Journal of Official Statistics, Vol. 13, No. 3, 1997, pp. 227–243

The Effect of Uncertainty of Migration on National Population Forecasts: The Case of the Netherlands

Joop de Beer[1]

Although in the Netherlands natural population growth (the excess of births over deaths) is larger than net migration, international migration has become the main source of uncertainty of forecasts of population growth, at least in the short and medium run. This calls for two types of response. First, obviously, one should try to reduce uncertainty by improving migration forecasts. Secondly, forecasters should provide an accurate indication of the degree of uncertainty of future migration and its implications on the uncertainty of forecasts of population size and age structure. On the basis of Dutch data, this article shows how the effect of the uncertainty of migration can be assessed both by analyzing historical forecast errors and by modelling forecast errors. With the help of Monte-Carlo simulations the effect of uncertainty of migration can be compared with that of fertility and mortality.

Key words: Population projections; forecast errors; international migration; time-series models; simulations.

1. Introduction

Whereas in the 1960s and 1970s fertility appeared to be the main source of uncertainty of population forecasts in the Netherlands, since the early 1980s international migration has become an important source of errors in forecasting population growth. On the basis of Dutch experience, this article examines to what extent uncertainty of future migration affects uncertainty of population forecasts.

In making national population forecasts for the Netherlands, Statistics Netherlands has responded in two ways to the increasing importance of forecast errors of migration. First, research projects were initiated for examining how the forecasts could be improved. Since as yet there is not a single method outperforming all other methods, in preparing migration forecasts an eclectic approach is followed. Section 2 gives a concise description. Second, Statistics Netherlands aims to provide an accurate indication of the degree of uncertainty of migration forecasts and its implications on the uncertainty of forecasts of total population and the age structure.

Section 3 examines the size of errors of migration forecasts. The validity of the low and high variations of the forecasts published since 1975 is examined. A time-series model of historical forecast errors is used to estimate the width of the forecast interval.

[1]Statistics Netherlands, Department of Population, P.O. Box 4000, NL-2270 J M Voorburg, The Netherlands.
Revised version of paper presented in the symposium 'Analysis of errors in demographic forecasts with implications on policy,' March 30–April 2, 1995, Koli (Finland). The views expressed in the article are those of the author and do not necessarily reflect the views of Statistics Netherlands.

Section 4 compares the effect of uncertainty of migration on population size and age structure with the effect of the uncertainty of fertility and mortality using simulations based on time-series models of forecast errors.

Finally Section 5 draws some conclusions.

2. Practice of Forecasting Migration in the Netherlands

2.1. *National population forecasts in the Netherlands*

The official Dutch National Population Forecasts are published by Statistics Netherlands. Since 1984 the forecasts have been updated annually. The reason for the annual update is that this guarantees the forecasts always to be in line with the most recent observations. Assumptions on the most plausible future development are specified for the period until the so-called target year. For the latest forecasts the target year is 2010. After the target year the migration number - as well as fertility and mortality rates - are held constant.

In the national population forecasts Statistics Netherlands takes uncertainty into account by publishing three variants. In addition to the medium variant, which is assumed to describe the most probable future, low and high variants are published. For each of the three components (migration, fertility, and mortality) low, medium, and high variants are specified. Of the possible 27 combinations of these variants only three are published regularly. The low variant is the variant with the lowest population growth. This variant is obtained by combining low values of fertility and immigration with high vales of mortality and emigration. Accordingly the high variant is the variant with the highest population growth.

The population forecasts are based on a cohort-component model. The model has as input net migration numbers by age and sex. However, the assumptions on future net migration are based on separate assumptions on immigration and emigration.

2.2. *Methods of forecasting migration*

In a survey of methods used for forecasting international migration in 30 industrialized countries George and Perreault (1992) conclude that most countries apply rather simple methods. A great majority of countries set future migration equal to the current level or assume zero net migration.

Statistics Netherlands follows an eclectic approach. Several methods are used in parallel: demographic analysis, explanatory analysis, time-series models, and consultation of external experts. On the basis of judgment the results of these analyses are combined into one forecast.

The judgmental assessments are made by a group of demographers in the Department of Population at Statistics Netherlands and discussed with an Advisory Committee which is composed of external experts. The final decision is taken by the Head of the Department of Population.

One example of demographic analysis is the estimation of the size of family reunion and family formation on the basis of such demographic variables as age, sex, marital status, and marriage duration (De Beer et al. 1993). One example of explanatory analysis is the identification of push and pull factors explaining separate types of migration. For example, migration between European countries turns out to be closely related to the unemployment rate in the country of origin and the country of destination.

Consultations of experts plays an important part in making migration forecasts as not all relevant information can be included in one quantitative model which is fitted to historical data. For example, the size of future migration may depend on the effectiveness of new immigration policies, which cannot be assessed on a purely empirical basis. The procedure for soliciting expert opinion is that several alternative assumptions on future migration are presented to a meeting of migration experts. On the basis of the discussion Statistics Netherlands selects one alternative.

2.3. *ARIMA models*

The time series of immigration and emigration can be modelled by ARIMA (Autoregressive Integrated Moving Average) models (Box and Jenkins 1970). For both time series first-order autoregressive models seem compatible with the pattern of the autocorrelations shown in Table 1.

