Statistisk sentralbyrå
Statistics Norway

A probabilistic forecast of the immigrant population of Norway

TALL SOM FORTELLER

Nico Keilman

DISCUSSION PAPERS

996

Discussion Papers No. 996, January 2023

Statistics Norway, Research Department

Nico Keilman

A probabilistic forecast of the immigrant population of Norway

Abstract:
We present a probabilistic forecast for the immigrant population of Norway and their Norwegian-born children ("second generation") broken down by age, sex, and three types of country background: 1. West European countries plus the United States, Canada, Australia, and New Zealand; 2. East European countries that are members of the European Union; 3. other countries.

First, we compute a probabilistic forecast of the population of Norway by age and sex, but irrespective of migration background. The future development of the population is simulated 3 000 times by stochastically varying parameters for mortality, fertility and international migration to 2060. We add migrant group detail using stochastically varying random shares to split up each result from the previous step into six sub-groups with immigration background, and one for the non-immigrants. The probabilistic forecast is calibrated against the Medium Variant of Statistics Norway's official population projection.

Keywords: stochastic forecast, immigrants, second generation, random share method

JEL classification: C150, J11

Acknowledgements: Discussions with Aslaug Foss in an early stage of the project are gratefully acknowledged. Astri Syse provided me with important data. Terje Skjerpen, Michael Thomas, Stefan Leknes, Ane Tømmerås, and Dinh Quang Pham gave useful comments.

Address: Nico Keilman, University of Oslo, Department of Economics, and Statistics Norway, Research Department. E-mail: nico.keilman@econ.uio.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

The Discussion Papers series presents results from ongoing research projects and other research and analysis by SSB staff. The views and conclusions in this document are those of the authors

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
https://www.ssb.no/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Sammendrag

Ingen prognose er eksakt, og derfor er det viktig å kvantifisere prognoseusikkerheten. Det betyr at prognosemakere bør beregne to typer resultater: for det første punktprognoser, som er så nøyaktige som mulig, og for det andre de statistiske fordelingene rundt punktprognosene.

Formålet med arbeidet rapportert her er å beregne statistiske fordelinger rundt prognoser for størrelse og alder- og kjønnsstruktur til innvandrerbefolkningen i Norge og deres barn. Som punktprognoser anvender vi resultatene av SSBs deterministiske befolkningsframskriving publisert i juli 2022. Vi brukte metoden med stokastiske andeler, som er brukt tidligere på husholdningsprognoser. Metoden starter med en stokastisk prognose for fremtidens befolkning fordelt på alder og kjønn. Hvert resultat av sistnevnte prognose er en stokastisk variabel. Denne variabelen blir kombinert med et sett med tilfeldige andeler, som deler opp hvert befolkningstall, gitt alder og kjønn, i tall for innvandrer-kategorier. Vi presenterer resultater for årene 2030, 2040, 2050 og 2060 for innvandrer-befolkningen i Norge og deres norskfødte barn («andre generasjon») fordelt på alder og kjønn. Vi skiller både innvandrere og deres barn etter tre kategorier som representerer landbakgrunn: 1. Vesteuropeiske land pluss USA, Canada, Australia og New Zealand; 2. Østeuropeiske EU-land; 3. andre land. Befolkningen uten migrasjonsbakgrunn utgjør en syvende gruppe.

Resultatene viser at noen få befolkningstrender som ble spådd for Norge til 2060 er ganske sikre: en sterk økning i størrelsen på innvandrerbefolkningen (nærmere bestemt de som tilhører gruppe 3) og av norskfødte barn av innvandrere. Når det gjelder alderstrukturene til innvandrere og deres barn, er prediksjonsintervallene rundt prognosene til disse personene i ettårsaldersgrupper så brede at det er lite informasjon i disse prognosene. For befolkningen som helhet (uavhengig av innvandrerbakgrunn) er prognosene for aldersstrukturen i ettårsaldersgrupper pålitelige frem til rundt 2040, med unntak av barn født etter 2022. For senere år blir intervallene svært brede for alle aldre. Men aldring er sikkert.

1. Introduction

Forecasts of the immigrant population are essential for government planning with respect to labour market and health policy, integration issues, and educational facilities. The future of this population sub-group is uncertain, but some developments are more likely than others. Therefore, probabilistic forecasts are a necessary tool for informed planning and decision making by policy makers.

Statistics Norway publishes projections for the population divided by age, sex, and migration back-ground at regular intervals. The most recent projections were published in July 2022; see Thomas and Tømmerås (2022). These projections are deterministic; uncertainty is accounted for by formulating several scenarios for the development of fertility, mortality, and international migration in the future. Whereas a scenario approach may be useful in case one is interested in future population trends based on a set of specific assumptions, the deterministic nature of the scenarios implies that uncertainty is not quantified. This makes it difficult for the user to select between the different scenarios. Also, when the user just selects the scenario results labelled as most likely by the producer of the projections, this may be a choice that is far from optimal. Take the example of a planner of educational facilities: under-predicting the number of schoolchildren may lead to hiring extra capacity, which may cost more than idle capacity in case of over-predictions. In such cases, the optimal choice is a trajectory a little or very much higher than the most likely trajectory - how much higher depends on the expected variation in the predictions. All this suggests a probabilistic forecast, not a deterministic one. Indeed, the Norwegian Ministry of Finance (more precisely, its Advisory Committee on Models and Methods), which is responsible for designing the country's long-term economic plans, has asked Statistics Norway to compute a probabilistic population forecast. However, one should note that the aim of a probabilistic forecast is not to present estimates of future trends that are more accurate than those computed in a deterministic forecast, but rather to give the user a more complete picture of prediction uncertainty.

A growing body of literature reports on stochastic demographic forecasts of various types, such as multi-country forecasts (see United Nations (2022) for all countries of the world, and Alho et al. (2006) for 18 European countries), forecasts for national populations (for early contributions see Lee and Tuljapurkar 1994; Alho 1998; Keilman et al. 2002), for regional populations (Wilson 2013a, 2013b), for households (Alders 1999, 2001; De Beer and Alders 1999; Scherbov and Ediev 2007; Alho and Keilman 2010; Christiansen and Keilman 2013; Keilman 2016), for the labour market (Fuchs et al. 2018), and for long-term care (Vanella et al. 2020). As to immigrant populations, a number of statistical agencies and individual authors have computed *deterministic* forecasts for this population sub-group (see Rees 2011 for a review), but very few have quantified the uncertainty surrounding future developments of immigrants. For exceptions, see Alders (2005) and Coleman and Scherbov (2005), to be discussed later. The aim of the current paper is to fill this gap, and to construct a probabilistic forecast for the migrant population of Norway.

Since our approach builds on methods frequently used in probabilistic household forecasts, we will discuss these below.

De Beer and Alders pioneered the field of stochastic household forecasts; see Alders (1999, 2001) and De Beer and Alders (1999), who applied their approach to data for the Netherlands. Alders and De Beer used stochastic simulation and combined a stochastic population forecast with forecasts of random shares. The shares distribute the population probabilistically over six household positions: individuals could live as a child with parents, live alone, live with a partner, live as a lone parent or in an institution, or belong to another category. Expected values for population variables and for the shares for specific household positions came from observed time series, but the statistical distributions of the shares were based on intuitive reasoning. Scherbov and Ediev (2007) combined a probabilistic population forecast for the population disaggregated by age and sex with random headship rates and applied their method to the case of Russia. A headship rate reflects the proportion of the population that is the head of a private household, for a given combination of age and sex (United Nations 1973; Jiang and O'Neill 2004). Scherbov and Ediev based a large part of their uncertainty distributions on intuition. Wilson (2013a, b) computed a probabilistic household forecast for Greater Sydney. Household parameters were modelled as random walks. Standard deviations of the random errors were based on judgement, due to the lack of past errors in estimates of living arrangements and households.

A problem connected to these probabilistic household forecasts is that uncertainty parameters were largely subjective judgemental. Alho and Keilman (2010) improved on this situation by estimating uncertainty parameters from data. Building on the random share method of De Beer and Alders, they applied their approach to Norwegian data. Yet simplifying assumptions had to be made, because only limited data were available. Christiansen and Keilman (2013) used long time series data of observed shares for Denmark and Finland, and formal time series methods to quantify the uncertainty connected to household shares in the future. Expected values of the shares came from a multi-state model of household dynamics. Keilman (2016) simplified the approach of Christiansen and Keilman, by modelling the age pattern of household shares in a given year by a Brass-type of model and constructing time series models for the parameters of the latter model.

Few probabilistic forecasts of immigrant populations exist. One has been reported by Alders (2005), but the author presented results only, not the method. The approach of Coleman and Scherbov (2005) relied heavily on expert opinions. The authors started with a deterministic cohort-component projection of the population of the UK from 2001 to 2100. The population was broken down into four ethnic groups: White, Asian, Black, and Mixed. High, Medium, and Low scenarios were formulated for future values of the total fertility rate, life expectancy at birth, and net migration. Subjectively chosen probabilities were assigned to the High-Low intervals for each of these three random variables in the years 2001, 2021, 2051, and 2100, while the Medium scenario was chosen as the mean of the distribution. The values at intermediate dates were determined using piecewise linear interpolation, and the results of 1 000 random simulations were analysed.

We present a probabilistic forecast for the immigrant population of Norway and their Norwegian born children ("second generation") broken down by age and sex. We adapt the random share method discussed earlier to data for the population with immigrant background. We distinguish both the immigrants and their children according to three groups of countries, see Section 2. The population without any migration background forms a seventh population subgroup. We start by updating an existing probabilistic cohort component type of population forecast of the population in Norway broken down by age and sex, but irrespective of migration background. The future development of the population is simulated 3 000 times by stochastically varying parameters for mortality, fertility and international migration for the years 2022 - 2060. We add migrant group detail to each simulation and use stochastically varying random shares to split up each result from the previous step into six sub-groups with immigration background, and one for the non-immigrants. The probabilistic forecast is calibrated against the Medium Variant of Statistics Norway's official population projection.

2. Immigrant population: definitions and issues

Whether a person is counted as an immigrant can be defined in several ways, and different definitions lead to different statistics. One could use rules based on nationality, on ethnicity, on having migrated to a different country, or simply on country of birth. Nationality is problematic, because persons may change nationality after migration. Thus, someone who used to be considered as an immigrant, becomes a non-immigrant simply as the result of a legal procedure. Ethnicity is problematic, because the issue can be sensitive and subjective, and difficult to define (Jacobs et al. 2009). An extremely simple rule is to consider as an immigrant anyone born outside the country. One consequence is that a child of native parents who temporarily resided abroad may be labelled as immigrant, and this is not very satisfactory in many cultural studies of migrants. Therefore, a narrower definition restricts immigrants to persons born abroad with one or both parents being foreign-born. Statistics Norway adds further restrictions for the number of grand-parents who were born abroad, see below. These types of restrictions are also helpful in case one defines the notion of "second generation". One possibility is to consider a person as second generation as soon he or she is born in the country with at least one parent and at least two grand-parents born abroad. Rules of this kind help to solve definitional problems in cases where one parent is an immigrant ("first generation"), whereas the other parent is not.

The definition of immigrant adopted here is the one used by Statistics Norway: see https://www.ssb.no/en/befolkning/innvandrere/statistikk/innvandrere-og-norskfodte-med-innvandrerforeldre. An immigrant is a person legally residing in Norway, who was born abroad to two foreign-born parents and four foreign-born grand-parents. Note that this definition does not in itself suggest any racial or cultural connotation, the criterion is place of birth of the parents and of grand-parents. Thus, a foreign-born person who migrates to Norway, and who has no, or only one foreign-born parent, or at most three foreign-born grand-parents, is not counted as immigrant. Of the 5.4 million persons who were registered in Norway on 1 January 2022, 819 000 were immigrants according to this definition. On the other hand, 898 000 persons, or 9.6 per cent more, were born abroad. The definition implies that a refugee or an asylum seeker is not counted as an immigrant until his or her application has been granted. Statistics Norway does not use the notion "second generation" but speaks instead of "Norwegian-born children with two immigrant parents". Immigrants and their Norwegian-born children together are denoted as "persons with immigrant background". One consequence of the definition for children is that a child with one immigrant and one native parent does not belong to the population with immigrant background.

Immigrants and their Norwegian-born children are classified according to country of origin. For immigrants this is the country of birth. For Norwegian-born immigrants, this is the parents' country of birth. If the parents are born in different countries, the mother's country of birth is used.

We have adopted the three country groups that Statistics Norway used in its population projection. This grouping is also part of Statistics Norway's standard classification of countries. Country group 1 comprises all the Western European countries, i.e. countries that were part of the 'old' EU (pre-2004) and/or the EFTA, as well as the US, Canada, Australia and New Zealand. Country group 2 comprises the eleven new EU countries in Eastern Europe (EU members in 2004 or later): Bulgaria, Croatia[1], the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, and Slovenia. Country group 3 comprises 'the rest of the world', e.g. the rest of Eastern Europe, Africa, Asia (including Turkey), South and Central America and Oceania (excluding Australia and New Zealand). See Thomas and Tømmerås (2022 p. 37, p. 153) for more details and a justification.

