historical forecasts. Formal or informal methods may be used to predict the errors for the current forecast. Keyfitz (1981) and Stoto (1983) were among the first to use this approach in demographic forecasting. They assessed the accuracy of historical forecasts for population growth rates. The Panel on Population Projections of the US National Research Council (NRC 2000) elaborated further on this idea and developed a statistical model for the uncertainty around total population in UN-forecasts for all countries of the world. Others have investigated and modelled the accuracy of predicted TFR, life expectancy, immigration levels, and age structures (Keilman 1997; De Beer 1997). There are two important problems. First, time series of historical errors are usually rather short, as forecasts prepared in the 1960s or earlier generally were poorly documented. Second, extrapolation is often difficult because errors may have diminished over successive forecast rounds as a result of better forecasting methods.

Irrespective of the method that is used to determine the prediction intervals for all future fertility, mortality and migration parameters, the next step is to apply these to the base population in order to compute prediction intervals for future population size and age pyramids. This can be done in two ways: analytically, and by means of simulation.

The *analytical approach* is based on a stochastic cohort component model, in which the statistical distributions for the fertility, mortality, and migration parameters are transformed into statistical distributions for the size of the population and its age-sex structure. Alho and Spencer (1985) and Cohen (1986) employ such an analytical approach, but they need strong assumptions. Lee and Tuljapurkar (1994) give approximate expressions for the second moments of the distributions.

The *simulation approach* avoids the simplifying assumptions and the approximations of the analytical approach. The idea is to compute several hundreds or thousands of forecast variants (“sample paths”) based on input parameter values for fertility, mortality, and migration that are randomly drawn from their respective distributions, and store the results in a database. Early contributions based on the idea of simulation are those by Keyfitz (1985), Pflaumer (1986, 1988), and Kuijsten (1988).

In order to illustrate that probabilistic forecasts are useful when uncertainty has to be quantified, I shall give an example for the population of Norway. I shall compare the results from a probabilistic forecast with those from a traditional deterministic one, prepared by Statistics Norway.

5 Probabilistic forecasts: an alternative to forecast variants

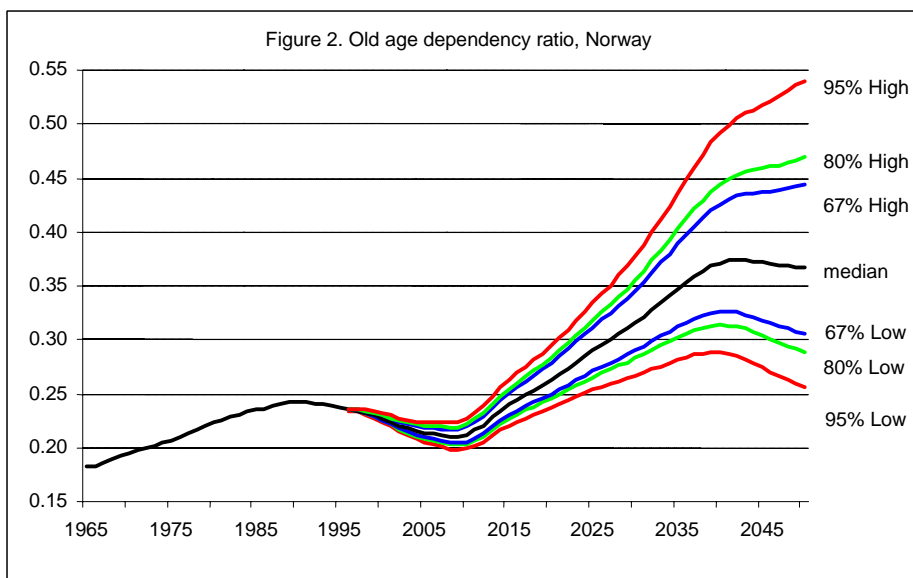
Technical details of the methods used to construct the probabilistic forecast are presented elsewhere (Keilman *et al.* 2001, 2002). Here I shall give a brief summary.

ARIMA time series models were estimated for observed annual values of the TFR, the life expectancy for men and women, and total immigration and immigration in Norway since the 1950s. Based on these ARIMA models, repeated stochastic simulation starting in 1996 yielded 5,000 sample paths for each of these summary parameters to the year 2050. The predictive distributions for the TFR and the life expectancy at birth were checked against corresponding empirical distributions based on historical forecasts published by Statistics Norway in the period 1969–1996. The predicted TFR, life expectancy, and gross migration flows were broken down into age specific rates and numbers by applying various model schedules: a Gamma model for age specific fertility, a Heligman-Pollard model for mortality, and a Rogers-Castro model for migration. Next, the results of the 5,000 runs of the cohort component model for the period up to 2050 were assembled in a data base containing the future population of Norway broken down by one-year age group, sex, forecast year (1997–2050), and forecast run. For each variable of interest, for example the total population in 2030, or the old age dependency ratio (OADR) in 2050, one can construct a histogram based on 5,000 simulated values, and read off prediction intervals with any chosen coverage probability.

The results showed odds equal to four against one (80 per cent chance) that Norway's population, now 4.5 million, will number between 4.3 and 5.4 million in the year 2025, and 3.7–6.4 million in 2050. Uncertainty was largest for the youngest and the oldest age groups, because fertility and mortality are hard to predict. As a result, prediction intervals in 2030 for the population younger than 20 years of age were so wide, that the forecast was not very informative. International migration showed large prediction intervals around expected levels, but its impact on the age structure was modest. In 2050, uncertainty had cumulated so strongly, that intervals were very large for virtually all age groups, in particular when the intervals are judged in a relative sense (compared to the median forecast).

Figure 2 shows the high and the low bound of the various prediction intervals for the old age dependency ratio, defined as the population 67 and over rela-

tive to that aged 20–66.³ The prediction intervals are those with 95 per cent, 80 per cent, and 67 per cent coverage. The median of the predictive distributions is also plotted. The intervals widen rapidly, reflecting that uncertainty increases with time. We see that ageing is certain in Norway, at least until 2040. In that year, the odds are two against one (67 per cent interval) that the OADR will be between 0.33 and 0.43, *i.e.*, at least 10 points higher than today’s value of 0.23. The probability of a ratio in 2040 that is lower than today’s is close to zero.



How do these probabilistic forecast results compare with those obtained by a traditional deterministic forecast? Statistics Norway’s most recent population forecast contains variants for high population growth and low population growth, among others (Statistics Norway 2005). The high population growth forecast results from combining a high fertility assumption with a high life expectancy assumption (*i.e.*, low mortality) and a high net immigration assumption. Likewise, the low growth variant combines low fertility with low life expectancy and low immigration. The forecast predicts a population aged 67 and over in 2050 between 1,095,000 (low growth) and 1,406,000 (high growth). However, the corresponding OADR-values are 0.409 for low population growth, and 0.392 for high population growth. Therefore, while there is a considerable gap between the absolute numbers of elderly in the two

³ The legal retirement age in Norway is 67.

variants, the relative numbers, as a proportion of the population aged 20–66, are almost indistinguishable. The interval for the absolute number thus reflects uncertainty in some sense, but the OADR-interval for the same variant pair suggests almost no uncertainty. On the other hand, the probabilistic forecast results in Figure 2 show a two-thirds OADR-prediction interval in 2050 that stretches from 0.31 to 0.44.⁴

This example illustrates that it is problematic to use forecast variants from traditional deterministic forecast methods to express forecast uncertainty. First, uncertainty is not quantified. Second, the use of high and low variants is inconsistent from a statistical point of view (Lee 1999, Alho 1998). In the high variant, fertility is assumed to be high in every year of the forecast period. Similarly, when fertility is low in one year, it is 100 per cent certain that it will be low in the following years, too. Things are even worse when two or more mortality variants are formulated, in addition to the fertility variants, so that high/low growth variants result from combining high fertility with high life expectancy/low fertility with low life expectancy. In that case, *any* year in which fertility is high, life expectancy is high as well. In other words, one assumes perfect correlation between fertility and mortality, in addition to perfect serial correlation for each of the two components. Assumptions of this kind are unrealistic, and, moreover, they cause inconsistencies: two variants that are extreme for one variable need not be extreme for another variable.

As a further illustration of the use of stochastic population forecasts when analyzing pension systems, let me consider the possibility of a flexible retirement age. When workers postpone retirement, they contribute longer to the pension fund, and the years they benefit from it become shorter (other factors remaining the same). Therefore I analyse the following question: which retirement age is necessary in Norway in the future in order to achieve a constant OADR (see also Section 4 of the chapter by Tuljapurkar in this volume for a similar analysis for the United States)? I will investigate two cases. First I assume a constant OADR equal to 0.24, which is the highest value observed in the past (around 1990, see Figure 2). Second, I assume an OADR equal to 0.18. This is the value in 1967, the year when the Norwegian pension system in its current form was introduced. Since the future age struc-

⁴ The median OADR-value of the stochastic forecast in 2050 (0.37) is lower than the medium value of Statistics Norway's forecast for that year (0.395). Life expectancy in 2050 rises to 86 years in Statistics Norway's forecast, but only to 82.3 years in the median of the stochastic forecast. The latter forecast was prepared four years earlier than Statistics Norway's forecast.

ture is uncertain, the retirement age necessary to obtain a constant OADR becomes a stochastic variable. Table 1 gives the results.

Table 1 Prediction intervals for retirement age, Norway

| | Median | 67 per cent interval | 80 per cent interval | 95 per cent interval |
|--------------------|--------|----------------------|----------------------|----------------------|
| OADR = 0.24 | | | | |
| 2010 | 64.8 | 64.5–65.2 | 64.4–65.3 | 64.1–65.5 |
| 2030 | 69.2 | 68.2–70.3 | 67.9–70.6 | 67.2–71.3 |
| 2050 | 71.9 | 69.4–74.4 | 68.7–75.1 | 67.0–76.8 |
| OADR = 0.18 | | | | |
| 2010 | 67.6 | 67.2–68.1 | 67.1–68.2 | 66.8–68.5 |
| 2030 | 72.0 | 71.0–73.0 | 70.7–73.3 | 70.1–74.0 |
| 2050 | 75.1 | 72.8–77.2 | 72.2–77.9 | 70.7–79.4 |

The table shows that the retirement age in Norway must increase strongly from its current value of 67 years, if the OADR were to remain constant at 0.24. The median of 71.9 years in 2050 indicates that the rise is almost 5 years. Yet the uncertainty is large here. In four out of five cases would the retirement age in 2050 be between 69 and 75 years. In the short run the situation is completely different. The age structure of the population of Norway is such that the retirement age can *decrease* to 2010, and yet the ratio of elderly to the population in labour force ages could remain constant. This finding is almost completely certain. Even the upper bound of the 95 per cent interval (65.5) is much lower than today’s retirement age.

If one would require an OADR as low as the one in 1967, the median age at retirement has to increase to no less than 75.1 years in 2050. A higher retirement age is necessary even in the short run: the median in 2010 is 67.6 years, and the lower bound to the 80 per cent prediction interval indicates that the probability that we may see an increase is about ten per cent or lower, given the assumptions made.

6 Challenges in probabilistic population forecasting

A probabilistic forecast extrapolates observed variability in demographic data to the future. For a proper assessment of the variability, one needs long series with annual data of good quality. The minimum is about fifty years, but a longer series is preferable. At the same time, one would ideally have a long series of historical forecasts, and estimate empirical distributions of observed forecast errors based on the old forecasts. There are very few countries that have so good data. Therefore, a major challenge in probabilistic forecasting is to prepare such forecasts for countries with poorer data. Two research directions seem promising. First, when time series analysis cannot be used to compute predictive distributions, one has to rely strongly on expert opinion. Lutz and colleagues (1996, 2001) have indicated how this can be done in practice. An important task here is a systematic elicitation of the experts' opinions, in order to avoid too narrow prediction intervals. Second, in case the data from historical forecasts are lacking, one could replace actual forecasts by naïve or baseline forecasts (Keyfitz 1981; 1998). Historical forecasts often assumed constant (or nearly constant) levels or growth rates for summary indicators such as the TFR, the life expectancy, or the level of immigration. Thus we can study how accurate past fertility forecasts would have been if they had assumed that the base value had persisted. Similarly, we can compute mortality errors based on an assumption of a linear increase in life expectancy. Such naïve error estimates would be expected to lead to conservative, that is, too large variability estimates, in some cases only slightly so but in others substantially.

Most applications of probabilistic forecasting so far focus on one country. Very few have a regional or an international perspective. One important exception is the work by Lutz, Sanderson, and Scherbov (1996, 2001), who used a probabilistic cohort component approach for 13 regions of the world⁵. For fertility and mortality, they combined the three methods mentioned in section 4 to obtain predictive distributions for summary indicators. An important challenge was the probabilistic modelling of interregional migration, because migration data show large volatility in the trends, are unreliable, not consistent between countries, or often simply lacking. In their 1996 study, Lutz and colleagues assumed a matrix of constant annual interregional migration flows, with the 90 per cent prediction bounds corresponding to certain high and low migration gains in each region. In the recent study, net migra-

⁵ Probabilistic forecasts of total population size for all countries of the world have been prepared by the Panel on Population Projections (NRC 2000), but these forecasts do not give age detail.

tion into the regions was modelled as a stochastic vector with a certain autocorrelation structure. A second challenge was the treatment of interregional correlations for fertility, mortality, and migration. Due to the paucity of the necessary data, these correlations are difficult to estimate. Therefore, the authors combined qualitative considerations with sensitivity analysis, and investigated alternative regional correlation levels.

Because of these data problems, the development of a sound method for probabilistic multiregional cohort component forecasting is an important research challenge. For sub-national forecasts, the problems are probably easier to overcome than for international forecasts, because the data situation is better in the former case, at least in a number of developed countries. The way ahead would thus be to collect better migration data, and to invest efforts in estimating cross-regional correlation patterns for fertility, mortality, and migration. An alternative strategy could be to start from a probabilistic cohort component forecast for the larger region, and to compute such forecasts at the lower regional level (by age and sex) by means of an appropriate multivariate distribution with expected values corresponding to the regional shares from an independently prepared deterministic forecast.

