Bringing Age Back In: Accounting for Population Age Distribution in Forecasting Migration

Nathan G. Welch
MITRE Corporation

Hana Ševčíková
University of Washington

Adrian E. Raftery
University of Washington

October 17, 2025

Abstract

Existing models of country-level net migration ignore the effect of population age distribution on past and projected migration rates. We propose a method to estimate and forecast international net migration rates for the 200 most populous countries, taking account of changes in population age structure. We use age-standardized estimates of country-level net migration rates and in-migration rates over quinquennial periods from 1990 through 2020 to decompose past net migration rates into in-migration rates and out-migration rates. We then recalculate historic migration rates on a scale that removes the influence of the population age distribution. This is done by scaling past and projected migration rates in terms of a reference population and period using a quantity we call the *migration age structure index* (MASI). We use a Bayesian hierarchical model to generate joint probabilistic forecasts of net migration rates over 5-year periods for all countries through 2100. We find that accounting for population age structure in historic and forecast net migration rates leads to narrower prediction intervals by the end of the century for most countries. Furthermore, accounting for population age structure leads to less out-migration among countries with rapidly aging populations that are forecast to contract most rapidly by the end of the century. This leads to less drastic population declines than are forecast without accounting for population age structure.

keywords: *migration, probabilistic, population, projection, forecast*

1 Introduction

Preparations for the next great demographic era are underway around the world (Kim, 2019; Director of National Intelligence, 2021). Decades of declining and then sub-replacement fertility created a demographic dividend, where the share of the working-age population far outweighed the share of children and older adults no longer participating in the workforce (Bongaarts, 2009). Now the average ages of populations around the world are increasing at the fastest rate in history (Bloom & Zucker, 2023). International migration may blunt the impact of this dynamic, but migration alone will not fully mitigate the realities of the next demographic era (RAND, 2005; Coleman, 2008). Accounting for past and forecast population age structure reveals that migration may be an even less effective mitigation than currently understood (Münz, 2013; Lee, 2011). However, existing forecasting methods fail to account for this eventuality, leaving the role of population age structure in migration dynamics to imprecise qualitative conjecture.

We propose a probabilistic net migration forecasting approach that accounts for the impacts that population age structure had on historic migration rate estimates. We use these to generate new probabilistic population forecasts for all countries. The resulting forecasts quantify how different the global population distribution could be after accounting for past and projected shifts in the population age structure.

The article is organized as follows. Section 2 reviews the existing work that is foundational to our study. Section 3 summarizes the data and methods used to generate age standardized migration forecasts. This is followed by an evaluation of the age-standardized migration method in Section 4. Then Section 5 summarizes differences between population forecasts with and without accounting for population age structure in the migration component of the forecasts. We conclude in Section 6 with a discussion of the contributions and limitations of our approach.

2 Background

Rogers & Castro (1981) identified a consistent pattern in the age profile of out-migrants and proposed a model migration age schedule. Even though modern migration estimation and forecasting methods use migration age schedules to account for the impact of migration on the sending and receiving populations, population age structure is rarely an explicit consideration in estimation or forecasting. This is reasonable over short-term forecast horizons given the relatively slow pace of population age structure changes, but less so over longer horizons.

Kupiszewski (2002) discussed differences in migration forecasts when the sending population age-structure is ignored. Specifically, it was argued that ignoring the limited number of migration-age individuals in Poland and future population aging contributed to unrealistically large out-migration forecasts from Poland to other European Union (EU) countries before accession to the EU in 2004.

Fertig & Schmidt (2005) argued that neglecting population age in calculating migration rates obscures the underlying population dynamics. They proposed an adjusted net migration rate that is calculated as the number of migrants aged 0-39 divided by the sending country population aged 0-39. This approach reduces distortions created by the population of older people who tend to make up a relatively small proportion of migrants, but effectively sets older age migration to zero and ignores the age structure of the critical 0-39 age group.

Kolk (2019) reviewed age-specific migration and the analogy to age-specific and total fertility rate definitions. It was argued that population-level measures should be adopted for both subnational and international migration estimation. The authors outlined key challenges to extending such methods to an international context, specifically properly accounting for the population at risk of in-migration and differences in the definition of an international migrant from country to country.

Skjerpen & Tønnessen (2021) incorporate the age profiles of people from three groups of origin countries to project Norwegian migration inflows through 2100. They find that incorporating origin population age structure in their models equates to a reduction of 0.1 children per woman in the total fertility rate by 2100, compared to official Norwegian assumptions. The authors also highlight that existing migration forecasting methods overlook the relative stability of age structure forecasts, which should be leveraged to improve migration forecasts.

Fully probabilistic population forecasting is an active area of demographic research (e.g. Hyndman & Booth, 2008; Raftery et al., 2012, 2014; Wiśniowski et al., 2015; Yu et al., 2023). Azose et al. (2016) developed a probabilistic net migration model and used it to produce probabilistic population forecasts for all countries through 2100. Welch & Raftery (2022) proposed a fully probabilistic population forecasting method that uses a probabilistic model of bilateral migration flows for all countries through 2045. However, none of these population forecasting methods systematically account for population age structure changes in both the historic data used to fit these models and in forecasting.

Raftery & Ševčíková (2023) proposed a method for accounting for population age structure in probabilistic migration and population forecasts, and applied it to very long-term population forecasts, to 2300, motivated by the problem of estimating the social cost of carbon. Like us, they used net migration data, which has advantages of data availability and analytical simplicity over methods based on more complete migration data, such as between-country flows or in- and out-migration flows for countries. However, a problem with this is that it does not take account of the different age distributions of in- and out-migrants (Rogers, 1990). They addressed this issue by assuming that during periods of net in-migration, all movement was classified as in-migration with no out-migration, and similarly, net out-migration periods were treated as solely out-migration. Here we develop a new method that avoids this unrealistic assumption while still using net migration data.

3 Data & Methods

Data

The United Nations defines a long-term migrant as a person who moved to a different country and stayed for at least one year (McAuliffe & Oucho, 2024). Recent methodological innovations make it possible to estimate globally consistent, country-level total migration inflows and outflows from Census counts of the number of people by country of birth in all countries (Abel, 2013; Azose & Raftery, 2019). Abel & Cohen (2019) found that the pseudo-Bayes method of Azose & Raftery (2019) to estimate bilateral migration flows led to the most accurate migration estimates among several extant methods. Unfortunately, pseudo-Bayes migrant flow estimates over 5-year periods are available for only six 5-year periods starting in 1990 and ending in 2020 (Abel & Cohen, 2019). Estimating and forecasting bilateral migration flows has a number of advantages (Welch & Raftery, 2022), but population age structure changes too slowly to solely rely on these estimates

for long-term migration forecasting.

The United Nation's World Population Prospects (WPP) publishes globally consistent net migration estimates, defined as the difference between migration inflows and outflows divided by population size, back to the 1950-1955 period (United Nations, 2022). The pseudo-Bayes flow estimates contain more detailed information in the form of bilateral flows. We use strengths of both data sets to develop a new combined data series.

Migration flow estimates are calculated with respect to specific revisions of the WPP. We use WPP 2019 revision (United Nations, 2019) and the associated flow estimates for all analyses. The 2022 revision transitioned from 5-year period to 1-year period estimates and forecasts (United Nations, 2022). The 2024 revision was the first to use a fully probabilistic migration forecasting model (United Nations, 2024). We use the 2019 revision to show the differences in long-term forecasts using the last available deterministic migration projections for 5-year periods in the WPP 2019 compared to a probabilistic migration model that ignores population age structure and one that explicitly accounts for past and future population age structure changes.

We fit a model, described in Section 3, to the 1990–2020 period from which both in-migration and net migration rates can be estimated for all countries with populations of 100,000 or more. This model is then used to approximate the contribution that in-migration made to the net migration rate prior to 1990. Our model-based approach is an efficient solution that allows for the estimation of globally consistent age-standardized in-migration and out-migration rates for 5-year periods starting in 1950 and running through 2020. The successful decomposition of net migration rates into in-migration and out-migration rates is critical to calculating migration rates adjusted for the population age structure that gave rise to those rates.

Age Standardization of Net Migration Rates

For reasons of data availability across all countries, and analytic simplicity, the United Nations (UN) uses all-age net migration as part of the basis for its population projections for all countries (United Nations, 2019). The WPP 2024 (United Nations, 2024) replaced a deterministic projection net migration method with the probabilistic approach of Azose & Raftery (2015). However, neither the UN's deterministic approach nor the probabilistic approach of Azose & Raftery (2015) takes account of historic or future changes in population age structure.

We propose to modify the approach of Azose & Raftery (2015) to account for population age structure. We do this by creating an age-standardized version of the net migration rate. This poses a challenge, as historical net migration data are not disaggregated by age, making it difficult to standardize net migration rates directly. This was one of the arguments of Rogers (1990) for not using net migration at all.

We propose a model-based approach to help address these concerns. We start by developing a method for decomposing historic net migration into in- and out-migration. This uses a subset of the data for which estimates of both in- and out-migration are available (i.e. 1990–2020), and uses this to estimate in- and out-migration for the entire data period (i.e. 1950–2020). We then develop a method for age-standardizing out-migration. This turns out to be very simple, relying on a quantity we call the *migration age structure index* (MASI). We extend this to in-migration, and hence to net migration. Finally we apply the probabilistic forecasting method of Azose & Raftery (2015) to the age-standardized net migration rates, and input the resulting forecasts to the overall

probabilistic population projection method.

Table 1 summaries the notation required for our methods. By convention, we refer to each period using the first year (inclusive) and the last year (exclusive). For example, the period 2000—2005 represents the interval from January 1, 2000, to December 31, 2004, i.e., [2000, 2005). Also, a "+" subscript in place of an index value denotes the marginal sum over that index.

Table 1: Notation for migration age-standardized migration method.