In Norway, the population statistics are based on the central population register. There are several quality issues relating to data on immigration and emigration. Vassenden (2015) gives detailed information. Vassenden reports a number of earlier studies have documented that many persons leave Norway, but do not notify the authorities. This means that they remain recorded in the population register as legally residing in the country. In many cases, authorities will discover, sometime after the actual emigration, that a person who is registered as having residence in Norway, no longer lives in the country. Indications could stem from several registers, such as the tax register, the welfare register, the educational register, or the employment register. In all cases, the record of the individual in question does not show any changes for a long time. Another indication is that the person no longer has a known address in Norway. The National Population Register has procedures for adjusting the status of persons who no longer reside in Norway ("administrative deregistration"). For the period 2004 – 2013, this concerned 26

[1] Croatia switched from country group 3 to country group 2 upon gaining EU membership in 2013.

per cent of all emigrations (Vassenden 2015). The share increased regularly, from 17 per cent in 2004 to 36 per cent in 2014. In 2019, however, there was a marked decline in the number of administrative deregistrations of individuals (Thomas and Tømmerås 2022, p. 102). These administrative procedures imply that statistics about immigrant stocks may lag behind actual developments, and that numbers are a few per cent too high.

3. Statistics Norway's projection

Statistics Norway has a long history of producing the official population projections for Norway, which goes back to at least 1969; see Texmon (1992). For many of the previous projections, future population trends were broken down by age, sex, and municipality of residence. However, as of the projections published in 2005, results for immigrant stocks were also included.

The most recent population projections were published in July 2022; see Thomas and Tømmerås (2022).[2] That report gives results on future trends in fertility, mortality, immigration and emigration, as well as population pyramids for the years 2022 – 2100. Immigrants from three country groups, Norwegian-born children with two immigrant parents, and the rest of the population were projected as separate groups. More detailed information is available from Statistics Norway's data base "StatBank"; www.ssb.no/en/befolkning/befolkningsframskrivinger/statistikk/nasjonale-befolkningsframskrivinger .

Different scenarios are provided for future fertility, life expectancy, and immigration. For each of these components, three different scenarios were created, labelled as High, Medium, and Low.[3] The main variant of the projections, labelled as "MMM", is based on a combination of medium fertility, medium life expectancy, and medium immigration. The MMM variant is the scenario considered as the most plausible. It should be noted that in the national population projections, immigrations and emigrations are calculated separately. Net migration constitutes the difference between the two. Whereas future immigration is estimated using a model, future emigration probabilities are based on observed emigration patterns. Thus, the projected emigration depends partly on the immigration assumption used. Relatively strong population growth ("HHH") results from combining high fertility assumptions with high life expectancy and high immigration, and low population growth ("LLL") is based on low assumptions for each of the

[2] This report gives results for the country as a whole. Regional projections for the population in municipalities were published in Leknes and Løkken (2022). This report is in Norwegian only, but a summary in English of findings and some background information is available at https://www.ssb.no/en/befolkning/befolkningsframskrivinger/artikler/municipal-population-projections-2022.

[3] There are four additional scenarios, primarily used for analytical purposes. Key assumptions here are constant immigration, constant life expectancy, no international migration, and equal immigration and emigration.

three components. The Main Variant projects a population size that grows from the current 5.4 million to 6.1 million in 2060 and 6.2 million in 2100. Population ageing continues: the share of persons aged 70 or more, which was around 6 per cent in 1950, is expected to increase further from today's 13 per cent today to 22 per cent in 2060 and 25 per cent in 2100. The number of young people (0 – 19) will remain fairly constant. By 2060, they will be outnumbered by the population aged 70+. The Main Variant also expects an increasing number of immigrants: 819 000 today to 1.18 million in 2060. The number of Norwegian-born to two immigrant parents is likely to more than double: 206 000 today and 437 000 in 2060.

4. Brief outline of the method

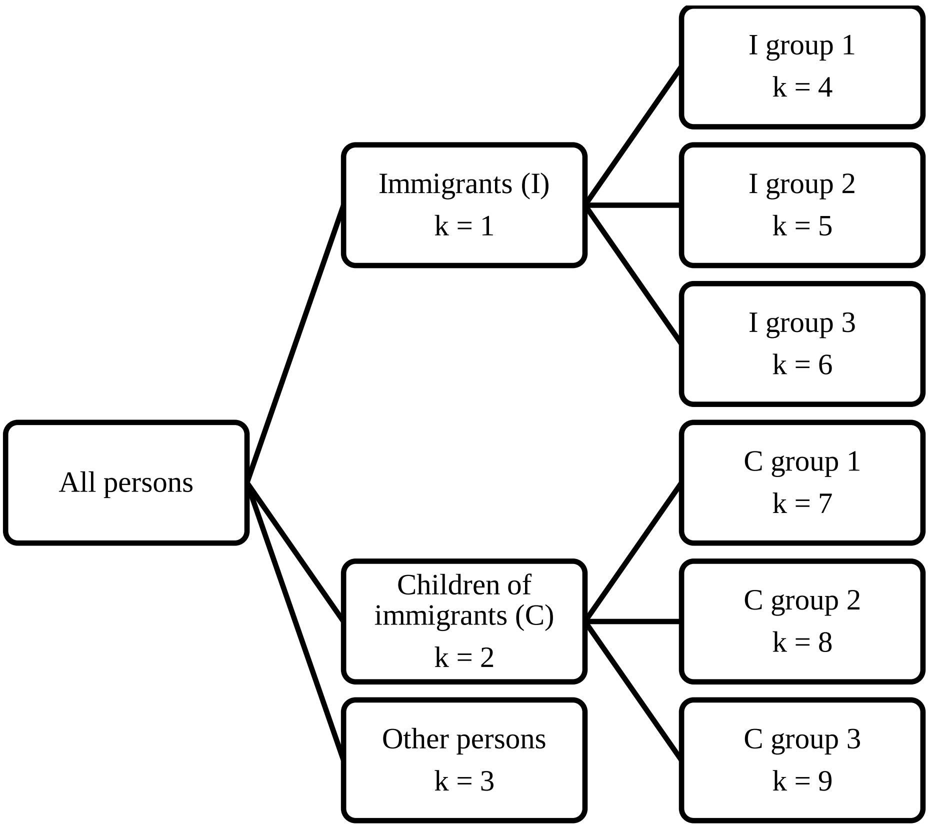
The first step consisted of stochastic simulations of a forecast model of the cohort component type for the population of Norway during the years 2022 – 2100, thereby updating an earlier stochastic forecast for the years 2020 – 2100 (Keilman 2020). Starting from the registered population broken down by age and sex as of 1 January 2022, the simulations were based on randomly chosen parameters for fertility, mortality, and international migration to 2100. This resulted in 3 000 trajectories for future numbers of men and women in Norway by one-year age groups for the years 2023 - 2100. The probability distributions for the parameters (fertility, mortality, international migration) were calibrated against corresponding numbers in the Medium Variant ("alternative MMM") of Statistics Norway's official population projection of 2022 (Thomas and Tømmerås 2022).

The second step was to add migrant group detail to the stochastic population forecast. Each simulated population number for a given age, sex, at a certain future year, was broken down into nine population subgroups according to immigration background, as defined in Section 2. This breakdown was achieved by means of shares that were randomly chosen from their assumed predictive distributions. The purpose of the modelling exercise was to obtain these distributions for the migrant group shares, disaggregated by sex and one-year age group. Each share, for a given year, age, sex, and migrant group, has an assumed normal probability distribution in the logit scale. In turn, this distribution was calibrated against the Medium Variant of Statistics Norway's 2022-based projection (Thomas and Tømmerås 2022). The result was a set of 3 000 trajectories for the population of Norway broken down by age, sex, and migrant group, for selected years: 2030, 2040, 2050, and 2060.

Following earlier work by Wilson (2013a, b), Christiansen and Keilman (2013), and Alho and Keilman (2010), we have adopted a tree-like structure when modelling the shares. Here we used two levels. First, the population (given age and sex) was divided into three groups: immigrants, Norwegian-born children of immigrants, and the rest of the population. Next, both the immigrants and their children were divided further into three country groups. This gave six groups of persons with an immigration background, in addition to the remaining part of the population (the members of which are without immigration background in the sense of the definitions of Section 2). Figure 1 shows the tree-like structure.

We modelled the shares for immigrants (k = 1) and their children (k = 2). We did not need to model the shares for the group of other persons (k = 3), because the three shares sum to one. Similarly, we modelled two of the three shares for immigrants (k = 4, 5, 6) and two of the three shares for the children of the immigrants (k = 7, 8, 9). For each group k and for both sexes, we have a table with observed values of the shares for the years 2000 – 2021, and ages 0 – 105. We assumed that the shares in each table can be written as a function of time and age. We extrapolated the function into the future and simulated predictive distributions for the extrapolated shares.

Figure 1. Tree-like structure of persons with immigration background. "Immigrants", "Children of immigrants", and country groups 1, 2, and 3 as defined in Section 2



5. Random shares

We write V(k,x,s,t) for the number of people in migrant group k = 1, 2,...,9 who are at age x = 0, 1,... and are of sex s = 1 or 2, at time t = 0,1, 2,.... The sum $\sum_k$V(k,x,s,t) gives the population W(x,s,t) of age x and sex s at time t, irrespective of migrant group. Migrant group k has share α(k,x,s,t) = V(k,x,s,t)/W(x,s,t) = α_k(x,s,t) in the population of age x and sex s at time t. The migrant groups are numbered as follows (cf. Figure 1): immigrants (k = 1), Norwegian-born children of immigrants (k = 2), other persons (k = 3), immigrants from country groups 1, 2, and 3 (k = 4, 5, and 6, respectively), and immigrants' children from country groups 1, 2, and 3 (k = 7, 8, and 9, respectively). Often, we will denote the various groups of interest by the following obvious codes: I for immigrants (k = 1), C for children of immigrants (k = 2), O for other persons (k = 3), I1, I2, and I3 for immigrants from country groups 1, 2, and 3 (k = 4, 5, and 6 respectively), and C1, C2, and C3 for immigrants' children from country groups 1, 2, and 3 (k = 7, 8, and 9 respectively).

For a given migrant group, year, and sex, we model the age profiles, in other words, the shares $\alpha_k(x,s,t)$ as a function of age. These age profiles are specified by means of a few parameters. The parameters may vary over time for men and women who belong to a certain migrant group. The focus is on finding appropriate functions for the age profiles, and appropriate time series models for the parameters of these functions.

5.1. Descriptive analysis for the period 2000 – 2021

Annual data on persons with immigration background with legal residence in Norway for the period 2010 – 2021 (1 January), broken down by sex, age (0 – 79 and 80+), and migrant group (k = 1 – 9) stem from the online databank of Statistics Norway; see https://www.ssb.no/en/statbank/table/13055/. In addition, we could dispose of similar data for the years 2000 – 2010.[4] We start with a descriptive analysis of the shares $\alpha_k(x,s,t)$.

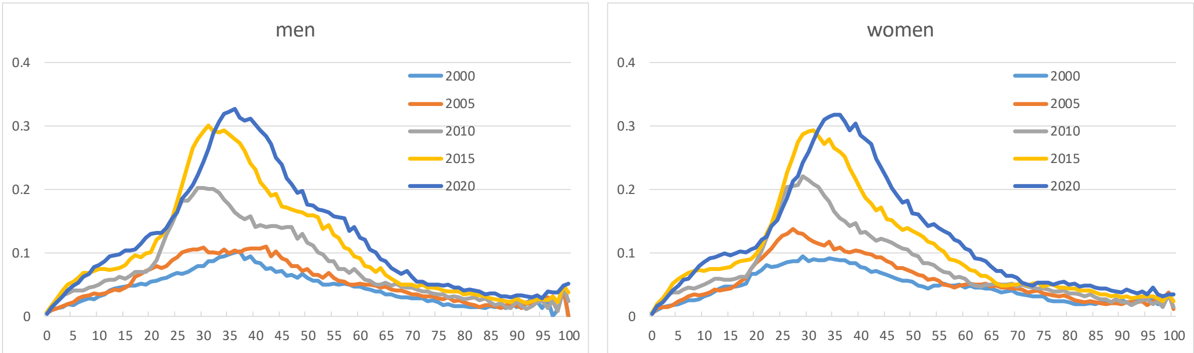
Figure 2 plots age profiles for the shares of immigrants $\alpha_1(x,s,t)$ for men and women aged 0 – 100 years for selected calender years. Note that these shares are aggregates over country groups. The shares are much larger between ages 20 and 60 than at other ages. The age profiles are very similar for men and women. We see a strong increase in the shares of immigrants after 2005, when new member countries joined the European Union, and immigration from these countries to Norway became easier. However, as we will see below, immigrants from group 3 countries contribute to this increase, too. The modes of the curves for 2010, 2015, and 2020 move systematically to higher ages over time, which suggests a cohort effect in the age profiles. Note that all age groups, including the youngest, concern persons born abroad. Age groups below 20, say, are children who immigrated, alone or together with one or both parents, or who came to Norway after adoption.

