European Demographic Forecasts Have Not Become More Accurate Over the Past 25 Years

Nico Keilman

Are recent population forecasts more accurate than those that were computed a few decades ago? Since the 1960s, better demographic data have become available, behavioral theories have been refined, and more advanced techniques of analysis and forecasting have been developed (Crimmins 1993; Preston 1993). The demographic and statistical literatures offer a continuous accumulation of knowledge. A reasonable assumption is that this progress has led to more accurate demographic forecasts. To provide a thorough check of the assumption of improved forecast performance, I have analyzed the accuracy of historical population forecasts produced by the national statistical agencies of 14 countries.

A growing literature evaluates national population forecasts *ex-post facto* against observed statistics (for instance Preston 1974; Calot and Chesnais 1978; Inoue and Yu 1979; Keyfitz 1981; Stoto 1983; Pflaumer 1988; Keilman 1997, 2001; National Research Council 2000; Keilman and Pham 2004). These studies have shown, among other things, that forecast accuracy is greater for short than for long forecast durations, and that it is greater for large than for small populations. These studies have also highlighted considerable differences in accuracy between regions and forecast components. Because of the extrapolative character of population forecasting, irregularities in observed variables are associated with large errors. Finally, poor data quality tends to go together with poor forecast performance.

I prefer to use the term "forecast" rather than "projection." Most statistical agencies speak of a projection, which indicates a purely conditional computation: how would the age pyramid evolve in the future assuming certain developments in fertility, mortality, and international migration? An analysis of the accuracy is useless in that case, because a projection is always

100 percent correct (except for computation errors). However, unless the agency clearly presents its assumptions as unrealistic, most users interpret the projection results as a forecast, indicating a likely development, given the current knowledge of the forecaster (Keyfitz 1972).

Official population forecasts

I have analyzed the accuracy of population forecasts produced by the national statistical agencies of Austria, Belgium, Denmark, Finland, France, Germany, Italy, Luxembourg, the Netherlands, Norway, Portugal, Sweden, Switzerland, and the United Kingdom. The forecasts were made in the period 1950–2001, and I have assessed whether recent forecasts are more accurate than earlier ones. I have restricted myself to forecasts produced by statistical agencies, because those forecasts were made with a single methodology (the cohort-component method), which is the standard forecasting methodology among population forecasters (O'Neill and Balk 2001), and because they were produced in stable institutional settings. Thus, I have a relatively homogeneous data set that provides a meaningful basis for error analysis.

I have labeled each forecast made by a statistical agency by its launch year, that is, the starting year of the forecast. Usually this is the year for which the starting population of the forecast is defined. An observed population age pyramid, taken from a census or a population register, is updated based on assumed parameter values for fertility, mortality, and migration for later years. These later years constitute what I call the forecast years or forecast period. By combining forecast year and launch year, one can derive forecast duration. For instance, a forecast with launch year 1960 that predicts a certain migration level for the year 1970 allows us to compute the migration error at a forecast duration of ten years ahead.

For fertility, I have analyzed forecast accuracy of the total fertility rate. The TFR in a certain year reflects the average number of children a woman just entering the childbearing years (conventionally taken as age 15) would have if fertility in that year remained constant for the next 35 years.

For mortality, I have used life expectancy at birth, or life expectancy for short. The life expectancy in a certain year is computed in a life table, based upon death rates for that year. It reflects the mean age at death of the hypothetical life table population, which is assumed to be exposed, over its life course, to the mortality risks as expressed by the observed death rates in question.

For international migration, I have analyzed forecast errors in the level of net migration in a certain year, that is, the difference between immigration and emigration numbers. In order to compare errors in net migration across countries, I have computed net migration in each country relative to the country's population size in 2000.

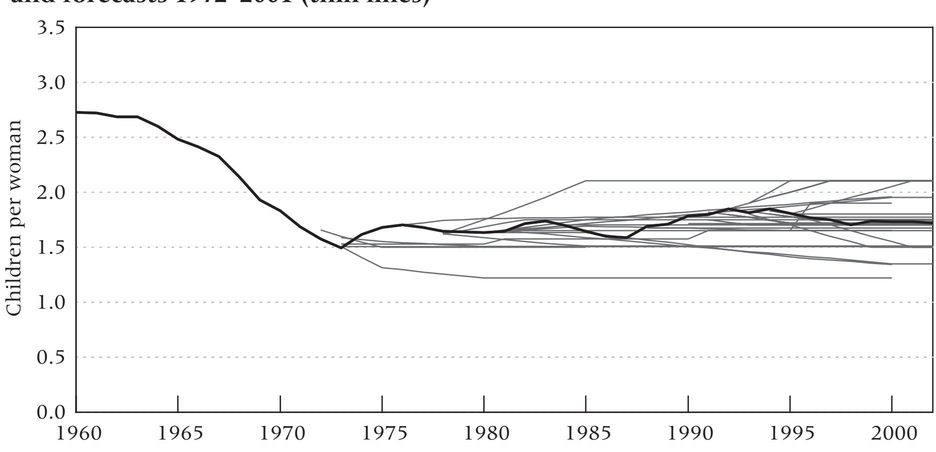
Data

I have collected data for the forecasts of the 14 countries mentioned above. All errors apply to the period from the launch year up to 2002. I used both published and unpublished sources (Keilman and Pham 2004). The observed values were taken from the 2002 edition of the *Demographic Yearbook* published by the Council of Europe, supplemented by national sources for later years. For Germany, I have computed fertility and mortality errors for the former West Germany for the period 1952–2002. For migration, I have computed errors for West Germany in forecasts made between 1952 and 1989, and for the reunified Germany in forecasts made since 1990.

