Nico Keilman, Dinh Quang Pham,
and Arve Hetland

Norway's Uncertain Demographic Future

Statistisk sentralbyrå • Statistics Norway
Oslo – Kongsvinger

Sosiale og økonomiske studier	Serien Sosiale og økonomiske studier omfatter nye forsknings-bidrag – monografier og redigerte arbeider – på de områder Statistisk sentralbyrå har forskningsvirksomhet. Analysemetoder og temavalg vil variere, men hovedsakelig vil arbeidene være av anvendt og kvantitativ natur med vekt på utnytting av SSBs data i analyser for samfunnsplanleggingsformål og til allmenn forståelse av sosial og økonomisk utvikling.
Social and Economic Studies	The series *Social and Economic Studies* consists of hitherto unpublished studies in economics, demography and other areas of research in Statistics Norway. Although the studies will vary in analytical methods and in subject matter, they tend to be applied studies based on quantitative analysis of the data sources of Statistics Norway. The research programmes from which the studies originate typically emphasize the development of tools for social and economic planning.

ISBN 82-537-5002-1
ISSN 0801-3845

Emnegruppe
02.03 Framskrivninger

Trykk: Lobo Media/500

Preface

This project was made possible by grant nr. 114055/730 from the Norwegian Research Council. We have benefited from useful discussions with Juha Alho, Joop de Beer, Helge Brunborg, Joel Cohen, Bill Bell, Wolfgang Lutz, Sergei Scherbov, Leiv Solheim, Ewa Tabeau, Evert van Imhoff and Lars Østby. Kluwer Academic Publishers kindly granted us the right to include in this report parts of our paper "Predictive intervals for age-specific fertility" that appeared in the European Journal of Population Volume 16 no 1 of 2000.

Oslo, August 2001

Abstract

Nico Keilman, Dinh Quang Pham, Arve Hetland

Norway's Uncertain Demographic Future

Social and Economic Studies 105 • Statistics Norway 2001

The demographic future of any population is uncertain, but some of the many possible trajectories are more probable than others. Therefore, an exploration of the demographic future should include two elements: a *range* of possible outcomes, and a *probability* attached to that range. Together, these two constitute a prediction interval for the population variable concerned. This report presents the findings of a research project, the aim of which was to compute prediction intervals for the future population of Norway broken down by age and sex to the horizon 2050.

We estimate that the odds are 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. This illustrates that uncertainty increases with time. There is a clear trade-off between greater accuracy (higher odds) and higher precision (narrower intervals). Odds of 19 against one (95 per cent chance) result in a wider interval: 4.1-5.7 million in 2025, and 3.2-7.3 million in 2050. The probabilistic population forecasts of the youngest and the oldest age groups show largest uncertainty, because fertility and mortality are hard to predict. As a result, prediction intervals in 2030 for the population younger than 20 years are so wide, that the forecast is not very informative. International migration shows large prediction intervals around expected levels, but its impact on the age structure is modest. In 2050, uncertainty has cumulated so strongly, that intervals are very large for virtually all age groups, in particular when the intervals are judged in a relative sense (compared to the median forecast). According to our statistical model, the expected accuracy of the total population size forecast published by Statistics Norway is somewhat below two-thirds on the long run, and a little above that level on the short run.

The results have been obtained on the basis of stochastic simulation of each of the three components of population change; fertility, mortality, and international migration. Simulation of the components relied heavily on three complementary methods:

- time series analysis for the historical development of key demographic indicators, such as the TFR, the life expectancy, and numbers of immigrants and emigrants;
- an analysis of historical forecast errors, assembled on the basis of forecasts produced by Statistics Norway since 1969;
- and finally expert judgement, which was used, for instance, to restrict the prediction interval for the TFR or that for the numbers of immigrants and emigrants to a reasonable range.

The predictions for each component were calibrated in such a way that the median coincided with the Medium Variant value of the 1999-based official population forecast of Statistics Norway.

The time series predictions indicated that assumptions on future TFR as employed by Statistics Norway in its official population forecasts have estimated coverage probabilities of only 46, 31, and 24 per cent in the years 2010, 2030, and 2050. The official mortality (i.e. life expectancy) assumptions have higher expected accuracy in 2050 (just over 60 per cent), but lower accuracy in the beginning of this century (just over a third in the period 2000-2010).

Sammendrag

Nico Keilman, Dinh Quang Pham, Arve Hetland

Norges usikre demografiske framtid

Sosiale og økonomiske studier 105 • Statistisk sentralbyrå 2001

Framtiden til enhver befolkning er usikker, men noen av de mange mulige utviklingene er mer sannsynlige enn andre. En befolkningsprognose bør derfor inneholde to elementer for hver variabel, for eksempel folkemengden på et framtidig tidspunkt: en spennvidde (intervall) med mulige resultater, og en sannsynlighet knyttet til dette intervallet. Til sammen utgjør disse elementene et prediksjonsintervall for denne variabelen. Denne rapporten dokumenterer resultatene til et forskningsprosjekt hvor formålet har vært å beregne prediksjonsintervaller for Norges framtidige befolkning etter kjønn og alder fram til 2050.

Ifølge våre beregninger er oddsen fire mot en (en sannsynlighet på 80 prosent) for at folkemengden i Norge, som nå er 4,5 millioner, vil ligge et sted mellom 4,3 og 5,4 millioner i 2025, og mellom 3,7 og 6,4 millioner i 2050. Dette illustrerer at usikkerheten øker med tiden. Større treffsikkerhet (høyere odds) går på bekostning av lavere presisjon (større intervaller). Oddsen er 19 mot en (95 prosent sjanse) for en folkemengde mellom 4,1 og 5,7 millioner i 2025, og mellom 3,2 og 7,3 millioner i 2050. Prognosen for barn og eldre viser størst usikkerhet, fordi fruktbarhet og dødelighet er vanskelig å predikere. Følgelig er prediksjonsintervallet i 2030 for befolkningen yngre enn 20 år så bredt at denne prognosen ikke særlig er informativ. Framtidig inn- og utvandring er også svært usikre, men de har en beskjeden effekt på aldersstrukturen. I 2050 har usikkerheten kumulert så sterkt, at intervallene er store for tilsynelatende alle aldre, særlig når man betrakter dem relativt sett, det vil si i forhold til gjennomsnittsprognosen. I 1999 publiserte Statistisk sentralbyrå (SSB) en prognose for blant annet landets befolkning fram til 2050. Ifølge vår modell er den forventede treffsikkerheten for den samlede folkemengden i SSBs prognose i underkant av 67 prosent på lang sikt, og litt høyere enn dette nivået på kort sikt.

Vi har beregnet våre resultater ved å ta utgangspunkt i stokastiske simuleringer for hver av de tre komponentene for befolkningsendring: fruktbarhet, dødelighet og inn- og utvandring. Simuleringene var basert på tre ulike komplementære metoder:

- tidsserieanalyser for den historiske utviklingen til visse demografiske nøkkelvariabler, for eksempel Samlet fruktbarhetstall (SFT), forventet levealder, og det årlige antall inn- eller utvandrere;
- en analyse av feil i SSBs gamle befolkningsprognoser siden 1969;
- subjektive oppfatninger av demografiske eksperter, for eksempel for å begrense prediksjonsintervallet til SFT eller inn- og utvandringsstrømmene til en realistisk verdi.

Prognosen for hver komponent har blitt kalibrert på en slik måte at medianverdien var sammenfallende med SSBs prognosealternativ for middels befolkningsvekst fra 1999-prognosen.

Vår tidsserieprognose antyder at SSBs anslag for den framtidige utvikling i SFT har en forventet treffsikkerhet på bare 46, 31 og 24 prosent henholdsvis i årene 2010, 2030 og 2050. Den offisielle dødelighetsforutsetningen (forventet levealder) har høyere forventet treffsikkerhet i 2050 (litt over 60 prosent), men lavere treffsikkerhet i årene 2000-2010 (litt over en tredel).

Contents

List of Figures

List of Tables

1. The need for stochastic forecasts

It is easier to predict the main results of the next population forecast for a certain country, than to predict the population itself. This is one of the main conclusions from analyses into the accuracy of national population forecasts published by statistical offices of Western countries after the Second World War. In other words: real developments for fertility, mortality and migration are much more volatile than usually assumed in forecasts. When studying the population forecasts of Western countries, one notes how strikingly similar the predicted values in a new forecast are compared to those in the previous forecast. The *actual* variables, however, often show a very different development. While forecasts are surprise-free, reality is not. The rapid fall in fertility in many Western countries in the 1970s came as a surprise for most population forecasters and other demographers. The result was too high birth rates, and an overestimation of the number of young children. Other forecast variables with large errors are the predicted numbers of elderly persons (in particular the oldest old), which were far too low in recent decades because of too pessimistic mortality assumptions, and the size of immigration, which is determined by largely unforeseeable political, economic and legal factors.

The reason that forecasts err is our limited understanding of demographic behaviour. Valid behavioural theories that explain birth, death, or migration to a sufficient degree have not yet been found. Existing theories have limited validity in time or space, or they are strongly conditional or partial, or both (Keyfitz 1982). When it is difficult to explain demographic processes, then it is even more problematic to predict them. The current practice among forecasters is to study regularities and irregularities in the historical developments of major demographic variables such as the Total Fertility Rate and the life expectancy, to understand observed trends, and to extrapolate them into the future.

Extrapolation, as next best to prediction on the basis of causal explanation, implies that population forecasts are inherently uncertain. Any serious forecaster will attempt to include that uncertainty in the forecast in such a way, that it will become clear to the user. The standard approach in national population forecasts is to formulate two or more sets of assumptions for those key variables of which the future development is difficult to predict. Examples are the Total Fertility

Rate and the Life Expectancy at Birth. This approach dates back to at least 1933, when Pascal Whelpton computed a population forecast for the United States, in which he presented several fertility variants. But the use of forecast variants is presumably much older, compare the intervals which Spengler (1935) reports for the results of a number of forecasts for the US. Nowadays, statistical agencies in 15 of the 18 member countries of the European Economic Area (EEA, that is the EU- and EFTA-countries, except Switzerland) produce forecasts with between two and four fertility variants (Eurostat 1997a, Table 35). In the forecasts made in the beginning of the 1990s in those countries, the low and high fertility variants defined intervals of between 0.3 and 0.6 children per woman wide for a forecast duration of approximately 10 years. Moreover, the first co-ordinated population forecast for all 18 EEA-countries has a difference between the high and the low fertility variant of 0.5 children per woman in the year 2035 (Eurostat 1997b, Table 3).

In spite of the general use of forecast variants to express uncertainty, there are two reasons why this approach is unsatisfactory from a statistical point of view. First, uncertainty is not quantified. Take the example of the current official population forecast of Norway. Statistics Norway (1999) assumes in that forecast that the period Total Fertility Rate in the year 2010 will be between 1.5 (low variant) and 2.1 (high variant) children per woman, with margins that are closer to the medium variant of 1.8 in the years prior to 2010. As a consequence, the number of children aged 6-12 in 2010 varies from 401 000 to 436 000, with a medium forecast of 419 000. However, no probability is attached to the interval 401 000-436 000. In other words, the user does not know how probable it is that the real number of children aged 6-12 in 2010 will lie between 401 000 and 436 000, or perhaps fall outside that range. But for an educational planner with an interest in primary schools it must be of great importance to know whether the estimated probability of the number of school children to lie between 401 000 and 436 000 in the year 2010 is 30, or 60, or perhaps even 90 per cent. In the former case, he should incorporate much more flexibility into the school planning process, than in the latter. Knowing the probabilities for distinct intervals is even more pressing in case the planner's loss function is asymmetric. This means that the costs incurred by a forecast that is too low by a certain number of children are different from those caused by a forecast that is too high by the same number. Hiring extra capacity on short notice may cost more, or less, than accepting a certain amount of excess supply of teachers and class-rooms. To minimize his expected losses, the planner will not use the medium forecast, but a number that is higher or lower, depending on his loss function and the probabilities involved.

Second, the use of variants is inconsistent, in the sense that two variants that are extreme for one variable are not extreme for some other variable. For instance, Statistics Norway has assumed high and low trajectories both for future fertility, mortality, and net immigration. The combination of high fertility with high life expectancy (i.e. low mortality), and high immigration defines the so-called high population growth variant. The low population growth variant combines low fertility with low life expectancy and low immigration. Now consider the old age dependency ratio (OADR), defined as the population aged 67 or over divided by the population aged 20-66[1]. In the year 2050, the projected population aged 67+ ranges from 1.076 to 1.244 million according to the low and the high population growth variants. This interval constitutes 31 per cent of the population aged 67+ according to the medium growth variant, in which medium fertility is combined with medium life expectancy and medium immigration. Similarly, the population aged 20-66 ranges from 2.970 to 3.459 million, or 32 per cent of the value according to the medium variant. Yet the OADR-range stretches only from 0.360 to 0.364, or 1.3 per cent of the OADR-value in the medium variant. This uncertainty is much less than one normally would expect. In fact, the interval is considerably larger when one compares two other projection variants, namely weak ageing (high fertility together with low life expectancy and high immigration) and strong ageing (low fertility combined with high life expectancy and low immigration). According to the latter two variants, the interval for the OADR in 2050 is 52 per cent of the OADR-value in the medium-ageing variant. These examples show that a variant pair that is extreme for some variable (e.g. the low and high population growth variant, when the population by age is analysed) need not be extreme for some other variable (e.g. OADR). The reason is that the traditional approach assumes perfect correlation both across components and over time. In the high population growth variant, *each year* when fertility is high, both life expectancy and immigration are high as well. At the same time, fertility (and also life expectancy and migration) is high in each year of the forecast period. From a statistical point of view, such perfect correlations across components and over time is unlikely.

Because the traditional approach is so unsatisfactory, some statistical agencies and individual scholars have in recent years attempted to compute stochastic population forecasts, which result in prediction intervals for future population size and age pyramids that quantify uncertainty around expected or median values. See for instance Hanika et al. (1997) and Lutz and Scherbov (1998a) for Austria, Lutz and Scherbov (1998b) for Germany, Alders and De Beer (1998) and De Beer and Alders (1999) for the Netherlands, and Alho (1998) for Finland. These methods have been inspired by earlier work on stochastic forecasts by for instance Lee and Tuljapurkar (1994) and Alho (1990). The first contribution to the field seems to be the work by Törnqvist (1949) for Finland. The present report relies heavily on those studies, yet we have attempted to improve certain aspects. New is, we believe,

- the presentation of prediction intervals for *cohort* fertility in Section 3.2;
- our analysis of imposing upper and lower limits to the period Total Fertility Rate and the consequences this has for the predictive distribution of the TFR, in Section 3.1;

[1] The legal retirement age is 67 in Norway.

- the emphasis on sampling errors for age-specific birth and death rates in Chapters 3 and 4;
- and the systematic comparison between simulated prediction intervals for the TFR, the life expectancy, and the age structure, and those constructed on the basis of historical forecasts.

2. Approaches to stochastic population forecasting

2.1. General aspects

The cohort-component approach is the standard approach in population forecasting (Keilman and Cruijsen 1992). In short, this approach starts from a table with a known population at some recent point in time, broken down by sex and one-year age groups. This so-called base population is moved forward in time, one year a time, so that all persons move up one age group. Assumed mortality rates by age and sex determine the number of survivors from the original population, one year later. Fertility rates broken down by age of the mother are applied to women in fertile ages, and the result is the number of live births during the year. An assumed sex ratio at birth, and sex-specific infant mortality rates determine surviving numbers of boys and girls in the youngest age class one year later. Finally, emigration rates and immigration numbers, both broken down by sex and age, result in flows of emigrants and immigrants. Emigrants are subtracted, and immigrants are added to the base population, after having been exposed to appropriate risks of mortality (immigrants only), of fertility (both immigrants and emigrants), and of emigration (immigrants only). The result is an updated population by sex and age, one year later and one year older. Repeated application of this process gives a forecast for a period as long as the forecaster wishes to have it.

The process described above requires hundreds of input parameters for each forecast year: 35 fertility rates by age of the mother, 100 mortality rates for each sex, and some 70 migration parameters for male and female immigration and emigration. Thus a forecast 50 years ahead, say, involves thousands of parameters. A *stochastic* forecast based on the cohort component approach requires in principle that one specifies the joint statistical distribution of all those parameters. In practice this is too complex a task, and one breaks up the joint distribution in several other distributions. In this respect, four types of correlation are important: correlation between components, correlation across age, correlation across sexes, and correlation across time (serial correlation).

In a Western country such as Norway, there is little or no reason to assume correlation between the components of fertility, mortality and migration[2]. Nor is there any empirical evidence of such correlation (Lee and Tuljapurkar 1994; Keilman 1997). Therefore, in the stochastic forecasts of the US, Austria, Germany, Finland, and the Netherlands, the three components were considered independent of each other (Lee and Tuljapurkar 1994; Lutz and Scherbov 1998a, 1998b; Alho 1998; De Beer and Alders 1999).

Correlation across time is important for each component. 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. This implies a strong autocorrelation for these two components. International migration is much more volatile, but economic, legal, political, and social conditions stretching over several years steer migration flows to a certain extent, and some degree of autocorrelation should be expected.

Men and women display similar behaviour regarding mortality and migration. This gives rise to a certain degree of correlation across sexes for these two components.

Correlation across age is strong for each component. The age profiles of fertility (by mother's age), mortality and migration are highly regular. Age-specific fertility has one top around 25-30 years. Age-specific mortality falls from rather high levels of infant and child mortality to low levels for teenagers, and rises subsequently to a maximum for the oldest old. The age pattern of migration is bi-modal, with tops for young adults and children of pre-school age, and much lower levels for intermediate and old ages.

A common approach in stochastic population forecasting is to assume some statistical distribution for a selected key parameter for each component in the future, both its level and development over time. Examples are the Total Fertility Rate (TFR) for fertility, the Life Expectancy at Birth for male and female mortality, and the annual number of net-immigrations. Three main methods to arrive at such a distribution may be distinguished: time-series methods, expert judgement, and extrapolation of observed errors in historical forecasts. 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 (or around the expected error), and the expected value is taken from a deterministic forecast computed in the traditional manner.

---

[2] 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. Western countries with extremely high immigration from developing countries may have a positive correlation between their levels of immigration and childbearing. In Norway this effect is negligible.

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 ARIMA (Autoregressive Integrated Moving Average)-models (Box and Jenkins 1970). Since ARIMA models consider the time series as stochastic, both point predictions and interval predictions are possible. Time series models were developed for short horizons. When applied for long-run population forecasting, the point forecast may become unreasonable, and/or the prediction intervals may become excessively wide. 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. 1986; Hanika et al. 1997; Lutz and Scherbov 1998a, 1998b). A group of experts is asked to indicate the probability that a key 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. the prediction intervals they give tend to be too narrow (Armstrong 1985).

Extrapolation of empirical errors requires observed errors from historical forecasts. Next, 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 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). An important problem is that time series of historical errors are usually rather short, as forecasts prepared in the 1950s or earlier generally have been documented not well enough.

The three approaches are complementary, and elements of all three are often used in actual practice. For instance, time series methods are inherently judgmental, in particular when a certain form of the extrapolation model or when the length of the historical series is chosen. The subjective decisions taken in this respect may have important consequences for the shape of the prediction intervals. Furthermore, prediction intervals, either obtained by time series methods or by expert opinion, are frequently checked against historical error patterns (e.g. Lee and Tuljapurkar 1994; Alho 1998).

Irrespective of the method that was 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. There are two common approaches to obtain such intervals: an analytical approach and a simulation approach.

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 that have been necessary hitherto in the analytical approach. The idea is simply 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. The results are stored in a database, and the predictive distribution for a certain variable follows immediately from the histogram of that variable. Early contributions based on the idea of simulation are those by Keyfitz (1985), Pflaumer (1986, 1988), and Kuijsten (1988).

2.2. The approach in this report

In the current project, the prediction intervals for age-specific fertility, mortality, immigration, and emigration have been computed using a common approach. A short general description is given below, whereas details are contained in chapters 3-5. Prediction intervals for the size of the population and its age-sex distribution have been determined by means of stochastic simulation.

The four components were assumed independent.[3] The approach relies heavily on time-series models for one or more key parameters for each component: Stochastic simulation was used to obtain future values of the key parameters and, for fertility and mortality, stochastic age patterns.[4] Each set of key parameters was assumed to follow a multivariate normal distribution, with a known vector of expected values and known covariance matrix. Multivariate normally distributed numbers were drawn from each distribution using Cholesky decomposition of the covariance matrices (Bratley et al. 1983). Repeated

[3] The correlation across the sexes for mortality (Chapter 4) and migration (Chapter 5) is accounted for. A possible correlation between immigration and emigration (for instance caused by return migration), or one between migration of men and women (family migration, family reunification) is ignored. Extensive empirical tests did not result in clearly interpretable patterns, see Chapter 5.
[4] The age patterns for immigration and emigration were assumed deterministic, see Chapter 5.

stochastic simulation resulted in 5 000 sample paths for each age-specific parameter.

For *fertility,* we assumed that a Poisson process generates births in each forecast year. The intensity of that process depends strongly on the age of the mother. Thus the parameter of the Poisson process, that is, the fertility rate, varies by age. For a given year or a given birth cohort of women, we assumed that the age pattern of fertility follows a Gamma curve. This is a mathematical function, which consists of a Gamma density and a scaling parameter. The Gamma curve has four parameters: the Total Fertility Rate (TFR), the Mean Age at Child-bearing (MAC), the Variance in that age (VAR), and the minimum age. The four parameters were estimated on the basis of data on Norwegian births in one-year age groups for each year in the period from 1900 to 1995. This resulted in a time series of parameter estimates. The series for three of the four parameters (TFR, MAC, and VAR) were modelled by means of a multivariate time series model of the ARIMA-type. The minimum age was kept constant at its value as estimated for recent years. Predictions were made by stochastic simulation for the period 1996-2050 for the remaining three parameters. Extreme values, for instance TFR-values lower than 0.5, or higher than four children per woman, were rejected, and the simulations were repeated until 5 000 time paths with admissible values were obtained. The Gamma curve was used to transform the parameter predictions back into future age-specific fertility rates.

For *mortality,* a first order homogeneous Markov process with constant intensity was assumed for each age and both sexes. The intensities were estimated by means of the corresponding age and sex-specific death rates. The predictions consisted of three steps. First, a life table calculation resulted in annual values of the life expectancy at birth for men and women. The two time series of life expectancies for the period 1945-1995 were modelled by means of a multivariate ARIMA model, and stochastic simulation resulted in prediction intervals for male and female life expectancies for the period 1996-2050. The constant term in the ARIMA model was adjusted in such a way that the expected life expectancy values in 2050 coincided with target values assumed by Statistics Norway in its official population forecast. Second, we assumed that for each year the age pattern of mortality could be described by means of a Heligman-Pollard (H-P) curve. This is a mathematical function with eight parameters. The parameters were estimated on the basis of data on Norwegian deaths for men and women in one-year age groups for each year in the period from 1900 to 1995. This resulted in two time series of parameter estimates, one for each sex. The time series for the years 1945-1995 were modelled by means of multivariate time series models of the ARIMA-type. Stochastic simulation resulted in 5 000 multivariate sample paths for the H-P parameters for the period 1996-2050. The H-P curve was used to transform the parameter predictions back into future age-specific mortality rates. For each year in the prediction period, a life table calculation summarized those rates into one indicator, viz. the life expectancy at birth. These life expectancies were assembled in a look-up table, together with the underlying

H-P parameter values. In the third and final step, H-P parameter values from the look-up table were assigned to life expectancy values for each sample path, predicted in the first step, by matching life expectancy values from the first and the second step, controlling for calendar year and sex. The result was an age pattern for male and female mortality for each year in the future and each simulation run.

For *international migration*, we distinguished between immigration and emigration flows. Annual numbers for each flow, observed for the period 1967-1997 were modelled by means of univariate ARIMA time series models. Stochastic simulation resulted in 5 000 sample paths with annual numbers of immigrants and emigrants for the period 1996-2050. The constant terms in the ARIMA models were adjusted in such a way, that target levels of annual immigration and emigration assumed by Statistics Norway were predicted. Predicted numbers of immigrants and emigrants were broken down by sex on the basis of randomly drawn shares for men and women. This resulted in four flows: immigration and emigration by sex. Age-specific numbers for each of the four flows were obtained on the basis of a Rogers-Castro (R-C) curve with six parameters[5]. The R-C curve was fitted to age-specific shares for each of the four flows in each calendar year. The resulting time series of R-C parameters were predicted into the future by means of simple extrapolation procedures. Predicted R-C parameters were used to transform predicted flows back into predicted numbers of immigrants and emigrants by sex and age. The age pattern for each flow was the same across sample paths.

Since the ultimate purpose was to generate stochastic population forecasts, much attention was given to an appropriate quantification of uncertainty. In the present approach, there are four main sources of uncertainty attached to future birth and death rates, and migration numbers:

1. sample variation in the historical age-specific rates;
2. errors in the parameter estimates of the age pattern curves (Gamma curve for fertility, Heligman-Pollard curve for mortality, Rogers-Castro curve for migration);
3. residual variance in each time series model;
4. estimation errors in the parameter estimates of each time series model.