For immigration the estimated model (based on observations in the period 1960–1994) is:

$$I_t = .70 I_{t-1} + 26{,}952 + e_{I,t}$$
$$\quad (.11) \qquad (10{,}054)$$

where I_t = immigration in year t and $e_{I,t}$ is a random term. Estimated standard errors are given in parentheses. As the autocorrelations of the residuals do not exhibit a systematic pattern, the model seems to be appropriate. This model implies that the long-run projection of immigration moves to a level of $26{,}952/(1 - .70)$ = ca. 90,000.

The fitted model for emigration is:

$$E_t = .84\ E_{t-1} + 10{,}597 + e_{E,t}$$
$$\quad (.13) \qquad (7{,}879)$$

where E_t = emigration and $e_{E,t}$ is white noise. This model implies that emigration moves to 66,000 in the long run. However, it should be noted that the constant term is relatively small compared with the standard error. If a model without a constant term is estimated,

Table 1. Autocorrelation coefficients, 1960–1994

	immigration	emigration	net migration
1	0.72	0.74	0.50
2	0.47	0.49	0.16
3	0.34	0.32	0.07
4	0.33	0.23	0.19
5	0.27	0.14	0.26
6	0.01	0.03	−0.01
7	−0.12	0.08	−0.20
8	−0.29	0.13	−0.40
9	−0.21	0.18	−0.21
10	−0.19	0.24	−0.11

the value of the autoregressive coefficient turns out to be one. This indicates that a random-walk model may be appropriate:

$$E_t = E_{t-1} + e_{E,t}.$$

The forecasts of this model are equal to the last observed value. Starting in 1994 the forecasts equal 78,000.

For net migration the first autocorrelation differs from zero, whereas the other autocorrelations are relatively small. This suggests a first-order moving average model. The fitted model is:

$$N_t = \underset{(4{,}103)}{21{,}846} + \underset{(.15)}{.48}\, e_{N,t-1} + e_{N,t}$$

The projections of this model two years or more ahead equal 22,000. This corresponds rather closely to the forecast resulting from separate forecasts of immigration and emigration based on first-order autoregressive models: 90,000 – 66,000 = 24,000. If the random-walk model projection of emigration is used, the projection of net migration becomes smaller: 90,000 – 78,000 = 12,000. Note that the random-walk projection of emigration is very sensitive to random fluctuations in the last observation year, in contrast with the long-run projections based on the autoregressive model.

In the 1994-based Dutch population forecasts net migration is assumed to move to 35,000 in the long run. One important reason for assuming a level that is considerably higher than the level projected by the ARIMA-models, is that the latter level is much lower than the observed level in recent years, except for the value in 1994. The conclusion of the consultation of external experts was that they considered it unlikely that future migration would be substantially lower than the current level. The assumed long-run value of 35,000 is based on the average net migration in the last ten observation years.

In order to compare the ARIMA-projections with previous official forecasts, ARIMA-models of net migration were estimated for the periods 1960–1980, 1960–1985, and 1960–1990. For all these periods, the autocorrelations are compatible with a first-order moving average model. The projections are shown in Figure 1a. The projections starting in 1980, 1985, and 1990 turn out to differ only slightly. Clearly the differences of the successive ARIMA-projections are much smaller than the differences between the official forecasts starting in the corresponding years, shown in Figure 1b.

When preparing the 1980- and 1985-based official forecasts no ARIMA-models were used. Although in preparing the 1990-based forecasts ARIMA-projections were taken into account, the projections were adjusted on the basis of judgment.

2.4. *ET-model*

In preparing the 1993-based forecasts a new time-series model was used: the ET-model (De Beer 1994). This model combines time-series extrapolation for the short run and judgment for the long-run target value. E stands for extrapolation and T for target.

The ET-model includes parameters that indicate the relative weight of the assumptions

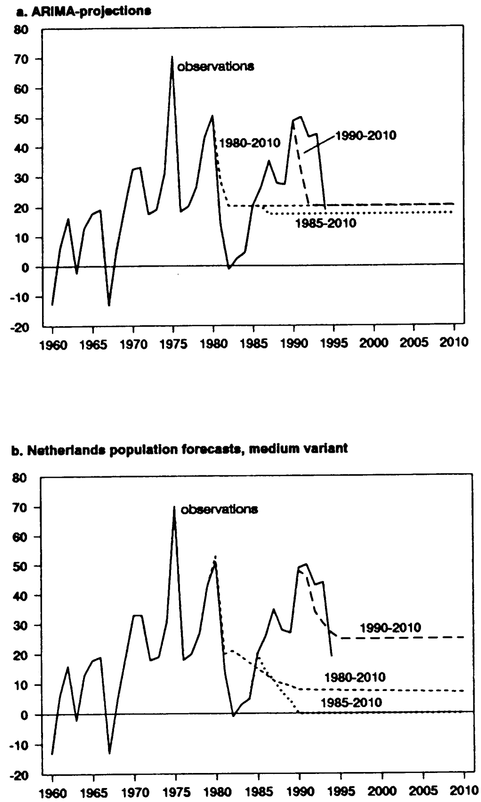

Fig. 1. Net migration (×1,000): ARIMA-projections and official forecasts

on short-, medium-, and long-run developments. The formulas of the model are given in the Appendix.

