DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 51, ARTICLE 11, PAGES 323–376
PUBLISHED 8 AUGUST 2024
http://www.demographic-research.org/Volumes/Vol51/11/
DOI: 10.4054/DemRes.2024.51.11

Research Article

A multidimensional global migration model for use in cohort-component population projections

Lucas Kluge **Orlando Olaya Bucaro**

Samir KC **Dilek Yildiz**

Guy Abel **Jacob Schewe**

Contents

Demographic Research: Volume 51, Article 11
Research Article

A multidimensional global migration model for use in cohort-component population projections

Lucas Kluge[1]

Orlando Olaya Bucaro[2]

Samir KC[3]

Dilek Yildiz[4]

Guy Abel[5]

Jacob Schewe[6]

Abstract

BACKGROUND
International migration is influenced by economic and social factors that change over time. However, given the complexity of these relationships, global population scenarios to date include only stylized migration assumptions that do not account for changes in the drivers of migration. On the other hand, existing projection models of international migration do not resolve all demographic dimensions necessary to interact with the cohort-component models typically used for population projections.

OBJECTIVE
Here we present a global model of bilateral migration that resolves these dimensions while also accounting for important external, economic, and social factors.

METHODS
We include age, education, and gender dependencies into a recently developed model of migration by origin, destination, and country of birth. We calibrate the model on bilateral

[1] Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany; Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany Email: kluge@pik-potsdam.de.
[2] Wittgenstein Centre (IIASA, VID/OEAW, WU) International Institute for Applied Systems Analysis, Laxenburg, Austria.
[3] Wittgenstein Centre (IIASA, VID/OEAW, WU) International Institute for Applied Systems Analysis, Laxenburg, Austria; Asian Demographic Research Institute, Shanghai University, Shanghai, China.
[4] Wittgenstein Centre (IIASA, VID/OEAW, WU) International Institute for Applied Systems Analysis, Laxenburg, Austria.
[5] Department of Sociology, Faculty of Social Sciences, University of Hong Kong, Hong Kong; Wittgenstein Centre (IIASA, VID/OEAW, WU) International Institute for Applied Systems Analysis, Laxenburg, Austria.
[6] Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany.

flow data, couple it to a widely used cohort-component population model, and project migration until 2050 under three alternative socioeconomic scenarios.

CONCLUSIONS
The extended model fits data better than the original migration model and is more sensitive to the choice of socioeconomic scenario, thus yielding a wider range of projections. Regional net migration flows projected by the model are substantially larger than in the stylized assumptions. The largest flows are projected in the most economically unequal scenario, while previously, the same scenario was assumed to have the smallest flows.

CONTRIBUTION
The results offer an opportunity to reconcile stylized migration assumptions with quantitative estimates of the roles of important migration drivers. The coupled migration-population modeling framework means that interactions between migration and other demographic processes can be captured, and the migration component can be evaluated in more detail than before.

1. Introduction

In many developed countries, migration is the dominant driver of demographic change, with fertility and mortality changing only slowly (UN DESA 2020). As a result, international migration cannot be neglected when investigating the demographic structure of both sending and receiving countries and specifically when estimating future population sizes and composition. State-of-the-art country-level population projections account for migration by specifying future net migration counts as a continuation of past levels or trends (UN DESA 2019b, 2022) or by assuming constant in- and out-migration rates at some percentage of past rates (KC and Lutz 2017; Lutz et al. 2018; Potančoková, Stonawski, and Gailey 2021). Some of these scenarios also involve a convergence toward zero net migration.

The advantage of such stylized assumptions is that they are easy to understand and circumvent many of the difficulties and uncertainties associated with forecasting future migration (Bijak 2010). Indeed, there are many complexities in the relationships between migration and its drivers, and between different drivers, that defy exact quantification or even a complete conceptual understanding given the paucity of related data (Bijak and Czaika 2020; Czaika and Godin 2022). On the other hand, a number of important interactions between migration and other variables are relatively well documented but are not taken into account in the projections mentioned above. For example, it has been shown that existing bilateral migrant networks, or diasporas, are strongly associated with increased migration, with the potential of creating a positive feedback between migrant

stocks and flows (Beine 2016; Hatton and Williamson 2005; Massey et al. 1993). While the magnitude of this feedback may be difficult to assess and subject to the influence of many competing effects, assuming constant migration counts or rates completely rules out any feedback.

More generally, it is not clear whether assuming present-day migration rates or counts (or any given percentage of them, for that matter) for, say, the year 2050, is consistent with other assumptions made in the population projections and underlying scenarios. For instance, educational attainment has been used to construct fertility rates (Lutz, Butz, and KC 2014). However, these same levels of education do not enter the migration assumptions, even though migration rates can differ significantly between low and highly skilled populations (Docquier and Marfouk 2004). The world in 2050 will also likely differ from today's in terms of income levels and economic inequalities, which might lead to changes in migration patterns that are not reflected in simple extrapolations of present-day migration rates or counts.

The goal of this paper is to replace the stylized migration assumptions in a state-of-the-art population projection model (the cohort-component model developed at the International Institute for Applied Systems Analysis (IIASA) (Lutz et al. 2018; Lutz, Butz, and KC 2017) for Horizon 2020's Future Migration Scenarios for Europe (FUME) project and model) with a mechanistic model of bilateral migration that accounts for key socioeconomic variables thought to drive migration. We build on a recently developed model of bilateral migration by country of birth, origin, and destination (Rikani and Schewe 2021) and extend it to account for the effects of age, education, and gender. These additional dimensions allow coupling the migration model with the FUME and give the possibility to explore specific demographic dimensions of future migration projections. We evaluate the extended model by measuring the explained proportion of variance in bilateral migration flow data and by comparing its results with those of the baseline model. We then present projections of bilateral migration flows obtained from the coupled migration–population projection model under three different scenarios and compare them with the stylized migration assumptions used previously. Furthermore, we provide the projected migration rates for each world region disaggregated by age, skill, and gender.

2. Methods and data

In this section, we first introduce the baseline migration model, followed by our extensions. Afterwards we briefly elaborate on the population projection model that provides dynamic population estimates, serving as input for our migration projections, followed by a section on the coupling process between both these models. Last, we introduce data used for calibration, evaluation, and projection, and present the measures used to determine the model's performance.

2.1 Migration model

2.1.1 Baseline model

We start with a deterministic, multiplicative model of bilateral international migration flows by country of birth, origin, and destination, describing migration as a function of the population at risk of migrating, the size of the bilateral diaspora, and average income levels in origin and destination countries (Rikani and Schewe 2021). The number of migrants born in k moving from country i to country j ($i \neq j$, $j \neq k$) is modeled as

$$M_{k,i\rightarrow j} = \sigma F(G_i) g_j^{\alpha_g} p_{k,j}^{\alpha_p} P_{k,i}, \tag{1}$$

where $P_{k,i}$ denotes the population stock of individuals born in k and residing in i, G_i the gross domestic product per capita (GDPpc) of the origin country, and $g_j = G_j/G_{Glob}$ the relative GDPpc of the destination country, where G_{Glob} denotes the global averaged GDPpc. $p_{k,j} = P_{k,j}/P_k$ is the relative diaspora (i.e., the proportion of people born in k and residing in j in terms of the total number of people born in k and residing outside of k). σ, α_g, α_p are fitting parameters.

The function $F(G_i)$ is given by

$$F(G_i) = F_{intent}(G_i) \cdot F_{resource}(G_i) = \frac{1}{1+\frac{G_i}{\hat{G}}} \frac{1}{1+e^{-\gamma(G_i-\tilde{G})}}, \tag{2}$$

where $\hat{G}$, $\tilde{G}$, and γ are fitting parameters. The first term $F_{intent}(G_i)$ is a hyperbolic function that yields higher values for low-income levels; it is a stylized representation of the prevalence of migration intentions, being stronger for poorer countries than for richer countries. The second term $F_{resource}(G_i)$ is a sigmoidal function that yields higher values for high-income levels; it is a stylized representation of the resources (financial and other) available to potential migrants to afford moving. Combining these two functions results in a hump shape with a maximum for intermediate levels of GDPpc, closely reproducing a nonparametric fit of the variation of estimated out-migration rates in terms of origin-country GDPpc (Rikani and Schewe 2021). This 'migration hump' has been documented in migration flow and stock data, and migration intentions and resource constraints have been suggested as causes, along with other factors (Hatton and Williamson 2005; De Haas 2007; Clemens 2014; Williamson 2014; Martin and Taylor 1993).

Apart from this nonlinear term, the form of Equation (1) is similar to 'gravity-type' regression models commonly used in econometric analyses of migration (e.g., Grogger and Hanson 2011; Ortega and Peri 2013; Beine 2016; Helbling and Leblang 2019; Wesselbaum and Aburn 2019). Both other predictor variables, destination income levels and

the size of existing bilateral diasporas, have been shown consistently to have a strong positive effect on bilateral flows; the former is thought to be related to economic opportunities as a major 'pull factor', while the latter is explained by the diaspora's role in lowering migration costs for newcomers and easing their access to the labor market, among other factors (Beine, Docquier, and Özden 2011; Beine 2016; Massey et al. 1993). Notably, the inclusion of the diaspora variable renders insignificant many other dyadic variables – such as common language, geographic distance, or colonial ties – that are important in other models (e.g., Ortega and Peri 2013). Another difference from common 'gravity' models is that, because of the nonlinear origin-GDP term, the model cannot be estimated in logs using linear least squares estimators. Instead, the untransformed model is estimated using nonlinear least squares, which also helps to accommodate zero-values inflow and stock data, possibly leading to biases in log-transformed models (Santos Silva and Tenreyro 2006).

This simple model neglects many other factors (drivers) that have been shown in some contexts to influence international migration, such as migration policies (Mayda 2010), political stability (Giménez-Gómez, Walle, and Zergawu 2019), or environmental change (Black et al. 2011). On the other hand, using country of birth as an additional dimension besides country of origin and country of destination, the model can account separately for out-migration and return-migration. Especially return-migration flows can be of significant magnitude (Azose and Raftery 2019). While out-migration is described by Equation (1), return-migration is modeled as a linear function of the population at risk:

$$M_{i,j\to i} = \kappa P_{i,j}, \tag{3}$$

with κ being a linear scaling factor. This is motivated by findings indicating that the number of return-migrants strongly correlates with the size of the corresponding diaspora (i.e., the relevant population at risk); whereas economic factors, for example, only have a limited impact on return flows (Rikani and Schewe 2021; Constant 2020; Battistella 2018).

While Equation (1) through (3) model migration flows at a given time step, population stocks evolve according to

$$P_{k,i}\,(t+1) = \begin{cases} \tilde{P}_{k,i}(t) \cdot \left(1 - r_k^{\dagger}\right) & k \neq i \\ \tilde{P}_{i,i}(t) \cdot \left(1 + r_i^{*} - r_i^{\dagger}\right) + \sum_{l\neq i} \tilde{P}_{l,i}(t) \cdot r_l^{*} & k = i \end{cases}, \tag{4}$$

where $\tilde{P}_{k,i}(t) = P_{k,i}(t) - \sum_{j\neq i} M_{k,i\to j} + \sum_{j\neq i} M_{k,j\to i}$ and $r_i^{\dagger}$, r_i^{*} being the country specific mortality and fertility rates, respectively. This system of equations allows for

projecting migration flows and population stocks over time. In the following, we present and test several modifications to these equations.

2.1.2 Extended model

To now extend this model, we first replace the destination GDPpc term in Equation (1) with a destination-to-origin GDPpc ratio. Second, instead of the general exponent used for the previous destination GDPpc term, we add gender-dependent exponents to the destination-to-origin GDPpc ratio. Third, we replace the general diaspora exponent with skill-dependent exponents. Next, we introduce age as another dimension, through age-specific multiplicative factors in Equation (1). Finally, we replace Equation (4) with a state-of-the-art cohort-component population projection model based on Lutz, Butz, and KC (2017) and Lutz et al. (2018) (FUME). From the perspective of the migration model, this last step allows endogenizing projections of age-specific fertility, mortality, and education transitions, and accounting for, for example, for the interaction between education and fertility, instead of assuming uniform fertility and mortality rates. From the perspective of the population model, previous migration assumptions, based on linear extensions of past mean migration rates or trends, are replaced with a mechanistic migration model that accounts explicitly for the roles of major drivers of global migration: diasporas, economic differences in countries of origin and destination, and natural population change. It also importantly accounts for interactions between different migration corridors.