Earlier studies of a similar nature, such as Bell (1992, 1997), De Beer (1989, 1992), Duchêne and Gillet-De Stefano (1974), Knudsen et al. (1993), and Thompson et al (1989) have ignored one or more of these sources. Our strategy has been, whenever possible, to take due account of all four. This gave reasonable prediction intervals for mortality, but for fertility the uncertainty was

[5] The retirement peak in observed immigration and emigration turned out to be negligible in the Norwegian data, see Chapter 5.

clearly too large, and prediction intervals were too wide. Therefore uncertainty connected to future fertility was reduced in an *ad-hoc* manner by subjectively decreasing prediction intervals for some key parameters.

For migration, observed patterns were so volatile, that the time series extrapolations for absolute numbers of immigration and emigration resulted in excessively wide prediction intervals after a few years already. Our strategy has been to fix the width of the intervals after an initial period of five years.

Time series methods as described in Section 2.1 have been used to generate prediction intervals for fertility, mortality and migration parameters, see Chapters 3-5. For the base population it was assumed that this was perfectly known. In countries with defective data this is obviously not a good assumption, and the base population should be considered stochastic as well.

In summary, our stochastic population forecasts are based on a combination of three different methods: time series extrapolation, inspection of observed errors in historical forecasts, and the use of expert judgement. Most attention was given to time series extrapolation, in order to obtain a correct initial specification of (co-)variances and autocorrelations. For mortality, this resulted in acceptable prediction intervals. But for fertility and international migration, time series extrapolation gave too wide intervals around key parameters, which were reduced in an *ad hoc* manner on the basis of subjective decisions. Observed errors for fertility, mortality and the age structure were used to check the plausibility of the resulting intervals.

Although subjective decisions regarding the width of prediction intervals, the choice of time series model, and the selection of the estimation period certainly will have influenced our final results, expert judgement played a less important role than in the stochastic forecasts mentioned in Section 2.1. By emphasizing time series extrapolation, we have tried to avoid too narrow intervals as a consequence of being too confident in one’s ability to predict demographic developments accurately.

3. Fertility

The methodology used for fertility extrapolations has been reported extensively in papers by Keilman and Pham (2000) and Keilman and Hetland (1999). Section 3.1 will summarize the main features.

3.1. Period fertility

3.1.1. Historical patterns

Assume that a group of $Y(x)$ women aged x give birth to $B(x)$ children in a certain year. Assume further that a Poisson process generates the births. Maximum Likelihood estimators for the intensity of the process, and its corresponding variance are

(3.1) $F_x = B(x)/Y(x)$, and

$$Var(F_x) = B(x)/Y^2(x) = F_x/Y(x)$$

The estimator F_x is the traditional age-specific birth rate.

We have used the Norwegian age-specific birth rates for the years 1900-1993 computed by Brunborg and Mamelund (1994). These were supplemented with rates for the years 1994 and 1995. The age range was from 16 to 44. When computing person years of exposure, we ignored mortality and international migration, and used the population of women at the beginning of each year, broken down in one-year age groups.

The array of 29 rates for each year can be summarized by means of a parametric curve, which is a function of age. Thanks to the regular shape of the age pattern of fertility, such a function contains only a few parameters, usually three to five. Various curves fit the data well: normal, lognormal, double exponential, Coale-Trussell, Hadwiger, polynomial, gamma, and logistic curves, to name the most important ones. Several authors have noted the attractive properties of the Gamma curve. The fit is usually good, and the parameters can be interpreted, after an appropriate transformation, in a straightforward way. See Bell (1997), Hoem et al. (1981), and Duchêne and Gillet-De Stefano (1974). Denote the fertility intensity for age x as f_x. The Gamma curve is defined as

(3.2) $$f_x = \frac{1}{\Gamma(\alpha_3)}\alpha_1\alpha_2^{\alpha_3}(x-\alpha_4)^{\alpha_3-1}\exp\left[-\alpha_2(x-\alpha_4)\right], \quad x \geq \alpha_4 .$$

The four parameters α_i are to be estimated from the data. $\Gamma(.)$ is the Gamma function defined by

$$\Gamma(p) = \int_0^\infty u^{p-1}\exp(-u)du.$$

α_1 represents the TFR, whereas α_4 is the minimum age at childbearing. The parameters α_2 and α_3 have no immediate demographic interpretation. However, f_x/α_1 is the Gamma density, with mean $\alpha_4+(\alpha_3/\alpha_2)$ and variance α_3/α_2^2. Therefore it is customary to introduce the following transformation

(3.3) $$\begin{aligned} \beta_1 &= \alpha_1 \\ \beta_2 &= \alpha_4 + \alpha_3/\alpha_2 \\ \beta_3 &= \alpha_3/\alpha_2^2 \\ \beta_4 &= \alpha_4. \end{aligned}$$

Hence parameters β_1 and β_4 have the same interpretation as α_1 (TFR) and α_4 (minimum age). β_2 represents the mean age at childbearing, while β_3 is the variance in that age. The β-parameters have been estimated by means of non-linear regression, by minimizing the following weighted sum of squares

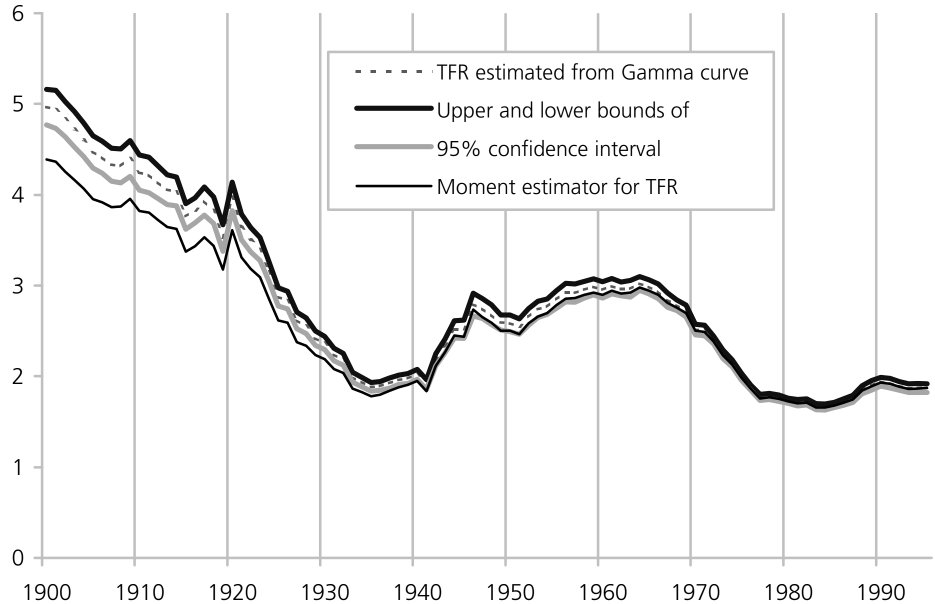
(3.4) $$\Sigma_x w_x(F_x - f_x)^2.$$

w_x is the inverse value of the variance of F_x, see expression (3.1), reflecting the "measurement error" for the rate F_x: a small variance indicates a precise estimate for the intensity, and vice versa.[6] Hence ages for which the variance is large get less weight in the regression than those with smaller variances. Weighted least squares estimation is approximately equivalent with Maximum Likelihood estimation of the parameters β_i, with births $B(x)$ and exposure time $Y(x)$ as data (Van Imhoff 1991).

[6] Since in the end we are concerned with errors in the projected numbers of births, we consider the *absolute* size of the measurement error, not the *relative* size, which would amount to $Var(F_x)/F_x$. Thus for some ages the measurement error is small simply because the intensity is low.

The four β-parameters and the corresponding covariance matrix, together with their variances were estimated on the basis of Norwegian birth rates for each of the years 1900-1995. Figures 3.1-3.3 present estimates for β_1, β_2, and β_3 with corresponding 95 per cent confidence intervals. The TFR, the mean age at childbearing, and the variance computed in the traditional demographic manner (i.e. moment estimators TFR=$\sum_x F_x$, $m=\sum_x x.F_x$/TFR and $s^2=\sum_x (F_x-m)^2/\text{TFR}^2$) are also given. The latter estimators are only influenced by Poisson variability, not by the fit of the Gamma curve. For the minimum age β_4 we found that estimates in recent years were invariably equal to the boundary value of zero (after initial values around 14-15 years of age in the first half of the century).[7]

Figure 3.1. Total Fertility Rate estimates and 95 per cent confidence interval



[7] Estimates for the minimum age β_4 fell below 14 in 1975 and decreased further to reach zero in 1991. During the same period, the estimates for the mean age at childbearing β_2 rose from 26.6 to 28.3 years. Together with the relatively low estimates for the Total Fertility Rate β_1 during these years (< 2), the predicted birth rates at ages below 16 are still negligible, in spite of the unrealistic estimate for the minimum age.

Figure 3.2. Estimates for mean age at childbearing and 95 per cent confidence interval

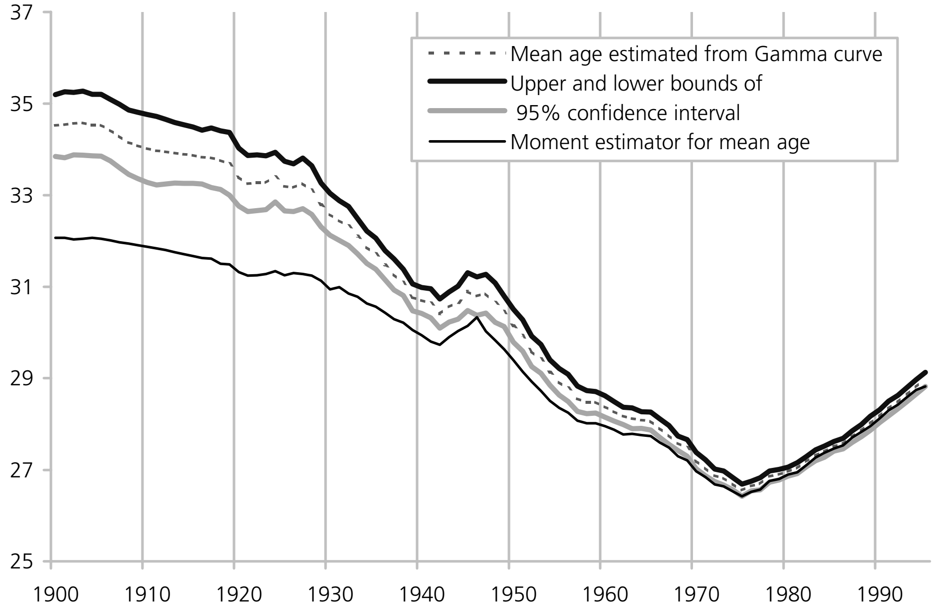
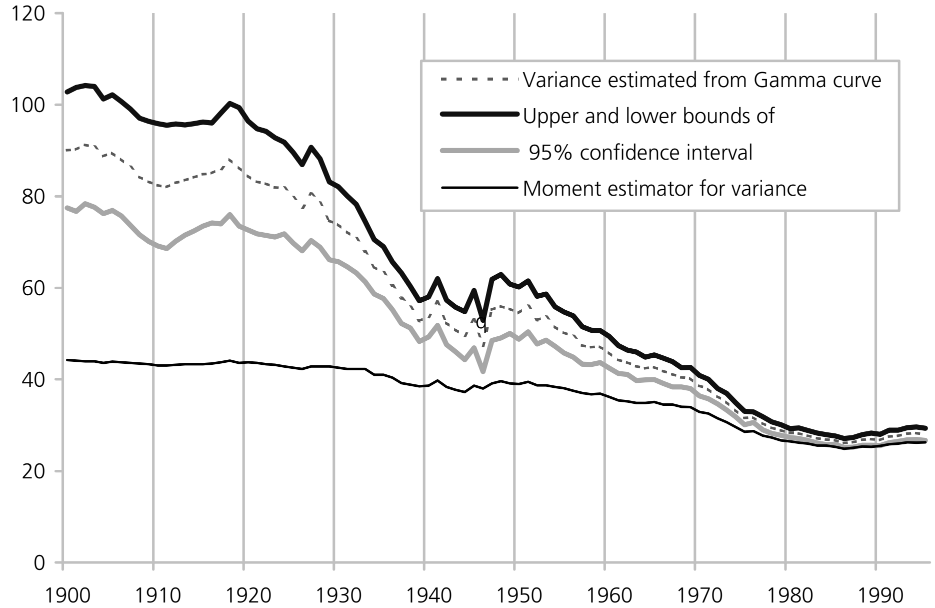


Figure 3.3. Estimates for variance in childbearing age and 95 per cent confidence interval



Similar to many other Western countries in this century, Norway had two periods with a strong fertility decrease (Figure 3.1). The first one, which started around 1880, ended in the 1930s, whereas the second one took place at the end of the 1960s and during the 1970s. The baby boom of the 1950s and 1960s was to a large extent the consequence of a decrease in the mean age at childbearing for women born in the years 1920-1945. This led not only to a fall in the period mean age (see Figure 3.2), but also to rather high period-TFR values (Figure 3.1). The period-TFR attained its minimum in the years 1983 and 1984, when it was as low as 1.66 children per woman. After a rise towards 1.9 children per woman in 1990, the TFR has been rather constant. But in recent decades, women get their children at increasingly higher ages, compare the strong rise in the period mean age at childbearing in Figure 3.2. Facilitated by modern contraceptive methods, growing shares of young Norwegian adults postponed the birth of their first child and took some form of education at the tertiary level during the 1970s and 1980s. Next they worked some years before they entered parenthood (Kravdal 1994). Much of the fertility decrease during this century was caused by a reduction of higher-parity births, which generally take place at high ages.

The 95 per cent confidence intervals for the β_i are rather wide in the years 1900-1925 and 1945-1965, indicating a relatively bad fit. After 1980, the fit is excellent. The traditional TFR (moment estimate) coincides with the estimates from the Gamma curve from 1970 onwards, and it falls within the 95 per cent confidence bounds from 1940. It is much lower in the first half of the century. However, one does not know whether the traditional TFR or our Gamma-based TFR are good approximations to the real (but unknown) Total Fertility, since both are only an *estimate* of the latter. Figure 3.4 illustrates how the fit improves over the years.

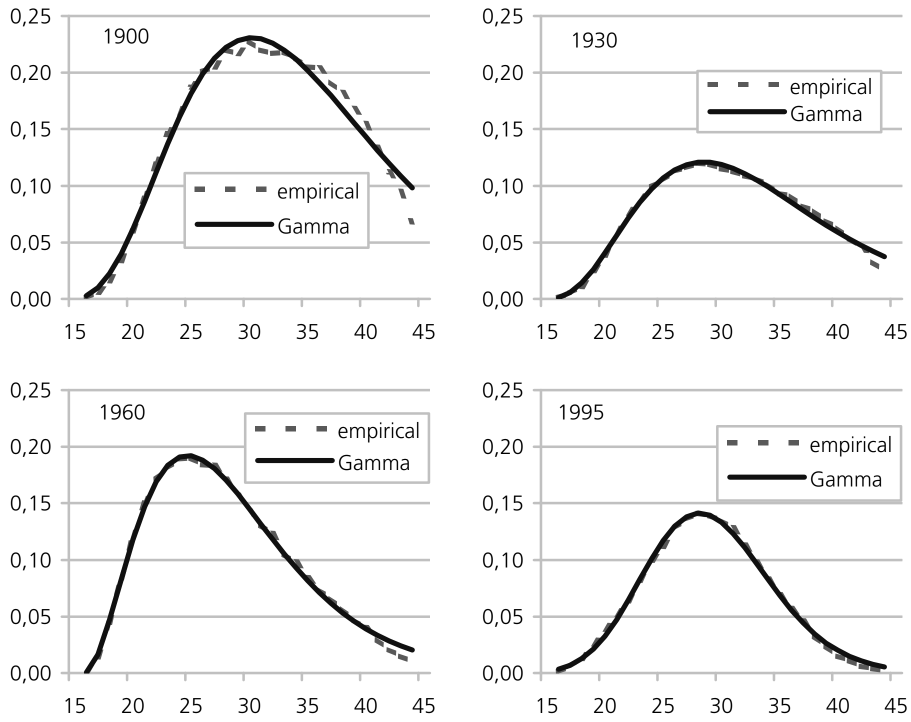
The result of the curve fitting exercise described above is a series of estimates for the four parameters of the Gamma curve for each year between 1900 and 1995, and the corresponding estimated covariance matrix for each year. A multivariate ARIMA model has been used to predict three of the four parameters: β_1, β_2, and β_3. As noted earlier, the minimum age of childbearing β_4 fell from 14 in 1975 to zero in 1991, and remained at that level since. We predict that β_4 will equal zero in the future, too.[8]

Let $C_t = (\ln(\beta_{1,t}), \ln(\beta_{2,t}), \ln(\beta_{3,t}))`$ be a column vector with the first three Gamma curve parameters in year t in logarithmic form. First differences of C led to stationarity, and we found that a multivariate ARIMA (1,1,0) model fitted the data well. The model is of the form

$$(3.5) \qquad Z_t = \phi Z_{t-1} + \varepsilon_t,$$

[8] Note that Thompson et al. (1989) find a similar (although steeper) drop in the minimum age.

Figure 3.4. Age-specific birth rates, empirical values and Gamma curve fit



where $Z_t=C_t-C_{t-1}$, ϕ is a fixed 3x3-matrix of coefficients, and $\varepsilon_t=(\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})$` is a multivariate normal column vector with mean 0 and constant covariance matrix Σ_ε. We deliberately omitted an intercept from the model, so that we avoid predicting an indefinitely increasing or decreasing pattern in β_t.[9] Estimation was done in such a way that years in which variances and covariances for the β-estimates were large, got less weight.

We have limited the estimations to the years 1945-1995. On the one hand, a long series is desirable on statistical grounds. On the other hand, there is little reason to believe that the childbearing behaviour of women in the first half of the century was so similar to that in more recent decades, that both can be captured by one model. Moreover, the fit of the Gamma curve was much better in the second half of this century than in the first half. We have also investigated the sensitivity of our predictions for choosing the shorter periods 1960-1995 and 1975-1995 (details not given here). The conclusion was that opting for the

[9] Preliminary tests on the basis of a univariate model for the Total Fertility Rate showed that this was an appropriate choice.

period 1945-1995 strikes a good balance between a high residual variance (1960-1995), and an imprecise estimate for the autoregressive coefficient (1975-1995).

Estimates for the elements of ϕ and corresponding standard errors are listed in Table 3.1. Diagonal elements are high and strongly significant. All but one ($\hat{\phi}_{31}$) of the off-diagonal elements turned out to be non-significant at the 5 per cent level in a trial calculation. These were set equal to zero, and the model was re-estimated with those restrictions, using Restricted Least Squares (Lütkepohl 1993).

The estimated covariances for the non-zero $\hat{\phi}_{ij}$ -elements are given in Table 3.2, whereas Table 3.3 contains estimates for the residual covariances Σ_{ε}. Much of the uncertainty, relatively speaking, concerns the TFR, as witnessed by the high estimate of $\sigma_{\varepsilon,11}$. This is explained by the large fluctuations in the TFR since 1945 (see Figure 3.1), whereas those in the mean age (Figure 3.2) or in the variance (Figure 3.3) were much smaller.

Table 3.1. Estimates of ϕ_{ij}

	$\hat{\phi}_{11}$	$\hat{\phi}_{12}$	$\hat{\phi}_{13}$	$\hat{\phi}_{21}$	$\hat{\phi}_{22}$	$\hat{\phi}_{23}$	$\hat{\phi}_{31}$	$\hat{\phi}_{32}$	$\hat{\phi}_{33}$
Estimate	0.6694	0	0	0	0.8852	0	0.0909	0	0.3089
Standard Error	0.1044	-	-	-	0.0735	-	0.0419	-	0.1337

Table 3.2. Covariance estimates for non-zero elements of $\hat{\phi}$ [1]

	$\hat{\phi}_{11}$	$\hat{\phi}_{22}$	$\hat{\phi}_{31}$	$\hat{\phi}_{33}$
		x10^{-3}		
$\hat{\phi}_{11}$	10.185	0.011	-0.354	-0.000
$\hat{\phi}_{22}$	0.011	4.547	0.023	-0.130
$\hat{\phi}_{31}$	-0.354	0.023	1.644	-2.484
$\hat{\phi}_{33}$	-0.000	-0.130	-2.484	16.588

[1] Covariances between other elements of $\hat{\phi}$ are zero.

Table 3.3. Estimates of $\Sigma_{\varepsilon}=(\sigma_{\varepsilon,ij})$[1]

$\hat{\sigma}_{\varepsilon,11}$	$\hat{\sigma}_{\varepsilon,12}$	$\hat{\sigma}_{\varepsilon,13}$	$\hat{\sigma}_{\varepsilon,22}$	$\hat{\sigma}_{\varepsilon,23}$	$\hat{\sigma}_{\varepsilon,33}$
		x10^{-3}			
0.703	0.005	0.105	0.007	0.015	0.309

[1] Only the upper triangular part of the symmetric matrix is given.

3.1.2. Future fertility

We simulated 5 000 sample paths for the vector β, each one from 1996 until 2050. For every sample path we drew one value for the matrix ϕ, and 55 values for the vector ε, one for each year. The estimates of both ϕ and ε follow a multivariate normal distribution. The mean and the covariance of $\hat{\phi}$ are given in Tables 3.1 and 3.2. The mean of $\hat{\varepsilon}$ is the null-vector, while its covariance estimates are contained in Table 3.3. Multivariate normally distributed numbers were drawn from these two distributions using Cholesky decomposition of the covariance matrices (Bratley et al. 1983).

As could be expected, the long-run simulations show excessively wide prediction intervals, see Figures 3.5-3.7. The average values for the three parameters in 2050 are 2.21 (TFR), 30.9 (mean age), and 28.3 (variance), while the medians are 1.86 children per woman, 30.2 years, and 27.6 years^2, respectively. The odds are two against one that the TFR in 2050 will lie between 1.1 and 3.3 children per woman, while the 95 per cent prediction interval is (0.5, 6.1) in that year.