$O_{i,t}$	Integer-valued outflow from origin i in period starting with year t
$O_{i,t,a}$	Integer-valued outflow from origin i in period starting with year t of age group a–$(a+4)$
$I_{i,t}$	Integer-valued inflow to destination i in period starting with year t
$N_{i,t}$	Integer-valued net migrant flow in country i in period starting with year t
$P_{i,t,a,s}$	Integer-valued population of origin i at the start of period t for age group a-$(a+4)$ and sex $s \in \{\text{male}, \text{female}\}$
$P_{i,t,+,+}$	Integer-valued population of origin i at the start of period t
$\tilde{P}_{i,t,+,+}$	Integer-valued population of origin i at risk of migration over period t to $t+5$, defined as $P_{i,t+5,+,+} - N_{i,t}$
$\text{OMR}_{i,t}$	$= O_{i,t}/\tilde{P}_{i,t,+,+}$. Out-migration rate for country i in the period starting in year t
$\text{OMR}_{i,t,a}$	$= O_{i,t,a}/\tilde{P}_{i,t,a,+}$. Out-migration rate for country i, period starting in year t and age group a
$G_{i,t}$	$= \sum_a \text{OMR}_{i,t,a}$. Gross Migraproduction Rate (GMR) for country i and period starting in year t
$\text{IMR}_{i,t}$	$= I_{i,t}/\tilde{P}_{i,t,+,+}$. In-migration rate for period starting in year t and country i
$\text{NMR}_{i,t}$	$= \text{IMR}_{i,t} - \text{OMR}_{i,t}$. Net migration rate for country i and period t
$\pi_{i,t,a}$	$= \tilde{P}_{i,t,a,+} / \sum_a \tilde{P}_{i,t,a,+}$. Share of population from age group a in period t
$\pi^*_{i,t,a}$	Reference population age distribution for country i in period t
R_a	Reference migration age schedule normalized so that $\sum_a R_a = 1$
$C_{i,t}$	$= \sum_a R_a \pi_{i,t,a}$. The migration age structure index (MASI), a scalar accounting for the population and migrant age structure in the migration rate at time t for country i
$\check{C}_t$	$= \sum_{i,a} R_a \pi_{i,t,a}$. MASI for the world at time t
$\text{OMR}^\star_{i,t}$	$= \text{OMR}_{i,t} \times C_{i,\text{baseline}}/C_{i,t}$. Age-standardized out-migration rate
$\text{IMR}^\star_{i,t}$	$= \text{IMR}_{i,t} \times \check{C}_{\text{baseline}}/\check{C}_t$. Age-standardized in-migration rate
$\text{NMR}^\star_{i,t}$	$= \text{IMR}^\star_{i,t} - \text{OMR}^\star_{i,t}$. Age-standardized net migration rate

Decomposing Historic Net Migration into In- and Out-migration

Net migration age-standardization takes place on the inflow and outflow components of the net migration rate. We use a mixed-effects model to relate the 1990–2020 in-migration rate, $\text{IMR}_{i,t}$, to the net migration rate, $\text{NMR}_{i,t}$, for each country i over these six 5-year periods, t, from 1990 through 2020. Our mixed-effects model is defined as follows:

$$\begin{aligned} \text{IMR}_{i,t} &= \beta_{0,i} + \beta_1 \max\left(\text{NMR}_{i,t}, 0\right) + \varepsilon_{i,t}, \\ \beta_{0,i} &\sim \text{Normal}\left(\beta_0, \sigma^2_{between}\right), \\ \varepsilon_{i,t} &\sim \text{Normal}\left(0, \sigma^2_{within}\right). \end{aligned} \tag{1}$$

This model uses a random intercept term to account for the average country-specific in-migration rate associated with the net migration rate for each period from 1990–2020.

Country-specific intercepts, $\beta_{0,i}$, are drawn from a distribution with global mean, β_0. The difference in a country's random intercept from the global mean, β_0, is determined by the average difference from the global mean for that country in all periods. Out-migration rates are implicitly defined by this model.

The random intercept model was fit using the linear mixed-effects model implementation in the *lme4* R package implementation (Bates et al., 2015). Starting with $\text{NMR}_{i,t}$, the estimated mean in-migration rate is $\text{IMR}_{i,t} = \beta_{0,i} + \beta_1 \max(\text{NMR}_{i,t}, 0)$. The net migration rate and estimated mean in-migration rate are then used to calculate the out-migration rate, $\text{OMR}_{i,t} = \text{IMR}_{i,t} - \text{NMR}_{i,t}$.

Figure 1 summarizes the 1990–2020 data, model fit, and association between estimated and actual country-level migration flow rates. Points in Figure 1(a) show the observed $\max(\text{NMR}_{i,t}, 0)$ on the horizontal axis and $\text{IMR}_{i,t}$ on the vertical axis. The discontinuity at $\text{NMR}_{i,t} = 0$ reflects the fact that the in-migration rate must be non-negative. In the $\text{NMR}_{i,t} \leq 0$ portion of the domain, the country-specific intercept, $\beta_{0,i}$, of model 1 establishes the minimum average magnitude of in-migration corresponding to a negative net migration rate for that country. When $\text{NMR}_{i,t} > 0$, the model in-migration rate increases from this minimum average in-migration rate by an average of β_1 per unit increase in $\text{NMR}_{i,t}$. The global mean model in-migration rate is shown by the blue line in Figure 1(a).

Many observations in Figure 1(a) are highly concentrated around some values. Locally Estimated Scatterplot Smoothing (LOESS) is a non-parametric regression method used to reveal underlying scatterplot trends without assuming a specific global form for the relationship (Cleveland, 1979). In this context, agreement between the LOESS line and the mean of model (1) indicates that the model recovers the general trend of the observed data. In the $\text{NMR}_{i,t} < 0$ portion of the domain, the similarity between the LOESS line and the $\beta_{0,i}$ values shows that the mean intercept term provides an accurate summary of the observed in-migration and net migration rates. Near $\text{NMR}_{i,t} = 0$, however, the disagreement between the model mean and LOESS line shows that many $\text{IMR}_{i,t}$ observations near $\text{NMR}_{i,t} = 0$ could be overstated with the model mean alone. The largest LOESS line departure occurs for large positive values of $\text{NMR}_{i,t}$. This is primarily due to the influence of large positive $\text{NMR}_{i,t}$ observations for a few countries.

Finally, Figure 1(a) shows two examples of country-specific average association between $\text{IMR}_{i,t}$ and $\text{NMR}_{i,t}$. The top line shows the model fit for Qatar (green), which experienced some of the highest migration rates in the world since 1990. The bottom line shows the model fit for China (red), a country with stable net out-migration over the same period. These two examples underscore the importance of a model capable of accommodating the significant variation observed between countries.

Figure 1(b) shows the fitted and observed $\text{IMR}_{i,t}$ for all countries from 1990 to 2020. The high level of agreement between fitted and observed in-migration rates with an $R^2 = 0.98$ is notable considering the simplicity of model (1). The agreement between fitted and observed out-migration rates in Figure 1(c) is lower with $R^2 = 0.91$, but is still strong. Taken together, Figures 1(b)-(c) show that our model-based decomposition of the net migration rate is reasonable for the periods where net migration, in-migration, and out-migration rates are all available.

Figure 2 shows the observed net migration rate for 5-year periods starting in 1950 and running through 2020 on the original scales for the United States, Germany, Turkey, and Saudi Arabia. These countries were selected to demonstrate how our model-based approach to estimating in- and out-migration rates compares to direct estimates of these

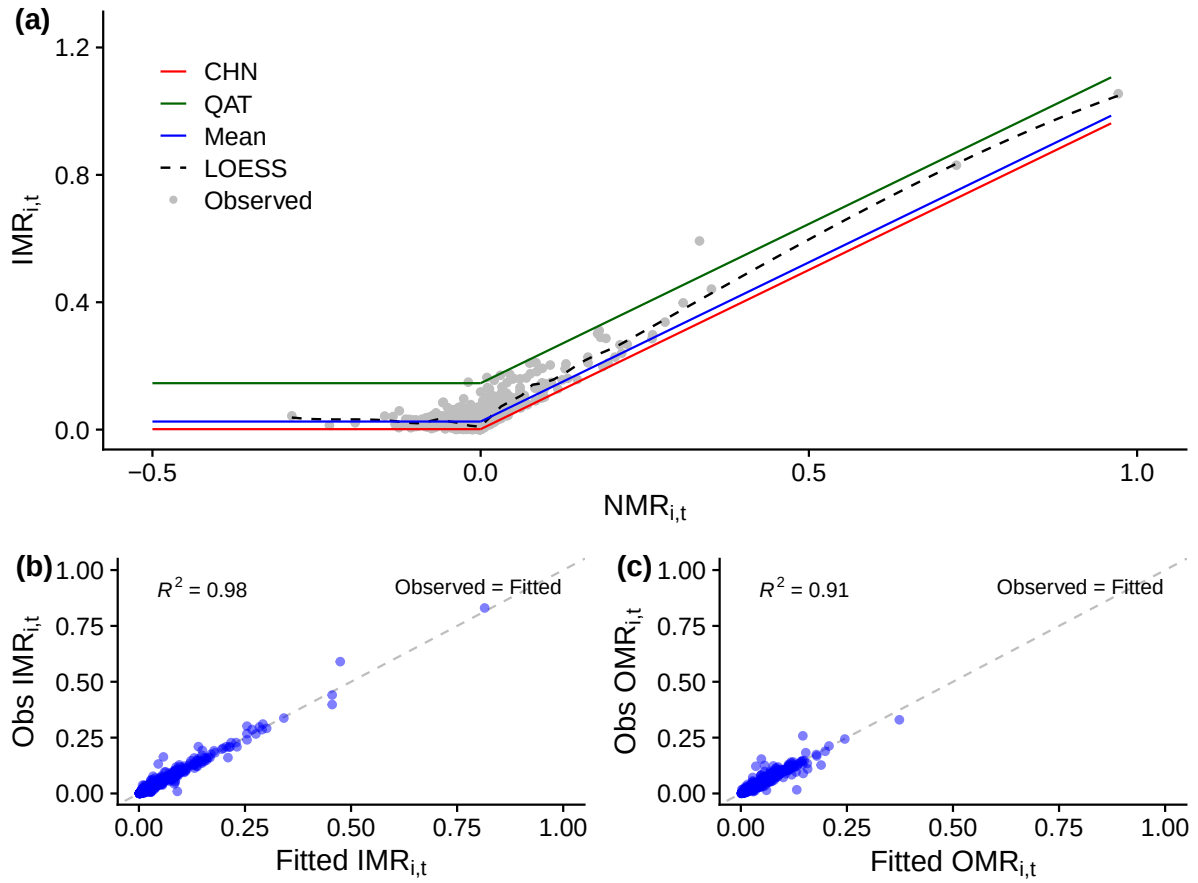

Figure 1: (a) Observed in-migration rates versus net migration rates for all countries (points) with mean model (blue), LOESS line (black dashed), and country-specific models for China (CHN, red) and Qatar (QAT, green); (b) Observed in-migration rates versus fitted in-migration rates compared to IMR Observed=Fitted (dashed line); (c) Observed out-migration rates versus fitted out-migration rates compared to OMR Observed=Fitted (dashed line) for 5-year-periods from 1990–2020.

quantities when the net migration rate was large and positive for a sustained period (United States), was more dynamic (Germany), experienced an abrupt departure from a stable historic norm (Turkey), or was generated in a Gulf Cooperation Council (GCC) member country (Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, United Arab Emirates). The in-migration and out-migration rates from 1990 to 2020 used to estimate our model are shown in the second and third columns (black dashed lines) along with the model-based in- and out-migration rate estimates for all 5-year periods (solid blue lines). Model-based estimates relating net migration rates to in- and out-migration rates are similar to the observed rates for 1990–2020. Estimated in-migration and outflow rates prior to the 1990–1995 period appear plausible for the periods where pseudo-Bayes estimates are available.