In order to apply the ET-model for making forecasts the values of five parameters have to be assessed: one parameter determines the weight of short-run changes, another parameter determines how rapidly the medium-term trend is being damped, the third parameter determines the length of the base period used to calculate the average change that is the basis of the medium-run projections, the fourth parameter is the target value and finally there is a parameter that determines how rapidly the forecasts move toward the target value.

The first three parameters determine to what extent the projections depend on past observations. The other two parameters determine the long-run direction of the forecasts.

When applying the ET-model judgment can play a role in choosing the target value. But also the values of the other parameters can be based on judgment, because the parameters are interpretable (in contrast with ARIMA-models that tend to have a black-box character). Instead of determining the values of the parameters on a priori grounds, they can also be obtained from fitting the model to the observed time series. Another possibility is to fix the values of some parameters a priori and to estimate the values of the other parameters. For example, it is possible to fix the target value on a priori ground and to estimate the values of the other parameters in such a way that they produce the most accurate forecasts in a certain sample period.

3. Uncertainty of Forecasts of International Migration

Uncertainty of forecasts of migration can be measured in two ways. First, the variance of forecast errors can be estimated on the basis of a statistical time-series model describing the movement of migration. In the preceding section ARIMA-models were discussed. Second, errors of historical forecasts can be analyzed. On the basis of a time-series model of historic forecast errors the variance of future forecast errors can be estimated.

In the Dutch population projection model net migration is used as input. Hence the analysis of forecast errors is applied to net migration rather than to immigration and emigration separately. As fluctuations in immigration are considerably larger than those in emigration, errors in forecasting immigration are the main source of errors in forecasting net migration.

As discussed in Section 2 net migration can be modelled by a first-order moving average model. The variance of the forecast errors of this model equals σ^2 for the first forecast year and $\sigma^2(1+\theta^2)$ for the subsequent forecast years, where σ^2 is the variance of the random term and θ is the moving-average coefficient. Assuming the random term to be normally distributed, the boundaries of the 67 per cent forecast interval are $\pm\sigma\sqrt{(1+\theta^2)}$. As for the time series of net migration the estimate of σ equals 16,000 and the estimate of θ equals .48, the forecast interval equals $\pm 18{,}000$.

Application of the moving average model implies that the width of the forecast interval does not increase with the forecast lead time after the second year into the projection period.

Assumptions on international migration have been included in the Dutch population forecasts since 1975 (see Figure 2). In the 1975-based forecasts two variants were specified. Hence there was no medium variant. From 1980 onwards three variants were specified. The medium variant was assumed to describe the most plausible future development. In the early 1980s there was a strong decrease of immigration. In the 1984-based forecasts net migration was assumed to remain low.

In 1984 immigration started to rise. At first, the forecasters assumed that the increase was only temporary. In subsequent years the short-term forecasts turned out to be too low. As a reaction, in 1986 the long-term migration assumption was revised upward. As migration continued to grow, some further upward revisions were made in later forecasts. In the 1990s net migration started to decline. As the experts assumed the decline to be temporary, the migration forecasts were not revised downward.

In sum it can be concluded that whereas the forecasts of the 1980s were too low, the forecasts of the 1990s were too high.

In the 1980s not only the level of net migration was underforecasted, but also the degree of uncertainty of the forecasts. Although in the early 1980s no explicit statement was made about the assumed probability that the interval between the low and high variants would cover the future observations, in later forecasts it was explicitly stated that the interval was assumed to correspond with 67 per cent probability. On the basis of this assumption, forecasts can be evaluated by examining whether the intervals cover two thirds of the observations.

The intervals between the low and high variants of the forecasts published in the first half of the 1980s turn out to cover only about 10 per cent of the observations.

Particularly in the 1984-based forecasts the interval for the long run was extremely