Not only regional forecasts, but also other types of population forecasts should be couched in probabilistic terms, such as labour market forecasts, educational forecasts, and household forecasts, to name a few. Very few of such probabilistic forecasts have been prepared. Lee and Tuljapurkar (2001) have investigated the expected accuracy of old age security funds forecasts in the United States. A major topic of research here is to analyse the relative contribution to uncertainty of demographic factors (fertility, mortality, migration) and non-demographic factors (labour market participation, educational attainment, residential choices).

References

- Alho, J. M. (1998) A stochastic forecast of the population of Finland. *Reviews* 1998/4. Statistics Finland, Helsinki.
- Alho, J. and B. Spencer (1985) Uncertain population forecasting. *Journal of the American Statistical Association* 80: 306–314.
- Armstrong, J. (1985) *Long-range forecasting: From crystal ball to computer*. New York: Wiley (2nd ed.).
- Ascher, W. (1978) *Forecasting: An appraisal for policy-makers and planner*. Baltimore: The Johns Hopkins University Press.
- Bandel Bäckman, J. and S. Schéele (1995) Uncertainty in population forecasts for small areas. Pp. 419–434 in C. Lundh (ed.) *Demography, Economy and Welfare*. Lund Studies in Economic History 1. Lund: Lund University Press.
- Boudon, R. (1986) *Theories of social change: A critical appraisal*. Cambridge: Polity Press.
- Calot, G. and J.C. Chesnais (1978) La prevision démographique. Dans A.-C. Decouflé *Traité élémentaire de prevision et de prospective*. Paris: PUF.
- Cohen, J. (1986) Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. *Demography* 23: 105–126.
- 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. and M. Alders (1999) Probabilistic population and household forecasts for the Netherlands Working Paper nr. 45, Joint ECE-Eurostat Work Session on Demographic Projections, Perugia, Italy, 3–7 May 1999.
- Hanika, A., W. Lutz and S. Scherbov (1997) Ein probabilistischer Ansatz zur Bevölkerungsvorausschätzung für Österreich, *Statistische Nachrichten* 12/1997: 984–988.
- 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).
- Henry, L. and H. Gutierrez (1977) Qualité des prévisions démographiques à court terme: Etude de l'extrapolation de la population totale des départements et villes de France 1821–1975, *Population* 32: 625–647.

Inoue, S. and Y.C. Yu (1979) United Nations new population projections and analysis of ex post facto errors. Paper Annual Meeting Population Association of America, Philadelphia, April 1979.

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

Keilman, N. (1998) How accurate are the United Nations world population projections? *Population and Development Review Supplement to Volume 24*: 15–41.

Keilman, N. (2000) Précision et incertitudes des prévisions nationales de population Pp. 33–67 in C. Wattelar et J. Duchêne (red.): *Le défi de l'incertitude: Nouvelles approches en perspectives et prospective démographiques. Actes de la Chaire Quetelet 1995*. Louvain-la-Neuve: Academia-Bruylant/L'Harmattan.

Keilman, N. (2001) Data quality and accuracy of United Nations population projections, 1950–95. *Population Studies* 55(2): 149–164.

Keilman, N. and H. Cruijssen, eds. (1992) *National Population Forecasting in Industrialized Countries*. Amsterdam etc.: Swets & Zeitlinger.

Keilman, N., D.Q. Pham, and A. Hetland (2001) Norway's uncertain demographic future. *Social and Economic Studies* 105. Oslo: Statistics Norway. Internet www.ssb.no/english/subjects/02/03/sos105_en.

Keilman, N., D.Q. Pham, and A. Hetland (2002) Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research* 6–15 May 2002, 409–453. Internet www.demographic-research.org/volumes/vol6/15/6-15.pdf.

Keilman, N. and D. Q. Pham (2004) Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area. Discussion Paper 386, Statistics Norway.

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

Keyfitz, N. (1982) Can knowledge improve forecasts? *Population and Development Review* 8: 729–751.

Keyfitz, N. (1985) A probability representation of future population. *Zeitschrift für Bevölkerungswissenschaft* 11: 179–191.

Kuijsten, A. (1988) Demografische toekomstbeelden van Nederland (Demographic scenarios for the Netherlands). *Bevolking en Gezin* 1988/2: 97–130.

Laplace, Pierre Simon de (1812–1829) *Œuvres* Vol. VII, *Théorie Analytique des Probabilités*.

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

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

Lee, R. and S. Tuljapurkar (1994) Stochastic population forecasts for the United States: Beyond High, Medium, and Low, *Journal of the American Statistical Association* 89: 1175–1189.

Lee, R. and S. Tuljapurkar (2001) Population forecasting for fiscal planning: Issues and Innovation. Pp. 7–57 in A. Auerbach and R. Lee (eds.). *Demographic change and fiscal policy*. Cambridge University Press.

Lutz, W., W. Sanderson, and S. Scherbov (1996) Probabilistic population projections based on expert opinion pp. 397–428 in W. Lutz (ed.). *The future population of the world: What can we assume today?* London: Earthscan (rev. ed.).

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

Lutz, W. and S. Scherbov (1998b) Probabilistische Bevölkerungsprognosen für Deutschland (Probabilistic population projections for Germany), *Zeitschrift für Bevölkerungswissenschaft* 23: 83–109.

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

Nagel, E. (1961) *The structure of science: Problems in the logic of scientific explanations*. New York: Harcourt, Brace and World.

National Research Council (2000) *Beyond six billion: Forecasting the world's population*. Panel on Population Projections. J. Bongaarts and R. Bulatao (eds.). Committee on Population, Commission on Behavioral and Social Sciences and Education. Washington DC: National Academy Press.
Internet <http://books.nap.edu/books/0309069904/html/index.html>.

- Pflaumer, P. (1986) Stochastische Bevölkerungsmodelle zur Analyse der Auswirkungen demographischer Prozesse auf die Systeme der sozialen Sicherung, *Allg. Statist. Archiv* 70: 52–74.
- Pflaumer, P. (1988) Confidence intervals for population projections based on Monte Carlo methods. *International Journal of Forecasting* 4: 135–142.
- Preston, S. (1974) An evaluation of postwar mortality projections in Australia, Canada, Japan, New Zealand, and the United States. *WHO Statistical Report* 27: 719–745.
- Sanderson, W. (1995) Predictability, complexity, and catastrophe in a collapsible model of population, development, and environmental interactions. *Mathematical Population Studies* 5: 233–279.
- Schéele, S. (1981) Osäkerhet i befolkningsprognoser. Stockholm: Stockholms Kommune Utrednings- och Statistikkontor.
- Smith, S.K., J. Tayman, and D. Swanson (2001) *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers (The Plenum series on demographic methods and population analysis).
- Statistics Norway (2005) *Framskrivning av folkemengden 2002/2050 (Population projections 2002/2050) NOS D 319*, Oslo: Statistics Norway.
- Stoto, M. (1983) The accuracy of population projections. *Journal of the American Statistical Association* 78: 13–20.
- Tayman, J., E. Schafer, and L. Carter (1998) The role of population size in the determination and prediction of forecast errors: An evaluation using confidence intervals for subcounty areas. *Population Research and Policy Review* 17: 1–20.
- Tuljapurkar, S. (1996) Uncertainty in demographic projections: Methods and meanings. Paper Sixth Annual Conference on Applied and Business Demography. Bowling Green, 19–21 September 1996.
- United Nations (2001) *World population prospects: The 2000 revision. Volume I: Comprehensive tables*. New York: United Nations.
- United Nations (2005) *World population prospects: The 2004 revision*. New York: United Nations.
- White, H.R. (1954) Empirical study of the accuracy of selected methods of projecting state populations. *Journal of the American Statistical Association* 29: 480–498.

Remarks on the Use of Probabilities in Demography and Forecasting

Juha M. Alho

Professor of Statistics at the University of Joensuu, Finland

1 Introduction

The concept of “probability” is used as a step in life table construction to get the expected number of survivors in a cohort. However, in traditional texts on demographic methods (*e.g.*, Shryock and Siegel 1976), variance in the number of survivors plays no role. Similarly, concepts of estimation, estimation error, and bias are routinely used, but standard error and sampling distribution are not (except in connection with sample surveys). Although statistically satisfactory accounts of the life table theory have existed for a long time (*e.g.*, Chiang 1968, Hoem 1970), a reason for neglecting population level random variability, and statistical estimation error, has been that the populations being studied are so large that random error must be so small as not to matter, in practice.

In contrast, when statistical methods started to become used in population forecasting in the 1970s, 1980s and 1990s, some of the resulting prediction intervals have been criticized as being implausibly wide. This view has not often been expressed in print, but Smith (2001, 70–71) provides an example. Others, especially sociologically minded critics have gone further and argued that due to the nature of social phenomena, the application of probability concepts in general, is inappropriate. On the other hand, demographers coming with an economics background have tended to find probabilistic thinking more palatable.

The purpose of the following remarks is to review some probability models, and show how the apparent contradiction arises. We will see that the basic principles have been known for decades. The basic cause of the difficulties – and disagreements – is that there are several layers of probabilities that can be considered. Consequently, it is essential to be explicit about the details of the model.

2 Binomial and Poisson Models

As emphasized by good introductory texts on statistics (*e.g.*, Freedman, Pisani and Purves 1978, p. 497), the concept of probability can only be made precise in the context of a mathematical model. To understand why one often might ignore other aspects of random variables besides expectation, let us construct a model for the survival of a cohort of size n for one year. For each individual $i = 1, \dots, n$, define an indicator variable such that $X_i = 1$, if i dies during the year, and $X_i = 0$ otherwise. The total number of deaths is then $X = X_1 + \dots + X_n$. We assume that the X_i 's are random variables (*i.e.*, their values are determined by a chance experiment). Suppose we make an assumption concerning their expectation

$$E[X_i] = q, I = 1, \dots, n, \tag{1}$$

and assume that

$$X_1, \dots, X_n \text{ are independent.} \tag{2}$$

It follows that X has a binomial distribution, $X \sim \text{Bin}(n, q)$. As is well known, we have the expectation $E[X] = nq$, and variance $\text{Var}(X) = n(q - q^2)$. Therefore, the coefficient of variation is $C = ((1 - q)/nq)^{1/2}$.

Now, in industrialized countries the probability of death is about 1 per cent and population size can be in the millions, so relative variation can, indeed, be small. For example, if $q = 0.01$ and $n = 1,000,000$, we have that $C = 0.01$. Or, the relative random variability induced by the model defined in (1) and (2) is about 1 per cent. Equivalent calculations have already been presented by Pollard (1968), for example.

One might object to the conclusion that relative variability is negligible on the grounds that (1) does not hold: surely people of different ages (and of different sex, socio-economic status etc.) have different probabilities of death. Therefore, suppose that

$$E[X_i] = q_i, \text{ with } q = (q_1 + \dots + q_n)/n. \tag{3}$$

In this case

$$\text{Var}(X) = nq - \sum_{i=1}^n q_i^2 \tag{4}$$

However, it follows from the Cauchy-Schwarz inequality that

$$n \sum_{i=1}^n q_i^2 \geq n^2 q^2. \quad (5)$$

Therefore, we have that the variance (4) is actually *less* than the binomial variance. The naive argument based on population heterogeneity simply does not hold.

Before proceeding further, let us note that, apart from substantive factors, heterogeneity of the type (3) is imposed on demographic data, because vital events are typically classified by age, so individuals contribute different times to the “rectangles” of the Lexis diagram. This is one reason why the basic data are typically collected in terms of rates, and a Poisson assumption is invoked. There is some comfort in the fact that if the assumptions (2) and (3) are complemented by the following assumptions: suppose the q_i 's depend on n and as $n \rightarrow \infty$, (i) $nq = \lambda > 0$, and (ii) $\max\{q_1, \dots, q_n\} \rightarrow 0$, then the distribution of X converges to $Po(\lambda)$ (Feller 1968, p. 282). The Poisson model is of interest, because under that model $E[X] = \lambda$ as before, but $\text{Var}(X) = \lambda > n(q - q^2)$. In other words, the Poisson model has a larger variability than the corresponding binomial models. Quantitatively the difference is small, however, since now $C = \lambda^{-1/2}$. If $n = 1,000,000$ and $q = 0.01$, then $\lambda = 10,000$, and $C = 0.01$, for example. Or, the relative variability is the same as that under the homogeneous binomial model, to the degree of accuracy used.

The usual demographic application of the Poisson model proceeds from the further identification $\lambda = \mu K$, where K is the person years lived in the population, and μ is the force of mortality. The validity of this model is not self-evident, since unlike n , K is also random. At least when λ is of a smaller order of magnitude than K , the approximation appears to be good, however (Breslow and Day 1987, pp. 132–133). As is well-known, the maximum likelihood estimator of the force of mortality is $\hat{\mu} = X/K$ with the estimated standard error of $X^{1/2}/K$. Extensions to log-linear models that allow for the incorporation of explanatory variables follow similarly.

Since the Poisson distribution has the variance maximizing property, and it provides a model for both the independent trials and occurrence/exposure rates, we will below restrict the discussion primarily to the Poisson case.

3 Random Rates

Since (1) is not the cause of the low level of variability in the number of deaths, we need to look more closely at (2). A simple (but unrealistic) example showing that there are many opportunities here is the following. Suppose we make a single random experiment with probability of success = q , and probability of failure = $1 - q$. If the experiment succeeds, define $X_i = 1$ for all i . Otherwise define $X_i = 0$ for all i . In this case we have, for example, that $X = nX_i$, so $E[X] = nq$ as before, but $\text{Var}(X) = n^2(q - q^2)$, and $C = ((1 - q)/q)^{1/2}$ independently of n . For $q = 0.01$ we have $C = 9.95$, for example, indicating a huge (nearly 1,000 per cent) level of variability.

More realistically, we may think that dependence across individuals arises because they may all be influenced by common factors to some extent, at least. For example, there may be year to year variation in mortality around a mean that is due to irregular trends in economics, epidemics, weather etc. If the interest would center on a given year, the model might still be $X \sim \text{Po}(\mu K)$, but if several years are considered jointly, then the year-to-year variation due to such factors would have to be considered. In this case, we would entertain a hierarchical model of the type

$$X \sim \text{Po}(\mu K) \text{ with } E[\mu] = \mu_0, \text{Var}(\mu) = \sigma^2.$$

In other words, the rate μ itself is being considered random, with a mean μ_0 that reflects the average level of mortality over the (relatively short) period of interest, and variance σ^2 that describes the year to year variation.