The data cover the period 1950 to 2002. The earliest launch year is 1950 (Luxembourg and the Netherlands). The last launch year is 2001 (Denmark and Finland). Some countries have produced population forecasts frequently, sometimes even annually (Austria, Denmark, Finland, the Netherlands, and the United Kingdom). Other countries updated their forecast only in connection with a population census (France, Portugal). This means that my three data sets for fertility, mortality, and migration are highly unbalanced pooled cross-sectional time series. That is, the data consist of a large number of time series for 14 countries, but the time series are of unequal length. For TFR, I have 4,847 observations in 308 time series; for life expectancy there are 5,562 observations (2,819 for women and 2,743 for men) in 386 time series; and for net migration 4,339 observations in 279 series.

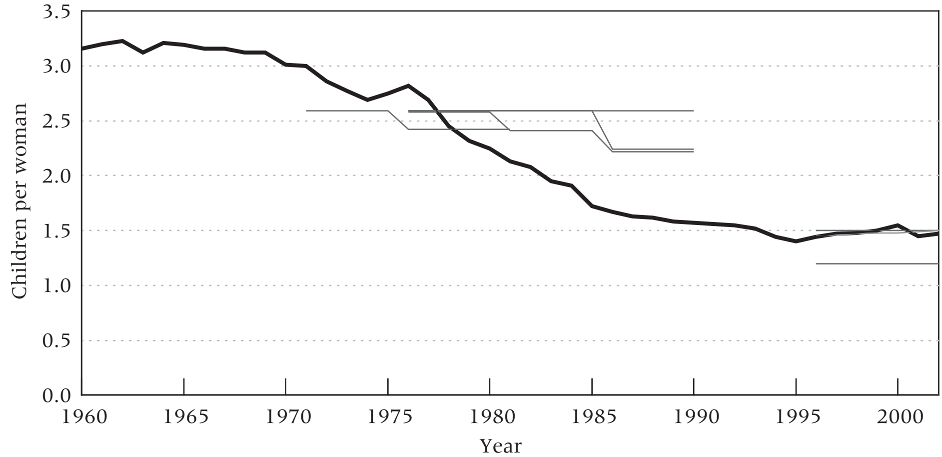
To give an impression of the magnitude of the errors, I have plotted in Figures 1–4 observed and forecast values of the total fertility rate and life

FIGURE 1 Total fertility rate, Finland, observed 1960–2002 (thick line) and forecasts 1972–2001 (thin lines)



SOURCE: See Keilman and Pham 2004.

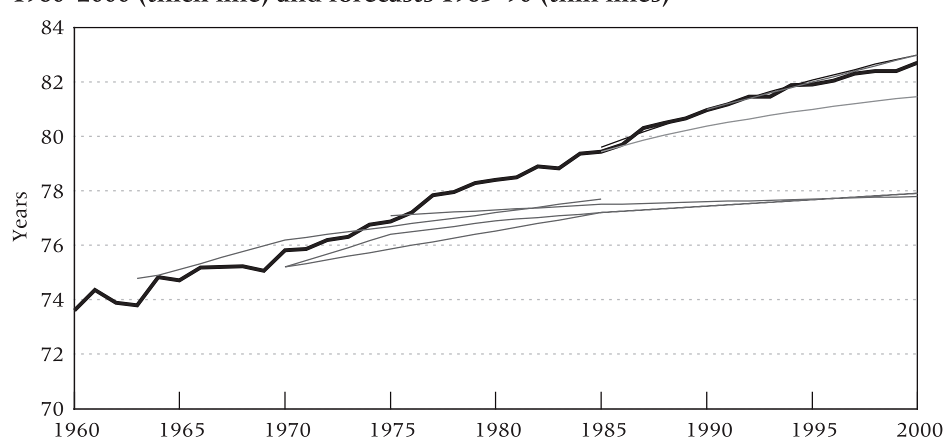
FIGURE 2 Total fertility rate, Portugal, observed 1960–2002 (thick line) and forecasts 1971–96 (thin lines)



NOTE: Observed values for 1965 and 1991 estimated by linear interpolation.
SOURCE: See Keilman and Pham 2004.

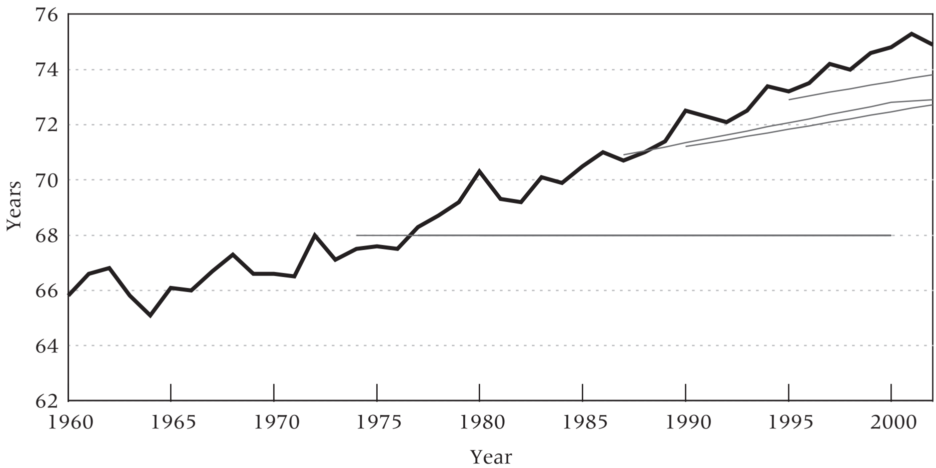
expectancy for subsequent forecasts of Finland, Portugal, France, and Luxembourg. I selected these countries because they represent forecasts with the highest (Finland, France) and the lowest (Portugal, Luxembourg) estimates of overall accuracy—the so-called country effects; see the Appendix. Apparently, it was easier to prepare accurate TFR forecasts for Finland (Figure 1) than for Portugal (Figure 2). One critical factor here is the quality of the data, which is relatively problematic for Portugal (Keilman and Pham 2004). Thus in many

FIGURE 3 Life expectancy for women in France, observed 1960–2000 (thick line) and forecasts 1963–90 (thin lines)



SOURCE: See Keilman and Pham 2004.

