Assumptions for long-term stochastic population forecasts in 18 European countries

Hypothèses de projections stochastiques à long terme des populations de 18 pays européens

Maarten Alders · Nico Keilman · Harri Cruijsen

Abstract   The aim of the 'Uncertain Population of Europe' (UPE) project was to compute long-term stochastic (probabilistic) population forecasts for 18 European countries. We developed a general methodology for constructing predictive distributions for fertility, mortality and migration. The assumptions underlying stochastic population forecasts can be assessed by analysing errors in past forecasts or model-based estimates of forecast errors, or by expert judgement. All three approaches have been used in the project. This article summarizes and discusses the results of the three approaches. It demonstrates how the—sometimes conflicting—results can be synthesized into a consistent set of assumptions about the expected levels and the uncertainty of total fertility rate, life expectancy at birth of men and women, and net migration for 18 European countries.

Keywords   Probabilistic forecast · Forecast assumptions · Time series · Empirical errors · Expert judgement · Scaled model of error

Résumé   Le but du projet 'Uncertain population of Europe' (UPE) était de calculer des projections de population stochastiques (probabilistes) à long terme des populations de 18 pays européens. Nous avons développé une méthodologie générale pour construire des distributions prédictives de fécondité, mortalité et migration. Les hypothèses sous-jacentes aux projections stochastiques de populations peuvent être élaborées en analysant les erreurs de projections passées, en effectuant

M. Alders
Statistics Netherlands, PO Box 4000, NL-2270 JM Voorburg, The Netherlands,
e-mail: mals@cbs.nl

N. Keilman (✉)
Department of Economics, University of Oslo, PO Box 1095, Blindern, Oslo N-0317, Norway
e-mail: nico.keilman@econ.uio.no

H. Cruijsen
DEMOCAST, Vluchtheuvelstraat 2, NL-6621 BK Dreumel, The Netherlands
e-mail: hvckroon@kabelfoon.nl

une modélisation pour estimer les erreurs de projection, ou par un jugement d'expert. Les trois approches ont été appliquées dans le projet, et leurs résultats sont résumés et discutés dans cet article. Nous démontrons que les résultats, parfois contradictoires, peuvent être synthétisés pour former un ensemble cohérent d'hypothèses concernant les niveaux attendus et l'incertitude autour de l'indice synthétique de fécondité, de l'espérance de vie à la naissance des hommes et des femmes, et de la migration nette dans 18 pays européens.

Mots-clés   projection stochastique · hypothèses de projection · séries temporelles · jugement expert · modèle des erreurs d'échelle flexible

1 The UPE project

Demographic trends in Europe have continued to take forecasters by surprise. Few predicted the rapid declines and ongoing low levels of fertility in the Mediterranean and former communist countries during the last three decades. Similarly, the sharp fall in death rates in countries where life expectancy at birth was already high (e.g. France, Italy and Sweden) was not foreseen by many. Finally, considerable and sometimes even massive migration flows came unexpectedly.

Although there is some hope that more detailed or comprehensive demographic studies may help to improve our understanding of the causes of these errors after the fact, there appears to have been an element of genuine surprise at the demographic trends mentioned above. Therefore, there is no reason to believe that such developments will be easier to predict in the near future than they were in the past. If population forecasts are to be used to formulate policies regarding the labour market, health care, economic development or pension systems, then the uncertainty involved should be quantified and included in those forecasts.

This was the purpose of the Uncertain Population of Europe (UPE) project: to compute stochastic population forecasts for 18 European countries, which we will denote as EEA+ countries. The group consists of the 15 members of the European Union prior to the joining of the new member states in 2004 (i.e. Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Ireland, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the UK), plus Norway, Iceland and Switzerland. Except for Switzerland, these countries make up the so-called European Economic Area, hence EEA+.[1] We have quantified uncertainty of the demographic forecast by applying the cohort-component book-keeping model for each country 3,000 times, with a deterministic jump-off population and probabilistically varying values for age- and sex-specific mortality, age-specific fertility and net migration by age and sex. The starting point was the population on 1 January 2003, by country, 1-year age group and sex. The forecast horizon was 2050. The method is based on the so-called scaled model for error, implemented in the program Program for Error Propagation (PEP). Brief verbal descriptions of this model and of PEP are contained in Appendix 1.

For each year, three main sets of assumptions were required:

---

[1] We omitted EEA member Liechtenstein for practical reasons.

1. Country-specific point predictions for age-specific rates of fertility, age- and sex-specific rates for mortality and numbers of net immigration broken down by age and sex. Assumptions of this kind are the same as those that statistical agencies formulate when they compute their deterministic population forecasts.
2. Country-specific uncertainty parameters for fertility and mortality rates and for migration numbers. More particularly, variances and first-order autocorrelations were needed for the logarithm of the fertility and the mortality rates and for the net-migration numbers.
3. Correlations across countries for fertility, mortality and migration.

We have derived these forecast assumptions from three separate sources:

1. Time-series analyses of age-specific and total fertility; age- and sex-specific mortality and life expectancy at birth; and net migration by age and sex, relative to total population size.
2. Analyses of historical forecast errors for total fertility, life expectancies and net migration.
3. Interviews with subject experts for fertility, mortality and migration.

The purpose of this article is to report on the assumption-making process. This process included many steps and we cannot describe them all. More information can be found at the UPE website, http://www.stat.fi/tup/euupe/, and in particular in the project report available at http://www.stat.fi/tup/euupe/upe_final_report.pdf. The website also contains forecast results for each of the 18 countries in the form of age and sex details for 10-year intervals to 2050. Alho et al. (2006) summarize the results.

The UPE project is the first attempt to combine information from these three sources in a systematic and balanced way. It shows that the three approaches are truly complementary. Earlier stochastic forecasts did combine elements of the three, but one of them was usually dominant—in many cases, a time series model. Lee and Tuljapurkar (1994) modelled the time series of the level parameter for US fertility obtained by means of the Lee–Carter method as an Autoregressive Integrated Moving Average or ARIMA (1,0,1) process with a constrained mean, subjectively chosen equal to 2.1. Alho (1998) compared prediction intervals for the total fertility rate (TFR) in Finland obtained by means of an ARIMA (1,1,0) model with those that result from the errors of so-called naïve forecasts (i.e. forecasts that assume the current TFR level is a reasonable forecast of the future TFR). He used a similar method for mortality. He also combined errors of naïve forecasts with time-series analysis and expert judgement in his crude assessments of forecast uncertainty for 12 large world regions (Alho, 1997). De Beer and Alders (1999) modelled the life expectancy of the Netherlands as a random walk with drift, and compared the resulting prediction intervals with those obtained from a time series of historical forecast errors for life expectancy. They also modelled the TFR in the Netherlands as a random walk (without drift) and calculated historical forecast errors for the TFR. The final assumptions rely heavily on a judgemental analysis distinguishing the TFR by parity (Alders & De Beer, 2004). Lutz, Sanderson, and Scherbov (2001) chose a certain level for the variance in the TFR in a target year. The variance was larger for regions with high fertility than for low-fertility regions. As to mortality, they generally assumed that life expectancies would increase between 0 and 4 years with 80% probability. These subjectively chosen distributions were combined with a

moving average time series process for the error in the TFR or life expectancy increase. At the same time, the authors aimed at producing prediction intervals that were at least as large as those published by the NRC panel for major world regions (National Research Council, 2000). Keilman, Pham, and Hetland (2002) modelled the log of the TFR in Norway as an ARIMA (1,1,0) model, but obtained unreasonably large prediction intervals for the TFR in the long run. In their simulations, they rejected TFR values larger than four children per woman. Their simulations for life expectancy were based on a complicated multivariate ARIMA model, the predictions of which were checked against observed errors in historical Norwegian life expectancy forecasts.

This article presents the approach that we followed for the point predictions and the prediction intervals for fertility, mortality and migration assumptions. We report the intervals in the form of 80% prediction intervals. In our view, 80% intervals give a better impression of forecast uncertainty than the more usual 95% intervals, which reflect extremes. Cross-national correlations are mentioned only briefly. Alho (2005) gives a more extensive report on that topic. Finally, we assumed independence across the components of fertility, mortality and migration.

In practice, we derived initial guesses for point predictions of model parameters and for uncertainty parameters from time-series analyses. These were adjusted, where necessary, based on historical forecast errors. We made further adjustments, sometimes of considerable magnitude, to reflect expert views.

2 Data issues

2.1 Principal data series needed

Since we applied the cohort-component approach, we needed long time series for age-specific fertility, mortality and net migration for each country. This required the following annual data:

- population on 1 January by sex and single years of age (0, 1,...,100 +);
- live births by sex;
- live births by single years of age of the mother (age at last birthday: 15, 16,...,49);
- deaths by sex and single years of age (age on 31 December: 0, 1,...,101 +);
- net migration by sex and single years of age (*idem*).

In addition, we needed internationally comparable time series for as many years as possible for the TFR, life expectancy at birth by sex and net migration. To facilitate comparisons across countries, we scaled net migration for each country by the population size on 1 January 2000. The base population is that on 1 January 2003.

2.2 Measurement problems

We have assumed that population statistics in all 18 countries are based on the de jure concept, which covers all people who have legal residence and/or usual residence in the country, even if they are temporarily abroad. The de jure population concept should be distinguished from the de facto population concept, which includes all people who are actually present in the country at a given moment in

time, regardless of whether they have legal and/or usual residence there. The latter population concept includes, for instance, all non-resident tourists and people without a legal residence permit; at the same time, it disregards residents who are abroad, such as tourists and people who have not reported emigration. In a multi-country project it is important to use one concept in order to avoid double counts and missing persons.

Countries that use population register information for producing annual population statistics seem to follow the de jure population concept (Eurostat, 2003). In our group of 18 countries, the national statistical offices of the following 13 countries use information from population registers: Austria, Belgium, Denmark, Finland, Germany, Iceland, Italy, Luxembourg, the Netherlands, Norway, Spain, Sweden and Switzerland. The majority of these countries also use the outcomes of population censuses, roughly once per decade. The five countries without a register (France, Greece, Ireland, Portugal and the UK) rely on the outcomes of population censuses, combined with information from vital registration systems or sample surveys to measure migration flows. All countries that carry out population censuses report that they follow the respective United Nations regulations, which recommend counting based on the de jure population concept.

However, in practice, countries may encounter various types of problems when attempting, without too much delay, to accurately determine or update the population age and sex structure according to the de jure concept. Most of these problems are caused by international migration, either directly or indirectly. Below we will briefly mention problems connected to (1) the residence status of people who experience a vital event or migration, (2) measurement and definition of international migration, (3) regularization of illegal or undocumented migrants and (4) post-census adjustments of population statistics. We will not discuss the accuracy of stock data for the oldest old, or measurement problems for vital events connected to different age definitions (age at last birthday, age on 31 December/1 January, etc.).

First, all countries draw up a birth certificate when a child is born and a death certificate when a person dies. Yet not all live births and deaths among the resident population will be counted. Births and deaths of residents who are temporarily abroad are either not registered at all or registered only with significant delays. At the same time, births and deaths of non-residents may be included in a country's population statistics. We know that half of the 18 countries systematically base their vital statistics on the de jure concept: Belgium, Denmark, Finland, Iceland, Luxembourg, the Netherlands, Norway, Sweden and Switzerland. The remaining nine countries work (or have worked until recently) with a mixture of de jure and de facto vital statistics measurement systems: Austria, France, Germany (births only), Greece, Ireland, Italy, Portugal, Spain and the UK (Eurostat, 2003). At least four of these (France (births only), Ireland, Portugal and the UK) handle this problem in a symmetric way: de jure births/deaths occurring abroad are excluded, while de facto births/deaths occurring in the country are included. Thus the errors compensate to a certain extent. In the remaining countries, there may be structural under-estimations or over-estimations in annual numbers of live births and deaths.