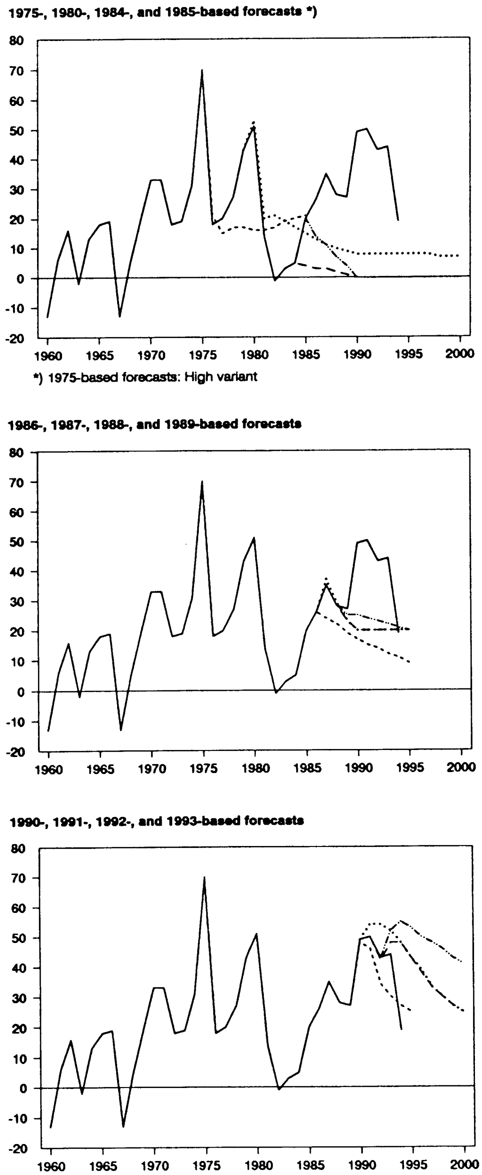

Fig. 2. Net migration (×1,000): Medium variant Dutch population forecasts

small (Figure 3). One reason is that in the long run the forecast interval was expected to describe uncertainty on average migration rather than on migration in separate years (De Beer 1993). As a consequence the forecast interval for the long run was smaller than for the short run. The major reason for the small forecast interval, however, is that the uncertainty on future migration was strongly underestimated. At the time the forecasts were made, net migration was about zero. Apparently the forecasters thought that the period of positive net migration that had started in the late 1960s had come to an end and they thought it very unlikely that immigration would start to rise again.

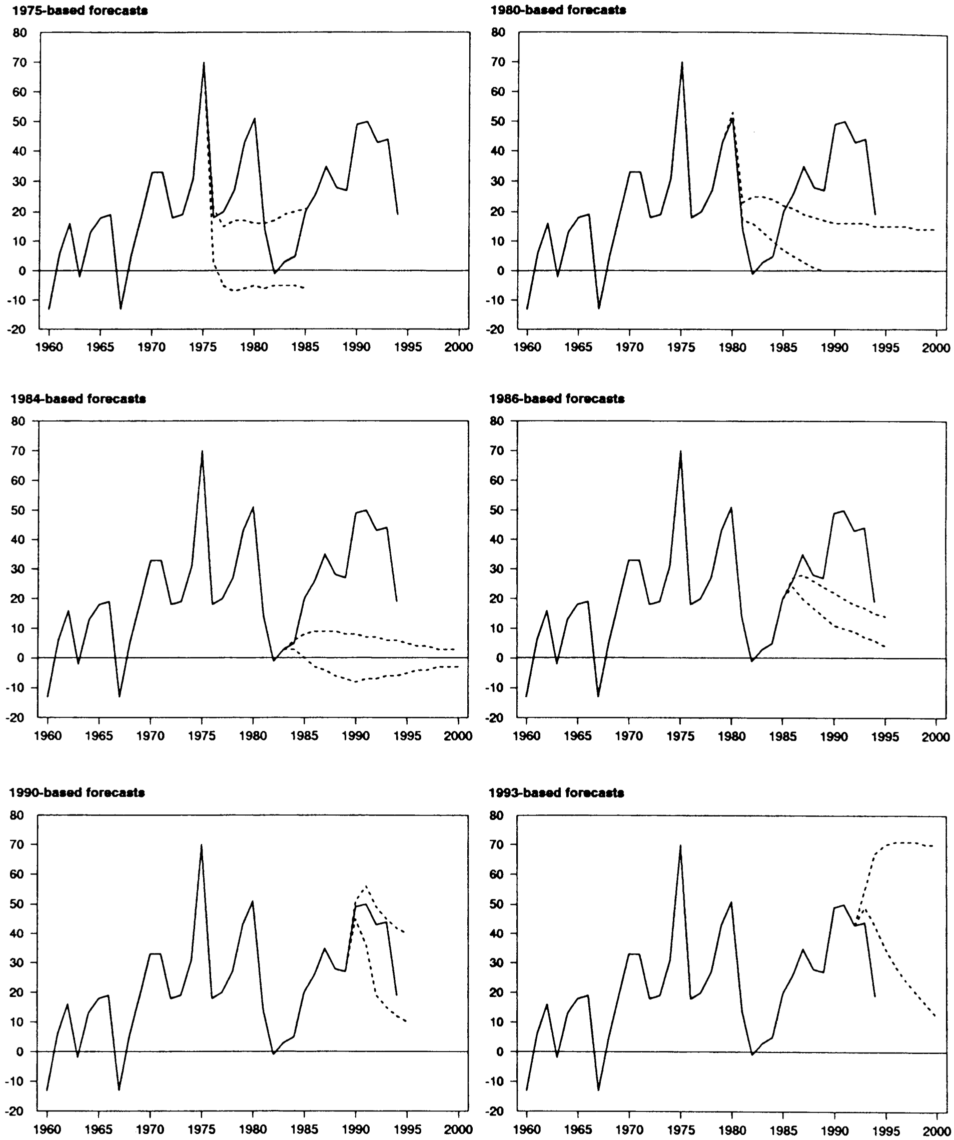

Fig. 3. Net migration (×1,000): Low and high variants

In the 1986-based forecasts the interval between the low and high variants was widened. This turned out to be too small also. The intervals of the forecasts published between 1986 and 1989 cover about one third of the observations. There was a further increase of the interval. If the same interval had been used for the subsequent forecasts, the interval would have covered almost 60 per cent of the observations in the years 1990–1994.