Extending the migration model to have age, education, and gender dimensions is a technical necessity for coupling it with the cohort-component population projection models. However, including age, education, and gender effects also reflect important heterogeneities observed in the real world. Additionally, it enables a more detailed analysis of the migration projections, allowing users to examine each of the 12 demographic cohorts (3 age, 2 sex, 2 education) separately.

It has been shown that all three factors have major impacts on migration behavior (Lutz et al. 2018; Kaczmarczyk and Okólski 2005; Bijak et al. 2005; Docquier and Marfouk 2004). The scale of migration differs significantly between different age groups (Raymer, De Beer, and Van der Erf 2011; Raymer and Wiśniowski 2018; Wiśniowski et al. 2016; Zagheni and Weber 2012): The highest mobility is typically observed for young adults in their twenties and early thirties. Mobility is lower for older age groups as well as for children, except for very young children who have not entered school yet. Although details vary, the general shape of this age profile can be found for many countries (Wiśniowski et al. 2016; Zagheni and Weber 2012; Nawrotzki and Jiang 2015).

Similarly, the educational attainment of an individual may have a significant impact on her/his mobility and prospects when choosing potential migration destinations (Docquier and Marfouk 2006). Research shows that there is a discrepancy when comparing the migration flows of high- and low-skilled workers (Arif 2022), – for example, esti-

mating more high-skilled individuals to change their residence than low-skilled workers (Docquier and Marfouk 2004; Dao et al. 2021; Massey et al. 1993). These differences can be caused by origin-related factors, like the financial capability of affording such a trip, as well as destination-related factors, such as job opportunities and wages (Cattaneo 2007).

Furthermore, literature shows that gender can play a major role when estimating the number of migrants. Considering only the total numbers, it has been observed that men move more often than women do (UN DESA 2022). This gender migration gap cannot be assumed to be uniform. The ratio of female to male migrants varies significantly with different origins and destinations. Reasons for such differences can be diverse. Some literature focuses on external factors, finding that the number of women migrating increases with women's economic freedom and women's rights (Neumayer and Plümper 2021). Considering individual motives, studies state that women move more often for educational purposes than men do but less often for business and economic-related reasons (Yilma and Regassa 2019; Williams 2009). Additionally, gender plays into other factors that are known to be important when considering migration choices – for example, that overqualification affects migrant women more than male migrants (Akgüç, M. and Parasnis, J. 2023).

We include a simple first-order representation of these effects in the global migration model. Replacing the destination-GDPpc term in Equation (1) with the destination–origin GDPpc ratio, and accounting for multiple education levels, age groups, and genders in both Equations (1) and (3), we write

$$M_{s,e,a,k,i\to j} = \sigma_a F(G_i) \left(\frac{G_j}{G_i}\right)^{\alpha_{g,s}} p_{k,j}^{\alpha_{p,e}} P_{s,e,a,k,i} \tag{5}$$

for out-migration and

$$M_{s,e,a,i,j\to i} = \kappa_a P_{s,e,a,i,j} \tag{6}$$

for return-migration.

While the model would mathematically allow it to be used for transit migration, it would take a considerable amount of computational power to calibrate such a model and even more to couple it with the existing population projection model. Therefore, transit migration is not included in this study.

Indices $e \in \{l, h\}$(l=low, h=high), $a \in \{$0–24, 25–64,65+$\}$, and $s \in \{f, m\}$(f=female, m=male) denote education level, age group, and gender, respectively.

The high education level corresponds to ISCED-97 levels 3–6, (i.e., at least upper-secondary education), whereas the low education level corresponds to ISCED-97 levels 0–2. Our choices for the delineation of education levels and age groups are based on the

structure of available datasets (Wittgenstein Centre for Demography and Global Human Capital 2018), as well as the goal to avoid unpopulated cohorts.

In this study, we propose a refinement in modeling migration flows by replacing the previously used metric of relative destination GDPpc with a novel GDPpc destination–origin ratio. This approach redefines the 'pull factor' by highlighting migration corridors that were previously underrepresented. By focusing on the GDPpc comparison between destination and origin countries, this method acknowledges destinations that are relatively more attractive in their regional context, even if they possess a globally lower GDPpc. This shift in perspective could be crucial in identifying and understanding migration patterns to destinations that, under the previous model based on relative destination GDPpc alone, might have been overlooked due to their less prominent global economic standing.

Furthermore, we included the education/skill dependency by implementing skill-dependent diaspora coefficients, assuming that already-existing migrant communities have differing effects on people based on their skill. These assumptions are supported by literature, such as Beine, Docquier, and Özden (2010), who find that countries with larger diasporas will attract a higher proportion of low-skilled migrants, or Kracke and Klug (2021), who hypothesize that larger diasporas increase the effects of overqualification.

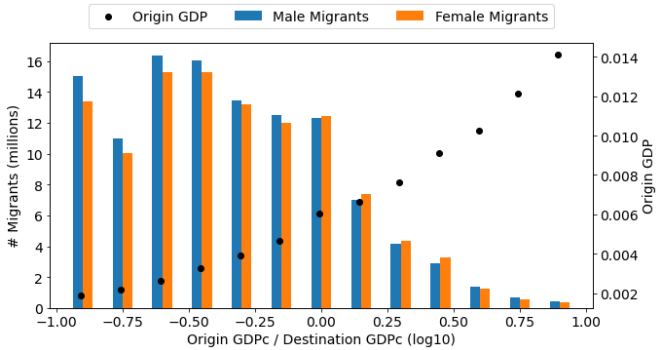
The effect of age on the size of migration flows is limited to the linear scaling factors σ_a and κ_a. This implies, in particular, the assumption that the reasons why migration rates differ between age groups are unrelated to any of the other model variables, (i.e., GDP or diasporas).

Our approach assumes different levels of mobility between three broad age groups – youth (below 25 years of age), working-age (25 to 64), and retired (65 and above) – allowing us to capture in a crude way the observed age profiles discussed above, with higher migration rates in the working-age population and lower migration in the young and the elderly populations. Because our parameters are globally uniform, we do not account for phenomena pertaining to specific migration corridors – for example, retirement migration, such as from Great Britain to Spain (Wiśniowski et al. 2016).

In our efforts to integrate gender dynamics into the migration model, we examined the sex-specific migration flow estimates provided by Abel and Cohen (2022). Our analysis sought to identify correlations between these estimates and the existing migration drivers in our model. Notably, we observed a higher proportion of male migrants in corridors where the GDPpc of the origin country was lower than that of the destination country (Figure 1). Conversely, the majority of migration flows where the destination GDPpc is lower than the origin's GDPpc are dominated by female migrants. Additionally, an increase in the average origin GDPpc was noted in migration corridors with a rising origin-to-destination GDPpc ratio. These observations are consistent with the findings of Neumayer and Plümper (2021), which suggest a trend of increased female migration from countries exhibiting greater gender-related economic freedom. Furthermore, these

patterns echo the research by Naveed et al. (2023), linking gender equality with economic growth.

Figure 1: **Migration flows by gender depending on-origin-to destination GDP per capita ratio and origin GDP per capita for the related bins**



Source: Migration data from F1, GDP data from G1, G2.
Notes: For data abbreviations please refer to Table A-1. Migration flows (indicated by orange and blue bars) were binned on the log scale. Origin GDP per capita (indicated by black dots) values were averaged for each bin.

To understand the effects of our model extensions, we additionally investigate three intermediate models: one accounting only for education but neither gender nor age,

$$M_{e,k,i\to j} = \sigma F(G_i)\left(\frac{G_j}{G_i}\right)^{\alpha_g} p_{k,j}^{\alpha_{p,e}} P_{e,k,i}, \tag{7}$$

$$M_{e,i,j\to i} = \kappa P_{e,i,j}; \tag{8}$$

one accounting only for gender,

$$M_{s,k,i\to j} = \sigma F(G_i)\left(\frac{G_j}{G_i}\right)^{\alpha_{g,s}} p_{k,j}^{\alpha_p} P_{s,k,i}, \tag{9}$$

$$M_{s,i,j\to i} = \kappa P_{s,i,j}; \tag{10}$$

and the last one accounting only for age,

$$M_{a,k,i\to j} = \sigma_a F(G_i) \left(\frac{G_j}{G_i}\right)^{\alpha_g} p_{k,j}^{\alpha_p} P_{a,k,i}, \tag{11}$$

$$M_{a,i,j\to i} = \kappa_a P_{a,i,j}. \tag{12}$$

2.2 Population projections

The population is projected using the FUME, based on the previously developed model at Wittgenstein Centre for Demography and Global Human Capital (WIC, a multi dimensional demographic cohort-component model (Rogers 1975; Keyfitz 1985) with the population disaggregated by age, sex, and education).

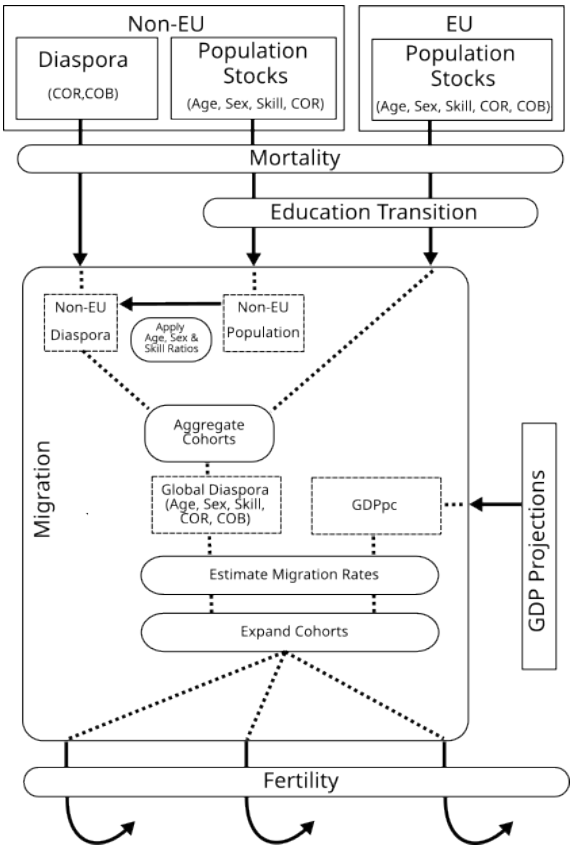
The WIC population projection model uses education-specific fertility, mortality, and educational transitions – informed by a global expert inquiry – to project the population in 201 countries by age, sex, and education (Lutz et al. 2018; Lutz, Butz, and KC 2017). The population in the projection is stratified into five-year age groups (0–4, 5–9, …., 100+), six levels of educational attainment (no education, some primary, primary, lower secondary, upper secondary, and post-secondary), and sex. Inputs to the projection model are education-, age-, and sex-specific mortality rates; education- and age-specific fertility rates; and education transitions by sex and age.

To incorporate the new migration model in the FUME population projection, the population projection model was expanded with the additional dimension of country of birth. The population is therefore subdivided by country of birth in addition to the above-mentioned dimensions of age, sex, and education. In the work of implementing country of birth in the projections, additional data were analyzed to produce country-of-birth-specific mortality, fertility, and education transition rates (Eurostat 2023; Statistics Sweden 2023). However, the analysis of available data showed small differences in mortality and fertility between native-born and foreign-born in countries with available data, such as the Nordic countries. Larger differences between groups are observed by education rather than by country of birth. Moreover, many countries lack the records necessary to produce country-of-birth-specific mortality, fertility, and education transition rates. Therefore, it is assumed in the population model that there are no mortality and fertility differences by country of birth.

2.3 Model coupling

In the following, we briefly summarize the coupling procedure between the migration and the underlying population projection model. A sketch of this process is displayed in Figure 2.

Figure 2: Scheme of the projection model



Notes: Sharp corners indicate datasets, round corner indicate processes. Dotted lines indicate processes that are performed within the migration model. COR = Country of residence; COB = Country of birth.