While the model-based decomposition is not perfect, Figure 2 shows that the mixed-effects net rate decomposition approach leads to plausible estimates of in-migration and out-migration rates. The correlation between the observed and predicted rates was over 0.97. Broad agreement between observed in-migration and out-migration rates for 1990–2020 periods suggests that net migration rates prior to 1990 can be similarly decomposed

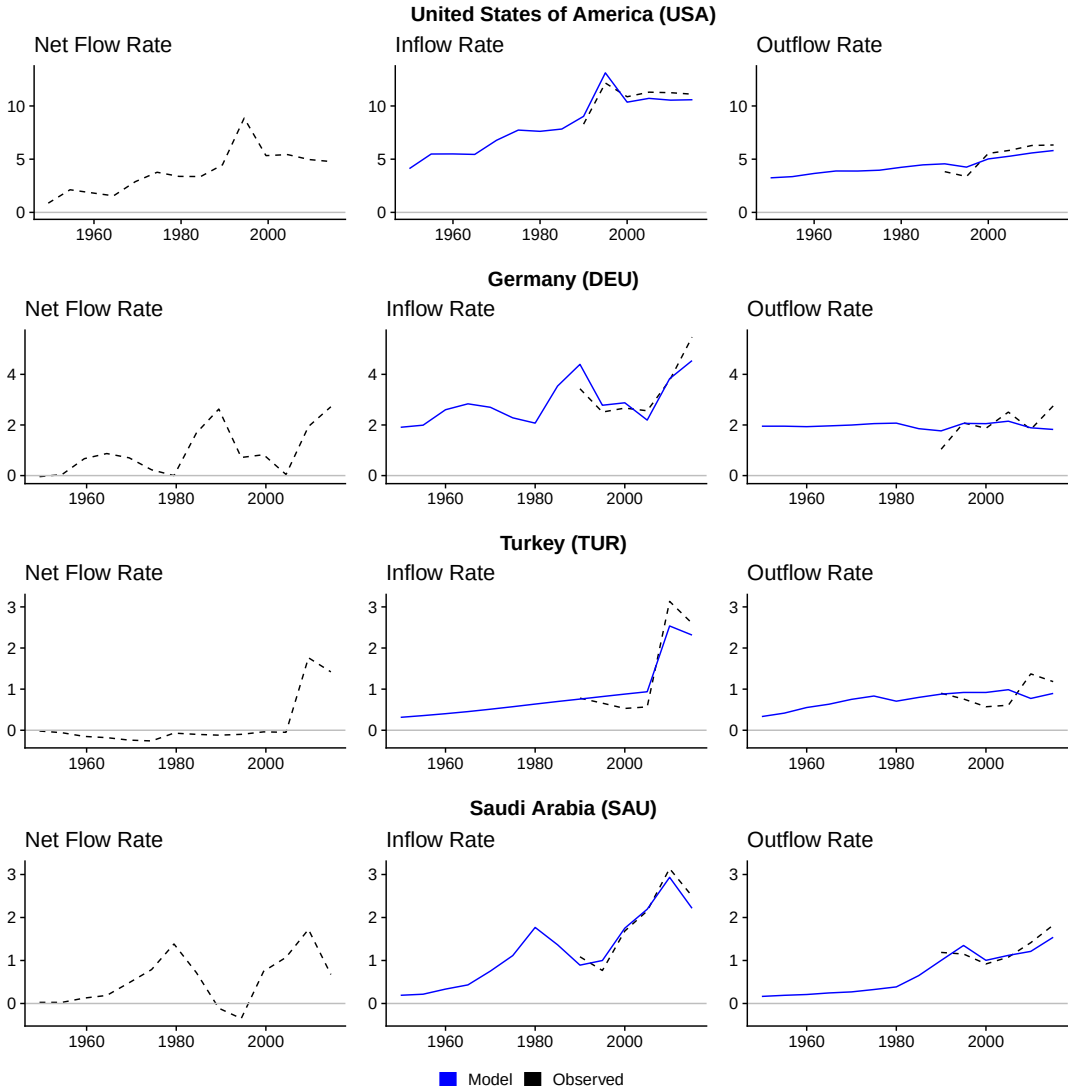

Figure 2: Observed net migration (left column), decomposed into in-migration (middle column), and out-migration rates (right column), on the scale of annual migrants per thousand people, compared to mixed-effects model estimates for the United States, Germany, Turkey, and Saudi Arabia. Solid blue lines show the model-based estimates. Dashed black lines show the observed migration rates used for the estimation. Migration flow estimates are shown at the midpoint of each 5-year period.

into in-migration and out-migration rates.

Raymer et al. (2023) propose an alternative approach to decomposing net migration into in-migration and out-migration. This method uses a fixed constant to approximate the magnitude of net migration attributed to total in-migration and out-migration in proportion to a country's population. This method could also be used to decompose the net flow into total inflow and outflow in place of model (1). However, model (1) offers a systematic approach to specifying a quantity similar to the fixed constant used in Raymer et al. (2023), but one that is country-specific, which our analyses suggest is needed.

Age Standardization of In- and Out-migration Rates

The net migration rate, $\text{NMR}_{i,t}$, for country i over period t–t+5 is defined as the difference between the in-migration rate, $\text{IMR}_{i,t}$, and out-migration rate, $\text{OMR}_{i,t}$:

$$\text{NMR}_{i,t} = \text{IMR}_{i,t} - \text{OMR}_{i,t} = \frac{N_{i,t}}{\tilde{P}_{i,t,+,+}} = \frac{I_{i,t} - O_{i,t}}{\tilde{P}_{i,t,+,+}}. \tag{2}$$

Here, $I_{i,t}$ denotes the total inflow count, $O_{i,t}$ the total outflow count, and $N_{i,t} = I_{i,t} - O_{i,t}$ the total net flow count over the period starting in year t. We define the denominator of the net migration rate, $\tilde{P}_{i,t,+,+} = P_{i,t+5,+,+} - N_{i,t}$, as the population of country i at the end of the period starting in year t before factoring in the change due to net migration. This denominator specifies the population at risk of out-migration over the period.

The definition of the net migration rate shown in equation (2) departs from more conventional specifications, e.g., Preston et al. (2001) definition stated in terms of person-years. We use this net migration rate specification to estimate model parameters on the same scale as the *bayesPop* probabilistic population forecasting implementation (Ševčíková & Raftery, 2016).

Historic net migration rate estimates are not disaggregated by age for most countries, but we aim to standardize migration rates to remove the effects of population age structure differences among countries in the same period and within countries across periods. Migration age patterns are known to be relatively consistent over time and country of origin (Rogers & Castro, 1981), but, as Rogers (1990) discussed, the rate defined in equation (2) obscures the influence of the sending and receiving country population age structures. A population age-standardized net migration rate should account for the age structure of the sending populations in both components of the net rate as the migrant age distribution and totals are primarily linked to the origin population age structure.

Let $\pi_{i,t,a}$ denote the population in age group a as a proportion of the total population in country i at time t, namely

$$\pi_{i,t,a} = \frac{\tilde{P}_{i,t,a,+}}{\sum_a \tilde{P}_{i,t,a,+}}\,. \tag{3}$$

Since the origin population of inflows to country i consists of every country other than i in period t, we approximate the proportion of the global population in age group a by

$$\breve{\pi}_{t,a} \approx \frac{\tilde{P}_{+,t,a,+}}{\tilde{P}_{+,t,+,+}} \quad \text{where} \quad \tilde{P}_{+,t,a,+} = \sum_{i,s} \tilde{P}_{i,t,a,s} \quad \text{and} \quad \tilde{P}_{+,t,+,+} = \sum_{i,s,a} \tilde{P}_{i,t,a,s}\,. \tag{4}$$

Rogers & Castro (1981) introduced the concept of the *Gross Migraproduction Rate* (GMR) for country i and period t, defined as $G_{i,t} = \sum_a \text{OMR}_{i,t,a}$, where $\text{OMR}_{i,t,a}$ represents the out-migration rate for country i over period t and age group a. This is a measure of overall migration that is not affected by population age distribution, similar to how the Total Fertility Rate (TFR) is a measure of overall fertility that is independent of the age distribution of women.

The values of $\text{OMR}_{i,t,a}/G_{i,t}$, which give the age pattern of age-specific out-migration rates, tend to be stable over time and place, reflecting a tendency for international migration to be largely concentrated among people aged 15–35 and their dependent children, peaking in the twenties, as pointed out by Rogers & Castro (1981). They also proposed

a parametric model for this pattern, the famous Rogers-Castro curve. As an approximation, we thus consider the situation where this ratio is constant over time and space, so that $\mathrm{OMR}_{i,t,a}/G_{i,t} = R_a$ for all i, t, where $\sum_a R_a = 1$. Under this assumption, the quantity R_a remains constant across all countries and time periods, but it could be modeled using a Rogers-Castro curve or estimated empirically.

We then have an exact result for the out-migration rate for a reference population with a given age distribution:

Theorem 1. *Consider a population that has the same age-specific out-migration rates as country i in period t, but a different population age distribution given by $\pi^*_{i,t,a}$. Then this population has out-migration rate*

$$OMR^*_{i,t} = OMR_{i,t} \frac{\sum_a \pi^*_{i,t,a} OMR_{i,t,a}}{\sum_a \pi_{i,t,a} OMR_{i,t,a}}. \tag{5}$$

If both populations have the same age-specific pattern of migration rates, R_a, then

$$OMR^*_{i,t} = OMR_{i,t} \frac{C^*_{i,t}}{C_{i,t}}, \tag{6}$$

where $C_{i,t} = \sum_a \pi_{i,t,a} R_a$ is the migration age structure index (MASI).

This indicates that one can age-standardize the out-migration rate to a given reference population using equation (6). This is a remarkably simple method, since it only involves multiplication by the MASI, $C_{i,t}$. The MASI involves only the population age distribution and the migration age-pattern, R_a, but not the age-specific migration rates themselves, which cancel. This is an important feature since past age-specific migration rates are not available for most countries.

We can age-standardize in-migration by viewing it as out-migration to country i from the rest of the world. We approximate the age distribution of the rest of the world by the age distribution of the world as a whole. This yields an age-standardized in-migration rate, $\mathrm{IMR}^*_{i,t}$. Finally, we obtain the age-standardized net migration rate as $\mathrm{NMR}^*_{i,t} = \mathrm{IMR}^*_{i,t} - \mathrm{OMR}^*_{i,t}$.

Typically the reference population age distribution will be the distribution in a particular year. In this article, we standardize to the age pattern of 2020, since this is the most recent census year for many countries. To calculate the age-standardized in- and out-migration rate for period t in terms of the 2020 population age structure, remove the population age structure effects in period t, and scale the migration rate in terms of the 2020 reference population:

$$\begin{aligned} \mathrm{IMR}^\star_{i,t} &= \mathrm{IMR}_{i,t} \frac{\check{C}_{2020}}{\check{C}_t} \\ \mathrm{OMR}^\star_{i,t} &= \mathrm{OMR}_{i,t} \frac{C_{i,2020}}{C_{i,t}} \end{aligned} \tag{7}$$

where $\check{C}_t$ denotes the MASI for the world at time t.

This implies that the age-standardized net migration rate is given by $\mathrm{NMR}^\star_{i,t} = \mathrm{IMR}^\star_{i,t} - \mathrm{OMR}^\star_{i,t}$. This measure places historic and forecast net migration rates in the context of the 2020 population age structure, removing variation in net migration rates attributable to differences in population age structure across periods. Additionally, this

specification demonstrates how to convert between age-standardized rates and historic or future rates relative to the reference population.

After estimating historic in- and out-migration rates calculated from model (1), age-standardized in-migration, out-migration, and net migration rates can be calculated for forecasting. To estimate each model, we use total migration across both sexes, accounting for differences in migration propensities between males and females in the forecasting procedure, following the approach outlined in Azose et al. (2016). While there is inherent uncertainty in past estimates of both net migration and bilateral flows, we do not address data quality differences across countries or unmeasured uncertainty in historical migration estimates.

Figure 3 summarizes the key steps to calculating the age-standardized net migration rate for one country. The top left portion of the diagram shows the mixed effects model (1) estimation using periods where inflow, outflow, and net flow estimates are all available (dashed black lines). Model-based estimates of historic in-migration and out-migration rates are then generated from this model (solid blue lines). The top right portion of the diagram shows how the MASI is constructed from a Rogers-Castro-like migration age schedule and the population age distribution at the start of each period. The bottom portion of the figure shows how age-standardized migration rates are calculated for base year 2020.