Figure 3 plots shares for immigrant men and women for each of the three country groups. The graphs illustrate that immigrants from country group 1 are less prevalent than those from

[4] For the years 2000 – 2010 we have data for ages 0 – 105, in one-year age groups. Since the shares for ages 80 – 105 are small and irregular, one is tempted to drop them from the analysis. However, we did not do so: as migrants who currently live in Norway get older, the shares for future years become much larger; see Section 5.4. For that reason, we imputed shares $\alpha_k(x,s,t)$ for ages 80 – 105 and years 2011 – 2021 by linear interpolation between 2010 and 2022. Data for the latter year for all ages 0 – 105 are available from Statistics Norway's 2022-based population projections; see https://www.ssb.no/en/statbank/table/13599/ .

elsewhere. Children and young adults have low shares for country groups 1 and 2, compared to group 3. This reflects the fact that many of the group 1 and 2 immigrants come as labour immigrants, whereas many group 3 immigrants have a background as refugee or asylum seeker, and family reunification is relatively frequent.

Figure 2. Age-specific shares of immigrants (k = 1) for men and women aged 0 – 100, selected years



After the enlargement of the European Union in 2004 with new Central and Eastern European member countries, immigration from that part of Europe increased considerably, as we can see in the plots for country group 2. However, immigrant shares for Western European countries (together with USA, Canada, Australia, and New Zealand) were rising slightly as well in this period, caused by peak immigration flows in the years 2007 – 2015. The curves for the remaining part of the world increase regularly. For a given country group, the age profiles show similar shapes for men and women.

Labour migration could be a factor that explains why men who belong to group 2 have somewhat higher shares in recent years than women. Note that the profiles in Figure 2, which are aggregates over country background, do *not* differ much between the sexes.

Next, we show a few plots with the age profiles for the shares of Norwegian-born children of immigrants. Figure 4 illustrates the findings irrespective of country group. The curves are very similar for boys/men and girls/women. The profiles increase regularly over time. Shares beyond age 50 are close or equal to zero. This reflects the fact that many of the immigrants came to Norway only a few decades ago, and hence their children who were born in Norway are relatively young.

Figure 3. Age-specific shares of immigrants from country groups 1 (k = 4), 2 (k = 5), and 3 (k = 6), for men and women aged 0 – 100, selected years

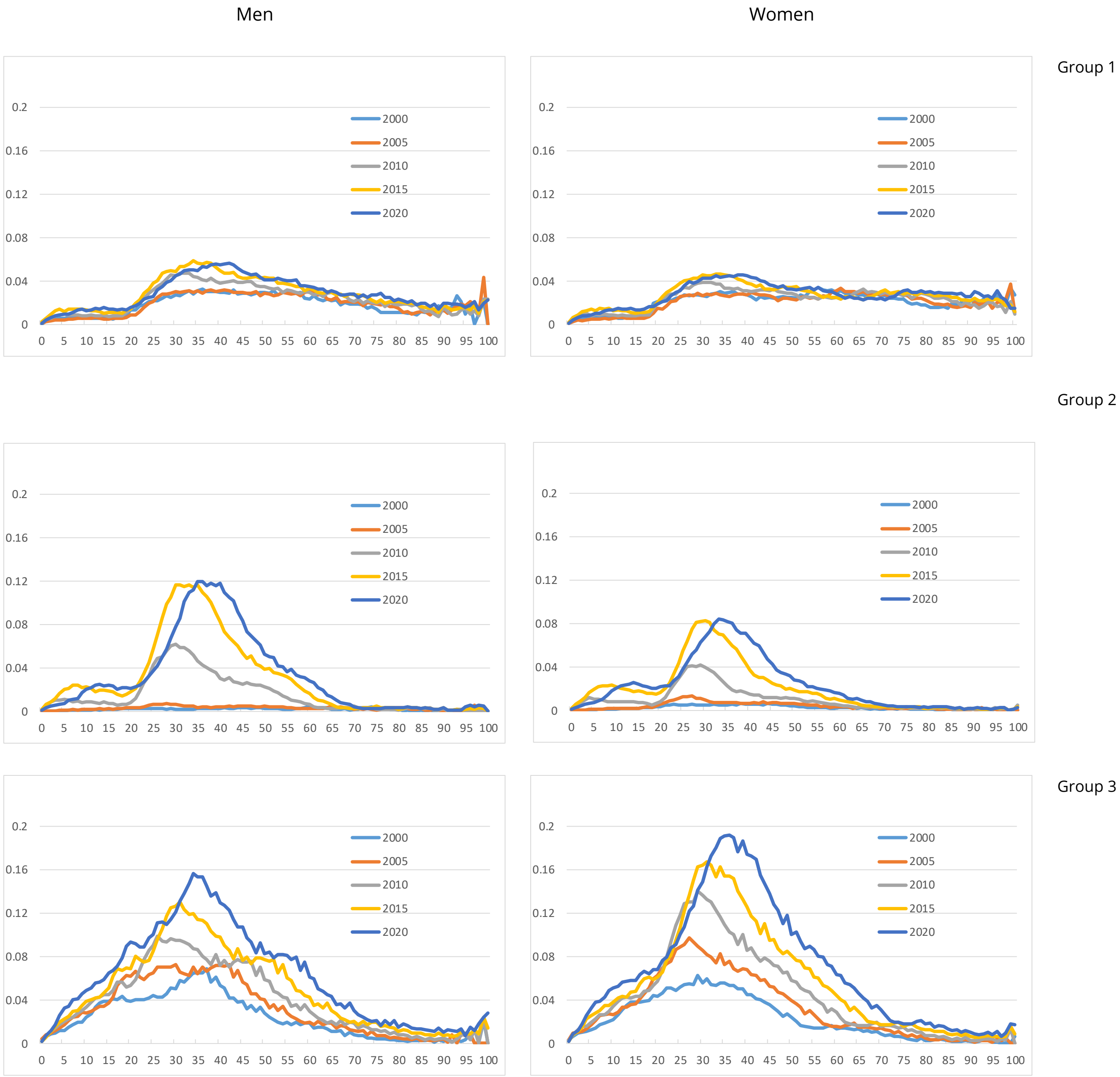
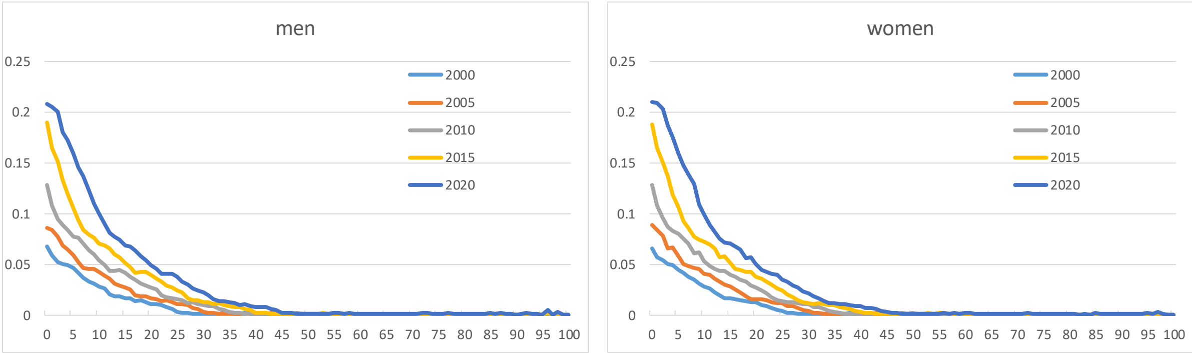


Figure 4. Age-specific shares of Norwegian-born children of immigrants (k = 2) for men and women aged 0 – 100, selected years



In Figure 5 we have added country group detail to the age profiles of immigrant children. The majority of these children have a country background in group 3. In all cases, we observe a more or less regular rise over time in the age profiles. For children from group 2, the increase did not start until around 2005, after the enlargement of the European Union.

5.2. Modelling the shares

To ensure that predicted shares are within the [0,1] interval, we have used a multinomial logit transformation. For immigrants (k = 1), Norwegian-born children (k = 2), and other persons (k = 3) and a given year t, age x, and sex s, define the transformed shares as

$$\beta_1 = ln\left(\frac{\alpha_1}{\alpha_3}\right) \text{ and } \beta_2 = ln\left(\frac{\alpha_2}{\alpha_3}\right), \quad \alpha_1 + \alpha_2 + \alpha_3 = 1$$

with α's as defined in the introduction to Section 5. The population subgroup "other" (k = 3) is arbitrarily selected as the benchmark.[5]

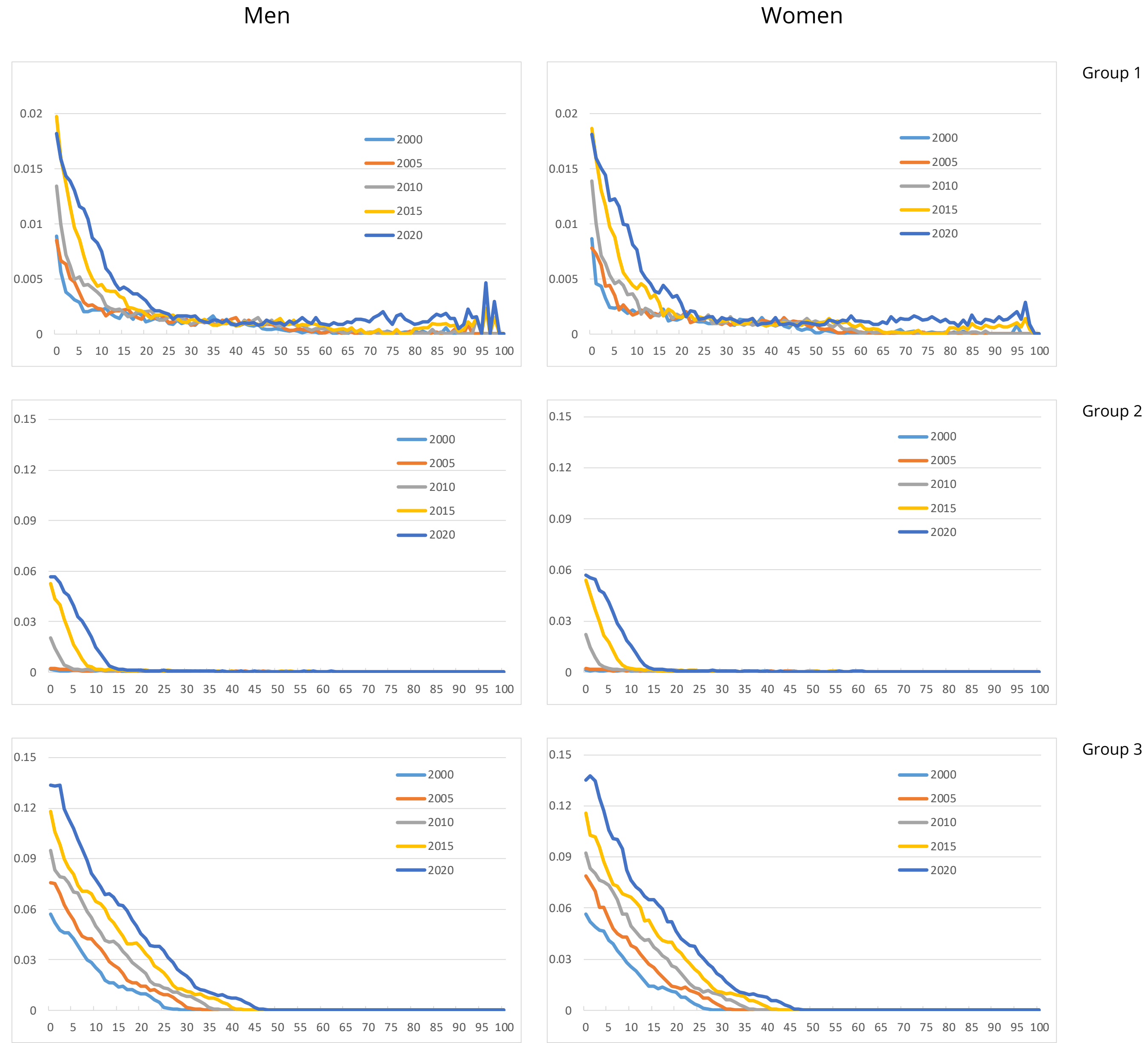
A second multinomial logit transformation defines country group specific shares of immigrants in the logit scale:

$$\beta_4 = ln\left(\frac{\alpha_4}{\alpha_6}\right) \text{ and } \beta_5 = ln\left(\frac{\alpha_5}{\alpha_6}\right), \quad \alpha_4 + \alpha_5 + \alpha_6 = \alpha_1,$$

using immigrants from country group 3 (k = 6) as the benchmark group.

[5] The choice of the benchmark category is arbitrary. For instance, choosing k = 2 as benchmark leads to the transformation $\gamma_1 = \ln(\alpha_1/\alpha_2)$ and $\gamma_3 = \ln(\alpha_3/\alpha_2)$. Then it follows that $\gamma_1 = \beta_1 - \beta_2$ and $\gamma_3 = -\beta_2$.

Figure 5. Age-specific shares of Norwegian-born children of immigrants from country groups 1 (k = 7), 2 (k = 8), and 3 (k = 9), for men and women aged 0 – 100, selected years.