Each time step of the coupled model consists of the following steps: (1) application of mortality and educational transition rates; (2) estimation and application of bilateral migration flows; and (3) application of fertility rates. The input data that the migration

model receives from the population projection model consist of population stocks by education, sex, age, and country of residence for non-EU countries, and population stocks by education, sex, age, place of residence, and country of birth for EU countries. These data have six education groups, 21 age groups, and two genders; to use them in the migration model, we aggregate into the two education groups and three age groups mentioned above. Separately, the population model also provides population stocks by country of birth and country of residence for all countries but is not broken down by education, sex, and age. However, to calculate return flows, the migration model requires migrant stocks by country of birth, age, education, and sex for non-EU countries too. To estimate these stocks, we start by calculating the fraction of the total population assigned to a specific age, sex, and education cohort for each country:

$$\chi_{s,e,a,i} = \frac{P_{s,e,a,i}}{\sum_{s,e,a} P_{s,e,a,i}}. \tag{13}$$

We then apply these fractions to the diaspora data $P_{k,i}$, which are available for each country:

$$P_{s,e,a,k,i} = \chi_{s,e,a,k}\, P_{k,i}. \tag{14}$$

That is, for non-EU countries, we assume that the age, sex, and education distribution of the total population applies equally to migrants and non-migrants.

After estimating the migration rates, we extend these rates to fit the more fine-grained cohort structure of the population projection model, by assuming that migration rates are equal within each age and education cohort of the migration model. For example, the migration rate calculated for the 0 to 24 years age group is applied equally to the 0 to 4, 5 to 9, 10 to 14, 15 to 19, and 20 to 24 years age groups in the population model.

Once migrants are attributed to their respective destinations, they are subject to the fertility, mortality, and education transition rates of the country of residence. We did not implement any mechanics for people to lose their education when migrating, which might occur if destination countries do not acknowledge certain foreign degrees or education.

Furthermore, it should be mentioned that in order to investigate the impact of different, isolated scenarios on the migration model, the population projections always follow the SSP2 assumptions (Lutz et al. 2018; Lutz, Butz, and KC 2017), while the migration model simulates SSP1, SSP2, and SSP3. SSP1, titled 'Taking the Green Road,' emphasizes environmental conservation, reduced reliance on resources, and diminishing inequality. SSP2, known as 'Middle of the Road,' depicts a scenario where current societal, economic, and technological trends progress at a moderate pace without significant shifts. SSP3, 'A Rocky Road,' portrays a scenario of a divided world marked by increas-

ing nationalism, regional conflicts, and diminished international cooperation, resulting in inequality and decreasing technological progress (O'Neill et al. 2014).

2.4 Data

We calibrate and evaluate the model on 172 countries (Table A-6). This number is the result of intersecting several datasets, such as population, GDP, or diaspora data. All datasets are summarized in Table A-1. Hereafter, we use the abbreviations listed in the table when discussing these datasets.

For the population data, we use datasets P1 and P2, describing bilateral migrant stocks obtained from the UN Department of Economic and Social Affairs and total population estimates by age, education, and sex from the Wittgenstein Centre for Demography and Global Human Capital (for further information on data from the Wittgenstein Centre, we refer to Lutz et al. (2018)).

The total population stocks are divided into 21 five-year age groups, ranging from 0 to 4 years to 100+ years; ten educational attainment groups ranging from no education to master and higher; and female and male. To match the cohort structure of the migration model, we aggregate the different levels of educational attainment and age groups to obtain high- and low-skilled population ratios for three broad age groups of 0 to 24yrs, 25 to 64yrs and 65+ years. No educational attainment level is indicated in P2 for age groups below age 15, given that these are children of school and preschool age. For the migration model, we assign these age groups to the educational attainment distribution of the respective 'parent' cohorts, assuming that children under 15 migrate with their parents:

$$P_{s,e,\leq 15,i} = P_{\leq 15,i} \frac{P_{e,25-64,i}}{P_{25-64,i}}. \tag{15}$$

The resulting numbers are then added to the high- and low-skilled cohorts for the 0 to 24 year age group:

$$P_{s,e,0-24,i} = P_{e,\leq 15,i} + P_{e,16-24,i}. \tag{16}$$

The resulting sex, age, and education distribution is then also applied to the bilateral stock data P1 according to country of birth – that is, the sex, age, and education distribution of a given immigrant group (diaspora) $P_{k,i}$ is assumed to be equal to that of the total population of country k. This procedure is repeated every time step; therefore individuals till the age of 15 will obtain their 'parents' education while persons older than 15 acquire their own education.

Historic, country-specific GDP data G1 come from Penn World Table and is reported in terms of 2005 purchasing power parity. Where data points are missing in G1, they are taken from a harmonized collection of GDP data, G2, and adjusted to 2005 purchasing power parity (Geiger 2018). For future projections, we use GDP projections under the Shared Socioeconomic Pathways (SSPs) G3 (Koch and Leimbach 2022). These are updates to earlier projections (Dellink et al. 2017) and include the impacts of the COVID-19 pandemic. We note that they do not account for the economic effects of the Russia–Ukraine war.

For calibration of the model's parameters, and evaluating the goodness of fit, we use three datasets. First, we use bilateral flows F1 (Abel and Cohen 2022), derived from UN migrant stocks using a pseudo-Bayesian method (Azose and Raftery 2019), allowing bilateral nonzero flows and therefore including return flows. Data come in five-year intervals, covering the period from 1990 to 2020. Additionally, we use F2 (Yildiz and Abel 2024), consisting of age-dependent emigration and immigration numbers, for extracting age-dependent migration profiles. F2 was created as described in Yildiz and Abel (2024). Age cohorts in F2 are divided into 15 five-year cohorts, ranging from 0 to 4 years to 65+ years. In addition to the migration flow estimates from F1 and F2, we use F3 data (DEMIG 2023) consisting of 34 reporting countries and from up to 236 countries of origin, including annual numbers over the 1946–2011 period. Before using the flows, we aggregate them into five-year time steps to fit the period from 1990 to 2010 and therefore be comparable to F1. In total the processed data that we use for calibration and evaluation consists of roughly 2,000 different migration corridors and about 6,000 migration flows for each male and female migrant. For comparison, F1 offers about 50,000 corridors and about 300,000 data points for both genders.

Last, we use data on bilateral refugee stocks R1, available from the UN High Commissioner for Refugees, to estimate non-refugee emigration in order to calibrate the migration hump function in Equation (2). For all other purposes, refugee flows are not explicitly excluded.

2.5 Performance measures

Here, we discuss two different performance measures used to evaluate our model. The first is the commonly used coefficient of determination, R^2, which can be defined as

$$R^2 = 1 - \frac{\sum_{i,t} \left(y_{i,t} - \hat{y}_{i,t}\right)^2}{\sum_{i,t} \left(y_{i,t} - \bar{y}\right)^2}, \tag{17}$$

with $\hat{y}_{i,t}$ being some estimated value at time t, $y_{i,t}$ indicating the observed value, $\bar{y}$ the mean of all observations, and i is some index depending on the dataset. The second measure is the mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n} \sum_{i,t} \left| \frac{y_{i,t} - \hat{y}_{i,t}}{y_{i,t}} \right|, \tag{18}$$

with $\hat{y}_{i,t}$ and $y_{i,t}$ being defined as for R^2.

We opted for the R^2 metric as a means to evaluate the model's efficacy concerning absolute migration flows. This choice is of particular interest as it emphasizes the absolute discrepancies, thereby harmonizing with the principles of the least squares calibration method. Additionally, to assess the relative variances between the flows, irrespective of their scale, we incorporated the MAPE as an auxiliary measure. To address the division-by-zero problem when using $y_{i,t} = 0$, we have implemented an adjustment by adding 1 to both the observed and predicted migration flows. This adjustment is grounded in the rationale that 1 represents the minimal discrete value that can be considered a valid number for migration flows. Moreover, given that migration data predominantly consists of zeros or values substantially greater than 1, this modification minimally impacts the overall metric. Using both these measures ensures a comprehensive analysis, capturing both the magnitude and proportionate differences in migration flows.

3. Results

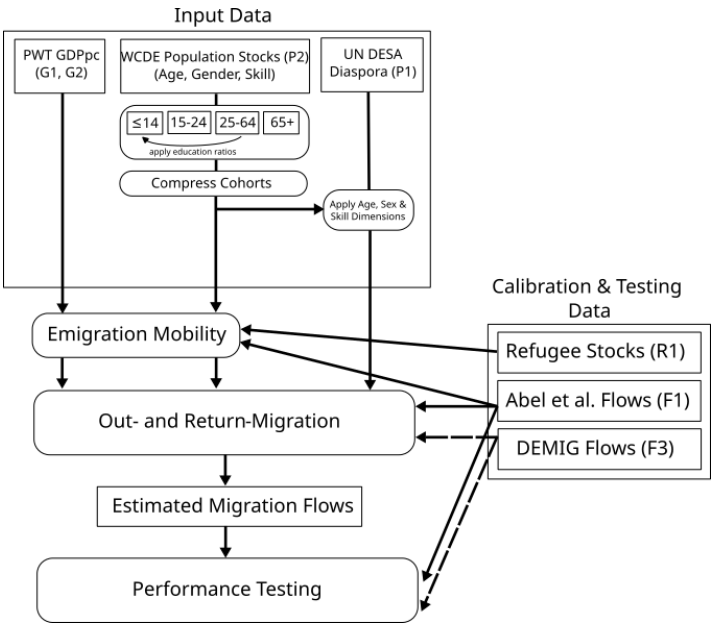
Our Results section consists of three subsections. We first discuss the calibration results and the performance of the migration model when compared to past migration data. Afterward, we analyze the effects of the model extensions, followed by a discussion of the future migration projections.

3.1 Calibration and goodness of fit

We calibrate the model in two steps, following Rikani and Schewe (2021). A schematic procedure of the calibration process is displayed in Figure 3. After preprocessing, as discussed in the Data section, we estimate the parameters of the migration hump function $F(G_i)$ by fitting Equation (2) to emigration flows (see Figure 3 "Emigration Mobility"). These emigration flows are obtained by aggregating the bilateral flows from F1 over destination countries. Moreover, we subtract refugee flows where data are available

in R1 because Equation (2) is meant to apply to only non-refugee migration. We obtain $\hat{G} = 36800(\pm 12200)$, $\tilde{G} = 923(\pm 52)$, and $\gamma = 0.00166(\pm 0.0011)$.

Figure 3: Scheme of the calibration process



Notes: Sharp corners indicate datasets, round corners indicate processes. Initial data sources are indicated in brackets.

Next, we calibrate the remaining model parameters by fitting the overall model to three different datasets, all containing bilateral flow data (see Figure 3 "Out- and Return-Migration"). The first set contains F1, being a set of global flow estimates that include out-migration as well as return-migration estimates. The second dataset is derived from F3 and includes only out-migration. The third dataset contains the migration numbers from F1 but uses only the migration corridors that are included in F3, which we refer to as $\text{F1}|_{\text{F3}}$. This selective approach allows for a focused analysis, specifically comparing data differences between F1 and F3. By doing so, we effectively exclude migration dynamics absent in F3, attributed to the missing migration corridors.

We calibrate all five different model approaches introduced in the Methods and Data section, including the original model introduced in Rikani and Schewe (2021), given in Equation (1); the separate education, sex, and age approaches, as described in Equations (7)–(11); and finally the full approach, combining all the previously mentioned

dimensions (Equation 5). We include all single-dimension approaches to see whether each of the additions is beneficial compared to the original model and how the coefficients between the several approaches differ.

We do not estimate all six age-dependent parameters σ_a and κ_a simultaneously, but instead we estimate only one parameter σ and κ for the cohort of 0 to 24 years. Simultaneously, we apply the average emigration rates per age group extracted from F2 to those two coefficients. We find that the mean emigration rates per age group are 0.0377 for 0 to 24 years, 0.0361 for 25 to 64 years, and 0.0188 for 65+ years. The reason we can not estimate all three parameters at the same time using F1 is explained in the Appendix).

All models are calibrated using the least squares estimation algorithm from the Python SciPy library (Pauli et al. 2020). The calibrated parameters and their confidence intervals for the models using F1 are presented in Table 1. For models calibrated with F3, these details are displayed in Table 2, and for the $\text{F1}|_{\text{F3}}$ corridor models, they can be found in Table 3. Table 4 displays performance values for all models across the different datasets. Notably, the parameters from the F3 calibration exclude any for κ, as this dataset does not encompass return-migration numbers.