Figure 3.5. Total Fertility Rate

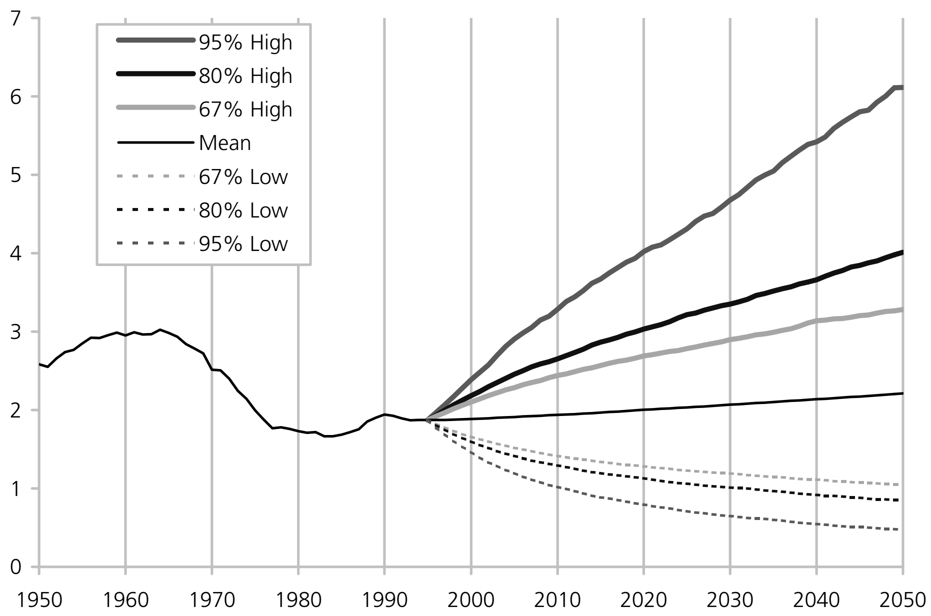


Figure 3.6. Mean age at childbearing

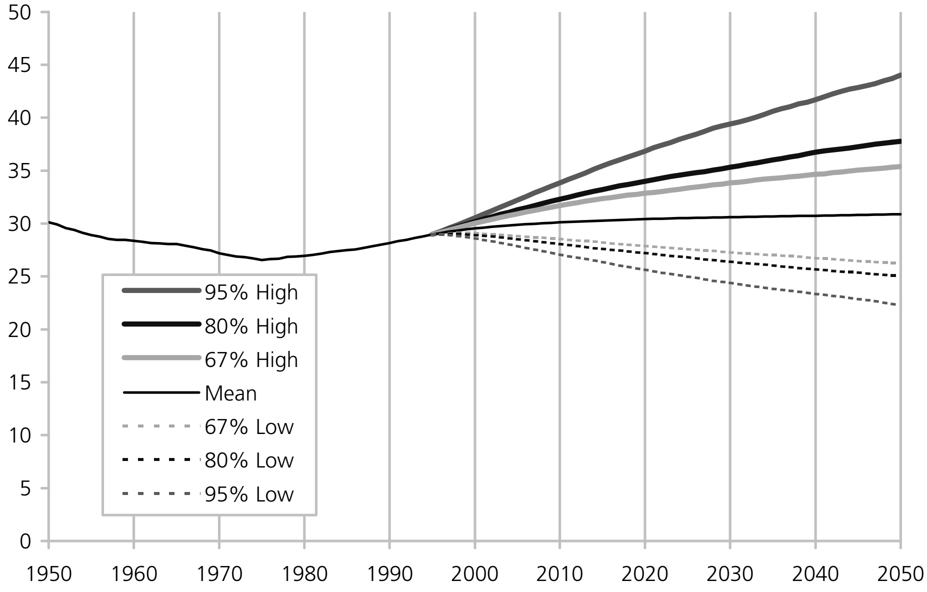
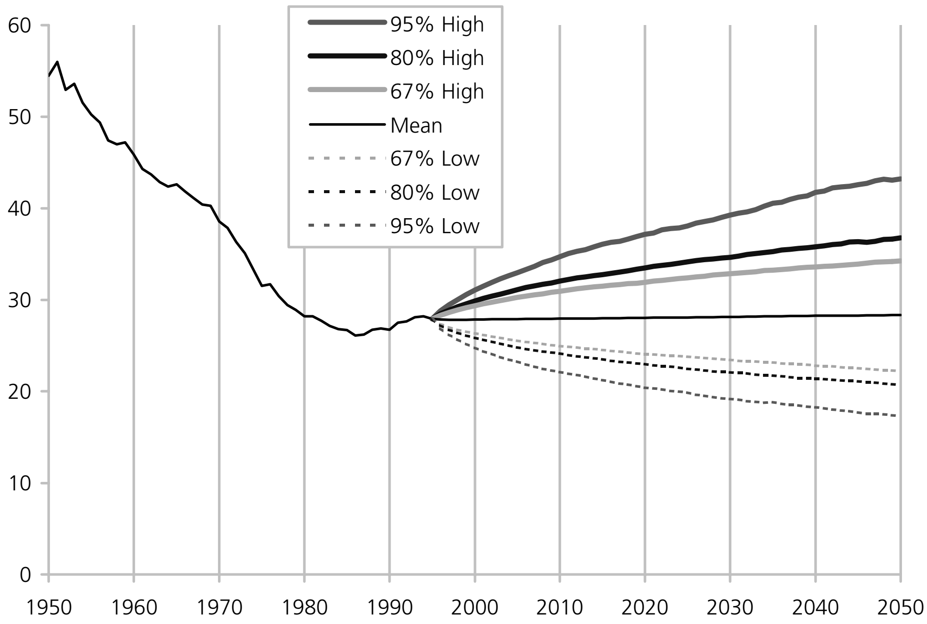


Figure 3.7. Variance in childbearing age



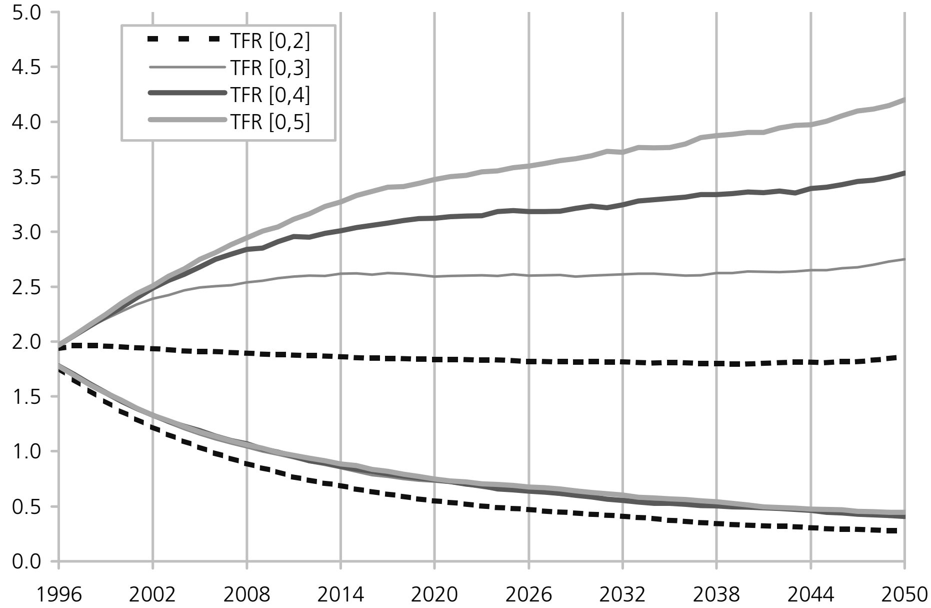
The expected probability that the TFR will exceed 6.1 children per woman in 2050 is only 2.5 per cent. Yet it is difficult to imagine such extremely high levels of childbearing in a country like Norway. The highest TFR ever recorded for Norway is (the moment estimate of) 4.8 in 1879, the highest value since 1845, see Brunborg and Mamelund (1994). With a TFR of 6.1 children per woman or more, fertility would exceed the historical maximum by more than one child, and it would exceed the current level in many less developed countries. Clearly, the model predictions for the middle of the next century cannot be considered as realistic. Now assume (somewhat subjectively) that a fertility level of more than four children per woman on the medium and long term should be rejected (even when the probability of such a level is only a few per cent). Then we see that the model gives reasonable results up to around 2020 or perhaps 2030. Beyond that, prediction intervals are too wide. One has to take recourse to other methods when predictions so far ahead are required. The easiest one is to assume that in 2030, say, uncertainty is already so large that it will not increase any more (e.g. Alho and Spencer 1997). In that case prediction intervals are constant after 2030. A more sophisticated one is to assume that there is an upper bound to fertility levels in Norway. This will be analysed below.

3.1.3. Rejecting extreme values

The reason why the ARIMA model on the long run produces wide prediction intervals for the period TFR is that it contains no extraneous information beyond the past data. Several factors are associated with the development of fertility after World War II in Norway, similar to other countries. Important elements in this respect are the introduction of modern contraception, the adoption of new norms and values with respect to childbearing and partnership, increased interest in tertiary education, and growing levels of labour force participation among women (Kravdal 1994; Lesthaeghe and Surkyn 1988). None of these factors is, or could be, explicitly modelled. It is reasonable to expect that they will be in operation in the future, too. Yet we have no idea to what extent these or other factors will constrain Norwegian fertility in the next century, be it period or cohort developments. Therefore, we used the ARIMA time series model to carry out a number of experimental simulations in which upper and lower bounds were imposed on the period TFR.

The logarithmic transformation of the TFR ensures that the predicted TFR never falls below zero. However, the predicted TFR virtually has no upper limit. Therefore, upper limits were defined ranging from 2 to 5, lower limits ranging from 0 to 1, and simulations were carried out for a certain combination of upper and lower TFR limit. Simulations that resulted, for any future year, in a predicted TFR outside the range defined by the upper and lower limits were rejected, and new simulations were generated until we had obtained 5 000 simulations with admissible values.

Figure 3.8. Period TFR, 95 per cent prediction interval, minimum TFR=0

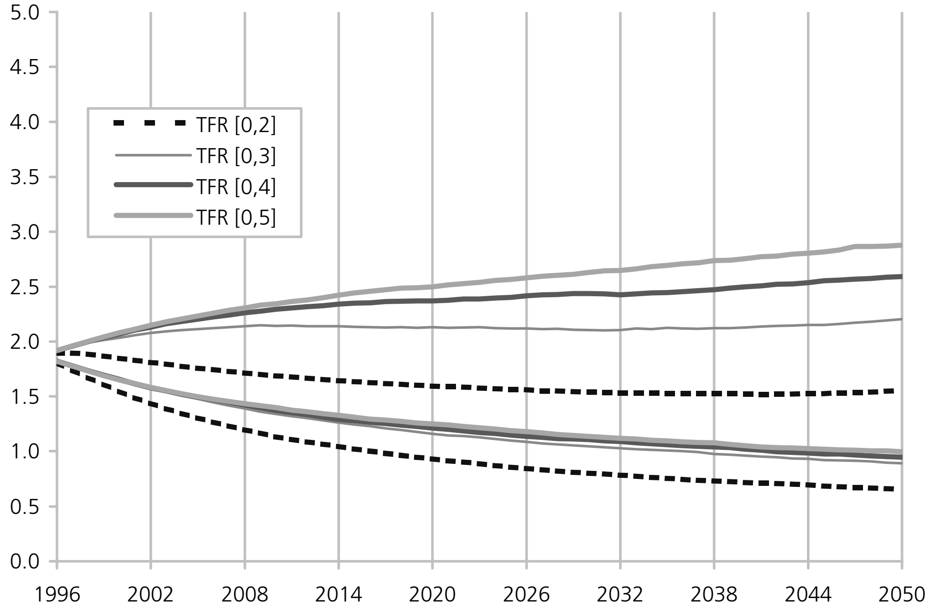


We experimented with upper limits for the period TFR of 2, 3, 4, and 5 children per woman, and lower limits equal to 0, 0.25, 0.5, 0.75, and 1 child per woman on average. Selected results are given in Figures 3.8 and 3.9.

Figures 3.8 and 3.9 show that imposing an upper TFR limit equal to 2 children per woman is unrealistic. It implies that the upper bound of the 95 per cent prediction interval is constant over the years, whereas the upper 67 per cent bound is pressed progressively downwards. Hence also the mean and the median TFR will fall over the years. With an upper limit equal to 3 children, the upper 67 per cent bound is almost constant, although it tends to decrease slightly between 2009 (2.15 children per woman) and 2031 (2.10). A more realistic pattern, with increasing upper bounds, is shown when the upper TFR limit is set equal to 4 or higher.

In the short run, the width of the 67 per cent prediction interval varies little when the upper limit becomes less strict. In the year 2010, the width is 0.80 children per woman when the TFR is limited to three children at the most, and 0.92 and 0.95 children per woman for maximum TFR values of 4 and 5, respectively. In the long run however, uncertainty (expressed in terms of the width of the 67 per cent prediction interval) grows faster when the maximum TFR is increased: from 1.31 children per woman for a maximum TFR of 3, to 1.64 and 1.88 children per woman for maximum TFRs of 4 and 5, respectively. The width of the 95 per cent prediction interval is much more influenced, also on the short run. In 2010 it amounts to 1.60, 1.92, and 2.06 children per woman for maximum TFR values of 3, 4, and 5.

Figure 3.9. Period TFR, 67 per cent prediction interval, minimum TFR=0

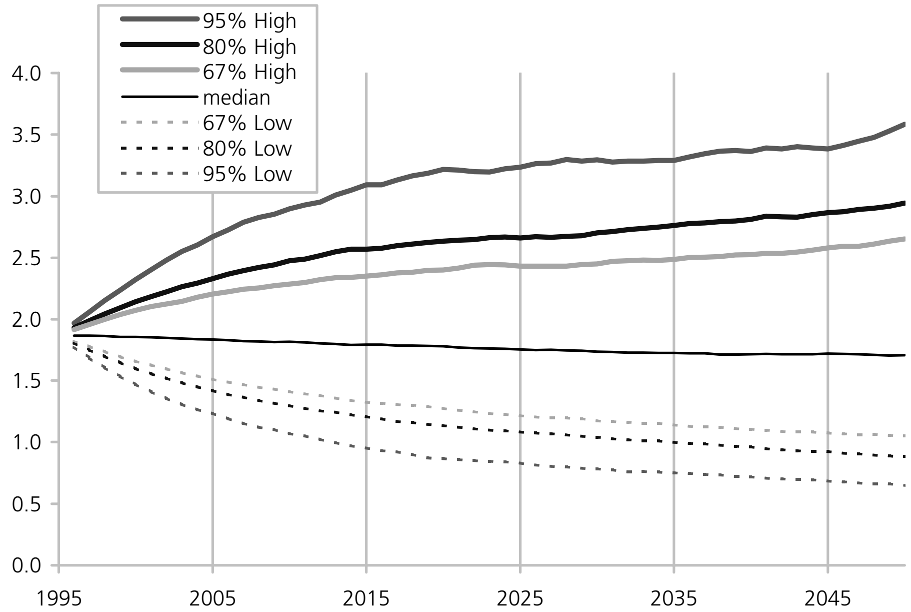


Thus we see that the 95 per cent prediction interval is much stronger influenced than the 67 per cent interval. The reason is that the upper and lower 2.5 per cent of the simulated values define the bounds of the 95 per cent interval. These bounds are rather sensitive to outliers, which is not the case for the 67 per cent bounds.

When the lower limit is successively increased from zero to one child per woman, both the upper and the lower bounds of the intervals are pressed upwards, the lower bounds somewhat more than the upper ones (figures not shown here). In other words, the intervals are slightly narrower than with a lower TFR limit set to zero, as expected. For instance, with a minimum TFR equal to one, the 67 per cent interval in the year 2010 is 0.72, 0.82, and 0.90 children per woman wide when the maximum TFR is set to 3, 4, or 5. In 2050 the corresponding ranges are 1.02, 1.42, and 1.72. In spite of these differences, the shapes of the prediction intervals remain the same as those in Figures 3.8 and 3.9.

On the basis of the experiments reported in this section, we decided to restrict the period TFR to values between 0.5 and 4.0 children per woman. At the same time, restrictions on other parameters were introduced. By the middle of the next century, the childbearing behaviour of Norwegian women may be very different from todays. Medical technology may have made it possible to postpone childbearing to ages well beyond 50. But even then, a mean age at childbearing of 50 years, or a variance in the age at childbearing of 400, is clearly unrealistic.

Figure 3.10. Prediction intervals for TFR, TFR restricted to (0.5, 4.0)



At the same time it is unrealistic to assume that teenage fertility has become so important that the mean age falls below 20. Thus we used the following additional restrictions: 20<MAC<40, and 0<VAR<250. Finally, a restriction was imposed on the elements of the matrix ϕ: each of those was required to lie between minus one and one, in order to ensure stationarity. Figure 3.10 illustrates the resulting prediction intervals for the TFR based on 5 000 simulations with admissible values.

For each simulation and each calendar year in the period 1996-2050, the Gamma curve transformed the four summary indicators into age-specific fertility rates. We have deliberately chosen not to censor rates beyond age 50. Recent medical advances have led to an increased demand for Assisted Reproductive Technology (such as *in vitro* fertilization) after age 30 in Western countries. We cannot exclude the possibility that in the middle of the next century childbearing will be an option for women older than 50. It has been suggested that frozen *ova* may be taken from the woman at age 22, say, and that these may be fertilized and implanted later (Beets 1996). The predicted rates, however, are small, and for practical reasons we used the age of 65 as the highest fertile age. At the other end of the age scale, there are some very low rates at ages below 15. These have been ignored in the cohort-component projections of Chapter 6.

Table 3.4. Fertility assumptions in Statistics Norway's 1996-based population forecast

	Total Fertility Rate (children per woman)			Mean age at childbearing (yrs)		
	Low variant	Medium variant	High variant	Low variant	Medium variant	High variant
1995[1]	1.86	1.86	1.86	28.79	28.79	28.79
2000	1.79	1.86	1.95	29.34	29.20	29.03
2010 and beyond	1.68	1.86	2.10	30.50	30.00	29.50

[1] Values computed from birth statistics.

Source: Statistics Norway (1997).

The High-Low range for the TFR in Statistics Norway's 1996-based population forecast is 0.42 children per woman in 2010, see Table 3.4. A comparison with Figure 3.10 tells us that the estimated coverage probability for that range is only 32 per cent. In later years, the coverage probability falls to levels of 20 per cent in 2030 and 15 per cent in 2050. The low coverage probability was one of the reasons why Statistics Norway increased the High-Low range for the TFR in its 1999-based forecast to [1.5, 2.1] for the years 2010 and later. This improved the estimated coverage probabilities to 46, 31, and 24 per cent in the years 2010, 2030, and 2050.

3.1.4. Logistic transformation

Lee (1993) has used a different method for restricting predicted TFR-values. He first transformed the annual TFR for the USA into

$$g_t = \ln\{(\mathrm{TFR}_t - L)/(U - \mathrm{TFR}_t)\}$$

where L and U are pre-specified lower and upper limits for the TFR. Next he identified and estimated an ARIMA-model for the transformed variable g_t. This logit transformation will produce a forecast for the TFR that will never exceed U or fall below L. However, as noted by Alho and Spencer (1997), such a model may have undesirable consequences. They demonstrated that when g_t follows a random walk process, then TFR_t will eventually be "absorbed" close to U or L for large enough t. This anomaly also showed up in our case. We selected $L=1.0$ and $U=3.1$, and identified a univariate ARIMA (2,1,0)-process for the logit transformed Norwegian TFR. (The maximum TFR-value in the period 1945-1995 was 3.02 in 1964. Hence an upper bound of 3.0 would cause the transformation to break down in that year.) Prediction intervals were computed analytically, assuming that the estimated coefficients of the ARIMA-process are equal to the real ones. In 2050, the bounds of the 67 per cent prediction interval (1.12, 2.86) were very close to those of the 95 per cent interval (1.01, 3.08). In the long run, the boundaries of any interval approach the upper and lower bounds U and L arbitrarily closely. The conclusion is that Lee's logit transformation cannot be used for constraining our prediction intervals.

3.1.5. Sensitivity analyses

The analysis in Section 3.1.1 differs from earlier studies in three respects. First, as Van Imhoff (1991) has noted, curve-fitting exercises almost invariably ignore the fact that a birth rate is not an observed quantity, but an estimate of the parameter of an underlying model. See, for example, Bell (1992, 1997); De Beer (1992); Duchêne and Gillet-De Stefano (1974); Knudsen et al. (1993); Thompson et al (1989); and the references they contain. (Note, however, Hoem (1976) and Hoem et al. (1981), who make a similar point, and establish an interesting link between weighted least squares estimation for the intensities F_x and minimum chi-square estimation for the counts of births B_x.) Second, ARIMA models for the TFR or other summary indicators show the same defect, compare, for example, Knudsen et al. (1993); Bell (1997); and Lee (1993). Finally, predictions on the basis of such time-series models assume that the parameters of the model are given, whereas in reality these are only estimates, with corresponding confidence intervals. How serious are these omissions and assumptions? To what extent do they lead to smaller confidence intervals and prediction intervals? We have opted for an empirical analysis of this question, and compared the results reported in the previous sections with corresponding results based on traditional assumptions. Below we summarize our findings for three cases: predictions with known ϕ-matrix, unweighted Gamma curve estimates, and unweighted ARIMA model estimates.

When the ϕ-matrix is known, simulation is unnecessary, and predictions for the TFR, the mean age, and the variance and corresponding prediction intervals can be computed analytically. We found that assuming the ϕ-matrix as given leads to 95 per cent bounds for the TFR and the mean age in 2050 that are too narrow by 8 per cent (0.5 child per woman) and 18 per cent (3.9 years), respectively. The 95 per cent bounds of the variance are little affected.

Unweighted Gamma curve estimates ignore the fact that low birth rates at young and old ages have small variances, and thus these rates should get more weight than high rates at intermediate ages. When rate variances are ignored, all weights w_x in expression (3.4) are chosen equal to one. The consequence is that the estimated TFR becomes higher compared to the weighted case, since the fitted curve follows the high rates more closely. To what extent the mean age and the variance are influenced, is an empirical matter. Fitting the Gamma curve using unweighted least squares to the data for the years 1945-1995, we found that estimates for all three parameters are higher than in the weighted case in almost every year. The difference is very small after 1970, but larger between 1945 and 1970, when the fit of the Gamma curve was somewhat less. In general, the results indicate that weighting has only had minor impact on the estimates for the TFR and the mean age, and a bit more for the variance in the first two decades after the war. However, the estimated parameter variances for the three parameters are relatively small in the unweighted case. This leads to narrow confidence intervals for many predicted age-specific rates. During the ages of high childbearing, the confidence intervals around the predicted age-specific rates are nearly half as large in the unweighted case compared to the weighted case. For younger and older ages the differences are much smaller.

Not only the birth rates, but also the three summary parameters are estimates, each with their own variance. We estimated a multivariate ARIMA (1,1,0) model for the three log-transformed parameters of interest. These parameters were computed on the basis of unweighted birth rates, and the ARIMA-model was estimated giving equal weights to the parameters in each year. The prediction intervals of such an "unweighted ARIMA-model" can be computed analytically, assuming that the estimated ϕ-matrix is the real one. We found a predicted TFR in 2050 which was almost the same as that in the weighted case, but the mean age, and particularly the variance increased to a higher equilibrium level. This is explained by the high estimates for ϕ_{22} (0.986, for the mean age) and ϕ_{32} (0.804, for the variance) that we obtained. The 95 per cent interval for the TFR is much narrower than in the weighted case (by 1.7 child in 2050), and that for the mean age a little so (by almost 4 years).[10] The consequence of ignoring weights thus is that we are too optimistic about the future TFR, in the sense that the prediction intervals are too narrow. At the same time we are too pessimistic regarding the future variance in the age at childbearing. In traditional cohort-component forecasting an error in the TFR is more important for the number of births, and hence for the population at young ages, than an error in the variance. If we would have ignored the sample variation in the historical age-specific birth rates in this particular empirical application, and also ignored the errors in the parameter estimates of the Gamma curve, we would have put too much confidence in subsequent births predictions.

3.1.6. Comparison with historical TFR-errors

Prediction intervals determine the *expected* errors in the *current* forecast. Investigating *observed* errors in *historical* forecasts can provide an independent check of the expected errors. We have analysed the errors in twelve TFR-forecasts that Statistics Norway has published between 1969 and 1996. (Errors in the mean age or in the variance have not been computed, because the assumed values for these indicators in historical forecasts have not been documented.) Assumed TFR-values for each forecast from the jump-off year until 1999 were compared with "observed" values (moment estimators). We have updated the data originally assembled by Texmon (1992), who collected, among others, TFR errors for the forecasts of 1969-1987 during the years 1969-1989. Most forecasts had more than one fertility variant, often two or three. In that case we included all variants in the data, because none of the forecasts, except two, contained a clear advice as to which of the variants was considered by Statistics Norway as the most probable one at the time of publication. Hence it was left to the user to pick one of the variants. The exceptions were the 1993- and 1996-based

[10] This empirical evidence is in line with theoretical results obtained by Koreisha and Fang (1999), who show that predictions generated by (univariate) ARMA models become more uncertain as a result of measurement errors, given certain conditions for the autocovariance function of the process.

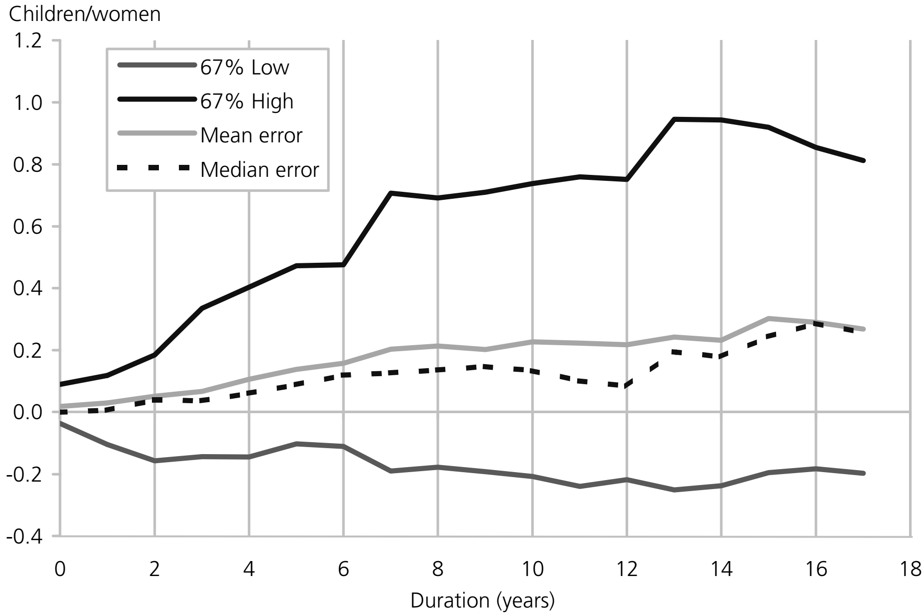
forecasts, for which it was clearly indicated that the medium fertility variant was considered as more probable than the high or the low variant (Statistics Norway 1994, 1997). However, users were also advised to investigate the consequences of choosing the other two variants as an input to their own plan or analysis. Only in case the user's conclusions would depend little on the choice for the two extreme variants, the user was advised to employ the medium variant. Thus we may assume that all variants have been used for the forecasts published between 1969 and 1996, although the middle one probably more often than the high or the low one (in case there were three variants).

The error in the TFR was simply defined as the assumed minus the "observed" value. Hence a positive or negative error indicates a value that is too high or too low. We had 28 series of TFR errors, and each series was ordered by forecast duration (the jump-off year was defined as duration 0). For each duration, the errors were ordered from low (including negative values) to high. Finally we selected, by linear interpolation, if necessary, two error values, such that one-sixth of the errors for each duration were lower, and one-sixth were higher than these values. Hence these two values can be interpreted as the bounds of an empirical 67 per cent "confidence" interval. Two out of three errors are within these bounds, and one-sixth of the errors is higher or lower.