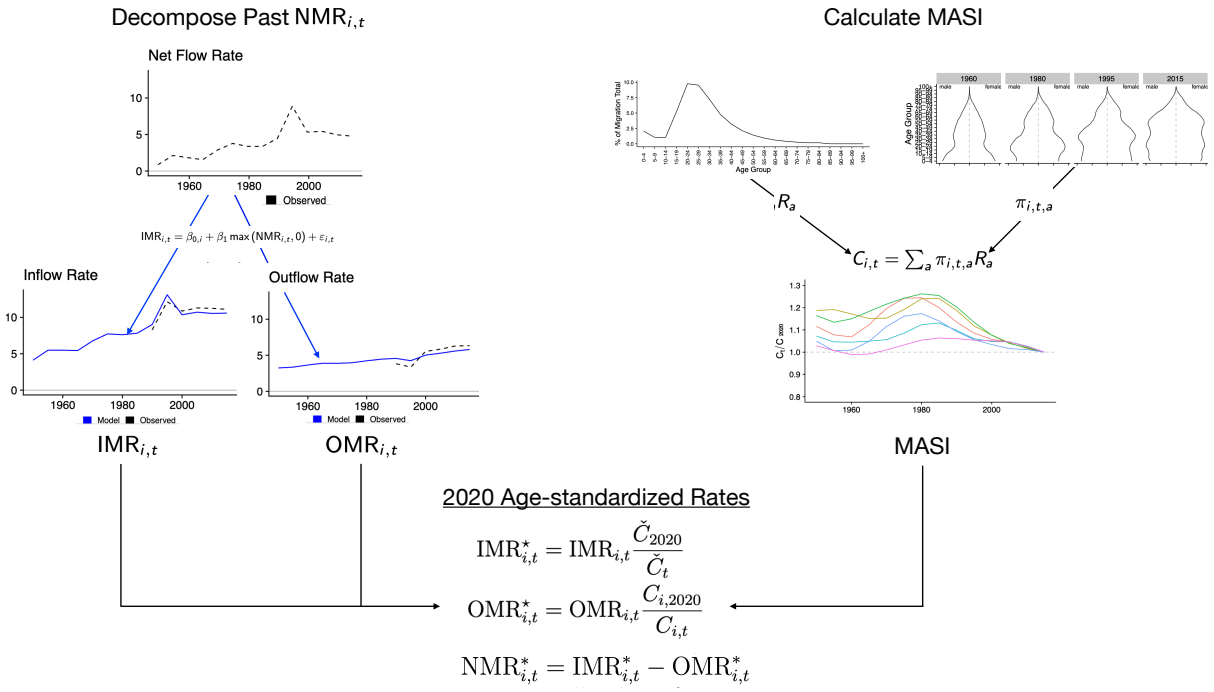

Figure 3: Age-standardized net migration rate process diagram showing the net migration rate decomposition method (top left), Migration Age Standardization Index (MASI) calculation using a model migration age schedule and population age distribution over time (top right) and 2020 base year age-standardized rate calculations (bottom). Migration flow estimates are shown at the midpoint of each 5-year period.

Net Migration Rate Model

We use the resulting rates to develop a probabilistic model for age-adjusted net migration rates. Using the above estimates of in-migration and out-migration rates, we compute

the age-standardized net migration rates, $\text{NMR}^{\star}_{i,t}$, for all countries i and for t from 1950 through 2020, using the 2020 population age structures as the baseline population, as shown in Section 3. While any baseline year is valid, the most recent population estimate is used as the baseline. We then fit the Bayesian hierarchical model of Azose & Raftery (2015). The top level of this Bayesian hierarchical model for the age-standardized net migration rate for country i in time period t is as follows.

$$\textbf{Level 1} \begin{cases} \left(\text{NMR}^{\star}_{i,t} - \mu_i\right) = \phi_i \left(\text{NMR}^{\star}_{i,t-1} - \mu_i\right) + \xi_{i,t} \\ \xi_{i,t} \overset{\text{ind}}{\sim} \text{Normal}\left(0, \psi_i^2\right) \end{cases} \tag{8}$$

Table 2 defines each term in Level 1 of the net migration rate model.

Table 2: Notation for Level 1 of Azose & Raftery (2015) net migration rate model.

$\text{NMR}^{\star}_{i,t}$	Age-standardized net migration rate for country i for period starting in year t
μ_i	Theoretical long-term average migration rate for country i
ϕ_i	Autoregressive parameter for age-standardized net migration rate for country i
$\xi_{i,t}$	Normally distributed random deviation term
ψ_i^2	Variance of random deviations for country i

Hyperparameter values for this model required no adjustment since the default specifications were broadly defined and our specification of the net rate is on the same scale as that of Azose & Raftery (2015). The full model is estimated using a Markov chain Monte Carlo (MCMC) algorithm. This yields a sample (indexed by $j \in (1, \ldots, J)$) of model parameter vectors, $\left(\mu_i^{(j)}, \phi_i^{(j)}, \psi_i^{2(j)}\right)$, for each country $i \in (1, \ldots, 200)$.

Forecasting

The goal is to forecast migration and population probabilistically by jointly generating a set of future migration and population trajectories. Smith et al. (2001) outline the features distinguishing a forecast from a projection. A projection describes a future outcome based on specific assumptions without assigning probabilities of realizing these outcomes. A forecast, however, arises from a statistical method that includes model-based probabilities of realizing these outcomes.

We use the same approach to forecasting fertility and mortality as the WPP 2019 (United Nations, 2019). Our migration forecasting method builds on the approaches of Azose & Raftery (2015) and Azose et al. (2016), with modifications to account for differences in population age distributions over time and across countries. Forecasts are generated through 2100 to compare the deterministic migration method from WPP 2019, the probabilistic age-agnostic migration method of Azose et al. (2016), and our probabilistic age-standardized migration method.

For each future trajectory $j \in \{1, \ldots, J\}$, we independently project the population for each country i by 5-year age group a through $a+4$ and sex s for year $t+5$, assuming no migration. A total of 21 age groups are included with age bins [0,5), [5,9), ..., [95,100), [100, ∞). The resulting projection, $\tilde{P}^{(j)}_{i,t+5,a,s}$, denotes a realization of the total population that would be observed in year $t+5$ had no one in the population migrated in or out of each country. This approximates the population at risk of migrating for the period t–$t+5$, stratified by age and sex.

We then independently generate an age-standardized net migration rate for the period t–$t+5$, $\text{NMR}_{i,t}^{\star(j)}$, by sampling from the previously estimated Bayesian hierarchical model of Azose & Raftery (2015). The sampled age-standardized net migration rate is subsequently decomposed into the age-standardized in-migration rate, $\text{IMR}_{i,t}^{\star(j)}$, and age-standardized out-migration rate, $\text{OMR}_{i,t}^{\star(j)}$, using the country-specific posterior predictive mean derived from model (1). Note that age-standardized coefficients for model (1), $\beta_0^\star$ and $\beta_1^\star$, are estimated from historical age-standardized rates ($\text{IMR}^\star$ and $\text{OMR}^\star$), ensuring that all rates are expressed on the same scale.

The mean age-standardized in-migration rate for trajectory j and country i is calculated as

$$\text{IMR}_{i,t}^{\star(j)} = \beta_{0,i}^\star + \beta_1^\star \max\left(\text{NMR}_{i,t}^{\star(j)}, 0\right). \tag{9}$$

The corresponding out-migration rate for the period starting in year t is given by

$$\text{OMR}_{i,t}^{\star(j)} = \text{IMR}_{i,t}^{\star(j)} - \text{NMR}_{i,t}^{\star(j)}. \tag{10}$$

Age-standardized in-migration and out-migration rates then need to be converted back to period-specific rates to calculate inflow and outflow counts corresponding to the projected population age structure:

$$\begin{aligned} \text{IMR}_{i,t}^{(j)} &= \text{IMR}_{i,t}^{\star(j)} \times \check{C}_t^{(j)} / \check{C}_{2020}, \\ \text{OMR}_{i,t}^{(j)} &= \text{OMR}_{i,t}^{\star(j)} \times C_{i,t}^{(j)} / C_{i,2020}. \end{aligned} \tag{11}$$

In-migration rates then are converted to a total inflow counts by multiplying the in-migration rate by the total country population for trajectory j at time t. These inflow totals are disaggregated by sex in proportion to the male and female population, following the approach of Azose et al. (2016), and by age using a Rogers-Castro-like migration age schedule. This yields the total inflow to country i in period t–$t+5$ for age group a and sex s for trajectory j, denoted $I_{i,t,a,s}^{(j)}$. Out-migration counts by age and sex, $O_{i,t,a,s}^{(j)}$, are calculated similarly.

While a Rogers-Castro-like age schedule is suitable for out-migration in most countries, the unique migration patterns in GCC countries require a different approach. In recent decades, GCC migration has been dominated by large inflows of male migrant workers employed on temporary visas, typically lasting two years or less, with little opportunity for long-term residency or citizenship. As these workers depart and are replaced by new arrivals, the overall age structure of the population remains relatively stable despite changes in migration volume.

To reflect this dynamic, we adjusted the migration age schedule to maintain the foreign-worker-dominated age structure in GCC countries. Rather than applying the standard Rogers-Castro-like model, we modified the out-migration schedule to emphasize outflows of older working-age migrants, aligning with the temporary nature of these populations. Specifically, the out-migration age schedule was derived as the difference between the age distribution of the population and the normalized Rogers-Castro-like schedule, effectively re-weighting outflows to reflect older workers leaving these countries.

This adjustment ensures that GCC migration patterns preserve the prime working-age structure (15–65) of the migrant population, preventing the aging-in-place of large migrant-worker cohorts. While out-migration rates have historically been much smaller than in-migration rates, these modifications better capture the dynamics of GCC migration and yield more plausible forecasts of population and migration trends through 2100 compared to using the approach used for all other countries.

Rebalancing global inflows and outflows each period was necessary to maintain global net-zero migration. While Azose & Raftery (2015) and Azose et al. (2016) proposed methods for this, their approaches do not directly apply to forecasts that generate inflow and outflow counts for each country. Adjusting global net migration counts without specifying the proportion allocated to inflows or outflows creates inconsistencies in the inflow and outflow trajectories relative to the global net flow. To address this, we resolved global net-zero deviations by adjusting country-level inflow and outflow totals in proportion to each country's population share of the global total.

The global rebalancing procedure is as follows. Let $\tilde{I}^{(j)}_{i,t,a,s}$ and $\tilde{N}^{(j)}_{i,t,a,s}$ denote the j^{th} inflow count and net migration count trajectories before global rebalancing, respectively, for country i, period t–$t+5$, age group a–$a+4$, and sex s. Then the adjusted inflow and outflow counts were calculated as

$$
\begin{aligned}
I^{(j)}_{i,t,a,s} &= \tilde{I}^{(j)}_{i,t,a,s} - w \left(\sum_{k \in K} \tilde{N}^{(j)}_{k,t,a,s} \right) \frac{\tilde{P}^{(j)}_{i,t,a,s}}{\sum_{k \in K} \tilde{P}^{(j)}_{k,t,a,s}}, \\
O^{(j)}_{i,t,a,s} &= \tilde{O}^{(j)}_{i,t,a,s} + (1-w) \left(\sum_{k \in K} \tilde{N}^{(j)}_{k,t,a,s} \right) \frac{\tilde{P}^{(j)}_{i,t,a,s}}{\sum_{k \in K} \tilde{P}^{(j)}_{k,t,a,s}},
\end{aligned}
\tag{12}
$$

where $w = 0.5$, evenly splitting the global net-zero deviations between inflows and outflows. The index K represents two normalization groups: GCC countries and their labor-supplying origins (Bangladesh, India, Indonesia, Philippines, Pakistan), and all other countries outside the GCC labor corridor.