Beyond the five model approaches already introduced, we further tested the performance of the original model with only a single modification: substituting the destination GDP driver with a destination-to-origin GDP ratio. This variation aims to ascertain whether this specific alteration enhances the overall effectiveness of the model despite adding further demographic dimensions. The performance values are displayed in Table 4 as well.

Table 1: Parameter estimates and 99% confidence intervals of different model approaches using F1

Parameter	Original (Equation (1))	Education (Equation (7))	Gender (Equation (9))	Age (Equation (11))	Full (Equation (5))
σ_1	0.198 ± 0.003	0.249 ± 0.004	0.187 ± 0.003	0.199 ± 0.003	0.266 ± 0.005
σ_2				0.191 ± 0.003	0.255 ± 0.005
σ_3				0.099 ± 0.003	0.133 ± 0.005
$\alpha_{p,h}$	0.914 ± 0.003	1.148 ± 0.009	0.961 ± 0.003	0.958 ± 0.003	1.159 ± 0.009
$\alpha_{p,l}$		0.903 ± 0.003			0.897 ± 0.003
$\alpha_{g,m}$	0.149 ± 0.007	0.199 ± 0.007	0.329 ± 0.008	0.271 ± 0.007	0.262 ± 0.007
$\alpha_{g,f}$			0.235 ± 0.008		0.119 ± 0.008
κ_1	0.125 ± 0.001	0.125 ± 0.001	0.129 ± 0.001	0.132 ± 0.001	0.132 ± 0.001
κ_2				0.127 ± 0.001	0.127 ± 0.001
κ_3				0.066 ± 0.001	0.066 ± 0.001

Table 2: Parameter estimates and 99% confidence intervals of different model approaches using F3

Parameter	Original (Equation (1))	Education (Equation (7))	Gender (Equation (9))	Age (Equation (11))	Full (Equation (5))
σ_1				0.130 ± 0.014	0.340 ± 0.037
σ_2	0.536 ± 0.147	0.313 ± 0.033	0.121 ± 0.013	0.125 ± 0.014	0.326 ± 0.037
σ_3				0.065 ± 0.014	0.170 ± 0.037
$\alpha_{p,h}$	0.829 ± 0.022	0.874 ± 0.029	0.790 ± 0.028	0.781 ± 0.028	0.835 ± 0.029
$\alpha_{p,l}$		0.780 ± 0.021			0.776 ± 0.021
$\alpha_{g,m}$	-1.1840 ± 0.186	-0.325 ± 0.067	-0.195 ± 0.084	-0.203 ± 0.080	-0.311 ± 0.071
$\alpha_{g,f}$			-0.148 ± 0.082		-0.374 ± 0.071

Table 3: Parameter estimates and 99% confidence intervals of different model approaches using F1 numbers but including only the migration corridors available in F3

Parameter	Original (Equation (1))	Education (Equation (7))	Gender (Equation (9))	Age (Equation (11))	Full (Equation (5))
σ_1				0.306 ± 0.011	0.401 ± 0.019
σ_2	0.09 ± 0.025	0.374 ± 0.018	0.293 ± 0.010	0.293 ± 0.011	0.384 ± 0.011
σ_3				0.153 ± 0.011	0.200 ± 0.011
$\alpha_{p,h}$	0.865 ± 0.008	1.068 ± 0.034	0.879 ± 0.011	0.872 ± 0.010	1.061 ± 0.033
$\alpha_{p,l}$		0.758 ± 0.011			0.744 ± 0.011
$\alpha_{g,m}$	0.887 ± 0.204	-0.122 ± 0.032	0.055 ± 0.029	0.034 ± 0.028	-0.132 ± 0.032
$\alpha_{g,f}$			0.041 ± 0.029		-0.182 ± 0.033
κ_1				0.175 ± 0.046	0.175 ± 0.042
κ_2	0.160 ± 0.043	0.160 ± 0.039	0.160 ± 0.043	0.168 ± 0.046	0.168 ± 0.046
κ_3				0.087 ± 0.046	0.087 ± 0.046

Table 4: Performance of different model variations evaluated using different flow datasets/variations

Dataset	Measure	Original (Equation (1))	GDPpc Ratio	Education (Equation (7))	Gender (Equation (9))	Age (Equation (11))	Full (Equation (5))
F1	R^2	0.6144	0.6321	0.6443	0.6351	0.6353	0.6531
F3	R^2	0.5225	0.5080	0.5528	0.5087	0.5103	0.5571
$F1|_{F3}$	R^2	0.9141	0.9295	0.9275	0.9125	0.9148	0.9303
F1	MAPE	0.44	0.33	0.35	0.33	0.33	0.36
F3	MAPE	3.18	1.89	1.72	1.88	1.97	1.75
$F1|_{F3}$	MAPE	0.75	0.98	1.68	0.97	1.01	1.84

Source: Bilateral flow data from F1 and F3.

In the calibration using the full F1 dataset, the model's performance, as indicated by the R^2 measure, improves when substituting the relative GDPpc approach with an origin-to-destination GDPpc ratio (Table 4, column "GDPpc Ratio"). Further enhancements are observed by individually incorporating age, sex, and skill dimensions, with the most significant improvement seen in the model integrating all three. The MAPE analysis reveals that incorporating gender, age, and GDPpc ratio yield the most accurate results, with all extended model versions significantly outperforming the original model.

Examining the model coefficients, the scaling parameters σ and κ maintain similar magnitudes across different approaches. All diaspora coefficients that are not skill-dependent have similar values and are slightly below one, therefore increasing the output, as the relative diaspora values are always smaller than one. In models with separate coefficients for high- and low-skill, the former exceeds one, while the latter falls below, indicating that relative diaspora output value is reduced for high-skill migrants and increased for low-skilled migrants. Moving on, gender-dependent coefficients indicate higher values for male migrants than females, making men move more from lower origin GDPpc countries to higher GDPpc countries, and women vice versa.

Utilizing the F3 dataset for calibration presents varying results. Scaling coefficients show greater variation and diaspora coefficients are marginally lower than those calibrated with the full F1 dataset. Models using education-dependent diaspora coefficients show the same behavior as for F1, where the high-skilled coefficient is larger than the low-skilled one. The approaches without education-dependent diaspora coefficients show similar magnitudes for the parameter as for the F1 calibration too. Contrary to the F1 dataset findings, GDPpc ratio coefficients are negative, but the male coefficient remains larger than the female. Calibrating with the F1 dataset but limiting to migration corridors present in the F3 dataset yields larger scaling coefficients for all models except the original. The GDPpc ratio coefficients are negative for the education and full approach, albeit less so compared to the F3 dataset. The GDP-ratio coefficients for the gender and age approach coefficients are found to be almost zero.

Assessing the R^2 performance with F3 and the $\text{F1}|_{\text{F3}}$ datasets reveals that the comprehensive model still delivers superior performance. However, unlike with the full F1 dataset, adding specific dimensions to the model reduces its efficacy compared to the original. Using the F3 dataset, both the GDPpc and gender approaches underperform relative to the original model. For $\text{F1}|_{\text{F3}}$, only the gender approach lags behind.

MAPE evaluations for F3 indicate that all model extensions show improvements over the original model with education and full approach yielding the best results. While $\text{F1}|_{\text{F3}}$ shows smaller percentage errors, here the original model performs best, while the education and full model perform worse than the other approaches. Comparing $\text{F1}|_{\text{F3}}$ with F3, we observe that the MAPE values for the full model are of a similar magnitude, while the R^2 values exhibit significant differences. The relative deviations between modeled and observed flows show similar distributions across the entire dataset, therefore yielding similar overall MAPE measures. However, when selecting the largest 1% of flows of each of the datasets – which account for more than 50% of the total flow volume – the full model demonstrates significantly lower errors with $\text{F1}|_{\text{F3}}$ (mean deviation of 38%) compared to F3 (mean deviation of 62%). Consequently, the model is more effective at estimating large flows in $\text{F1}|_{\text{F3}}$ but less so for those in F3. Therefore the total deviation is smaller for $\text{F1}|_{\text{F3}}$, yielding better overall R^2 measures.

The improved performance metrics observed for datasets using F1 data, compared to those for F3, may be attributed in part to the fact that P1 was used for estimating the F1 data and at the same time also informs our relative diaspora calculations. Despite incorporating various factors and combining P1 with P2 in our relative diaspora assessments, it is important to acknowledge this potential correlation when evaluating the datasets.

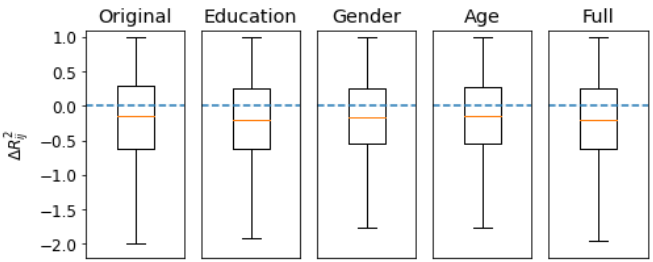
Because the spatial variation is much larger than the temporal variation, these model fits primarily reflect spatial patterns. We additionally test the models' ability to explain variations over time, by replacing y_i in Equation (17) with

$$\frac{M_{ijt} - M_{ij(t-1)}}{M_{ij(t-1)}} \tag{19}$$

and calling the resulting measure ΔR^2_{ij}. That is, we measure the models' R^2 in terms of the variation in bilateral migration flow from one period to the next. Due to the limited amount of time series data in F3, we perform this evaluation for only F1. Neither the original model nor the extended models are able to explain temporal changes found in individual bilateral flow time series, as indicated by below-zero ΔR^2_{ij} for a majority of flows (Figure 4). This is the case for 'gravity'-type migration models in general (Beyer, Schewe, and Lotze-Campen 2022), and continues to be an important caveat with the models presented here, as well as other recent model developments (Qi and Bircan 2023).

Projections from such models are informed primarily by spatial patterns in the historical data and must not be interpreted as predictions.

Figure 4: Model performance in terms of temporal changes in bilateral migration flows



Source: Data to test the performance from F1.
Notes: Dashed blue horizontal lines indicate zero, orange lines the median value of R^2_{ij}, boxes the interquartile range, and whiskers the 98.5th and the 1.5th percentile.

3.2 Effect of model extensions

Both the original and the extended models show a similar global pattern of deviations from the bilateral flow estimates F1 for the most recent period, with the full model overestimating emigration and immigration in some countries where the original model underestimates them (Figures 5a–d).

When we examine Figures 5e and 5f, which illustrate the comparative accuracy of our original and full models against the F1 estimates, it becomes evident that the original model excels in predicting emigration numbers for the majority of countries. In contrast, the full model demonstrates superior accuracy in estimating immigration flows. However, it's important to recognize that these figures indicate only which model is more closely aligned with the estimates and do not precisely convey the extent of deviation from the actual figures. For instance, the F1 estimates place Canada's emigrant count at around 1 million. The original model's prediction is approximately 985,000, while the full model suggests about 915,000, indicating that both models yield relatively accurate estimates. To account for this shortcoming, we include Figures 5g and 5h, which display the total difference between full and original model. We note that this comparison relates to a single time period only.

Furthermore, the recent enhancements to our model permit a nuanced analysis of migration projections with an unprecedented level of demographic granularity. It enables the examination of migration rates differentiated by sex, skill level, and age. These de-

mographic dimensions are not only pivotal in understanding migration flows but are also incorporated into the projection model, including, for example, the use of age-specific mortality rates and educational transitions. Consequently, this advancement allows us to explore migration flows with a new degree of complexity, offering insights into the multifaceted interplay between demographic variables and migration patterns.

The addition of age, sex, and education dimensions to our model permits a more granular analysis of global migration patterns than was possible with previous projection models. Indeed, projected emigration and immigration 'rates'[7] differ markedly between male and female, high- and low-skilled, and age groups (Figures A-1–A-8). Such differences, especially between age groups, are also visible in the F2 data, highlighting the importance of including these distinctions in models.