FIGURE 4 Life expectancy for men in Luxembourg, observed 1960–2002 (thick line) and forecasts 1974–95 (thin lines)



NOTE: Three forecasts, with launch years 1974, 1977, and 1980, assumed a constant value of 68 years.
SOURCE: See Keilman and Pham 2004.

cases the TFR forecast for Portugal was already in error in the first forecast year. A second factor is that observed trends are relatively stable in Finland starting in the mid-1970s. Concerning life expectancy, French forecasts (for women; see Figure 3) predicted the upward trend much more accurately than Luxembourg forecasts (for men; see Figure 4).

Forecast errors

For each forecast year, I computed the error in the total fertility rate as the difference between the TFR value according to the forecast and the observed value. I have done likewise for errors in life expectancy and net migration. The absolute value of the error is an often-used measure of forecast accuracy: it tells us by how much the forecast went wrong, regardless of whether it was too high or too low (e.g., NRC 2000). Henceforth I use the term “absolute error,” where “absolute” is to be understood as “omitting the sign.” In principle one could compare the accuracy of recent forecasts with that of older forecasts by comparing the mean absolute errors across launch years. This would not give a correct impression of the accuracy of subsequent forecasts, however. The reason is that recent forecasts have a shorter lifetime than older ones. It is more difficult to predict demographic behavior 20 years ahead than five years ahead. Thus forecast errors increase with forecast duration, and this fact alone causes errors in recent forecasts to be relatively small. Indeed, the mean absolute errors for fertility and mortality in Table 1 are smaller for recent launch years than for earlier ones.

The statistical analysis controls for this so-called duration effect. It also takes account of the fact that demographic behavior may be more difficult to predict for some forecast years and forecast periods than for others ("period effect"). An example is the 1960s and 1970s, when the baby boom suddenly came to an end in many countries and fertility fell unexpectedly fast. Also included are controls for country and forecast variant. In many cases, the statistical agency published more than one variant for fertility, and sometimes also for mortality and migration. Frequently these variants take the form of a middle, high, and low series. (In some cases there are two or four forecast variants.) Sometimes the variants are to be interpreted as uncertainty variants, in other cases as alternative scenarios or alternative futures. By introducing a control for the forecast variant, I account for the fact that the accuracy of the middle variant has to be judged differently compared to the accuracy

TABLE 1 Observed absolute errors (difference between observed and forecast variable, irrespective of sign): Average values for various levels of the independent variables

	Absolute error in		
	TFR (children per woman)	Life expectancy (years)	Net migration (per 1,000)
Forecast (launch year)			
1950–54	0.301	3.739	2.807
1955–59	0.280	NA	2.696
1960–64	0.622	2.738	2.474
1965–69	0.831	2.286	2.009
1970–74	0.373	2.730	2.458
1975–79	0.214	2.682	2.440
1980–84	0.183	1.456	2.487
1985–89	0.155	0.727	2.458
1990–94	0.121	0.730	1.871
1995–99	0.077	0.509	1.664
2000+	0.048	0.225	2.315
Period			
1950–54	0.209	2.073	1.930
1955–59	0.250	3.285	1.833
1960–64	0.351	2.443	3.146
1965–69	0.214	1.425	2.349
1970–74	0.319	1.087	3.204
1975–79	0.390	1.534	1.953
1980–84	0.313	1.729	1.445
1985–89	0.262	1.698	2.260
1990–94	0.221	1.500	3.278
1995–99	0.221	1.667	2.015
2000+	0.210	1.577	2.167

/...

TABLE 1 (continued)

	Absolute error in		
	TFR (children per woman)	Life expectancy (years)	Net migration (per 1,000)
Variant			
Low	0.193	1.720	2.949
Middle	0.262	1.715	1.979
High	0.283	0.482	2.238
Other	0.234	0.418	4.238
Sex			
Men		1.617	
Women		1.591	
Country			
Austria	0.220	1.572	2.492
Belgium	0.116	0.947	1.669
Denmark	0.127	1.168	1.456
Finland	0.520	2.749	3.318
France	0.153	2.056	3.723
West Germany	0.317	1.430	1.218
Germany			0.920
Italy	0.259	0.549	1.337
Luxembourg	0.192	3.125	5.200
Netherlands	0.299	1.222	1.647
Norway	0.276	1.309	1.200
Portugal	0.359	1.824	6.954
Sweden	0.243	1.638	1.477
Switzerland	0.217	1.443	3.488
United Kingdom	0.323	3.834	1.764