This adjustment ensures global net migration for each age group and sex equals zero, with net migration for trajectory j in country i during t–$t+5$ given by $N^{(j)}_{i,t,a,s} = I^{(j)}_{i,t,a,s} - O^{(j)}_{i,t,a,s}$. As in Azose & Raftery (2015), these adjustments led to minimal changes to the raw forecasts.

We then recalculated the balanced age-standardized net migration rate, $\text{NMR}^{\star(j)}_{i,t}$, as

$$
\text{NMR}^{\star(j)}_{i,t} = \left(\frac{I^{(j)}_{i,t,+,+}}{\tilde{P}^{(j)}_{i,t+5,+,+}} \right) \times \check{C}_{2020} / \check{C}^{(j)}_{t} - \left(\frac{O^{(j)}_{i,t,+,+}}{\tilde{P}^{(j)}_{i,t+5,+,+}} \right) \times C_{i,2020} / C^{(j)}_{i,t}, \tag{13}
$$

which will be the rate jump-off for sampling in the next time period. Finally, we account for net migration in the population projection generated at the start of the forecasting routine,

$$
P^{(j)}_{i,t+5,a,s} = \tilde{P}^{(j)}_{i,t+5,a,s} + N^{(j)}_{i,t,a,s}, \tag{14}
$$

before moving to the next forecast period.

4 Validation

Age-standardized forecast efficacy was evaluated in terms of out-of-sample performance. Figure 4 shows the observed and four-period migration and population forecasts for the United States, El Salvador, South Africa, and Saudi Arabia starting in the 2000-2005 period through the 2015-2020 period. Forecasting methods include 1995-2000 migration rate persistence (effectively the current UN method) for every period, the age-agnostic method of Azose et al. (2016), and the age-standardized method described above.

Rows of the figure show the migration age structure index (MASI) ratio, i.e. the ratio of the MASI for the period to the MASI for the baseline year 2000, net migration rate, and

population for each country, while columns correspond to the four countries. Prediction intervals for both the age-agnostic and age-standardized method included the observed population and net migration rates in all cases. Population and migration forecasts using the age-standardized approach were as good or better than the age-agnostic forecasts by the end of the period for all cases except in El Salvador in terms of agreement with the observed quantities. Persistence of 1995-2000 migration rates performed especially poorly for El Salvador and Saudi Arabia.

The selected countries illustrate the differences in migration and population forecasts with and without age-standardization. The United States was included for its scale of migration, El Salvador for its rapidly aging population and high out-migration rate, South Africa for the relatively minor contribution migration makes to its population dynamics compared to fertility and mortality, and Saudi Arabia as a representative GCC country.

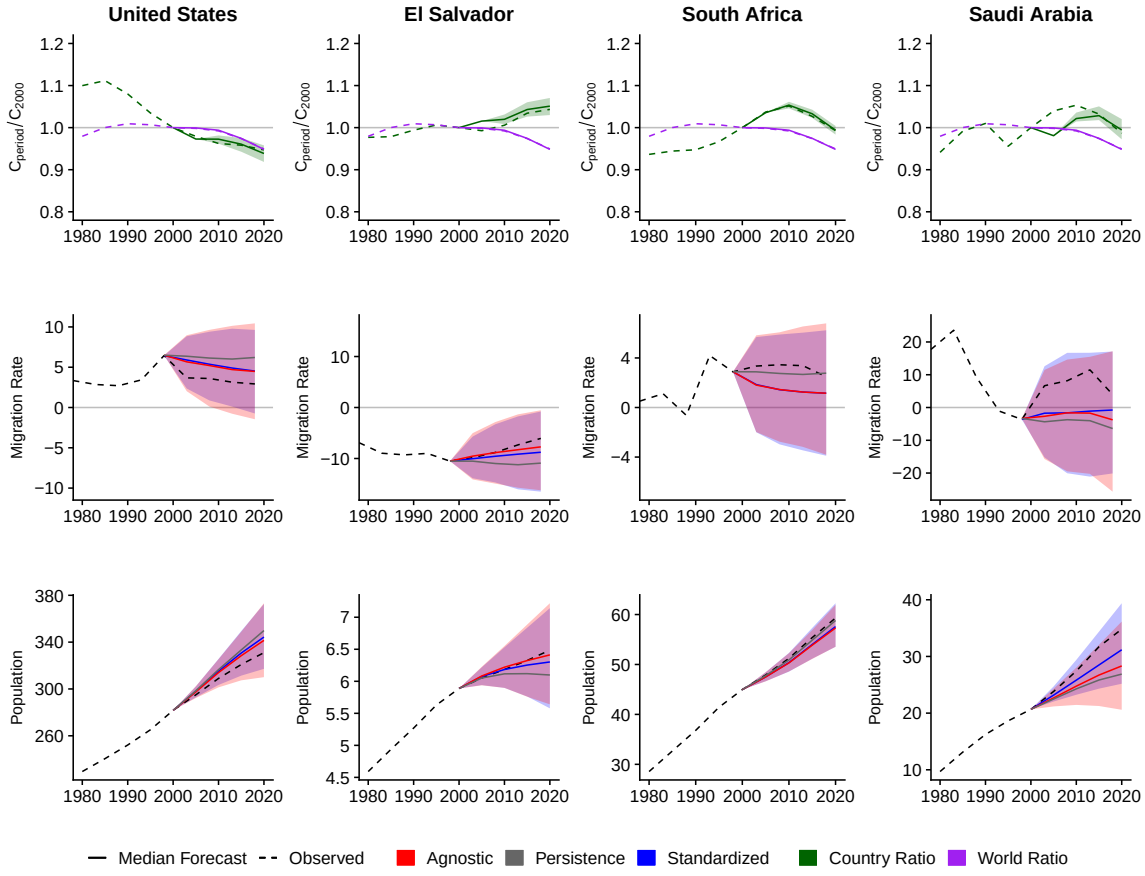

Figure 4: Migration age structure index (MASI) ratio for each country (■) and the globe (■) with base-year 2000, out-of-sample validation forecast of population (millions of people), and age-standardized and age-agnostic net migration rate (net annual migrants per thousand), for four countries. Forecasts use probabilistic age-standardized net migration (■), probabilistic age-agnostic net migration (■), observed fertility, and observed mortality. Dashed lines in each plot indicate the observed values. Solid lines indicate the median forecast. Shaded regions show the 80% prediction interval. Migration models were fit to 1950–2000 data. Forecasts are for the 2000–2005, 2005–2010, 2010–2015 and 2015–2020 periods. Migration flow estimates and forecasts are shown at the midpoint of each 5-year period.

Age-standardized and age-agnostic forecasts largely agree with one another since pop-

ulation age structure differences make little impact over such a short period of time. Note that the proportion of the migration age population increased relative to the baseline population in El Salvador. The population aged 0–15 was the largest proportion of the population in 2000; however, the proportion of the population aged 0–15 fell during the validation period and the large 0–15 population cohort aged into the prime migration age cohorts by the end of the 2015–2020 period. As a result, El Salvador's age structure ratio increased from the 2000 baseline period. The age structure ratio forecast outpaced the observed age structure ratio for the first two periods (2000–2010), but the prediction interval contained the observed age structure ratio for the last two periods (2010–2020).

Global age ratio forecasts closely matched the true global age structure, but country-level population age ratio forecasts were mixed. Prediction intervals for the MASI ratio captured all or most of the observed values for the United States and South Africa. For Saudi Arabia, the forecast MASI ratio was lower than the true value for the first two periods but included the observed value by the end of the forecast period. In El Salvador, the forecast MASI ratio consistently exceeded observed values from 2000 to 2010. These discrepancies are explained by the higher net migration observations compared to the forecasts for those countries.

Global and country-level MASI ratio values equal one for the baseline population structure (2000 in this case) by definition. When the global MASI ratio exceeds the country-level MASI ratio, then we expect higher in-migration using the age-standardized method compared to the age-agnostic method. Conversely, when the country-level MASI ratio is below 1, we expect lower net out-migration compared to the baseline period using the age-standardized method and a model that does not account for population age structure.

The United States MASI ratio remained below both the global MASI ratio and baseline MASI ratio throughout the forecast period. This combination should correspond to less out-migration from the United States as the global population had a net positive supply of migration-age people and the United States had a shrinking net positive supply of migration age people compared to the baseline population. Indeed, the age-standardized net migration forecast for the United States was shifted towards higher net in-migration compared to the age-agnostic forecast.

The forecast MASI ratios for El Salvador and South Africa exceeded both the global MASI ratio and their respective baseline ratios, indicating that the age-standardized migration forecast should be shifted towards more net out-migration compared to the age-agnostic forecast. The age-standardized net migration rate was indeed more negative than the age-agnostic migration rate for El Salvador. However, age-standardized and age-agnostic net migration forecasts were indistinguishable for South Africa, reflecting the relatively small share of population change attributable to migration there.

The forecast MASI ratio for Saudi Arabia exceeded the global MASI ratio and the baseline ratio, indicating that the age-standardized migration forecast should be shifted towards more net out-migration compared to the age-agnostic forecast if the forecasting methods were otherwise identical. However, the age-standardized approach for GCC countries uses a distinct methodology from all other countries, making direct comparisons to the age-agnostic approach less relevant. Still, net migration forecasts generated from the age-standardized method were more similar to the observed rates. Since migration is the main contributor to population change in GCC countries, the small improvement in the age-standardized forecast led to a substantial improvement in the population forecast compared to the age-agnostic approach.

Multiple forecast horizons were evaluated, generating predictions one to four periods ahead of the last observed data. Out-of-sample forecasts used the last observed population age structure as the baseline, fit each model to data available prior to the first forecast period, and generated forecasts for each period through the 2015-2020 period. One-period-ahead forecasts were generated for 2000, 2005, 2010, and 2015; two-period-ahead forecasts for 2000, 2005, and 2010; three-period-ahead forecasts for 2000 and 2005; and the four-period-ahead forecast for 2000 only. Five-period-ahead forecasts were not evaluated as it is not possible to fit the mixed-effects model (1) with only one period of in-migration and out-migration rate data.

We evaluated the median of the J country-specific forecasts in terms of the Mean Absolute Error (MAE), the Log Mean Absolute Error (LMAE), and the Mean Absolute Scaled Error (MASE). Let $f_{i,t}$ denote the net migration rate forecast for country i and period starting in year t, and $r_{i,t}$ the true net migration rate. The LMAE is defined as follows:

$$l(y) = \text{sign}(y)\left[\log\left(|y| + c\right) - \log c\right] \text{ with } c > 0$$

$$\text{sign}(y) = \begin{cases} 1 & y > 0 \\ 0 & y = 0 \\ -1 & y < 0 \end{cases}$$

$$LMAE_t = \frac{1}{200}\sum_{i=1}^{200} \left|l(f_{i,t}) - l(r_{i,t})\right|. \tag{15}$$

We used $c = 1$. The LMAE formulation prevents large errors in a few forecasts from dominating the error metric.

The MASE, recommended by Hyndman & Koehler (2006), evaluates forecast performance in terms of the mean ratio of errors from a proposed method (numerator) and errors from a naïve method (denominator), such as persistence forecast. Denominator errors are calculated from in-sample data, e.g. mean net migration rate errors generated from the 1950-2000 data using the naïve method, and compared to errors calculated from out-of-sample forecasts from a proposed method, e.g. mean net migration rate forecast errors in the 2000-2020 data.