Second, a more significant measurement problem relates to a range of difficulties in estimating, in a consistent manner, de jure international migration flows. For

instance, Poulain, Debuisson, and Eggerickx (1990) have extensively documented the fact that definitions of immigration and emigration vary substantially within Europe. Until now, only the statistical agencies in the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) have succeeded in establishing a mutual, internationally consistent system of migratory flows occurring within their region. Furthermore, in spite of ongoing national and international efforts, a few EU countries do not measure international migration flows on an annual basis. France, Greece (emigration only), Ireland, Portugal and the UK lack a population registration system and therefore have to estimate annual migration flows using various indirect sources. Only when the results of a new population census become available one can try to make improved re-estimations.

A third problem is connected to unreported emigration in countries with a population register. For example, annual numbers of people who left the Netherlands without reporting their move to the population register of the municipality where they had lived have increased over the past 20 years from less than 5,000 to well over 35,000. Meanwhile, annual registered emigration has increased only slowly, to a level of around 70,000 people in 2003.

Fourth, measuring international migration accurately is difficult because of increasing numbers of illegal or undocumented migrants. Contemporary regularization programmes in Greece, Italy, Portugal and Spain show that millions of people can enter and stay in the European Union for years without a legal residence permit. They are able to do so in spite of the extension and reinforcement of border controls and the development and implementation of much stricter rules and higher penalties for hiring illegal or semi-legal employees. In addition, rules for asylum seekers, seasonal workers and migration resulting from family reunion and/or family formation have become more restrictive. This may have led to more illegal migrants. Hence the de jure population has become increasingly different from the de facto population.

Measuring international migration accurately is also difficult because whether a person is considered an international migrant or not depends on the intended length of stay in the country of destination. It is reasonable to assume that processes of globalization and individualization have changed the character of migration. First, both the magnitude and the share of short-term migration due to asylum, study, work or family formation have drastically increased over the past 2–3 decades. At the same time, the number and proportion of those who intend to migrate more or less permanently have become less important. This implies that increasing numbers of international migrants tend to shift from one category to another over the course of their life. However, migration measurement systems record only the current reason for migration; they are not able to capture a move which depends on circumstances earlier in the migrant's life.

These four groups of problems connected to international migration imply that it is difficult to compare demographic data across countries and over time. However, very little is known about the magnitude of the errors involved. Section 2.3 gives numerical examples for a few selected countries. A systematic investigation of the consequences of these measurement problems for population forecasts was beyond the scope of the UPE project and thus we have not quantified these errors. This means that the prediction intervals are too narrow by this error source alone, although we do not know by how much.

2.3 Data availability and data quality

National statistical agencies possess the longest demographic time series. However, as already mentioned in the previous section, national series may have different practices for calculating or estimating rates and summary indicators. Furthermore, national historical series are not always easily available or well documented.

Over the past two to three decades, internationally harmonized demographic time series have become available. Examples are the well-known international demographic databases of the United Nations (Population Division), the Council of Europe (CoE) and the Statistical Office of the European Communities (Eurostat). The CoE and Eurostat have been substantially supported by the work of the European Demographic Observatory (Observatoire Demographique Européenne or ODE) in Paris. ODE has successfully implemented an internationally accepted standardized system of calculating age-specific fertility and mortality rates, TFRs and life expectancies (SYSCODEM; for a detailed description, see Eurostat, 2005a). Another important international database is the Human Mortality Database of the University of California, Berkeley (USA), and the Max Planck Institute for Demographic Research (Germany).

Unfortunately, on international migration no comprehensive, internationally harmonized database exists. The international migration database compiled by Eurostat since the beginning of the 1990s had closed down at the time we were carrying out our project, due to a large number of inconsistencies and missing data.

We have used the following main data sources in the UPE project:

- TFR: Chesnais (1992) and Council of Europe (2002).
- Life expectancy at birth: Council of Europe (2002) and the Human Mortality Database of the University of California, Berkeley (USA), and the Max Planck Institute for Demographic Research (Germany).
- Net migration: Council of Europe (2002).

In a few cases, these international sources have been supplemented with information from national sources. Occasionally, for Germany and the UK, sub-national series have been applied, describing the situation for the Federal Republic of Germany (FRG) and for England and Wales, respectively. Country-specific details are contained in Keilman and Pham (2004).

Some time series are very long (e.g. TFRs for Finland since 1776; life expectancies for France since 1806), while others are short (e.g. life expectancies for Ireland since 1985). For all 18 countries considered, the annual series for net migration start in 1960.

In order to generate the detailed set of quantitative assumptions on age-specific fertility and mortality, we constructed a separate international database covering the period 1990–2003, mainly using figures taken from Eurostat's database NewCronos (as available during spring 2004). The same source supplied us with data on net migration by age and sex for countries with a population register. Finally, NewCronos, combined with demographic now-casts for 2003, also gave us figures for the initial population on 1 January 2003 (Eurostat, 2004).

With respect to the key indicators, we can state that the time series on net migration are by far the weakest. Annual figures for migration have been generally estimated based on the difference between total population growth and natural

growth. Thus they include measurement errors connected to all three components of change. The ongoing practice of using different definitions and measurement systems on international migration and/or the application of different post-census re-estimation procedures and population counts have obviously led to a considerable number of international inconsistencies and strong trend shifts. The most striking examples are:

- After the population census of 1999, France re-estimated for the period 1995–1999 an average annual crude net migration level of around –0.2 per 1,000 population in 2000, whereas all other EU countries reported crude net migration levels during the second half of 1990s of at least 1.5. Since the year 2000, France has assumed a crude net migration level close to 1 per 1,000—more or less similar to the levels provisionally estimated before the census of 1999.
- Before its latest population census, held in October 2002, Italy reported a total net migration of almost 1.5 million people for the period 1991–2001; however, based on the 2002 census counts, the total net migration for this period appeared to be no more than 0.7 million people.
- The 2002 issue of *Recent Demographic Developments in Europe* reports 'observed' net migration to Portugal in multiples of 1,000 for each year since 1992 (Council of Europe, 2002). The 1998 issue reports net migration for the years 1991–1997 even in multiples of 5,000. For the years 1993–1997, there is little agreement between the two time series of net-migration numbers.
- Some countries show large differences between pre-census around 2000/2001 population figures by sex and age, and census outcomes. In a few cases (e.g. France, Italy and the UK), relative deviations amount to well over 5%. Especially for the age groups 20–30 and 80+, the latest census results reveal that pre-census estimates were too high.

3 Historical forecast errors

We collected information on errors in historical forecasts by the national statistical agencies of the following 14 countries: Austria, Belgium, Denmark, Finland, France, Germany,[2] Italy, Luxembourg, the Netherlands, Norway, Portugal, Sweden, Switzerland and the UK. Most of the forecasts date from the period 1960–2000, although some early ones go back to the 1950s. We have used both published and unpublished sources. We selected the TFR, life expectancy at birth and net migration (i.e. the difference between immigration and emigration) as indicators for the three demographic components of change. Keilman and Pham (2004) give details of the data collection process and the quality of the data.

The data set is restricted to forecasts produced by statistical agencies. An important reason for this choice was the fact that the forecasts were made with a single methodology, namely the cohort component method of population

---

[2] More precisely, the FRG between 1952 and 1989 and the reunified Germany from 1990 onwards. For fertility and mortality, we have forecast errors for the (former) FRG for the period 1952–2002. For migration, we have errors for the FRG in forecasts made between 1952 and 1989, and for Germany in forecasts made since 1990. All errors apply to the period from the launch year up to 2002.

forecasting. Indeed, this is the standard forecasting methodology among population forecasters (Keilman & Cruijsen, 1992). A second reason was that the forecasts were produced in stable institutional settings. Thus we have a relatively homogeneous data set, which provides a meaningful basis for error analysis.

We computed annual forecast errors as the simple difference between forecast value and corresponding observed value for each of the three indicators. Thus a positive error indicates that the forecast was too high, a negative error that it was too low. In many cases, variant assumptions were used in a specific forecast. For example, the 1990 forecast for Norway includes a low, a medium and a high assumption for fertility. Variant assumptions were also frequently made for the components of mortality and migration. In that case, we included all variants in our data set, because very few of the forecast reports contained clear advice as to which of the variants the statistical agency considered the most probable at the time of publication. Hence it was left to the user to pick one of them. We can assume that all the variants have been used, although the middle one probably more often than the high or the low ones (in cases where there were three variants).[3]

Figure 1 plots the mean absolute error (MAE) and the mean error (ME) in the TFR. The means are computed across countries, forecast periods and forecast variants, but controlling for forecast duration. The MAE reflects forecast accuracy. It tells us how far off the forecast was, irrespective of the sign of the error. The ME reflects forecast bias. Figure 1 shows that the TFR forecasts made in the 14 countries since the 1950s were wrong by an average of 0.3 children per woman for a forecast horizon of 15 years ahead and by 0.4 children per woman for 25 years ahead. They already differ from the actual TFR by 0.06 in the first year. In the very long run, all forecasts were too high, since the ME coincides with the MAE; for short- and medium-term forecasts, there was some compensation between positive and negative errors, since the ME is lower than the MAE. Figure 1 reflects the well-known fact that fertility was over-predicted in many European countries in the late 1960s and the 1970s, when actual fertility fell rapidly.

Figure 2 shows the MAE and the ME for life expectancy. There are hardly any differences between the means for men and women. Therefore we have plotted the curves for only one sex. Life expectancy has systematically been under-predicted, by more than 2 years for forecasts 15 years ahead and by 4.5 years for 25 years ahead. Nearly all forecasts had life expectancy too low, and so mortality too high, since the curves for the MAE and the ME are almost perfectly symmetric around zero.

Errors in scaled net migration are summarized in Fig. 3. A number of historical projections have ignored migration, particularly the earliest ones. It is reasonable to assume that many users will have considered them as the statistical agency's best guess regarding the country's future population. Therefore we have assumed that the implicit forecast hypothesis for international migration was a net migration level of zero. Hence the signed error was simply equal to minus the observed net migration in those cases.

Net migration levels have been consistently under-predicted in historical forecasts. In a number of cases, the reason is that migration was omitted from the forecast, while actual net migration was positive. In other cases, the net migration

---

[3] For some countries, we had enough data to check the implications of this choice. For Norway, the standard deviation in the observed TFR errors based on all forecast variants was very close to that based on main variants only. For Sweden, the all-variants standard deviations were approximately 10% higher than those based on main variants.
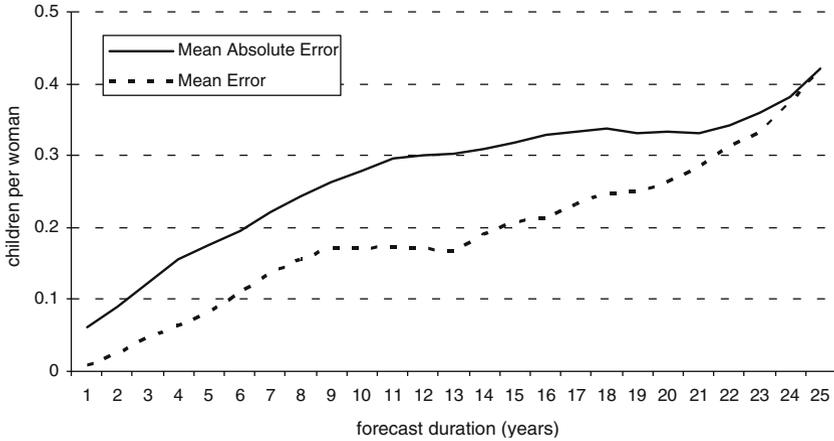
Fig. 1 Errors in TFR forecasts
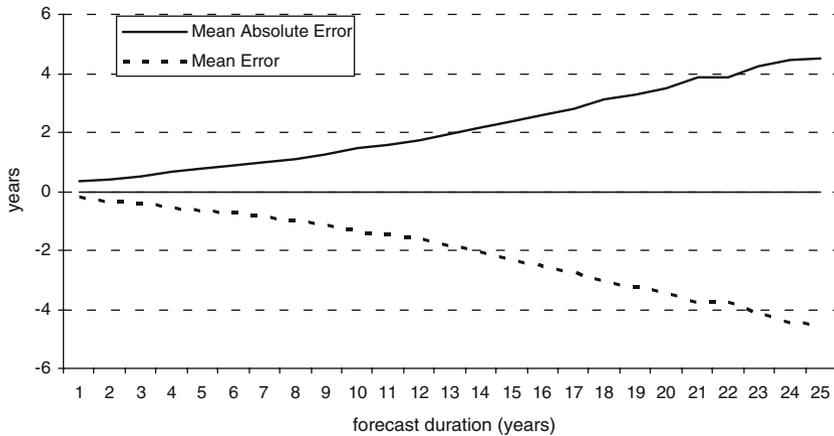
Fig. 2 Errors in life expectancy forecasts

assumption was simply too low. We found two distinct groups of countries. One group consists of Austria, Germany, Luxembourg, Portugal and Switzerland. The countries in this group have MEs well above the average. The forecasts for Austria, Germany and, to some extent, Switzerland were less accurate than the average, because of large immigration flows after the fall of the Berlin Wall in 1989. Luxembourg is a small country in which the level of migration in itself is high, so large migration forecast errors frequently occur. The large errors for Portugal are explained by the fact that migration statistics are not as reliable as those in other EEA countries (see Sect. 2). Countries of the other group, which consists of Belgium, Denmark, Finland, France, Italy, the Netherlands, Norway, Sweden and the UK, show much smaller errors in their migration forecasts.