Note: vertical scales differ between country groups

Finally, define logit-transformed shares for Norwegian-born children of immigrants from country groups 1 (k = 7) and 2 (k = 8) as:

$$\beta_7 = ln\left(\frac{\alpha_7}{\alpha_9}\right) \text{ and } \beta_8 = ln\left(\frac{\alpha_8}{\alpha_9}\right), \quad \alpha_7 + \alpha_8 + \alpha_9 = \alpha_2,$$

where children from country group 3 (k = 9) are the benchmark group.

The back-transformation from shares β_k (k = 1, 2, 4, 5, 7, 8) to the set of corresponding α's is straightforward; see Section 5.4.

The result of the logit transformation of shares is six sets of β's for immigrants and children, broken down by age (0 – 105 years), sex (men and women), and calendar year (2000 – 2021). This means that we have a total of 6 x 106 x 2 x 22 = 27 984 β-values, or 4 664 for each migrant group.

We assume that each β is normally distributed, with mean and variance that may depend on k, x, s, and t. The challenge is to predict them to future years, and to find the variances of the prediction errors. The predictions themselves follow from the Medium Variant of Statistics Norway's official forecast. We reduce the dimensionality of the problem and summarize the β's for a given k by a few parameters.

Inspection of Figure 2 – 5 suggests that there is no simple function of age with few parameters that describes the age profiles well (possibly with the exception of Norwegian-born children in Figure 4). Therefore, we have used a very general approach, and assumed

$$\beta_k(x,s,t) = a_k(t) + b_k(x,s) + e_k(x,s,t),\ k = 1, 2, 4, 5, 7, 8 \qquad (1)$$

The function $b_k(x,s)$ is commonly known as the *standard age profile* and the model describes how β in a certain year differs from the standard. This so-called relational approach has been used in the context of mortality (Brass 1971, De Beer 2012), fertility (Booth 1984, Zeng et al. 2000, De Beer 2011), and nuptiality (Coale and Trussell 1974). The well-known Brass relational model is a special case of model (1), namely one for a fixed time t. It was originally intended for modelling age-specific survival and can be written as $Y(x) = a + b \cdot Y^S(x) + e(x)$. Here, Y(x) is the logit-transformed probability of survival from birth to age x, while $Y^S(x)$ is some standard age pattern of survival, also in logit form. a and b are coefficients to be estimated from the data, and e(x) is an error term. Changing parameter a shifts the age pattern up or down relative to the standard, while b changes its slope. See e.g. Preston et al. (2001, pp. 199 – 201) for a thorough discussion.

To allow the maximum of flexibility, we adopted initially a non-parametric approach, and specified both $a_k(t)$ and $b_k(x,s)$ in expression (1) as a sum of terms, one for each year t (t = 2000, 2001, ..., 2021) and one for each age x (x = 0, 1, ..., 105). In addition, we assumed different age profiles for men compared to women, whereas Figures 2 – 5 suggested that the time effect $a_k(t)$ would be independent of sex. For a given migrant group, we assumed

$$\beta(x,s,t) = \sum_{i=2000}^{2021} a_i 1_i(t) + \sum_{i=0}^{105} b_{i,s} 1_i(x,s) + e(x,s,t) \qquad (2)$$

Here, the indicator function $1_i(j)$ equals 1 for i = j, and 0 otherwise. The coefficients a_t and $b_{x,s}$ are to be estimated from the data; they represent the time effects and the age effects, respectively, of the array β(x,s,t). For instance, for immigrants from country group 1, we found a positive trend in the coefficients a_t (t = 2000, 2001, ..., 2021). This implies that this migrant group has become more prevalent, compared to the immigrants from country group 3 (the reference group).

Model (2) contains many parameters. In order to reduce the risk of overfitting, we have attempted to estimate a more parsimonious model. The original α-shares relate to stocks of persons. Therefore, they tend to change slowly over time, although there are some irregularities, too. The same is true for the β's. As we will argue in Section 5.3, the time effects for the various groups showed very regular upward or downward trends, with two exceptions (immigrants and children from country group 2; see below). A special situation occurs when the time effect is a linear function of time. In that case model (2) can be simplified to

$$\beta(x,s,t) = A_0 + B_1 t + \sum_{i=0}^{105} b_i 1_{i,s}(x,s) + e(x,s,t) \tag{3}$$

This implies that the first difference of β with respect to time equals

$$\Delta\beta(x,s,t) = \beta(x,s,t) - \beta(x,s,t-1) = B_1 + d(x,s,t), \tag{4}$$

where d(x,s,t) = e(x,s,t) – e(x,s,t – 1). Model (4) represents a Random Walk with Drift (RWD). The time-increment in each β of a given age equals a constant value ("drift") plus a random term. However, it is unlikely that the time-increments are the same for each age. A more flexible model is

$$\Delta\beta(x,s,t) = A_1 + B_1.\beta^S(x,s) + d(x,s,t), \tag{5}$$

where $\beta^S(x,s)$ is a standard age pattern in the spirit of the Brass model, to be defined below. Note that model (5) for the increments Δβ(x,s,t) is consistent with a model for β(x,s,t) that includes an interaction effect between time and age (in addition to a time effect parameterized as a straight line).

One has to be prepared for error terms d(x,t) that are auto-correlated, because d(x,t) is the difference between two error terms. One solution to this problem is to extend the model as follows

$$\Delta\beta(x,s,t) = A_1 + B_1.\beta^S(x,s) + d(x,s,t)$$

(6)

$$d(x,s,t) = \rho.d(x,s,t-1) + u(x,s,t) ,$$

where u(x,t) is a random error term, and ρ is a first-order autocorrelation coefficient.

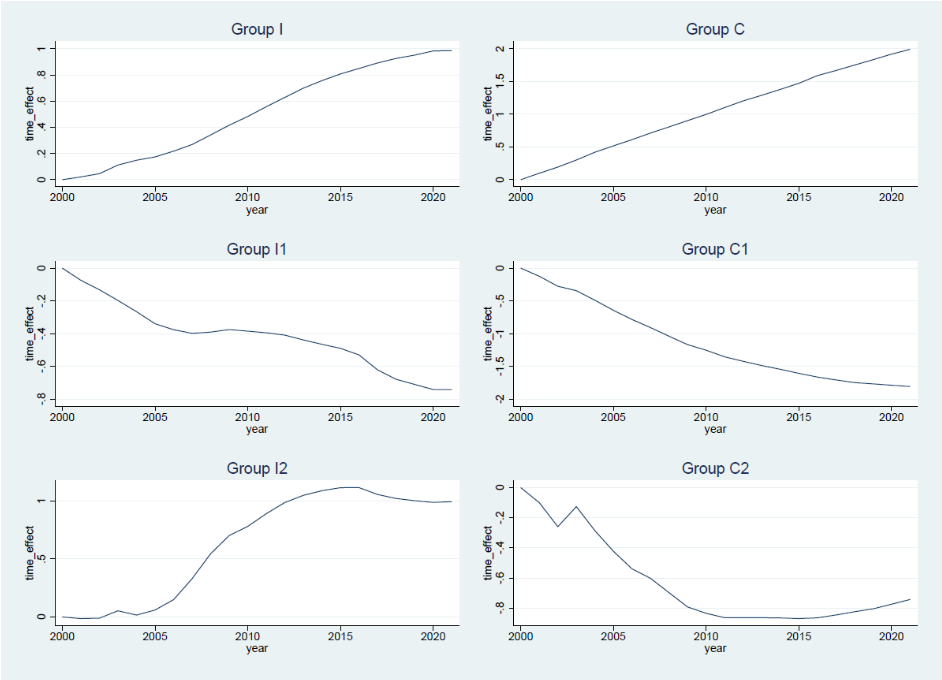
5.3. Model estimates

5.3.1 Estimates of model (5)

One can use regression by Ordinary Least Squares (OLS) for estimating time effects and age effects of model (2). To this end, one introduces a number of dummy variables as independent variables and selects one age and one year as reference age and reference year for the estimates. The dummy variables are assembled in a set of vectors, consisting of zeros and ones. This matrix is known as the "design matrix". For a given k, this results in estimates $\hat{\beta}_{x,s}^{OLS}$ for the age effects. Due to the nature of the OLS-solution, these estimates are identical, up to an additive constant, to the average of observed β-values across time, i.e. $\hat{\beta}_{x,s}^{OLS}$ equals a constant plus $\beta^S(x,s)$, where $\beta^S(x,s) = \sum_t \beta(x,s,t)/22$. The constant depends upon choice of reference year, and whether or not the model includes an intercept. Similarly, estimated time effects are identical, up to a constant, to the average of observed β-values across ages.

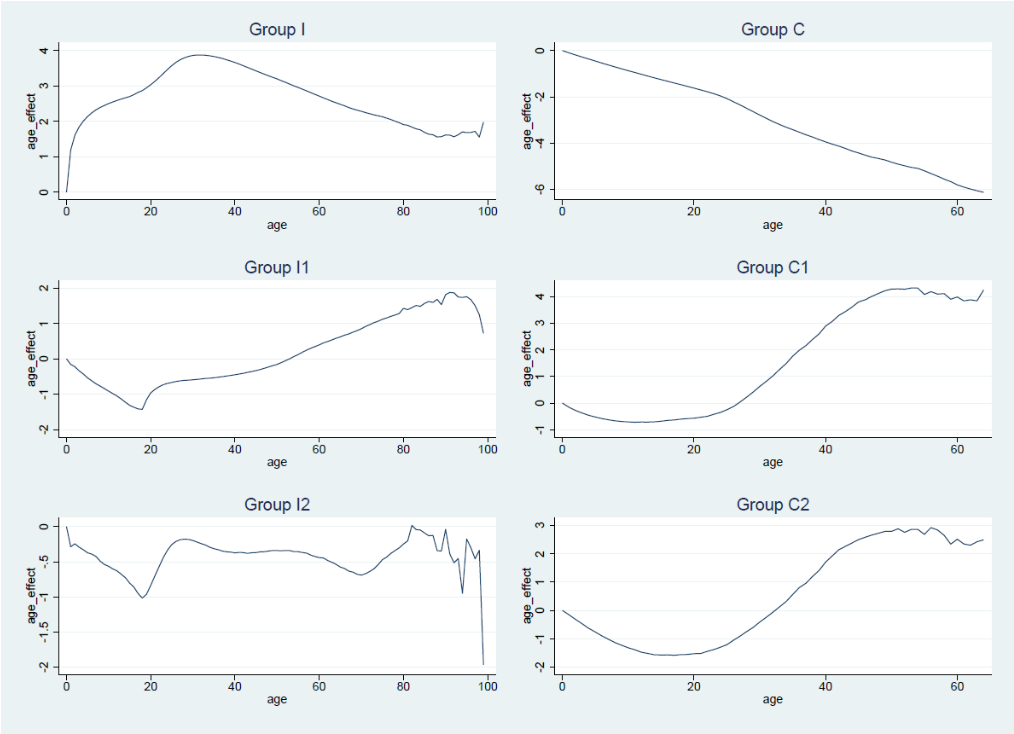
Estimation of the parameters of model (5) was done in two steps. First, we computed the time effects a_t and age effects $b_{x,s}$ in model (2) for each group by taking age-averages and time-averages, respectively, of observed β-values as outlined above; see Figures 6 and 7. Results for men and women were very close. Hence, Figures 6 and 7 show time effects and age effects for the two sexes combined. The age effects were very irregular at high ages, due to the small numbers involved, in particular for Groups C, C1, and C2. Therefore, we restricted computation of age effects for the latter three groups to ages below 70. When interpreting the results, one should keep in mind that for the group in question, the results are relative to both a reference year (year 2000 in Figure 6) or reference age (age 0 in Figure 7), *and* the share in the benchmark group. As an example, take time effects for immigrants irrespective of country group, i.e. group I, in Figure 6. Across all ages 0 – 105, β-values ("prevalence") for persons in this group increase faster than did the values for members of the benchmark group "other" (k = 3). Indeed, a large part of population growth in Norway during the first two decades of this century was due to an immigration surplus. Between 2000 and 2021 the population increased by 947,000 persons – international migration accounted for 62 per cent of the growth (Statistics Norway 2022). Country group 2 includes 10 countries in Central and Eastern Europe that joined the European Union in 2004, and two countries that became members in 2007. This explains the steep increase in the time effect for immigrants from these countries (group I2). The curve flattens out around 2015, when many refugees from Syria started to come to Europe, implying that the reference group, i.e. group I3, became more prevalent. Figure 6 shows also that immigrants from group 3 (compared to immigrants from country group 1) and their children became more prevalent during the period 2000 – 2021, as reflected in falling trends in time effects for groups I1, C1, and C2. Except for migrants and children from country group 2, the trends are very regular.

