

Chapter 9

Erroneous Population Forecasts



Nico Keilman

9.1 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 9.1 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 Fig. 9.1 appear small. The highest forecast came out in 1973. That forecast numbered 6.49 billion, only 6% 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% 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 30-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

N. Keilman (✉)

Department of Economics, University of Oslo, Oslo, Norway

e-mail: nico.keilman@econ.uio.no

© The Author(s) 2019

T. Bengtsson, N. Keilman (eds.), *Old and New Perspectives on Mortality*

Forecasting, Demographic Research Monographs,

https://doi.org/10.1007/978-3-030-05075-7_9

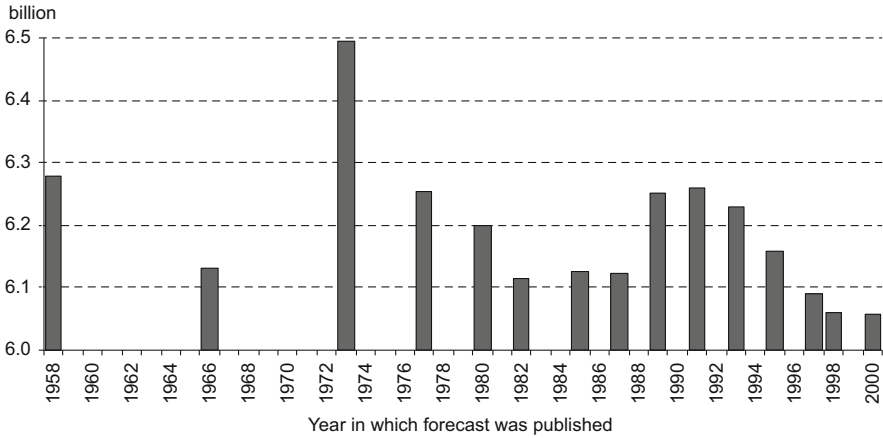


Fig. 9.1 Zooming in on the year 2000 world – population at the end of the twentieth century

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 subnational 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 Cruijssen 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.

9.2 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 1812–1829). 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.

9.3 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; 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.

9.3.1 *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% on average for 5-year ahead forecasts, to 9% 15 years ahead, and to 14% 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) %. 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: 5%. 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 5 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 5 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 (5 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.

9.3.2 *Forecasts Are More Accurate for Large Than for Small Populations*

A size effect in empirical errors at the sub national level was established already 50 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% 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.

9.3.3 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% or more for forecasts 15 years into the future.

9.3.4 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 Sect. 9.3.3, 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.

9.4 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

¹A cohort component forecast that has 1-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 50 years, a forecast for a period of 50 years, say, still requires that one specify the joint statistical distribution of $(7 + 40 + 28) \times 10 = 750$ parameters.

²For 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.

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% 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, b) perfect autocorrelation was assumed for the summary parameters (total fertility, life expectancy, and net migration). This 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, b). 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% coverage versus 95% or 99% (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.

9.5 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 RogersCastro model for migration. Next, the results of the 5000 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 1-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 5000 simulated values, and read off prediction intervals with any chosen coverage probability.

The results showed odds equal to four against one (80% 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 9.2 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 relative to that aged 20–66.³ The prediction intervals are those with 95%, 80%, and 67% 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% 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.

³The legal retirement age in Norway is 67.

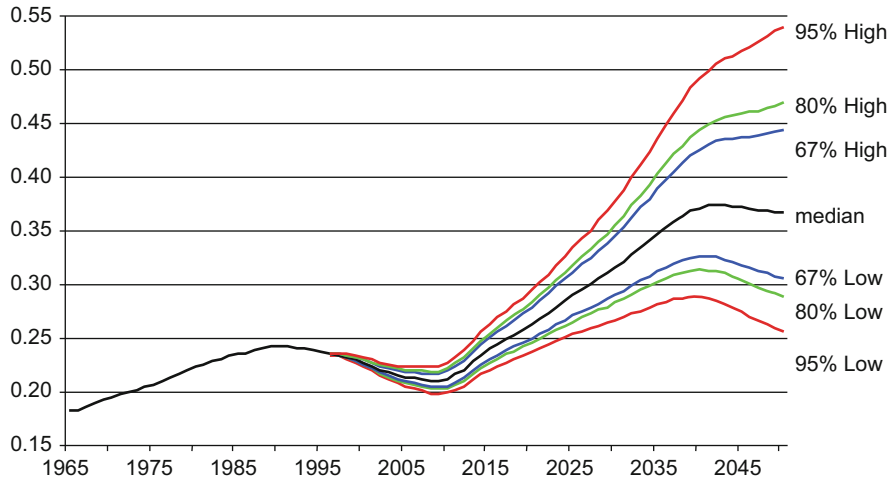


Fig. 9.2 Old age dependency ratio, Norway

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 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 Fig. 9.2 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

⁴The 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 4 years earlier than Statistics Norway’s forecast.