In summary, historical forecasts in the region on average assumed levels of future fertility that were too high and levels of mortality and immigration that were too low. Both forecast bias (reflected by the ME) and forecast inaccuracy (MAE) increased regularly with forecast duration.
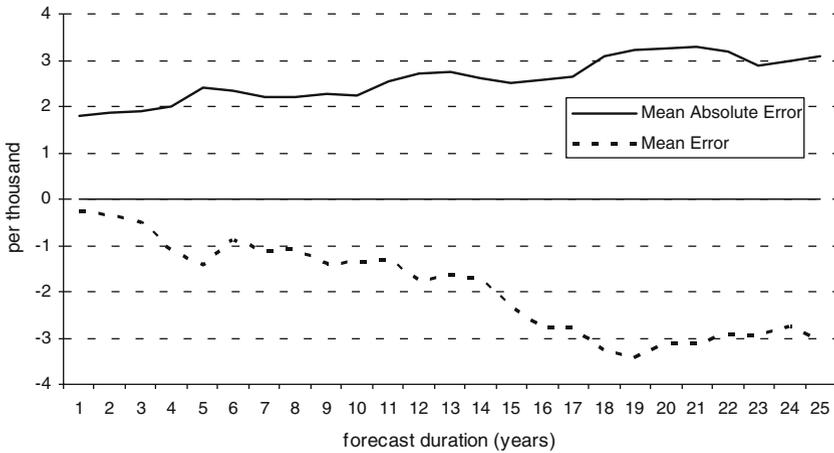
**Fig. 3** Errors in net migration forecasts

4 Time-series analysis

The purpose of the time-series analysis was to compute expected values (point predictions) and prediction intervals to 2050 for fertility, mortality and net migration in each country. We applied two types of time series models: (1) a naïve model, in which we assumed constant levels for the TFR and net migration, or constant reductions in the age-specific death rates; (2) a more advanced model, using ARIMA and GARCH types of model (Generalized Autoregressive Conditional Heteroscedasticity, GARCH). This second approach was used for the TFR, life expectancy and net migration. We will briefly present the main features of the time-series analyses in terms of predicted values and 80% intervals in 2050. For the ARIMA and GARCH models, these intervals are determined by the statistical distribution of the residual term and those of the parameter estimates. For the naïve time series models, we computed empirical errors for each calendar year as the difference between the naïve prediction and the actual value for that year.

4.1 Fertility

Figure 4 plots the TFR in the 18 countries. Here our interest is in the overall trend. The countries show a similar pattern in 20th-century TFRs, which reflect the demographic transition, followed by the effects of the economic recession in the 1930s and the baby boom in the 1950s and 1960s. Major events, such as the First and Second World Wars and the outbreak of Spanish influenza in 1918–1919, are clearly reflected in the series for most countries. In the 20th century, many countries show a tendency towards lower variability in TFRs. Also, inter-country differences had become quite small in the 1990s.[4]

An important question is how much of the data one should use in the modelling. Several issues are at stake here. First, Box and Jenkins (1970, 18) suggest at least 50

---

[4] This is true for the differences between countries in the original TFR scale. However, *relative* differences between countries have become larger since the mid-1960s. See also Sect. 6.
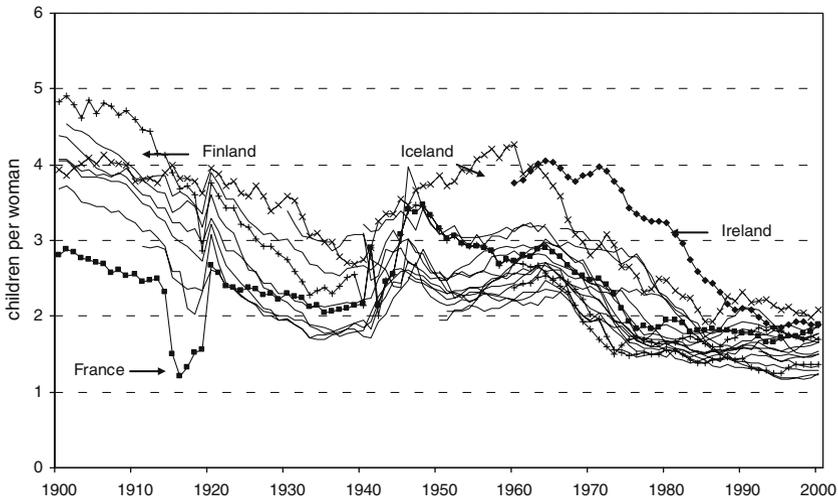
Fig. 4 TFR in 18 European countries

observations for ARIMA-type time series models, although annual models (in contrast to monthly time series) probably need somewhat shorter series. Second, the quality of the data is better for the 20th century than for earlier years. This is particularly true for the denominators of the fertility rates (i.e. the annual numbers of women by single years of age). Third, we can question the relevance of data as far back as the mid-1800s. Current childbearing behaviour is very different from that of women in the 19th century. Fourth, our ultimate goal is to compute long-term predictions for some 50 years ahead, which necessitates a long series.

The ultimate choice is necessarily a subjective one that includes a good deal of judgement and arbitrariness. We believe that we can strike a reasonable balance between conflicting goals by selecting the 20th century as the basis for our models. An analysis solely based on the last 50 years, say, would be unfortunate: it would include the baby boom of the 1950s and early 1960s, but not the low fertility of the 1930s, to which the boom was at least partly a reaction. A base period stretching back into the 19th century would be hampered by problems of data quality, and it would unrealistically assume that the same model could capture demographic behaviour over such a long period. In a sensitivity analysis for Denmark, Finland, Norway and Sweden we also experimented with base period 1945–2000. For Norway and Finland we found 95% prediction intervals that were smaller (by 1.4 and 0.5 children per woman on average, respectively) than those we have accepted for further analysis (see below). For Denmark and Sweden they were larger (by 0.8 and 1.2 children per woman, respectively).

We have long data series for nine countries: Denmark, Finland, France, Iceland, the Netherlands, Norway, Sweden, Switzerland, and England and Wales.[5] We have

---

[5] Available time series for the observed values of the TFR and life expectancy are rather short for the UK (England, Wales, Scotland and Northern Ireland). The situation is much better for England and Wales: annual TFR series are available from 1911 and annual life expectancy values from 1841. Thus we have assumed that variability and predictability of fertility and mortality in the UK in the 20th century were the same as those in England and Wales.

estimated time series models for the TFR based on a whole century of data for these nine. Time series models for the remaining nine countries were estimated based on annual TFR data for the years 1950–2000. This was the case for Austria, Belgium, Germany, Greece, Ireland, Italy, Luxembourg, Portugal and Spain.

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

We have applied an ARCH time series model to the log-transformed TFR. Let Z_t be the logarithm of the TFR in year t. Then the model is:

$$Z_t = C + \phi Z_{t-1} + v_t + \eta_1 U_{1,t} + \eta_2 U_{2,t} + \eta_3 U_{3,t} + \eta_4 U_{4,t} + \eta_5 U_{5,t}$$
$$v_t = \psi_1 v_{t-1} + \psi_2 v_{t-2} + \cdots + \psi_m v_{t-m} + \varepsilon_t$$
$$\varepsilon_t = \left(\sqrt{h_t}\right) e_t \tag{1}$$
$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$$

where $e_t \sim N(0,1)$. This is the AR(m)-ARCH(q) model. The outliers caused by the two world wars and by the outbreak of Spanish flu are handled by between two (Denmark, Iceland, the Netherlands, Sweden) and five (Switzerland) dummy variables $U_{i,t}$. In addition we have $\omega > 0$ and $\alpha_i \geq 0$.

The maximum number of terms m included in the autoregressive expression of v_t was initially set equal to 10, but few of the ψ estimates turned out to be significantly different from zero. In practice, m was restricted to two. Similarly, estimates for α_i suggested that the order (q) of the CH part of the model could be restricted to one. We tested the residuals for normality, independence and constant variance. Details are given in Keilman and Pham (2004).

For the nine countries with long time series for the TFR, two sets of prediction intervals up to 2050 were constructed: one based on the annual data series 1900–2000, another based on annual figures observed during the period 1950–2000.

We assessed the robustness of the prediction intervals by applying several simpler time series models (e.g. a pure AR(m) model) on long series of data for Denmark, Finland, Norway and Sweden. Based on these sensitivity tests, we concluded that the ARCH model in expression (1) gives a useful and reliable description of the development in the TFR in the four countries in the previous century. Given the similarity of trends, we have assumed that this is also the case for the other countries.

Application of the ARCH type of model to the annual TFR series of all 18 countries for the period 1950–2000 led to the conclusion that the CH part of model (1) was needed only for Belgium, Germany, and England and Wales. Obviously, in

most countries the TFR level was less volatile during the second half of the 20th century than during the first half. In addition, due to the recent sharp fall in fertility, the constant term had to be omitted for Greece, Ireland, Italy, Portugal and Spain.

We used the model to compute prediction intervals for the future TFR up to 2050. Since we cannot be certain that the estimated coefficients are equal to the real ones, we used simulation to obtain these intervals. In each of the 5,000 simulation runs, parameter values were drawn from a multivariate normal distribution, with expectation equal to the parameter estimates for model (1), together with the estimated covariance matrix. The possibility that a pandemic as devastating as Spanish flu or a war with consequences as catastrophic as either the First or the Second World War could occur during the prediction period was included in the simulations based on data since 1900. For each dummy variable, we first drew a random number from the binomial distribution with a probability of 'catastrophe' equal to 1/101. Next, the starting year for the catastrophe was determined based on a random draw from the uniform distribution on the interval [2001, 2050]. Finally, its effect followed from the estimated expectation and variance of the dummy coefficient.

The ARCH predictions for the TFR in the year 2050 for the nine countries with long data series vary from 1.3 children per woman for Switzerland to 1.9 children per woman for France and the Netherlands. The 80% prediction intervals in 2050 are between 1.1 (Switzerland) and 1.4 (Finland, Iceland, Norway) children per woman wide. These intervals are narrower than corresponding intervals based on (unconstrained) ARIMA-type time series models: see, for instance, Thompson, Bell, Long, and Miller (1989) and Keilman et al. (2002). The reason is that our model (1) explicitly takes account of the reduced variability in the TFR over time, whereas ARIMA models assume constant variance.