Figure 6. Time effects a_t. Year 2000 is reference year ($a_{2000} = 0$)



Age effects are more irregular and difficult to interpret. As an example, take immigrants irrespective of country group (group I) in Figure 7. Across all years 2000 - 2021, their β-values, i.e. their prevalence compared to the benchmark group, are much larger for adult than for young ages. Indeed, from Figures 2 and 4 we can conclude that across all years, β is roughly equal to ln(0.01/(1-0.01-0.1)) = -4.5 at age 0, but that it is approximately ln(0.25/(1-0.25-0)) = -1.1 at age 35. When we shift the curve with the age effects upwards, such that the effect is zero for the reference persons of age 0, we find large positive age effects for adults in this group. The regular age pattern for Norwegian-born children of immigrants (group C; k = 2) was already noted in connection with their observed shares in Figure 4.

Figure 7. Age effects b_x, for men and women combined. Age 0 is reference age (b_0 = 0)



Because of the regular time trends in estimated time effects, in a second step we adopted model (5) as a good representation of $\beta_k(x,s,t)$ for groups I, I1, C, and C1 for the period 2000 – 2021. For groups I2 and C2, a more or less constant time effect since 2015 seems to be a better basis for extrapolation into the future. For each group, we used for the age profile $\beta^S(x)$ the mean of observations $\sum_t \beta(x,t)/22$, as before.

Table 1 shows results irrespective of sex, as the estimates differed very little between men and women. In addition, the table gives estimates of the covariance between the estimators of the two parameters, and of the variances of the error terms d(x,s,t), to be used later.

The estimates of A_1 and B_1 are difficult to interpret, but note that all have very small standard errors, as reflected in high t-values. Yet the proportion of variance explained by the model (not shown in the table) is very low, typically eight per cent or less. In an attempt to improve on this, we included a possible cohort effect. The data series are short, just 22 years (six years for groups I2 and C2), which makes a full cohort analysis across ages 0 – 105 (groups I, I1, and I2) or even ages 0 – 69 (groups C, C1, and C2) impossible. Instead, for groups I, I1, C, and C1 we constructed an approximate "cohort standard" $\beta^C(x)$ as the average of observed β(x,t)-values along cohort lines in the β(x,t)-table, centred on the year 2010. This means we took the average of observed β(x+i,2010+i)-values for $\beta^C(x)$, i = -10, -9, -8, ..., +9, +10, +11, in other words $\beta^C(x) = \Sigma_i$ β(x+i,2010+i)/22. β(x,t)-values for "missing" ages and years (an upper left triangle for ages 0 – 9 and years 1990 – 1999, and a lower right triangle for ages 106 – 116 and years 2022 - 2032) had to be ignored, and averages $\beta^C(x)$ for these ages were computed for fewer than 22 β(x,t)-values. This is the procedure for men and women of groups I and I1. Cohort standards for groups C and C1 were computed in a similar way but involved fewer ages (0 – 69). Hence, they have missing lower right triangles for ages 70 – 80 (and years 2022 – 2032). Adding pseudo cohort standards computed along these lines does not improve the results: for each group, and both for men and women, we found that cohort standards are strongly correlated with the period standards $\beta^S(x)$ for men and women in groups I and I1 (correlations of 63 and 98 per cent). For children in groups C and C1, correlations between period and cohort standards are 91 and 99 per cent. Hence, it is not possible to asses an independent effect of the cohort standards. A possible explanation is that the effect of cohort standards is already included in model (5), because, as was noted before, this model is consistent with a model for β(x,t) that includes an age-time interaction. The latter interaction may be viewed as a way of expressing cohort effects (e.g. Luo and Hodges 2020, and the references therein).

Table 1. Parameter estimates for model (5). Data for the years 2000 – 2021 (groups I, I1, C, and C1); for the years 2015 – 2021 for groups I2 and C2. Men and women combined

		A_1		B_1			
Group	k	Estimate	t-value	estimate	t-value	cov(A_1,B_1)	σ_k^2
I	1	0.094	12.5	0.017	5.5	0.0231E-3	0.013
I1	4	-0.042	-10.5	-0.021	-4.9	0.0140E-3	0.036
I2	5	-0.146	-6.9	-0.087	-6.5	0.2776E-3	0.018
C	2	0.066	6.0	-0.006	-2.0	0.0335E-3	0.046
C1	7	-0.218	-8.7	-0.072	-6.8	0.259E-3	0.102
C2	8	-0.262	-4.0	-0.107	-4.9	1.3910E-3	0.129

Note: Student t-values based on robust standard errors.

For all six groups the residuals, when plotted in a histogram, showed a very symmetric shape, although a qq-plot indicated heavier tails than a normal distribution would imply.

5.3.2 Estimation of modified versions of model (5)

We considered a number of modifications of model (5), both as an attempt to improve the model, and as a check of the robustness of our findings.

We checked whether an AR(1) error term would improve the fit, and estimated model (6) by using the Prais-Winsten estimation procedure (Greene 2003). The first-order auto-regression coefficient ρ might be different for men and women, and hence estimations were done for each sex separately. All estimates of ρ turned out to be negative, but the values were moderate to small, i.e. between -0.41 and -0.0, with an average value of -0.24. Since the time series are short (22 or 7 years) and hence a parsimonious model with few parameters is to be preferred, we decided to ignore a possible auto-regression in the error term and take the results of Table 1 as the starting point for further analysis. The consequence is that estimators for the parameters in Table 1 still are unbiased, but that standard errors and t-values are incorrect. On average (across the 24 standard errors for two sexes, two model parameters, and six groups), the standard errors in model (5) are 24 per cent higher than the standard errors of the corresponding models with auto-correlated errors. Thus, we are a bit conservative, in the sense that we use standard errors of the estimates that are a little high. This is appropriate, given the often-observed underestimation of uncertainty in the prediction of densities (Armstrong et al. 2015, Makridakis et al. 2019).

Some support for the decision to ignore possible auto-correlation in the error term comes also from re-estimating model (5) using the Newey-West method (Greene 2003). This approach computes robust standard errors for the estimated model parameters taking both heteroscedasticity and auto-correlation of the error term into account. Following usual practice, we specified a maximum lag of $T^{0.25} = 22^{0.25} = 2.17 = 2$ years, where T is the number of years in the data series. Estimates of A_1 and B_1 were the same as those in Table 1, as expected, whereas Newey West standard errors were approximately 10 – 20 per cent lower than the robust standard errors underlying the t-values in Table 1.

A second attempt to improve the model was to add a quadratic term $[\beta^S(x,s)]^2$ for the standard age profile as an independent variable to model (5). In four out of six cases (groups I, I1, I2, and C), the estimate of the coefficient of the quadratic term was not significantly different from zero, at the five per cent level. For group C1, the estimate of the linear term $[\beta^S(x,s)]$ became non-significant, while that of $[\beta^S(x,s)]$ was significant (robust Student t-value equal to 3.4). Only in the case of group C2 was there a significant contribution of both the linear and the quadratic standard age profile. Since the model improved not systematically by adding a quadratic term, we decided to use a linear term only, as in expression (5).

We have used data for the years 2000 – 2021 (2015 – 2021 for groups I2 and C2). During that period, there were several shocks in international migration flows to and from Norway. These shocks may have had an effect on immigrant shares in later years. In 2020, the COVID-19 pandemic led to strong travelling restrictions, which limited international migration movements. In 2015, the war in Syria caused many refugees and asylum seekers in Europe. For Norway, this implied a large immigration flow from country group 3. Finally, the enlargement of the European Union in 2004 and later years led to increased labour immigration to Norway from country group 2. One may account for such a shock by including a dummy variable in the model, which takes the value one in the year the shock occurred, and zero for all other years. When the dummy variable turns out to be significant, it will reduce the residual variance of the model. Dummy variables of this kind were very useful in the models for immigration to Norway analysed by Cappelen et al. (2022). Nevertheless, we have not included them in our model (5). The main reason is that one cannot exclude the possibility that there will be events in the future that have a similar bearing on immigrant shares as those mentioned above. Hence, we do not attempt to reduce the residual variance, on purpose. Again, this is appropriate, as it reflects a cautious and conservative attitude towards density forecasting (Armstrong et al. 2015, Makridakis et al. 2019).

Finally, we checked if ignoring data for some early or recent years would have a strong impact on the estimates. We found very little effect for estimates of groups I, I1, C, and C1, when data for the years 2000 and 2001, or for 2020 and 2021 are omitted from our data series. For groups I2 and C2, the estimates in Table 1 are based on data for the years 2015 – 2021 only. With so few years of data, it is not useful to re-estimate the model with the years 2015 – 2016, or 2020 – 2021 omitted.

5.4. Predicted shares

Starting from a known value β(x,s,T) for a given group k, a future value h years ahead (h = 1, 2, …) is

$$\beta(x,s,T+h) = \beta(x,s,T) + h.\left(A_1 + B_1.\beta^S(x)\right) + \sum_{j=1}^{h} d(x,s,T+j) \quad (7)$$

The h-step ahead forecast E[β(x,s,T+h)] is estimated as β(x,s,T) + h. $(\widehat{A}_1+\widehat{B}_1.\ \beta^S(x))$, where we have replaced A_1 and B_1 by their estimated values. The forecast error F(x,s,T+h) equals β(x,s,T+h) – E[β(x,s,T+h)]. Given our assumptions, its variance Var[F(x,s,T+h)] can be estimated as

$$\mathrm{Var}\left[\sum_{i=1}^{h} d(x,T+i) - h.\left(\widehat{A}_1 + \widehat{B}_1.\ \beta^S\ (x)\right)\right] =$$

$$= h.\widehat{\sigma_s^2} + h^2.\mathrm{Var}\left[\widehat{A}_s\right] + h^2.\left(\beta^S\ (x)\right)^2 \mathrm{Var}\left[\widehat{B}_1\right] - 2.h.\beta^S\ (x).\mathrm{Cov}\left[\widehat{A}_1, \widehat{B}_1\right], \quad (8)$$

where σ_s^2 is the variance, for a given group k, of the error term $d_k(x,s,t)$ of model (5).

5.4.1 Correlations

When predicting the shares, one has to take into account possible correlations in several dimensions. The logit-shares $\beta_k(x,s,t)$ may be correlated across ages, across sexes, and between migrant groups. Since we model each β as a Random Walk with Drift process, it has independent increments and zero autocorrelation. We estimated correlations across migrant groups, ages, and between men and women from the residuals of model (5).

The residuals for six migrant groups have (6 x 5)/2 = 15 pairwise correlations. Of these, eight were negative, seven were positive. Thirteen correlation estimates turned out to be moderate or low: between -0.265 and +0.125. Eight estimates are not significantly different from zero at the five per cent level. Quite strong correlations are those between I1 and I2 (0.519), and between C1 and C2 (+0.337). The mean and the median values of the fifteen correlations are 0.019 and -0.0075, respectively. There is no clear pattern in the fifteen estimates: some are positive, others are negative – most correlations are modest or small, two of them are large. Since these results are hard to interpret, we have assumed that migrant groups are uncorrelated.

Table 2. Correlations between men and women, by migrant group

I	I1	I2	C	C1	C2
0.4681	0.5337	0.4589	0.0290	0.1685	0.3200

Table 2 shows pairwise correlations between men and women for the six migrant groups. They are higher for migrant groups I, I1, and I2 than for children of groups C, C1, and C2. One explanation is the following. The correlations derive from the residuals of model (5), which describes first differences in β-transformed shares. Since the shares reflect stocks, their first differences derive from changes in stocks. For migrant groups I, I1, and I2, the larger part of the changes stems from immigration, while mortality plays a minor role, because the migrants are relatively young. At the macro level, immigration for these groups is positively correlated between men and women. Shares for children groups C, C1, and C2 change due to mortality and outmigration (and fertility for age 0), because all children are born in Norway. In this case the numbers involved are much smaller, and hence the changes are more volatile and less systematic than changes caused by immigration for groups I, I1, and I2. In the simulations, we used the average correlations for groups I, I1, and I2 (0.4869) and groups C, C1, and C2 (0.1725).

Errors are possibly correlated across ages. An assumption of a first-order auto-regression (AR1) process for the errors in the age dimension has been used in similar earlier work (Alho and Keilman 2010, Christiansen and Keilman 2013). Under this model, the AR1-parameter is equal to the correlation between neighbouring ages. Table 3 gives estimated correlations by migrant group.

Table 3. Correlations across ages, by migrant group

I	I1	I2	C	C1	C2
-0.2032	-0.2876	0.1219	-0.1769	-0.2095	-0.2810

Five out of six migrant groups show estimates around -0.2. The negative values are surprising. They suggest that when a β-value for a certain age x is larger than expected, the values for neighbouring ages (x-1) and (x+1) are smaller than expected. The reason for this finding is unclear, but for groups I, I1, and I2 it might be associated with the volatility of annual migration flows. Note, however, that all correlations Corr[d(x,t),d(x+1,t)] are computed period-wise. As mentioned earlier, a cohort effect is visible in the shares for a number of groups. Indeed, *cohort-wise* correlations Corr[d(x,t),d(x+1,t+1)] turned out to be positive and strong, around 0.8 for groups I, I1, and I2, and 0.95 – 0.99 for groups C, C1, and C2. Since the results in Table 3 are difficult to interpret, and the values are modest to small, we have assumed that the β-values are uncorrelated across ages in a given future year, given sex and migrant group.