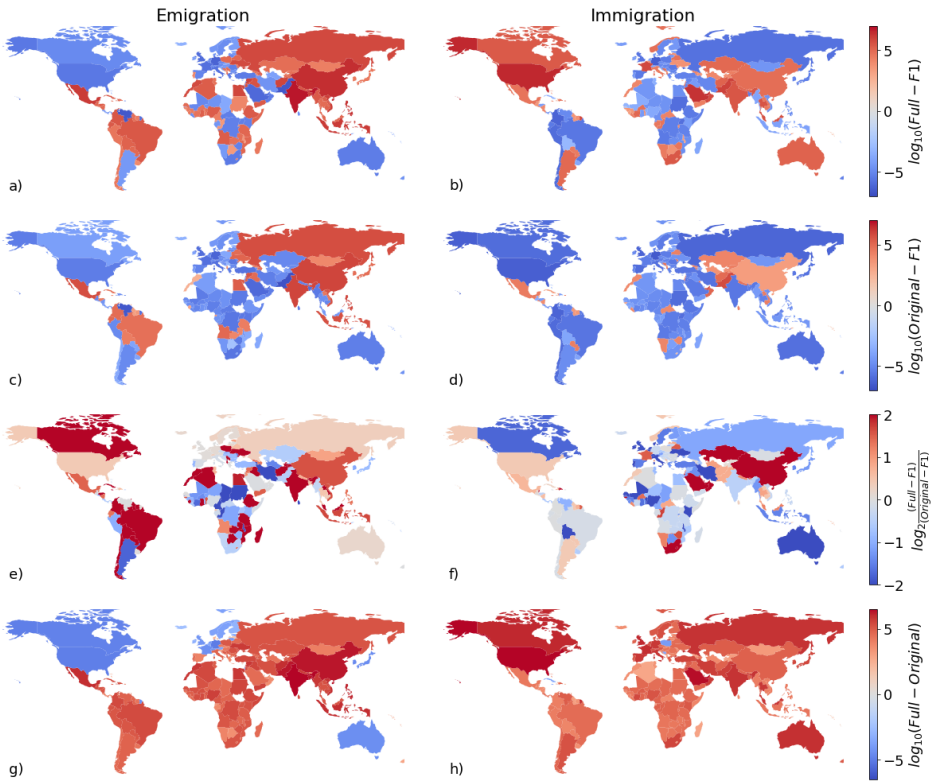
In assessing the accuracy of our model estimates (Figures A-1–A-8) against empirical data (F1 and F2) spanning the 2015–2019 period, notable trends emerge across global regions. Our model captures the observed mobility patterns with respect to age, particularly highlighting the heightened mobility of young and working-age cohorts. However, nuanced variations are evident within different regions. Notably, discrepancies arise in the immigration 'rates' for the 65+ years cohort in Latin America and the Caribbean and the Middle East and North Africa, where the model estimates the older migrants to be less mobile than the other two cohorts while the observed data shows them to be more mobile in these regions. In terms of the total magnitude of the migration rates, the model generally indicates higher migration rates than F2. Exceptions are emigration from sub-Saharan Africa and immigration to Latin America and Caribbean, South Asia, and sub-Saharan Africa, where the model underestimates the rates observed in the data. Examining gender-specific trends, our model consistently predicts higher emigration and immigration 'rates' for men than for women across all regions, agreeing with the majority of empirical data points. This pattern does not fit emigration and immigration 'rates' in East Asia and Pacific, as well as emigration rates in East Europe and Central Asia and Middle East and North Africa, where the model estimates higher male than female migration rates, but the observed data shows the opposite. Despite maintaining the overarching trend of greater male mobility, our model tends to overestimate gender differences and overestimate the total flow rate.

After examining the emigration and immigration figures produced by both models, we shift our focus to the estimated migration flows, categorizing them according to income groups as defined by the World Bank. Detailed data for these flow charts are available in the Appendix (Tables A-2–A-5). Initially, when analyzing the aggregate numbers, it is evident that the full model is more proficient in estimating the total volume of migration flows. Moreover, in terms of migration to and from low, lower-middle, and upper-middle income countries, the full model's estimates are generally more aligned

[7] Immigration rates are not proper demographic rates as their denominators (populations of destination countries) are not correct populations at risk of migrating.

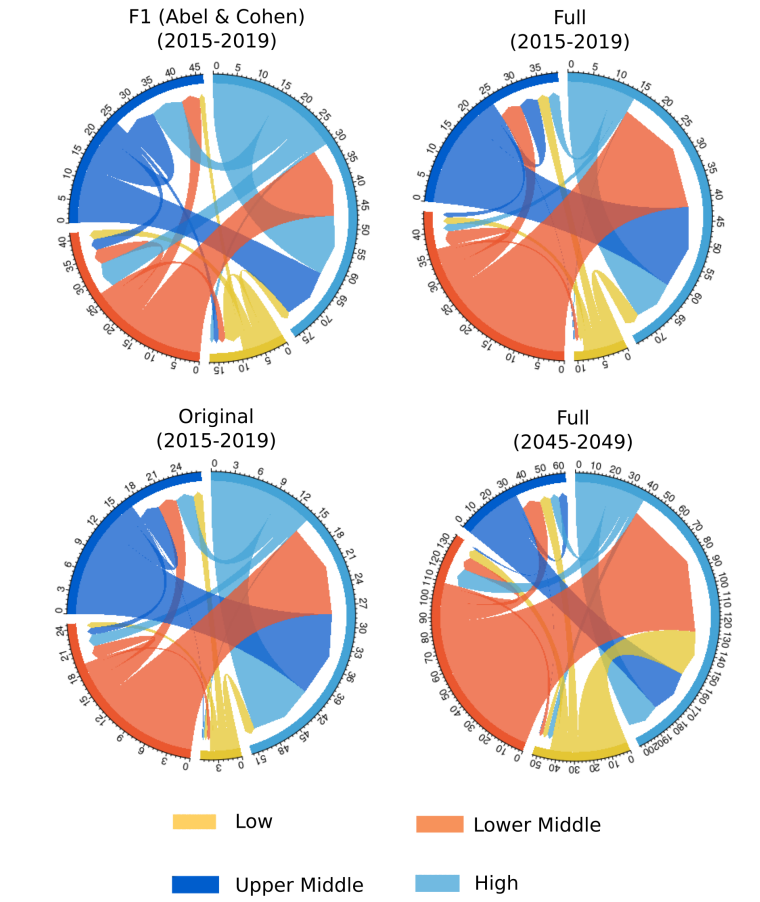
with the F1 benchmark. However, the model tends to overestimate high-income countries as destinations and slightly underestimate them as origin regions. Additionally, we have incorporated results from the full model for the period 2045–2049 to examine temporal dynamics. This analysis reveals a notable expansion in the migration corridor from lower and lower-middle-income to high-income countries during this timeframe.

Figure 5: Comparison of emigration and immigration flows between age and education model, original model, and F1, for the 2015–2019 period



Notes: Absolute difference (logarithmic) between the full model and F1 (panels a and b) and between the original model and F1 (panels c and d). To estimate these values, we first calculate the difference, take the absolute value, apply the logarithm, and then reapply the sign. Panels e and f: ratio of the difference between the original model and F1 compared to the difference between the full model and F1. Blue indicates where the full model is closer to the estimates of F1; red indicates where the original model is closer. Panels g and h display the difference between the full and the original model.

Figure 6: Flows of the original and full model and F1 for the 2015–2019 period and of the full model for 2045–2049



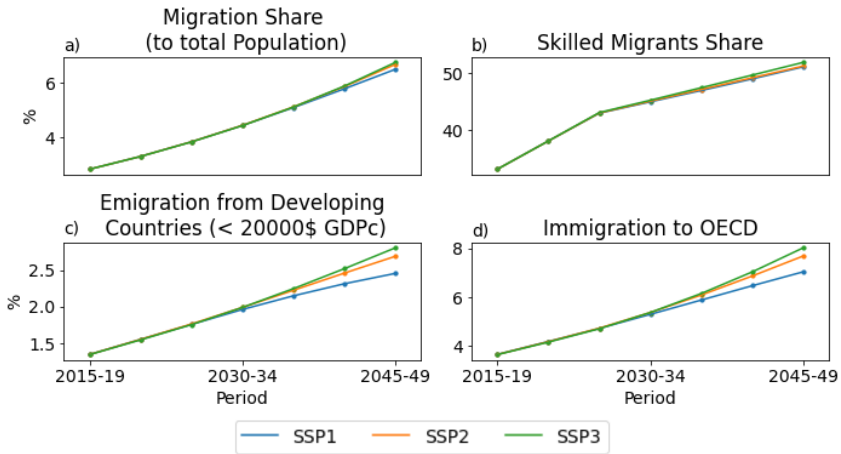
Notes: Flows are aggregated by World Bank income level and are displayed for the time steps 2015–2019 and 2045–2049, respectively. Flow numbers are given in millions, and the total flow numbers are displayed in Tables A-2–A-5.

3.3 Future migration projections

We project migration flows globally for 2015–2049 under three different scenarios – SSP1, SSP2, and SSP3 – using the full model, coupled to the population projection model.

The global share of migrants is projected to increase from about 3% to about 6% of the global population under all three scenarios (Figure 7a). Note, however, that since the projected total population differs between the scenarios, total migrant stocks differ more in absolute numbers than in percentage terms. The share of skilled migrants in total migration flows is also projected to increase in a very similar way in all scenarios (Figure 7b). This is because the same education transition rates are applied in all scenarios. We note that in Figures A-1–A-8 the low-skilled migration rates are higher; therefore, we conclude that the high-skilled population share is rising drastically. Larger differences between the SSPs are found for the rate of emigration from developing countries and the rate of immigration to OECD countries (Figures 7c and 7d): These are projected to increase most strongly in SSP3 (the most economically unequal scenario), while they stabilize around mid century in SSP1 (the inclusive development scenario).

Figure 7: Projection results of the full model for SSP1, SSP2, and SSP3 on a global level



Note: Panel a displays global migration share (migrant stocks divided by total population), panel b the share of high-skilled migrants (high-skilled migration flows to total migration flows), panel c the percentage of emigrants of the total population leaving countries with a GDPpc $<$ \$20,000, and panel d the percentage of immigrants of total population arriving in OECD countries obtained from the full model. Dots indicate the value of a given time period. SSP1: emphasis on environmental protection, lower resource intensity, and reduced inequality; SSP2: continuation of current trends; SSP3: more pronounced inequality and slower technological advancement.

Regional migration flows can diverge substantially more between the three scenarios, and they do so more than in the original model (Figures 8–11). In particular, the extended model projects a larger spread in emigration flows in many net sending regions

and in immigration flows in net receiving regions. This results in a larger spread in net migration between the scenarios in all regions.

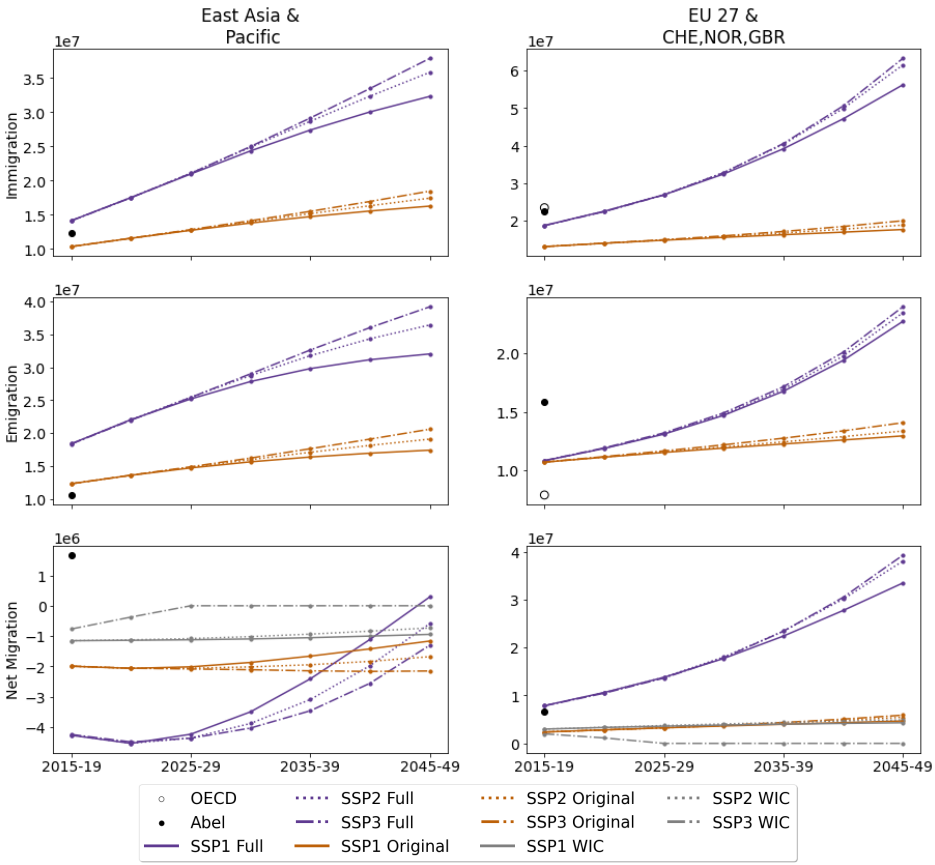
For instance, the extended model projects net migration in the Middle East and North Africa region in 2045 to be twice as large under SSP3 as under SSP1, whereas the original model projects significantly less differences between the scenarios. This increased sensitivity to the scenario assumptions – which include economic and demographic development – is likely because the extended model considers the destination–origin GDPpc ratio, while the original model considered only the destination GDPpc (relative to global average GDPpc) as an economic pull factor. Lower-income economies generally grow faster than high-income economies, and their growth rates differ more between the SSPs. This means that cross-country inequalities tend to differ between the SSPs more than the relative destination GDPpc.