When the ARCH model is fitted to the shorter time series 1950–2000 in all 18 countries, the point predictions in 2050 show a larger range: from 1.1 (Greece, Italy, Spain) to 2.0 (Belgium) children per woman. The widths of the 80% prediction intervals range from 0.7 (Greece) and 0.8 (Portugal) to 1.7 (Austria, Germany) and 2.1 (Sweden) children per woman. For the nine countries involved, the prediction intervals based on short time series are (with the exception of Finland) at least as wide, and for the Netherlands and Sweden much wider, than the intervals based on long series.

The naïve model assumes that a TFR value as observed for year t, TFR(t), gives a forecast for k years later, TFR($t + k$), as TFR($t + k$) = TFR(t), k = 1, 2, 3, ..., 50. For each forecast duration k, we estimated empirical error patterns by varying the base year t. For nine countries we had long data series, and thus empirical error distributions that were based on many data points, even for a forecast horizon of 50 years. For countries with short series, pooling was necessary. We found that predictions of 50 years ahead had empirical 80% prediction intervals between 1.6 and 2.2 children per woman wide.

4.2 Mortality

Figure 5 shows the life expectancy at birth for men and women in the 18 countries. Major interruptions to the upward trend, caused by two world wars and Spanish flu, are clearly visible. The time series show less variability in the second half of the 20th century than in the first half. In addition, differences between countries appear to

men
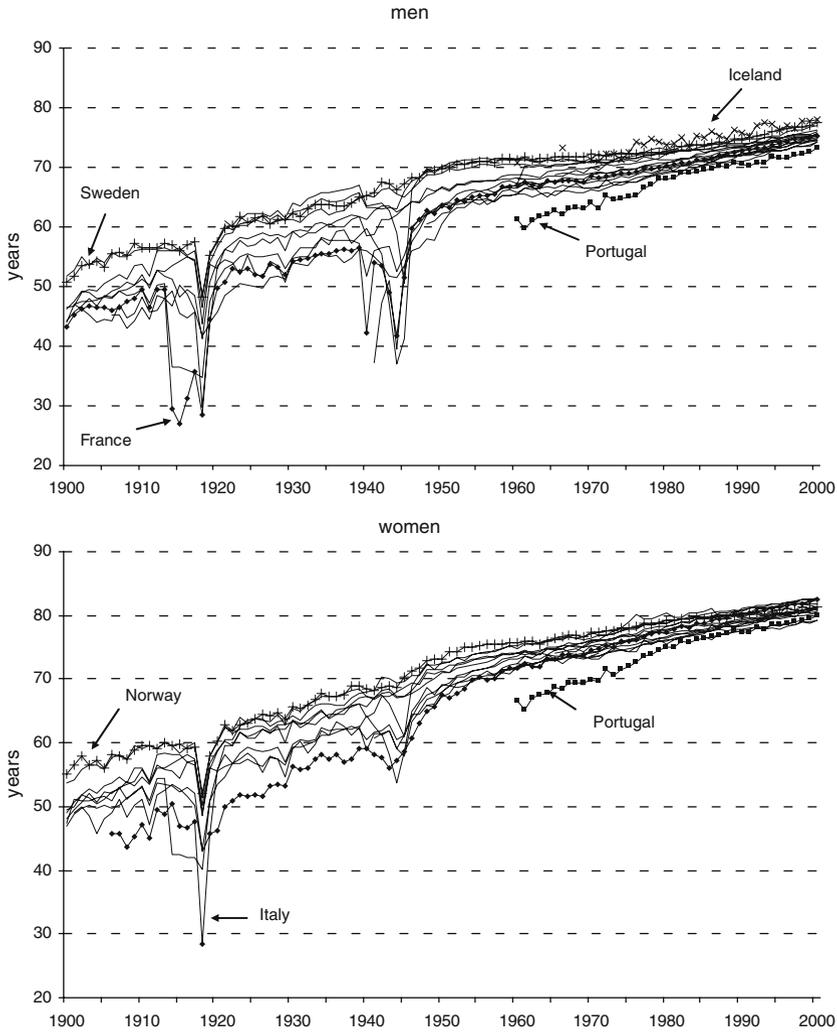
women

Fig. 5 Life expectancy at birth in 18 European countries: men and women

become smaller. The series vary a great deal in length across the countries. For 11 countries, we have estimated time series models of the ARCH type based on long series, most often for the period 1900–2000. In a second analysis, applied to all 18 countries, we used data for the period 1960–2000. Finally, we have applied a naïve model that assumes a constant decrease in age-specific death rates.

The time series models applied belong to the group of GARCH models: that is, models that are slightly more general than the ARCH models employed for the TFR. All models were estimated for men and women separately.

Let $e_{0,t}$ represent the life expectancy at birth in year t, and define $\nabla e_{0,t}$ as $e_{0,t} - e_{0,t-1}$. The model is:

$$\nabla e_{0,t} = C + \phi \nabla e_{0,t-1} + v_t + \sum_j \eta_j U_{j,t}$$

$$v_t = \psi_1 v_{t-1} + \psi_2 v_{t-2} + \cdots + \psi_m v_{t-m} + \varepsilon_t$$

$$\varepsilon_t = \left(\sqrt{h_t}\right) e_t, \text{ where } e_t \sim N(0,1), \text{ and} \tag{2}$$

$$h_t = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \gamma_j h_{t-j}$$

This is the AR(m)-GARCH(p,q) regression model. For the nine countries with long data series, the AR parameter m varied between zero (men and women in Denmark, men in Switzerland and women in England and Wales) and four (men and women in Italy). This parameter reflects the number of terms in the autoregressive expression for v_t. The maximum values of p and q, reflecting the number of moving average terms and autoregressive terms in the expression for h_t, were one (all cases, except French women, for whom it was zero) and two (men in Belgium, men and women in France) respectively.

The time series models indicate that between 2000 and 2050 life expectancy at birth for men and women is expected to rise by between six and 13 years. Across countries and sexes, the average annual increase amounts to 0.2 years. This is in line with historical developments. Long-range (50 years) 80% prediction intervals are 3–9 years wide, with women from England and Wales at the lower end of the spectrum, and Danish men and women at the upper end. Differences between predictions based on long and short time series appear to be small, particularly for men.

The naïve (constant-decline) model assumes that the rate of decline during the past 30–35 years for age-specific mortality rates (as long as it is not negative) observed in each country will continue in the coming 50 years. The result is an exponentially declining trend for age-specific mortality, for most ages, for all countries. This model predicts that between 2000 and 2050 life expectancy at birth for men will rise by well over four (Denmark) to almost 10 years (Finland and Germany). For women the future gains in longevity are generally expected to be slightly lower. The respective 80% prediction intervals are almost 11 years.

4.3 Migration

Net migration poses a greater challenge than total fertility or life expectancy, for two reasons:

1. the observed trends are strongly volatile, due to political and economic developments, and changes in legislation;
2. the data situation is problematic—time series of observed net migration are rather short, and the data quality may be questioned in some cases (see Sect. 2).

The variable of interest is the level of net immigration per 1,000 inhabitants (population 2000). Figure 6 plots this variable for the period 1960–2000. Compared to the other countries, Portugal experienced extraordinarily high levels of emigration between 1964 and 1973, mainly due to labour migration to other European countries. The fall of the Berlin Wall and the war in the former Yugoslavia led to large immigration flows into German-speaking countries in the 1990s.
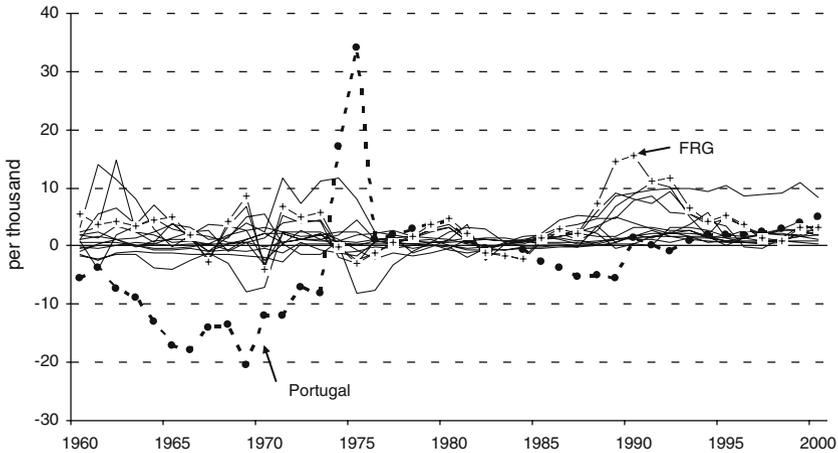
Fig. 6 Net migration to 18 European countries

We modelled net migration in three ways: as an autoregressive process, as a linear trend model and as a naïve model that assumes constant net migration. The predictions from the first two models indicate that the total net migration level in 2050 to the EEA+ countries may range between 600,000 and 2 million. Country-specific predictions for 2050 are generally between zero and 10 per 1,000. This is somewhat higher than the bands plotted in Fig. 6, because for many countries we identified a significant upward trend in net migration. The estimated trend is moderate for Denmark, Italy, Luxembourg, the Netherlands, Norway and Spain, while Finland, Greece, Portugal, and England and Wales show a strong trend. The autoregressive model led to reasonable 80% prediction intervals: between 2.4 (Denmark) and 14.1 (Luxembourg) promille points wide, although Portugal was the exception (33.9, due to a bad model fit).

Naïvely assuming constant net migration levels as from 2000 would result in a total net migration for the period 2000–2050 for the EEA+ of well over 57 million people. Ten years ahead 80% prediction intervals range between 2 (France) and 24 (Portugal) per 1,000 inhabitants (population 2000) under this model.

5 Expert views

The basic idea in the UPE project is that the past is the key source of information for the future. For the expected levels of mortality, fertility and international migration in about 50 years from now, as well for the assessment of the uncertainty, the experience of the past is analysed and used. The probability of events that have not yet occurred, however, cannot be based on an analysis of past events only. For example, the uncertainty of mortality forecasts depends partly on the probability of medical breakthroughs that may have a substantial impact on survival rates. An argument for and an assessment of the probability of the occurrence of such circumstances and/or events and their impact on demographic components are needed to determine the uncertainty of the forecast. Demographic experts may be

requested to point out these possibilities and assess how these factors and determinants influence the uncertainty of the future.

Following the statistical analyses described in Sects. 3 and 4, and after some exploratory work on the systematic eliciting of expert's opinions, a series of one-day, in-depth interviews was organized with four experts on European demographic developments: two on fertility, one on mortality and one on international migration.[6] We selected the experts based on two requirements. First, each should have sufficient knowledge of the relevant demographic developments in the 18 countries involved. Second, each should have a basic understanding of forecast uncertainty. These two requirements limited the group of potential experts considerably. Further practical aspects, such as time and budgetary constraints, resulted in our choice of four experts.

The purpose of the interviews was to obtain an independent assessment of future demographic trends and the associated uncertainty. We presented graphs, one for each of the 18 countries, to the experts. Life expectancy was used as the summary measure of mortality, TFR as the summary measure of fertility and net migration of migration. Each graph showed observed values for the years 1900–2000 (if available), two or three point forecasts up to 2050 and two or three prediction intervals. We formulated those forecasts based on the results of the time-series analyses and the analyses of historical forecast errors, but amended them in the light of demographic and non-demographic factors that were omitted from these analyses. The primary task of the experts was to suggest revisions to point forecasts and prediction intervals, to give arguments for the suggested revisions and to assess the uncertainty they could foresee for the future as compared to the past. Their role was solely advisory; they are in no way committed to the results of the UPE project.

The interviews started with a general question on the ideas or arguments of the experts concerning (qualitative) developments that they think are important for the future in their field of expertise. Subsequently, for each country we asked three specific questions:

1. Is one of the point forecasts OK? What are the arguments that justify the preference of one over the other? What are the arguments favouring some other alternative?
2. Is one of the widths of the intervals OK? What are the arguments that justify the preference of one interval over the other? What are the arguments favouring some other alternative?
3. Is the future more or less uncertain than the past? Why?