In this case we have that

$$\begin{aligned} \text{Var}(X) &= E[\text{Var}(X | \mu)] + \text{Var}(E[X | \mu]) \\ &= \mu_0 K + \sigma^2 K^2. \end{aligned} \tag{7}$$

It follows that $\text{Var}(\hat{\mu}) = \sigma^2 + \mu_0/K$. This result is of fundamental interest in demography, because if K is large, then the dominant part of the error is due to the annual variability. If the interest centers (as in the production of official population statistics) on a given year, with no regard to other years, we would be left with the pure Poisson variance μ_0/K , which is often small. An exception is the oldest-old mortality, where Poisson variation is always an issue, because for ages high enough K will always be small and μ_0 large.

However, when the interest centers on the time trends of mortality, and eventually on forecasting its future values, then the year to year variation σ^2 must

be considered. Under model (6) this is independent of population size K . This is a realistic first approximation, but we note that model (6) does not take into account the possibility that a population might consist of relatively independent subpopulations. In that case, populations having many such subpopulations would have a smaller variance than a population with no independent subpopulations.

4 Handling of Trends

Consider now two counts. Or assume that for $i = 1, 2$, we have that

$$X_i \sim \text{Po}(\mu_i K_i) \text{ with } E[\mu_i] = \mu_{0i}, \text{Var}(\mu_i) = \sigma_i^2, \text{Corr}(\mu_1, \mu_2) = \rho. \quad (8)$$

Repeating the argument leading to (7) for covariances yields the result

$$\text{Corr}(X_1, X_2) = \rho / \{(1 + \mu_{01} / \sigma_1 K_1) (1 + \mu_{02} / \sigma_2 K_2)\}^{1/2}. \quad (9)$$

Or, the effect of Poisson variability is to decrease the correlation between the observed rates. We note that if the K_i 's are large, the attenuation is small. However, for the oldest old the K_i 's are eventually small, and the μ_{0i} 's large, so attenuation is expected.

In concentrating on Poisson variability that is primarily of interest in the assessment of the accuracy of vital registration, demographers have viewed annual variation as something to be explained. Annual changes in mortality and fertility are analyzed by decomposing the population into ever finer subgroups in an effort to try to find out, which are the groups most responsible for the observed change. Often partial explanations can be found in this manner, but they rarely provide a basis for anticipating future changes (Keyfitz 1982). To be of value in forecasting, an explanation must have certain robustness against idiosyncratic conditions, and persistence over time. This leads to considering changes around a trend as random.

One cause for why some demographers find statistical analyses of demographic time-series irritating seems to lie here: what a demographer views as a phenomenon of considerable analytical interest, may seem to a statistician as a mere random phenomenon, sufficiently described once σ^2 is known. [This tension has counterparts in many parts of science. Linguists, for example, differ in whether they study the fine details of specific dialects, or whether they try to see general patterns underlying many languages.]

In forecasting, the situation is more complex than outlined so far. In mortality forecasting one would typically be interested taking account of the nearly universal decline in mortality, by making further assumptions about time trends. For example, suppose the count at time t is of the form $X_t \sim \text{Po}(\mu_t K_t)$, such that

$$\log(\mu_t) = \alpha + \beta t + \xi_t, \text{ where } E[\xi_t] = 0, \text{ Cov}(\xi_t, \xi_s) = \sigma^2 \min\{t, s\}. \quad (10)$$

If the ξ_t 's have normal distributions, under (10) we would have that $E[\mu_t] = \exp(\alpha + \beta t + \sigma^2 t/2) \equiv \mu_{0t}$. (This model is closely related to the so-called Lee-Carter model.)

One reason that makes (10) more complicated than (8), is that μ_{0t} involves parameters to be estimated, so standard errors become an issue. Especially, if $\text{Var}(\hat{\beta})$ is large, this source of error may have a considerable effect for large t , because it induces a quadratic term into the variance of error, whereas the effect of the random walk via σ^2 is only linear.

The way $\text{Var}(\hat{\beta})$ is usually estimated from past data assumes that the model specified in (10) is correct. Therefore, probabilistic analyses based on (10) are conditional on the chosen model. What these probabilities do not formally cover is the uncertainty in model choice itself (Chatfield 1996).

One should pay attention to model choice because it is typically based on iteration, in which lack of fit is balanced against parametric parsimony (*cf.*, Box and Jenkins 1976). One would expect error estimates calculated after a selection process to be too small, because of potential overfitting. Yet, a curious empirical fact seems to be that statistical time-series models identified and estimated in this manner, for example demographic time-series, often produce prediction intervals that are rather wide, and even implausibly wide in the sense that in a matter of decades they may include values that are thought to be biologically implausible.

A possible explanation is that the standard time-series models belong to simple classes of models (*e.g.*, (10) can be seen as belonging to models with polynomial trends with once integrated, or $I(1)$, errors) and the identification procedures used are tilted in favor of simple models within those classes. This shows that although judgment is certainly exercised in model choice, it can be exercised in a relatively open manner that tends to produce models that are too simple rather than too complex. When such models are estimated from the data, part of the lack of fit is due to modeling error. Therefore, the estimated models can actually incorporate some aspects of modeling error.

Modeling error can sometimes be handled probabilistically by considering alternative models within a larger class of models, and by weighting the results according to the credibility of the alternatives (*e.g.*, Draper 1995). Alho and Spencer (1985) discuss some minimax type alternatives in a demographic context. A simpler approach is to use models that are not optimized to provide the best possible fit obtainable. In that case the residual error may capture some of the modeling error, as well.

5 On Judgment and Subjectivity in Statistical Modeling

“One cannot be slightly pregnant”. In analogy, it is sometimes inferred from this dictum that if judgment is exercised in some part of a forecasting exercise, then all probabilistic aspects of the forecast are necessarily judgmental in nature. In addition, since judgment always involves subjective elements, then the probabilities are also purely subjective. I believe these analogies are misleading in that they fail to appreciate the many layers of probabilities one must consider.

First, the assumption of binomial or Poisson type randomness is the basis of grouped mortality analyses, and as such implicitly shared by essentially all demographers. It takes some talent to see how such randomness could be viewed as subjective.

Second, although models of random rates are not currently used in descriptive demography, they are implicit in all analyses of trends in mortality. Such analyses use smooth models for trends, and deviations from trends are viewed as random. The validity of alternative models can be tested against empirical data and subjective preferences have relatively little role.

On the other hand, models used in forecasting are different in that they are thought to hold in the future, as well as in the past. Yet, they can only be tested against the past. However, even here, there are different grades. In short term forecasting (say, 1–5 years ahead), we have plenty of empirical data on the performance of the competing models in forecasting. Hence, there is an empirical and fairly formal basis for the choice of models. In medium term forecasting (say, 10–20 years ahead), empirical data are much more scant, and alternative data sets produce conflicting results of forecast performance. Judgment becomes an important ingredient in forecasting. In long-term forecasting (say, 30+ years ahead), the probabilities calculated based on any statistical model begin to be dominated by the possibility of modeling error and beliefs concerning new factors whose influence has not manifested

itself in the past data. Judgment, and subjective elements that cannot be empirically checked, get an important role. Note that the binomial/Poisson variability, and the annual variability of the rates, still exist, but they have become dominated by other uncertainties.

In short, instead of viewing probabilities in forecasting as a black and white subjective/objective dichotomy suggested by the “pregnancy dictum”, we have a gradation of shades of gray.

6 On the Interpretation of Probabilities

A remaining issue is how one might interpret the probabilities of various types. Philosophically, the problem has been much studied (*e.g.*, Kyburg 1970). It is well-known that the so-called frequency interpretation of probabilities is not a logically sound basis for defining the concept of probability. (For example, laws of large numbers presume the existence of the concept of probability for their statement and proof.) However, it does provide a useful interpretation that serves as a basis of the empirical validation of statistical models we have discussed from binomial/Poisson variation to short and even medium term forecasting. For long term forecasting it is less useful, since we are not interested in what might happen if the history were to be repeated probabilistically again and again. We only experience one sample path.

It is equally well-known that there is a logically coherent theory of subjective probabilities that underlies much of the Bayesian approach to statistics. This theory is rather more subtle than is often appreciated. As discussed by Savage (1954), for example, the theory is prescriptive in the sense that a completely rational actor would behave according to its rules. Since mere mortals are rarely, if ever, completely rational, the representation of actual beliefs in terms of subjective probabilities is a non-trivial task.

For example, actual humans rarely contemplate all possible events that might logically occur. If a person is asked about three events, he might first say that A is three times as likely as B, and B is five times as likely as C; but later say that A is ten times as likely as C. Of course, when confronted with the intransitivity of the answers, he could correct them in any number of ways, but it is not clear that the likelihood of any given event would, after adjustment, be more trustworthy than before.

Actual humans are also much less precise as “computing machines” than the idealized rational actors. Suppose, for example, that a person P says that his uncertainty about the life expectancy in Sweden in the year 2050 can be re-

presented by a normal distribution $N(100, 82)$. One can then imagine the following dialogue with a questioner Q:

Q: So you think the probability that life expectancy exceeds 110 years is over 10 per cent?

P: I don't know if I think that. If you say so.

Q: Why can't you say for sure?

P: Because I can't recall the 0.9 quantile of the standard normal distribution right now.

(Upon learning that it is 1.2816, and calculating $100 + 1.2816 \cdot 8 = 110.25$ P then agrees.)

Both difficulties suggest that any “subjective” probability statements need to be understood in an idealized sense. To be taken seriously, a person can hardly claim that he or she “feels” that some probabilities apply. Instead, careful argumentation is needed, if one were to want to persuade others to share the same probabilistic characterization (Why a mean of 100 years? Why a standard deviation of 8 years? Why a normal distributional form?).

7 Eliciting Expert Views on Uncertainty

Particular problems in the elicitation of probabilistic statements from “experts” are caused by the very fact that an expert is a person who should know how things are.

First, representing one's uncertainty truthfully may be tantamount to saying that one does not really know, if what he or she is saying is accurate. A client paying consulting fees may then deduce that the person is not really an expert! Thus, there is an incentive for the expert to downplay his or her uncertainty.

Second, experts typically share a common information basis, so presenting views that run counter to what other experts say, may label the expert as an eccentric, whose views cannot be trusted. This leads to expert flocking: an expert does not want to present a view that is far from what his or her colleagues say. An example (pointed out by Nico Keilman) is Henshels (1982, 71) assessment of the U.S. population forecasts of the 1930s and 1940s. The forecasts came too high because according to Henshel the experts talked too much to each other. Therefore, a consideration of the range of expert opinions may not give a reasonable estimate of the uncertainty of expert judgment.

Economic forecasting provides continuing evidence of both phenomena. First, one only has to think of stock market analysts making erroneous pre-

dictions with great confidence on prime time TV, week after week. Second, one can think of think-tanks making forecasts of the GDP. Often (as in the case of Finland in 2001), all tend to err on the same side.

Of course, an expert can also learn to exaggerate uncertainty, should it become professionally acceptable. However, although exaggeration is less serious than the underestimation of uncertainty, it is not harmless either, since it may discredit more reasonable views.

A third, and much more practical problem in the elicitation of probabilities from experts stems from the issues discussed in Section 5. It is difficult, even for a trained person, to express one's views with the mathematical precision that is needed. One approach that is commonly used is to translate probabilities into betting language. (These concepts are commonly used in the Anglo-Saxon world, but less so in the Nordic countries, for example.) If a player thinks that the probability is at least p that a certain event A happens, then it would be rational to accept a $p : (1 - p)$ bet that A happens. (*I.e.*, if A does not happen, the player must pay p , but if it does happen, he or she will receive $1 - p$. If the player thinks the true probability of A occurring is $\rho \geq p$, then the expected outcome of the game is $\rho(1 - p) - (1 - \rho)p = \rho - p \geq 0$.)

This approach has two problems, when applied in the elicitation of probabilities from experts. First, it is sometimes difficult to convince the experts to take the notion of a gamble seriously when they know that the "game" does not really take place. Even if the experts are given actual money with which to gamble, the amount may have an effect on the outcome. The second problem, in its standard form, the gambling argument assumes that the players are risk neutral. This may only be approximately true if the sums involved are small. If the sums are large, people tend to be more risk adverse (Arrow 1971). Moreover, experimental evidence suggests (*e.g.*, Kahneman *et al.* 1982) that people frequently violate the principle of maximizing expected utility.

The betting approach has been used in Finland to elicit expert views on migration (Alho 1998). In an effort to anchor the elicitation on something empirical, a preliminary time series model was estimated, and the experts were asked about the probability content of the model based prediction intervals for future migration. Experts had previously emphasized the essential unpredictability of migration, but seeing the intervals they felt that future migration is not as uncertain as indicated. The intervals were then narrowed down using the betting argument. In this case the use of an empirical bench mark may have lead to a higher level of uncertainty than would otherwise have been obtained.

References

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

Alho J.M. and Spencer B.D. (1985) Uncertain Population Forecasting. *J. Amer. Statist. Ass.* 80, 306–314.

Arrow K.J. (1971) *Essays in the Theory of Risk-bearing*. Chicago: Markham Publishing Co.

Box G.E.P. and Jenkins G.M. (1976) *Time Series Analysis, Revised Ed.* San Francisco: Holden-Day.

Breslow N.E. and Day N.E. (1987) *Statistical Methods in Cancer Research, Vol. II.* Lyon: I.A.R.C.

Chatfield C. (1996) *The Analysis of Time Series*. London: Chapman and Hall.

Chiang C.L. (1968) *Introduction to Stochastic Processes in Biostatistics*. New York: Wiley.

Draper D. (1995) Assessment and Propagation of Model Uncertainty (with discussion). *J. R. Statist. Soc., Series B*, 45–97.

Feller W. (1968) *An Introduction to Probability and Its Applications, Vol: I* (3rd ed.). New York: Wiley.

Freedman D., Pisani R. and Purves R. (1978) *Statistics*. New York: Norton.

Henshel R.L. (1982) Sociology and Social Forecasting. *Annual Review of Sociology* 8, 57–79.

Hoem J. (1970) *Grunnbegreper i formell befolkningslære*. Oslo: Universitetsforlaget.

Kahneman D., Slovic P. and Tversky A. (Eds.) (1982) *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kyburg H.E. (1970) *Probability and Inductive Logic*. London: Macmillan.

Lee R.D. and Auerbach A.J. (2001) *Demographic Change and Fiscal Policy*. Cambridge: Cambridge University Press.