The migration flows projected by the extended model for the period 2015–2019 are closer to the F1 estimates for that period (filled circles in Figures 8–10) for about half the regions, while the original model is closer for the other half. This again reflects the small and spatially heterogeneous improvement in model fit discussed earlier.

More surprisingly, our migration projections differ qualitatively from the stylized migration assumptions previously employed in the WIC population projection model. There, SSP3 was assumed to exhibit the lowest migration rates (in fact transitioning to zero net migration by 2030) because of the "emphasis on security and barriers to international exchange" in the SSP3 narrative (KC and Lutz 2017); while medium migration levels were assumed in SSP1 and SSP2. In contrast, we find the largest migration flows in SSP3 due to the high levels of economic inequality in this SSP. Moreover, even net migration in SSP1, while lower than in the other SSPs according to our projections, is still higher than in any of the SSPs in the WIC stylized assumptions. That is, our model depicts a considerably more mobile world than those assumptions, and one where international migration is strongly shaped by economic inequalities. In general, it can be observed that the full model estimates significantly higher migrant numbers for the future compared to both the original and WIC data.

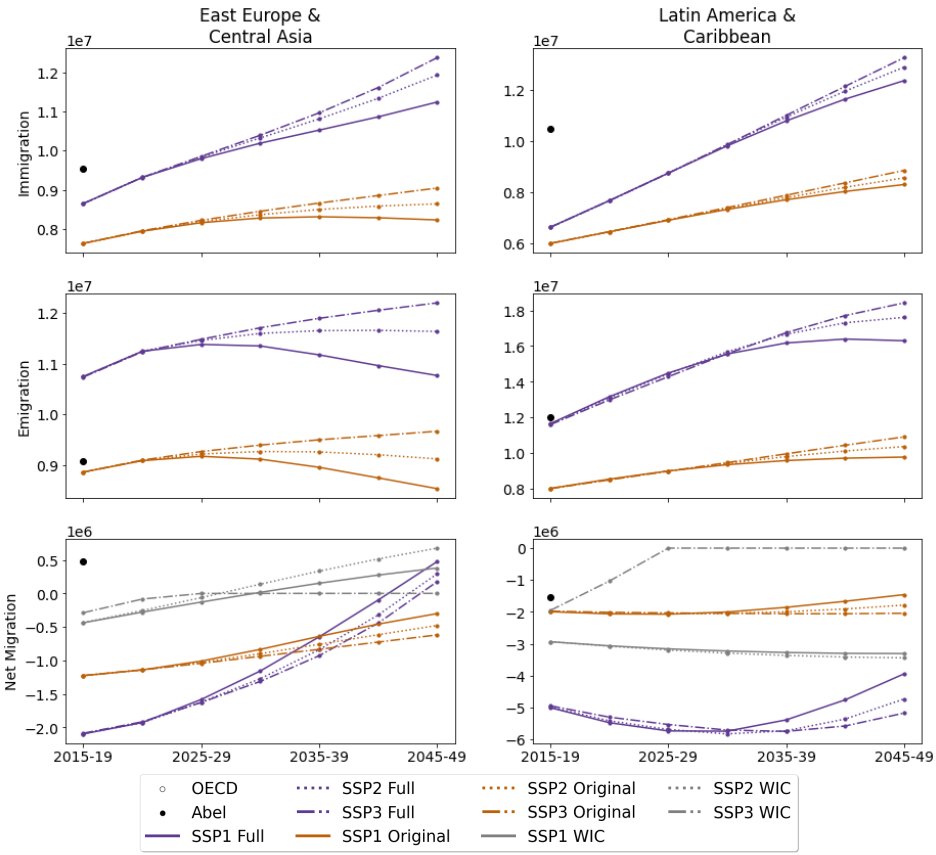
These differences highlight the effect of accounting explicitly for drivers of migration, such as GDPpc differentials, which were ignored in the stylized assumptions. On the other hand, our model does not consider any effects of immigration policies and control measures, and how these might change in the future. While some studies suggest that the effects of immigration policies and controls on flows are more limited than anticipated by policymakers (Schon and Leblang 2021) and are in turn dependent on economic conditions and the size of bilateral migrant stocks (Helbling and Leblang 2019), they are nevertheless important, and accounting for them might further improve projection models like the one presented here. In fact, since our model already includes economic variables and bilateral migrant stocks, it could readily account for the interactions of these variables with immigration policies.

Figure 8: Projected immigration, emigration, and net migration flows by world region (East Asia and Pacific and EU 27 and CHE, NOR, GBR)



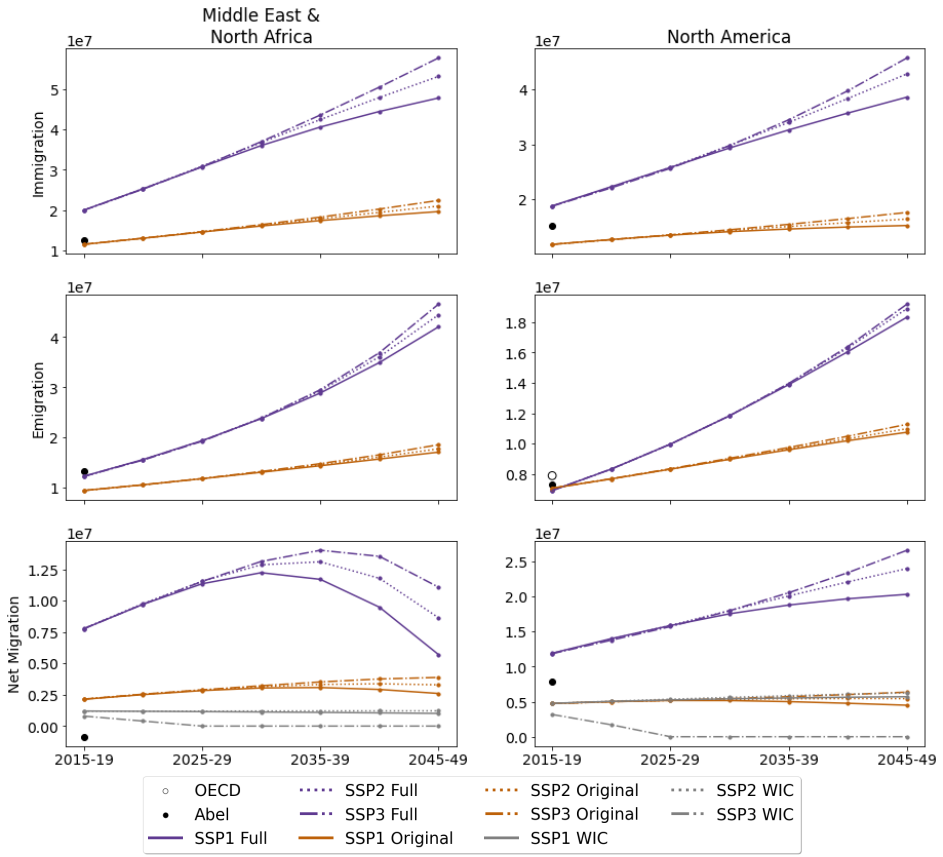
Notes: Five-year immigration (top), emigration (middle), and net migration (bottom) between 2015 and 2049 for the original model (orange) and the full model (purple). Bottom panels include net migration assumptions from the Wittgenstein Centre population model (Wittgenstein Centre for Demography and Global Human Capital 2018) for comparison (gray). Filled black circles indicate flow estimates from F1 for the 2015–2019 period, open circles indicate immigration data from (OECD Stat 2022) (excluding return-migration). See Table A-6 for region definitions. Regional figures include only flows between regions, not flows between countries within a region. SSP1: emphasis on environmental protection, lower resource intensity, and reduced inequality; SSP2: continuation of current trends; SSP3: more pronounced inequality and slower technological advancement.

Figure 9: Projected immigration, emigration, and net migration flows by world region (East Europe and Central Asia and Latin America and Caribbean)



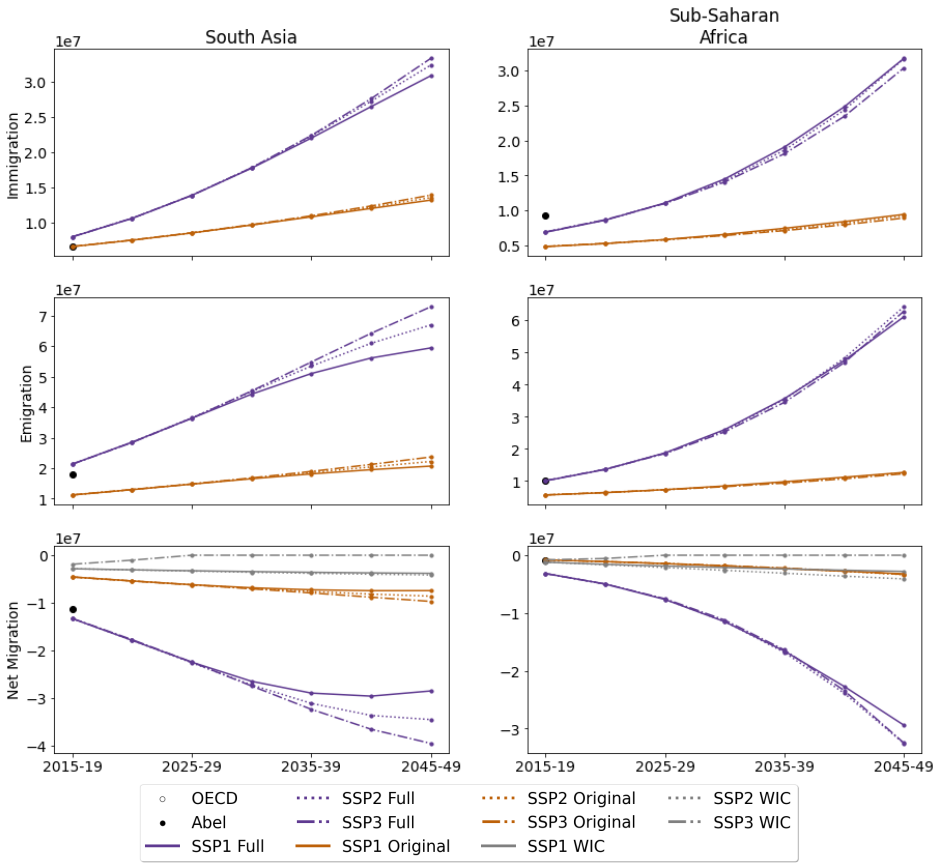
Notes: Five-year immigration (top), emigration (middle), and net migration (bottom) between 2015 and 2049 for the original model (orange) and the full model (purple). Bottom panels include net migration assumptions from the Wittgenstein Centre population model (Wittgenstein Centre for Demography and Global Human Capital 2018) for comparison (gray). Filled black circles indicate flow estimates from F1 for the 2015–2019 period, open circles indicate immigration data from (OECD Stat 2022) (excluding return-migration). See Table A-6 for region definitions. Regional figures include only flows between regions, not flows between countries within a region. SSP1: emphasis on environmental protection, lower resource intensity, and reduced inequality; SSP2: continuation of current trends; SSP3: more pronounced inequality and slower technological advancement.