An example of the type of information submitted to the expert, before the interview took place, is given in Fig. 7. The figure shows the life expectancy of women in Austria for the period 1950–2000 and three different forecasts. Each forecast consists of a set of point predictions and 80% predictive intervals. One forecast is based on the GARCH time series model (see Sect. 4), another on the naïve model of constant reductions in mortality (see Sect. 4), while the third combines a GARCH-based point prediction with intervals derived from empirical errors. The procedures for deriving the second and third types of interval are described in Appendix 2.

---

[6] One interview was done by email.
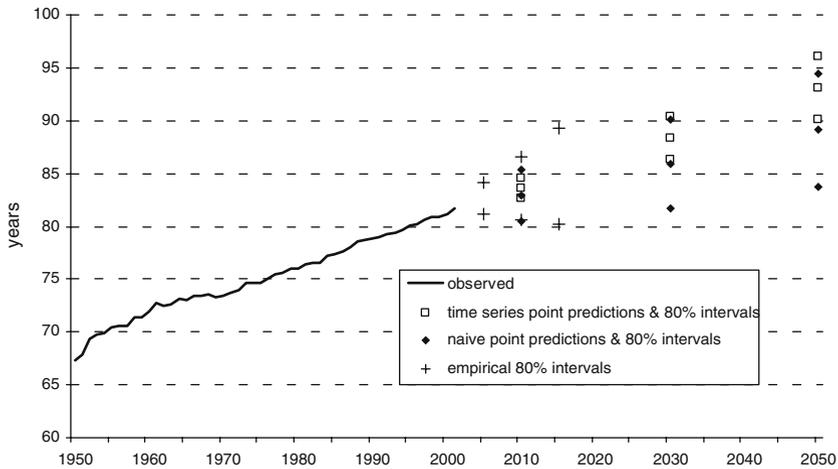
Fig. 7 Life expectancy of women in Austria

The experts either gave their own point forecasts or chose one of the alternatives presented to them in the material. The experts on mortality and migration gave 80% prediction intervals around these forecasts, based on their insights into future as compared to past uncertainty. The first fertility expert labelled his upper and lower limits for future fertility as 'expert margins', which in his view do not represent any level of uncertainty. The second fertility expert gave his views on the proposed point forecast, prediction intervals and future as compared to past uncertainty.

The experts provided numerous useful justifications and insights with regard to the most likely demographic future developments and the uncertainty around these trends. Here we give a few examples.

5.1 Mortality

The mortality expert expressed the following views regarding current and future developments in mortality:

1. The improvement in age-specific mortality has gradually shifted from young ages to older ages. During the past decade, an acceleration of decline (especially in ages 80–100) has been observed in several countries, notably France, Germany, Italy, Japan, Switzerland and Spain. However, in some other countries, such as Denmark, Finland, Norway, the Netherlands and the UK, the progress has been slow.
2. For females, the best-practice value of life expectancy has increased, by 0.25 years per calendar year in the past 160 years. It is not likely that life expectancy in EEA+ countries will permanently increase at a slower pace. Corrective action would be taken by the government if a country began to fall too far behind. An example of this is Denmark, where committees have been appointed to investigate means of reducing dangerous behaviour (e.g. smoking and alcohol consumption, both of which can be influenced by education and regulation) and the inadequacy of past health investment. So, reductions in the prevalence of smoking, say, are expected to have a rapid effect on

cardio-vascular morbidity, a major cause of death. For other diseases, such as lung cancer, the long period of latency will attenuate the effect of behavioural changes.

3. Life expectancy of individual countries has sometimes increased faster than the best-practice life expectancy and sometimes slower. Countries close to the best-practice level are expected to have a slightly slower increase, while those further away from the best-practice level are expected to have a slightly faster increase. Thus we can assume some degree of convergence in life expectancies across countries.

4. The empirically observed level of *average* uncertainty in Europe, which includes the effects of wars, epidemics, penicillin, etc., is appropriate for the downside or lower limit of the prediction interval. However, possible future medical advances may bring unexpected gains in life expectancy. Examples include the cure of cancer, the prevention of Alzheimer's disease, improvements in cardio-vascular health through the rejuvenation of the heart via stem cell therapy, and improvements of pharmaceuticals based on genetic understanding. Some even consider the possibility of slowing down the pace of ageing feasible. The effect of possible acceleration in biomedical technologies is not reflected in past developments. Thus the upper limit should be about twice as far from the median as the lower limit, or 11 years above the median.

5.2 Fertility

The first fertility expert provided the following list of key factors in future reproductive behaviour:

1. Postponement of childbirth, followed by a later catching up at a higher age, is the most important direct determinant of fertility developments. Postponement behaviour is clear and universal in Europe, but this is not the case for catching-up behaviour.

2. There is a north–south divide in Europe. The north, and especially the Scandinavian countries, was the forerunner. North European countries were the first to postpone childbearing (visible in the data from the early 1970s) and the first to recuperate. In the German-speaking countries and in the south of Europe there was postponement too, but there was a much weaker recuperation, if it is there at all (at least visible in the data we have so far).

3. A number of explanatory factors account for the new pattern of family formation and for concomitant postponement. The general ones are:

- increased female education and female economic autonomy;
- rising and high consumption aspirations that created the need for a second income in households and also fostered female labour participation;
- increased investments in career developments by both sexes, in tandem with increased competition in the workplace;
- rising 'post-materialist' traits such as self-actualization, ethical autonomy, freedom of choice and tolerance for the non-conventional;
- a stronger focus on the quality of life, with an increasing appetite for leisure as well;
- a retreat from irreversible commitments and a desire for maintaining an 'open future';

- finally, rising probabilities of separation and divorce, and hence a more cautious 'investment in identity'.

One of the fertility experts had problems with the fact that statistical models were chosen which did not include our *present* knowledge of key factors determining fertility levels. According to him, variances based on historical forecasts cannot be used for prediction intervals of expected futures. The other fertility expert, on the other hand, was sure that the past is a good guide to assessing future uncertainty, and that a volatile past is a good predictor of a volatile future in fertility levels, as long as sensible models were used in the past for forecasting and that present knowledge is incorporated.

5.3 Migration

The expert pointed out that in general and for the EEA+ as a whole, the future is less uncertain than the past for migration, because experience has taught us that sharp changes in net migration tend to fade out fairly soon. He provided the following principal factors for determining migration developments in the coming 50 years:

1. The economic developments in countries of the EEA+, and in the EEA+ area as a whole, are the most important condition or determinant driving migration. If the economic engine starts rolling again—and the recession is short and/or over soon—demand for labour will rise. The national economies in many countries cannot supply all the demand for labour. People will come first from other EEA+ countries, but also and primarily from outside the present EEA+ to fill the gaps or seize opportunities that are there. However, demand will not be met completely, because rigid economies and wage systems will keep unemployment high. Business cycles will lead to fluctuations in migration flows.
2. The ageing of the EEA+ population is the second important force that induces a demand for labour migrants.
3. Developments in the global south and east will continue to put (enormous) pressures on the gates of the wealthy EEA+.
4. The expansion of the EU by 10 countries will have a temporary effect (immigration boom, fading out, followed by return).
5. Historical ties and destinations will keep their relevance when living conditions can be improved by moving abroad. Examples are UK migration to Australia, USA and Canada, and southern Europe (the last group for the wealthy and healthy).

6 Synthesis

6.1 General issues

Most demographic developments start smoothly, last for a long time and therefore evolve gradually. Principal trends such as declining family size, increasing childlessness, later motherhood, increasing life expectancy and net immigration levels can

easily last five decades or more. However, there were and will be turning points. In addition, sudden trend shifts, short periods of acceleration or deceleration, and incidental distortions have been observed or may arise due to a significant change of 'environmental' circumstances, including the introduction of new and more effective means of planning and control (e.g. the pill, medicines and therapies to combat major diseases, more restrictive legislation concerning asylum seekers).

In addition to the time dimension, most demographic trends have a spatial dimension. For example, the trend towards later motherhood started in Scandinavia (females born in and around 1942), spread rapidly to western and central Europe, and reached southern Europe for women born in the 1950s and 1960s. Due to various political, cultural and economic factors, considerable international differences still exist with regard to fertility, mortality and migration levels and patterns. However, the overall trend within the group of 18 countries considered is not one of divergence—in fact, is sometimes even one of convergence. Below we will deal with the variability in fertility, mortality and migration across countries, as opposed to the variations over time in previous sections. Figure 8 gives the variability in the TFR across countries for each year from 1900 to 2000. It illustrates that fertility has had a few periods of short-lived divergence, but that the overall pattern is stable.[7] For cohort fertility, there is a clear tendency towards convergence for women born since 1945, but from the 1960 generation differences across countries do not diminish any more (see Fig. 9). International differences in life expectancy at birth have become smaller, although for women the trend has stabilized in recent years (see Fig. 10). For the remaining life expectancy at age 60 in the old EU15, the international differences are stable for women from 1970; for men they have decreased since that year (see Fig. 11). Finally, only a few countries among our group of 18 have experienced an emigration surplus in recent years; in the 1950s and 1960s there were many more.

Will the trends towards convergence between countries continue? In other words, can we expect demographic continuities in the short, medium or long run, or are there strong reasons for assuming discontinuities, leading to new, reversed trends? We have assumed that current trends in the demographic indicators we have analysed, including the trends towards stable or smaller differences between countries, will last for a few decades more. However, as in the past, short periods of accelerating, stagnating or even reversing trends may occur. These discontinuities or changes in the speed of a trend are not predictable and are therefore treated as random fluctuations around an expected value or median value.

In Table 1, we have summarized the principal assumptions concerning future fertility, mortality and migration trends and patterns that we have adopted. We based these assumptions on information from our three sources: the analysis of historical forecast errors, time series predictions and the views of the experts. The table shows, for the year 2049, national point forecasts and 80% prediction intervals for the TRF, life expectancy at birth and crude net migration rate (expressed per 1,000 of population in 2000). In the following sections we explain the reasoning behind these assumptions.

---

[7] *Relative* differences in TFR across countries have become larger since the mid-1960s. Around 2000, the coefficient of variation in the log of the TFR was 0.30–0.35, about twice as much as in the 1960s.
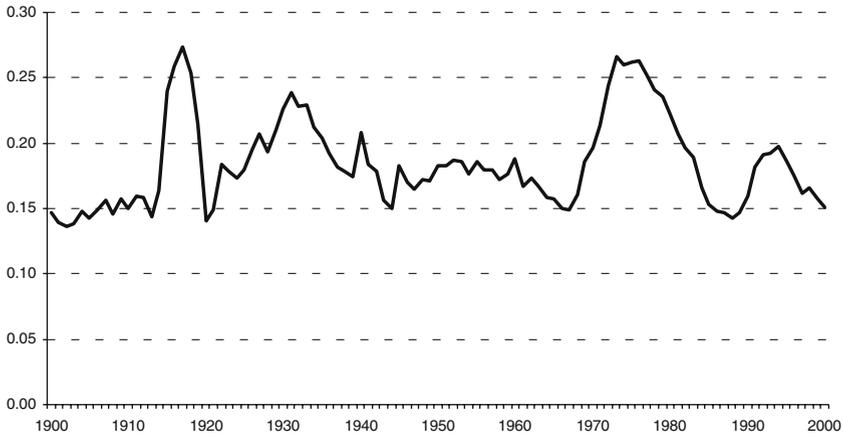
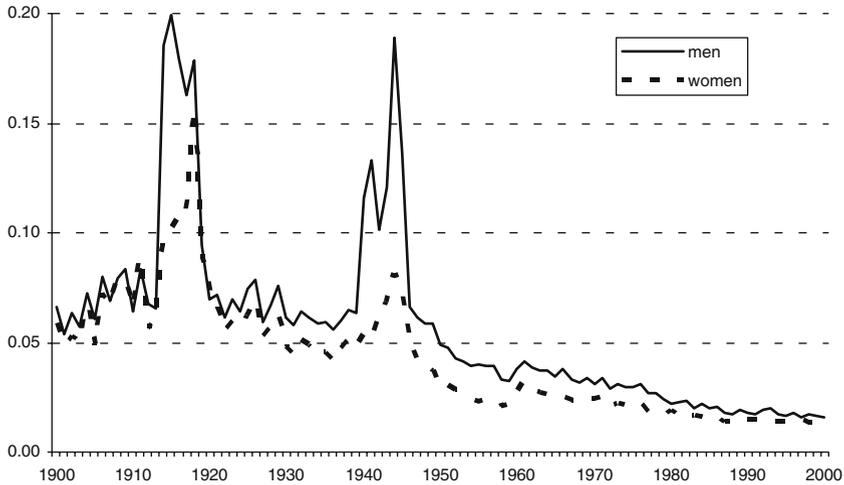**Fig. 8** Coefficient of variation in TFR across 18 European countries

Fig. 9 Coefficient of variation in Completed Cohort Fertility across 18 European countries

6.2 Fertility

Past trends, contemporary levels and recent explanatory research indicate that there is a clear geographic divide in fertility levels in Europe. The northern and western EEA+ countries are experiencing levels of the completed fertility rate (CFR) and TFR of about 1.8 children per woman, whereas the Mediterranean and German-speaking countries are moving towards historically low levels of around 1.4 children per woman.