5.4.2 Predictions

As noted before, we did not use model (5) for predicting future values of the $\beta_k(x,s,t)$, but took them directly from the Medium Variant of Statistics Norway's population projection published in 2022. This way we have aligned our prediction intervals with the official projection. We used projection results for 2030, 2040, 2050 and 2060, computed shares $\alpha_k(x,s,t)$ for these years, and took the β-transformed values of these "target shares" as expected values of the $\beta_k(x,s,t)$.

Figures 8 and 9 extend the shares $\alpha_k(x,s,t)$ for groups I and C in Figures 2 and 4 with future values.

Figure 8. Age-specific shares of immigrants (k = 1) for men and women aged 0 – 100, selected years

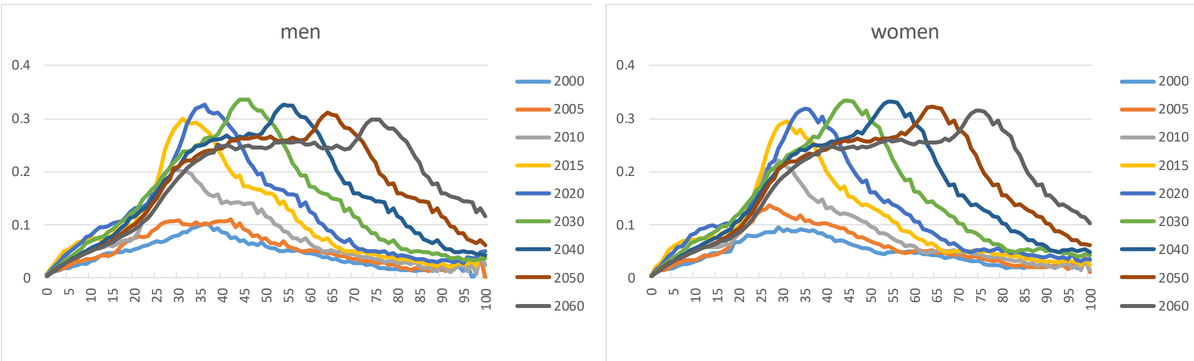
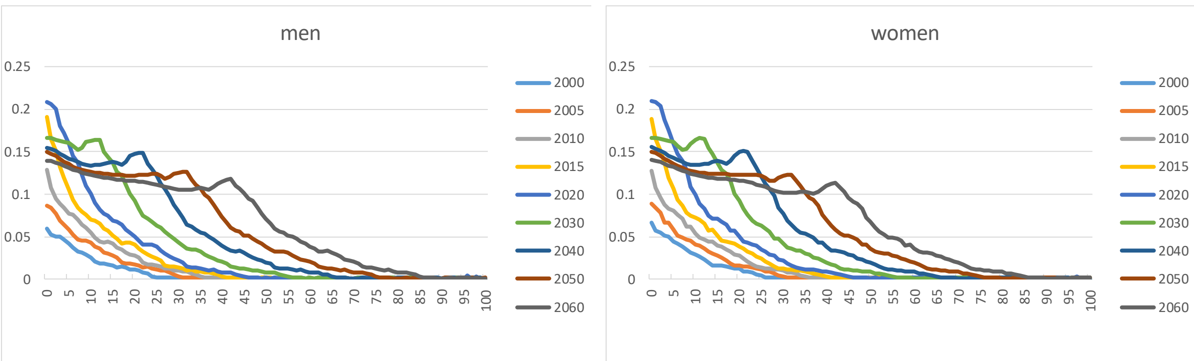


Figure 9. Age-specific shares of Norwegian-born children of immigrants (k = 2) for men and women aged 0 – 100, selected years



Statistics Norway predicts strong ageing among men and women belonging to group I born around 1985 – 1990, who show large shares (Figure 8). Plots based on numbers of immigrants by age (instead of shares) have similar shapes; see Figure 1.17 in the projection report (Thomas and Tømmerås 2022). The shares of younger cohorts level off at around 25 per cent as soon as they reach adult ages. As to children (Group C in Figure 9), a new pattern seems to emerge after 2020. Historical curves show a regular decline with age, and, at the same time, an increase over time. The latter increase disappears for future years, whereas the decline in the age direction becomes a bit irregular. The explanation is to be found in the shares of children of group C3, who make up by far the largest shares in group C (cf. Figure 5). The fertility of immigrant women from Asia and Africa, who form a major sub-group of mothers of C3-children, has systematically declined during the years 2011 – 2021, with a particularly strong drop starting in 2017.[6] Statistics Norway has extrapolated this decrease to future years, which results in age patterns of shares that decline over time. The small peaks in the age patterns for given years reflect the large number of children to which these women gave birth in the years 2014 – 2016. The start of these peaks is already visible, with a little imagination, in the curves for the year 2020 – both in Figure 9 for all children, and in Figure 5 for children of group 3. The patterns in Figure 9 for shares are in line with those in Figure 1.19 of the projection report (Thomas and Tømmerås 2022), based on absolute numbers.

Target values of the shares $\alpha_k(x,s,t)$ for the years 2030, 2040, 2050, and 2060 were transformed into $\beta_k(x,s,t)$-values, using the expressions of Section 5.2. The latter values served as expected values for the predictive distributions of $\beta_k(x,s,t)$. The variances of $\beta_k(x,s,t)$ follow from expression (8), and we assumed normality, as stated before. The distributions were simulated based on N=3 000 random draws for each of the four future years, and each combination of k, x, and s.

Expressions for the back-transformation from $\beta_k(x,s,t)$ to $\alpha_k(x,s,t)$ are readily derived, based on the definitions in Section 5.2. Temporarily suppressing x, s, and t, one finds

$$\alpha_1 = \frac{\exp(\beta_1)}{1+\exp(\beta_1)+\exp(\beta_2)},$$

$$\alpha_2 = \frac{\exp(\beta_2)}{1+\exp(\beta_1)+\exp(\beta_2)},$$

$$\alpha_3 = \frac{1}{1+\exp(\beta_1)+\exp(\beta_2)},$$

$$\alpha_4 = \frac{\alpha_1.\exp(\beta_4)}{1+\exp(\beta_4)+\exp(\beta_5)},$$

$$\alpha_5 = \frac{\alpha_1.\exp(\beta_5)}{1+\exp(\beta_4)+\exp(\beta_5)}, \qquad (8)$$

[6] The Total Fertility Rate of African women living in Norway was 2.78 in 2017, 2.55 in 2018, 2.40 in 2019, 2.20 in 2020, and 2.15 in 2021. For women from Asia (incl. Turkey) the decline was weaker. (https://www.ssb.no/en/statbank/table/12481/tableViewLayout1/).

$$\alpha_6 = \frac{\alpha_1}{1+\exp(\beta_4)+\exp(\beta_5)},$$

$$\alpha_7 = \frac{\alpha_2.\exp(\beta_7)}{1+\exp(\beta_7)+\exp(\beta_8)},$$

$$\alpha_8 = \frac{\alpha_2.\exp(\beta_8)}{1+\exp(\beta_7)+\exp(\beta_8)},$$

$$\alpha_9 = \frac{\alpha_2}{1+\exp(\beta_7)+\exp(\beta_8)}.$$

More formally, we assumed that for a given combination of k, x, and s, the distribution of β in a future year t is $N(\mu,\sigma^2)$, where μ is the β-transformed value of the target share α, and σ^2 follows from expression (8). 3 000 random numbers β^r (r = 1, 2 ,..., 3000) were drawn from this distribution, and each β^r was transformed to a corresponding α^r.[7] This resulted in 3 000 simulations for each share $\alpha_k(x,s,t)$, for nine population subgroups (k = 1 – 9), 101 ages (x = 0, 1, 2, ..., 100+), men and women (s = 1, 2), and four years (t = 2030, 2040, 2050, 2060).

For a given migrant group, age, sex, and year, the 3 000 predicted shares α^r (r = 1, 2, ..., 3 000) were multiplied with 3 000 simulated population numbers W^r (irrespective of migrant group; see the introduction to Section 5) from a stochastic population forecast, resulting in 3 000 numbers V^r. The stochastic population forecast updates a similar forecast for Norway published in 2020 (Keilman 2020; see also Foss 2012). We replaced the jump-off population from the previous stochastic forecast by the registered population broken down by age and sex as of 1 January 2022. Next, we used age- and sex-specific rates and numbers for fertility, mortality, and net migration from the Medium Variant of Statistics Norway's 2022 national projections as point predictions for the updated stochastic forecast. Finally, uncertainty parameters for fertility, mortality, and net migration, i.e. variances for vital rates and migration numbers, as well as (auto-)correlations between these rates and numbers, were taken from the previous stochastic forecast. We assumed relatively high variances for vital rates and migration numbers for the years 2022 - 2026, due to uncertainty about the effects of the Covid-19 pandemic and the war in the Ukraine. See Keilman (2020) for a number of details.

The multiplication $\alpha^r.W^r = V^r$ implicitly assumes that the random variables for the share α and the population number W are uncorrelated. This is a reasonable assumption, since both represent population *stocks* (population as such for the population numbers, immigrants and their children for the shares). *Changes* in the stocks are probably correlated. Given age and sex, migration and mortality cause changes in the population as such and changes in the share of immigrants (fertility of immigrant women has no impact on the shares of immigrants, because the

---

[7] Random numbers were drawn such that they reflected the assumed correlation between men and women. When stochastic variables X, Y, and Z all have a standard normal distribution, then we form new stochastic variables (X+aY) and (Z+aY). Their correlation is $a^2/(1+a^2)$, for some number $a = \sqrt{(r/(1-r)}$, where r is the required positive correlation coefficient. Negative correlation results from taking (X+aY) and (Z-aY).

Norwegian-born children of these women form a separate group). Likewise, fertility causes changes in the number of new-born children and in the share of Norwegian-born children of immigrants. Hence, we may expect a correlation between changes in the population as such and changes in shares of immigrants. However, these effects are of second order importance, compared to the stocks represented by population numbers and shares for immigrants and their Norwegian-born children. For this reason, we have ignored correlation between the two sets of random variables.

The result of the multiplication of random shares with random population trajectories was a set of simulated values V^r (r = 1, 2, ..., 3 000) for the population broken down by sex (men, women), age (0, 1, 2, ..., 99, 100+), and seven categories defined by migration background (immigrants and Norwegian born children, both for three country sub-groups, and other persons) for each of the years 2030, 2040, 2050, and 2060. It turned out that in many cases, the mean of the simulated values $(\Sigma_r V^r)/3000$ for a certain combination of age, sex, migrant group, and calendar year was higher than the corresponding target value V from the official projections. In some cases, the target value was even below the 10-th percentile, or larger than the 90-th percentile, of the set of V^r-values. The discrepancies were larger for 2050 and 2060 than for 2030 or 2040. The difference between the mean of the simulated V^r-values and the target value V is caused by the shares α^r computed from the exponential back-transformation defined in expression (8). The details are complicated, but an approximate argument can illustrate this. Assume that a random variable X has a normal distribution $N(\mu,\sigma^2)$. Define a new random variable as Y = exp(X). Then Y has a log-normal distribution with expected value equal to $\exp(\mu + ½\sigma^2)$, which is larger than $\exp(\mu)$ by a factor $\exp(½\sigma^2) > 1$. Although the situation in our case is a bit more complicated, with a logit transformation and several random variables simultaneously, the argument is similar. The random variable Y above corresponds to our share α, and X corresponds to β. Each α^r is an exponential transformation of a simulated β^r, yet the mean across all 3 000 α^r-values differs from the exponentially transformed mean of β^r-values, which corresponds with the expectation μ. The discrepancy is larger, the larger the variance of the β-estimate is.

We proportionally adjusted each simulated number V^r, given age, sex, migrant group, and calendar year, by the ratio of the corresponding target value V and the mean of the 3 000 simulated values. This led to a mean value across the simulations that is equal to the target value.

6. Main results

6.1 Total population

The results for the stochastic forecast show median predicted population sizes in 2030, 2040, 2050 and 2060 equal to 5.66, 5.88, 6.02 and 6.05 million, respectively. To compare, the Medium Variant of Statistics Norway's population projection gives 5.66, 5.89, 6.03, and 6.11 million for these four years. These numbers are the same as the average values for our simulations, as one could expect. The 80 per cent prediction intervals are - in millions - [5.57-5.75], [5.66-6.13], [5.63-6.46], and [5.50-6.77] for these four years. On the other hand, results for the Low and the High Variants of the official projection show much wider intervals – for example, [5.18-7.09] million for 2060. The wide intervals are caused by the way Statistics Norway constructed the Variant projections. For example, in the High projection Variant it is assumed that fertility is high in *all* future years, and vice versa for the Low Variant. Similar assumptions are implicit in the High and Low Variants of life expectancy and of net migration. The stochastic forecast for the population by age and sex assumes that fertility, mortality, and net migration do not have perfect autocorrelation. This means that birth rates may be higher than expected in one year, but lower the year thereafter, and similarly for death rates and migration numbers. Moreover, fertility, mortality, and migration are stochastically independent of each other.