The MASE for k-period-ahead forecasts, based on T_k forecast periods starting with period t_0–$t_0 + 5$ and S_k in-sample periods starting in period s_0–$s_0 + 5$, is defined as

$$\text{MASE}_k = \frac{\frac{1}{200\times T_k}\sum_{i=1}^{200}\sum_{\{t_0\}_k}\left|r_{i,t_0+5(k-1)} - f_{i,t_0+5(k-1)}\right|}{\frac{1}{200\times S_k}\sum_{i=1}^{200}\sum_{\{s_0\}_k}\left|r_{i,s_0+5k} - r_{i,s_0}\right|}. \tag{16}$$

This formula computes the ratio of average k-period ahead forecast errors for all possible horizons k in the in-sample and out-of-sample periods. For the $k = 1$ (one-period-ahead forecasts), $\{t_0\}_1 = \{2000, 2005, 2010, 2015\}$, $T_1 = 4$, $\{s_0\}_1 = \{1950, 1955, \ldots, 1990\}$, and $S_1 = 9$. For the $k = 4$ four-period-ahead forecast, $\{t_0\}_4 = \{2000\}$, $T_4 = 1$, $\{s_0\}_4 = \{1950, 1955, \ldots, 1975\}$, and $S_4 = 6$.

Forecast calibration was evaluated in terms of 95% prediction interval coverage and average prediction half-interval width. A well-calibrated model's 95% prediction intervals should include approximately 95% of the true net migration rate observations. Intervals containing fewer than 95% of true values are too narrow, underestimating forecast variation, while those containing more than 95% are too wide, overestimating variation. The

prediction half-interval width is half the difference between the upper and lower prediction interval quantiles. Models that are accurate, well-calibrated, and have narrower half-interval widths are preferred over those with wider intervals but similar accuracy and calibration. Both coverage calculations are approximations

Table 3 summarizes out-of-sample predictive performance of the age-standardized, age-agnostic, and net rate persistence methods. Errors are calculated as the difference between the observed value and the median of 2,000 posterior predictive distribution draws. Rate persistence uses the last observed net rate in each country as the forecast. The best score for each method is shown in bold font.

Table 3: Mean predictive performance of different methods for net migration rate (migrants per thousand period person years): mean absolute error (MAE), log mean absolute error (LMAE) with $c = 1$, mean absolute scaled error (MASE), 95% prediction interval coverage, and mean prediction interval half-width (HW).

	Method	MAE	LMAE	MASE	Cover	HW
5 Years	Persistence	4.02	0.68	1.42	—	—
	Agnostic	**3.44**	**0.65**	**1.25**	**93**	10.47
	Standardized	3.54	**0.65**	1.28	**93**	**10.39**
10 Years	Persistence	5.33	0.88	1.63	—	—
	Agnostic	**3.86**	**0.76**	**1.16**	**91**	11.63
	Standardized	3.93	0.77	1.17	90	**11.44**
15 Years	Persistence	5.16	1.00	1.44	—	—
	Agnostic	**3.49**	0.83	**1.10**	**92**	12.32
	Standardized	3.51	**0.82**	1.11	**92**	**11.85**
20 Years	Persistence	4.77	1.02	1.27	—	—
	Agnostic	2.91	0.82	1.00	**94**	12.54
	Standardized	**2.86**	**0.79**	**0.98**	**94**	**12.04**

The age-standardized and the age-agnostic model accuracy were similar across all horizons, but age-standardized forecast accuracy overtook age-agnostic forecast accuracy as the number of periods between the last observations and the forecast period increased for all metrics. The 95% prediction interval coverages were effectively indistinguishable across all forecast horizons, except the two-period-ahead forecast. Both methods led to slight undercoverage compared to the 95% nominal value. However, the age-standardized prediction interval half-widths were narrower for every forecast horizon. The similarity between the age-agnostic and age-standardized models over short horizons is expected, as population age structure changes gradually. Both probabilistic models outperformed the rate persistence model in all horizons.

The age-agnostic and net rate persistence models were fit using the *bayesPop* R package (Ševčíková & Raftery, 2016), while the age-standardized model was implemented using a custom implementation of the *bayesPop* R package.

5 Results

Figure 5 summarizes probabilistic population forecasts in four countries using the age-agnostic migration model from Azose et al. (2016), the population forecasts using the age-standardized migration model, and the projections from the WPP 2019 (United Nations, 2019). Columns of the figure correspond to the same country. Rows of the figure show forecasts through 2100 for the MASI ratio, the net migration rate, and total population.

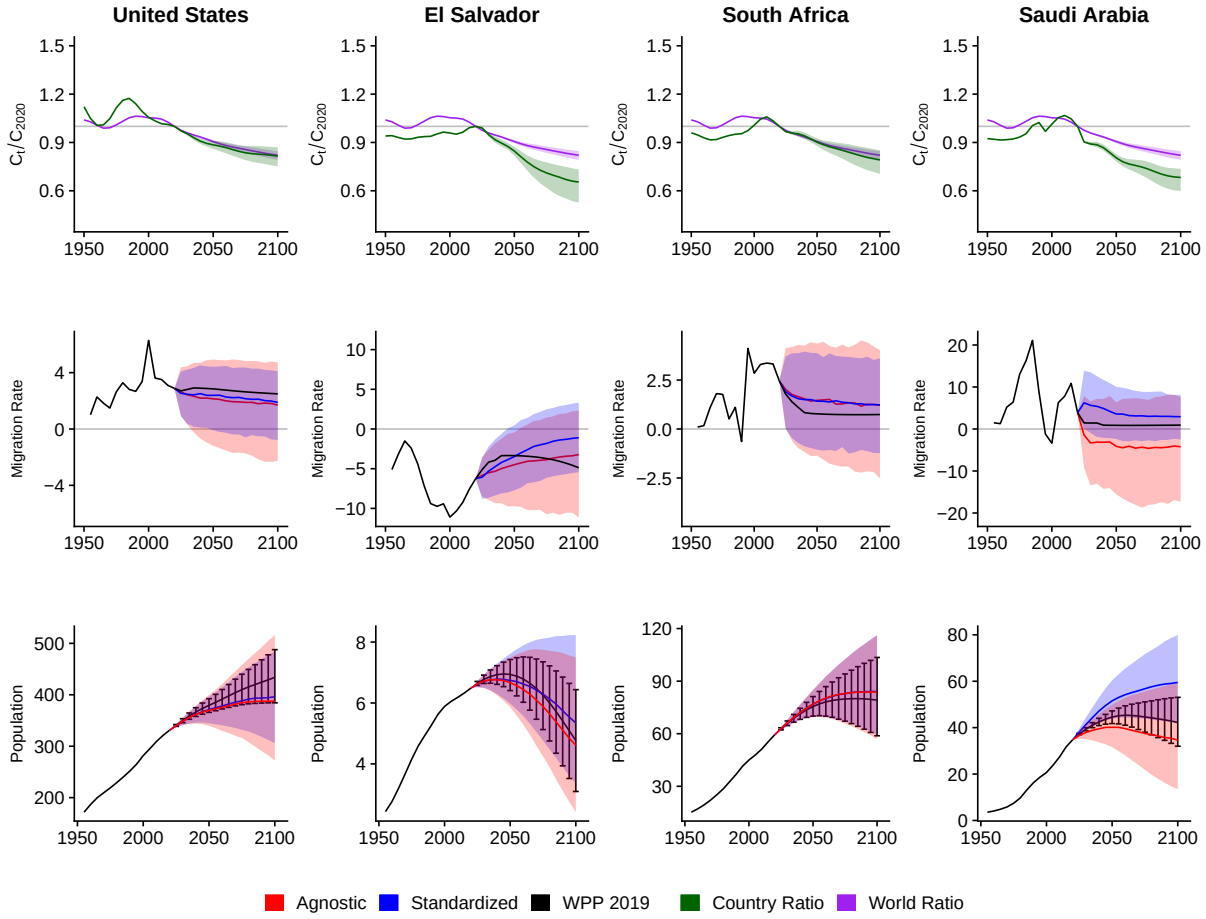

Figure 5: 2020 MASI ratios (top row), net migration rate as net annual migrants per thousand (middle row), and probabilistic forecast of population (in millions) age-standardized and age-agnostic (bottom row) for the United States, El Salvador, South Africa, and Saudi Arabia. Forecasts use probabilistic net migration (■=age-standardized and ■=age-agnostic), as well as probabilistic fertility, and mortality. Migration flow estimates and forecasts are shown at the midpoint of each 5-year period. The MASI ratio plots show the values for each country (■) and the world (■). Solid lines in each plot indicate the observed and median forecast. Shaded regions show the 80% prediction interval. Forecasts start in the 2020-2025 period.

The MASI ratios in the first row of Figure 5 summarize the differences in the population age structure from the 2020 baseline for each country (green) compared to the population age structure for the world (purple). Countries with a higher migration-age population than the 2020 population baseline rise above 1. Countries aging faster than the world population fall below the global index. Forecast population age structures in the United States and South Africa generally follow the global aging trend.

When the country-level MASI ratio is similar to the global ratio, the median net migration forecasts from the age-standardized model and age-agnostic model should be more similar compared to countries where the country-level and global MASI ratios diverge. The population age structures of El Salvador and Saudi Arabia are forecast to age much faster than the global average. In these cases, the age adjustment shifts the migration forecast towards higher net inflows.

Median net migration forecasts from both probabilistic net migration models are similar in many countries, but long-term age-standardized net migration forecast intervals (shaded blue regions) tend to be narrower than age-agnostic (shaded red regions) intervals. This is due to the population age normalization step used in the age-standardized method, which removes variation in historic net migration rates caused by differences in population age structure before fitting the Azose & Raftery (2015) model. Removing variation in past net migration estimates attributable to population age structure reduces the variance in the net migration model parameter estimates and hence the migration and population forecasts. Prediction intervals (80%) for both probabilistic net migration models include the United Nations' WPP 2019 projections through 2100 for all countries shown in Figure 5.

Median net migration forecasts from the age-agnostic and age-standardized methods are most similar in countries where population age indices align closely with the global index. The United States and South Africa demonstrate this trend in Figure 5.

In countries with population age structures that diverge from the global norm, such as El Salvador, the age-standardized method should produce markedly different forecasts. El Salvador's population is projected to age more rapidly than the global average from 2020 to 2100, resulting in a median age-standardized forecast with less net out-migration compared to the age-agnostic model. This adjustment yields a higher population forecast for the age-standardized model. Additionally, the age-standardized 80% prediction interval shows greater uncertainty about whether El Salvador's population will peak by 2100, with a higher upper bound compared to the age-agnostic forecast.

Figure 5 also shows that Saudi Arabia's population age structure is forecast to change about as fast as El Salvador's compared to the 2020 baseline population age structure. Saudi Arabia's net migration forecast is also shifted towards higher net in-migration, but this leads to a higher net positive migration forecast compared to the age-agnostic model. The age-standardized net positive migration forecast is better aligned to the historic net rate data in Saudi Arabia and the UN's WPP 2019 net migration forecast.

Differences in Saudi Arabia's migration and population forecast using the age-standardized approach reflect both population age structure effects and adjustments to modeling unique features of migration in GCC countries. These adjustments yield a substantially higher median population forecast and prediction interval. The UN WPP 2019 population forecast falls to the bottom of the age-standardized prediction interval, compared to its position near the upper middle of the age-agnostic model interval.