The northern and western cluster of countries comprises Belgium, Denmark, Finland, France, Iceland, Ireland, Luxembourg, the Netherlands, Norway, Sweden and the UK. The Mediterranean and German-speaking cluster consists of Austria, Germany, Greece, Italy, Spain and Switzerland. Portugal is the only EEA+ country that cannot easily be classified: its fertility trends and levels are somewhere in the middle.

For the period 2003–2049, it is assumed that these clusters will remain. The northern and western EEA+ countries will continue to achieve a TFR level of 1.8

Fig. 10 Coefficient of variation in life expectancy at birth across 18 European countries
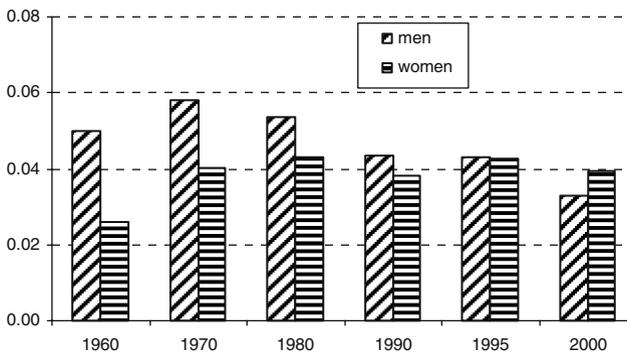
Fig. 11 Coefficient of variation in remaining life expectancy at age 60 across EU15 countries

children per woman. The TFR in Portugal will rise to a level of 1.6 children per
woman, whereas the TFR of other EEA+ countries will persist or slightly increase to
a level of 1.4 children per woman. This gives a coefficient of variation in 2049 of 0.11,
slightly lower than the current value.[8] The 80% intervals in 2049 range from about
1.1 to 2.8 children per woman for the northern cluster and from 0.9 to 2.2 children
per woman for the other cluster. With respect to the timing of fertility, we assumed
that the mean age at motherhood on a period basis would continue to increase in all
countries and eventually converge at a level of 31 years, to be reached by the year
2017.

These key assumptions have been made for the following reasons. The northern
and western countries were the first where women delayed childbearing and the first
where they showed catching-up behaviour. In the southern European and German-
speaking countries, there was also a delay but much less catching-up behaviour. In

---

[8] The coefficient of variation in the log of the TFR is 0.23 in 2049, which is lower than the value in
2000 (0.33): see footnote 7.

Table 1 Summary of assumptions for the TFR, life expectancy at birth and net migration in 18 European countries: point forecasts and limits of 80% prediction intervals in 2049

	TFR			Net migration (per 1,000 population in 2000)		
	Point forecast	Lower limit	Upper limit	Point forecast	Lower limit	Upper limit
Austria	1.40	0.89	2.20	3.5	–1.0	8.0
Belgium	1.80	1.14	2.84	2.0	–0.6	4.6
Denmark	1.80	1.15	2.82	2.0	–0.6	4.6
Finland	1.80	1.15	2.82	1.5	–1.1	4.1
France	1.80	1.15	2.83	1.5	–3.0	6.0
Germany	1.40	0.88	2.21	3.5	–1.0	8.0
Greece	1.40	0.90	2.18	4.5	–3.2	12.2
Iceland	1.80	1.14	2.85	1.5	–3.6	6.6
Ireland	1.80	1.15	2.83	3.5	–2.3	9.3
Italy	1.40	0.89	2.20	4.5	–1.3	10.3
Luxembourg	1.80	1.14	2.84	6.0	–1.7	13.7
Netherlands	1.80	1.15	2.82	3.0	0.4	5.6
Norway	1.80	1.16	2.80	3.5	0.9	6.1
Portugal	1.60	1.02	2.51	4.5	–3.2	12.2
Spain	1.40	0.89	2.21	4.5	–1.3	10.3
Sweden	1.80	1.12	2.89	3.0	0.4	5.6
Switzerland	1.40	0.90	2.18	3.5	0.9	6.1
UK	1.80	1.16	2.80	3.5	–1.0	8.0
	Life expectancy at birth, males			Life expectancy at birth, females		
	Point forecast	Lower limit	Upper limit	Point forecast	Lower limit	Upper limit
Austria	84.4	80.3	88.8	88.7	85.1	92.5
Belgium	84.2	79.4	89.2	88.3	84.1	92.9
Denmark	83.2	78.3	88.3	87.3	82.5	92.4
Finland	84.7	80.0	89.4	88.7	84.9	93.4
France	85.5	80.6	90.6	89.7	85.5	94.1
Germany	84.9	79.8	90.5	89.1	84.7	94.0
Greece	82.8	78.2	87.2	86.9	83.1	91.0
Iceland	85.9	81.8	90.2	89.9	85.1	95.7
Ireland	84.7	80.1	89.6	89.9	85.5	95.1
Italy	85.7	81.4	90.4	89.8	85.8	94.3
Luxembourg	85.2	79.9	91.8	89.4	84.7	95.3
Netherlands	82.5	78.1	87.1	86.4	82.4	91.0
Norway	83.7	79.3	88.2	87.9	83.8	92.2
Portugal	84.2	79.1	89.6	88.4	84.1	93.3
Spain	85.9	81.1	91.4	90.1	85.9	94.9
Sweden	84.7	80.3	89.4	88.7	84.2	94.3
Switzerland	85.3	81.1	89.6	89.4	85.7	93.8
UK	83.4	78.7	88.3	87.5	83.3	92.2

the latter countries, one-child families became quite popular, and we expect that this will remain the case, due to relatively poor childcare and housing facilities.

These key assumptions on fertility are based largely on time series models applied to long series of observations. We used the experts' views and reduced the prediction intervals based on time series models and past forecast errors for the short and medium term. The main reason is that the models were not applied in such a way as to take into account the relatively low volatility of the TFR during the last one or

two decades. Therefore, the 80% prediction intervals for the short and medium term are expected to be smaller than those predicted by the models. The naïve model gave us a standard deviation of the relative error in the TFR in 2050 equal to about 0.35 children per woman. The point predictions differ between 1.4 and 1.8, and hence the widths of the 80% prediction intervals differ between 1.3 and 1.7 children per woman. Based on historical estimates from countries with long series, we concluded that the TFR behaved essentially as a random walk (i.e. a process with independent increments). PEP parameters for autocorrelation in error increments were set accordingly.

6.3 Mortality

The assumptions for mortality were based largely on an extrapolation of age-specific mortality rates. For each of the 18 countries we assumed that eventually the rate of improvement of mortality rates will converge towards a common rate of decline. The decline starts from recent country-specific values and changes in a linear fashion over time towards the common rate of decline, which is to occur by the year 2030. The eventual rate of decline was empirically estimated from data from Austria, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland and the UK during the latest 30-year period observed. In some countries, the extrapolation procedure would imply diverging developments of male and female life expectancies. This is in contrast with observations in the last two or three decades. It seems plausible to assume that the gender gap in life expectancy will continue to decline as differences between men and women in lifestyle habits (e.g. smoking) become smaller. For this reason, we made a proportional adjustment such that the gender gap equals 4 years in the target year. In the case of Ireland alone, the gender gap is assumed to equal 5 years due to strongly diverging trends in the recent past.

The basic assumption of ongoing international convergence in mortality improvement implies that we expect that in countries with an exceptionally fast rate of decline in the past, the rate of decline will slow down to some extent. On the other hand, in countries with a modest rate of improvement in the past the decline is expected to catch up with the European average rate of improvement. There are several reasons to justify the extrapolation procedure described above. The most important is that it takes into account the country-specific developments. Developments differ strongly between the European countries. For most countries, there are no reasons to believe that these developments will reverse in the next few years. For the long run, however, countries adjust to the global European trend. This 'global' European trend incorporates all structural improvements that have been achieved in mortality. These assumptions imply that, especially for males, the differences between countries are becoming smaller. In 2002, the difference between the country with the lowest male life expectancy (Portugal) and the highest (Iceland) was about 4.7 years. It is assumed that this difference will decrease to 3.3 years (lowest for the Netherlands and highest for Spain). For females, differences are decreasing only slightly. The international differences in life expectancy in 2049 imply a coefficient of variation of 0.012, for both men and women. This is in line with the historical trend (see Fig. 10).

The resulting expected gains in life expectancy at birth for men during the period 2002–2049 vary between 6.5 (the Netherlands) to well over 10 years (Luxembourg, Portugal and Spain). For women slightly smaller improvements are expected, varying from 5.7 (the Netherlands) to 9.6 (Ireland).

The 80% prediction intervals are about 50% wider than the model-based intervals. This is mainly because of the views of the expert, who stated that it is not unlikely that unprecedented medical breakthroughs will happen. The assumed 80% intervals in 2049 range from 7.4 years for Austrian females to almost 12 years for males in Luxembourg. As explained by Alho (2005), we assumed a value of 0.05 for the autocorrelation in the error increments of the death rates, independent of age or sex.

6.4 International migration

Forecasting international migration was seriously hampered by the data situation. Available international time series are rather short and in some cases of poor quality. This implies that, more than for fertility and mortality, expert knowledge will need to be involved and that prediction intervals will be wide.

We took the results from the linear trend model applied to all 18 countries as a starting point and adjusted this initial assumption downwards using qualitative arguments. Finally, we incorporated country-specific differences.

For many countries, the time-series analysis indicated a significant upward linear trend in net migration. The linear trend model applied to the total of the 18 countries results in an increase of scaled net migration from almost 3 per 1,000 in 2000 to more than 5 per 1,000 in 2050.[9] This level is even higher than the peak that was reached in 1992 of 3.6 per 1,000. However, it does not seem plausible to assume such a continuation of the linear trend. Part of it is due to the increase in the number of asylum seekers in the 1980s and 1990s. In most recent years the number of asylum seekers has been much lower than it was around 1992, which is partly due to restrictive migration policies in some countries. Although it is not unlikely that the numbers of asylum seekers observed in 1992 will be seen again in the future, it does not seem very plausible that structurally higher levels will be reached. This can be affected by moves to deal with refugees closer to their countries of origin, discussions about EU-regulated asylum policies (quotas) and the rather abrupt changes in attitude, and accompanying unprecedentedly restrictive policies, to asylum migration in countries like Denmark and the Netherlands.