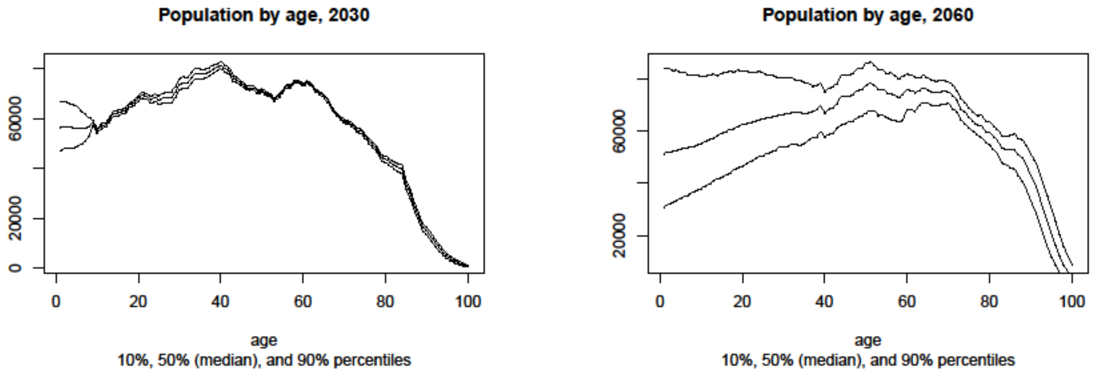
Uncertainty differs strongly between age groups. Prediction intervals are very narrow until roughly 2040, except for children born in the years 2022 – 2039. This means that forecasts of adults and elderly are rather certain during the first few decades of the forecast period. For later years, uncertainty increases gradually for all age groups. As an illustration, Figure 10 shows the median value and 80 per cent prediction intervals for the age distributions in 2030 and 2060.

The results of the stochastic forecast for total population and for the population pyramid are very similar to those of the stochastic forecast published in 2020. For instance, the 80 per cent prediction interval for total population size in 2060 was [5.5-7.0] million in the earlier forecast (Keilman 2020, p. 179). For this reason, the focus in this section will be on the findings for immigrants and their children.[8]

---

[8] Detailed results for population size or age structure are available upon request.

Figure 10. Age distribution, 2030 and 2060. The upper and lower curves are 90 per cent and 10 per cent percentiles (upper and lower bounds of the 80 per cent prediction intervals) of the predictive distribution. The middle curves represent median values



6.2. Immigrants

Table 4 gives median values, as well as upper and lower bounds of 67 and 80 per cent prediction intervals for the size of the population sub-group of immigrants (irrespective of country group) for selected years between 2030 and 2060.

Table 4. Number of immigrants in 2022 (registered), and 2030, 2040, 2050, 2060

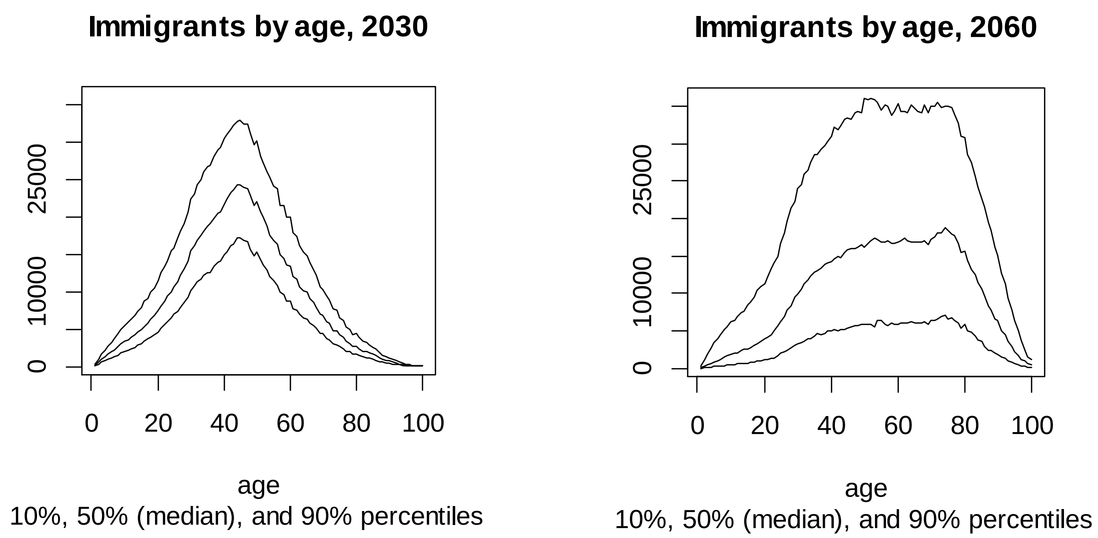
	2022	2030	2040	2050	2060
Median	819	956	1064	1140	1179
67% prediction interval		[920-992]	[1007-1124]	[1065-1219]	[1081-1283]
80% prediction interval		[908-1003]	[988-1143]	[1042-1247]	[1054-1314]
Medium Variant		956	1065	1143	1182
Low-High interval		[902-1034]	[960-1218]	[977-1403]	[949-1585]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations. Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands.

Table 4 shows a strong growth in the number of immigrants in the next four decades: the median value in 2060 is 44 per cent higher than the current 819,000. The lower bounds of the 80 per cent intervals tell us that the increase is almost certain. Chances are 90 per cent (odds nine to ten) that there will be at least 920,000 immigrants in 2030, and 1.054 million in 2060 – numbers that are much higher than today's number. However, we are not at all certain about the speed of the increase and how many more immigrants there will be, since the 80 per cent interval for 2060 is rather wide: 22.1 per cent of the median value ((1314-1054)/1179). Expressed this way, uncertainty grows regularly from 9.9 per cent in 2030, to 14.6 and 18.0 per cent in 2040 and 2050. At the same time, the interval between Statistics Norway's Low and High Variants indicates unduly large uncertainty.

The age distributions in Figure 11 suggest that even in 2030, predicted numbers for immigrants aged 30 – 60, say, have so wide prediction intervals that the results for one-year age groups bear little information. By 2060, this is the case for ages between 10 and 90, roughly speaking. In other words, in case one needs information about the age structure of immigrants in the future, this can only be in the form of broad age groups in order to be meaningful.

Figure 11. Age distribution of immigrant population, 2030 and 2060. The upper and lower curves are 90 per cent and 10 per cent percentiles (upper and lower bounds of the 80 per cent prediction intervals) of the predictive distribution. The middle curves represent median values



Tables 5-7 repeat the layout of Table 4, but present separate results for three sub-groups of immigrants according to country background (groups I1, I2, and I3).

Table 5. Number of immigrants from country group 1 in 2022 (registered), and 2030, 2040, 2050, 2060

	2022	2030	2040	2050	2060
Median	167	179	187	192	192
67% prediction interval		[166-192]	[169-208]	[168-217]	[166-220]
80% prediction interval		[163-196]	[164-215]	[162-227]	[158-231]
Medium Variant		179	188	193	193
Low-High interval		[175-187]	[178-209]	[176-230]	[168-252]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations, Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands

Table 6. Number of immigrants from country group 2 in 2022 (registered), and 2030, 2040, 2050, 2060

	2022	2030	2040	2050	2060
Median	203	224	240	247	243
67% prediction interval		[209-239]	[217-265]	[217-279]	[209-278]
80% prediction interval		[204-245]	[210-274]	[207-290]	[199-290]
Medium Variant		224	241	248	244
Low-High interval		[209-236]	[208-264]	[196-285]	[174-301]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations, Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands.

The median forecast in Table 5 and the expected value/Medium Variant suggest a slight increase in numbers of immigrants from country group 1, although the growth seems to flatten out by 2050. However, we are uncertain whether there will be an increase, because the 80 per cent prediction interval to 2060 covers the current value of 167,000. The 80 per cent interval is 38.0 per cent wide, relatively speaking. In other words, uncertainty is larger for this sub-group than for all immigrants in Table 4, as one could expect.

Statistics Norway notes a small increase for immigrants from country group 2 to around 2050, and a slight fall to 2060. Our median value in Table 6 shows the same trajectory, but the prediction intervals indicate that the development may have been very different, once we will know the actual numbers. Uncertainty is large, with a relative width of the 80 per cent interval in 2060 equal to 37.4 per cent.

Table 4 suggests that there will be more immigrants in the future. At the same time, Tables 5 and 6 indicate that it is not certain that the growth concerns immigrants who belong to country groups 1 or 2. Therefore, eventual growth must come from the remaining group 3. The results in Table 7 confirm this. With lower bounds for the 80 per cent intervals in the years 2030 – 2060 that are well above the current number (449,000), the conclusion must be that chances are less than 10 per cent that this immigrant group will not increase. Although we can be quite sure that there will be an increase, we do not know how strong the growth will be. The reason is that the 80 per cent interval in 2060 is rather wide. It amounts to 27.9 per cent of the median value, which indicates a bit more uncertainty than the results for all immigrants in 2060 in Table 4.

Table 7. Number of immigrants from country group 3 in 2022 (registered), and 2030, 2040, 2050, 2060

	2022	2030	2040	2050	2060
Median	449	552	635	700	741
67% prediction interval		[528-578]	[592-679]	[643-761]	[666-824]
80% prediction interval		[520-585]	[579-694]	[624-780]	[644-851]
Medium Variant		552	636	702	744
Low-High interval		[518-611]	[573-745]	[605-887]	[607-1032]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations, Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands.

6.3. Norwegian-born children of immigrants

Statistics Norway projects a strong increase in the number of Norwegian-born children of immigrants. The results in Table 8 confirm this. The median value more than doubles from 2022 to 2060. One can be quite certain about an increase: the lower bound of the 80 per cent interval in 2060 is 350,000, which is 70 per cent higher than today's number of 206,000 children. Note that the Low-High interval of the official projections agrees quite well with the 80 per cent prediction intervals.

Table 8. Number of Norwegian-born children of immigrants in 2022 (registered), and 2030, 2040, 2050, 2060

	2022	2030	2040	2050	2060
Median	206	262	328	381	431
67% prediction interval		[244-283]	[293-364]	[335-439]	[367-504]
80% prediction interval		[238-290]	[284-377]	[322-459]	[350-531]
Medium Variant		263	329	387	437
Low-High interval		[247-282]	[285-376]	[314-469]	[337-565]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations, Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands.

Forecast results for Norwegian-born children of immigrants with one-year age group detail (Figure 12) are not reliable for most ages: ages up to 40 in 2030, and up to 65 in 2060, roughly speaking.

Figure 12. Age distribution of Norwegian-born children of immigrants, 2030 and 2060. The upper and lower curves are 90 per cent and 10 per cent percentiles (upper and lower bounds of the 80 per cent prediction intervals) of the predictive distribution. The middle curves represent median values

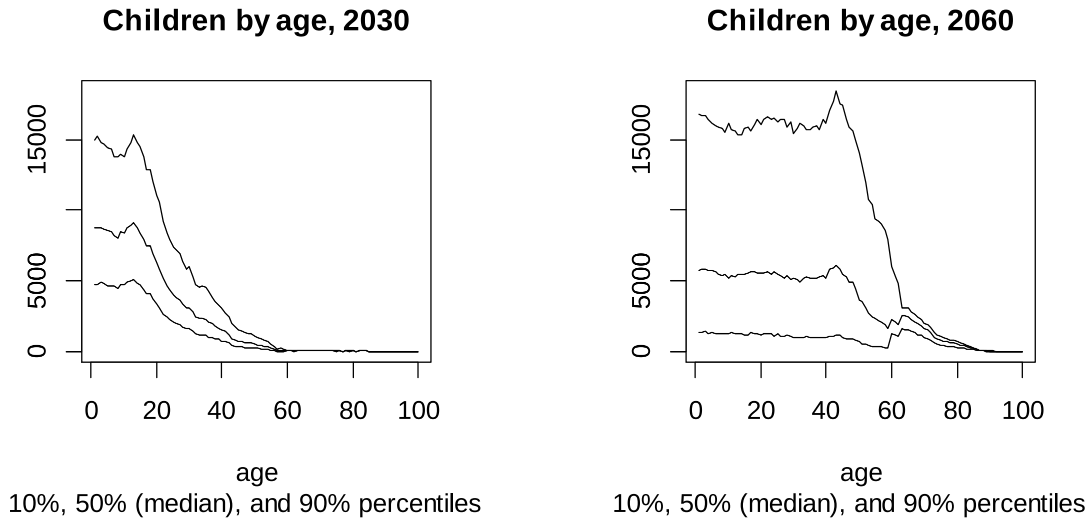


Table 9. Number of Norwegian-born children of immigrants from country group 1 in 2022 (registered), and 2030, 2040, 2050, 2060

	2022	2030	2040	2050	2060
Median	17	20	24	26	29
67% prediction interval		[17-23]	[19-29]	[21-34]	[22-38]
80% prediction interval		[17-24]	[18-31]	[19-36]	[21-40]
Medium Variant		20	24	27	30
Low-High interval		[19-21]	[21-27]	[23-33]	[24-38]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations, Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands.

When the group of Norwegian-born children of immigrants is split up by country background, the increases for children in groups 1 and 2 to 2060 are quite reliable, although the numbers involved are small, and the 80 per cent intervals in 2060 are very wide, relatively speaking – 65.5 and 57.6 per cent of the median for groups 1 and 2, respectively; see Tables 9 and 10.