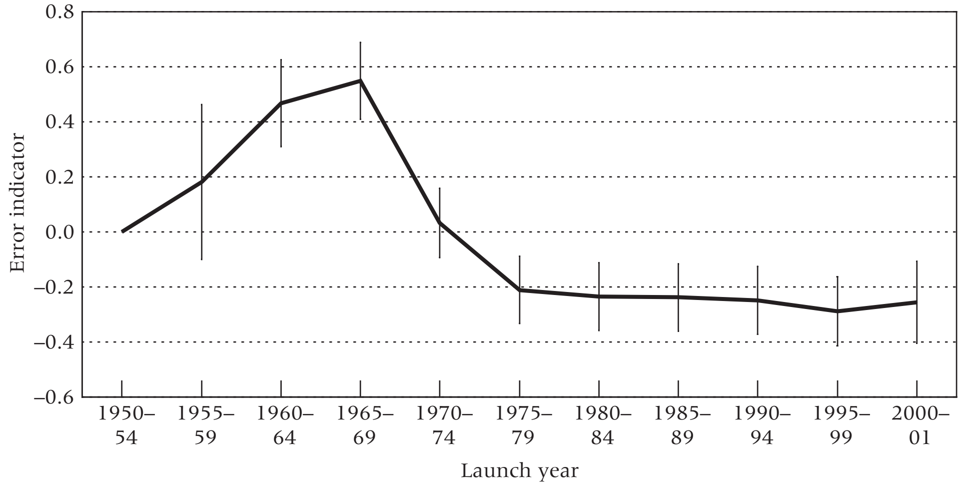
of the other variants. The statistical analysis results in an error indicator for each launch year (i.e., an estimate of the "launch year effect"), which summarizes forecast errors for the forecasts produced in that year, controlling for the effects of forecast duration, of period, of forecast variant, and of country. The method is described in the Appendix.

Table 1 gives average values of observed absolute errors for various categories of the control variables.

Estimation results

Figure 5 shows estimates of the forecast error indicator for fertility forecasts that were made in consecutive five-year periods. The indicator values as such cannot be interpreted; one can only compare them across launch years. The

FIGURE 5 Estimated values of forecast error indicator (thick line) and 95 percent confidence intervals (thin lines) for total fertility rate forecasts



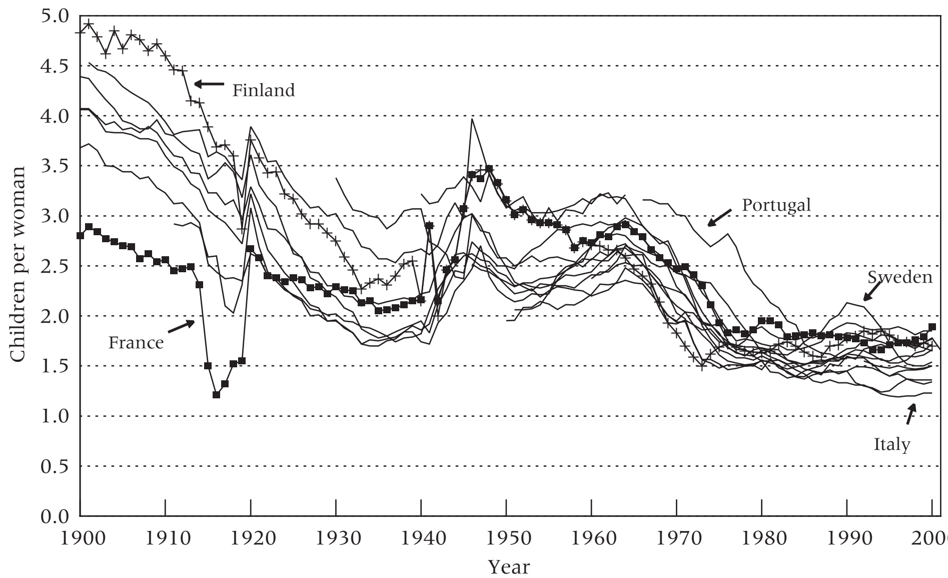
NOTE: The dependent variable is ln[0.3+abserror(TFR)], where abserror(TFR) denotes the absolute error in the TFR (i.e., the difference between observed and forecast TFR, irrespective of sign). The model also controls for period, duration (parameterized by the sum of a straight line and a square root function), country, and forecast variant. Launch years 1950–54 are reference years. (See Appendix.)
SOURCE: Appendix Table 1.

earliest period, launch years 1950–54, is the reference period. A low value for a given launch year reflects small forecast errors, compared to a launch year with larger values. Thus an assumed improvement in forecast accuracy should be reflected in a falling trend in the indicator. Improvements clearly occurred after the 1960s, when the baby boom came to an end (see Figure 6), but the improvements from the end of the 1970s to the end of the 1990s are too small to be statistically significant. After 1999 there is even a slight (but insignificant) increase in the error indicator.

When I ignore the fact that fertility may be more difficult to predict in some periods than in others, the results are qualitatively the same as those displayed in Figure 5. However, when I ignore the fact that recent forecasts have a shorter lifetime than earlier ones, the forecast error indicator falls continuously since the end of the 1960s. Thus we can conclude that the shorter lifetime of later forecasts is one of the factors that explain why the average errors of recent forecasts are smaller than those of older ones (as Table 1 shows); but when we take this factor into account the accuracy improvement disappears. One might also assume that different degrees of predictability in different periods are an additional explanation, but the data are inconclusive about this.