With respect to labour migration, the ageing problem is often mentioned as a pull factor for migration. Some developments may, however, temper this phenomenon, such as the increasing participation of women and minority populations in the labour force and the export of labour. Moreover, in some sending countries, such as the new EU member states, ageing will be even more problematic than in the EEA. Next to asylum and labour migration, family-related migration is a major source of migration. Family reunification and family formation are important motives for

---

[9] Net migration for the aggregate of the 18 countries was computed, for each year, as the sum of net migration in absolute numbers (and thus unscaled), relative to total population in the EEA+ in the year 2000. This aggregate net migration masks gross flows between the countries of the EEA+. It only reflects the net effect of immigration to and emigration from the EEA+ as a whole. There are no reliable data on gross flows between the countries. Thus we are not able to say how much of the gross flows is masked as a result of this aggregation across 18 countries.

immigrants to enter the EEA. As migrant populations are growing in the countries of the EEA, family-related migration is expected to remain important.

Based on these observations, it is assumed that scaled net migration for the total EEA will increase, but not as much as estimated by the linear trend model. Instead, a target level of 3.5 per 1,000 is assumed in 2049. This level takes into account some of the trend that is observed in historic data and is almost equal to the historic maximum that was reached in 1992.

The target level is used as a starting point for country-specific assumptions. Three, not very strict, clusters of countries are distinguished:

countries below average: Belgium, Denmark, Finland, France and Iceland
countries close to average: Austria, Germany, Ireland, the Netherlands, Norway, Sweden, Switzerland and the UK
countries above average: Greece, Italy, Luxembourg, Portugal and Spain

Migration levels close to the average are motivated as follows. Net migration in Austria is assumed to increase to the average level. As noted by the migration expert, Austria was the gateway to Europe from the east in 1990s. As a result, it has a large foreign population which can attract new migrants. The high flows of *Aussiedler* and refugees, which made Germany the most important receiving European country around 1990, are probably over. Moreover, labour migration from central and eastern Europe is more balanced nowadays. The relatively high unemployment has a negative impact on net migration. For Ireland the future depends on whether the Celtic Tiger boom continues or collapses. It is assumed that net migration will level off to the average. In the Netherlands the economic situation, together with the restrictive policies in recent years, has led to decreasing migration numbers. In 2003 net migration was even negative. However, the Netherlands will remain attractive to immigrants, due to the large migrant populations already there. Even so, a slightly lower level than the average is assumed (3 per 1,000), which is partly due to limited absorbing capacity given the high population density. Norway and Sweden are still relatively generous in admission of asylum seekers, which will probably continue. Since the migration expert foresees more future restrictions in Sweden, a slightly lower target level is assumed for this country. Switzerland shows less enthusiasm for foreigners at present and will try to keep net migration below the high levels that they have experienced in the past. The UK has become a country of immigration and will probably stay that way. Asylum seekers are expected to continue arriving and the labour market is easy to enter. According to the expert, there will not be an increasing level of migration but rather a continuing high level. Currently the level is about 3 per 1,000, which is assumed to rise to a level of about 3.5 per 1,000.

The assumptions for countries with target levels below average are motivated as follows. In Belgium net migration has been structurally lower than in neighbouring countries such as Germany and the Netherlands. However, an increase is foreseen, partly because of the important flow of labour elite migration focused on its EU role. Although an increase will be assumed, the target level (2 per 1,000) will not be as high as for Germany and the Netherlands. Also for Denmark a level of 2 per 1,000 is assumed. Denmark admitted a lot of asylum seekers around 1995, but there has been a very clear backlash in recent years. Moreover, the observed levels are generally lower than in countries that are assumed to move towards the European average.

This also applies for Finland, which has experienced relatively low levels of net migration in the past. Since the expert does not foresee large flows from Russia (except from some Estonians), a level of 1.5 per 1,000 is assumed, which is still well above recent levels. France is one of the countries for which the data quality was questioned by the migration expert. This probably has to do with the way France treats the francophone migrants in the statistics. Still, a rather low level of 1.5 per 1,000 is assumed. The continuing high unemployment discourages immigration. Moreover, since ageing is less than in countries such as Germany, the demand for foreign labour is assumed to be less than in Germany. Net migration in Iceland is very volatile. One of the key issues is that net migration is highly influenced by the US military base in this country.

For the southern European countries and Luxembourg, future net migration is assumed to be higher than average. For all these countries but Luxembourg, there are serious data problems which hamper proper forecasting. Italy is one of the gateways to Europe for migrants from Africa and the Balkans (in particular Albania). At present, it is unclear whether these migrants stay in Italy or move northwards. Italy seems rather relaxed about the inflows of migrants. Portugal, on the other hand, is the gateway for migrants from countries like Brazil, Angola and Mozambique. Spain has recently been confronted with massive immigration flows from Latin America. It is assumed that the southern European countries will remain the main gateway to Europe, irrespective whether migrants move on to the north. A target level of 4.5 per 1,000 is assumed. Luxembourg, currently by far the most affluent EU country, is a special case, with very high net migration levels and a large non-native population. It is assumed that the target level is higher than in the southern countries: 6 per 1,000.

With respect to the 80% prediction intervals, we took the results from the autoregressive time series model as the starting point. We reduced these intervals for countries with good registrations. This implies that intervals are smallest in the Nordic countries and broadest in the southern European countries (see Table 1). To make consistent assumptions, we clustered the 18 countries. As to autocorrelations, these differed between countries between 0.13 and 0.56, with a median value across the 18 countries of 0.22.

6.5 Age patterns

The previous three sections discussed the levels of fertility (TFR), mortality (life expectancy at birth) and migration (net migration). Point forecasts for fertility rates by age of the mother, mortality rates by age and sex, and numbers of net migration by age and sex were obtained as follows.

For fertility, empirical fertility rates in 1-year age groups for the year 2002 were smoothed across ages and next extrapolated to 2050 with two constraints. First, their sum had to be equal to 1.4, 1.6 or 1.8, depending on the regional cluster the country belongs to. Second, the mean age at motherhood on a period basis would reach a level of 31 years by the year 2017.

For mortality, a jump-off value of age- and sex-specific mortality was established by smoothing the observed values of years 1998–2002 and adjusting for increase during the period to match the level of 2002. For all countries, rates for ages 95+ were computed using information from younger ages. For Germany, Greece,

Portugal and the UK, additional special estimations were made to establish starting values for the highest ages. The point forecast for age- and sex-specific mortality was calculated by starting from the jump-off value and applying an age-specific rate of decline during years $t = 2002\text{–}3$, 2003–2004, ... , 2048–2049, to the value obtained until then. A country-specific initial rate of decline was estimated and a linear change towards the eventual rate of decline, common for all countries, to occur by the year 2030. The initial rate of decline was empirically estimated from years 1993–2002. The values were constrained to be non-negative and smoothed separately for males and females before use.

The age structure of net migration was assumed to start from a national pattern estimated from data in 1990–2000, and to change linearly to an average pattern, estimated from Austria, Belgium, Denmark, Iceland, the Netherlands, Norway, Sweden and Switzerland, after 10 years and then held constant for the rest of the forecast period. For Greece the average pattern was used from the start.

6.6 Cross-national correlations

We estimated cross-national correlations from correlation patterns in historical forecast errors and from the residuals of the time series models. We used an eigenvalue analysis (factor analysis) for the correlation matrices relating to the errors in total fertility and the life expectancy at birth, and to observed net migration. The analysis suggested for fertility a contrast between the Mediterranean countries (Greece, Italy, Portugal and Spain) and the other countries. For mortality, we found two groups of countries: Portugal and Spain, then all the other countries. The factor analysis for net migration resulted in three regions: one consisting of Austria, Germany and Switzerland; a second consisting of Greece, Italy, Portugal and Spain; and a third consisting of the remaining countries. Alho (2005) gives more details of the cross-national correlations. These correlations are relevant for the results published for the EEA+ as a whole, not for the forecasts of the individual countries.

7 Conclusions

The UPE population forecasts by sex and age differ significantly from previous sets of population projections compiled by Eurostat and the UN, and from national population forecasts produced by national statistical agencies in terms of both how the most likely future demographic development is assessed and how the uncertainty of forecasting is taken into account.

Although national population forecasters typically and increasingly do assess trends in other countries, recent past developments in the country in question still receive heavy attention. While this may improve accuracy in the very short term, in the longer run diverging trends lead to large differences in the demographic outlook that are incompatible with the shared economic, cultural and social norms among the 18 EEA+ countries considered. The UPE project attempted to acknowledge the recent developments in formulating the most likely future development for the first few forecast years. However, eventually, and in particular for mortality, the demographic developments were assumed to conform to average trends of the area. This does not mean that a strong convergence hypothesis has been imposed, but it

keeps the otherwise divergent trends in check. This corresponds to what Eurostat has applied during the compilation of the 1990-based, 1995-based and 1999-based long-term national population scenarios, and our experience suggests that this practice should be continued.

However, our assessment of the most likely future trends differs from the past practice of Eurostat and the UN, along with many national statistical agencies. A key question regarding fertility is whether the low levels of the past two decades in the Mediterranean and German-speaking countries will continue, or whether this is a temporary phenomenon related to the timing of births. Along with Eurostat, and as opposed to the UN, the UPE team concluded that while some recuperation is likely, there is no evidence that fertility will rise significantly from the current levels. Although the current levels are the lowest in recorded history, the causes of the decline are poorly understood and we cannot rule out the possibility that there will be further declines. Therefore the UPE team expects that the TFR will most likely remain close to recently observed levels and the average age at motherhood will increase further.

As regards mortality, the UPE project shows that virtually all official national and international population forecasts over the past four to five decades have considerably underestimated the gain in life expectancy at birth. Most demographic forecasters simply did not or could not believe that the decline in age-specific mortality would persist. Therefore they generally expected a slowdown of the improvement in life expectancy, eventually leading to stagnation. This erroneous assumption has led to a systematic under-estimation of the surviving populations, especially in the oldest ages. The UPE team expects that it is more likely that current rates of decline will continue, thus leading to a larger future population than predicted by the official agencies. It also notes that even more optimistic forecasts would be obtained if, instead of age-specific mortality, life expectancy were to be taken as the variable to be predicted.

As regards migration, we can draw similar conclusions. Net migration flows have been continuously under-estimated. In addition, recent forecasts by Eurostat, the UN and several national agencies assume moderate levels of future net migration. In contrast to mortality, this is a more recent phenomenon, covering the past two decades or so. For a number of countries, the migration data are of much lower quality than data on fertility or mortality, so an assessment of past trends is on weaker ground. The UPE team assumes that the level of migration, primarily from outside the EEA+, will exceed the current levels to some extent. However, we have not simply assumed that the observed increasing trend will continue. Instead, country-specific target levels of migration have been specified on a judgemental basis. The consequence is that our forecasts of net migration are considerably higher than those made by official agencies.

The high assumptions for mortality and international migration imply that the UPE forecast predicts larger numbers of working-age populations in European countries, and clearly larger numbers of the elderly, than the most recent forecasts by the UN and Eurostat. Fertility is slightly lower, but the net result is that the UPE forecast predicts a population decline that comes later than the declines predicted by the UN or Eurostat (see Alho, Cruijsen, & Keilman, 2006). UPE expects a modest annual growth rate of 0.2% for the population in the EEA+ countries in the years 2003–2050. The 2004 revision by the UN predicts that the population in the EEA+ countries will decrease in the years 2030–2050 from 407 million to 400 million, after

an initial increase from the current level of 392 million (United Nations, 2005). In contrast, the UPE forecast anticipates 427 million inhabitants by 2050. Eurostat (2005b) predicts that the 15 member countries of the former EU will have a population of 384 million in 2050, 7% less than the UPE prediction for these countries.