Figure 3.11 shows the bounds of the 67 per cent interval for the TFR errors, together with the mean and the median error. The lower bound is approximately –0.2, close to zero. The error was larger than –0.2 in five out of every six cases was; most often it was positive. This reflects the fact that the strong fertility decline in the 1970s (see Figure 3.1) came as a surprise for Norwegian population forecasters, as was the case for demographers in many other Western countries. The distance between the mean or median error on the one hand, and the upper bound on the other, is larger than that between mean/median and lower bound, indicating that large errors were more frequent than small ones in the historical forecasts. The width of the 67 per cent interval grows from 0.13 children per woman in the jump-off year to 1.1 at duration 15. The historical error in Figure 3.11 increase somewhat faster than the expected ones do. After 15 years into the forecast period, the 67 per cent prediction interval in Figure 3.10 is only 0.9 children wide, but the general agreement between the two types of error is rather good. Thus the analysis in this section supports the main conclusions concerning the width of the prediction TFR-intervals in Section 3.1.3, namely that rejecting TFR-values outside a range of [0.5-4.0] children per woman results in realistic intervals[11].

[11] One may argue that using historical errors as a check for expected errors is too pessimistic, since one ignores improvements in forecast methodology. Indeed, average errors for the forecasts made in the 1980s are smaller than those for the forecasts from the 1970s. But this is to be expected, because the early forecasts are of longer duration, and hence more uncertain. When one wishes to estimate the average error in subsequent forecasts, controlling for this duration effect, a multivariate model is needed. An analysis of this kind carried out for Norwegian fertility forecasts showed very little improvement over subsequent forecast rounds (Keilman 1997). Period effects dominate the error pattern.

Figure 3.11. Empirical errors in historical TFR-forecasts, forecasts 1969-1996, period 1969-1999



3.2. Cohort fertility

3.2.1. Introduction

The predictions presented so far are pure period predictions. This is in line with the current literature on stochastic forecasts and related topics. Compare, for instance, the period TFR predictions by Thompson et al. (1989) and Bell (1992) for the US 1984-2020, Lee (1993) and Lee and Tuljapurkar (1994) for the US 1990-2065, Hanika et al. (1997) and Lutz and Scherbov (1998a) for Austria 1996-2010, and (1998b) for Germany 1995-2030, and finally Alho (1998) for Finland 1998-2050. Little is known, however, regarding the intervals for the *cohort* analogue of the TFR, i.e. the Completed Cohort Fertility (CCF). On the one hand, the CCF often displays a smoother time trend than the TFR (Akers 1965, 417). Therefore one might assume that it is easier to predict the CCF than the TFR (De Beer, 1992), which suggests that the CCF prediction interval is relatively narrow. But on the other hand, it takes some 30 years after women had their first child, before they complete childbearing. Hence current young cohorts display massive incomplete data problems, and this would tend to widen the intervals.

There are various ways of predicting the CCF. The starting point is a table of empirical age-specific fertility rates for a number of years. The time series of empirical CCFs may be modelled directly by univariate ARIMA methods. This ignores the information contained in fertility rates for young cohorts in recent years. De Beer (1985) identifies and fits a so-called Cohort ARIMA (CARIMA) model to the table of fertility rates. The method combines two ARIMA models: one in the age direction, and one in the cohort direction. Next De Beer extrapolates age-specific fertility rates. Willekens and Baydar (1984) fit an age-period-cohort model (APC model) to the fertility table, and extrapolate the period and the cohort effects separately. They estimate age effects, and assume that these are constant over time. Combining extrapolated period effects and cohort effects, together with the constant age effects results in predicted fertility rates by age. Thompson et al. (1989) use a multivariate ARIMA model for the three parameters of a gamma curve describing the age pattern of period fertility.

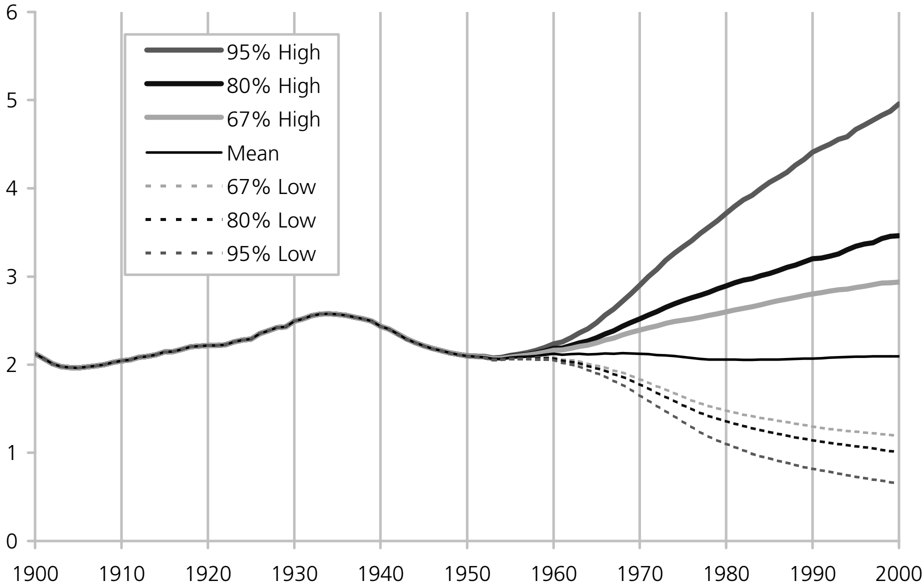
Although none of the authors mentioned here addressed this issue, the CCF could have been found simply by rearranging extrapolated fertility rates by birth cohort. This is the method adopted in this Section, which elaborates on the approach adopted by Thompson et al. More specifically, of the four sources of variance mentioned in Section 3.1, we will not only take the residual variance in the time series model into account (as is usually done), but also the other three.

This Section presents findings for prediction intervals of Norwegian cohort fertility. Section 3.2.2 shows such intervals obtained by rearranging predicted age specific fertility rates from a period to a cohort perspective, on the basis of unrestricted period fertility simulations. In Section 3.2.3 cohort patterns are explored with restricted period fertility. Finally, Section 3.2.4 compares the high and low cohort fertility assumptions from the 1996-based population forecast by Statistics Norway with the prediction intervals.

3.2.2. Prediction intervals for cohort fertility

Age-specific fertility rates predicted for the years 1996-2050 (see Section 3.1) were complemented with observed rates for the period 1900 to 1995, and the whole array of age-specific rates from 1900 to 2050 was rearranged by birth cohort. This led to one time-series for the CCF and other cohort fertility indicators - partly observed, and partly extrapolated. The process was repeated for each of the 5 000 simulations described in Section 3.1. Figure 3.12 gives the prediction intervals for the CCF. These are based on the *unrestricted* period fertility simulations reported in Section 3.1.2. CCF-prediction intervals based on *restricted* period fertility simulations will be reported in Section 3.2.3.

Figure 3.12. Completed Cohort Fertility. Birth cohorts 1900-2000



First, note that the mean value for the CCF is almost constant at 2.1 children per woman for cohorts born after 1960. Second, the CCF displays, on the long run, prediction intervals of the same width as those for the period TFR in Figure 3.5, provided one takes the different time scales into account. For instance, the two-thirds CCF-interval for women born in the year 2000 ranges from 1.19 to 2.94 children per woman. This agrees closely with the bounds of the TFR-interval in the year 2030, the year in which those women will attain their mean age at childbearing: 1.19 and 2.90. The difference between Figures 3.5 and 3.12, however, concerns the intervals for the birth generations of the 1960s and 1970s. Still comparing CCF-intervals for women born in a given year with TFR-intervals 30 years later, we see that the CCF-interval for women born in the mid-1960s is three times as wide as that for the TFR in the mid-1990s. But the widths converge quickly for younger cohorts and later years: 25 per cent wider for cohorts born around 1970, compared to years around 2000, and only 10 per cent wider for cohorts born around 1980/years around 2010[12].

The TFRs before 1996 in Figure 3.5 were not extrapolated, but estimated on the basis of observed data up to 1995. Hence prediction intervals for years up to

[12] If we shift the two time scales over a period of 28 years, instead of 30, the differences between Figures 3.5 and 3.12 remain essentially the same. Yet initial differences are larger, but on the long run the two intervals still have the same width.

1995 are negligibly narrow: there is very little uncertainty due to estimation. But at the same time, the CCF for cohorts born before 1996-30=1966 is still a bit uncertain, which is reflected in the prediction intervals for those cohorts in Figure 3.12. For example, the 67 per cent interval is 0.01 children wide for women born in 1954. This reflects the fact that these women had not yet entirely completed their childbearing in 1995, when they were 41 years old. Younger cohorts had more years to go in 1995, and hence the intervals become increasingly wider: the 67 per cent interval is 0.08 for cohort 1960, and 0.26 children per woman for women born in 1965. It takes the CCF interval some 16 years to grow from 0.1 to 1.0 children wide - uncertainty in the TFR grows over the same range during a shorter period: 13 years.

In summary, we can say that on the long run, the CCF is as uncertain as the TFR, but the uncertainty in the CCF starts earlier, and therefore increases slower than that in the TFR. This is a mere reflection of the fact that as time goes by, more and more young birth cohorts are exposed to the risk of childbearing over a longer span of their reproductive life.

Figure 3.13. CCF, 95 per cent prediction interval, minimum TFR=0

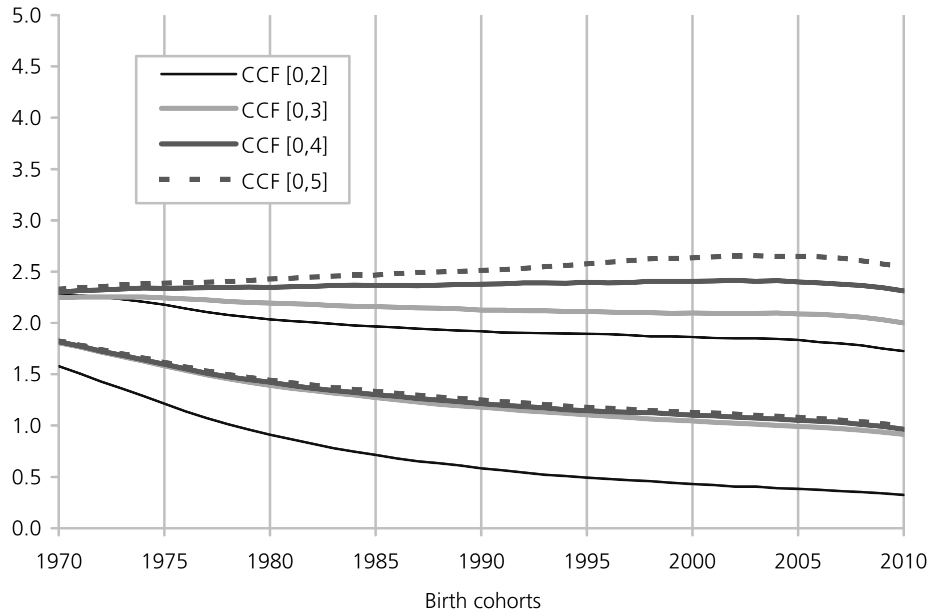
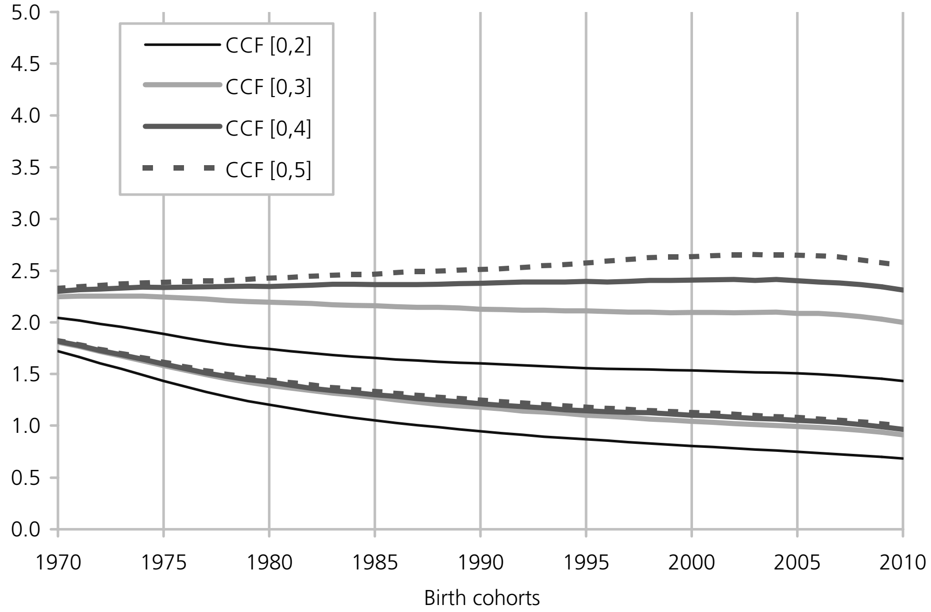


Figure 3.14. CCF, 67 per cent prediction interval, minimum TFR=0



3.2.3. CCF-prediction intervals when the period TFR is restricted

Figures 3.13 and 3.14 show how the prediction intervals for the Completed Cohort Fertility change when we select various limits for the period TFR. Again we see that a TFR ceiling equal to 3 children per woman would lead to a slight decrease in the upper 67 per cent and 95 per cent bounds, which is unreasonable. Selecting a maximum TFR of 4 or 5 children per woman gives a more realistic pattern. Note that the slight fall in the upper bound after birth cohort 2005 is an anomaly in the simulations: the simulations stop in the year 2050, so that cohorts born in 2000 or later are progressively truncated at their higher ages. As with the period TFR, the width of the 67 per cent interval on the short run changes relatively little when the maximum TFR is increased from 4 to 5 children.

Prediction intervals based on a minimum TFR equal to 1 are a bit narrower than those in Figures 3.13 and 3.14, in particular because the lower bounds of the intervals are lifted up.

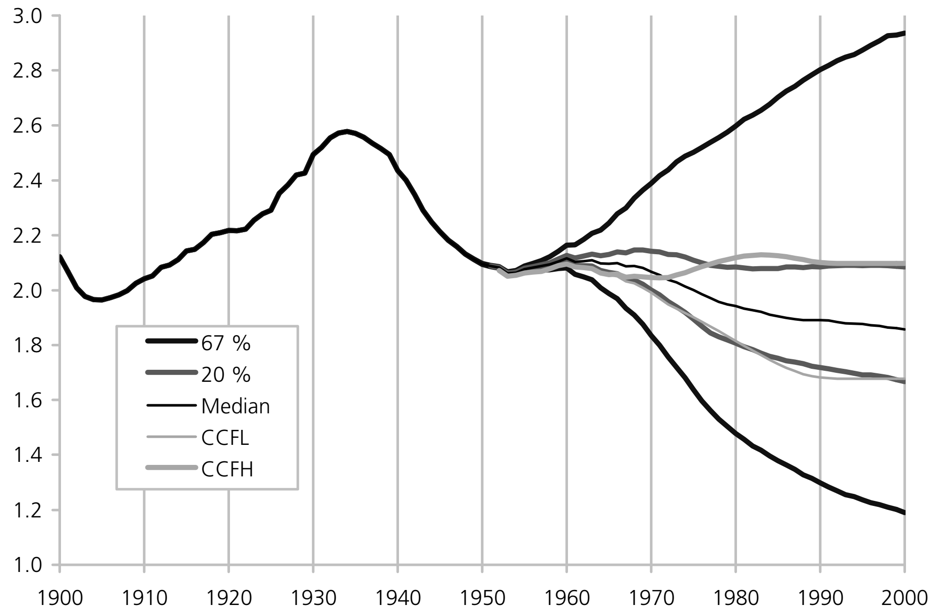
3.2.4. A comparison with CCF-assumptions in Statistics Norway's population forecast

How do the CCF prediction intervals compare with the high and the low variants for cohort fertility in the 1996-based population forecast of Statistics Norway? In that forecast, the year 2010 was chosen as the target year for fertility, that is, the year after which no change was assumed for fertility parameters. During the years 1996-2010, the period TFR was assumed to be more or less constant at a level of between 1.7 and 2.1 children per woman, while a further increase in the mean age was foreseen from 28.8 years in 1995 to 29.5-30.5 years, see Table 3.4 in Section 3.1.3.

Figure 3.15 shows the CCF assumed by Statistics Norway, together with prediction intervals obtained by simulation. We notice that for cohorts born after 1975, no more than about 20 per cent of the simulations were covered by the official high-low intervals. In other words, the expected probability that the future CCF will fall between the margins defined by Statistics Norway is only one out of five. The corresponding probability for the (unrestricted) period TFR is a bit higher, but still very low: only 29 per cent in the year 2010.

Thus we see that Norwegian forecasters have been quite optimistic when they assumed so narrow bounds between the low and the high fertility variants in the official forecast. One may argue that this conclusion is a consequence of the fact that the prediction intervals in Figures 3.5, 3.12, and 3.15 are far too wide. However, as mentioned in Section 3.1, independent evidence in the form of errors in historical TFR-forecasts supports the prediction intervals for the period TFR, at least to the year 2010.

Figure 3.15. Completed cohort fertility, assumed in Statistics Norway's 1996 forecast (CCFL and CCFH), and prediction intervals



Hence the interval defined by the high and low CCF-variants in Statistics Norway's population forecast of 1996 had an expected accuracy of no more than approximately 20 per cent, given our model. Is this conclusion affected by the choice of the maximum TFR values in the model? We found that the width of the 20 per cent prediction interval changes very little, when the maximum TFR value is set to 4 or 5 children per woman, as compared to the virtually unrestricted TFR simulations in Figures 3.13 and 3.14. Selected results are given in Table 3.5. For birth cohort 2000, the interval becomes 0.34-0.38 children wide, when a maximum TFR of 4-5 is chosen. This is somewhat less than the width of 0.42 children per woman for the 20 per cent interval in Figure 3.15. In other words, the expected accuracy of the official high-low interval is somewhat higher than 20 per cent, but still very limited. This may be one of the reasons why the high-low TFR interval was changed to [1.5-2.1] in the 1999-forecast of Statistics Norway.

Table 3.5. Width of 20 per cent prediction intervals for the CCF

	TFR<10	TFR<5	TFR<4
Cohort			
1970	0.14	0.12	0.12
1980	0.28	0.25	0.24
1990	0.37	0.32	0.31
2000	0.42	0.38	0.34

4. Mortality

4.1. The approach in general

Prediction intervals for death rates by age and sex were obtained on the basis of the following method.

- Death probabilities for ages 0, 1, 2, … 100 by sex for the years 1945-1995 were taken from Mamelund and Borgan (1996). We assumed that these probabilities were generated by a first order homogeneous Markov process with constant mortality intensity over each one-year age interval.
- The predictions consisted of three steps.
  - First, a life table calculation resulted in annual values of the life expectancy at birth for men and women for the period 1945-1995. The two time series of life expectancies were modelled by means of a bivariate ARIMA model, and stochastic simulation resulted in prediction intervals for male and female life expectancies for the period 1996-2050. Section 4.2 gives details.
  - Second, we assumed that for each year the historical age pattern of mortality could be described by means of a Heligman-Pollard (H-P) curve. This is a mathematical function with eight parameters, see Section 4.3. This resulted in two time series of parameter estimates for the years 1945-1995, one for each sex. The time series were modelled by means of multivariate time series models of the ARIMA-type. Stochastic simulation yielded 5 000 multivariate sample paths for the parameters for the period 1996-2050. The H-P curve was used to transform the parameter predictions back into future age-specific mortality rates. For each year in the prediction period, a life table calculation summarized those rates into one indicator, viz. the life expectancy at birth. These life expectancies were assembled in a look-up table, together with the underlying H-P parameter values.
  - In the third and final step, H-P parameter values and age/sex-specific death rates from the look-up table were assigned to life expectancy values for each sample path, predicted in the first step, by matching life expectancy values from the first and the second step, controlling for calendar year and sex.

In a trial calculation, the death rates that resulted from the H-P extrapolations in step 2 were directly employed in the stochastic cohort component simulations. This, however, led to unacceptably large prediction intervals for the life expectancy at birth. For instance, in 2050, the 95 per cent interval was 12.2 years for men and 19.1 years for women (4.5 and 9.4 years for the 67 per cent intervals). At the same time we obtained correlation coefficients between male and female mortality (as measured by the correlations of estimates for time series coefficients for corresponding H-P-parameters for men and women) of between 5 and 15 per cent this way. This indicated an unrealistically weak correlation between the sexes. On the other hand, by predicting the life expectancy at birth directly, the 95 per cent intervals became narrower, and correlations across the sexes increased considerably, see below. Thus, the direct predictions of the life expectancy were accepted, and step 3 above was used to assign age-specific death rates to the life expectancy values of step 2. By the nature of the approach, the correlation across time of H-P parameters and age-specific death probabilities was not preserved. Therefore, these correlations were checked for death probabilities, by comparing one-step ahead autocorrelations for selected ages as observed in the historical data with similar autocorrelations computed for the simulations. Details will be given in Section 4.3.

4.2. A multivariate ARIMA model for the life expectancy at birth

The following ARIMA (2,0,0) model was estimated for the log of male and female life expectancies for the period 1945-1995.

$$\begin{bmatrix} ln(e_t^M) \\ ln(e_t^F) \end{bmatrix} = \begin{bmatrix} K^M \\ K^F \end{bmatrix} + \begin{bmatrix} \phi_1^M & 0 \\ 0 & \phi_1^F \end{bmatrix} \begin{bmatrix} ln(e_{t-1}^M) \\ ln(e_{t-1}^F) \end{bmatrix} + \begin{bmatrix} \phi_2^M & 0 \\ 0 & \phi_2^F \end{bmatrix} \begin{bmatrix} ln(e_{t-2}^M) \\ ln(e_{t-2}^F) \end{bmatrix} + \begin{bmatrix} \varepsilon_t^M \\ \varepsilon_t^F \end{bmatrix} \tag{4.1}$$

Parameter estimates are given in Table 4.1.

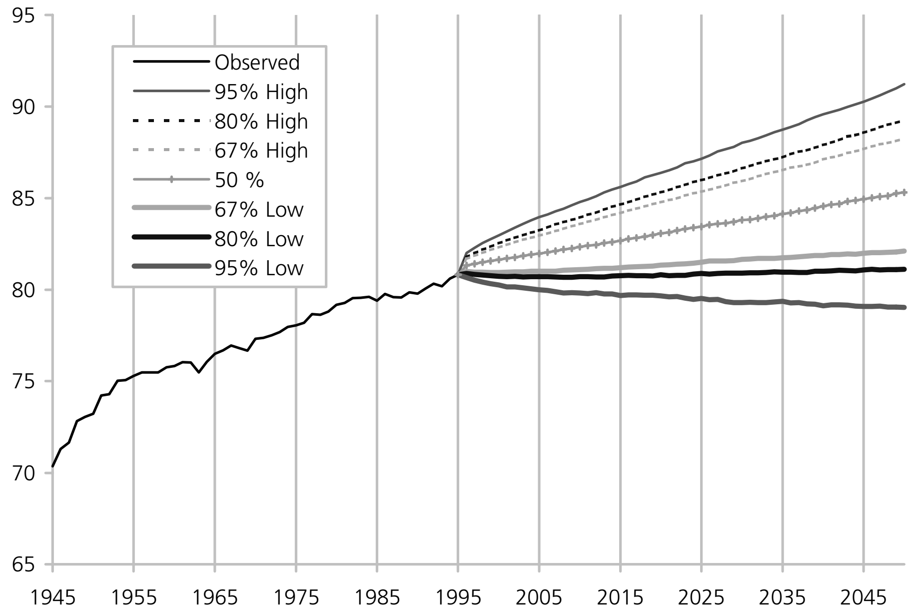
Table 4.1. Parameter estimates for model (4.1). Estimated standard errors in parentheses

	ϕ_1	ϕ_2	K
Men	0.9518	0.0488	-0.00145
	(0.0998)	(0.0999)	
Women	0.9089	0.0918	-0.00247
	(0.1146)	(0.1147)	

The ϕ_2-estimates for men and women were not significantly different from zero. Yet these have been retained in order to avoid autocorrelation in the residual. The constant vector (K^M,K^F) was not estimated, but computed directly. To begin with, model (4.1) did not contain a constant - in other words, (K^M,K^F) was equal to the null vector. Predictions to 2050 resulted in values for the life expectancy that were much higher than those used by Statistics Norway in its population forecast: model (4.1) predicted 86.7 years for men and 95.8 years for women, whereas the official values are 80 and 84.5 years, respectively. Therefore, the null vector was adjusted in such a way that model (4.1) gave the official predictions in 2050. The adjustment method, which is a variant of Lee's (1993) method for univariate ARIMA models, is described in Appendix 4.1.