For mortality forecasts, too, I selected the earliest period (launch years 1950–54) as the reference period. None of the mortality forecasts in my data set has launch years 1955–59, hence I cannot compute errors in life

FIGURE 6 Total fertility rate in 14 European countries, ca. 1900–2000

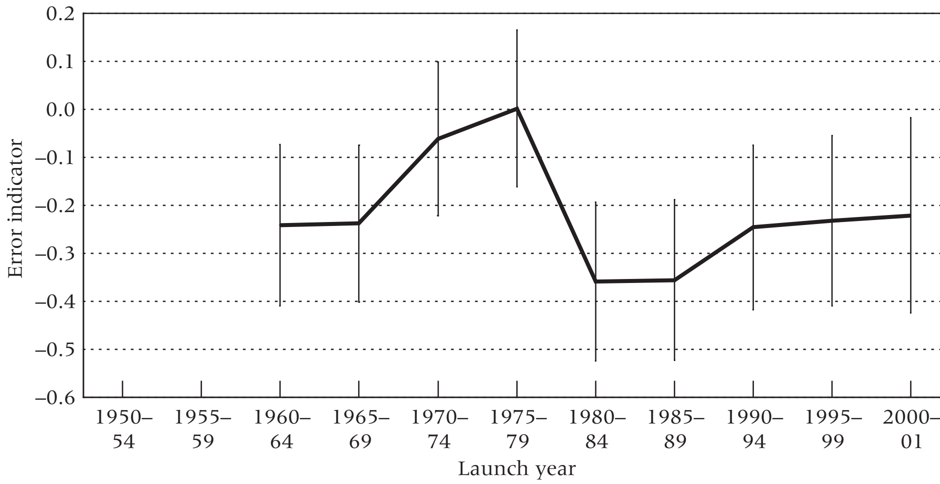


SOURCE: See Keilman and Pham 2004.

expectancy forecasts for those years. Figure 7 shows that life expectancy forecasts clearly became more accurate between the end of the 1970s and the beginning of the 1980s. The reason is that during the 1950s and 1960s, survival chances for adult and elderly men were no longer improving as they had in earlier decades. In some countries middle-aged men even saw their survival chances falling. This unfavorable trend, strongly related to life styles after World War II, was in large part caused by high mortality due to neoplasms, cardiovascular diseases, and motor vehicle accidents (Preston 1974). Around 1970 the stagnation of the life expectancy of men came to an end. Population forecasters were slow to observe the new trend, but by 1980 many forecasts assumed increasing life expectancies (Ascher 1978). Yet, Figure 7 shows that life expectancy forecasts became less accurate after the 1980s, although the increase in the curve is not statistically significant. Population forecasters systematically underpredicted the rises in life expectancy (Keilman 1997).

My statistical analysis of the errors in net migration forecasts revealed that those forecasts have not become more accurate since the early 1950s. For launch years after 1955 and up to 2000, the forecast error indicator that I computed is never significantly different from zero (see estimated forecast effects in the last column of Appendix Table 1). If anything, this means that later net migration forecasts were not any better—nor worse—than the forecasts prepared in the period 1950–54.

FIGURE 7 Estimated values of forecast error indicator (thick line) and 95 percent confidence intervals (thin lines) for life expectancy at birth (LEB) forecasts



NOTE: The dependent variable is ln[1+abserror(LEB)], where abserror(LEB) denotes the absolute error in the LEB (i.e., the difference between observed and forecast LEB, irrespective of sign). The model also controls for period, duration (parameterized by the sum of a straight line and a square root function), country, forecast variant, and sex. Launch years 1950–54 are reference years. None of the forecasts in the dataset has launch year 1955–59. (See Appendix.)
SOURCE: Appendix Table 1.

Probabilistic population forecasts: Quantifying our ignorance

Forecast errors are inherent to forecasting, and here I show that errors in demographic forecasts have not become smaller in recent decades. No indications lead us to believe that forecast accuracy will improve substantially in the near future. Nor are there reasons why prospects for improvements in the accuracy in industrialized countries other than the 14 analyzed here are very different. This implies that the users of a forecast, who are often interested in the future size and age distribution of the population, should be informed of how accurate that forecast is. The errors in fertility, mortality, and migration forecasts translate into errors in forecasts of population size and the age pyramid. A probabilistic forecast computes the future population and its age distribution in the form of probability distributions, not just numbers. It includes the forecaster's assessment of the most likely developments (similar to the middle variant in a traditional forecast), in addition to a range of possible deviations from the most likely path, and how probable such deviations are. Thus a probabilistic forecast informs the user how uncertainty varies across age groups or between the sexes. When a policymaker is able to take forecast uncertainty into account, better decisionmaking is likely to result. As soon as the expected costs involved in wrong decisions are known, an optimal strategy can be chosen.

Unfortunately, nearly all official forecasts are deterministic, not probabilistic—Statistics Netherlands is the only known exception (Alders and de Beer 1998). But demographers and statisticians have developed methods to calculate probabilistic forecasts; see the recent book by Alho and Spencer (2005) and the special issue of the *International Statistical Review* (Lutz and Goldstein 2004) for comprehensive reviews. One of the key considerations is the role of experts in defining uncertainty. In some applications, their role is modest, and time-series methods and other statistical techniques are applied to model the stochastic processes. Examples of this approach are the probabilistic forecasts for the United States by Lee and Tuljapurkar (1994), and for 18 European countries by Alho et al. (2006). In other applications—for example, one for the population of the world (Lutz et al. 1997)—experts play a more important role and define the variance of key parameters (such as the TFR or the life expectancy at birth). Originally this approach gave less emphasis to formal time-series methods, but such methods have also been included in more recent applications: see Lutz et al. (2001) for world population and Lutz and Scherbov (2004) for India.