Figure 10: Projected immigration, emigration, and net migration flows by world region (Middle East and North Africa and North America)



Notes: Five-year immigration (top), emigration (middle), and net migration (bottom) between 2015 and 2049 for the original model (orange) and the full model (purple). Bottom panels include net migration assumptions from the Wittgenstein Centre population model (Wittgenstein Centre for Demography and Global Human Capital 2018) for comparison (gray). Filled black circles indicate flow estimates from F1 for the 2015–2019 period, open circles indicate immigration data from (OECD Stat 2022) (excluding return-migration). See Table A-6 for region definitions. Regional figures include only flows between regions, not flows between countries within a region. SSP1: emphasis on environmental protection, lower resource intensity, and reduced inequality; SSP2: continuation of current trends; SSP3: more pronounced inequality and slower technological advancement.

Figure 11: Projected immigration, emigration, and flows by world region (South Asia and sub-Saharan Africa)



Notes: Five-year immigration (top), emigration (middle), and net migration (bottom) between 2015 and 2049 for the original model (orange) and the full model (purple). Bottom panels include net migration assumptions from the Wittgenstein Centre population model (Wittgenstein Centre for Demography and Global Human Capital 2018) for comparison (gray). Filled black circles indicate flow estimates from F1 for the 2015–2019 period, open circles indicate immigration data from (OECD Stat 2022) (excluding return-migration). See Table A-6 for region definitions. Regional figures include only flows between regions, not flows between countries within a region. SSP1: emphasis on environmental protection, lower resource intensity, and reduced inequality; SSP2: continuation of current trends; SSP3: more pronounced inequality and slower technological advancement.

Abel (2018) develops a set of net migration assumptions along the SSPs by assuming constant present-day net migration in SSP2 and scaling net migration in the other SSPs according to the deviations in the global distribution of wealth – measured by GDP – rel-

ative to SSP2. This can be considered an intermediate approach between the stylized assumptions in Lutz et al. (2018) and our calibrated migration model, taking the role of changing economic inequalities into account while still covering a deliberately large range of outcomes. In fact, for 2050 their scenarios cover both positive and negative net migration in Europe and Africa and include near-zero net migration in Asia and North America. The ranges are narrower in our projections, but interestingly, they often coincide approximately with the higher-end scenarios of Abel (2018), perhaps corroborating the order of magnitude of plausible migration trends, even though our full model still estimates higher net migration numbers for regions like Europe or North America.

4. Discussion and conclusion

This study aimed to develop a global migration model including age, education, and gender dimensions for coupling population projection models to estimate future migration and population scenarios. We extended a recently developed global migration model by adding mobility profiles for three different age groups, as well as education- and gender-specific dynamics, allowing us to differentiate between male and female as well as high- and low-skilled migrants. Furthermore, we coupled our migration model with the FUME developed by IIASA, allowing us to estimate future migration projections in a demographically detailed manner while only using the GDP projections as a scenario-dependent driver.

The model is calibrated to global bilateral flow estimates, which are subject to uncertainties related to the migrant stock data they are based on and to the method used to derive flows from stock data. To review this problem, we also tested and calibrated the model on a smaller subset of migration flow records. The estimated model parameters when using the global dataset are consistent with previous estimates, (e.g., a positive coefficient on the wage ratio when treating the GDPpc ratio as a skill-independent wage ratio) (Dao et al. 2021; Bertoli and Moraga 2013). Contrary to the global dataset, we obtain negative coefficients for the GDPpc ratio when calibrating on the subset of estimated migration flows and actual migration records, including only a limited range of migration corridors. An explanation for this phenomenon could be the restricted scope of destination countries in the dataset. The majority of these corridors are within the EU/Schengen area. Given that these nations are relatively wealthy, GDPpc ratios within this subset may be smaller compared to those on a global level. This disparity in GDPpc ratios is likely a contributing factor. Additionally, the prevalence of intra-European migration within this dataset can significantly influence the model's dynamics. Such migration flows, predominantly occurring between countries with no visa restrictions, might lead to less pronounced economic pull factors. Consequently, this could result in a shift in the

coefficient sign, reflecting the unique economic and migratory landscape of the EU as compared to more diverse global migration patterns.

Moreover, our results, which indicate higher migration rates among the younger and intermediate age groups compared to older populations, are in line with previous studies (Raymer, De Beer, and Van der Erf 2011; Raymer and Wiśniowski 2018; Wiśniowski et al. 2016). The observed marginally higher mobility in the younger cohort compared to the working-age cohort, as evidenced by the coefficients presented in Table 1, can be attributed to our cohort delineation. To generate similarly sized groups, we divide age groups at age 25 and 65. This is congruent with age categorizations utilized by institutions such as the United Nations and the Wittgenstein Centre. Age mobility estimates (Wiśniowski et al. 2016) indicate that mobility peaks at ages 20 to 30. Consequently, our classification incorporates one segment of the most mobile population within the youngest (<25 years) cohort, while another segment is incorporated into the working-age (25 to 65 years) cohort. This, together with the small number of age groups, blurs the peak at age 20 to 30, but our results are still consistent with existing literature and datasets (F3 and Wiśniowski et al. 2016). Additionally, finding that diasporas tend to attract more lower-skilled migrants (Beine, Docquier, and Özden 2010; Kracke and Klug 2021) and that women are more likely to emigrate from countries with higher GDPpc (Neumayer and Plümper 2021; Naveed et al. 2023) also aligns with existing research in the field.

The inclusion of a GDPpc destination–origin ratio term has made the model more sensitive to assumptions about national economic development compared to the original model that considered only destination GDPpc. This results in a wider spread of migration projections between different SSPs. Whether this is indeed more realistic is difficult to ascertain empirically given that the extended model fits the (uncertain) past data only slightly better than the original model; however, the GDPpc ratio term does account for the current understanding of economic pull factors better than a destination-only term and is often used in empirical migration models to reflect these factors (e.g., Docquier 2018; Burzyński et al. 2022; Shayegh, Emmerling, and Tavoni 2022).

While the incorporation of the GDPpc ratio has improved the model's overall effectiveness, future enhancements could benefit from incorporating a skill-based dimension. Specifically, considering wages across different skill levels could provide valuable insights. However, a significant challenge lies in the lack of harmonized global data on wage variations by skill level. Our attempts to estimate such wages did not yield notable improvements in the model's performance. This highlights an area for potential future research, contingent on the availability of more comprehensive and standardized data.

Our model is designed as a potential alternative to more ad hoc migration assumptions in population projection models. When coupled to the FUME, it results in markedly different migration flows compared to the migration assumptions previously used in that model. We find generally larger emigration, immigration, and net migration flows when using our model. Moreover, we project the largest net migration flows under the SSP3

scenario, which previously was assumed to have the smallest net migration flows. This is because of the high economic inequality in that scenario, which in our model results in high levels of migration but was not considered in the stylized migration assumptions for SSP3. Those assumptions, in turn, did consider political barriers to migration, which are missing from our model. Nevertheless, the comparison highlights the importance of accounting for economic drivers of migration in population projections. By using a mechanistic model instead of stylized assumptions, both the factors included and factors omitted can be made explicit.

Migration in our model is a function of economic and demographic drivers. Coupling with a state-of-the-art cohort-component population projection model improves the representation of the demographic drivers, such as natural population change. While not considered in this study, scenario-specific assumptions about changes in fertility, mortality, and education transition and attainment rates can be included and explored in the coupled model (just as in the standard IIASA/WIC population model), in addition to changes in GDPpc levels (Potančoková, Stonawski, and Gailey 2021). This makes the coupled model a powerful tool for demographic scenario studies, resolving migration as well as other demographic processes at a relatively high level of detail and complexity.

Our extended migration model is subject to some of the same limitations as the original model it is based on, and indeed other global migration models. It is not able to reproduce the temporal dynamics of migration flows, instead relying primarily on spatial variations to estimate parameters that are then used to project into the future (Beyer, Schewe, and Lotze-Campen 2022). This remains a fundamental problem and might not easily be overcome given the limited amount and quality of existing migration data and the complexity of migration processes, which means that migration projections should be interpreted with caution. Furthermore, we wish to emphasize that this paper, along with Rikani and Schewe (2021), relies on migration data derived using the methods from Azose and Raftery (2019). These methods, which are based on the analysis of internal migration flows, consequently leave room for improvement. Another fundamental limitation is the omission of forced migration and more generally migration related to non economic factors. Refugee flows can contribute substantially to changes in migrant stocks, thereby also influencing more regular migration flows. However, accounting for these flows would require a very different model that accounts for hard-to-predict factors, such as conflict risk (Fransen and Haas 2022).

The factors that are included in the model are represented in a simplistic fashion such as to make the model tractable and quantifiable given the available data. As such, the model could be improved in many respects. For instance, it would be desirable to account for differences in attainable wages between migrants and non-migrants (Chiswick 2018; Adsera and Chiswick 2007), or between men and women (Carling 2005; Boyd and Grieco 2003). Ultimately, the model presented here should improve global-scale population projections by making their migration components more process-based, transpar-

ent, and detailed. A lot of challenges remain when it comes to better understanding the complex interactions between social, economic, political, and environmental factors and migration, and not all of these challenges can be addressed with mechanistic models. Furthermore, the model should also be tested across a broader range of scenarios, including those with more severe climate impacts than those assumed by the current SSP scenarios.

5. Acknowledgments

The work was supported within the framework of the European Union Horizon 2020 programme (grant agreement numbers 870649 (FUME) and 869395 (HABITABLE)) and by the BMBF (Förderkennzeichen 01LS2001A).

References

Abel, G.J. (2018). Non-zero trajectories for long-run net migration assumptions in global population projection models. *Demographic Research* 38(54): 1635–1662. doi:10.4054/DemRes.2018.38.54.

Abel, G.J. and Cohen, J.E. (2022). Bilateral international migration flow estimates updated and refined by sex. *Scientific Data* 9(173). doi:10.1038/s41597-022-01271-z.

Adsera, A. and Chiswick, B.R. (2007). Are there gender and country of origin differences in immigrant labor market outcomes across European destinations? *Journal of Population Economics* 20(3): 495–526. doi:10.1007/s00148-006-0082-y.

Akgüç, M. and Parasnis, J. (2023). Occupation–education mismatch of immigrant women in Europe. *Social Indicators Research* 170: 75–98. doi:10.1007/s11205-023-03066-0.

Arif, I. (2022). Educational attainment, corruption, and migration: An empirical analysis from a gravity model. *Economic Modelling* 110: 105802. doi:10.1016/j.econmod.2022.105802.

Azose, J.J. and 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. doi:10.1073/pnas.1722334116.

Battistella, G. (2018). Return migration: A conceptual and policy framework. New York: Center for Migration Studies (Migration Policy Report Perspectives on the Content and Implementation of the Global Compact for Safe, Orderly and Regular Migration): 3–14. https://cmsny.org/publications/scalabrini-policy-report-2018.

Beine, M. (2016). The role of networks for migration flows: An update. *International Journal of Manpower* 37(7): 1154–1171. doi:10.1108/IJM-01-2016-0013.

Beine, M., Docquier, F., and Özden, Ç. (2010). Diaspora effects in international migration: Key questions and methodological issues. *Swiss Journal of Economics and Statistics* 146(4): 639–659. doi:10.1007/BF03399331.

Beine, M., Docquier, F., and Özden, Ç. (2011). Diasporas. *Journal of Development Economics* 95(1): 30–41. doi:10.1016/j.jdeveco.2009.11.004.

Bertoli, S. and Moraga, J.F.H. (2013). Multilateral resistance to migration. *Journal of Development Economics* 102: 79–100. doi:10.1016/j.jdeveco.2012.12.001.

Beyer, R.M., Schewe, J., and Lotze-Campen, H. (2022). Gravity models do not explain, and cannot predict, international migration dynamics. *Humanities and Social Sciences Communications* 9(1). doi:10.1057/s41599-022-01067-x.

Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view*, vol. 24. Berlin/Heidelberg: Springer Science and Business Media. doi:10.1007/978-90-481-8897-0.

Bijak, J. and Czaika, M. (2020). Assessing uncertain migration futures: A typology of the unknown. Southampton/Krems: University of Southampton and Danube University Krems (QuantMig Project Deliverable D1.1). https://www.quantmig.eu/res/files/QuantMig%20D1.1%20Uncertain%20Migration%20Futures%20V1.1%2030Jun2020.pdf.