Increased immigration and increased life expectancies imply that the number of people aged 65 and over comes out rather high in the UPE forecast. There are large differences between UPE and UN forecasts in terms of assumed life expectancy for a number of countries, including France, Italy and Germany. For these three countries, the UN expects 17.11 million, 18.09 million and 22.38 million people aged 65+ in 2050, respectively. Equivalent expected values from the UPE forecast are much higher: 18.20 million, 19.29 million and 25.02 million. These high UPE numbers will clearly have implications for the welfare system, including state old-age pensions and health care systems.

Past population scenarios by Eurostat and the UN, together with forecasts of most national statistical agencies, have tried to handle the uncertainty of forecasting by presenting alternative variants. Although this approach can be helpful in some planning connections, these variants do not give a logically consistent description of forecast uncertainty. The UPE project has used a stochastic approach instead. In this approach, the forecaster recognizes that the most likely future development, or the point forecast, is not likely to be correct, and uses probability theory to describe the level of uncertainty around the most likely development. A probability distribution incorporating these two components is called a predictive distribution. In theory, it has been known how to formulate a predictive distribution for 50 years or so, but for both technical and substantive reasons, it has only been possible to produce stochastic forecasts of the type considered here until recently. The phenomenal increase in the speed of computing has largely removed the technical obstacles during the past decade.

A final conclusion is that the parameter values of the predictive distributions of future fertility, mortality and migration can be successfully derived from a methodology that combines the findings of three existing methods: analysis of observed errors in past forecasts, model-based estimates of forecast errors and the eliciting of expert opinions. Earlier studies on stochastic population forecasting have relied heavily on only one of the methods mentioned. The UPE project has demonstrated that by means of an overarching argument-based approach, the outcomes of the three methods can be applied for assumption-making. A creative mixture of both simple and advanced time series models, estimation techniques and expert knowledge can solve problems caused by the limited availability of historical population forecasts and a general lack of reliable, internationally comparable data series on international migration.

Appendix 1. The scaled model of error and PEP

This appendix gives an informal explanation of the main characteristics of the scaled model of error, as it is implemented in the computer program Program for Error Propagation (PEP). Technical descriptions of a general nature can be found in Alho and Spencer (1997, 2005), while Alho et al. (2006) give statistical details for the current application. The PEP manual, which is available at http://joyx.joensuu.fi/~ek/pep/pep.htm#PEP, contains a technical description of the scaled model of error. Finally, http://joyx.joensuu.fi/~ek/pep/peptutorial.pdf gives a tutorial for PEP.

Conventional population forecasts are frequently made by means of the cohort-component approach (e.g. Shryock, Siegel, & Associates, 1976). The approach starts from a recently observed table of the population by age and sex, called the jump-off population. The method updates this table for future years by repeatedly applying assumed death rates for each combination of age and sex, and birth rates specific for ages of the women in childbearing years. International migration can be handled in a number of ways. The simplest is to add, for every year in the future, net-migration numbers broken down by age and sex to the survivors of the population computed for the previous year.

Stochastic (or probabilistic) forecasts are carried out in a similar way, but now future fertility and mortality rates and numbers for net migration are considered as random variables (e.g. Alho & Spencer, 2005). Their distribution can be specified in alternative ways. The scaled model of error assumes that the rates are normal in the log scale and net-migration numbers are normal in the original scale. The normal distribution requires that one specifies the mean of the distribution as a measure of location, and the standard deviation as a measure of spread (or scale) around the mean, to reflect forecast uncertainty.

In PEP, the variances of the logarithms of the future age-specific fertility and mortality rates are represented in terms of the following model.

Mortality and fertility

Suppose that the true age-specific rate in age x during forecast year $t > 0$ consists of a product of two terms: a point forecast for that rate and a relative error. In the log scale, the error in year t and age x consists of a sum (over the years 1, 2, ... , t) of error increments, which have the expectation zero. The error increments consist of two random parts. One is specific for age but independent of time, while the other one depends on age and time. Their sum is scaled up or down by a deterministic scale specific for age and time, hence the name *scaled model*. The model assumes that for each age/time combination, the variances and covariances of the two random error parts of each error increment are such that the variance of the error increment itself is equal to the scale.

In the UPE application, we had to specify the correlation of the age-specific part of the error increment. It represents the correlation in age-specific mortality and fertility across age. By the nature of the model, the autocorrelation between error increments (for a given age) in the time dimension is constant (i.e. independent of the length of the time lag). Under a random walk model, the error increments would be uncorrelated, for example. Together, the autocorrelation for a given age and the

scale for a given age and time determine the variance of the relative error. The correlation of errors across components is assumed to be zero in the model.

International migration

The representation of error in net migration in PEP is done in absolute terms, using variables of the same type as the error increments described above, but now the error increments depend on time and sex. Dependence on age is not stochastic, but assumed to be deterministic, and it is given by a fixed distribution over age, for men and women separately. The assumption of perfect dependence is motivated by the quality of migration data, which is too poor in most countries to merit a more refined approach.

    The key properties of the scaled model are:

- since the choice of the scales is unrestricted, any sequence of non-decreasing error variances can be matched—in particular, heteroscedasticity can be allowed;
- for fertility and mortality, any sequence of cross-correlations over ages can be handled using an appropriate AR(1) model;
- any sequence of autocorrelations for the error increments can be matched.

Empirical specification

Under suitable conditions, the variances of the relative errors in the scaled model are a quadratic function of forecast lead time with two parameters (since the error variance at lead time zero is zero). We fitted a second degree polynomial without constant to the empirical error variances broken down by lead time (based on naïve errors), estimated the two parameters and solved for the two random parts of the error increments. In order to discount the values of outliers caused by, for example, wars, we used median errors rather than average errors as the basis for the estimation of the parameters of the scaled model. Moreover, in assessing the magnitude of the errors we did not subtract the mean—in other words, we included the possible forecast bias in the error estimates. This is because we do not believe that biases can be avoided in the future either. Alho et al. (2006) give details.

Appendix 2: Constructing predictive intervals from naïve errors and empirical errors

Figure 7 shows, for the case of Austria, two point forecasts for the level of life expectancy up to 2050, and three different prediction intervals. This appendix explains how these were computed. Prediction intervals for other countries, and those presented to the experts for fertility and migration, were derived in a similar manner.

Naïve forecast

We have assumed that the rate of decline during the past 30–35 years for age-specific mortality rates (as long as it is not negative) observed in each country will continue

in the coming 50 years and called it a *naïve* forecast. The result is an exponentially declining trend for age-specific mortality for most ages and for all countries.

European naïve 80% prediction interval (L & H)

We presented the combined European experience of naïve errors, because of the instability of individual country estimates. The experience of 11 countries (Austria, Denmark, Finland, France, Germany (West), Italy, the Netherlands, Norway, Sweden, Switzerland, and England and Wales) resulted in sets of naïve forecast errors, sorted by lead time. For each lead time $k = 1, 2, ..., 30$ we made two assumptions. First, the absolute values of the errors have a half-normal distribution. Second, the signed errors have zero expectation. With these assumptions we estimated the standard deviation of the signed errors, $SD(k)$, from the median of the absolute errors, $MdAE(k)$, according to:

$$SD(k) = MdAE(k)/0.675,$$

where 0.6745 is the 75% fractile of the normal distribution. Next we computed the constants $1.2816*SD(k)$, where 1.2816 is the 90% fractile of the normal distribution. These constants were added to/subtracted from the point forecasts obtained previously, to obtain the upper and lower limit of the 80% prediction intervals. For lead time 10 (or the forecast value 10 years ahead) the constant $1.2816*SD(10)$ has the value 1.63 years, and for lead time 50 the constant $1.2816*SD(50)$ is 5.42 years. This means that—based on the European experience of the past 30 years—the distance between the lower and the upper end point of the 80% confidence interval of a forecast for the coming 50 years is almost 11 years.

GARCH time series model forecast

The second forecast used the rate of decline for life expectancy itself based on the GARCH *time series* model (with constant) and fitted these to the time series of data. Future life expectancy values were simulated based upon the fitted model, drawing parameter values (i.e. the constant, the GARCH coefficient and the dummy coefficients) from the estimated distributions. The point forecast gives the mean value of 5,000 simulated trajectories for the period 2000–2050.

GARCH-based 80% prediction interval (L & H)

These curves represent the lower and upper limits of the 80% prediction intervals obtained on the basis of the 5,000 simulations. These intervals allow for stochastically varying volatility (=variance of the innovation term).

Empirical 80% prediction interval

These curves represent the lower and upper limits of an 80% prediction interval based on empirical forecast errors (i.e. errors observed in historical forecast). The same procedure was followed as the one described above for the naïve errors. Most of the collected forecasts were made after the 1960s. This implies that empirical time

series were rather short and that extrapolating them beyond 20 years would be too hazardous.

References

Alders, M., & de Beer, J. (2004). Assumptions on fertility in stochastic population forecasts. *International Statistical Review, 72*, 65–79.

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

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

Alho, J. (2005). Stochastic forecasting in a multiregional setting. Paper delivered at Second International Institute of Forecasters, workshop on Stochastic Demographic Forecasting, University of Salamanca, Spain, 28–29 July 2005.

Alho, J. M., Alders, M., Cruijsen, H., Keilman, N., Nikander, T., & Pham, D. Q. (2006). New forecast: Population decline postponed in Europe. *Statistical Journal of the UNECE, 23*(1), 1–10.

Alho, J., Cruijsen, H., & Keilman, N. (2006). Empirically based specification of forecast uncertainty. Paper delivered at the Conference on Demographic Uncertainty and the Sustainability of Social Welfare Systems, Kittilä, Finland, 20–23 March 2006.

Alho, J. M., & Spencer, B. D. (1997). The practical specification of the expected error of population forecasts. *Journal of Official Statistics, 13*, 204–225.

Alho, J., & Spencer, B. (2005). *Statistical demography and forecasting*. New York: Springer.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*, 307–327.

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

Chesnais, J.-C. (1992). *The demographic transition: Stages, patterns, and economic implications*. Oxford: Clarendon Press.

Council of Europe (2002). *Recent demographic developments in Europe 2002*. Strasbourg: Council of Europe Publishing.

De Beer, J., & Alders, M. (1999). Probabilistic population and household forecasts for the Netherlands. Working paper no. 45, Joint ECE–Eurostat work session on demographic projections, Perugia, Italy, 3–7 May 1999.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica, 50*, 987–1008.

Eurostat. (2003). Demographic statistics: Definitions and methods of collection in 31 European countries. Luxembourg, Eurostat working papers, population and social conditions, 3/2003/E/no. 25.

Eurostat. (2004). First demographic estimates for 2003. Statistics in focus, population and social conditions, Theme 3, 1/2004.

Eurostat. (2005a). Methodology for the calculation of Eurostat's demographic indicators, Luxembourg. Eurostat working papers and studies, population and social conditions, KS-CC-04–004.

Eurostat. (2005b). Population projections 2004–2050: EU25 population rises until 2025, then falls: Working age population expected to decrease by 52 million by 2050. Eurostat News Release 48/2005, 8 April 2005.

Keilman, N., & Cruijsen, H. (Eds.). (1992). *National population forecasting in industrialized countries*. Amsterdam etc.: Swets & Zeitlinger (NIDI/CBGS Publications no. 24).

Keilman, N., & Pham, D. Q. (2004). Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area. Discussion Paper no. 386, Oslo: Statistics Norway.

Keilman, N. W., Pham, D. Q., & Hetland, A. (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. Demographic Research, 6–15 May 2002, 409–53. Internet www.demographic-research.org/volumes/vol6/15/6–15.pdf.

Lee, R., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association, 89*(1), 175–189.

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

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

Poulain, M., Debuisson, M., & Eggerickx, Th. (1990). *Project d'harmonisation des statistiques de migration internationale au sein de la Communauté Européenne* (5 vols). Institut de Démographie, Louvain-la-Neuve.

Shryock, H. S., Siegel, J. S., & Associates (1976). *The methods and materials of demography*, condensed edition by E. G. Stockwell. New York etc.: Academic Press.

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

United Nations. (2005). World population prospects, the 2004 revision, New York: United Nations.