A number of practical issues are also connected to computing probabilistic forecasts. A primary one is the quality of the base data. Probabilistic forecasts commonly apply the cohort-component approach, that is, they start from a known base population in the form of a population age pyramid and apply random parameters for fertility, mortality, and migration to that base. But the base population is not always known with certainty, thus one should distinguish between uncertainty related to the base population and uncertainly related to future trends of fertility, mortality, and migration. In many industrialized countries the data are of good quality, and information about the base population is reliable. But there are exceptions. For example, recent censuses in France, Italy, and the United Kingdom produced population numbers for some age groups that were different from pre-census estimates by more than 5 percent (Alders et al. 2007). In the long run, uncertainty about the base population is less important than wrong assumptions regarding fertility, mortality, and migration, but uncertainly in the short run may be substantial (NRC 2000). A second issue is that the uncertainty parameters for probabilistic forecasts are themselves uncertain. Frequently, they result from extrapolations of observed uncertainty statistics, either model-based extrapolations or more intuitive ones. Thus, one strategy for forecasters is to be cautious and not underestimate the uncertainty of the forecast (Alho et al. 2006). In spite of these and other problems, current methods for probabilistic forecasts have been shown to give meaningful results (e.g., Alho et al. 2006; Lutz and Goldstein 2004; Lee 1998; Lee and Tuljapurkar 1994), and forecast agencies should consider adopting them. So far, they have dealt with forecast uncertainty by introducing a few forecast variants. Uncertainty is not quantified in this manner, because no probabilities are attached to these variants. Moreover, such variants are inconsistent from a statistical point of view (e.g., Lee 1998).

A practical issue is that users have to know how to handle forecast results in the form of probability distributions, rather than one number. In the short run, forecast uncertainty is not critical, at least for most forecast results at the country level. But in the long run, users should be aware of the costs attached to employing a forecast result that turns out to be too high or too low later on. They should ask themselves whether an immediate decision based on the uncertain forecast is necessary, or whether they can wait to see if a new forecast shows less uncertainty. In case an immediate decision is required, they should check whether overpredictions are more costly than underpredictions, and base their decisions on such an assessment.

Appendix: Methods

Model

I analyze whether recent forecasts are more accurate than older ones. For fertility, I estimated a regression model with the error in the total fertility rate as the dependent variable, and similarly for mortality (life expectancy) and migration (scaled net migration). Forecast launch years are included as independent variables in the form of dummy variables. I also control for the effects of forecast period, forecast duration, country, forecast variant, and, for mortality, also for sex.

Since the focus of this analysis is on forecast accuracy, I have modelled the *absolute* value of the forecast error. The absolute error tells us by how much the forecast went wrong, irrespective of whether it was too high or too low. The absolute error reflects forecast accuracy, while the signed error would have reflected forecast bias.

The dependent variable in the model is the natural logarithm of the absolute error. I used a logarithmic transformation because the absolute error is non-negative by definition. Thus the estimates show relative effects rather than absolute ones. The models for fertility, mortality, and migration are all of the form

$$\ln(a + E_{f,p,d,c,v}) = \beta_0 + F_f X_f + P_p X_p + \beta_1 d + \beta_2 \sqrt{d} + C_c X_c + V_v X_v + \varepsilon_{f,p,d,c,v}, \qquad (1)$$

where $E_{f,p,d,c,v}$ is the absolute error for the forecast with launch year f in calendar year (period) p at forecast duration d for country c and forecast variant v. The independent variables X_f, X_p, X_c, and X_v are dummy variables that represent the forecast (launch year), period, country, and variant, respectively. F_f, P_p, C_c, and V_v are coefficients to be estimated. They represent the effects of forecast launch year f, period p, country c, and variant v. The duration effect is parameterized as a sum of a linear and a square root function of forecast duration. These particular forms were chosen after some experimentation; they give satisfactory fits for fertility, mortality, and migration. β_0, β_1, and β_2 are parameters to be estimated, a is a predefined constant (see below), and ε is the residual, for which the usual properties are assumed. For the life expectancy model, there is an additional dummy variable X_s and a corresponding coefficient S_s for sex.

Interpretation of effects

The *forecast (launch year) effect* gives us the contribution of the launch year to the forecast error, irrespective of forecast period, forecast duration, the country in which the forecast was prepared, and the forecast variant. It is estimated for various launch years *f*. If the assumption of improvement in forecast accuracy over time is correct, the estimates of the effect for recent launch years will be smaller than those for earlier years. The forecast effect is called "forecast error indicator" in the main text.

The *period effect* tells us how difficult it was to predict fertility, mortality, or migration for a certain period, irrespective of launch year, duration, country, and variant. I assume that the period effect is independent of the effects of other explanatory variables—in particular, country. The reason is that the total fertility rate and life expectancy have largely shown a very similar pattern across the 14 countries in the period 1950–2002. For migration, the situation is somewhat different, hence I have included a few interactions between period effects and country effects (see below).

The *duration effect* accounts for the fact that forecast accuracy declines with increasing forecast horizon. Twenty years into the future it is more probable than only five years into the future that conditions that have an impact on fertility, mortality, and migration will have changed relative to the period in which the forecast was prepared. The duration effect is assumed to be independent of the effects of launch year, forecast period, country, and variant.

The *country effect* reflects the idiosyncrasies of the various countries related to the production of a population forecast, such as the quality of the available data, the number and skills of the forecasting staff, and so on. Also, the country effect captures the fact that large populations are easier to forecast than smaller populations, other things being equal.

The *variant effect* tells us whether there was a substantial difference in errors for middle variants, high and low variants, or other variants, other things being the same. Forecasts that had only one variant were coded as middle variant.