Figure 6 presents past and median forecast MASI ratios for the globe and selected countries by UN region. The MASI ratio provides a concise summary of historical and forecast population age structure dynamics. The MASI ratio contextualizes a country's structural migration potential relative to the reference population and global trends by weighting age groups using a Rogers-Castro-like migration age schedule. As a result, the MASI ratio provides a more relevant measure of migration-related population age dynamics than metrics like average age, which are less directly tied to structural migration potential, or summaries like population pyramids, which are harder to interpret across

periods and countries.

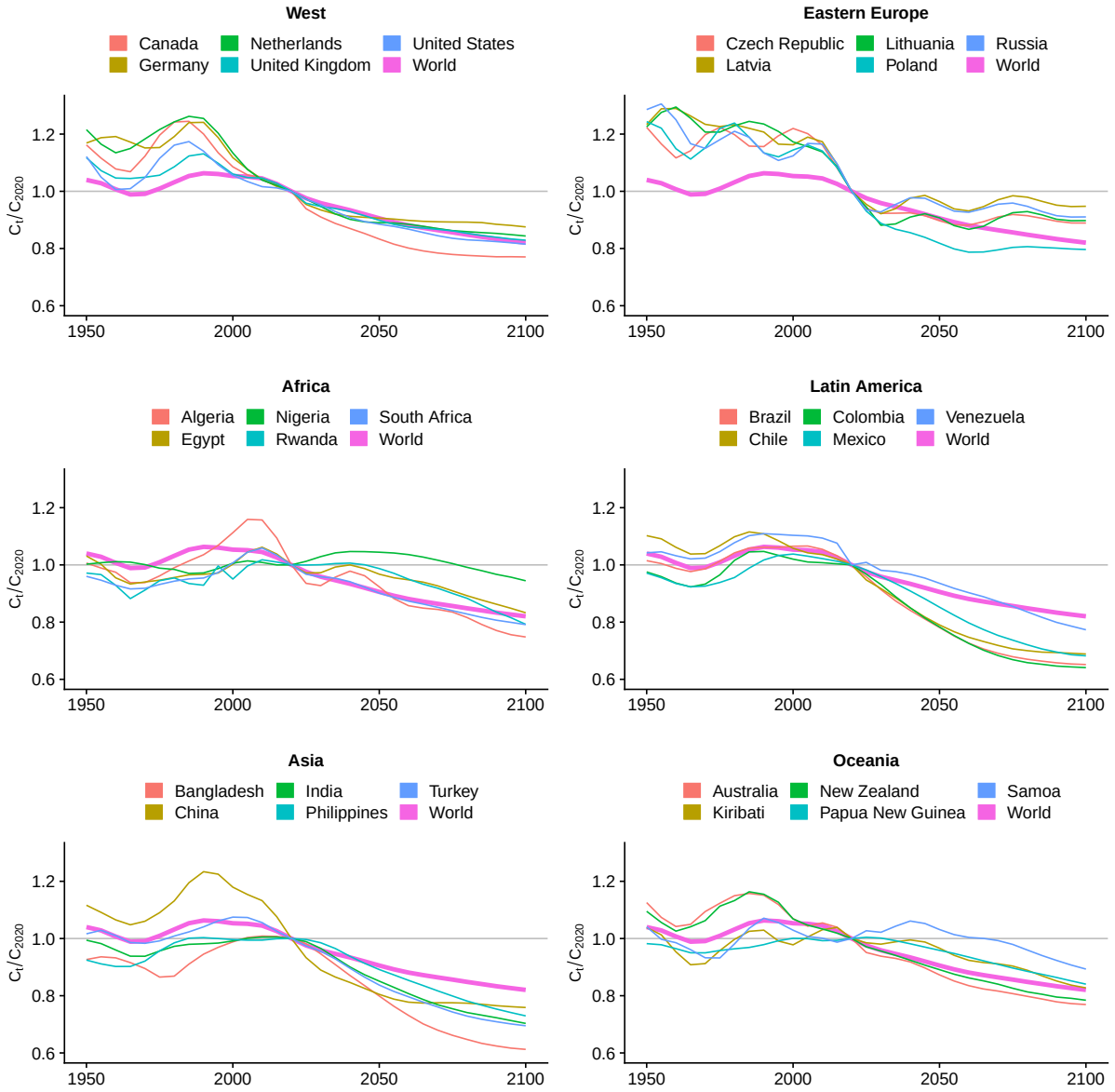

Figure 6: Historic and median country-level forecasts of MASI ratios for 2020 baseline by region compared to the world.

Figure 6 shows that the median population age structure forecasts in Western countries align with the global aging trend. This indicates that the that the underlying force of out-migration in Western countries will closely match the global average relative to 2020, leading to similar net migration forecasts from the age-standardized and age-agnostic methods.

In contrast, the migration force from Latin America is projected to decline sharply in the coming decades compared to the 2020 baseline and the global trend, resulting in higher net migration forecasts under the age-standardized method. For some countries, such as Nigeria (Africa) and Samoa (Oceania), the population age index has yet to peak, indicating that out-migration forecasts from these countries are underestimated in the age-agnostic method.

Age-standardized migration forecasts result in less severe population declines in countries facing significant demographic challenges. Figure 7 highlights the age index, net migration, and population forecasts for large countries expected to experience drastic population contractions by 2100.

The age-standardized forecast predicts less steep population declines compared to the age-agnostic model, as aging populations have fewer individuals of prime migration age. The age index indicates that the force of out-migration decreases as these countries age relative to the global population, shifting net migration forecasts toward higher in-migration. Median age-standardized migration forecasts align more closely with the UN's WPP 2019 migration and population projections than the age-agnostic forecasts.

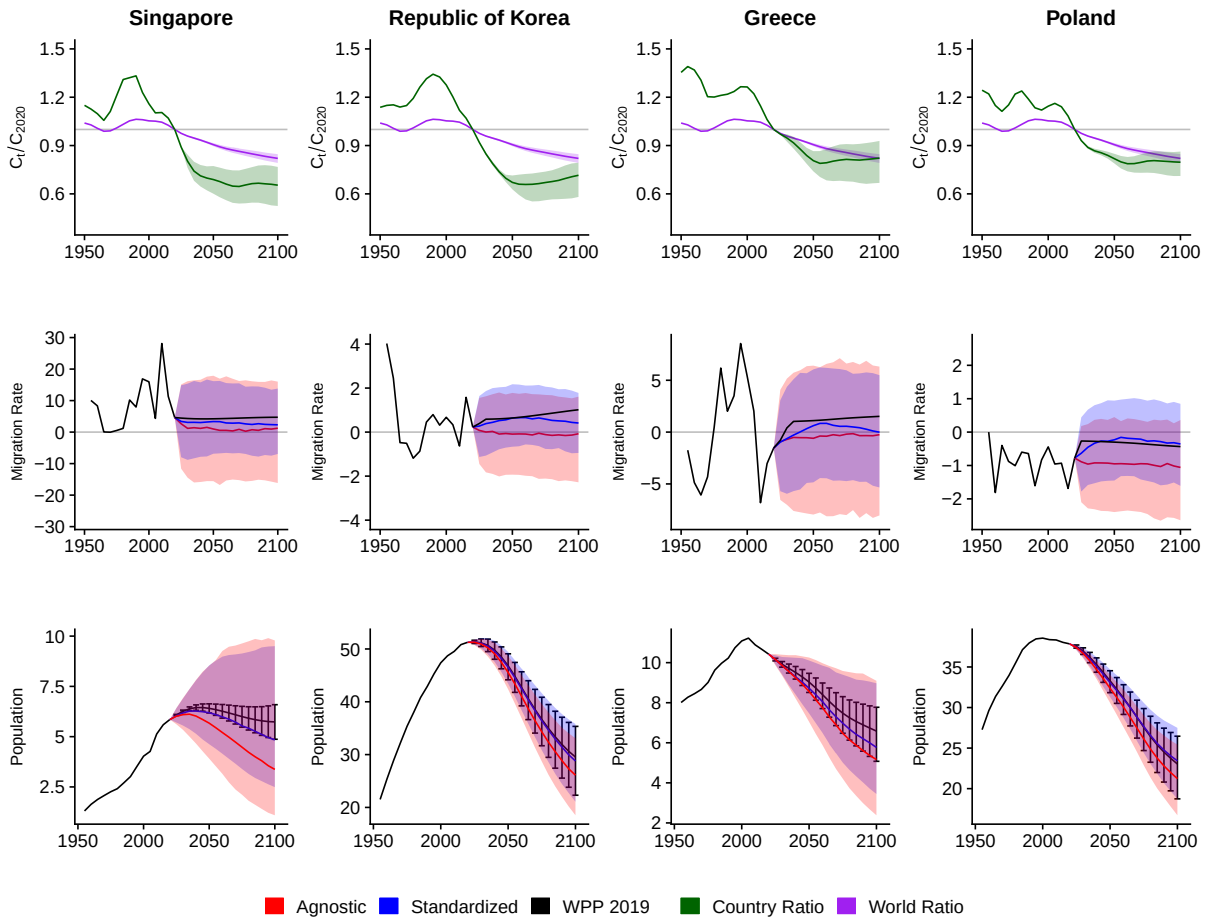

Figure 7: 2020 base-year MASI ratios (top row), age-standardized and age-agnostic net migration rate as net annual migrants per thousand (middle row), and probabilistic forecast of population (in millions) age-standardized and age-agnostic (bottom row) for Singapore, South Korea, Greece, Poland. Forecasts use probabilistic net migration (■=age-standardized and ■=age-agnostic), fertility, and mortality. Migration flow estimates and forecasts are shown at the midpoint of each 5-year period. Age-index ratio plots show the age structure ratios for each country (■) and world (■). Solid lines in each plot indicate the observed and median forecast. Shaded regions show the 80% prediction interval. Forecasts start in the 2020-2025 period.

Figure 8 shows the population forecast differences by 2100 for each country across three different migration models: age-agnostic and age-standardized probabilistic forecasts and the deterministic method from WPP 2019. Each point shows the deviation

of the median age-agnostic or age-standardized forecast from the median WPP 2019 projection, allowing for simultaneous comparison of all three models.

Points near the origin indicate close agreement among all three methods. Points along the horizontal dashed line indicate agreement between median age-agnostic forecast and WPP 2019 that differ from the median age-standardized forecast. Points on vertical dashed line indicate agreement between the median age-standardized forecast and WPP 2019 that differ from the median age-agnostic forecast. Points on diagonal dashed line, far from the origin, indicate agreement between the median age-agnostic and age-standardized forecasts that differ from WPP 2019. Regions bounded by the dashed lines, labeled I–VI, summarize other forecast combinations.