Table 10. Number of Norwegian-born children of immigrants from country group 2 in 2022 (registered), and 2030, 2040, 2050, 2060

	2022	2030	2040	2050	2060
Median	33	43	54	61	66
67% prediction interval		[36-51]	[44-65]	[48-75]	[53-81]
80% prediction interval		[34-53]	[41-69]	[45-80]	[49-87]
Medium Variant		43	54	62	67
Low-High interval		[40-46]	[46-61]	[48-72]	[49-83]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations, Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands.

Table 11 presents results for children of group 3. These children constitute the large majority of all Norwegian-born children of immigrants. Indeed, Table 11 indicates more than a doubling between 2022 and 2060, at least in terms of the median forecast. Relative uncertainty in 2060 is large: 47.0 per cent.

6.4. Population without immigration background

The population consisting of persons without an immigration background ("other persons") is likely to remain more or less constant in size between 2022 and 2060, from 4.4 million to between 4.0 and 5.0 million, according to the 80 per cent interval in Table 12. The relative width of that interval is 23.7 per cent, slightly more than that of the entire population (21.0 per cent).

Table 11. Number of Norwegian-born children of immigrants from country group 3 in 2022 (registered), and 2030, 2040, 2050, 2060

	2022	2030	2040	2050	2060
Median	156	198	249	292	334
67% prediction interval		[182-216]	[219-280]	[253-343]	[280-399]
80% prediction interval		[177-222]	[211-293]	[241-360]	[265-422]
Medium Variant		199	250	297	339
Low-High interval		[187-214]	[218-288]	[244-364]	[265-444]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations, Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands.

Table 12. Number of persons without migration background in 2022 (registered), and 2030, 2040, 2050, 2060

	2022	2030	2040	2050	2060
Median	4400	4442	4487	4490	4447
67% prediction interval		[4372-4512]	[4335-4649]	[4236-4763]	[4086-4874]
80% prediction interval		[4351-4535]	[4286-4704]	[4159-4865]	[3971-5027]
Medium Variant		4442	4493	4504	4483
Low-High interval		[4360-4512]	[4252-4674]	[4095-4807]	[3890-4938]

Note: Median value, lower and upper bounds of 67 per cent and 80 per cent prediction intervals based on 3 000 simulations, Medium Variant and Low and High Variants of Statistics Norway's projection of 2022. Numbers in thousands.

Observe that the interval between high and low results of this population sub-group in the official projection agrees quite well with the 80 per cent interval of our simulations. On the other hand, we found rather wide intervals for total population size (Section 6.1) and the immigrant population (Table 4) in the official projections. The total population comprises immigrants, children of immigrants, and the remaining population. Children contribute little in this respect. Hence, one may suspect that the reason for this discrepancy is to be found in the intervals for international migration in the official projections. An exact analysis is complicated, but we get some insight when we compare fertility, mortality, and migration parameters in Statistics Norway's High, Medium and Low projection variants; see Table 13.

Table 13 gives key parameter values for the year 2060 of the official projection. The comparison is very indirect, because the variables in Table 13 have very different metrics: children per woman for fertility, years of life for mortality, and number of persons for net migration. But the table suggests a very large distance between the High and the Low Variant relative to the Medium Variant for net migration. The relative distance is 215 per cent, much more than the intervals for the Total Fertility Rate (35 per cent) or the Life expectancy of men and women (4 – 5 per cent).

Table 13. Key projection parameters for the year 2060 in the 2022 population projection of Statistics Norway

	High Variant	Medium Variant	Low Variant	(H – L)/M (%)
Total fertility rate (c/w)	1.90	1.70	1.31	34.7
Life expectancy men (years)	91.0	88.9	86.4	5.2
Life expectancy women (years)	92.7	90.9	88.8	4.3
Net migration (number of persons per year)	26,800	10,800	3,600	214.8

Source: Thomas and Tømmerås (2022, p. 39).

7. Conclusions

No forecasts are exact, and so it is important to provide some measure of the forecast uncertainty. Therefore, forecasters should compute two types of results: first, point forecasts, which are as accurate as possible, and second, the statistical distributions around the point forecasts (Makridakis et al. 2019).

Here we report work that aimed at computing statistical distributions around forecasts of the size and age and sex structure of the migrant population of Norway and their children. As point forecasts we took the results of Statistics Norway's deterministic population projection published in July 2022. We used the method of random shares, which has been applied earlier to household forecasting. The method starts with a stochastic forecast of the future population broken down by age and sex. Each result of the latter forecast is a random variable. This variable is combined with a set of random shares that divide each population number, given age and sex, into numbers for migrant categories. We present results for the years 2030, 2040, 2050, and 2060 for the immigrant population of Norway and their Norwegian-born children ("second generation") broken down by age and sex. We distinguish immigrants and their children grouped by three categories representing country background: 1. West European countries plus the United States, Canada, Australia, and New Zeal-land; 2. East European countries that are members of the European Union; 3. other countries. The population without any migration background forms a seventh population subgroup.

The results show that a few population trends that were predicted for Norway to 2060 are quite certain: strong increases in the size of the immigrant population (more specifically those who belong to group 3) and of Norwegian-born children of immigrants. As to the age structures of immigrants and their children, the prediction intervals around the forecasts of these persons in one-year age groups are so wide that there is little information in these forecasts. For the population as a whole (irrespective of migrant background), forecasts for the age structure in one-year age groups are reliable up to around 2040, except for children born after 2022. For later years, the intervals become very wide for all ages. But ageing is certain.

Meanwhile, one should keep in mind that these results are based upon two important assumptions: our best guess is the trajectory predicted by Statistics Norway in its Main Variant, and the variation in future numbers is similar to the variation as observed in the past twenty years.

References

Alders, M. (1999). Stochastische huishoudensprognose 1998–2050 [Stochastic household forecast 1998–2050]. Maandstatistiek van de Bevolking, 47(11), 25–34.

Alders, M. (2001). Huishoudensprognose 2000–2050: Veronderstellingen over onzekerheidsmarges [Household forecast 2000–2050: Assumptions on uncertainty intervals]. Maandstatistiek van de Bevolking, 49(8), 14–17.

Alders, M. (2005). Allochtonenprognose 2004–2050: Belangrijkste uitkomsten [Projections of the foreign-origin population in the Netherlands 2004–2050]. Bevolkingstrends 53(1), 32–42.

Alho, J. (1998). A stochastic forecast of the population of Finland. Reviews 199/4. Helsinki: Statistics Finland.

Alho J., Alders M., Cruijsen H., Keilman N., Nikander T., Pham D.Q. (2006). New forecast: Population decline postponed in Europe. Statistical Journal of the United Nations Economic Commission for Europe 23(1), 1-10.

Alho, J., Keilman, N. (2010). On future household structure. Journal of the Royal Statistical Society Series A, 173(1), 117–143. doi:10.1111/j.1467-985X.2009.00605.x .

Armstrong, J. S., Green, K. C., Graefe, A. (2015). Golden Rule of Forecasting: Be Conservative. Journal of Business Research, 68 (8), 1717-1731. http://dx.doi.org/10.1016/j.jbusres.2015.03.031

Booth, H. (1984). Transforming the Gompertz for fertility analysis: the development of a standard for the relational Gompertz. Population Studies, 38(3), 495–506.

Brass, W. (1971). On the scale of mortality. In W. Brass (Ed.), Biological aspects of demography (pp. 69–110). London: Taylor & Francis.

Cappelen Å., Skjerpen T., Thomas M. (2022). International migration: Assumptions and results. Pp. 97 – 136 in M. Thomas and A. Tømmerås, Norway's 2022 national population projections. Oslo: Statistics Norway.

Christiansen, S. G., Keilman, N. (2013). Probabilistic household forecasts based on register data—The case of Denmark and Finland. Demographic Research, 28, 1263 – 1302. doi:10.4054/DemRes.2013.28.43.

Coale, A., Trussell, J. (1974). Model fertility schedules: variations in the age structure of childbearing in human populations. Population Index, 40(2), 185–258.

Coleman D., Scherbov S. (2005). Immigration and ethnic change in low-fertility countries – towards a new demographic transition? Population Association of America Annual Meeting, Philadelphia.

De Beer, J. (2011). A new relational method for smoothing and projecting age specific fertility rates: TOPALS. Demographic Research, 24, 409–454.

De Beer, J. (2012). Smoothing and projecting age-specific probabilities of death by TOPALS. Demographic Research, 27, 543–592.

De Beer, J., Alders, M. (1999). Probabilistic population and household forecasts for the Netherlands. Joint Economic Commission for Europe–EUROSTAT Work Session on Demographic Projections, Perugia, 3–7 May (Working paper 45).

Foss, A. (2012). Stokastiske befolkningsprognoser for Norge 2012-2060 [Stochastic population forecasts for Norway 2012-2060]. Økonomiske analyser 2/2012, 30–34.

Fuchs, J., Söhnlein, D., Weber B., Weber E. (2018). Stochastic Forecasting of Labor Supply and Population: An Integrated Model. Population Research and Policy Review 37, 33–58. https://doi.org/10.1007/s11113-017-9451-3.

Greene W.H. (2003). Econometric analysis (5th ed.). New Jersey: Prentice Hall

Jacobs D., Swyngedouw M., Hanquinet L., Vandezande V., Andersson R., Horta A.P.B., Berger M., Diani M., Ferrer A.G., Giugni G., Morariu M., Pilati K., Statham P. (2009). The challenge of measuring immigrant origin and immigration-related ethnicity in Europe. International Migration and Integration 10, 67-88. https://doi.org/10.1007/s12134-009-0091.

Jiang, L., O'Neill, B. (2004). Toward a new model for probabilistic household forecasts. International Statistical Review, 72(1), 51–64.

Keilman, N. (2016). A combined Brass-Random Walk approach to probabilistic household forecasting: Denmark, Finland, and the Netherlands 2011-2041. Journal of Population Research, 33(2), 1–27.

Keilman, N. (2020). A probabilistic forecast for the population of Norway. Pp. 177 – 182 in A. Syse, M. Thomas, R. Gleditsch Norway's 2020 population projections National level results, methods and assumptions. Oslo: Statistics Norway.

Keilman N., 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 – 454.

Lee, R., Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium and low. Journal of the American Statistical Association 89, 1175–1189.

Leknes S., Løkken S.A. (2022). Befolkningsframskrivinger for kommunene 2022 [Population projections for municipalities 2022]. Report 2022/30. Oslo, Statistics Norway.

Luo L., Hodges, J.S. (2020). The age-period-cohort-interaction model for describing and investigating inter-cohort deviations and intra-cohort life-course dynamics. Sociological Methods and Research 51(3), 1164–1210.

Makridakis S., Hyndman R.J., Petropoulos, F. (2019). Forecasting in social settings: The state of the art. International Journal of Forecasting 36(1), 15 – 28, https://doi.org/10.1016/j.ijforecast.2019.05.011.

Preston, S., Heuveline, P., Guillot, M. (2001). Demography: Measuring and modelling population processes. Oxford: Blackwell.

Scherbov, S., Ediev, D. (2007). Probabilistic household projections based on an extension of headship rates method with application to the case of Russia. Joint Economic Commission for Europe–EUROSTAT Work Session on Demographic Projections, Bucharest, 10–12 October (Working paper 16).

Statistics Norway 2022. Statbank Table nr 06913: Population and population changes, by contents and year.

Texmon, I. (1992). Norske befolkningsframskrivinger 1969 – 1990 [Population projections for Norway 1969 – 1990]. In: O. Ljones, B. Moen, and L. Østby (eds.) Mennesker og modeller. Sosiale og Økonomiske Studier no. 78. Oslo: Statistics Norway.

Thomas, M.J., Tømmerås, A.M. (2022). Norway's 2022 national population projections: Results, methods and assumptions. Report 2022/28. Oslo: Statistics Norway.

United Nations. (1973). Methods of projecting households and families. New York: United Nations.

United Nations (2022). World Population Prospects 2022: Summary of Results. UN Department of Economic and Social Affairs, Population Division DESA/POP/2022/TR/NO. 3.

Vanella, P., Heß, M., Wilke, C.B. (2020). A probabilistic projection of beneficiaries of long-term care insurance in Germany by severity of disability. Quality and Quantity 54, 943–974. https://doi.org/10.1007/s11135-020-00968-w.

Vassenden, K. (2015). Om kvaliteten på den norske inn- og utvandringsstatistikken [On the quality of the Norwegian immigration and emigration statistics]. Notater 2015/17. Oslo: Statistics Norway.

Wilson, T. (2013a). Quantifying the uncertainty of regional demographic forecasts. Applied Geography, 42 (August), 108–115.

Wilson, T. (2013b). The sequential propensity household projection model, Demographic Research, 28(24), 681–712. http://www.demographic-research.org/volumes/vol28/24/28-24.pdf.

Zeng, Y., Wang, Z., Ma, Z., Chen, C. (2000). A simple method for estimating α and β: An extension of Brass Relational Gompertz Fertility model. Population Research and Policy Review, 19(6), 525–549.