Keyfitz N. (1982) Can knowledge improve forecasts? *Population and Development Review* 8, 719–751

Pollard J.H. (1968) A note on Multi type Galton-Watson Processes with Random Branching Probabilities. *Biometrika*, 55, 589–590.

Savage L.J. (1954) *The Foundations of Statistics*. New York: Wiley.

Shryock H.S. and Siegel J.S. (1976) *The Methods and Materials of Demography*. New York: Academic Press.

Smith J.P. (2001) Comment, in Lee and Auerbach (2001), pp. 69–72.

An Expert Knowledge Approach to Stochastic Mortality Forecasting in the Netherlands

Maarten Alders and Joop de Beer

Senior Statisticians, Department of Statistical Analysis, Statistics Netherlands

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, even if the assumptions are expected to describe the expected future according to the forecasters. It is important that users of forecasts are aware of the degree of uncertainty. In order to give accurate information about the degree of uncertainty of population forecasts Statistics Netherlands produces stochastic population forecasts. 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.

In the Dutch population forecasts the assumptions on the expected future changes in mortality primarily relate to life expectancy at birth. In the most recent Dutch forecasts assumptions underlying the medium variant¹ are based

¹ The description ‘medium variant’ originates from the former practice when several deterministic variants were published. Since no variants are published anymore it does not seem appropriate to speak of a medium variant anymore. However, abandoning this terminology would make users think that the medium variant is different from the expected value. For this reason we will still use ‘medium variant’ while we mean the expected value.

on a quantitative model projecting life expectancy at birth of men and women for the period 2001–2050. The model describes the trend of life expectancy in the period 1900–2000 taking into account the effect of changes in smoking behaviour, the effect of the rectangularization of the survival curve and the effect of some other factors on changes in life expectancy at birth. Since the model is deterministic, it cannot be used directly for making stochastic forecasts. For that reason, the assumptions underlying the stochastic forecasts are based on expert judgement, taking into account the factors described by the model.

This paper examines how assumptions on the uncertainty of future changes in mortality in the long run can be specified. More precisely, it discusses the use of expert knowledge for the specification of the uncertainty of future mortality. Section 2 briefly describes the methodology underlying the Dutch stochastic population forecasts. Section 3 provides a general discussion on the use of expert knowledge in (stochastic) mortality forecasting. Section 4 applies the use of expert knowledge to the Dutch stochastic mortality forecasts. The paper ends with the main conclusions.

2 Stochastic population forecasts: methodology

Population forecasts are 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.

Based on statistical models of fertility, mortality, and migration, statistical forecast intervals of population size and age structure can be derived, either analytically or by means of simulations. In order to obtain a forecast interval for the age structure of a population analytically a stochastic cohort-component model is needed. Application of such models, however, is very complicated. Analytical solutions require a large number of simplifying assumptions. Examples of applications of such models are given by Cohen (1986) and Alho and Spencer (1985). In both papers assumptions are specified of which the empirical basis is questionable.

Instead of an analytical solution, 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-gender-specific fertility, mortality and emigration rates, and immigration numbers are specified. Each draw results in a population by age and gender at the end of each year. Thus the simulations provide a distribution of the population by age and gender in each forecast year.

To perform 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, 1998).

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 in three ways:

- a. an analysis of errors of past forecasts may provide an indication of the size of the variance of the errors of new forecasts;
- b. estimates of the variance can be based on a statistical (time-series) model;
- c. on the basis of expert judgement values of the variance can be chosen.

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. However, in publications the role of judgement is not always made explicit.

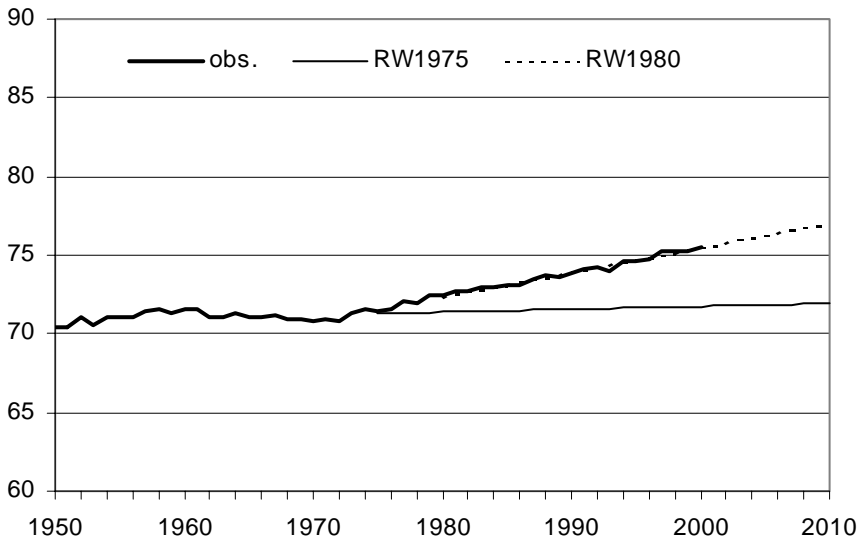
a. An analysis of errors of past forecasts

The probability of a forecast interval can be assessed on the basis of a comparison with 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 popula-

tion 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.

The question whether forecast accuracy has been increasing is important for assessing to what extent errors of past forecasts give an indication of the degree of uncertainty of new forecasts. One problem in comparing old and new forecasts is that some periods are easier to forecast than others. Moreover, a method that performs well in a specific period may lead to poor results in another period. Thus one should be careful in drawing general conclusions on the size of forecast errors on the basis of errors in a given period. For example, since life expectancy at birth of men in the Netherlands has been increasing linearly since the early 1970s, a simple projection based on a random walk model with drift would have produced rather accurate forecasts. Figure 1 shows that a forecast that would have been made in 1980 on the basis of a random walk model in which the intercept is estimated by the average change in the preceding ten years would have been very accurate for the period 1980–2000. However, this does not necessarily imply that forecasts of life expectancy are very certain in the long run. If the same method would have been used for a forecast starting in 1975 the forecast would have been rather poor (Figure 1). Thus simply comparing the forecast errors of successive forecasts does not tell us whether recent forecasts are ‘really’ better than preceding forecasts.

Figure 1 Life expectancy at birth, men, projections of random walk (RW) with drift



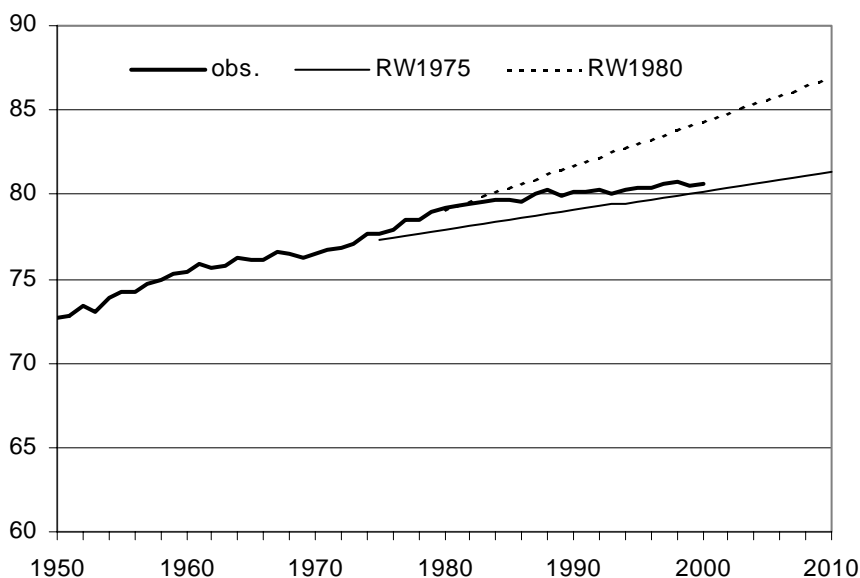
The fact that a forecast of life expectancy of men made in 1980 is more accurate than a forecast made in 1975 does not imply that recent forecasts are more accurate than older forecasts. This is illustrated by Figure 2 which shows projections of life expectancy of women. Projections of the life expectancy at birth of women based on a random walk model starting in 1980 are less accurate than projections starting in 1975.

In order to be able to assess whether new forecasts are ‘really’ better than older ones, we need to 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, in order to assess whether errors of past forecasts provide useful information about the uncertainty of new forecasts, it is important not only to measure the size of the errors but also to take into account the explanation of the errors. One main explanation of the poor development of life expectancy of men in the 1960s is the increase in smoking in previous decades, whereas the increase in life expectancy in subsequent years can partly be explained by a decrease in smoking. As women started to smoke some decades later than men, the development of life expectancy of women was affected negatively

not until the 1980s. This explains why a linear extrapolation of the trend in life expectancy of women starting in 1980 leads to overestimating the increase in the 1980s and 1990s. On the other hand a linear extrapolation of the trend in life expectancy of men starting in 1975 leads to underestimating the increase in subsequent years.

Figure 2 Life expectancy at birth, women, projections of random walk (RW) with drift

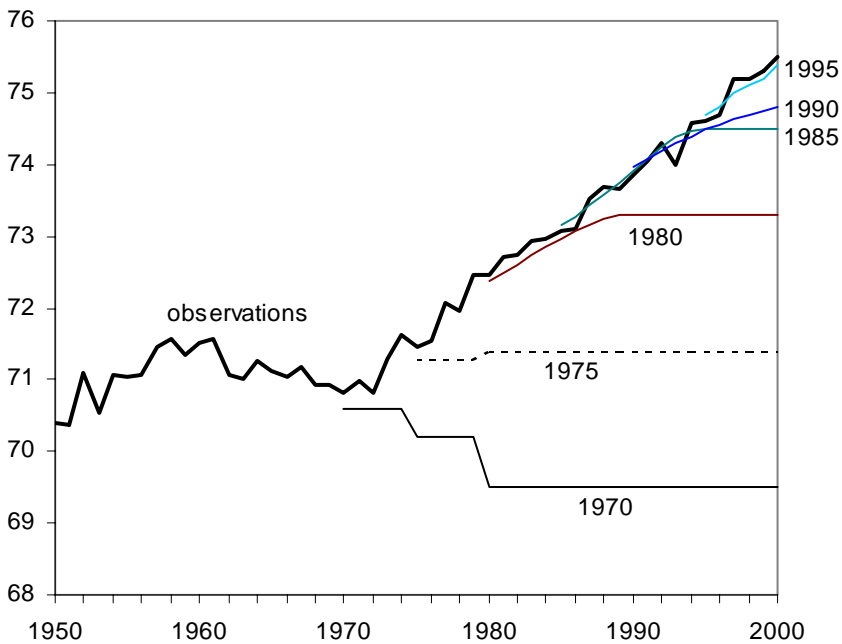


The question to what extent an analysis of past errors provides useful information about the degree of uncertainty of new forecasts depends on the question how likely it is that similar developments will occur again. The 1970-based forecasts were rather poor because forecasters did not recognize that the negative development was temporary (Figure 3). 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 the errors of the 1970-based forecasts. For that reason, the degree of uncertainty of new forecasts can be based on errors of forecasts that were made after 1970.

The decision which past forecasts to include is a matter of judgement. Thus, judgement plays a role in using errors of past forecasts for assessing the uncertainty of new forecasts. 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-run 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 run (Lutz, Goldstein and Prinz 1996). Forecast errors for the very long run result from forecasts made a long time ago. Figure 3 shows that the 30 and 25 years ahead forecasts made in 1970 and 1975 respectively are rather poor, so including these forecasts in assessing the uncertainty of new forecasts may lead to overestimating the uncertainty in the long run. Figure 3 suggests that the forecasts made in the 1980s and 1990s may well lead to smaller errors in the long run than the forecasts made in the 1970s, since the former forecasts are closer to the observations up to now than the latter forecasts were at the same forecast interval.

Figure 3 Life expectancy at birth, men, observations and historic forecasts



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

Rather than calculating errors in forecasts that were actually published, 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.

b. Model-based estimate 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 and 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 does apparently not regard the time-series model as correct. Moreover, time-series forecasting models were developed for short horizons, and they are not generally suitable for long run forecasts (Lee 1998). 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 judgement to constraint 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.

c. Expert 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 events that have not yet occurred. For example, the uncertainty of long-term forecasts of mortality depends on the probability of technological breakthroughs that may have a substantial impact on survival rates. Even though these developments may not be assumed to occur in the expected variant, an assessment of the probability of such events 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 judgement of the forecaster. With regard to mortality the assessment of the probability of unprecedented events like medical breakthroughs cannot be derived directly from models. Lutz, Sander-son, Scherbov and 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 per cent 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 the assessment of the degree of uncertainty of mortality forecasts is primarily based on expert judgement, taking into account errors of past forecasts and model-based estimates of the forecast errors.

3 Using expert knowledge

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, *e.g.*, 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 and Prinz 1996).

Forecasts of mortality can be based on extrapolation of trends in mortality indicators or on an explanatory approach. In both cases forecasters have to make a number of choices. In projecting future changes in mortality on the basis of an extrapolation of trends, one important question is which indicator is to be projected. If age- and gender-specific mortality rates are projected one may choose to assume the same change for each age (ignoring changes in the age pattern of mortality) or one may project each age-specific rate separately (which may result in a rather irregular age pattern). Instead of projecting separate age-specific mortality rates one may project a limited number of parameters of a function describing the age pattern of mortality, *e.g.*, the Gompertz curve or the Heligman-Pollard model. One disadvantage of the Heligman-Pollard model is that it includes many parameters that cannot be projected separately. This makes the projection process complex. On the other hand, the disadvantage of using a simple model with few parameters is that such models usually do not describe the complete age pattern accurately. Another possible forecasting procedure is based on a distinction of age, period and cohort effects. If cohort effects can be estimated accurately, such models may be appropriate for making forecasts for the long run. However, one main problem in using an APC model is that the distinction between cohort effects and the interaction of period and age effects for young cohorts is difficult. Finally the indicator most widely used in mortality forecasting is life expectancy at various ages, especially life expectancy at birth. Using life expectancy at birth additional assumptions have to be made about changes in the age pattern of mortality rates.

In addition to the choice of the indicator to be projected, other choices have to be made. One main question 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 or more years. Another important question is the choice of the extrapolation procedure: linear or non-linear. This question is difficult to be answered on empirical grounds: different mathematical functions may describe observed developments about equally well, but may lead to quite different projections in the long run. In summary, judgement plays an important role in extrapolations of mortality.