The constant *a*

In some cases, the absolute error is zero or close to zero, and thus the logarithm of the absolute error would be strongly negative. I added a small constant to the absolute error, which improved the symmetry of the residuals. After some experimentation, and using a QQ plot to check the normality of the residuals, I found that $a = 0.3$ (fertility), $a = 1$ (mortality), and $a = 1$ (migration) gave satisfactory results.

Autocorrelation

After model (1) was estimated by OLS, the empirical residuals showed strong positive autocorrelation, in particular for fertility and mortality (first-order autocorrelations for the ε-residuals were 0.855 for fertility, 0.838 for mortality, and 0.491 for migration). Thus I used the Prais–Winsten estimator (see Greene 2003) and transformed the dependent variable of model (1) as follows.

For a given forecast with launch year f, country c, and forecast variant v, there is a time series of forecast errors $\{E_{f,p+d,d,c,v}\}$, with $d = 0, 1, 2, \ldots$. Construct the transformed variables

$$E^*_{f,p,0,c,v} = \sqrt{1-\hat{\rho}^2}\,E_{f,p,0,c,v}, \text{ and}$$

$$E^*_{f,p+d,d,c,v} = E_{f,p+d,d,c,v} - \hat{\rho}_\varepsilon E_{f,p+d-1,d-1,c,v}, \; d=1, 2, 3, \ldots$$

where $\hat{\rho}_\varepsilon$ is an estimate of the first-order autocorrelation of the ε-residuals. Construct transformed independent variables similarly (including a transformed constant term). Next estimate model (1) with the transformed variables. The transformed model is formally equivalent to (suppressing unnecessary indexes) $\varepsilon_d = \rho_\varepsilon \varepsilon_{d-1} + u_d$, where u_d is a residual term with the usual properties.

The results in Figures 5 and 7 for fertility and mortality, and those reported for migration, are all based upon the transformed variables; see also Appendix Table 1. After transformation, the first-order autocorrelations for the u-residuals were 0.284, –0.068, and 0.108 respectively. The migration model includes interaction effects for Portugal with the period 1970–74, and for Austria, West Germany, and Germany with the period 1990–94. Compared to the other countries in the data set, Portugal experienced extraordinarily high levels of emigration between 1964 and 1973, mainly due to labor migration to other European countries. This suggests an interaction term between the country effect for Portugal and the period effects. The earliest forecast for Portugal in the data set has launch year 1971. For Germany/West Germany and Austria, interaction terms are included with the period 1990–94, because the fall of the Berlin Wall induced large immigration flows into German-speaking countries in the early 1990s. Appendix Table 1 gives all estimated effects, while Figures 5 and 7 plot the launch year effects for fertility and mortality.

In addition to transformations based on $\hat{\rho}_\varepsilon$ equal to 0.855 (fertility) and 0.838 (mortality), I also experimented with four additional transformations based on $\hat{\rho}_\varepsilon$ equal to 0.75 and 0.95, for both fertility and mortality. None of these led to a conclusion that is qualitatively different from the earlier one: since the early 1980s there has been no improvement in the forecasting performance of fertility and mortality forecasts in the 14 countries.

Random effects

Model (1) can be interpreted as one in which period effects and country effects are fixed parameters. These may be correlated with the other independent variables. If the unobserved country- and time-specific heterogeneity can be assumed to be realizations of a random process and uncorrelated with the other independent variables, a random effects model may be a more powerful representation (Greene 2003). For the case of fertility, I experimented with two random effects models: one in which country effects are assumed random, and one in which period effects are random. Both models resulted in the same conclusion as the one following from Figure 1: since 1980 there has been no clear improvement in the accuracy of fertility forecasts.

APPENDIX TABLE 1 Estimated effects for Prais–Winsten transformed variables of model (1), applied to absolute errors in the TFR, in life expectancy at birth, and in scaled net migration

	TFR	Life expectancy	Net migration
Forecast (launch year) effects			
1950–54 (ref.)	0	0	0
1955–59	0.1814	NA	–0.0304
1960–64	0.4682***	–0.2415**	–0.0057
1965–69	0.5480***	–0.2383**	0.0646
1970–74	0.0317	–0.0610	–0.0044
1975–79	–0.2115***	0.0018	–0.0155
1980–84	–0.2362***	–0.3594***	0.0728
1985–89	–0.2387***	–0.3560***	0.0726
1990–94	–0.2507***	–0.2467***	–0.0423
1995–99	–0.2889***	–0.2325**	–0.1065
2000+	–0.2566***	–0.2210**	0.0893
Period effects			
1950–54 (ref.)	0	0	0
1955–59	0.0045	–0.0004	–0.0947
1960–64	–0.0076	–0.0675	–0.0088
1965–69	–0.1546**	–0.1347	–0.0847
1970–74	0.0230	–0.1863**	–0.1260
1975–79	0.1531**	–0.2156**	–0.2589
1980–84	0.1543**	–0.2088**	–0.3711
1985–89	0.1439**	–0.2863***	–0.2554
1990–94	0.1489**	–0.3223***	–0.1062
1995–99	0.1485**	–0.3846***	–0.2284
2000+	0.1358**	–0.4017***	–0.2465
Duration (linear)	0.0015	0.0548***	0.0036
Duration (square root)	0.0927***	0.0127*	0.0911***
Variant effects			
Low	–0.0200	0.1693***	0.2295***
Middle (ref.)	0	0	0
High	0.1172***	0.0526**	–0.0020
Other	0.0491	–0.0212	–0.0072
Sex			
Men (ref.)		0	
Women		–0.0576***	
Country effects			
Austria	–0.0510	0.3181***	0.0259
Belgium	–0.1707**	0.1105*	0.0295
Denmark	–0.1576***	0.3840***	–0.0599
Finland	–0.2186***	0.2844***	–0.2314***
France	–0.1966***	–0.0976*	–0.3959***
West Germany	–0.2065***	0.1477***	0.2081***
Germany			0.3778***
Italy	–0.0198	0.0551	–0.1146
Luxembourg	–0.1816***	0.5109***	0.6803***
Netherlands	–0.1488***	–0.0044	–0.0934*
Norway	–0.1434***	0.0693	–0.1699***
Portugal	0.1992**	0.3250***	1.0174***
Sweden	–0.0798*	0.0845*	–0.1234**
Switzerland	–0.1369**	0.2575***	0.4520***
United Kingdom (ref.)	0	0	0