One problem of using historic forecast errors for assessing uncertainty of forecasts is that there are only few data on long-run forecast errors. For example, there are only three observations for 10-years ahead forecast errors. Clearly this is not a solid basis for assessing a forecast interval. Hence a forecast interval for the long run needs to be based on a model of forecast errors. One problem in using an ARIMA-model of net migration to assess forecast intervals is that the ARIMA-projections do not correspond to the medium variant of the official forecasts. As discussed above, the official forecasts depend heavily on judgment. Even though ARIMA-projections are taken into account in making the forecasts, subjective and qualitative elements play a decisive role. First, the choice of models is subjective to some extent. Second, the model-based projections are adjusted on the basis of qualitative information that is not included in the models. In analyzing historic forecast errors, the effect of judgment is taken into account. Hence uncertainty of forecasts is to be assessed on the basis of a model of historic forecast errors of net migration rather than on the basis of a time-series model of net migration itself. The autocorrelations of the forecast errors of net migration of all official forecasts published since 1975 taken together suggest that an autoregressive model may be appropriate. Fitting a first-order autoregressive model to the forecast errors of all forecasts simultaneously yields

$$f_{t+i} = \underset{(.07)}{.89}\, f_{t+i-1} + e_{f,t+i} \qquad i > 1$$

where f_{t+i} is the error of the forecast for i years ahead made in year t and $e_{f,t+i}$ is a random term. A constant term added to the model turns out to be not significantly different from zero.

The standard deviation of the random term equals 14,000. On the basis of this model forecast intervals can be calculated, assuming the random term to be normally distributed. The standard deviation of forecast errors with lead time j equals $\sigma\sqrt{\Sigma_{i=1}^{j}\varphi^{2i-2}}$, where φ is the autoregressive coefficient. In contrast with the ARIMA-model of net migration, according to this model the forecast interval increases with the forecast lead time. For example, the interval 15 years ahead is 1.6 times as wide as the interval with lead time two years. For forecasts 15 years ahead the interval is $\pm$30,000.

4. Effect of Errors of Migration Forecasts on Population Forecasts

When comparing the effect of uncertainty of migration on the size and age structure of the population with the effect of uncertainty on fertility and mortality, two elements play a role: (1) the degree of uncertainty of the three components and (2) the age groups affected by the three components.

First, the degree of uncertainty of net migration differs from that of fertility and mortality. An indication of the degree of uncertainty of migration compared with the uncertainty of births and deaths can be obtained by examining fluctuations in the time series of migration, births and deaths. Assuming that if a time series of annual numbers is characterized by a constant change, forecasts are more certain than if there are irregular fluctuations (of course this certainty may only be apparent: trends can be interrupted), the degree of uncertainty can be measured by the standard deviation of observed changes. Figure 4 shows the standard deviation of changes by the number of years between the observations on the basis of Dutch data for the period 1950–1994. The figure shows that in the short run immigration is the most uncertain component, due to very strong annual fluctuations. However, in the long run the number of births is much more uncertain. The uncertainty of immigration does not rise strongly with the interval. One explanation is that after some years of increase there tends to be a fall, and *vice versa*. Fluctuations of the number of deaths are relatively small. Hence the number of deaths is relatively easy to forecast, even in the long run.

Second, migration affects different age groups than births and deaths. Errors in forecasting fertility affect only age groups not yet born in the initial year. Errors in forecasting mortality mainly affect the eldest age groups. The age pattern of net migration shows a peak between ages 18 and 30 (Figure 5). For older ages net migration is rather low. As a consequence uncertainty on migration only affects the older age groups in the long run. Although for young age groups there clearly is positive net migration, the size is relatively small compared with the number of births: for young ages the size of net migration is less than one per cent of the annual number of births. Hence the effect of uncertainty of migration on the youngest ages is small compared with the uncertainty of births.

The effect of errors of forecasts of migration in comparison with errors of forecasting fertility and mortality can be assessed by examining historic forecast errors. Alternatively, forecast intervals can be assessed.

4.1. Historic forecast errors

Figure 6 shows errors of forecasts of the population by age on 1 January 1995 from forecasts starting in 1980 and 1985. Errors are defined by observations minus forecasts.

For the 1980-based forecasts the errors of migration and fertility largely balanced each other out: the forecast of fertility was too high and the forecast of migration too low. As a result the forecast of total population size for 1995 was rather accurate.

However, for the separate age groups the results were not nearly as good. For example, for age group 0–14 years the forecast was 4 per cent too high, whereas for age group 20–39 years the forecast was 3 per cent too low. Aggregated over age groups the total error caused by migration was larger than the error caused by fertility. The errors that can be attributed to mortality were clearly smaller than those of fertility and migration.