Instead of an extrapolation of trends forecasts of mortality 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 mortality (*e.g.*, changes in living conditions, life style, health care, safety measures, etc.) and of assumptions about both the impact of these determinants on the development of mortality and future changes in the determinants, it is concluded in which direction mortality may change. Clearly, if no quantitative model is specified, the assumptions about the future change in mortality 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.

In most developed countries life expectancy has been rising during a long period. Therefore, in assessing the uncertainty of forecasts of mortality the main question does not seem to be whether life expectancy will increase or decrease, but rather how strongly life expectancy will increase and how long the increase will continue. Basically, three types of change may be assumed. Firstly, one may assume a linear increase in life expectancy (which is not the same as a linear decrease of age-specific mortality rates) or a linear decline in the logarithm of the age-specific mortality rates. Such a trend may be explained by gradual improvements due to technological progress and increase in wealth. Secondly, one may assume that the rate of change is declining. For example, one may assume that the increase in life expectancy at birth will decline due to the fact that mortality rates for the youngest age groups are already so low that further improvements will be relatively small. More generally, the slowing down of the increase in life expectancy is related to the rectangularisation of the survival curve. Thirdly, it may be assumed that due to future medical breakthroughs life expectancy may increase more strongly than at present.

The assumption about the type of trend is not only relevant for the specification of the medium variant but also for assessing the degree of uncertainty of the forecasts. Obviously, if one assumes that trends will continue and that the uncertainty only concerns the question whether or not the rate of change will be constant or will decline gradually, the uncertainty of the future value of

life expectancy is much smaller than if one assumes that life expectancy may change in new, unprecedented directions due to medical breakthroughs.

As discussed above, one problem in determining the long-run trend in life expectancy is the choice of the base period. If one fits a mathematical function to the observed time series of life expectancy in a given period, the results may be quite different than if a model is fitted to another period. Either the estimated values of the parameters of the function may differ or even another function may be more appropriate. One cause of the sensitivity of the fitted function to the choice of the period is that part of the changes in life expectancy are temporary. For example, the increase in smoking by men in the Netherlands starting before the Second World War to a level of about 90 per cent in the 1950s and the decline in the 1960s and 1970s to a level of about 40 per cent had a significant effect on the trend in life expectancy: it caused a negative development in life expectancy in the 1960s and an upward trend in the 1980s and 1990s. It is estimated that smoking reduced life expectancy at birth around 1975 by some four years. If the percentage of smokers stabilizes at the present level, the negative effect of smoking can be expected to decline to two years. This pattern of change in mortality due to smoking is one explanation why the choice of the base period for projecting mortality has a strong effect on the extrapolation. It makes a lot of difference whether the starting year of the base period is chosen before the negative effect of smoking on life expectancy became visible or around the time that the negative effect reached its highest value. Another example of transitory changes is the decline in mortality at young ages. In the first half of the 20th century the decline in mortality of newborn children was much stronger than at present. As a consequence, life expectancy at birth increased more strongly than at present. If these transitory changes are not taken into account in fitting a function to the time series of life expectancy, the long-run projections may be biased as temporary changes are erroneously projected in the long run.

In order to avoid these problems, Van Hoorn and De Beer (2001) developed a model in which the development of life expectancy over a longer period, 1900–2000, is described by a long-term trend together with the assumed effects of smoking, the rectangularization of the survival curve, the introduction of antibiotics after the Second World War, the increase and subsequent decline in traffic accidents in the 1970s and changes in the gender difference due to other causes than smoking. The long-run trend is described by a negative exponential curve. Figure 4 shows the assumed effects of selected determinants on the level of life expectancy. Figure 5 shows that the model fits the data very well (see the appendix for a more extensive description of the model). Figure 5 also shows projections up to 2050. This model projects a

smaller increase in life expectancy of men than, *e.g.*, a linear extrapolation of the changes in the last 25 years or so would have done.

Since the model is deterministic, it cannot be used directly for making stochastic forecasts. The projections of the model are uncertain for at least two reasons. Firstly, it is not sure that the model is specified correctly. Several assumptions were made about effects on life expectancy which may be false. Secondly, in the future new unforeseen developments may occur that cannot be specified on the basis of observations. For example, future medical breakthroughs may cause larger increases in life expectancy than what we have seen so far. For this reason expert knowledge is necessary to estimate the probability and the impact of future events that have not occurred in the past.

Figure 4 **Effects on life expectancy at birth**

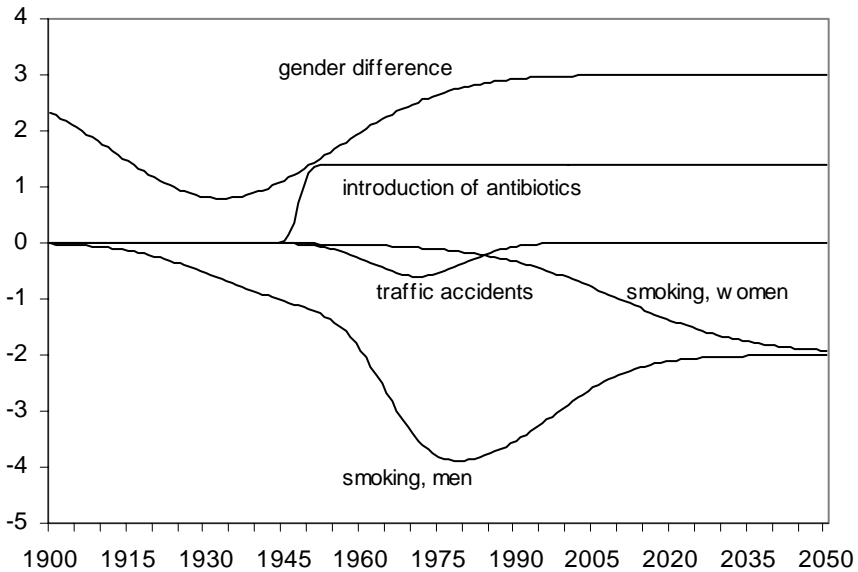


Figure 5 Life expectancy at birth: observations and model



4 Expert knowledge in the Dutch stochastic mortality forecasts

For making stochastic forecasts of mortality it is assumed that the projections of the model described in section 3 correspond with the expected values of future life expectancy. Assuming future life expectancy to be normally distributed, assumptions need to be made on the values of the standard deviations of future life expectancy.

As mentioned in section 2 both an analysis of previous forecasts 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 run. Alternatively, forecast errors for the long run 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 the standard error of forecast errors 50 years ahead is estimated at two years. This implies that the 95 per cent forecast interval for the year 2050 equals eight years. Alternatively, the standard error of forecast errors can be projected on the basis of a time-series model describing the development of life expectancy. The development of life expectancy at birth for men and women in the Netherlands can be described by a random walk model with drift (Lee and Tuljapurkar 1994,

model mortality in the United States as a random walk with drift too). The width of the 95 per cent forecast interval produced by this model for the year 2050 equals 12 years. Thus, on the basis of the time-series models of life expectancy and models of forecast errors of life expectancy it can be expected that the 95 per cent forecast interval of life expectancy in 2050 will be around 8–12 years. The decision which interval is to be used is based on judgement. Judgement ought to be based on an analysis of the processes underlying changes in life expectancy. The judgemental assumptions underlying the Dutch forecasts are based on four considerations.

(1) It is regarded highly likely that the difference in mortality between men and women will continue to decrease. This difference has arisen in the past decades largely because of differences in smoking habits. As smoking habits of men and women have become more similar, the gender difference in life expectancy is assumed to decrease in the forecasts.

(2) Changes in life expectancy at birth are the result of changes in mortality for different age groups. In assessing the degree of uncertainty of forecasts of life expectancy, it is important to make a distinction by age as the degree of uncertainty of future changes in mortality differs between age categories. The effect of the uncertainty about the future development of mortality at young ages on life expectancy at birth is only small, because of the current, very low levels of mortality at young ages. On the basis of the current age specific mortality rates, 95.3 per cent of live born men and 96.6 per cent of women would reach the age of 50. Clearly, the upper limits are not far away. According to the medium variant of the 2000 Dutch population forecasts the percentage of men surviving to age 50 will rise to 97.0 per cent in 2050 and the percentage of women to 97.4 per cent. A much larger increase is not possible. A decrease does not seem very likely either. That would, *e.g.*, imply that infant mortality would increase, but there is no reason for such an assumption. The increase in the population with a foreign background could have a negative effect on mortality, since the infant mortality rates for this population group are considerably higher than those for the native population. However, it seems much more likely that infant mortality rates for the foreign population will decline rather than that they would increase. Furthermore, the effect on total mortality is limited. Another cause of negative developments at young ages could be new, deadly diseases. The experience with AIDS, however, has shown that the probability that such developments would have a significant impact on total mortality in the Netherlands (in contrast with, *e.g.*, African countries) does not seem very large. A third possible cause of negative developments at young ages would be a strong increase in accidents or suicides. However, there are no indications of such developments. Thus it can be concluded that the effect of the uncertainty about mortality at young

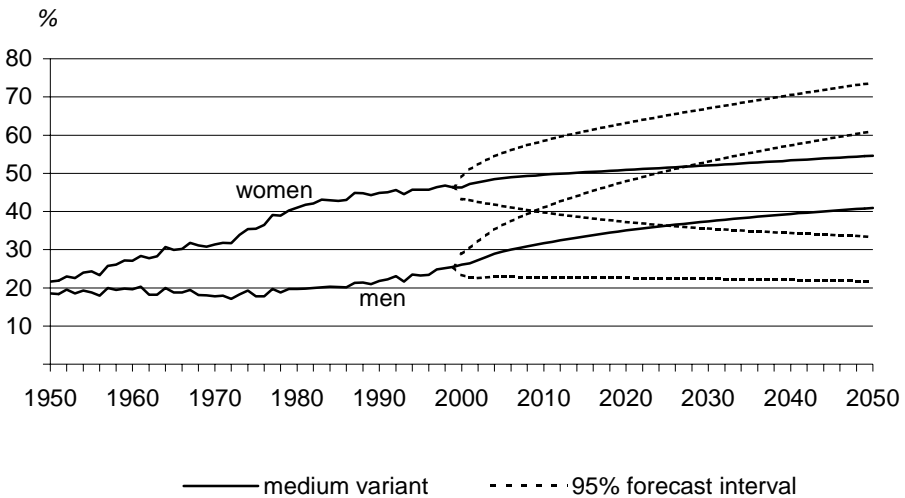
ages on the uncertainty about the future development of life expectancy at birth is limited.

(3) As regards older age groups one main assumption underlying the Dutch mortality forecasts is that the main cause of the increase in life expectancy at birth is that more people will become old rather than old people becoming still older. This implies the assumption that the survival curve will become more rectangular, an assumption based on an analysis of changes in age-specific mortality rates. The development of mortality rates for the eldest age groups in the 1980s and 1990s has been less favourable than for the middle ages. Another reason for assuming ‘rectangularisation’ of the survival curve is that expectations about a large increase in the maximum life span seem rather speculative, and even if they would become true, it is questionable whether their effect would be large during the next 50 years or so. A very strong progress of life expectancy can only be reached if life styles would change drastically or if medical technology would generate fundamental improvements (and health care would be available for everyone). Assuming a tendency towards rectangularisation of the survival curve implies that uncertainty about the future percentage of survivors around the median age of dying is relatively high. If the percentage of survivors around that age would be higher than in the medium variant (*i.e.*, if the median age would be higher), the decrease in the slope of the survival curve at the highest ages will be steeper than in the medium variant. Thus, the deviation from the medium variant at the highest ages will be smaller than around the median age. This implies that the degree of uncertainty associated with forecasts of life expectancy at birth mainly depends on changes in the median age of dying rather than on changes in the maximum life span. According to the medium variant of the 2000 population forecasts the percentage of survivors at age 85 in the year 2050 will be little under 50 per cent for men and slightly over 50 per cent for women. For that reason the degree of uncertainty of the mortality forecast is based on the assessment of a forecast interval at age 85.

(4) The last consideration concerns the important point of discussion whether medical breakthroughs can lead to an unexpectedly strong increase of life expectancy. Even in case of a significant improvement of medical technology, it will be questionable to what extent this future improvement will lengthen the life span of present generations. It should be kept in mind that the mortality forecasts are made for the period up to 2050, and thus primarily concern persons already born. Experts who think that a life expectancy at birth could reach a level of 100 years or higher usually do not indicate when such a high level could be reached. It seems very unlikely that this will be the case in the period before 2050.

The four considerations discussed above are used to specify forecast intervals. According to the medium variant assumptions on the age-specific mortality rates for the year 2050, 41 per cent of men will survive to age 85 (Figure 6). According to the present mortality rates, little more than 25 per cent of men would reach age 85. Because it is assumed to be unlikely that possible negative developments (*e.g.*, a strong increase in smoking or new diseases) will predominate positive effects of improvement in technology and living conditions during a very long period of time, the lower limit of the 95 per cent forecast interval for the year 2050 is based on the assumption that it is very unlikely that the percentage of survivors in 2050 will be significantly lower than the current percentage. For the lower limit it is assumed that one out of five men will survive to age 85. This would imply that the median age of death is 77.5 years. The upper limit of the forecast interval is based on the assumption that it is very unlikely that about two thirds of men will survive past the age of 85. The median age of death would increase to 88 years. Currently, only 16 per cent of men survive past the age of 88. A higher median age at dying than 88 seems thus very unlikely.

Figure 6 Survivors at age 85; medium variant and 95% forecast interval



As discussed above, the medium variant assumes that the gender difference will become smaller. This implies that life expectancy of women will increase less strongly than that of men. This is in line with the observed development since the early 1980s. Consequently, the probability that future life expectancy of women will be lower than the current level is higher than the

corresponding probability for men. The lower limit of the 95 per cent forecast interval corresponds with a median age at dying of 81 years, which equals the level reached in the early 1970s. This could become true, *e.g.*, if there would be a strong increase in mortality by lung cancer and coronary heart diseases due to an increase in smoking. The upper limit of the forecast interval is based on the assumption that three quarters of women will reach age 85. This would imply that half of women would become older than 91 years. This is considerably higher than the current percentage of 21. It does not seem very likely that the median age would become still higher.