Table 9.1 Prediction intervals for retirement age, Norway

	Median	67% interval	80% interval	95% 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

low in one year, it is 100% 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 Sect. 12.4 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 Fig. 9.2). 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 structure is uncertain, the retirement age necessary to obtain a constant OADR becomes a stochastic variable. Table 9.1 gives the results.

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% 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% prediction interval indicates that the probability that we may see an increase is about 10% or lower, given the assumptions made.

9.6 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 50 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 et al. (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; Alho 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 et al. (1996, 2001), who used a probabilistic cohort component approach for 13 regions of the world.⁵ For fertility and mortality, they combined the three methods mentioned in Sect. 9.4 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% prediction bounds corresponding to certain high and

⁵Probabilistic 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.

low migration gains in each region. In the recent study, net migration 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*. Helsinki: Statistics Finland.
- Alho, J., & Spencer, B. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, 80, 306–314.
- Armstrong, J. (1985). *Long-range forecasting: From crystal ball to computer* (2nd ed.). New York: Wiley.
- Ascher, W. (1978). *Forecasting: An appraisal for policy-makers and planner*. Baltimore: The Johns Hopkins University Press.
- Bandel Bäckman, J., & Schéele, S. (1995). Uncertainty in population forecasts for small areas. In C. Lundh (Ed.), *Demography, economy and welfare. Lund studies in economic history* (Vol. 1, pp. 419–434). Lund: Lund University Press.
- Boudon, R. (1986). *Theories of social change: A critical appraisal*. Cambridge: Polity Press.
- Calot, G., & Chesnais, J. C. (1978). La prevision démographique. In A.-C. Dans (Ed.), *Decoufflé 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., & Alders, M. (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., Lutz, W., & Scherbov, S. (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).
- Inoue, S., & Yu, Y. C. (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*, 24(Supplement), 15–41.
- Keilman, N. (2000). Précision et incertitudes des prévisions nationales de population 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* (pp. 33–67). 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., & Cruijssen, H. (Eds.). (1992). *National population forecasting in industrialized countries*. Amsterdam etc.: Swets & Zeitlinger.
- Keilman, N., & Pham, D. Q. (2004). *Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European economic area* (Discussion Paper 386). Statistics Norway.
- Keilman, N., Pham, D. Q., & Hetland, A. (2001). *Norway's uncertain demographic future* (Social and Economic Studies 105). Oslo: Statistics Norway. Internet www.ssb.no/english/subjects/02/03/sos105_en
- Keilman, N., Pham, D. Q., & Hetland, A. (2002, May 6–15). Why population forecasts should be probabilistic – Illustrated by the case of Norway. *Demographic Research*, 409–453. Internet www.demographic-research.org/volumes/vol6/15/6-15.pdf
- 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, & D. Ahlburg (Eds.), *Frontiers of population forecasting* (Supplement to Vol. 24 of Population and development review) (pp. 156–190). New York: Population Council.
- Lee, R., & Tuljapurkar, S. (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: Cambridge University Press.
- Lutz, W., & Scherbov, S. (1998a). An expert-based framework for probabilistic national population projections: The example of Austria. *European Journal of Population*, 14, 1–17.
- Lutz, W., & Scherbov, S. (1998b). Probabilistische Bevölkerungsprognosen für Deutschland (Probabilistic population projections for Germany). *Zeitschrift für Bevölkerungswissenschaft*, 23, 83–109.
- Lutz, W., Sanderson, W., & Scherbov, S. (1996). Probabilistic population projections based on expert opinion. In W. Lutz (Ed.), *The future population of the world: What can we assume today?* (rev. ed., pp. 397–428). London: Earthscan.
- Lutz, W., Sanderson, W., & Scherbov, S. (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* (Vol. 29, p. 716). New York: Harcourt, Brace and World.
- National Research Council. (2000). Beyond six billion: Forecasting the world's population. Panel on population projections. In J. Bongaarts & 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. *Allgemeine Statistisches 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 Utredningsoch Statistikkontor.
- Smith, S. K., Tayman, J., & Swanson, D. (2001). *State and local population projections: Methodology and analysis* (The Plenum series on demographic methods and population analysis). New York: Kluwer Academic/Plenum Publishers.
- 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., Schafer, E., & Carter, L. (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. (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.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