In initial year 1985 the forecasts of both fertility and migration were too low. Hence the forecast error of total population size was much larger than that of the 1980-based forecast. The error of the fertility forecast resulted in high relative errors for the age group 0–9 years: the forecast was 8 per cent too low. The error of the migration forecast resulted

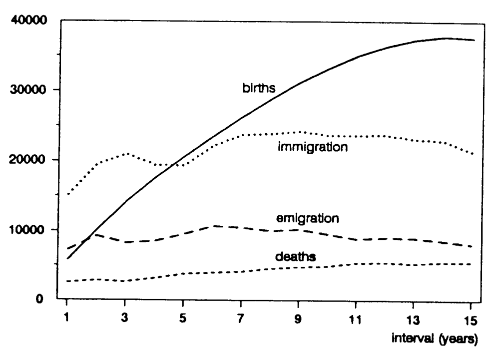

Fig. 4. Standard deviation of changes

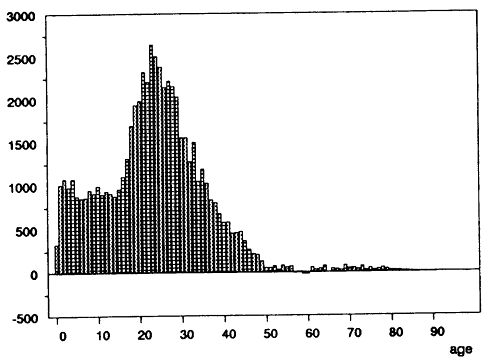

Fig. 5. Age pattern of net migration, 1994

in lower relative errors for the other age groups. For example, for age groups 10–44 years the forecast was 3 per cent too low. However, aggregated over the age groups the total forecast error caused by migration was much higher than that of fertility. For mortality there was a relatively large error for the oldest age groups. The forecast for people aged 85 years or older was 5 per cent too high. However, in absolute numbers the error of the mortality forecast was small compared to fertility and migration.

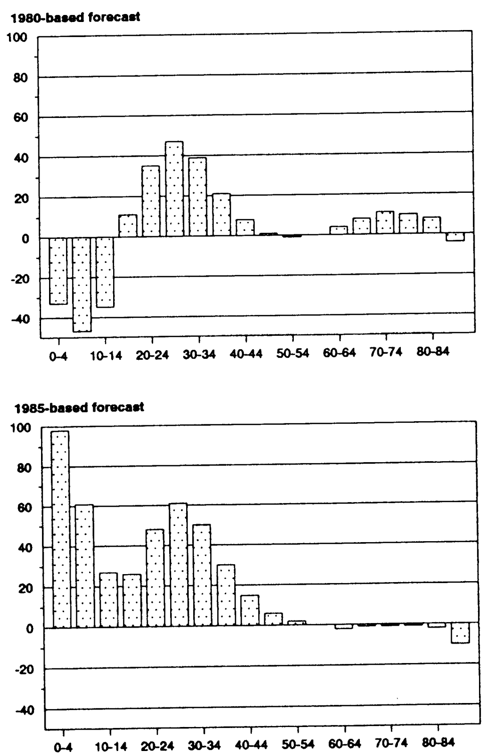

Fig. 6. Errors (×1,000) of forecasts of the population on 1 January 1995

4.2. Forecast intervals

The high variant of the Dutch population forecasts is based on the high variants of fertility, net migration, and life expectancy. One way of comparing the effect of the uncertainty of migration with that of fertility and mortality would be to compare the high variant with other variants, *viz.* variants in which the high variant of one component is combined with medium variants of the other two components (see e.g., Long 1991). One problem of such an approach is that the variants are based on a deterministic cohort-component model. In the high variant there is a perfect correlation between forecast errors in successive years. Although forecast errors are serially correlated, there is no perfect correlation.

Rather than calculating deterministic high and low variants, Monte-Carlo simulations can be used to assess forecast intervals of the size and age structure of the population. For that purpose ARIMA-models of forecast errors can be used. Section 3 discussed a first-order autoregressive model of forecast errors of net migration. This model was based on all Dutch population forecasts published since 1975. In previous population forecasts no assumptions on migration were included. Similarly, ARIMA-models of forecast errors of fertility and mortality can be identified on the basis of the population forecasts published since 1975.

For fertility, errors of forecasts of the period total fertility rate (TFR) are modelled and for mortality, errors of the life expectancy at birth of males and females. These forecast errors can be modelled by a random walk. Under the assumption that the age pattern of net migration is constant, the ARIMA-model of forecast errors of net migration is used to make simulations of net migration numbers by age. Similarly, assuming the age pattern of fertility to be constant, the model of forecast errors of the TFR is used to make simulations of age-specific fertility rates. On the basis of cubic polynomials describing the relationship between deviations of life expectancy from the medium variant and deviations in age-specific mortality rates, simulations of life expectancy can be used to calculate simulations of age-specific mortality rates. The simulations of age-specific migration numbers and fertility and mortality rates can be used to simulate the age structure of the population. The simulations are based on the assumption that the random terms of the ARIMA-models are normally distributed white noise.

Four types of simulations were done: one simulation of all three components simultaneously (assuming fertility and mortality rates and migration numbers to be mutually independent), and three separate simulations of fertility, mortality, and migration, assuming the other two components to be equal to the medium variant. On the basis of 2,000 replications of each type of simulation, 67 per cent forecast intervals were calculated.