The intervals for the percentage of survivors at the age of 85 for the intermediate years are assessed on the basis of the random walk model (Figure 6).

On the basis of these upper and lower limits of the 95 per cent forecast interval for percentages of survivors at age 85, forecast intervals for percentages of survivors at the other ages are assessed, based on the judgemental assumption that for the youngest and eldest ages the intervals are relatively smaller than around the median age (Figure 7). The age pattern of changes in mortality rates in the upper and lower limit are assumed to correspond with the age pattern in the medium variant.

The assumptions on the intervals of age-specific mortality rates are used to calculate life expectancy at birth. These assumptions result in a 95 per cent forecast interval for life expectancy at birth in 2050 of almost 12 years. For men the interval ranges from 73.7 to 85.4 years and for women from 76.7 to 88.5 years (Figure 8). The width of these intervals closely corresponds with that of the interval based on the random walk with drift model of life expectancy at birth mentioned before.

Figure 7 Survival curves; medium variant and 95% forecast interval

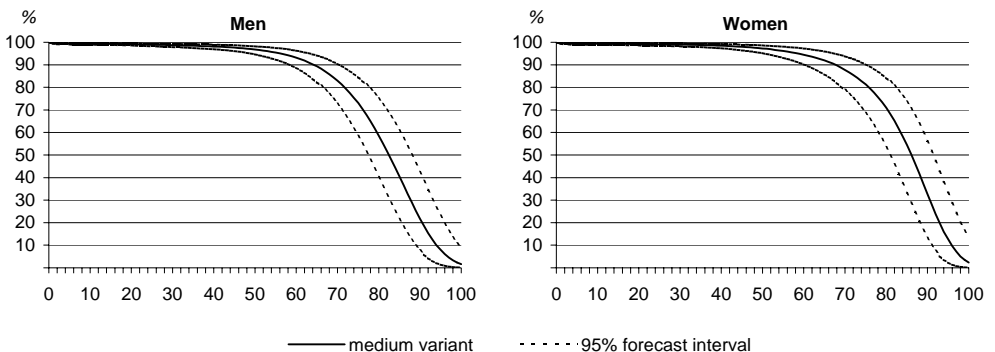
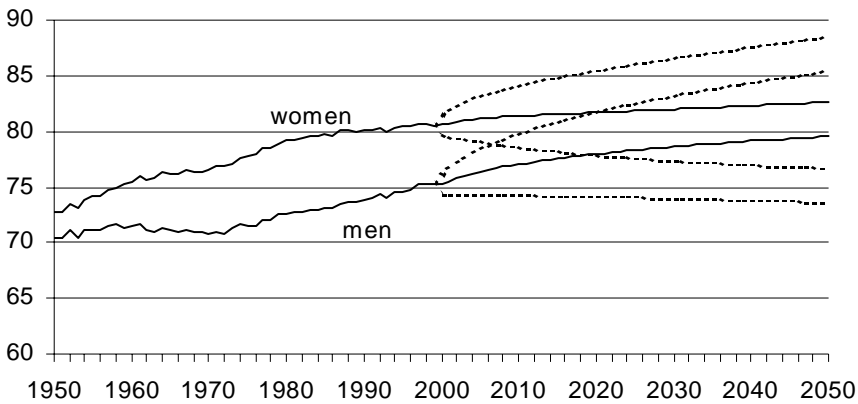


Figure 8 Life expectancy at birth; medium variant and 95% forecast interval



The intervals for the Netherlands are slightly narrower than the intervals for Germany specified by Lutz and Scherbov (1998). They assume that the width of the 90 per cent interval equals 10 years in 2030. This is based on the assumption that the lower and upper limits of the 90 per cent interval of the annual increase in life expectancy at birth equal 0 and 0.3 years respectively. This would imply that the width of the 90 per cent interval in 2050 equals about 15 years.

5 Conclusions

Long-term developments in mortality are very uncertain. To assess the degree of uncertainty of future developments in mortality 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.

The most recent Dutch mortality forecasts are based on a model that forecasts life expectancy at birth. Implementation of the model is based on literature and expert knowledge. The model includes some important determinants of mortality, such as the effect of smoking and gender differences. Since the model is deterministic, it cannot be used for stochastic forecasting. Therefore, an expert knowledge approach is followed. This approach can be described as ‘argument-based forecasting’. Basically, four quantitative assumptions are

made: (1) the difference in mortality between men and women will continue to decrease, (2) the effect of uncertainty about mortality is limited at young ages and is highest around the median age of dying, (3) the effect of medical breakthroughs on the life span will be limited up to 2050, and (4) more people will become old rather than old people will become still older (rectangularisation of the survival curve). Based on these assumptions target values for the boundaries of 95 per cent forecast intervals are specified. It appears that the width of the 95 per cent interval of life expectancy at birth in 2050 is almost twelve years, both for men and women. This interval closely resembles the interval based on a random walk model with drift. It is about four years wider than the interval based on a time-series model of errors of historic forecasts.

References

- Alho, J.M., 1998, A stochastic forecast of the population of Finland. *Reviews* 1998/4, Statistics Finland, Helsinki.
- Alho, J.M. and B.D. Spencer, 1985, Uncertain population forecasting. *Journal of the American Statistical Association* 80, 306–314.
- Box, G.E.P. and J.M. Jenkins, 1970, *Time series analysis. Forecasting and control* (Holden-Day, San Francisco).
- Cohen, J.E., 1986, Population forecasts and confidence intervals for Sweden: a comparison of model-based and empirical approaches. *Demography* 23, 105–126.
- 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.
- Harvey, A.C., 1989, *Forecasting, structural time series models and the Kalman filter* (Cambridge University Press, Cambridge).
- Keilman, N.W., 1990, *Uncertainty in population forecasting: issues, backgrounds, analyses, recommendations* (Swets & Zeitlinger, Amsterdam).
- Lee, R. and S. Tuljapurkar, 1994, Stochastic population forecasts for the United States: beyond high, medium, and low. *Journal of the American Statistical Association* 89, 1175–1189.
- Lee, R., 1998, Probability approaches to population forecasting. In: W. Lutz, J.W. Vaupel, D.A. Ahlburg (eds.), *Frontiers of Population Forecasting. A supplement to Volume 24, Population and Development Review* (Population Council, New York), 156–190.
- Lutz, W., J.R. Goldstein and C. Prinz, 1996, Alternative approaches to population projection. In: W. Lutz (ed.), *The future population of the world. What can we assume today?* (Earthscan, London), 14–44.
- Lutz, W., W. Sanderson, S. Scherbov and A. Goujon, 1996, World population scenarios for the 21st century. In: W. Lutz (ed.), *The future population of the world. What can we assume today?* (Earthscan, London), 361–396.

Lutz, W. and S. Scherbov, 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.

Van Hoorn, W. and J. De Beer, 2001, Bevolkingsprognose 2000–2050: prognosemodel voor de sterfte. *Maandstatistiek van de bevolking, Statistics Netherlands*, 49(7), 10–15.

Appendix

An explanatory model for Dutch mortality

There are several ways to explain mortality. One approach is to assume a dichotomy of determinants of mortality - internal factors and external factors. For instance age, sex and constitutional factors are internal, whereas living and working conditions as well as socio-economic, cultural and environmental conditions are external. Other factors, such as life styles and education are partly internal, partly external.

An alternative approach takes the life course as a leading principle for a causal scheme. Determinants that act in early life are placed in the beginning of the causal scheme, those that have an impact later in life are put at the end. In this way heredity comes first and medical care comes last. Factors like life styles are in the middle. In the following scheme this approach is elaborated, though some elements of the first approach are used also. Eight categories of important determinants are distinguished:

- A. Heredity (including gender)
- B. Gained properties (education, social status)
- C. Life styles (risk factors like smoking, relationships)
- D. Environment (living and working conditions)
- E. Health policy (prevention of accidents, promotion of healthy life styles)
- F. Medical care (technological progress and accessibility of cure and care)
- G. Period effects (wars, epidemics)
- H. Rectangularity of survival curve

Interactions of gender with other factors should be taken in account in forecasting mortality because a lot of differences between men and women exist. Categories A, B, C and D reflect heterogeneity in mortality in the population, while groups E, F and G reflect more general influences. As life expectancy is the dependent variable in the explanatory model, a supplementary factor (H) is needed which is dependent on the age profile of the survival curve. When the survival curve becomes more rectangular, a constant increase in life expectancy can only be achieved through ever-larger reductions of mortality rates.

Most of the eight categories listed above contain many determinants. Of course it is not possible to trace and quantify all determinants. The selection of variables is based on the following criteria:

1. there is evidence about the magnitude of the effect and about changes over time;
2. independence of effects; and
3. there is a good possibility to formulate assumptions for the future.

In category C (life styles), smoking is a good example of a suitable explanatory factor since there is considerable evidence about the prevalence of smoking in the population and the effect on mortality. In category E the effect of safety measures on death from traffic accidents is an example of an independent and relatively easy factor to estimate. The same holds for category F for the introduction of antibiotics which caused a sharp drop in mortality by pneumonia.

On the contrary, general medical progress is not a very suitable factor, because there is much uncertainty and divergence of opinions about the impact on life expectancy. The effect is hard to separate from that of social progress, growth of prosperity, cohort-effects etc.

Part of the variation in mortality (life expectancy) can be modelled by separate effects, the rest is included in the trend. It must be stressed that the model does not quantify the effect of the determinants on causes of death (for instance smoking on death rates of lung cancer and heart diseases), but directly links them with overall mortality (life expectancy).

Six determinants that meet the three criteria were included in the model. Figure 4 in the paper shows the assumptions about the effect of these determinants on life expectancy at birth in the observation period (1900–2000) and the forecast period (2001–2050).

A. Heredity: gender difference. Part of the gender difference in mortality can be attributed to differences in smoking behaviour (see C). Gender differences due to other factors were not constant through the entire 20th century. However, since 1980 the difference seems to have stabilised. The change in the gender difference in life expectancy in the 20th century can be described by a log normal curve.

B. Gained properties. There are strong differences in mortality between social groups. However, these effects were not included in the model. One reason is that there is a strong correlation with, e.g., life styles, living and work-

ing conditions and access to medical care. Thus, the second criterion mentioned above is violated.

C. Life style: smoking. On the basis of the literature the different effects for men and women are estimated. The effects can be quantified with a combination of a normal and a logistic curve. Smoking largely explains the quite different developments of life expectancy for men and women in the past decades. As smoking habits of men and women become more similar, in the forecasts the gender difference in life expectancy is assumed to decrease and finally become three years (in favour of women).

D. Environment. These effects can be characterised as gradual long-term changes. It is hard to distinguish these effects from a general long-term trend. Hence these factors are not included in the model as separate effects but are included in the trend.

E. Health policy: traffic accidents. In the early 1970s measures were taken to improve safety of traffic. As a result the number of deaths by traffic accidents declined. This effect is modelled as a deviation of the trend. Around 1970 the number of traffic accidents reached its highest level. A log normal curve appears to be appropriate for describing changes through time.

F. Medical care: introduction of antibiotics. After the Second World War the introduction of antibiotics caused a sharp decline of mortality by pneumonia. A logistic function describes the rise in life expectancy and the flattening out to a constant level.

G. Period effects: outliers. The Spanish flue and the Second World War caused sharp negative deviations of life expectancy, which are modelled by dummies.

H. Rectangularity. The shape of the survival curve has an important impact on the pace of the increase of life expectancy. If the survival curve in year t is more rectangular than in year s , a given reduction in all age-specific death rates in both years will result in a smaller increase in life expectancy in year t than in year s . We define the rectangularity effect as the growth of life expectancy in year t divided by the growth in the year 1895 caused by the same percentage of decline of all age-specific mortality rates. A linear spline function is used to smooth the results. The increase in life expectancy in 1995 appears to be only 40 per cent of that in 1895.

The new model contains a lot of parameters and simultaneous estimation can be problematic (unstable estimates). Therefore, the parameters were estima-

ted in two steps. In the first step values of parameters of the functions describing the effect of smoking, traffic accidents, the introduction of antibiotics, and rectangularity were chosen in such a way that the individual functions describe patterns that correspond with available evidence. In the second step the values of the trend parameters and the outliers were estimated on the basis of non-linear least squares and some values of parameters fitted in the first step were ‘fine-tuned’.

Several functions were tested to fit the trend. A negative exponential curve appears to fit the development since 1900 best (see Figure 5 in the paper). This function implies that there is a limit to life expectancy. However, according to the fitted model this is not reached before 2100.

The model is:

$$e_{0,g,t} = T_t + S_{g,t} + V_t + A_t + G_{g,t} + \sum_{j=1917}^{1919} u_{j,g,t} + \sum_{j=1940}^{1945} u_{j,g,t} + \varepsilon_{g,t}$$

$$T_t = a_0 + H_t R_t$$

$$H_t = a_1 e^{-a_2(t-t_1)}$$

$$R_t = 1 - b_1(t-t_2) - b_2(t-t_3) D_{1,t}$$

$$D_{1,t} = 0 \text{ if } t < t_3 \text{ and } D_{1,t} = 1 \text{ if } t \geq t_3$$

$$S_{g,t} = c_{g,1} e^{-c_{g,2}(t-t_4)^2} + \frac{c_{g,3}}{1 + c_{g,4} e^{-c_{g,5}(t-t_5)}} + \frac{c_{g,6}}{1 + c_{g,7} e^{-c_{g,8}(t-t_6)}} + c_{g,9}$$

$$V_t = d_1 e^{-d_2 \ln(t/t_7)^2}$$

$$A_t = \frac{f_1}{1 + f_2 e^{-f_3(t-t_8)}} + f_4$$

$$G_{g,t} = (h_1 e^{-h_2 \ln(t/t_9)^2} + h_3) D_{2,g}$$

$$D_{2,g} = 1 \text{ if } g = \text{female and } D_{2,g} = 0 \text{ if } g = \text{male}$$

$$u_{j,g,t} = 1 \text{ if } t = j \text{ and } u_{j,g,t} = 0 \text{ in other years,}$$

where $e_{0,g,t}$ is life expectancy at birth for gender g in year t , T is trend, S is the effect of smoking, V is the effect of traffic accidents, A is the introduction of antibiotics, G is the unexplained gender difference, u are outliers, ε is error, H is slope of the trend and R is the effect of the rectangularity of the survival curve.