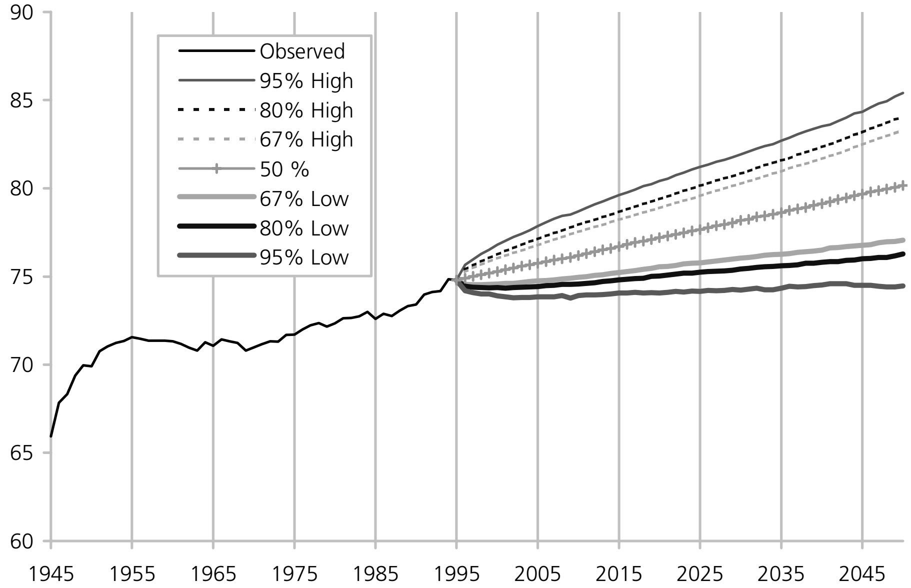
The correlation between corresponding ϕ-estimates for men and women turned out to be 45 per cent, both for ϕ_1 and ϕ_2. We also experimented with a bivariate ARIMA (1,1,0) model for the log of the life expectancies, and found a correlation of ϕ-estimates across the sexes equal to 62 per cent. (This model was rejected, because it left too much autocorrelation in the residuals for women.) Although the real correlation across the sexes for mortality is not known[13], these values seem to be much more reasonable than those obtained on the basis of ϕ-matrices for the H-P parameters (5-15 per cent, see above). This gives further support to modelling life expectancies directly, instead of deriving future life expectancy values from predicted H-P parameters.

Figure 4.1. Prediction intervals for life expectancy at birth. Men



[13] Alho (1998) estimates for mortality in Finland in the period 1900-1994 a correlation of 80 per cent between male and female age-specific death rates.

Figure 4.2. Prediction intervals for life expectancy at birth. Women



Prediction intervals for the life expectancy at birth by sex are shown in Figures 4.1 and 4.2. These are obtained on the basis of 5 000 simulations, in which the ϕ_1-matrix, the ϕ_2-matrix and the ε-vector were treated as multivariate normal variables with known means and covariances, and the *K*-vector was assumed fixed. The latter assumption implies that the prediction intervals for the life expectancy are too narrow, but we do not know by how much. On the other hand, in Section 4.5 we shall see that the intervals are wider than what could be expected on the basis of empirical errors in forecasts since 1969. Therefore it was decided to accept the intervals as displayed in Figures 4.1 and 4.2.

The 95 per cent intervals appear to be 11.1 (men) and 12.2 (women) years wide in 2050. In both cases this implies a relative width (compared to the median value) of 14 per cent. The correlation between male and female life expectancy turns out to be 0.50 in 2050, decreasing from 0.63 in 1996, 0.56 in 2010 and 0.51 in 2030.

Excess mortality of men, compared to women, is likely to decrease further in the future. In the period 1991-1995, the life expectancy at birth for men was 6 years lower than that for women, a slight reduction compared to the difference of 6.8 years which was observed in the mid-1980s. In 2010, the gap is reduced to an expected 5.7 years (the 95 per cent prediction interval for this difference equals (3.3, 8.0)), and it drops further to 5.2 years in 2030 (1.1, 9.4) and 4.7 years in 2050 (-1.3, 10.7). The lower bound of the 95 per cent interval falls below zero in 2040, implying a 2.5 per cent chance for higher female than male mortality.

In 1999 Statistics Norway published its most recent population forecast, in which it is assumed that the life expectancy of men will increase from 75.5 in 1999 to between 77 and 83 years in 2050, and for women from 81.2 to 81.5-87.5 years. Assuming that our model is correct, we can predict the expected accuracy of the official life expectancy forecasts. In 2050, this accuracy turns out to be 64 per cent for men, and 62 per cent for women. For both sexes, it is much lower in the beginning of this century (36 per cent in 2010 for men, and 35 per cent for women in 2000) but increases steadily to the 2050-values.

The fact that we accounted for estimation variance of the matrices ϕ_1 and ϕ_2, and not only the residual ε_t, had a stronger effect for women than for men. For instance, when the matrices are assumed as given and only the residual is stochastic, the 95 per cent prediction interval for the male life expectancy in 2050 is 9.5 years wide – compare this value with the interval of 11.1 years assuming stochastic ϕ-matrices. For women the interval becomes 9.2 years wide in 2050 (12.2 years with stochastic ϕ-matrices).

Our results indicate similar uncertainty for the life expectancy at birth as that estimated by De Beer and Alders (1999) for the case of the Netherlands. These authors modelled the time series of life expectancies for the two sexes separately based on a random walk with drift, and found a 95 per cent prediction interval of 12 years wide in 2050, both for men and for women.

Intervals presented by Tuljapurkar et al (2000) for the G7 countries (Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States) are much smaller than ours are. Their 90 per cent intervals of combined-sex life expectancy at birth in 2050 range from a minimum of 2.8 years for Canada to a maximum of 7.5 years for the UK. These intervals are based on a one-parameter Lee-Carter model for age-specific mortality for the two sexes combined. The single parameter was modelled as a random walk with drift. The authors used an abridged life table with five-year age classes up to 80-84. Ages 85 years and higher were lumped into one age class (except for Japan). The age and sex aggregation, which reduces random fluctuations, may have caused these relatively narrow intervals.

4.3. The Heligman-Pollard curve

Demographers, statisticians and actuaries have long been occupied with finding a suitable "law of mortality", i.e. a mathematical representation of age-specific mortality - most often (but not exclusively) in terms of the death rates. A large number of such laws have been proposed, a process that started with De Moivre in 1725, and that continues until the present day (Hannerz 2001). For a recent review, see Tabeau (2001). Some of the laws are restricted to adult or old age mortality (Coale-Kisker, Himes-Preston-Condran, Gompertz, Perks, Weibull). Tabeau (2001) and Boleslawski and Tabeau (2001) compare some 27 of such laws. Relational models, such as Brass' logit model, and the Lee-Carter model, are also considered. Among the laws and models, the Heligman-Pollard curve is widely applied (Rogers and Gard 1991; McNown and Rogers 1989; Hartmann 1987; Kostaki 1992a,b). Compared to other models, the H-P curve has the advantage that it pairs accuracy in prediction with flexibility, in particular when describing mortality changes over time. Hartmann (1987) used Swedish mortality data for the period 1900-1970, and concluded that the H-P curve is a useful model for making population projections, one reason being the fact that it accommodates for changing age patterns of mortality as the level of mortality changes.

We have used the following formula for the Heligman-Pollard curve

$$(4.2) \qquad q_x = A^{(x+B)^C} + D\,exp\left\{-E(\log x - \log F)^2\right\} + \frac{GH^x}{1+GH^x}, \ x \neq 0,$$

where q_x is the one-year probability of dying at age x, and the eight parameters *A-H* are to be estimated from the data (Heligman and Pollard 1980). The three terms model childhood mortality (*A,B,C*), the accident hump (*D,E,F*), and adult through old age mortality (*G,H*). Since *B* is very small for developed countries (indeed, we found after some experimentation that it could be put equal to zero for the case of Norway), *A* effectively represents infant mortality, and *C* the rate of decline of child mortality. *F* is the locality of the accident hump, *E* its spread (a large *E* reflects a narrow hump), and *D* its severity. The third term is a modified Gompertz curve, which decelerates at higher ages. *G* is the base level and *H* the increase in adult mortality.

Estimation was carried out by means of Relative Least Squares, using the Marquardt estimation method. We minimized the sum

$$\sum_x (1-\frac{q_x}{\hat{q}_x})^2 \ , x = 1, 2, 3, \dots 97,$$

with $\hat{q}_x$ the empirical probability. The Marquardt method requires first derivatives of expression (4.2) with respect to the unknown parameters. These were calculated by means of the program MAPLE. Ages 0, 98, 99 and 100 were treated differently[14].

[14] Since the estimate of *B* was found to be zero, q_0 cannot be estimated on the basis of the HP-curve, not even when we restrict ourselves to the first term for childhood mortality: in that case q_0 would invariably be equal to one. Therefore we used a different procedure, and assumed simply that $q_{0,t} \approx q_{\alpha,t-1}$. For men we found that α=0.1100 fitted the data well, and for women α=0.3715. Empirical q_x-values for $x \geq 98$ are highly irregular. Therefore these were not included in the curve fitting exercise. Predicted values for ages 98 and 99 are based on age-extrapolations of the HP-curve, fitted for ages 1-97. For age x=100, the cohort-component method requires a death probability for the *open* age group 100+. This probability was set equal to 0.5 for men and women in each year. For the life tables, q_{100} was set equal to 1.

We met considerable difficulties when estimating the parameters. Often the algorithm did not converge, or resulted in estimates that were dependent upon the initial values. Fitting the curve not to the original series of empirical probabilities $\hat{q}_x$, but to a three-term moving average of that series in the age dimension solved this problem. Indeed, few of the earlier applications of the HP-curve have successfully applied probabilities for one-year age groups *and* single calendar years. Some averaging over time, in the age dimension, or both seems to be necessary, compare Heligman and Pollard (1980), Kostaki (1992a,b), Bell (1997), and the references therein. For an exception, see Tabeau et al. (1999).

The initial value for each parameter was found by splitting the curve up into three parts, and estimating them separately (for each year and sex):

- $q_x = A^{(x+B)^C}$ for $x = 1,2,...10$
- $q_x = D\,exp\left\{-E(log\,x - log\,F)^2\right\}$ for $x = 11,12,...30$
- $q_x = \dfrac{GH^x}{1+GH^x}$ for $30 < x < 98$.

After some experimentation we found that the starting values obtained this way for the year 1995 could serve as starting values for the whole curve during the entire period. The fit for women at high ages was unsatisfactory. In many years, predicted rates for ages 80 or higher were lower than empirical rates. Therefore the third part of the curve was replaced by a pure Gompertz curve (GH^x), so that formula (4.2) for women reads

$$(4.3) \qquad q_x = A^{(x+B)^C} + D\,exp\left\{-E(log\,x - log\,F)^2\right\} + GH^x,\ x \neq 0.$$

Expression (4.3) gave a satisfactory fit for women for all years. In the following, when we refer to the Heligman-Pollard curve, we refer to expression (4.2) for men and expression (4.3) for women (with B=0 and E=1, see below).

The estimate for the parameter B turned out to be insignificant at the 5 per cent level, both for women and for men. Hence this parameter was set equal to zero for the whole period. For women, the estimate of E was close to one for many years, and hence was fixed at that value. This left us with seven parameters for men, and six for women.

4.4. A multivariate ARIMA model for the parameters of the Heligman-Pollard curve

The result of the calculations reported in Section 4.3 was a series of estimates $\hat{A}_t, \hat{C}_t, \hat{D}_t, \hat{E}_t, \hat{F}_t, \hat{G}_t, \hat{H}_t$ for men, and $\hat{A}_t, \hat{C}_t, \hat{D}_t, \hat{F}_t, \hat{G}_t, \hat{H}_t$ for women, together with their estimated covariance matrices $\hat{\Omega}_t$, t = 1945, ..., 1995.

The next step was to fit a multivariate ARIMA model to the time series of (log-transformed) parameters. A (1,1,0) model was found to fit well. For men, the model can be written as

$$(4.4)\quad \begin{bmatrix} \nabla\log A_t \\ \nabla\log C_t \\ \nabla\log D_t \\ \nabla\log E_t \\ \nabla\log F_t \\ \nabla\log G_t \\ \nabla\log H_t \\ 1 \end{bmatrix} = \begin{bmatrix} \Phi_{AA} & \Phi_{AB} & \Phi_{AD} & \Phi_{AE} & \Phi_{AF} & \Phi_{AG} & \Phi_{AH} & K_A \\ \Phi_{CA} & \Phi_{BB} & \Phi_{CD} & \Phi_{CE} & \Phi_{CF} & \Phi_{CG} & \Phi_{CH} & K_C \\ \Phi_{DA} & \Phi_{DB} & \Phi_{DD} & \Phi_{DE} & \Phi_{DF} & \Phi_{DG} & \Phi_{DH} & K_D \\ \Phi_{EA} & \Phi_{EB} & \Phi_{ED} & \Phi_{EE} & \Phi_{EF} & \Phi_{EG} & \Phi_{EH} & K_E \\ \Phi_{FA} & \Phi_{FB} & \Phi_{FD} & \Phi_{FE} & \Phi_{FF} & \Phi_{FG} & \Phi_{FH} & K_F \\ \Phi_{GA} & \Phi_{GB} & \Phi_{GD} & \Phi_{GE} & \Phi_{GF} & \Phi_{GG} & \Phi_{GH} & K_G \\ \Phi_{HA} & \Phi_{HB} & \Phi_{HD} & \Phi_{HE} & \Phi_{HF} & \Phi_{HG} & \Phi_{HH} & K_H \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \nabla\log A_{t-1} \\ \nabla\log C_{t-1} \\ \nabla\log D_{t-1} \\ \nabla\log E_{t-1} \\ \nabla\log F_{t-1} \\ \nabla\log G_{t-1} \\ \nabla\log H_{t-1} \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{A,t} \\ \varepsilon_{B,t} \\ \varepsilon_{D,t} \\ \varepsilon_{E,t} \\ \varepsilon_{F,t} \\ \varepsilon_{G,t} \\ \varepsilon_{H,t} \\ 0 \end{bmatrix}$$

or $Z_t = \Psi Z_{t-1} + \varepsilon_t$ for short. $\nabla\log A_t$ represents the first difference $\log \hat{A}_t - \log \hat{A}_{t-1}$. ε_t is assumed to be a multivariate white noise process $N(0, \Sigma_\varepsilon)$. For women, the model is similar, except that rows and columns for the parameter E are not included. Note that the model contains an intercept term (constants K_A, ... , K_H), as opposed to the fertility model in Chapter 3. The reason is that we want to allow the life expectancy to increase for the whole projection period up to 2050.

Estimation of the Φ_{ij}-coefficients was done in various steps. First, for each sex the full matrix was estimated by Ordinary Least Squares.[15] Next, in order to keep the number of estimates to a minimum, most of the estimates that were not significant at the 5 per cent level were set equal to zero, and the model was re-estimated using Restricted Least Squares.[16] After that, initial projections were made for the parameters *A-H* up to 2050, and these were converted into age-

[15] This means that correlation across the sexes in the age pattern of mortality was ignored. It is, however, included in the life expectancy.
[16] In case none of the estimates for a certain parameter (*A-H*) was significant, the corresponding estimate on the main diagonal of Ψ was retained. Estimates for the constants K_A-K_H were retained as much as possible, with a view to the adjustment method to be described below.

specific probabilities of dying and life tables. Finally the constants K_A-K_H were adjusted (using the method of Appendix 4.1) such that the resulting adjusted life table in 2050 predicted a life expectancy at birth of 84.5 years for women and 80 years for men. Table 4.2 gives the estimates for the non-zero elements of the Ψ-matrices for men and women.

Model predictions for the HP-parameters showed that the level parameters for childhood mortality (*A*), adolescent mortality (*D*), and adult mortality (*G*) are likely to fall further. For men, the reductions in *A* and *G* are stronger than that in *D*, so that expected age-specific mortality up to 2030 still shows a slight accident hump. More detailed information will be given in the next section.

Table 4.2. Estimates of non-zero elements of the Ψ-matrices for men and women, and corresponding standard errors. *K*-estimates in parentheses are adjusted for e_0-target value

	Men		Women	
	Estimate	Standard error	Estimate	Standard error
Φ_{AA}	-0.616	0.064	-0.436	0.096
Φ_{CC}	-0.549	0.074	-0.509	0.077
Φ_{DD}	-0.238	0.107	-0.542	0.089
Φ_{EE}	-0.644	0.073	-	-
Φ_{FF}	-0.398	0.094	-0.744	0.117
Φ_{GG}	-0.412	0.071	-	-
Φ_{HH}	-0.439	0.071	-0.454	0.125
Φ_{AD}	-	-	0.093	0.028
Φ_{DA}	-	-	0.526	0.222
Φ_{FA}	0.110	0.037	-	-
Φ_{FC}	-0.130	0.049	-	-
Φ_{FD}	0.056	0.026	-	-
Φ_{FG}	-	-	0.391	0.150
Φ_{HF}	-	-	-0.004	0.001
Φ_{HG}	-	-	-0.005	0.002
K_A	-0.073	0.013	-0.054	0.011
	(-0.275)		(-0.028)	
K_C	-	-	-	-
	(-3.17E-8)		(-2.97E-7)	
K_D	-0.021	0.019	-	-
	(-0.080)		(1.11E-4)	
K_E	0.035	0.036	-	-
	(0.132)			
K_F	-	-	-	-
	(-0.000)		(2.3E-5)	
K_G	-0.003	0.003	-0.013	0.007
	(-0.010)		(-0.007)	
K_H	-	-	-	-
	(2.54E-8)		(-2.05E-7)	

Table 4.3. Autocorrelation in death probabilities for selected age/sex combinations

	Simulations 1996-2050			Observed 1945-1995
	Mean	Median	2.5% and 97.5% percentiles	
Men				
Age 1	0.815	0.819	(0.726, 0.885)	0.859
Age 25	0.538	0.578	(0.193, 0.757)	0.513
Age 50	0.503	0.562	(0.003, 0.843)	0.448
Age 75	0.690	0.823	(0.057, 0.937)	0.595
Age 99	0.304	0.299	(-0.073, 0.676)	-0.026
Women				
Age 1	0.815	0.819	(0.726, 0.885)	0.875
Age 25	0.079	0.068	(-0.159, 0.357)	0.793
Age 50	0.314	0.320	(-0.050, 0.656)	0.690
Age 75	0.683	0.787	(0.161, 0.928)	0.889
Age 99	0.167	0.149	(-0.164, 0.551)	0.504

At ages 1-10, predicted death probabilities for girls were almost invariably higher than those for boys. This is explained by the fact that in the period 1980-1995, the parameter *A* falls more steeply for boys than for girls. Consequently the ARIMA model, which puts much weight on recent observations, predicts a steeper fall in the *A*-parameter for boys than for girls, which in turn causes lower death probabilities for boys. This was considered unrealistic, and therefore we decided to reduce girls' probabilities to the level of those for boys for ages 1-10, in those simulations where girls in this age group would have higher mortality than boys. The latter turned out to be the case in virtually all simulations. An experimental calculation showed that this *ad hoc* correction of female mortality would result in a life expectancy that is higher by 0.1 years at most, which was considered acceptable.

As mentioned in Section 4.1, by the nature of the approach we have adopted for mortality simulation, the correlation across time of H-P parameters and age-specific death probabilities was not preserved in the simulations. Therefore, this correlation was checked for death probabilities, by comparing one-step ahead autocorrelations for selected ages as observed in the historical data with similar autocorrelations computed for the simulations. Each simulation results in one autocorrelation value, given sex and age. Table 4.3 gives summary statistics for the autocorrelations in the simulations, as well as historical values.

The table shows that the historical autocorrelation structure for male mortality is rather well preserved in the simulations, while the autocorrelation values for females generally are low compared to historical values.[17] This means that simulated death probabilities for women vary somewhat stronger over time than historical probabilities did. For the life expectancy however, the historical autocorrelation structure was preserved, by the very use of model (4.1).

[17] At age 1, the autocorrelation does not differ between the sexes, for the reason described earlier in this section.

4.5. Prediction intervals for age and sex specific mortality

Figures 4.3 and 4.4 illustrate how the age pattern of mortality and the prediction intervals around the death probabilities will change over the years. For men, the accident hump disappears between 2010 and 2030, while for women this is already the case in the years before 2010. The reason for irregular mortality for women around age 10 has been explained in the previous section. The probabilities involved are very low. Note that prediction intervals are wider for women than for men, which can be attributed to a relatively good fit of the ARIMA-model for the male H-P parameters. Table 4.2 shows that the time-series model for the parameters *A* and *H* in particular are more satisfactory for men than for women. Figures 4.3 and 4.4 suggest that the intervals narrow when age increases, but this is an artefact of the logarithmic scale. Except for some irregularities at the lowest ages, intervals become wider at more advanced ages. For instance, in 2050, the 95 per cent interval widens regularly from a minimum of 7.3E-5 (at age 1, both sexes) to 0.25 for men and 0.30 for women, both at age 99.[18]

The results in this section are based on stochastic simulation, in which both the distribution of the residual ε_t *and* that of the coefficient matrix Ψ were taken into account. This approach acknowledges uncertainty sources 3 and 4 mentioned in Section 2.2. However, in contrast to the fertility simulations, sources 1 and 2 were *not* considered in the mortality simulations. More specifically, when model (4.4) was estimated, the annual values for the Heligman-Pollard parameters *A-H* were treated as data. In reality, these values are estimates, and their precision varies over time. When model (4.4) is estimated, years in which the covariances for parameters *A-H* are large should get less weight, compared with years in which covariances are low. This can be achieved by Weighted Least Squares (WLS) estimation of (4.4), with weights equal to the inverse of the square root of the covariance matrix. But predictions for the year 2050 of the parameters *A-H* based on an ARIMA model estimated by WLS gave unrealistically high values for parameters *C*, *F*, and *H*, both for men and women. The reason is that the covariance matrix of the estimates *A-H* is extremely unbalanced. In turn, this leads to an unbalanced matrix of weights: for women in 1995, for instance, the weights range from 0.0012 for *G*, to 741 for *H*. Therefore, model (4.4) was estimated with OLS. The consequence of ignoring uncertainty sources 1 and 2 is that prediction intervals in this section are too narrow – although it is unknown by how much.

[18] The *relative* intervals become narrower when age increases.