/…

APPENDIX TABLE 1 (continued)

	TFR	Life expectancy	Net migration
Interaction effects			
Portugal and period 1970–74			0.6856***
Austria and period 1990–94			0.7379***
West Germany and period 1990–94			0.6056***
Germany and period 1990–94			0.2637
Constant	–0.8880***	0.6192***	0.8038***
AR(1)–coefficient	0.855	0.838	0.491
Residual standard error	0.105	0.125	0.375
R^2	0.704	0.722	0.702
N	4,847	5,562	4,339

* p<0.1, ** p<0.05, *** p<0.01.

Note

Parts of this analysis were written during my stay at the Centre for Advanced Studies in Oslo, the hospitality of which I gratefully acknowledge. Early versions were presented at the Second International Institute of Forecasters' Workshop on "Stochastic Demographic Forecasting," Universidad de Salamanca, July 2005, and at the Econometrics Seminar, Department of Economics, University of Oslo, October 2006. I have benefited from helpful comments from Pilar Poncela, Marek Brabec, and other participants.

References

Alders, Maarten and Joop de Beer. 1998. "Kansverdeling van de bevolkingsprognose" (Probability distribution of population forecasts), *Maandstatistiek van deBevolking* 46: 8–11.

Alders, Maarten, Nico Keilman, and Harri Cruijsen. 2007. "Assumptions for long-term stochastic population forecasts in 18 European countries," *European Journal of Population* 23: 33–69.

Alho, Juha et al. 2006. "New forecast: Population decline postponed in Europe," *Statistical Journal of the UN E. C. E.* 23: 1–10.

Alho, Juha, and Bruce Spencer. 2005. *Statistical Demography and Forecasting*. New York: Springer.

Ascher, William. 1978. *Forecasting: An Appraisal for Policy Makers and Planners*. Baltimore: Johns Hopkins University Press.

Calot, Gérard and Jean-Claude Chesnais. 1978 "La prévision démographique," in A.-C. Decouflé, *Traité élémentaire de prévision et de prospective*. Paris: Presses Universitaires de France.

Crimmins, Eileen. 1993. "Demography: The past 30 years, the present, and the future," *Demography* 30: 579–591.

Greene, William. 2003. *Econometric Analysis.* Upper Saddle River, NJ: Prentice-Hall.

Inoue, S. and Y. C. Yu. 1979. "United Nations new population projections and analysis of ex post facto errors," paper presented at the Annual Meeting Population Association of America, Philadelphia, April.

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

———. 2001. ″Data quality and accuracy of United Nations population projections, 1950–95,″ *Population Studies* 55: 149–164.

Keilman, Nico and Dinh Quang Pham. 2004. "Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area." Discussion Paper 386, Statistics Norway, Oslo.

Keyfitz, Nathan. 1972. "On future population," *Journal of the American Statistical Association* 67: 347–363.

———.1981. "The limits of population forecasting," *Population and Development Review* 7: 579–593.

Lee, Ronald. 1998. "Probabilistic approaches to population forecasting,' in Wolfgang Lutz, James Vaupel, and Dennis Ahlburg (eds.), *Frontiers of Population Forecasting*. Supp. *Population and Development Review* 24: 156–190.

Lee, Ronald and Shripad Tuljapurkar. 1994. "Stochastic population forecasts for the United States: Beyond high, medium, and low," *Journal of the American Statistical Association* 89: 1175–1189.

Lutz, Wolfgang and Joshua Goldstein. 2004. "Introduction: How to deal with uncertainty in population forecasting?," *International Statistical Review* 72: 1–4.

Lutz, Wolfgang, Warren Sanderson, and Sergei Scherbov. 1997. "Doubling of world population unlikely," *Nature* 387: 803–805.

———. 2001. "The end of world population growth," *Nature* 412: 543–545.

Lutz, Wolfgang and Sergei Scherbov. 2004. "Probabilistic population projections for India with explicit consideration of the education–fertility link," *International Statistical Review* 72: 81–92.

National Research Council (NRC). 2000. *Beyond Six Billion: Forecasting the World's Population*. Panel on Population Projections. John Bongaarts and Rodolfo A. Bulatao, eds. Washington DC: National Academy Press.

O'Neill, Brian and Deborah Balk. 2001. "World population futures," *Population Bulletin* 56(3).

Pflaumer, Peter. 1988 "Confidence intervals for population projections based on Monte Carlo methods," *International Journal of Forecasting* 4: 135–142.

Preston, Samuel. 1974. ″An evaluation of postwar mortality projections in Australia, Canada, Japan, New Zealand and the United States," *WHO Statistical Report* 27: 719–745.

———. 1993. "The contours of demography: Estimates and projections," *Demography* 30: 593–606.

Stoto, Mike. 1983. "The accuracy of population projections," *Journal of the American Statistical Association* 78: 13–20.