Table 2 shows the upper bounds of the forecast intervals as a percentage of the medium variant. The top half of the table indicates that for forecasts some 20 years ahead migration is the main source of uncertainty on total population size. This does not apply to all age groups. Uncertainty on the number of people aged 0–19 years can be mainly attributed to fertility, whereas the uncertainty on the number of people aged 60 years or over is due to uncertainty on mortality. In the longer run the effect of the uncertainty of fertility increases more strongly than that of migration. In 2055 fertility is the main source of uncertainty on age category 0–39 years, whereas the uncertainty on age group 40–79 years can mainly be attributed to migration.

Table 2. Percentage difference between upper bounds of 67% forecast intervals and medium variant, based on simulations of fertility, mortality, amd migration, initial year in 1995

	Fertility	Mortality	Migration	All components
	Based on errors of forecasts published since 1975			
2015				
0–19 years	5	0	4	7
20–39 years	0	0	6	6
40–59 years	0	0	2	2
60–79 years	0	1	0	1
80+ years	0	5	0	5
Total	1	0	3	3
2035				
0–19 years	12	0	7	14
20–39 years	5	0	8	9
40–59 years	0	0	7	7
60–79 years	0	1	2	2
80+ years	0	9	0	9
Total	4	1	5	6
2055				
0–19 years	20	0	9	22
20–39 years	11	0	9	15
40–59 years	5	0	9	10
60–79 years	0	2	7	7
80+ years	0	12	3	13
Total	8	1	8	11
	Based on errors of forecasts published since 1950			
2015				
0–19 years	11	0	4	12
20–39 years	0	0	6	6
40–59 years	0	0	2	2
60–79 years	0	1	0	1
80+ years	0	6	0	6
Total	3	0	3	4
2035				
0–19 years	25	0	7	27
20–39 years	10	0	8	12
40–59 years	0	0	7	7
60–79 years	0	2	2	3
80+ years	0	10	0	10
Total	8	1	5	9

Table 2. continued

	Fertility	Mortality	Migration	All components
	Based on errors of forecasts published since 1950			
2055				
0–19 years	45	0	9	45
20–39 years	22	0	9	25
40–59 years	10	0	9	13
60–79 years	0	2	7	7
80+ years	0	14	3	15
Total	17	2	8	19

The forecast intervals shown in the top half of Table 2 are based on an analysis of forecast errors of the population forecast published since 1975, the reason being that, as remarked above, migration was not included in the forecast until 1975. One problem is, however, that the sharp decline of fertility rates in the late 1960s and early 1970s is not included in the sample period. Hence the uncertainty on fertility may be underestimated.

Keilman (1990) presents forecasts of the total fertility rate and life expectancy at birth of the 1950-, 1965-, and 1970-based forecasts. If the errors of these forecasts are taken into account, it turns out that the standard deviation of the error term in the random walk model describing forecasts errors of the total fertility rate is twice the standard error estimated on the basis of the forecast published since 1975. For life expectancy of males the standard error increases by 40 per cent and that of females by 20 per cent. Obviously this results in a larger effect of the uncertainty of fertility and mortality.

The bottom half of Table 2 shows that for the 20 years ahead forecasts the effect of uncertainty of fertility and migration on total population size is equal. In the longer run the effect of the uncertainty of fertility is much larger than that of migration. Fertility is the main source of uncertainty for all age groups born in the forecast period.

Table 2 shows that migration has a less pronounced effect on the age structure than both fertility and mortality. Even in the very long run the boundaries of the 67 per cent forecast interval caused by migration are within 10 per cent of the medium variant for all age groups, in contrast with the forecast interval caused by fertility for the young age groups and the interval caused by mortality for the eldest age group.

In comparing the top and bottom half of Table 2 one important question is, of course, which results reflect the 'true' degree of uncertainty. Obviously, the size of uncertainty cannot be quantified in a purely objective way. The assessment of uncertainty is partly based on judgment: to what extent will future forecast errors be similar to historic errors? Should the very large forecast errors of fertility in the 1960s be taken into account in assessing the uncertainty of the most recent forecasts? It can be argued that because of the improvement of forecast methods and the increase of knowledge on factors explaining changes in fertility the errors of recent forecasts will be smaller than those of forecasts published several decades ago.

5. Conclusions

Since the 1980s the role of international migration in population forecasting in the Netherlands has gained importance. First, the share of net migration in total population growth has increased. Second, as annual fluctuations in migration are larger than those in births and deaths, international migration has become an important source of uncertainty for the population forecasts. Forecast errors of migration account for more than half of the errors of the forecasts of population growth in the 1980s and early 1990s.

For Statistics Netherlands the increasing importance of migration as a source of uncertainty of population forecasts has lead to two types of response. First, various research projects were initiated in order to improve forecasts of migration. Second, Statistics Netherlands attempts to provide insight in the degree of uncertainty of the forecasts. For that reason high and low variants are published next to the medium variant.

Generally users of population forecasts pay relatively little attention to the high and low variants and focus their attention on the medium variant. Many users seem to underestimate the degree of uncertainty of forecasts. In presenting forecast intervals with an explicit statement on the corresponding probability it can be made clear to the users that there is a trade-off between the width of the forecast interval and the probability. Obviously, the forecasters can provide a forecast with a small interval, but it should be made clear to the users that this implies that the probability that the interval will cover the true value is low. For policymakers it is important to take the degree of uncertainty into account in decision-making if the costs of too low and too high forecasts are asymmetrical. If the costs of a decision based on a forecast that is too low are much higher than the costs due to a forecast that is too high, it may be advisable to use the high variant rather than the medium variant. This decision depends on weighing the probability of the variants and the width of the margin between the variants.