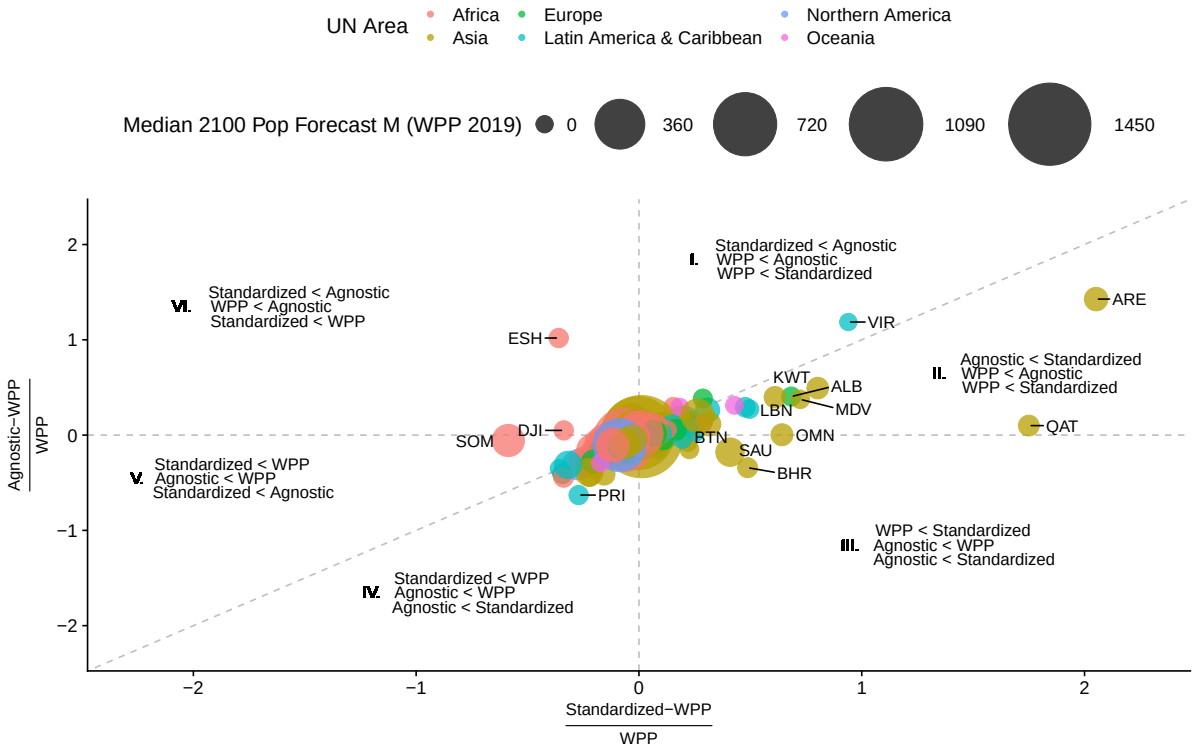

Figure 8: Three-way comparison of median 2100 population forecasts using age-standardized migration model, age-agnostic migration model, and median WPP 2019 forecast. Point color indicates UN Area, point diameter indicates WPP 2019 median population forecast size (millions), and regions I-VI show direction of forecast differences using the age-agnostic and age-standardized migration models.

Figure 8 shows that all three methods differ in Western Sahara (ESH). For the United Arab Emirates (ARE) and all GCC countries, the median age-agnostic and age-standardized methods forecasts are larger than WPP 2019, with the median age-standardized forecast being the highest.

The age-standardized method generates fewer negative net migration trajectories than the age-agnostic method, aligning with three decades of sustained positive net migration dominated by foreign workers supporting rapidly expanding economies. Net migration rate forecasts under the age-standardized approach maintain this positive trend, driving population growth in GCC countries.

While the age-standardized forecasts are higher than the age-agnostic forecasts, the resulting population age structures in most GCC countries are more plausible and closely

align with WPP 2019 projections.

6 Discussion

The link between migration and age is well established, yet few forecasting methods explicitly account for overall population age structure. We propose a statistical approach that incorporates population age structure changes into net migration forecasts for all countries, building on and addressing key limitations of Raftery & Ševčíková (2023).

Our approach addresses several criticisms of using net migration as a unit of analysis. By decomposing net flow rates into in-migration and out-migration rates, our model alleviates the migration age schedule issues identified by Rogers (1990), who noted that small net migration counts can obscure large, offsetting inflows and outflows. While net migration remains the unit of analysis, we use historic country-specific data to estimate in- and out-migration rates.

A bilateral migration flow model among all countries as in Welch & Raftery (2022) eliminates theoretical challenges induced by the use of net migration rate as the unit of analysis. However, the computational complexity of migration flow forecasting and limited historic bilateral flow data reduce the viability of this alternative for longer-term forecasting.

Probabilistic forecasts provide a systematic approach to incorporating uncertainty, but unexpected shocks to long-term trends can still arise without warning. Over the last decade alone, events such as the COVID-19 pandemic, Syria's civil war, Russia's invasion of Ukraine, and Venezuela's economic collapse have disrupted long-term norms. While new demographic patterns may emerge from these crises, their full effects should be integrated into updated forecasts as data become available.

Our proposed methods address several limitations, though others remain for future research. Using the global population age structure to normalize in-migration rates could be refined to better reflect the age structures of the most important origins. We adopted the global average for analytical tractability and to account for the potential evolution of migration corridors. Out-of-sample validation supports the reasonableness of this approximation.

We do not account for uncertainty in historic migration flow estimates used to calculate age-standardized net migration rates. The posterior predictive mean from model (1) was used to decompose net migration rates into in- and out-migration rates, as using the full posterior distribution added significant computational complexity without improving validation metrics or forecasts. Additionally, we apply a standard Rogers-Castro migration age schedule for non-GCC countries, which minimizes forecast uncertainty from schedule variation. Future work could address these approximations.

Our methods build on recent advances (Azose & Raftery, 2015), address key challenges of using net migration as the unit of analysis (Rogers, 1990), and provide a computationally efficient alternative to migration flow forecasting (Welch & Raftery, 2022). We also propose a method tailored to the unique dynamics of Gulf Cooperation Council countries, offering more accurate net flow forecasts for countries in this region. Lastly, the MASI ratio serves as a concise measure of population age structure, capturing the pace of aging over the coming decades.

References

Abel, G. J. (2013, January-June). Estimating Global Migration Flow Tables Using Place of Birth Data. *Demographic Research*, *28*, 505-546.

Abel, G. J., & Cohen, J. E. (2019, June). Bilateral International Migration Flow Estimates for 200 Countries. *Nature*, *6*(82), 1-13.

Azose, J. J., & Raftery, A. E. (2015). Bayesian Probabilistic Projection of International Migration. *Demography*, *52*(5), 1627–1650.

Azose, J. J., & Raftery, A. E. (2019). Estimation of Emigration, Return Migration, and Transit Migration Between All Pairs of Countries. *Proceedings of the National Academy of Sciences*, *116*(1), 116–122.

Azose, J. J., Ševčíková, H., & Raftery, A. E. (2016). Probabilistic Population Projections with Migration Uncertainty. *Proceedings of the National Academy of Sciences*, *113*(23), 6460–6465.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bloom, D. E., & Zucker, L. M. (2023, June). *Aging is the Real Population Bomb* (Tech. Rep.). International Monetary Fund.

Bongaarts, J. (2009). Human Population Growth and the Demographic Transition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1532), 2985–2990.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. doi: 10.2307/2286407

Coleman, D. (2008). The Demographic Effects of International Migration in Europe. *Oxford Review of Economic Policy*, *24*(3), 452–476.

Director of National Intelligence. (2021). *The Future of Migration* (Assessment No. NIC-2021-02486). United States Director of National Intelligence.

Fertig, M., & Schmidt, C. M. (2005). Aggregate-level Migration Studies as a Tool for Forecasting Future Migration Streams. In *International Migration* (pp. 129–156). Routledge.

Hyndman, R. J., & Booth, H. (2008). Stochastic Population Forecasts Using Functional Data Models for Mortality, Fertility and Migration. *International Journal of Forecasting*, *24*(3), 323–342.

Hyndman, R. J., & Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*, *22*(4), 679–688.

Kim, C. J. (2019). *Aging Societies: Policies and Perspectives* (Tech. Rep.). Asian Development Bank Institute.

Kolk, M. (2019). Period and Cohort Measures of Internal Migration. *Population*, *74*(3), 333–350.

Kupiszewski, M. (2002). How Trustworthy Are Forecasts of International Migration Between Poland and the European Union? *Journal of Ethnic and Migration Studies*, *28*(4), 627–645.

Lee, R. (2011). The Outlook for Population Growth. *Science*, *333*(6042), 569–573.

McAuliffe, M., & Oucho, L. A. (Eds.). (2024). *World Migration Report 2024*. Geneva: International Organization for Migration (IOM).

Münz, R. (2013, September). *Demography and Migration: An Outlook for the 21st Century* (Policy Brief No. 4). Migration Policy Institute.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Malden, Massachusetts: Blackwell.

Raftery, A. E., Alkema, L., & Gerland, P. (2014). Bayesian Population Projections for the United Nations. *Statistical Science*, *29*(1), 58–68.

Raftery, A. E., Li, N., Ševčíková, H., Gerland, P., & Heilig, G. K. (2012). Bayesian Probabilistic Population Projections for All Countries. *Proceedings of the National Academy of Sciences*, *109*(35), 13915–13921.

Raftery, A. E., & Ševčíková, H. (2023). Probabilistic Population Forecasting: Short to Very Long-Term. *International Journal of Forecasting*, *39*(1), 73-97.

RAND. (2005). *Population Implosion? Low Fertility and Policy Responses in the European Union* (Research Brief). Cambridge, UK: RAND Corporation.

Raymer, J., Guan, Q., Shen, T., Hertog, S., & Gerland, P. (2023, December). *Modelling the Age and Sex Profiles of Net International Migration* (Technical Report No. UN DESA/POP/2023/TP/No. 7). New York: Department of Economic and Social Affairs, Population Division.

Rogers, A. (1990). Requiem for the Net Migrant. *Geographical Analysis*, *22*(4), 283–300.

Rogers, A., & Castro, L. J. (1981). *Model Migration Schedules* (Research Report No. 81-30). Laxenburg, Austria: International Institute for Applied Systems Analysis.

Skjerpen, T., & Tønnessen, M. (2021). Using Future Age Profiles to Improve Immigration Projections. *Population Studies*, *75*(2), 255-267.

Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

United Nations. (2019). *World Population Prospects 2019: Methodology of the United Nations Population Estimates and Projections* (Tech. Rep. No. ST/ESA/SER.A/425). New York: Department of Economic and Social Affairs, Population Division.

United Nations. (2022, July). *World Population Prospects 2022: Methodology of the United Nations Population Estimates and Projections* (Tech. Rep. No. UN DESA/POP/2022/TR/NO.6). New York: Department of Economic and Social Affairs, Population Division.

United Nations. (2024, July). *World Population Prospects 2024: Methodology of the United Nations Population Estimates and Projections* (Tech. Rep. No. UN DESA/POP/2024/DC/NO.10). New York: Department of Economic and Social Affairs, Population Division.

Ševčíková, H., & Raftery, A. E. (2016). bayesPop: Probabilistic Population Projections. *Journal of Statistical Software*, *75*(5), 1–29.

Welch, N. G., & Raftery, A. E. (2022). Probabilistic Forecasts of International Bilateral Migration Flows. *Proceedings of the National Academy of Sciences*, *119*(35), e2203822119.

Wiśniowski, A., Smith, P. W., Bijak, J., Raymer, J., & Forster, J. J. (2015). Bayesian Population Forecasting: Extending the Lee-Carter Method. *Demography*, *52*(3), 1035–1059.

Woetzel, J., Madgavkar, A., Rifai, K., Mattern F., Bughin, J., Manyika, J., Elmasry, T., di Lodovico, A., Hasyagar, A. (2016, December). *People on the Move: Global Migration's Impact and Opportunity* (Report). Mckinsey Global Institute.

Yu, C. C., Ševčíková, H., Raftery, A. E., & Curran, S. R. (2023). Probabilistic County-Level Population Projections. *Demography*, *60*(3), 915–937.