Figure 4.3. Death probabilities for selected years and corresponding prediction intervals. Men

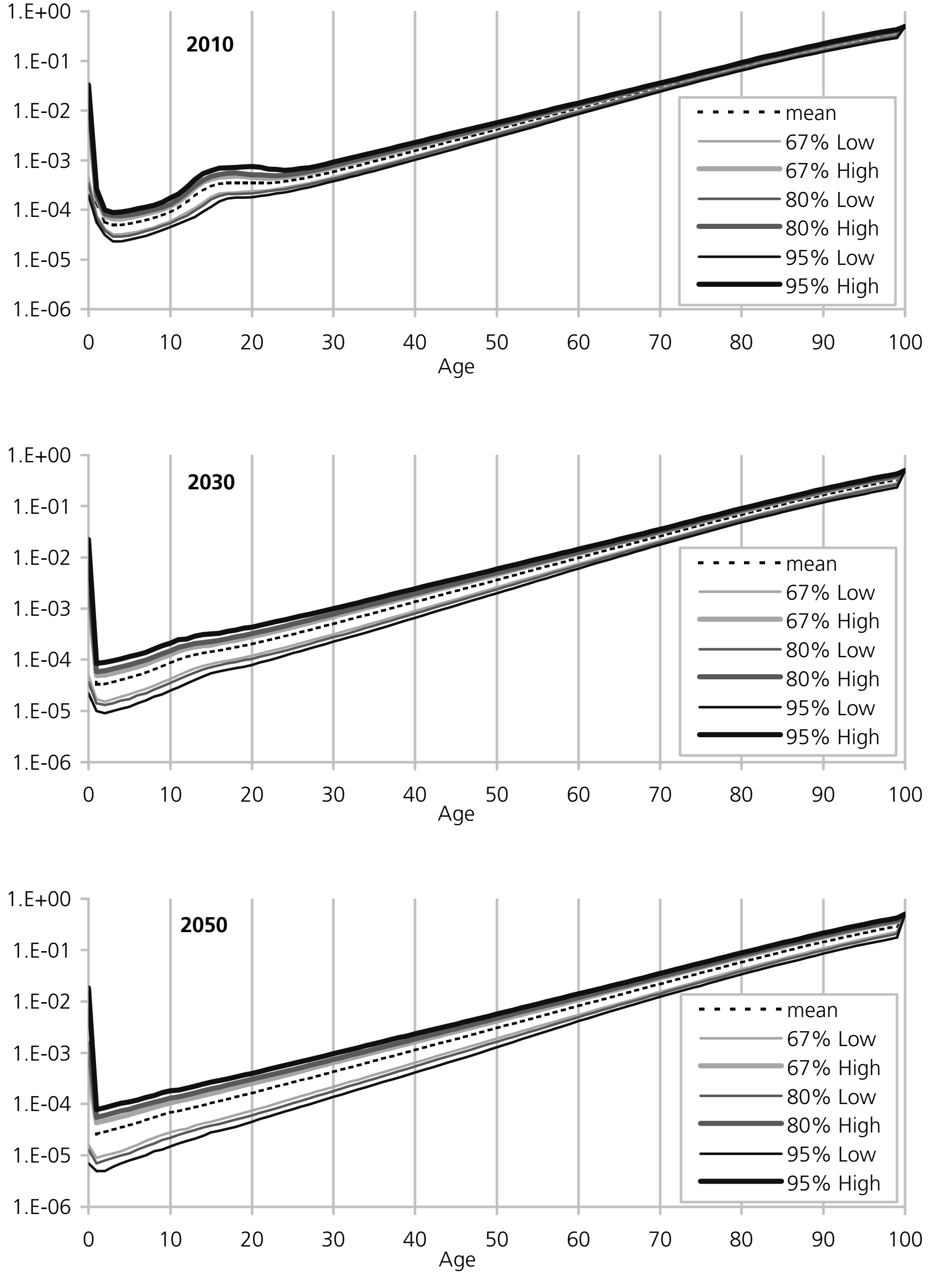
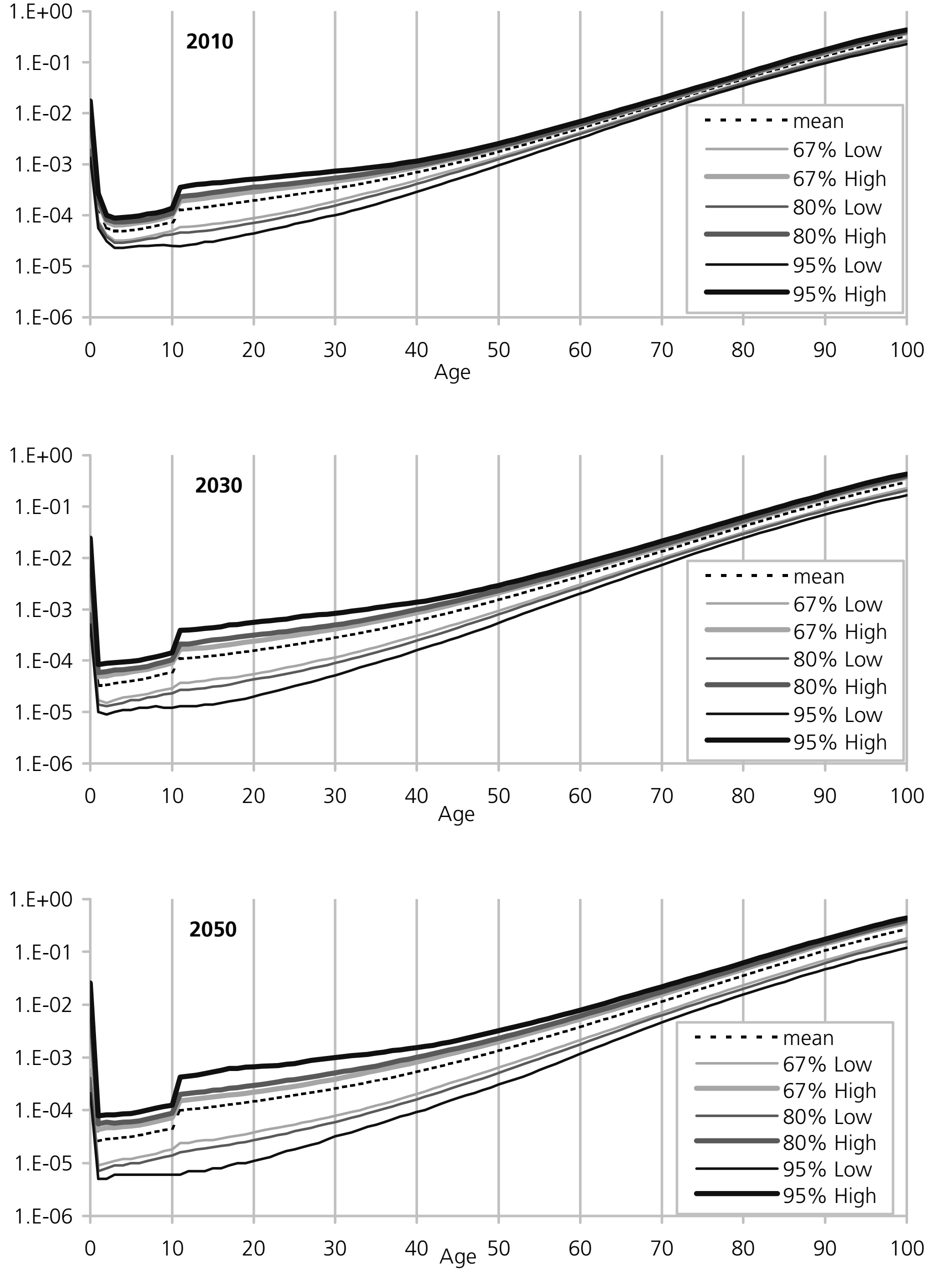


Figure 4.4. Death probabilities for selected years and corresponding prediction intervals. Women



4.6. Comparison with historical errors in the life expectancy

Figures 4.5 and 4.6 summarize empirical errors for extrapolated life expectancy of men and women. Eleven forecasts with base years between 1969 and 1996 were analysed in a way similar to that described in Section 3.1.7 for the TFR. Extrapolated life expectancy values were compared with observed values for the years 1969-1999. The figures show mean and median errors that become increasingly more negative for longer forecast durations. Hence Norwegian forecasters have been too pessimistic, in that they assumed too low life expectancy values in past forecasts. After 15 years, life expectancies were too low by two years on average, both for men and for women. The empirical 67 per cent interval is much wider for women than for men, reflecting the unexpectedly strong improvement in female mortality since 1969. Quite remarkable is the fact that the intervals do not widen with increasing projection length. This is explained by the extrapolation method for life expectancy. For forecasts made between 1969 and 1982, no improvement in mortality was foreseen, and life expectancy was kept constant at its most recently observed value.[19] In reality life expectancy increased more or less regularly. The consequence is that life expectancy errors show a time path for subsequent forecasts, which runs parallel to that of the mean error.

Because of the limited number of empirical errors, the intervals are somewhat irregular. Yet it is not unreasonable to say that at durations of 10-15 years, the 67 per cent interval is approximately 0.7 years wide for men, and roughly 1.8 years for women. Thus prediction intervals in Section 4.2 are wider than historical intervals, and this empirical evidence would imply that one should restrict the life expectancy predictions of Section 4.2 to narrower bounds. On the other hand, these bounds should have been wider, if we would have been able to take estimation errors for the constant vector (K^M,K^F) into account, see Section 4.2. Moreover, bounds around age- and sex-specific death rates should have been wider than what we could obtain, since sample variation in the historical age-specific rates and estimation errors in H-P parameters could not be included in the models. For these reasons we decided to accept the life expectancy predictions of Section 4.2.

[19] For the forecast of 1982, life expectancy was kept constant beginning in 1991.

Figure 4.5. Empirical errors in life expectancy forecasts for men. 1969-1999

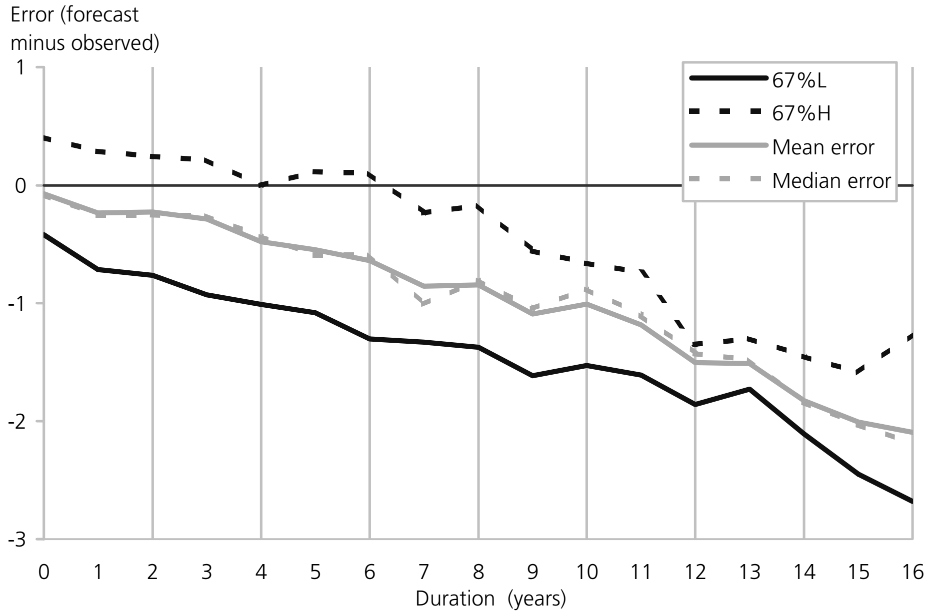
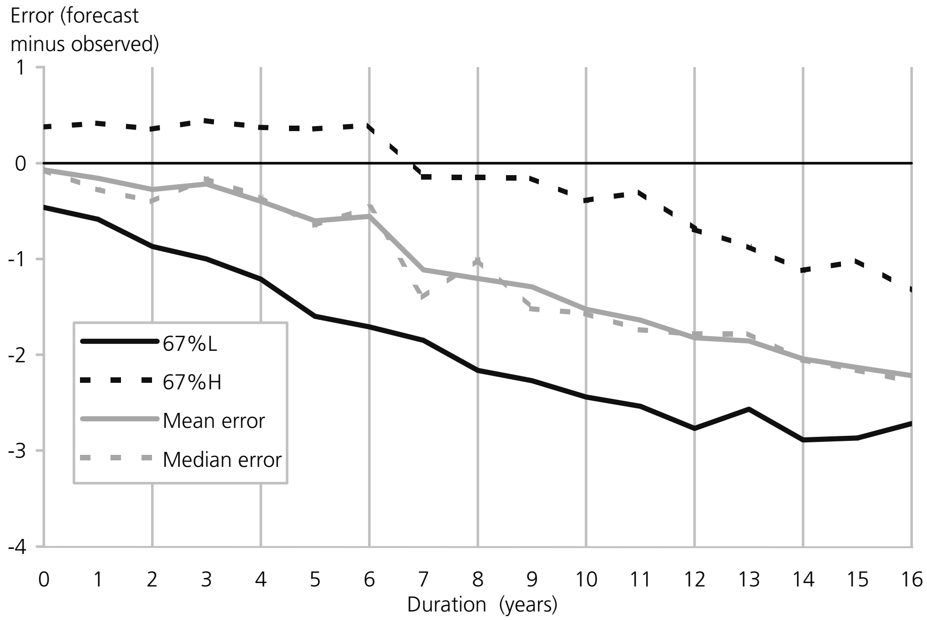


Figure 4.6. Empirical errors in life expectancy forecasts for women. 1969-1999



Appendix 4.1. An adjustment method for the intercept of the multivariate time series model

The purpose of the adjustment is to predict a pre-specified value for the life expectancy at birth in 2050, on the basis of the multivariate ARIMA (2,0,0) model given in expression (4.1).

Rewrite model (4.1) in state-space form as

$$\begin{bmatrix} \ln(e_t^M) \\ \ln(e_t^F) \\ \ln(e_{t-1}^M) \\ \ln(e_{t-1}^F) \end{bmatrix} = \begin{bmatrix} K^M \\ K^F \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_1^M & 0 & \phi_2^M & 0 \\ 0 & \phi_1^F & 0 & \phi_2^F \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \ln(e_{t-1}^M) \\ \ln(e_{t-1}^F) \\ \ln(e_{t-2}^M) \\ \ln(e_{t-2}^F) \end{bmatrix} + \begin{bmatrix} \varepsilon_t^M \\ \varepsilon_t^F \\ 0 \\ 0 \end{bmatrix}$$

or $W_t=C+\Phi W_{t-1}+\varepsilon_t$ for short. Thus W_t is a vector with log-transformed life expectancies by sex at times t and t-1, C is a vector of constants, and Φ is a matrix of coefficients. Taking expected values, and given C and Φ, the l-step ahead forecast for W_t is

$$W_{t+l} = \left(\sum_{j=0}^{l-1} \Phi^j \right) C + \Phi^l W_t \text{, or}$$

$$C = \left(\sum_{j=0}^{l-1} \Phi^j \right)^{-1} \left(W_{t+l} - \Phi^l W_t \right).$$

For the actual computations, we selected t=1995, and l=55. Starting values W_{1995} were known from the life tables. Estimates for the matrix Φ were as in Table 4.2 (omitting the K-estimates). The target values for the life expectancies were 84.5 years for women, and 80 years for men, both to be reached in 2050.

5. Migration

5.1. The approach in general

Of the three components of population change, international migration is the one for which uncertainty is largest, because unforeseen economic, political, and legal developments, both at the national and the international level have a strong effect on this component. Few attempts have been reported to model immigration and emigration as a stochastic process. Alho and Spencer (1995) propose to model the error of age-specific migration by defining correction factors that modify age-specific survival rates, to account for migration. They distinguish between immigration and emigration, and discuss the correlation between emigration errors and immigration errors. De Beer (1997) reports on the experiences with time series models for immigration and total emigration to and from the Netherlands over the period 1960-1994. Both for immigration and for emigration, an ARIMA (1,0,0)-model was fitted; net-migration was best described by an ARIMA (0,0,1)-model. The long-term predictions for the fitted models for immigration and emigration corresponded rather closely with the long-term net migration prediction. The net migration model was quite stable when fitted to the three periods 1960-1980, 1960-1985, and 1960-1990. Alho (1998) fitted an ARIMA (0,1,1) model to the logarithms of immigration and emigration for Finland, 1945-1995.

Numbers of migrations to and from Norway are modest relative to the resident population, even in those age classes where migration is most frequent. For this reason, and because of the largely unpredictable nature of migration, we have adopted an extremely simple approach for the computation of prediction intervals for immigration and emigration. Separate procedures were adopted for the level and the age pattern of migration. First, two time series models were estimated on the basis of gross flow data for the period 1958-1997, one for immigration and one for emigration. These were adjusted in such a way that they predicted the same migration flows as those used by Statistics Norway in its latest official population forecast. Next, both flows were broken down by sex, using male/female shares as observed in recent years. Finally, each of the four flows was broken down into one-year age groups on the basis of age-specific shares. The latter shares were found by means of simple extrapolations of the parameters of Rogers-Castro age schedules as estimated for the years 1967-1997. Observed errors in net migration for historical forecasts do not show any sensible pattern. Therefore these errors are not shown here.

5.2. Immigration and emigration flows for men and women

Data on gross flows by sex are available from 1958 onwards. We constructed four time series of data, i.e. two flows for each sex. Initially, we assumed that family migration and return migration should be reflected as lagged and unlagged cross-correlations between the four series. However, no systematic patterns could be discovered. The strongest meaningful cross-correlations were

- Emigration in a certain year and immigration one to three years later: correlations of 0.7-0.8 for men, and 0.6-0.7 for women (migration connected to work and education between Norway and neighbouring countries).
- Immigration of men and immigration of women one to three years later: correlations of between 0.75 and 0.85 (family reunification).

Other cross-correlations were low (immigration followed by emigration; emigration of one sex followed by emigration of the other sex) or could not be given any meaningful interpretation.

The unclear correlation patterns are explained by the fact that the character of migration to and from Norway has changed considerably over the past decades. Although migration to and from other countries in Western Europe and Northern America has always been a stable component in Norwegian migration flows, the remaining part of migration has been much more volatile. The 1960s and early 1970s saw immigration of unskilled workers from less developed countries, particularly from Pakistan. During the second half of the 1970s and the first half of the 1980s, family reunification of these labour immigrants was important. It was also the time when the demand for skilled labour in the oil industry increased. The second half of the 1980s saw several short-term flows of refugees and asylum seekers, in particular from Iran and Chile (1987-1988) and Sri Lanka (since 1987). After 1987, immigration from less developed countries decreased steadily. The war in the former Yugoslavia led to many refugees and asylum seekers in 1993. The fact that the reasons for migration and countries of origin and destination changed rather frequently during the past 40 years means that a stable correlation pattern for the whole period is improbable. Obviously, migration data broken down by reason for migration (labour migration, family reunification, refugee) and country of origin or destination would probably reveal more stable correlation patterns, but such data are not available in sufficient detail (broken down by age and sex) for a long enough period. Note that Alho (1998), in his analysis of Finnish migration data for the period 1945-1995, found no clear cross-correlation between immigration and emigration either.

We have assumed that immigration and emigration are independent. Given the weak empirical correlation between the two flows, this assumption means that the uncertainty in total net migration is a little too high in our simulations. The following models were estimated.

Table 5.1. Estimates of migration model parameters. Estimated standard errors in parentheses

	Immigration		Emigration	
C	0.533	(0.361)	0.013	(0.016)
ϕ	0.948	(0.037)	-	-
θ	0.304	(0.165)	-	-
ε	0[1]	(0.095)	0[1]	(0.099)

[1] Assumed value.

(5.1) Immigration: $\ln(i_t) = C + \phi\ln(i_{t-1}) + \varepsilon_t - \theta\varepsilon_{t-1}$

(5.2) Emigration: $\ln(u_t) = C + \ln(u_{t-1}) + \varepsilon_t$

Estimates are given in Table 5.1.

Both for immigration and for emigration, the constant *C* was adjusted in such a way that predicted levels of i_t and u_t corresponded to values assumed by Statistics Norway in its official population forecast.[20] The two models were used for predicting total immigration and emigration for the period 1995-2050. Due to the rapid growth in the prediction intervals, it was decided to keep all predictions constant after five years, see Figures 5.1 and 5.2.

Predicted numbers of immigrations and emigrations for each year were broken down by sex on the basis of corresponding sex shares as observed on average for the period 1967-1997. For men the average shares were 51.3 per cent for immigration, and 51.0 per cent for emigration. Corresponding standard deviations were 1.99 and 1.74 per cent. In the late 1950s and in the first half of the 1960s, more women than men migrated to and from Norway each year. These women had relatively often Norwegian nationality, and travelled abroad for reasons of education. In the 1970s and later, labour migration, family reunification and refugees dominated the migration pattern, with a higher share of men than women. We assume that the latter pattern will continue. For each year and each flow, male and female shares were drawn on the basis of a binomial distribution (approximated by a normal distribution), with expected values and standard deviations as given above.

[20] In the spirit of Appendix 4.1, one can show that for the ARMA(1,1)-process of expression (5.1), the following holds: $C = \left(\hat{Z}_{1995}(l) - \hat{\phi}^l Z_{1995} + \hat{\theta}\hat{\phi}^{l-1}\hat{\varepsilon}_{1995}\right) \Big/ \sum_{j=0}^{l-1}\hat{\phi}^l$, where $Z_t = \ln(i_t)$, $Z_{1995}(l)$ is the l-step ahead 1995-based prediction of Z_t, and a hat denotes an observed or estimated value. In the calculation of C, l was chosen equal to five years.

Figure 5.1. Immigration, observed and predicted

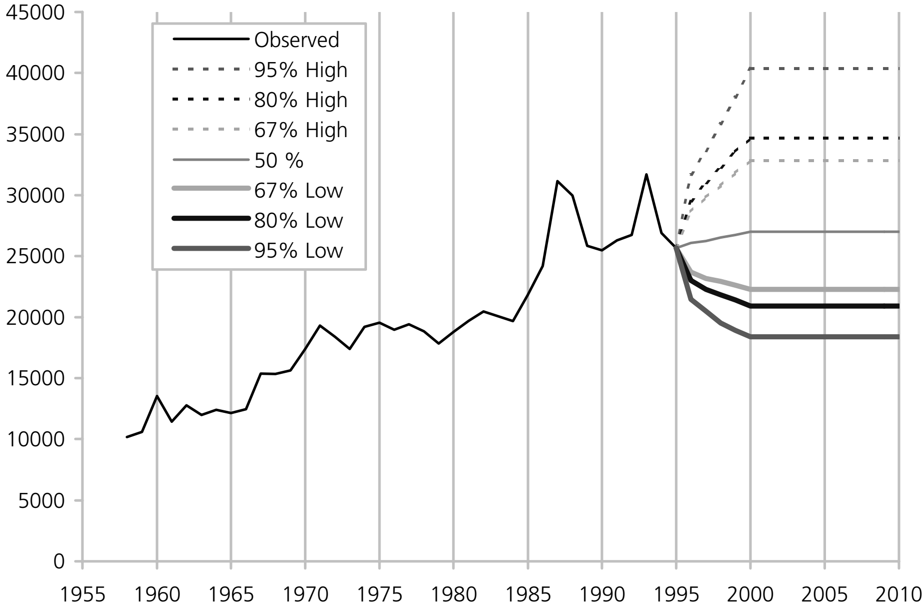
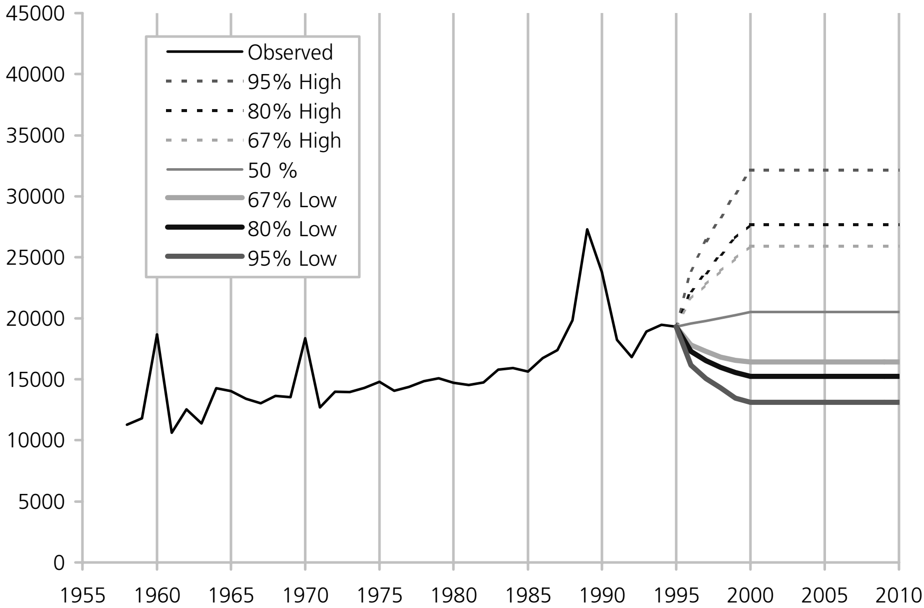


Figure 5.2. Emigration, observed and predicted



5.3. Age patterns of immigration and emigration

The age patterns of immigration and emigration for men and women have been treated in a similar manner. Below we will report the method used for immigration of men.

Let $i^m(x)$ denote the share of male immigration at age x in total male immigration in a certain year (the time index is suppressed here). The following double exponential model, which is due to A. Rogers and L. Castro (1986), has been assumed for the age pattern:

$$i^m(x) = a_1 e^{-\alpha_1 x} + a_2 e^{-\alpha_2 (x-\mu_2) - e^{-\lambda_2 (x-\mu_2)}}, \; x=2, 3, \ldots, 80$$

The two terms model migration of young children (parameters a_1 and α_1) and of adults (parameters a_2, α_2, μ_2, and λ_2). Retirement migration, which is included as a third term in the Rogers-Castro (R-C) model, is hardly or not visible in the Norwegian data, and hence we ignored it. By omitting ages 0 and 1 from the estimation, we achieved a good fit for ages over 50. The way we treated ages 0 and 1 is described below. We used non-linear weighted least squares regression to estimate the six parameters. In order to obtain a satisfactory fit for the prime migration ages, weights equal to 100 were used for ages 26 through 32, whereas weights for other ages were set to one.

The six parameters have been estimated for the period 1967-1997, and a simple extrapolation method has been used to predict their values for the years until 2050. Based on visual inspection, we decided to use exponential decline for decreasing parameters, and a constant pattern for parameters that showed no clear time trend. This generated predictions for $i^m(x)$ for ages 2-80. Inspection of the age pattern for future years still revealed the typical bi-modal shape that characterizes migration, although the pattern was slightly different from that in recent years. Each share was adjusted proportionally, so as to ensure that the sum over all ages between 2 and 80 would be equal to 0.96, viz. the average sum in the base period. For ages $x>80$, we assumed that $i^m(x)=0$. For ages zero and one, we predicted the shares in the following way:

$$i^m(x,t) = i^m(2,t).i^m(x,1995)/i^m(2,1995), \quad x=0,1; \; t=1995, 1996, \ldots, 2050.$$

Thus shares at ages zero and one were assumed to grow or decrease with the same rate as that for age 2. A final proportional adjustment guaranteed that the shares for all ages 0-80 add up to one.