Stochastic Forecasts of Mortality, Population and Pension Systems

Shripad Tuljapurkar

Professor, Stanford University, Stanford

1 Introduction

This paper discusses the construction of stochastic forecasts of human mortality and fertility rates and their use in making stochastic forecasts of pension funds. The method of mortality analysis was developed by Lee and Carter (1992), henceforth called the LC method. Lee and Tuljapurkar (1994) combined the LC method with a related fertility forecast to make stochastic population forecasts for the US. Tuljapurkar and Lee (1999) and Lee and Tuljapurkar (2000) combined these population forecasts with a number of other forecasts to generate stochastic forecasts of the US Social Security system.

My goal is to explain the distinctive features, strengths, and shortcomings of the stochastic method rather than to explain the method. I begin with a discussion of stochastic forecasts and their differences from scenario forecasts. Then I discuss mortality forecasts using Swedish mortality data, including a new forecast for Sweden. I go on to consider briefly how population forecasts are made and their use in modeling pension systems.

2 Stochastic forecasts

A population forecast made in year T aims to predict population $P(t)$ for later years, where P includes numbers and composition. The information on which the forecast is based includes the history of the population and of environmental factors (economic, social, etc.). Every forecast maps history into a prediction. Scenario forecasts rely on a subjective mapping made by an expert, whereas stochastic forecasts attempt to make an explicit model of the historical dynamics and project this dynamic into the future. Stochastic forecasts may rely partly on a subjective mapping as well. What are the pros and cons of the two approaches?

When historical data contain a strong “signal” that describes the dynamics of a process, it is essential to use the signal as a predictive mechanism. Equally, it is important to include information that is not contained in the signal – this residual information is an important element of uncertainty that should be reflected in the forecast. The LC method shows that there is such a signal in mortality history. When there is no strong signal in the historical data, a subjective prediction may be unavoidable. Fertility history tends to reveal relatively little predictive signal. Even here, uncertainty ought to be included because history does tell us about uncertainty, and we can estimate the variability around a subjective prediction.

The use of history to assess uncertainty certainly does make assumptions about persistence in the dynamic processes that drive the variables we study. This does not imply that we assume an absence of surprises or discontinuities in the future. Rather it assumes that all shocks pass through a complex filter (social, economic, and so on) into demographic behavior, and that future shocks will play out in the same statistical fashion as past shocks. I would not abandon this assumption without some demonstration that the filtering mechanisms have changed – witness for example the stock market bubble in the US markets in 1999–2000 and its subsequent decline. It may be useful to think about extreme scenarios that restructure aspects of how the world works – one example is the possibility that genomics may change the nature of both conception and mortality in fundamental ways – but I regard the exploration of such scenarios as educational rather than predictive.

I argue strongly for the systematic prediction of uncertainty in the form of probability distributions. This position does not argue against using subjective analysis where unavoidable. One way of doing a sound subjectively based analysis is to follow the work of Keilman (1997, 1998) and Alho and Spencer (1997) and use a historical analysis of errors in past subjective forecasts to generate error distributions and project them. The practice of using “high-low” scenarios should be avoided. Uncertainty accumulates, and must be assessed in that light. In my view, the best that a scenario can do is suggest extreme values that may apply at a given time point in the future – for example, demographers are often reluctant to believe that total fertility rate (TFR) will wander far from 2 over any long interval, so the scenario bounds are usually an acceptable window around 2, such as 1.5 to 2.2. Now this may be plausible as a period interval in the future but in fact tells us nothing useful about the dynamic consequences of TFR variation over the course of a projection horizon.

Uncertainty, when projected in a probabilistic manner, provides essential information that is as valuable as the central location of the forecast. To start

with, probabilities tell us how rapidly the precision of the forecast degrades as we go out into the future. It can also be the case that our ability to predict different aspects of population may differ, and probability intervals tell us about this directly. Probabilities also make it possible to use risk metrics to evaluate policy: these are widely used in insurance, finance, and other applications, and surely deserve a bigger place in population-related planning and analysis.

3 Mortality forecasts

The LC method seeks a dominant temporal “signal” in historical mortality data in the form of the simplest model that captures trend and variation in death rates, and seeks it by a singular-value decomposition applied to the logarithms $\log m(x,t)$ of central death rates. For each age x subtract the sample average $a(x)$ of the logarithm, and obtain the decomposition

$$\log m(x,t) - a(x) = \sum_i s_i u_i(x) v_i(t).$$

On the right side above are the real singular values $s_1 \geq s_2 \dots \geq 0$. The ratio of s_1^2 to the sum of squares of all singular values is the proportion of the total temporal variance in the transformed death rates that is explained by just the first term in the singular-value decomposition.

In all the industrialized countries that we have examined, the first singular value explains well over 90 per cent of the mortality variation. Therefore we have a dominant temporal pattern, and we write

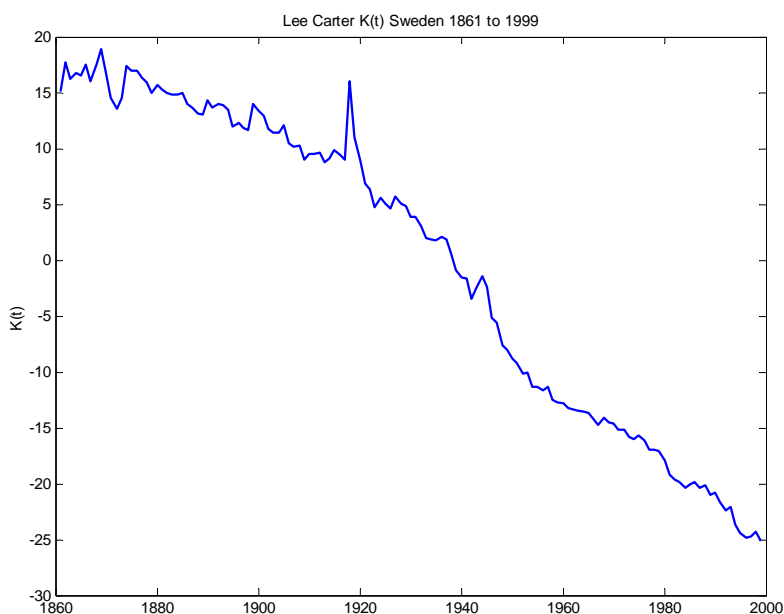
$$\log m(x,t) = a(x) + b(x) k(t) + E(x,t).$$

The single factor $k(t)$ corresponds to the dominant first singular value and captures most of the change in mortality. The far smaller variability from other singular values is $E(x,t)$.

The dominant time-factor $k(t)$ displays interesting patterns. Tuljapurkar, Li and Boe (2000) analyzed mortality data in this way for the G 7 countries over the period from approximately 1955 to 1995. They found that the first singular value in the decomposition explained over 93 per cent of the variation, and that the estimated $k(t)$ in all cases showed a surprisingly steady linear decline in $k(t)$. The mortality data for Sweden from 1861 to 1999 constitute one of the longest accurate series, and a similar analysis in this case reveals two regimes of change in $k(t)$. The estimated $k(t)$ for Sweden is shown in Figure 1. There is steady decline from 1861 to about 1910 and after 1920

there is again steady decline but at a much faster rate. Note that the approximately linear long-term declines are accompanied by quite significant short-term fluctuations. It is possible that we can interpret period-specific fluctuations in terms of particular effects (*e.g.*, changes in particular causes of death) but it is difficult to project these forward. For example, the change in the pattern in the early 1900s is consistent with our views of the epidemiological transition, but we do not know if the future will hold such a qualitative shift. Within the 20th century we take the approach of using the dominant linearity coupled with superimposed stochastic variation.

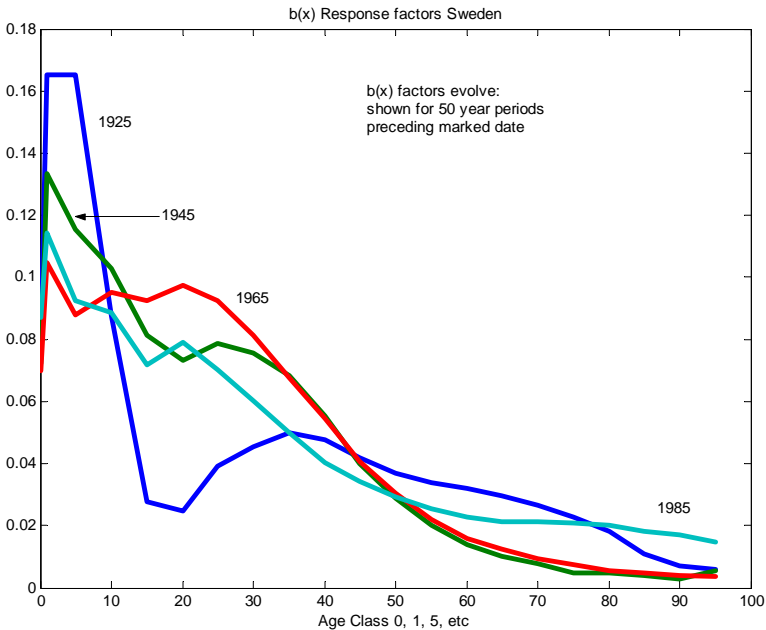
Figure 1



Mortality decline at any particular age x is proportional to the signal $k(t)$ but its actual magnitude is scaled by the response profile value $b(x)$. Figure 2 shows the $b(x)$ profiles computed for Swedish data using 50 year spans preceding the dates 1925, 1945, 1965, and 1985. Note that there is a definite time evolution, in which the age schedules rotate (around an age in the range 40 to 50) and translate so that their weight shifts to later ages as time goes by. This shifting corresponds to the known sequence of historical mortality decline starting with declines initially at the youngest ages and then in later ages over time. An intriguing possibility is that temporal changes in the $b(x)$ schedules may be described by a combination of a scaling and translation – a

sort of nonlinear mapping over time. An important matter for future work is to explore the time evolution of the $b(x)$, even though it appears (see below) that one can make useful forecasts over reasonable time spans of several decades by relying on a base span of several decades to estimate a relevant $b(x)$.

Figure 2



What accounts for the regular long-term decline in mortality that is observed over any period of a few decades? It is reasonable to assume that mortality decline in this century has resulted from a sustained application of resources and knowledge to public health and mortality reduction. Let us assume, as appears to be the case, that societies allocate attention and resources to mortality reduction in proportion to observed levels of mortality at different ages (*e.g.*, immunization programs against childhood disease, efforts to reduce cardiovascular disease at middle age). Such allocation would produce an exponential (proportional) change in mortality, though not necessarily at a constant rate over time. Over time, the rate of proportional decline depends on a balance between the level of resources focused on mortality reduction, and their marginal effectiveness. Historically, the level of resources has increased over time but their marginal effectiveness has decreased over time (because, for example, we are confronted with ever more complex causes of mortality that require substantial resources or new knowledge). The observa-

tion of linearly declining $k(t)$ – roughly constant long-run exponential rates of decline – implies that increasing level and decreasing effectiveness have balanced each other over long times. It is of course possible that the linear pattern of decline we report has some other basis. For the future, we expect a continued increase in resources spent on mortality reduction, and a growing complexity of causes of death. The balance between these could certainly shift if there were departures from history – for example, if new knowledge is discovered and translated into mortality reductions at an unprecedented rate. But this century has witnessed an amazing series of discoveries that have altered medicine and public health, and there is no compelling reason why the future should be qualitatively different. Therefore, I expect a continuation of the long-run historical pattern of mortality decline.

The LC method uses the long-term linear decline in $k(t)$ to forecast mortality. A naive forecast based on the long-run trend is not sensible because the short-term variation will accumulate over time, so it is essential to employ a stochastic forecast. In LC, the stochastic term $E(t)$ is modeled as a stationary noise term, and this procedure leads to forecasts for Sweden as shown in Figure 3, for life expectancy at birth, e_{00} , and in Figure 4 for life expectancy at age 65, e_{65} . In both cases we use a 50-year span of historical data prior to a launch date of 1999. The intervals shown are 95 per cent prediction intervals for each forecast year. Notice that there are separate forecasts for each sex, as well as a combined-sex forecast. The joint analysis of the two sexes in an LC framework has not been fully resolved, although Li *et al.* (2004) suggest one method for doing this.