One problem in assessing the degree of uncertainty of population forecasts is that the forecasts depend heavily on judgment. Hence the forecast interval of future population size cannot simply be based on statistical time-series models of migration, fertility, and mortality. In analyzing historic forecast errors the effect of judgment is taken into account, as historic forecasts are based on judgment. Hence time-series models of forecast errors of migration, fertility, and mortality rather than models of the time series of the components themselves should be the basis of an assessment of uncertainty of forecasts.

On the basis of statistical time-series models of forecast errors of fertility, mortality, and migration, simulations can be used to assess the effect of the uncertainty of the three components on the future size and age structure of the population.

On the basis of the errors of the population forecasts published since 1975 it can be concluded that the effect of the uncertainty of migration on total population size and on most separate age groups clearly is larger than that of fertility and mortality some 20 years ahead. In the longer run the effect of uncertainty of fertility increases more strongly than that of migration. For forecasts some 60 years ahead the effect of the uncertainty of fertility and migration on total population size is equal.

Appendix. The ET-model

The ET-model is built up of three components. The short-run component is based on the change in the last observation year, the medium-run component is based on the average change during a longer period and the long-run component is based on a target value. The forecasts are a weighted average of these three components. The weights vary with the forecast lead time: the weight of the short-run component decreases with the lead time, whereas the weight of the long-run component increases. The weight of the medium-run component first increases and then decreases.

The ET-model is defined as follows:

$$y_{t+i} = \alpha_i(y_{t+i-1} + d_t) + \beta_i(y_{t+i-1} + \lambda_i c_t) + \gamma_i y_{t+I}$$

where:

y = variable to be forecasted;
t = last year of base period;
$t + I$ = the target year
y_{t+i} = forecast of y in year $t + i$ based on observations up to and including year t
$d_t = y_t - y_{t-1}$
$c_t = (y_t - y_{t-k})/k$
$\gamma_i = \mu^{I-i}$ if $i \leq I$
$= 1$ otherwise
$\alpha_i = \varphi^i$ if $\varphi^i + \mu^{I-1} \leq 1$
$= 1 - \mu^{I-i}$ otherwise
$\beta_i = 1 - \alpha_i - \gamma_i$
$\lambda_i = \theta^i$
$k > 1; 0 \leq \mu \leq 1; 0 \leq \varphi \leq 1; 0 \leq \theta \leq 1$

The short-run component depends on the value of d_t, i.e., the change in the last observation year. The medium-run component depends on the value of c_t, i.e., the average change in the last k observation years. The long-run component is based on the target value y_{t+I}.

The weight of the long-run component (γ) increases exponentially with the forecast lead time and equals 1 in the target year ($i = I$). The speed with which the target value is reached depends on the value of the parameter μ. If μ is close to 1, the forecast will move rapidly into the direction of the target value. If μ is small the target value will only affect the direction of the forecasts in the long run and as a result the short- and medium-run forecasts are completely determined by the changes in the observation period.

The weight of the short-run component (α) declines exponentially. The sum of α and γ should not exceed 1. The speed of the decline of α depends on the value of the parameter φ. If the value of φ is close to 1, the change in the last observation year has a strong effect on the forecasts. The opposite applies if φ is close to zero.

The weight of the medium-term component (β) is 1 minus both other weights. Hence this weight increases in the medium run (as α declines) and decreases in the long run (as γ increases). However, if φ or μ is close to one, the medium-run component hardly affects the forecasts.

The parameter λ determines the extent to which the average change in the last k observation years will be continued in the forecast period. If θ is smaller than one, λ declines exponentially. As a result the forecast will follow a damped trend. Note that the effect of the parameter λ largely depends on the value of φ and μ.

6. References

Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day.

De Beer, J. (1993). Forecast Intervals of Net Immigration: The Case of the Netherlands. Journal of Forecasting, 12, 585–599.

De Beer, J. (1994). A Judgemental Time-series Forecasting Model. Paper presented at the Fourth Meeting of the Eurostat Working Party on Demographic Projections, 30–31 May, Mondorf-les-Bains, Luxembourg.

De Beer, J., Kuijper, H., Noordam, R., Prins, K., and Sprangers, A. (1993). The Linking of Immigrant Stock and Flow Data in the Netherlands; Present and Future Possibilities. Statistical Journal, 10, 321–334.

George, M.V. and Perreault, J. (1992). Methods of External Migration Projections. In National Population Forecasting in Industrialized Countries, N. Keilman and H. Cruijsen (eds.). Amsterdam, Swets and Zeitlinger, 87–103.

Keilman, N.W. (1990). Uncertainty in National Population Forecasting: Issues, Backgrounds, Analyses, Recommendations. Amsterdam, Swets and Zeitlinger.

Long, J. F. (1991). The Relative Effects of Fertility, Mortality, and Migration on Projected Age Structure. In Future Demographic Trends in Europe and North America, W. Lutz, (ed.). London, Academic Press, 503–522.

Received June 1995
Revised September 1996