6. Prediction intervals for Norway's future population

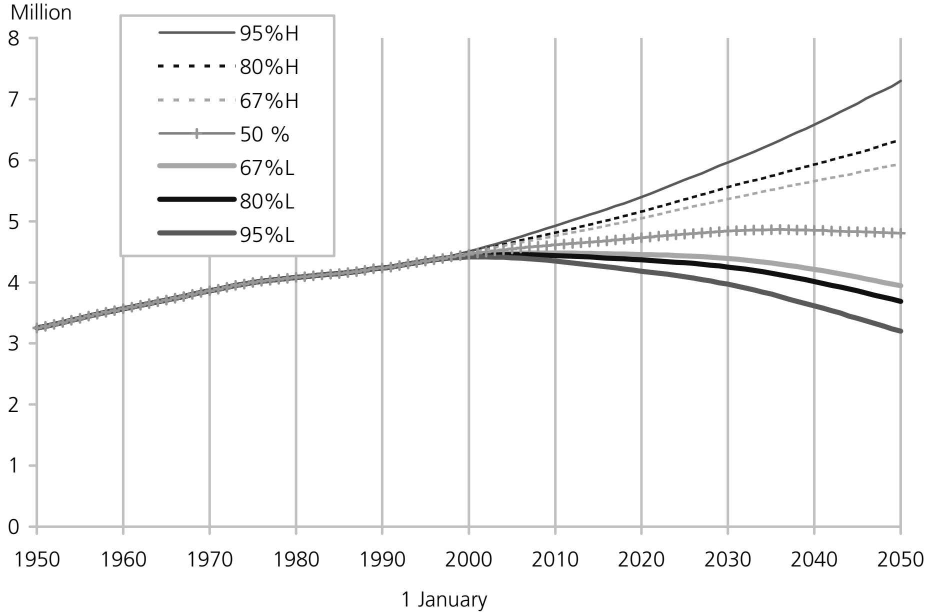
This chapter presents results of the 5 000 simulations for the cohort component model as described in the previous chapters. To summarize: the simulations were based on the following assumptions:
- future expected values for the TFR, the life expectancy, and numbers of immigrants and emigrants are in conformity with the medium variant of Statistics Norway’s 1999-based population forecast;
- prediction intervals for the TFR, the mean age at childbearing, the variance in that age, and the numbers of immigrants and emigrants have been restricted;
- for each component of change, uncertainty in time series coefficients has been taken into account, in addition to residual variance.

Because of the restrictions mentioned above, a total of approximately 8 000 simulations had to be carried out, of which some 3 000 were rejected. The simulations covered the years 1996-2050.

6.1. Total population

Figure 6.1 shows prediction intervals for total population size. We have selected an expected forecast accuracy of 67 per cent (odds two against one), 80 per cent (odds four against one) and 95 per cent (odds 19 against one). Note how fast uncertainty increases when we look further into the future: the intervals widen rapidly. The odds are two against one that the Norwegian population will number between 3.9 and 6 million in 2050. Compared to the median forecast (50 per cent) of 4.8 million in 2050, this two-thirds prediction interval is 43 per cent wide. There is a clear trade-off between greater accuracy (larger odds) and higher precision (narrower intervals). For instance, odds of 19 against one (95 per cent probability) produce an interval between 3.2 and 7.3 million in 2050. This interval is twice as wide as the 67 per cent interval: 88 per cent, compared to the median forecast.

Figure 6.1. Population size



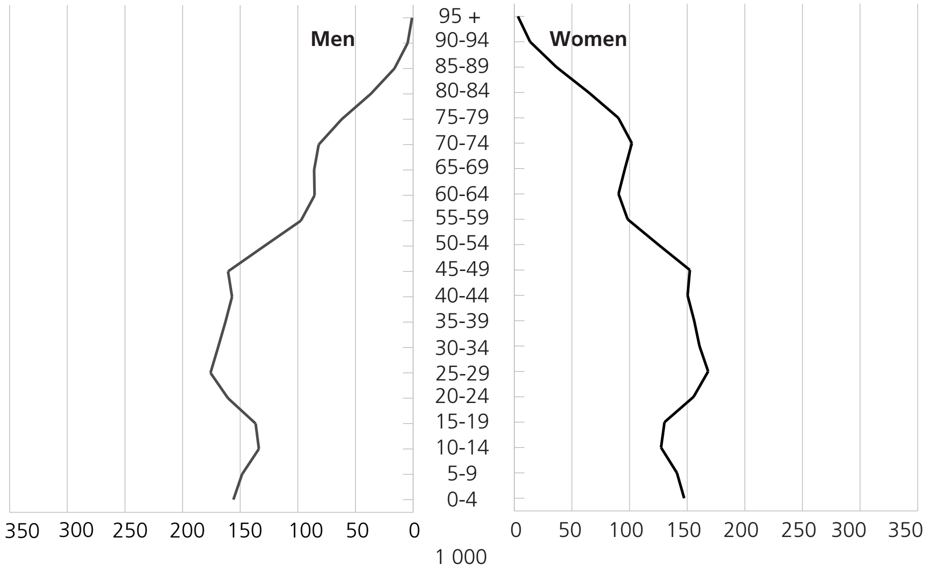
Statistics Norway’s two most recent population forecasts have base years 1996 (Statistics Norway 1997) and 1999 (Statistics Norway 1999). How do the results of those official forecasts compare to our stochastic simulations? The 1996 forecast predicted between 4.3 and 6.0 million inhabitants in Norway in 2050. According to our simulations, the expected probability that this will be the case is less than 56 per cent. The expected accuracy of the high-low interval of the official forecast is somewhat larger in the first forecast years: 77 per cent in 2000, and 66 per cent in 2010. The 1999 forecast had wider fertility margins than the previous one, and predicted between 4.2 and 6.3 million inhabitants in 2050. This resulted in a slightly higher expected accuracy in 2050 (62 per cent). Thus, the expected accuracy of the official forecast is somewhat below two-thirds on the long run, and a little above that level on the short run. The same conclusion was reached earlier on the basis of an entirely different method, namely the extrapolation of observed forecast errors in old population forecasts (Statistics Norway 1994, 33).

In case we would have assumed that time series coefficients for fertility, life expectancy, and numbers of immigrants and emigrants were given, instead of estimated, only residual variance would have been taken into account, and prediction intervals would have been narrower. A new set of 5 000 simulations with only residual variance resulted in a 95 per cent prediction interval for total population size in 2050 equal to (3.3, 7.2) million. The interval is 162 000 persons narrower than the corresponding interval in Figure 6.1.

6.2. Age pyramids

Figures 6.2-6.5 present age pyramids for the years 1996, 2010, 2030, and 2050. They show very clearly the age pattern of uncertainty: in an absolute sense, prediction intervals are wide for young age groups, and narrow for the elderly[21]. This reflects the fact that fertility and mortality have very different impact on the age structure, with international migration taking an intermediate position. Note also that the interval for the age group 0-4 widens rapidly between 2010 and 2030, because most of the parents of the youngest age group in 2030 themselves were born during the forecast period, which increases uncertainty. As a result, intervals in 2030 under the age of 20 are so wide, that the forecast is not very informative. In 2050 this is the case for virtually all age groups, in particular when the intervals are judged in a relative sense (compared to the median forecast).

Figure 6.2. Observed population by age and sex. 1996



[21] *Relative* intervals, that is intervals as a ratio of the median forecast, show a different pattern, see below.

Figure 6.3. Simulated population. 2010[1]

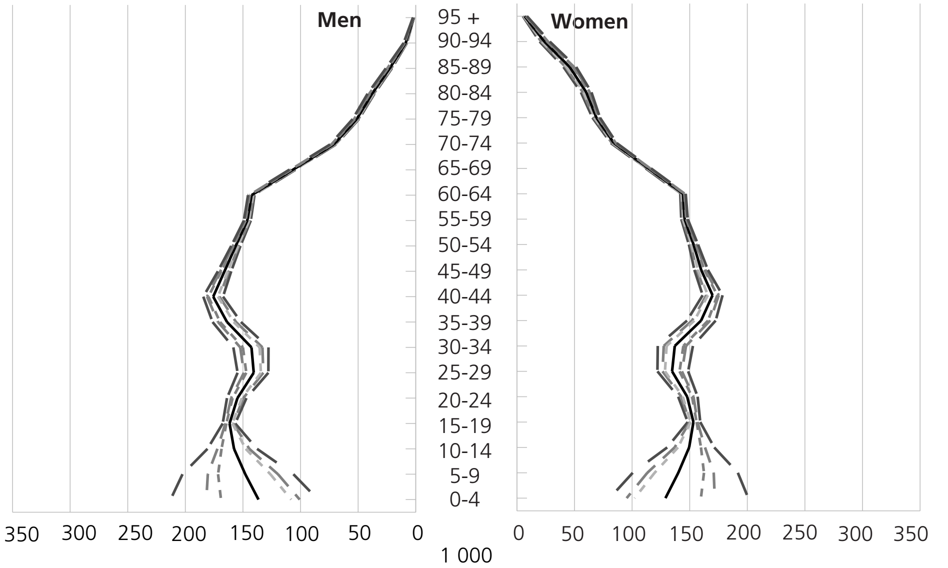
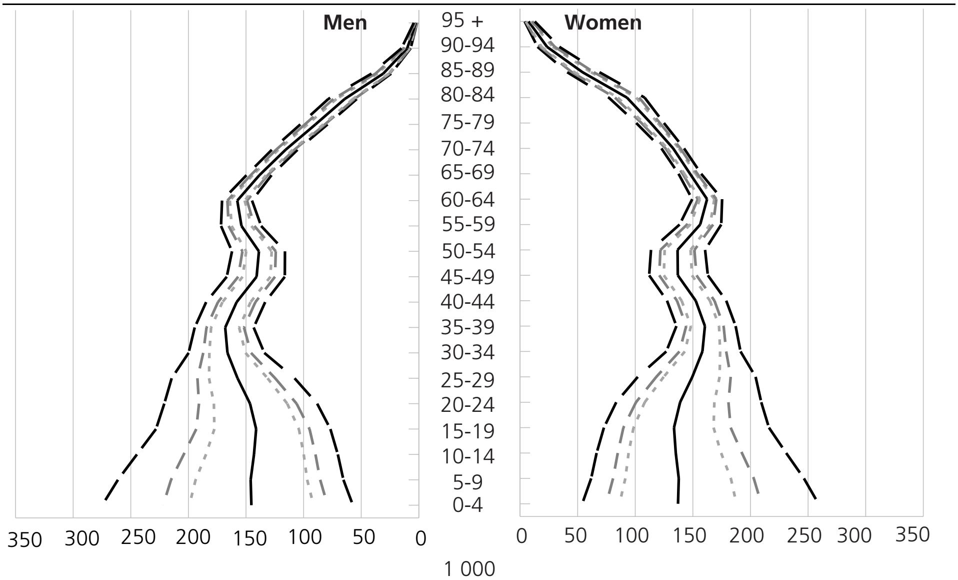
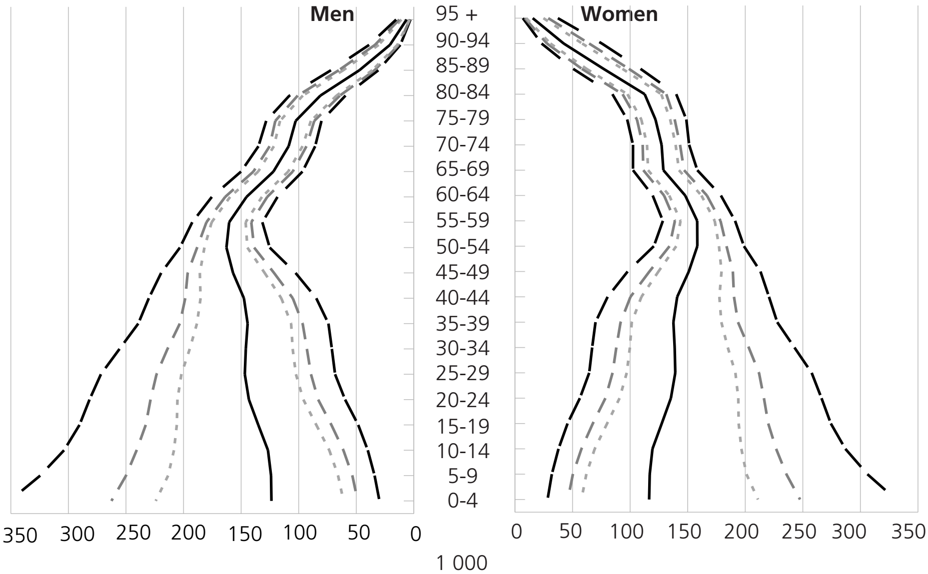


Figure 6.4. Simulated population. 2030[1]



[1] The solid line represents the median forecast. It is surrounded by (from inside to outside) the 67 per cent interval, the 80 per cent interval, and the 95 per cent interval.

Figure 6.5. Simulated population. 2050[1]



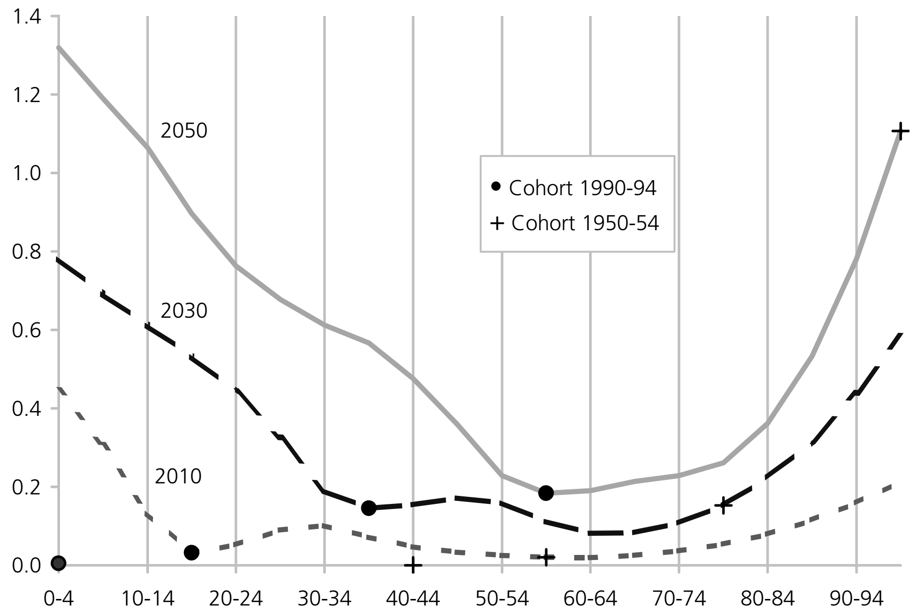
[1] The solid line represents the median forecast. It is surrounded by (from inside to outside) the 67 per cent interval, the 80 per cent interval, and the 95 per cent interval.

Figure 6.6 illustrates how expected forecast errors propagate through the age structure. It gives the width of the two-thirds prediction interval for men in 2010, 2030, and 2050, relative to the median forecast[22]. The figure demonstrates that relative uncertainty due to fertility is largest, even larger than that for the oldest old, caused by mortality. With increasing forecast duration, the uncertainty in age groups between 20 and 45, which is caused by international migration, drowns in fertility and mortality uncertainty.

The values for birth cohorts 1950-1954 and 1990-1994 are marked. The latter men, who were born immediately before the forecast's base year, have lowest uncertainty among all cohorts in the forecast. Those who were born earlier are confronted with larger uncertainty due to international migration (for example birth cohort 1975-1980, who were aged 30-34 in 2010), or due to high mortality (for instance cohort 1950-1954). The interval for this latter cohort grows exponentially over the cohort's life course. Men who were born after 1996 have also relatively wide prediction intervals, due to uncertainty in fertility. Note, for example, men born in the years 2005-2009, who were aged 0-4 in 2010 and 40-44 in 2050.

[22] The pattern for women is similar, the largest differences occurring for elderly women. For instance, at age 95+, women have a relative 67 per cent prediction interval that is 0.27, 0.61, and 0.97 per cent times the corresponding median value in 2010, 2030, and 2050. See also Figure 6.8 for the year 2010.

Figure 6.6. Relative width of 67 per cent prediction interval. Men



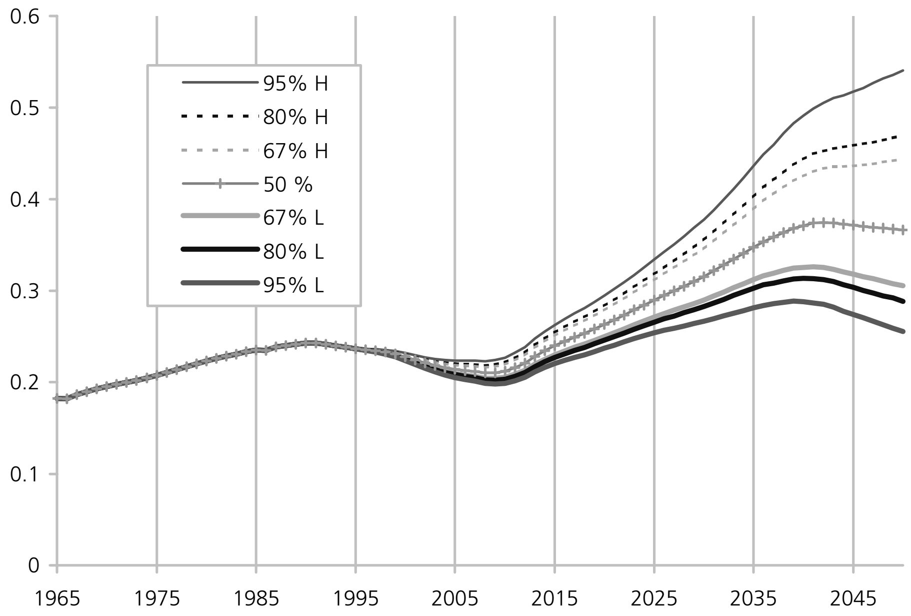
6.3. Dependency ratios

Ageing is certain, at least until around 2040. In that year, the odds are two against one that the old age dependency ratio (OADR), defined as the population 67[23] and over relative to that aged 20-66, will be between 0.33 and 0.43, i.e. at least 10 points higher than today's value of 0.23, see Figure 6.7. When we state that ageing is certain, that is because the probability of a ratio in 2040 which is lower than today's, is close to zero.

Statistics Norway has assumed high and low trajectories both for future fertility, mortality, and net immigration. The combination of high fertility with high life expectancy (i.e. low mortality), and high immigration defines the so-called high population growth variant. The low population growth variant combines low fertility with low life expectancy and low immigration. Now consider the OADR in Statistics Norway's official population forecast. We will use this variable to illustrate that the traditional way of computing high, medium and low variants in a deterministic cohort-component forecast gives a false impression of uncertainty. In Chapter 1 we noted that the high-low intervals for the numerator and the denominator of the OADR in the official forecast constitute 31 per cent and 32 per cent of the value according to the medium variant. At the same time, the OADR-range is only 1.3 per cent of the OADR-value in the medium variant. This

[23] The legal retirement age in Norway is 67.

Figure 6.7. Old age dependency ratio



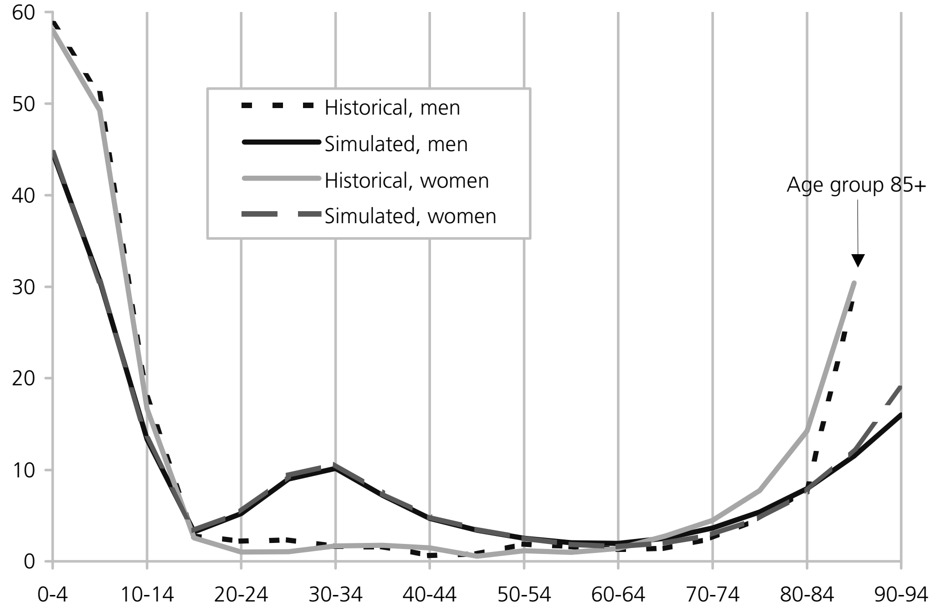
narrow interval suggests much less uncertainty for the OADR than a true stochastic forecast does. For instance, Figure 6.7 shows relative prediction intervals in 2050 equal to 38 (67 per cent), 49 (80 per cent), and 78 (95 per cent) per cent wide, compared to the median OADR forecast. The reason for too narrow intervals in the official forecast is that perfect correlation is assumed, both across components and over time. This unrealistic assumption implies that traditionally computed forecast variants cannot be given a statistical interpretation.

The 95 per cent prediction interval for the OADR stretches from 0.255 to 0.540 in 2050 - it is 0.285 wide. The corresponding interval that results when only residual variance is acknowledged, and time series coefficients are assumed given, is 0.257 wide.

6.4. Comparison with historical errors in the age structure

How do the prediction intervals around the age structure compare with empirical intervals computed on the basis of historical forecasts? We have assembled data on observed forecast errors in the age structure (five-year age groups, by sex) in the forecasts published by Statistics Norway with base years between 1969 and 1985, extending Texmon's database mentioned in Section 3.1.6. We have restricted ourselves to a forecast duration of 15 years. Forecast variants were

Figure 6.8. Relative width of 67 per cent interval. Forecast duration = 15 years



given equal weight. The forecast value for each age group was compared to its actual value, and the percentage error was computed.[24] For each age group up to 80-84 we had fifteen error values, for age group 85 and over 12. Errors were sorted by size, and bounds of the two-thirds empirical interval were obtained by interpolation, if necessary. Figure 6.8 plots the width of the two-thirds intervals for men and women, and compares these historical intervals with the simulated prediction intervals for the year 2010.

There is rather close agreement between the historical pattern and the simulated one. In both cases, errors are high for young ages, and moderate for the elderly. Historical variation is relatively large for ages 0-9 and 80 and over. For the youngest age groups this is explained by the sharp fall in the TFR in the 1970s and the modest increase at the end of the 1980s. Historical variation is relatively large for the elderly because of the strong increase in survival chances since 1970, in particular for women. In the simulated fertility and mortality trajectories up to 2010, such sudden developments are less likely (but not excluded). For ages 25-39, prediction intervals are wider than one would expect

[24] For the forecasts with base years 1977 and 1979, not all results were published for five-year age groups. Deviant age groups were 0-6, 7-12, 13-15, 16-19, 60-63, 64-66, and 67-69. Forecast results for these age groups were transformed into five-year results by means of linear interpolation based upon observed population data for one-year age groups.

on the basis of historical forecast errors. This reflects the fact that the variation in year-to-year immigration has increased since the beginning of the 1990s, see Figure 5.1. It is unlikely that the smooth developments that were observed until the mid-1980s will repeat themselves in the period up to 2010.

This empirical comparison suggests that prediction intervals for young and old ages are plausible, while the relatively wide intervals for the intermediate age groups can be justified.

7. The use of stochastic population forecasts

Population forecasts in general are widely used in various planning situations, such as schooling, health care, and pension systems. In the short run, the uncertainty expressed by stochastic forecasts is relatively limited[25], and a user who has a five-year planning horizon, say, may safely use a deterministic forecast computed in the traditional way. In such cases, the point forecast (expected value) is likely to be of more interest than the prediction intervals. But in the long run, the expected error in the forecast results becomes increasingly important, and planners who are interested in the age structure of the population 30 years or more into the future, should take uncertainty seriously. We shall give two examples, one from health care, and the other one from old age pensions.

In a recent analysis of the future demand for hospital beds in Norway, the authors concluded that in 2050, the population will need between 2.7 and 4.0 million person days in hospital annually (Paulsen et al., 1999). This implies an increase by 45-79 per cent compared with the current number of 1.9 million person days in hospital per year. The results were obtained on the basis of the low and the high population forecast of Statistics Norway, combined with certain assumptions regarding health parameters such as the mean duration of hospital stay for future patients, and the mean number of patients compared to the whole population. Many of these parameters are age-specific, and the increase in the demand for hospital beds in this analysis is entirely a consequence of population ageing (the authors assumed a slight decrease in the average duration in hospital per patient). The demand interval in 2050 is 1.3 million person days, or 38 per cent of the medium forecast. The majority of the patients is 65 years or older. For this age group, two-thirds prediction intervals increase sharply with age, from 21 per cent for the 65-69s, to 26-36 per cent for the 80-84s, and 65-78 per cent for the 90-94s, see Figure 6.6. This suggests that the accuracy of the 2.7-4.0

[25] For some very specific forecast results, such as the number of centenarians, uncertainty may be considerable also on the short run. The same holds true for stochastic forecast results at a low regional level, for instance municipalities.

million forecast is probably lower than two-thirds, and the Paulsen et al. projection results should be treated with great caution[26]. Clearly, a fully stochastic forecast of the future demand for hospital beds should also take into account the fact that the relevant medical parameters are stochastic. The covariance structure of these parameters, and possible correlations with population variables has yet to be explored.