Figure 3

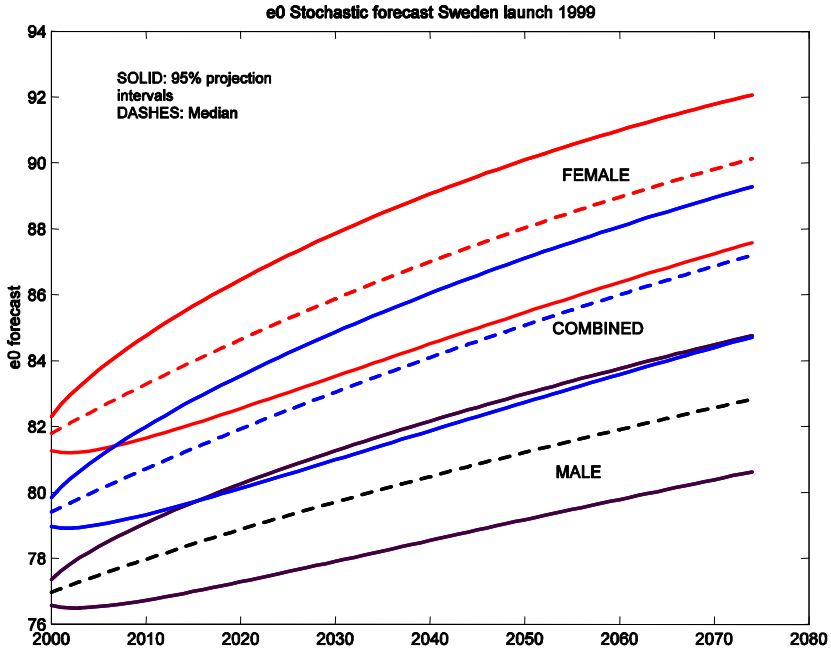
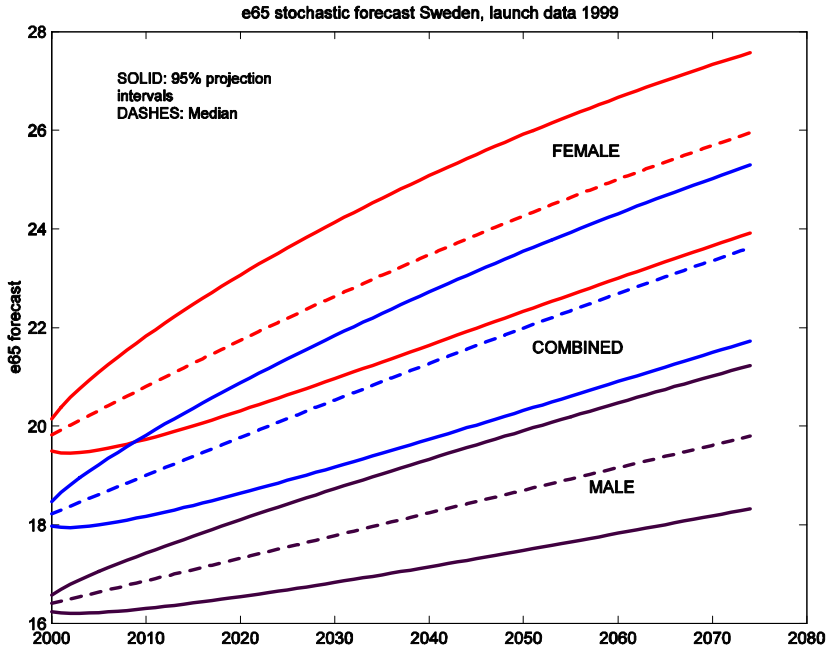


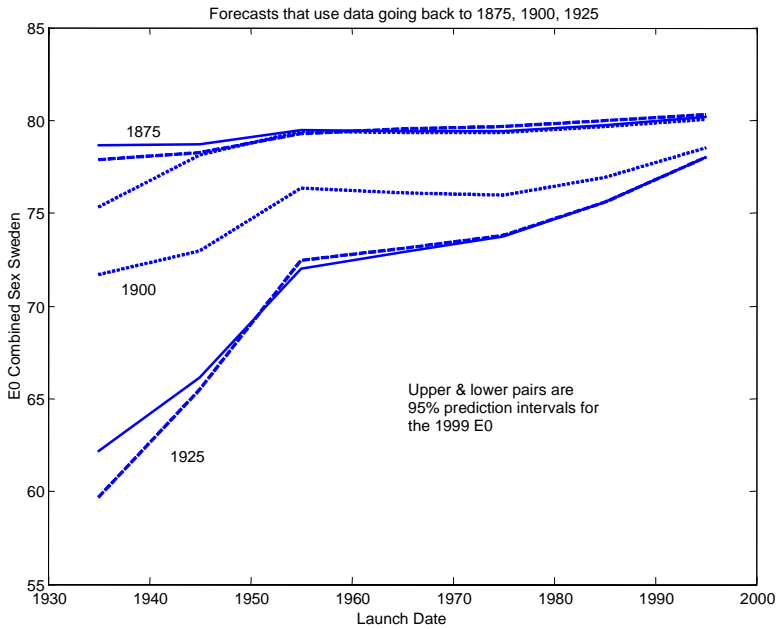
Figure 4



Some previous comments on the LC method have asserted that it is simply equivalent to a linear extrapolation in the log scale of the individual rates at each age, but it is not. For one thing, the extrapolations would include elements of the $E(t)$ terms in each age, and these may be larger at some ages than at others. For another, I take the stochastic variation seriously as an integral part of the forecast, and the realized long run trend can be rather different depending on where in the prediction interval one ends up. Without this variability, the forecasts would not be terribly useful over long horizons.

To illustrate the robustness of the LC method, Lee and Miller (2001) have analyzed the performance of the method using internal validation. A more extensive analysis for Sweden echoes their finding that the method is surprisingly robust. To illustrate, I use different base periods to forecast e_0 in 1999. I first select a starting base year, say 1875, and then a launch year which is chosen from the set 1935, 1945, ..., 1995; this gives a total of seven forecasts starting in 1875. We expect that a forecast for 1999 using the 1875 to 1935 base period would be much less accurate than a forecast which uses the 1875 to 1995 base period. The object of the exercise is to see whether the projection intervals for e_0 in 1999 will decrease in some systematic way as we include more recent (relative to 1999) history and whether they speak to the accuracy of the method. Figure 5 plots the projection intervals obtained in this way, using each of three starting years (1875, 1900, or 1925) and the seven launch years indicated above, so for each starting year we have an upper and lower prediction "fan" for e_0 in 1999. The figure shows that as we use more recent histories, we close in on the true 1999 value of e_0 of 79.4 years – the 95 per cent prediction interval brackets the true value most of the time which is impressive especially when compared with the historical performance of scenario forecasts. From a practical point of view, the prediction interval width is under seven years for launch dates from 1960 to 1980 and any of the starting base years. This means that we may expect a reasonable performance from LC forecasts for as far as 40 years into the future.

Figure 5



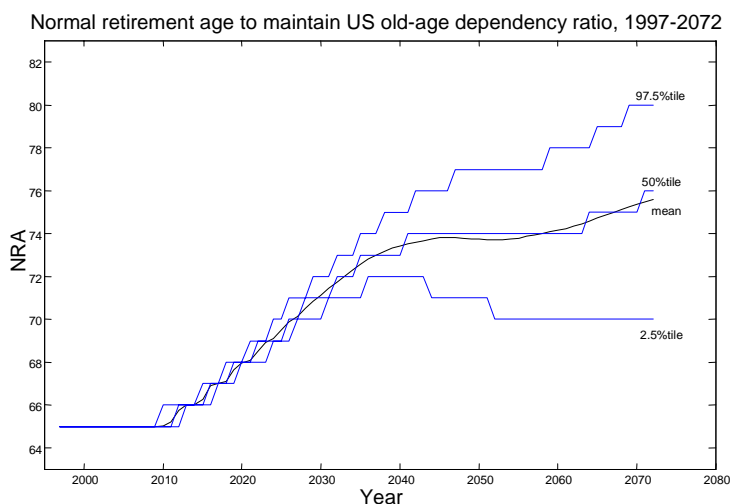
4 From population to pension systems and policy

For a population forecast we must supplement mortality forecasts with similar forecasts for fertility and if necessary for immigration. These elements can then be combined in the usual cohort-component procedure to generate stochastic population forecasts. Fertility forecasts pose special challenges because there does not seem to be a strong temporal pattern to fertility dynamics. Lee and Tuljapurkar (1994) use time series models for fertility to make stochastic forecasts for the US. Their simple models have been considerably extended by Keilman and Pham (2004) who suggest several ways of modeling and constraining the volatility of fertility forecasts.

How can stochastic forecasts be used in analyzing pension policy? At a purely demographic level, it is well known that the old-age dependency ratio is the key variable that underlies pension costs. As the old-age dependency ratio for a population increases, the more retirees-per-worker there are in the population, which implies greater stress on a pay-as-you-go pension system which relies on today’s workers to pay the benefits of today’s retirees. An interesting insight into the demographic impact of aging on the dependency ratio can be created by asking the following question. Suppose that the age at which people retire is, *e.g.*, 65. If this “normal retirement age” age cutoff

could be changed arbitrarily, how high would we have to raise it in order to keep the dependency ratio constant? If we have a population trajectory forecast, then we can simply compute in each year the retirement age, say $R(t)$, at which the old-age dependency ratio would be the same ratio as in the launch year. When we have stochastic forecasts, there is, in each forecast year t , a set of random values $R(t)$; in our analysis we look for the integer value of $R(t)$ that comes closest to yielding the target dependency ratio. Figure 6 shows the results of computing these stochastic $R(t)$ for the US population. What is plotted is actually three percentiles of the distribution of $R(t)$ in each year, the median value, and the upper and lower values in a 95 per cent projection interval. The plots show some long steps because the dependency ratio distribution changes fairly slowly over time. The smooth line shows the average value of $R(t)$ for each forecast year, which is surprisingly close to the median. Observe, for example, that there is a 50 per cent chance that the “normal retirement age” would have to be raised to 74 by 2060 in order to keep the dependency ratio constant at its 1997 value. There is only a 2½ per cent chance that the “normal retirement age” of 69 years would suffice. Given that current US Social Security policy is only intended to raise the “normal retirement age” to 67 years, and that even the most draconian proposals would only raise it to 69 years, we conclude that changes in the “normal retirement age” are very unlikely to hold the dependency ratio constant. Anderson, Tuljapurkar and Li (2001) present similar results for the G7 countries. In Tuljapurkar and Lee (1999) there are additional examples of how stochastic forecasts can be combined with objective functions to analyze fiscal questions.

Figure 6



To go beyond this type of analysis we need a full model of the structure of a pension system which may be “fully funded” or “pay-as-you-go” or some mixture. Many systems, in order to operate with a margin of security, are modified versions of pay-as-you-go systems that include a reserve fund. In the United States the OASDI (Old Age Survivors and Disability Insurance, or Social Security) Trust Fund, the holdings of the system are federal securities, and the “fund” consists of federally-held debt. A fund balance earns interest, is subject to withdrawals in the form of benefit payments, and receives deposits in the form of worker contributions (usually in the form of tax payments). Lee and Tuljapurkar (2000) discuss such models for the US Social Security system and also for other fiscal questions. The dynamics of such models proceed by a straightforward accounting method. Starting with a launch year (initial year) balance, we forecast contributions and benefit payments for each subsequent year, as well as interest earned. This procedure yields a trajectory of fund balance over time. Future contributions depend both on how many workers contribute how much to the system. Future benefit payments depend on how many beneficiaries receive how much in the future. Our population forecasts do not directly yield a breakdown in terms of workers and retirees. Therefore, we estimate and forecast per-capita averages by age and sex, for both contributions and benefits. We combine these age and sex-specific “profiles” with age and sex-specific population forecasts to obtain total inflows and outflows for each forecast year.

Contribution profiles evolve over time according to two factors. First, increases in contributions depend in turn on increases in the real wage. We forecast real wage increases stochastically (as described below), and contributions increase in proportion to wages. Second, changes in the labor force participation rates also affect contributions; we forecast labor force participation rates deterministically. Benefit profiles evolve over time in response to several factors. In our model of the U.S. Social Security system, we disaggregate benefits into disability benefits and retirement benefits. Retirement benefit levels reflect past changes in real wages because they depend on a worker's lifetime wages. Also, legislated or proposed changes in the Normal Retirement Age (the age at which beneficiaries become eligible to collect 100 per cent of their benefits) will reduce benefits at the old NRA.

Demographic variables are obviously not the only source of uncertainty facing fiscal planners; there are sizable economic uncertainties as well. Taxes and future benefits usually depend on wage increases (economic productivity) and funds can accumulate interest or investment returns on tax surpluses. Our models combine uncertainty in productivity and investment returns by converting productivity to real 1999 dollars, subtracting out increases in the CPI. We then model productivity rates and investment returns stochastically.

There is substantial correlation between interest rates on government bonds and returns to equities, so it is important to model these two variables jointly. For our historical interest rate series we use the actual, effective real interest rate earned by the trust fund, and for historical stock market returns we use the real returns on the overall stock market as a proxy. These two series are modeled jointly as a vector auto-regressive process.

Our stochastic model allows us to simulate many (1000 or more, usually) trajectories of all variables and obtain time trajectories of the fund balance from which we estimate probabilities and other statistical measures of the system's dynamics. This method may be used to explore the probability that particular policy outcomes are achieved, for example, that the "fund" stays above a zero balance for a specified period of years, or that the level of borrowing by the fund does not exceed some specified threshold.

Acknowledgements

My work in this area involves a long-term collaboration with Ronald Lee and Nan Li. My work has been supported by the US National Institute of Aging.

References

- Alho, J, and Spencer, B. D. (1997) The Practical Specification of The Expected Error of Population Forecasts, *Journal of Official Statistics*, 13, 201–225.
- Anderson, M., S. Tuljapurkar, and N. Li (2001) How accurate are demographic projections used in forecasting pension expenditure? IN T. Boeri, A. Borsch-Supan, A. Brugiavini, R. Disney, A. Kapteyn and F. Peracchi, eds., *Pensions: More Information, Less Ideology*. Kluwer Academic Publishers, pp. 9–27.
- 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 (Supplement):15–41.
- Keilman, N and D. Q. Pham (2004) Time series errors and empirical errors in fertility forecasts in the Nordic countries. *International Statistical Review* 72(1), 5–18.
- Lee, R. and L. Carter (1992) Modeling and Forecasting U.S. Mortality, *Journal of the American Statistical Association* 87(419), 659–671.
- Lee, R. and T. Miller (2001) Evaluating the Performance of the Lee-Carter Approach to Modeling and Forecasting Mortality, *Demography*, 38(4), 537–549.
- Lee, R. and S. Tuljapurkar (1994) Stochastic Population Projections for the U.S.: Beyond High, Medium and Low, *Journal of the American Statistical Association*, 89(428), 1175–1189.
- Lee, R. and S. Tuljapurkar (2000) Population Forecasting for Fiscal Planning: Issues and Innovations, in A. Auerbach and R. Lee, eds., *Demography and Fiscal Policy* (Cambridge University Press).
- Li, N., R. Lee, and S. Tuljapurkar (2004) Using the Lee-Carter Method to Forecast Mortality for Populations with Limited Data. *International Statistical Review*, 72: 19–36.
- Tuljapurkar, S. and R. Lee (1999) Population Forecasts, Public Policy, and Risk, *Bulletin of the International Statistical Institute, Proceedings*, 52nd annual meeting, ISI, Helsinki, August 1999.
- Tuljapurkar, S., N. Li and C. Boe (2000) A universal pattern of mortality decline in the G7 countries, *Nature* 405, 15 June 2000, 789–792.

This volume is the second in a series on mortality forecasting reporting proceedings of a series of workshops, organized by the Stockholm Committee on Mortality Forecasting and sponsored by the Swedish Social Insurance Agency.

The current volume addresses the issue of probabilistic models – why are mortality forecasts erroneous, what are the underlying statistical mechanisms, and how is stochastic mortality forecasting done in practice? Empirical illustrations are given for Sweden, the Netherlands, and the USA.

[Stockholm Committee
on Mortality Forecasting](#)

Professor Edward Palmer
Professor Tommy Bengtsson
Professor Juha M. Alho
Professor Kaare Christensen
Professor Nico Keilman
Professor James W. Vaupel