Fredriksen (1998) describes a microsimulation model that is able to simulate, among others, future public pension benefits and contributions of the Norwegian population. Norwegian public pensions are of the "Pay as you go" (PAYG)-type. In one set of simulations starting in 1993, the author shows that the contribution rate will rise from 16 to 23 per cent in 2030. The simulations are based upon a number of demographic and non-demographic (labour market participation, disability, earnings, retirement) assumptions. Next he analyses the consequences for the contribution rate of various alternative assumptions. The rate turns out to be between 21 and 25 per cent in 2030, depending on high or low disability risks in the population, or high or low labour force participation among women. The 4 percentage points difference implies a relative interval of 16 per cent of the reference value for the contribution rate. Because of the PAYG nature of the pensions, it is relevant to inspect the Old Age Dependency Ratio[27]. Then we see (Figure 6.7) that the OADR has a relative two-thirds prediction interval of 18 per cent in 2030 (the 80 per cent and 95 per cent intervals are 23 per cent and 35 per cent wide, respectively). This means that the bandwidth defined by possible policy options (denoted as “reasonable” by Fredriksen) is smaller than the uncertainty implied by the demographics alone. In other words, the effects of labour market changes and disability changes will likely drown in inherent population uncertainty at the horizon 2030[28].

These two examples show how important it is to take population forecast uncertainty seriously. However, an often-heard objection is that projections of the kind mentioned here are purely conditional “what-if” calculations, and that it is unnecessary to consider expected forecast errors. “*If* the population would

[26] For 2010 and 2030, the authors obtain intervals of 13 and 25 per cent of the medium forecast, respectively. Thus the general conclusion of low accuracy is not substantially altered on the short and medium term.

[27] In case fund based pension systems are considered, life course uncertainty of the type illustrated in Figure 6.6 would be relevant.

[28] Fredriksen investigates also “packages” of alternative assumptions. For instance, in an “ageing alternative” he combines low population growth with high disability risks. A “growth alternative” combines high population growth with low disability risks and high female labour force participation. The contribution rate varies between 16 and 30 per cent in 2030, when either of these packages is chosen. This implies a relative variation of 58 per cent, much larger than the prediction interval for the OADR. In this case the sensitivity analysis has clear policy relevance. One may object, however, that the implicit assumption of perfect correlation between the components (in the ageing alternative, *each* year when fertility is high, *also* disability risks are high, *and* mortality *and* immigration are low) is not very realistic.

show this or that trend, *what* would the consequences be for the health care system?" Any deviations from the assumed demographic path are not of primary importance. We do not agree with this objection. Since population size is, in practice, a continuous variable, the probability that the population will follow this or that trend, is zero. This reduces the policy relevance of many deterministic planning studies, unless one can demonstrate that demographic uncertainty is negligible. In all other cases one is forced to think in terms not of point forecasts, but of intervals - and intervals imply expected accuracy and statistical analyses. Therefore the traditional approach based on deterministic variant projections is inconsistent from a statistical point of view, as demonstrated by the intervals for the OADR in Section 6.3.

The format in which stochastic population forecasts are made available to the users is very different from that employed for traditional deterministic forecasts. The latter type of results can be included in tables in printed reports. If results for one-year or five-year age groups are published, the user who is interested in a specific larger age group can find the corresponding number by simply adding the age-specific results. A stochastic forecast, however, is presented in the form of predictive distributions. *Each forecast result has its own distribution.* Thus, the bounds of the prediction interval of a larger age group are generally *not* equal to the sum of the bounds of the constituent ages. For instance, the lower and upper bounds of the 95 per cent interval of total population in 2050 are 3.20 and 7.29 million, respectively, see Figure 6.1. On the other hand, the sums of the corresponding bounds for men and women broken down in five-year age groups in Figure 6.5 are 2.77 and 8.00 million. The former interval [3.20, 7.29] is the correct one. It is narrower than the results obtained by simple addition, because the results for the various age groups are correlated. At time t, age groups x_1 and x_2 are correlated, for three reasons. First, births occurring in years $(t\text{-}x_1)$ and $(t\text{-}x_2)$ are correlated. The time series model for fertility in Chapter 3 takes account of the fact that when fertility is high in a certain year, it is also likely high the year thereafter. Although the correlation between calendar years that are further apart is weaker than the one-year correlation, it is not entirely zero. In other words, nearby age groups are strongly correlated, ages further apart only weakly. The second reason is related to the chances of survival for the two birth cohorts $(t\text{-}x_1)$ and $(t\text{-}x_2)$ until time t. These survival chances are correlated, again due to the time series character of the mortality model. Finally, immigration for the two birth cohorts is correlated. Thus, when the prediction interval of two or more age groups is to be determined, correlation across ages should be taken into account. In traditional forecasts, one assumes perfect correlation across time for the components of change. For instance, in the high variant, fertility is high in the first forecast year, and it is also high in all the other years. Because of this perfect correlation, the sum of age-specific results for the high variant coincides with the high variant result for a sum of age groups, and similarly for the low variant. But equivalent high-low variants for age groups do not define bounds that have equal probability coverage, because the variants assume perfect correlation. In stochastic forecasts, the correlation is less than perfect and simple addition of prediction bounds does not give the correct bounds for the aggregate. (Expected values may be added, of course.)

The consequence is that stochastic forecasts should be made available to the user in the form of a database (Alho 1990), from which the user can construct the prediction intervals of any age group he is interested in. An example is the database of Statistics Netherlands, which can be accessed from the Internet, see http://statline.cbs.nl/statweb/index_NL.stm.

Finally, one should keep in mind that all the results presented here are computed on the basis of a number of linked stochastic models, with specific assumptions concerning the prediction intervals for future fertility, mortality, and migration. The expected values of the distributions coincide with those assumed by Statistics Norway in its official population forecast. But the intervals may be larger or smaller in other applications than the current one, depending on the particular models and assumptions used.

8. Conclusions

The demographic future of any population is uncertain, but some of the many possible trajectories are more probable than others. Therefore, an exploration of the demographic future should include two elements: a *range* of possible outcomes, and a *probability* attached to that range. Together, these two constitute a prediction interval for the population variable concerned. This report gives the results of a research project of which the aim was to compute prediction intervals for the future population of Norway broken down by age and sex to the horizon 2050.

Chapter 6 contains the main results. We estimate that the odds are 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 3.7-6.3 million in 2050. This illustrates that uncertainty increases with time. There is a clear trade-off between greater accuracy (larger odds) and higher precision (narrower intervals). Odds of 19 against one (95 per cent chance) result in a wider interval: 4.1-5.7 million in 2025, and 3.2-7.3 million in 2050. The probabilistic population forecasts of the youngest and the oldest age groups show largest uncertainty, because fertility and mortality are hard to predict. As a result, prediction intervals in 2030 under the age of 20 are so wide, that the forecast is not very informative. In 2050 this is the case for virtually all age groups, in particular when the intervals are judged in a relative sense (compared to the median forecast). The expected accuracy of the total population size forecast published by Statistics Norway is somewhat below two-thirds on the long run, and a little above that level on the short run.

Prediction intervals for different population variables cannot simply be combined to find the interval for an aggregated variable. For instance, the lower limit of the 67 per cent prediction interval for the total population in a certain future year is *not* equal to the sums of the corresponding lower limits for men and women. The reason is that predicted numbers of men and women are correlated, caused by the assumed correlation between male and female mortality, and that between male and female migration. The consequence is that each separate forecast result has its own prediction interval, which cannot be computed on the basis of intervals for more detailed results.

The results in Chapter 6 have been obtained on the basis of stochastic simulation of each of the three components of population change, viz. fertility, mortality, and international migration. Simulation of the components relied heavily on three complementary methods:

- time series analysis for key demographic indicators, such as the TFR, the life expectancy, and numbers of immigrants and emigrants;
- an analysis of historical forecast errors, assembled on the basis of forecasts produced by Statistics Norway since 1969;
- and finally expert judgement, for instance to restrict the prediction interval for the TFR or that for the numbers of immigrants and emigrants to reasonable ranges.

Much emphasis has been given in this study to time series analysis, because the underlying statistical model facilitates an appropriate treatment of forecast uncertainty. For fertility and migration, however, the time series resulted in unrealistically wide prediction intervals. This is the reason why the intervals had to be narrowed in a subjective manner. Historical forecast errors for the TFR, the life expectancy at birth, and the age structure for men and women have been used to check the plausibility of the prediction intervals' width.

The predictions for mortality and international migration were calibrated in such a way that the median coincided with the Medium Variant value of the 1999-based official population forecast of Statistics Norway. Our time-series predictions for the TFR agreed closely with the official ones.

The time-series predictions indicated that assumptions on future TFR as employed by Statistics Norway in its official population forecasts have estimated coverage probabilities of only 46, 31, and 24 per cent in the years 2010, 2030, and 2050. The official mortality assumptions have higher expected accuracy in 2050 (just over 60 per cent for the life expectancy), but lower values in the beginning of this century (around 35 per cent in the period 2000-2010).

The prediction intervals presented in this report have been computed on the basis of certain assumptions about the stochastic nature of fertility, mortality, and international migration. Although these assumptions appear reasonable, it is not at all certain that they are entirely correct. Different assumptions will lead to different prediction intervals. Thus the intervals should not be interpreted too strictly. Rather they indicate the order of magnitude of forecast uncertainty. Careful monitoring of the forecast results may provide an answer to the question whether the prediction intervals are correct. In case a significantly higher share than two-thirds of the observed population numbers lies inside the 67 per cent prediction intervals, this signals intervals that are too wide, and vice versa.

The results apply to the national level only. Therefore the present approach ought to be extended. Regional population forecasts should also be presented in the form of prediction intervals, but an appropriate methodology to compute such regional intervals has yet to be developed. Two major problems have to be solved in regional stochastic population forecasting: first, to specify correct regional correlation structures for fertility, mortality and migration, and second, to specify a correct statistical model for interregional migration.

References

Akers, D. (1965) “Cohort fertility versus parity progression as methods of projecting births” *Demography* **2**: 414-428.

Alders, M. and J. de Beer. (1998) “Kansverdeling van de bevolkingsprognose” (“Probability distribution of population forecasts”) *Maandstatistiek van de Bevolking* **46**: 8-11.

Alho, J. (1990) “Stochastic methods in population forecasting” *International Journal of Forecasting* **6**: 521-530.

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

Alho, J. and B. Spencer (1985) “Uncertain population forecasting” *Journal of the American Statistical Association* **80**: 306-314.

Alho, J. and B. Spencer (1997) Statistical demography and forecasting. Unpublished manuscript.

Armstrong, J. (1985) *Long-range forecasting: From crystal ball to computer.* New York: Wiley (2nd ed.)

Beets, G. 1996. “Does the increasing age at first birth lead to increases in involuntary childlessness?” Proceedings European Population Conference Milano 4-8 September 1995, Volume 2. Milano: FrancoAngeli, 15-30.

Bell, W. (1992) “ARIMA and Principal Components models in forecasting age-specific fertility” pp. 177-200 in N. Keilman and H. Cruijsen (eds.) *National Population Forecasting in Industrialized Countries.* Amsterdam etc.: Swets & Zeitlinger.

Bell, W. (1997) “Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates”. *Journal of Official Statistics* **13**: 279-303.

Boleslawski, L. and E. Tabeau (2001) “Comparing theoretical age patterns of mortality beyond the age of 80” pp. 127-155 in Tabeau et al. (2001).

Box, G. and G. Jenkins (1970) *Time series analysis: Forecasting and control.* San Francisco: Holden Day.

Bratley, P., B. Fox, and L. Schrage (1983) *A guide to simulation*. New York: Springer-Verlag.

Brunborg, H. and S.-E. Mamelund (1994) *“Kohort- og periodefruktbarhet i Norge 1820-1993”* (“Cohort and Period Fertility for Norway 1820-1993”) Report no. 94/27. Oslo: Statistics Norway.

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. (1985) “A time-series model for cohort data” *Journal of the American Statistical Association* **80**: 525-530.

De Beer, J. (1989) "Projecting age-specific fertility by using time-series models", *European Journal of Population* **5**: 315-346.

De Beer, J. (1992) “General time-series models for forecasting fertility” pp 148-175 in N. Keilman and H. Cruijsen (eds.) *National Population Forecasting in Industrialized Countries.* Amsterdam etc.: Swets & Zeitlinger.

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. Internet www.unece.org/stats/documents/1999.05.projections.htm.

Duchêne, J. and S. Gillet-De Stefano (1974) “Ajustement analytique des courbes de fécondité générale”, *Population et Famille* **32**: 53-93.

Eurostat (1997a) “Latest national population forecasts for the countries of the European Economic Area and Switzerland - Reviewed by a set of tables”, Background Document Meeting Working Party on Demographic Projections. Luxembourg, September 1997.

Eurostat (1997b) “Beyond the predictable: Demographic changes in the EU up to 2050”, *Statistics in Focus, Population and Social Conditions* 1997/7.

Fredriksen, D. (1998) Projections of population, education, labour supply, and public pension benefits: Analyses with the dynamic microsimulation model MOSART. Social and Economic Studies 101. Oslo: Statistics Norway.

Hanika, A., W. Lutz and S. Scherbov (1997) "Ein probabilistischer Ansatz zur Bevölkerungs-vorausschätzung für Österreich", Statistische Nachrichten **12**/1997: 984-988.

Hannerz, H. (2001) "Manhood trials and the law of mortality" *Demographic Research* **4**. Internet http://www.demographic-research.org.

Hartmann, M. (1987) "Past and recent attempts to model mortality at all ages" *Journal of Official Statistics* **3**: 19-36.

Heligman, L. and J. Pollard (1980) "The age pattern of mortality" Journal of the Institute of Actuaries **107**: 49-80.

Hoem, J. (1976) "On the optimality of modified chi-square analytic graduation" Scandinavian Journal of Statistics **3**: 89-92.

Hoem, J., H. O. Hansen, D. Madsen, J. Løvgreen Nielsen, E. M. Olsen, and B. Rennermalm (1981) "Experiments in modelling recent Danish fertility curves" Demography **18**: 231-244.

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

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

Keilman, N. and A. Hetland (1999) "Simulated confidence intervals for future period and cohort fertility". Working Paper nr. 6, Joint ECE-Eurostat Work Session on Demographic Projections, Perugia, Italy, May 1999. Internet www.unece.org/stats/documents/1999.05.projections.htm.

Keilman, N. and D.Q. Pham (2000) "Predictive intervals for age-specific fertility". *European Journal of Population.* **16**(1): 41-66.

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.

Knudsen, C., R. McNown, and A. Rogers (1993) "Forecasting fertility: An application of time series methods of parameterized model schedules" *Social Science Research* **22:** 1-23.

Koreisha, S. and Y. Fang (1999) "The impact of measurement errors on ARMA prediction" *Journal of Forecasting* **18**: 95-109.

Kostaki, A. (1992a) *Methodology and applications of the Heligman-Pollard formula.* Lund: Department of Statistics, University of Lund.

Kostaki, A. (1992b) "A nine-parameter version of the Heligman-Pollard formula" *Mathematical Population Studies* **3**: 277-288.

Kravdal, Ø. (1994) "The importance of economic activity, economic potential and economic resources for the timing of first births in Norway" *Population Studies* **48:** 249-267.

Kuijsten, A. (1988) "Demografische toekomstbeelden van Nederland" *Bevolking en Gezin* **1988/2**: 97-130.

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. 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.

Lesthaeghe, R. and J. Surkyn (1988) "Cultural dynamics and economic theories of fertility change" *Population and Development Review* **14**: 3-45.

Lütkepohl, H. (1993) *Introduction to multiple time series analysis*. Berlin etc.: Springer-Verlag (second edition).

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.

McNown, R and A. Rogers (1989) Time-series analysis forecasts of a parameterised mortality schedule. Pp. 107-123 in P. Congdon and P. Batey (eds.) *Advances in regional demography: information, forecasts, models*. London: Belhaven Press.

Mamelund, S.-E. and J.-K. Borgan (1996) Kohort- og periode dødelighet I Norge 1846-1994 ("Cohort and Period Mortality in Norway 1846-1994"). Reports no. 96/9. Oslo: Statistics Norway.

National Research Council (2000) *Beyond six billion: Forecasting the world's population.* Panel on Population Projections. John Bongaarts and Rodolfo Bulatao (eds.) Committee on Population, Commission on Behavioral and Social Sciences and Education. Washington DC: National Academy Press.

Paulsen, B., B. Kalseth, and A. Karstensen (1999) "Eldres sykehusbruk på 90-tallet: 16 prosent av befolkningen – halvparten av sykehusforbruket" ("The elderly's use of hospital services in the 1990s: 16 per cent of the population – half the use of hospital services") SINTEF rapport STF78 A99527. Trondheim: SINTEF Unimed.

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.

Rogers, A. and L. Castro (1986) "Migration". Pp. 157-206 in A. Rogers and F. Willekens (eds.) *Migration and Settlement: A Multiregional Comparative Study.* Dordrecht: Reidel Publ. Co.

Rogers, A. and K. Gard (1991) Applications of the Heligman/Pollard model mortality schedule. *Population Bulletin of the United Nations* **30**: 79-105.

Spengler, J. (1935) "Population prediction in nineteenth century America" *American Sociological Review* **1**: 905-921.

Statistics Norway (1994) *Framskriving av folkemengden 1993-2050* ("Population Projections 1993-2050"). Oslo/Kongsvinger: Statistics Norway.

Statistics Norway (1997) *Framskriving av folkemengden 1996-2050* ("Population Projections 1996-2050"). Oslo/Kongsvinger: Statistics Norway.

Statistics Norway (1999) "Fortsatt befolkningsvekst. Befolkningsframskrivinger. Nasjonal og regionale tall, 1999-2050" ("Population growth continues.

Population forecasts: national and regional results 1999-2050") *Dagens statistikk* 17.11.1999. Internett http://www.ssb.no/folkfram.

Stoto, M. (1983) "The accuracy of population projections" *Journal of the American Statistical Association* **78**: 13-20.

Tabeau, E., P. Ekamper, C. Huisman, and A. Bosch (1999) "Improving overall mortality forecasts by analysing cause-of-death, period and cohort effects in trends. *European Journal of Population* **15**: 153-183.

Tabeau, E. (2001) "A review of demographic forecasting models for mortality" pp. 1-32 in Tabeau et al. (2001).

Tabeau, E. A. van den Berg Jets, and C. Heathcote, eds. (2001) Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective. Dordrecht: Kluwer Academic Publishers.

Texmon, I. (1992) "Norske befolkningsframskrivinger 1969-1990" ("Norwegian population projections 1969-1990") pp. 285-311 in O. Ljones, B. Moen, and L. Østby (eds.) Mennesker og Modeller. Oslo/Kongsvinger: Statistics Norway.

Thompson, P., W. Bell, J. Long, and R. Miller (1989) "Multivariate time series projections of parametrized age-specific fertility rates", *Journal of the American Statistical Association* **84**: 689-699.

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.

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.

Törnqvist, L. (1949) "Om de synspunkter, som bestämt valet av de primäre prognos-antagendena" ("On the points of view that determined the choice of the main forecast assumptions") pp. 69-75 in J. Hyppolä, A. Tunkelo, and L. Törnqvist *Beräkninger rörende Finlands befolkning, dess reproduktion och framtida utveckling* Helsinki, Statistiska Meddelanden, Utgivna av Statistiska Centralbyrån nr. 38 (in Swedish and Finnish).

Van Imhoff, E. (1991) Profile: A program for estimating the coefficients of demographic age-intensity profiles. NIDI-Report no. 15. The Hague: NIDI.

Willekens, F. and N. Baydar (1984) "Age-period-cohort models for forecasting fertility" Working Paper no. 45, The Hague: Netherlands Interuniversity Demographic Institute.

De sist utgitte publikasjonene i serien Sosiale og økonomiske studier
Recent publications in the series Social and Economic Studies

83 Ingvild Svendsen: Empirical Tests of the Formation of Expectations – A Survey of Methods and Results. 1993-52s. 75 kr. ISBN 82-537-3948-6

84 Bjørn E. Naug: En økonometrisk analyse av utviklingen i importandelene for industrivarer 1968-1990. *An Econometric Analysis of the Development of Manufacturing Import Shares 1968-1990.* 1994-78s. 95 kr. ISBN 82-537-3955-9

85 Einar Bowitz og Ådne Cappelen: Prisdannelse og faktoretterspørsel i – norske næringer. *Price Formation and Factor Demand in Norwegian Industries*. 1994-177s. 125 kr. ISBN 82-537-4024-7.

86 Klaus Mohn: Modelling Regional Producer Behaviour – A Survey. *Modellering av regional produsentatferd – En litteraturoversikt*. 1994-71s. 95 kr. ISBN 82-537-4042-5.

87 Knut A. Magnussen: Old-Age Pensions, Retirement Behaviour and - Personal Saving. A Discussion of the - Literature. *Alderspensjon, pensjoneringsatferd og privat sparing. En diskusjon av litteraturen.* 1994-69s. 95 kr. ISBN 82-537-4050-6.

88 Klaus Mohn, Lasse S. Stambøl og Knut Ø. Sørensen: Regional analyse av arbeidsmarked og demografi – Drivkrefter og utviklingstrekk belyst ved modellsystemet REGARD. *Regional Analysis of Labour Market and Demography with the Model – REGARD.* 1994-165s. 125 kr. ISBN 82-537-4082-4.

89 Nils Martin Stølen: Wage Formation and the Macroeconomic Functioning of the Norwegian Labour Market. *Lønnsdannelse og den makroøkonomiske funksjonsmåten til det norske arbeidsmarkedet.* 1995-306s. 180 kr. ISBN 82-537-4141-3.

90 Øystein Kravdal: Sociodemographic Studies of Fertility and Divorce in Norway with Emphasis on the Importance of Economic Factors. *Sosiodemografiske studier av fruktbarhet og skilsmisse i Norge med vekt på betydningen av økonomiske faktorer.* 1994-267s. 155 kr. ISBN 82-537-4088-3.

91 Tom Kornstad: Empirical Life Cycle Models of Labour Supply and Consumption. *Empiriske livsløpsmodeller for arbeidstilbud og konsum.* 1995-115s. 110 kr. ISBN 82-537-4166-9.

92 H.C. Bjørnland: Trends, Cycles and Measures of Persistence in the Norwegian Economy *Trender, konjunktursvingninger og varighet av sjokk i norsk økonomi*. 1995. 109s. 110 kr. ISBN 82-537-4220-7

93 Å. Cappelen, R. Choudhury, T. Eika: Petroleumsvirksomheten og norsk økonomi 1973 1993 *The Oil Industry and the Norwegian Economy 1973 1993. 1996*. 128s. 110 kr. ISBN 82-537-4287-8

94 K.O. Aarbu, B. Lian: Skattereformen og delingsmodellen: En empirisk analyse *The Norwegian tax reform and the capital income imputation method: An empirical analysis*. 1996. 94s. 95 kr. ISBN 82-537-4297-5

95 T.J. Klette, A. Mathiassen: Vekst og fall blant norske industribedrifter: Om nyetablering, nedlegging og omstilling *Growth and turnover among Norwegian manufacturing plants*. 1996. 112s. 110 kr. ISBN 82-537-4298-3

96 K.H. Alfsen, T. Bye, E. Holmøy (eds.): MSG EE: An Applied General Equilibrium Model for Energy and Environmental Analyses *MSG EE: En anvendt generell likevektsmodell for energi og miljøanalyser*. 1996. 171s. 125 kr. ISBN 82-537-4342-4

97 A. Barstad: Store byer, liten velferd? Om segregasjon og ulikhet i norske storbyer Big Cities *Little Welfare? Segregation and Inequality in Norwegian Cities*. 1997. 153s. 125 kr. ISBN 82-537-4402-1

98 T.O. Thoresen: Mikrosimulering i praksis. Analyser av endring i offentlige overføringer til barnefamilier *Tax Benefit Model in Use. Analysing Changes in the Public Policy towards Families with Children*. 1998. 102s. 135 kr. ISBN 82-537-4527-3

99 K.E. Rosendahl: Social Costs of Air Pollution and Fossil Fuel Use – A Macroeconomic Approach *Samfunnsøkonomiske kostnader av luftforurensning og fossile brensler – En makroøkonomisk tilnærming*. 1998. 147s. 135 kr. ISBN 82-537-4542-7

100 R. Kjeldstad: Enslige forsørgere: Forsørgelse og levekår før og etter overgang til en ny livsfase *Single Parents: Social and Economic Adjustment before and after Transition to single Parenthood*. 1998. 162s. 155 kr. ISBN 82-537-4548-6

101 D. Fredriksen: Projections of Population, Education, Labour Supply and Public Pension Benefits. Analyses with Dynamic Microsimulation Model MOSART *Framskrivinger av befolkning, utdanning, arbeidsstyrke og folketrygdens pensjonsutgifter. Analyser utført med mikrosimulerings-modellen MOSART*. 1998. 123s. 135 kr. ISBN 82-537-4572-9

102 T. Bye, M. Hoel og S. Strøm: Et effektivt kraftmarked - konsekvenser for kraftkrevende næringer og regioner. 1999. 104s. 165 kr. ISBN 82-537-4698-9

103 O. Bjerkholt: Kunnskapens krav. Om opprettelsen av Forskningsavdelingen i Statistisk sentralbyrå. 2000. 296s. 260 kr ISBN 82-637-4856-6

104 H. Dale-Olsen og Dag Rønningen: Jobb- og arbeidskraftsstrømmer i Norge og OECD. En komparativ analyse av jobb- og arbeidskraftsstrømmer med fokus på årsaker *Job and worker flows in Norway and OECD. A comparative analysis of job and worker flows. Facts and explanations.* 2000. 186s. 210 kr. ISBN 82-537-4866-3