Bijak, J., Kupiszewska, D., Kupiszewski, M., and Saczuk, K. (2005). Impact of international migration on population dynamics and labour force resources in Europe. Warsaw: Środkowoeuropejskie Forum Badań Migracyjnych. http://www.cefmr.pan.pl/docs/cefmr_wp_2005-01.pdf.

Black, R., Adger, W.N., Arnell, N.W., Dercon, S., Geddes, A., and Thomas, D. (2011). The effect of environmental change on human migration. *Global Environmental Change* 21: S3–S11. doi:10.1016/j.gloenvcha.2011.10.001.

Boyd, M. and Grieco, E. (2003). Women and migration: Incorporating gender into international migration theory. Washington, D.C.: Migration Information Source. https://www.migrationpolicy.org/article/women-and-migration-incorporating-gender-international-migration-theory.

Burzyński, M., Deuster, C., Docquier, F., and de Melo, J. (2022). Climate change, inequality, and human migration. *Journal of the European Economic Association* 20(3): 1145–1197. doi:10.1093/jeea/jvab054.

Carling, J. (2005). Gender dimensions of international migration. *Global Migration Perspectives* 35(1): 1–26. https://www.iom.int/sites/g/files/tmzbdl486/files/2018-07/gmp35.pdf.

Cattaneo, C. (2007). The self-selection in the migration process: What can we learn? Castellanza: LIUC Università Cattaneo (LIUC Papers in Economics 52). https://www.researchgate.net/publication/5163890_The_Self-Selection_in_the_Migration_Process_What_Can_We_Learn.

Chiswick, B.R. (2018). The effect of Americanization on the earnings of foreign-born men. In: Suárez-Orozco, M., Suárez-Orozco, C., and Qin-Hilliard, D. (eds.). *Interdisciplinary perspectives on the new immigration*. Milton Park: Routledge: 111–136.

Clemens, M.A. (2014). Does development reduce migration? In: Lucas, R. (ed.). *International handbook on migration and economic development*. Cheltenham: Edward Elgar Publishing: 152–185.

Constant, A.F. (2020). Time-space dynamics of return and circular migration. In: Zimmermann, K. (ed.). *Handbook of labor, human resources and population economics*. Cham: Springer. doi:10.1007/978-3-319-57365-6$_1$07 − 1.

Czaika, M. and Godin, M. (2022). Disentangling the migration-development nexus using QCA. *Migration and Development* 11(3): 1065–1086. doi:10.1080/21632324.2020.1866878.

Dao, T.H., Docquier, F., Maurel, M., and Schaus, P. (2021). Global migration in the twentieth and twenty-first centuries: The unstoppable force of demography. *Review of World Economics* 157(2): 417–449. doi:10.1007/s10290-020-00402-1.

De Haas, H. (2007). Turning the tide? Why development will not stop migration. *Development and Change* 38(5): 819–841. doi:10.1111/j.1467-7660.2007.00435.x.

Dellink, R., Chateau, J., Lanzi, E., and Magné, B. (2017). Long-term economic growth projections in the shared socioeconomic pathways. *Global Environmental Change* 42: 200–214. doi:10.1016/j.gloenvcha.2015.06.004.

DEMIG (2023). DEMIG (2015) DEMIG C2C, version 1.2, limited online edition. Oxford: International Migration Institute, University of Oxford. https://www.migrationinstitute.org/data/demig-data/demig-c2c-data.

Docquier, F. (2018). Long-term trends in international migration: Lessons from macroeconomic model. *Economics and Business Review* 4(1): 3–15. doi:10.18559/ebr.2018.1.1.

Docquier, F. and Marfouk, A. (2004). Measuring the international mobility of skilled workers (1990–2000): Release 1.0. Washington, D.C.: World Bank (World Bank Policy Research Working Paper 3381). doi:10.1596/1813-9450-3381.

Docquier, F. and Marfouk, A. (2006). International migration by education attainment, 1990–2000. In: Ozden, C. and Schiff, M. (eds.). *International migration, remittances and the brain drain*. London: Palgrave Macmillan: 151–200.

Eurostat (2023). Eurostat Database. https://ec.europa.eu/eurostat/data/database.

Feenstra, R.C., Inklaar, R., and Timmer, M.P. (2015). The next generation of the Penn world table. *American Economic Review* 105(10): 3150–3182. doi:10.1257/aer.20130954.

Fransen, S. and Haas, H. (2022). Trends and patterns of global refugee migration. *Population and Development Review* 48(1): 97–128. doi:10.1111/padr.12456.

Geiger, T. (2018). Continuous national gross domestic product (GDP) time series for 195 countries: Past observations (1850–2005) harmonized with future projections accord-

ing to the shared socio-economic pathways (2006–2100). *Earth System Science Data* 10(2): 847–856. https://essd.copernicus.org/articles/10/847/2018/.

Giménez-Gómez, J.M., Walle, Y.M., and Zergawu, Y.Z. (2019). Trends in African migration to Europe: Drivers beyond economic motivations. *Journal of Conflict Resolution* 63(8): 1797–1831. doi:10.1177/0022002718823907.

Grogger, J. and Hanson, G.H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics* 95(1): 42–57. doi:10.1016/j.jdeveco.2010.06.003.

Hatton, T.J. and Williamson, J.G. (1998). *The age of mass migration: Causes and economic impact*. Oxford: Oxford University Press on Demand.

Hatton, T.J. and Williamson, J.G. (2005). What fundamentals drive world migration? In: Borjas, G. and Crisp, J. (eds.). *Poverty, international migration and asylum: Studies in development economics and policy*. London: Palgrave Macmillan: 15–38. doi:10.1057/9780230522534_2.

Helbling, M. and Leblang, D. (2019). Controlling immigration? How regulations affect migration flows. *European Journal of Political Research* 58(1): 248–269. doi:10.1111/1475-6765.12279.

International Organization for Migration (2019). World migration report 2020. Geneva: IOM Publications Platform. https://publications.iom.int/books/world-migration-report-2020.

Kaczmarczyk, P. and Okólski, M. (2005). *International migration in Central and Eastern Europe: Current and future trends*. Paper presented at Expert group meeting on international migration and development, New York, USA, July 7, 2005. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/unpd_egm_200507_p12_kaczmarczykokolski.pdf.

KC, S. and Lutz, W. (2017). The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100. *Global Environmental Change* 42: 181–192. doi:10.1016/j.gloenvcha.2014.06.004.

Keyfitz, N. (1985). *Applied mathematical demography*. Springer Texts in Statistics. New York: Springer. doi:10.1007/978-1-4757-1879-9.

Koch, J. and Leimbach, M. (2022). Update of SSP GDP projections: Capturing recent changes in national accounting, PPP conversion and Covid-19 impacts. *Ecological Economics* 206. doi:10.2139/ssrn.4011838.

Kracke, N. and Klug, C. (2021). Social capital and its effect on labour market (mis) match: Migrants' overqualification in Germany. *Journal of International Migration and Integration* 22: 1573–1598. doi:10.1007/s12134-021-00817-1.

Lutz, W., Butz, W.P., and KC, S. (2014). *World population and human capital in the twenty-first century: Executive summary*. Oxford: Oxford University Press.

Lutz, W., Butz, W.P., and KC, S. (2017). *World population, human capital in the twenty-first century: An overview*. Oxford: Oxford University Press. doi:10.1093/oso/9780198813422.001.0001.

Lutz, W., Goujon, A., KC, S., Stonawski, M., and Stilianakis, N. (2018). Demographic and human capital scenarios for the 21st century: 2018 assessment for 201 countries. Luxembourg: Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC111148/jrc_cepam_report_demographic_and_hc_scenarios_pdf.pdf.

Martin, P. and Taylor, J. (1993). The anatomy of a migration hump. In: Taylor, J. (ed.). *Development strategy, employment and migration: Insights from models*. Paris: OECD: 43–62.

Massey, D.S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Taylor, J.E. (1993). Theories of international migration: A review and appraisal. *Population and Development Review* 19(3): 431–466.

Mayda, A.M. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics* 23(4): 1249–1274. doi:10.1007/s00148-009-0251-x.

Naveed, A., Ahmad, N., Naz, A., and Zhuparova, A. (2023). Economic development through women's economic rights: A panel data analysis. *International Economics and Economic Policy* 20: 257–278. doi:10.1007/s10368-023-00560-1.

Nawrotzki, R.J. and Jiang, L. (2015). Indirectly estimating international net migration flows by age and gender: The community demographic model international migration (CDM-IM) dataset. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 48(3): 113–127. doi:10.1080/01615440.2014.999150.

Neumayer, E. and Plümper, T. (2021). Women's economic rights in developing countries and the gender gap in migration to Germany. *IZA Journal of Development and Migration* 12(1). doi:10.2478/izajodm-2021-0013.

OECD Stat (2022). Oecd.stat international migration database. https://stats.oecd.org/Index.aspx?DataSetCode=MIG.

Ortega, F. and Peri, G. (2013). The effect of income and immigration policies on international migration. *Migration Studies* 1(1): 47–74. doi:10.1093/migration/mns004.

O'Neill, B.C., Kriegler, E., Riahi, K., Ebi, K.L., Hallegatte, S., Carter, T.R., Mathur, R., and Van Vuuren, D.P. (2014). A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Climatic Change* 122: 387–400. doi:10.1007/s10584-013-0905-2.

Pauli, V., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods* 17: 261–272. doi:10.1038/s41592-019-0686-2.

Potančoková, M., Stonawski, M., and Gailey, N. (2021). Migration and demographic disparities in macro-regions of the European Union, a view to 2060. *Demographic Research* 45(44): 1317–1352. doi:10.4054/DEMRES.2021.45.44.

Qi, H. and Bircan, T. (2023). Modelling and predicting forced migration. *PLoS ONE* 18(4): e0284416. doi:10.1371/journal.pone.0284416.

Raymer, J., De Beer, J., and Van der Erf, R. (2011). Putting the pieces of the puzzle together: Age and sex-specific estimates of migration amongst countries in the EU/EFTA, 2002–2007. *European Journal of Population/Revue Européenne de Démographie* 27(2): 185–215. doi:10.1007/s10680-011-9230-5.

Raymer, J. and Wiśniowski, A. (2018). Applying and testing a forecasting model for age and sex patterns of immigration and emigration. *Population Studies* 72(3): 339–355. doi:10.1080/00324728.2018.1469784.

Rikani, A. and Schewe, J. (2021). Global bilateral migration projections accounting for diasporas, transit and return flows, and poverty constraints. *Demographic Research* 45(4): 87–140. doi:10.4054/DemRes.2021.45.4.

Rogers, A. (1975). Spatial migration expectancies. Laxenburg: IIASA (IIASA Research Memorandum). https://pure.iiasa.ac.at/id/eprint/459/.

Santos Silva, J.M.C. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics* 88(4): 641–658. doi:10.1162/rest.88.4.641.

Schon, J. and Leblang, D. (2021). Why physical barriers backfire: How immigration enforcement deters return and increases asylum applications. *Comparative Political Studies* 54(14): 2611–2652. doi:10.1177/00104140211024282.

Shayegh, S., Emmerling, J., and Tavoni, M. (2022). International migration projections across skill levels in the shared socioeconomic pathways. *Sustainability* 14(8): 4757. doi:10.3390/su14084757.

Statistics Sweden (2023). Statistics Sweden Database. https://www.scb.se/en/.

UN DESA (2019a). International migrant stock 2019. New York: UN DESA. https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates19.asp.

UN DESA (2019b). *World Population Prospects 2019: Methodology of the United Nations population estimates and projections*. Tech. rep., New York: UN DESA. https://population.un.org/wpp/Publications/Files/WPP2019_Methodology.pdf.

UN DESA (2020). International migration 2020 highlights. New York: UN DESA (ST/ESA/SER.A/452). https://www.un.org/development/desa/pd/news/international-migration-2020.

UN DESA (2022). World population prospects 2022: Methodology of the United Nations population estimates and projections. New York: UN DESA (UN DESA/POP/2022/TR/NO. 4). https://population.un.org/wpp/Publications/Files/WPP2022_Methodology.pdf.

UN High Commissioner for Refugees (2020). Refugee population statistics database [online]. https://www.unhcr.org/refugee-statistics/download/.

Wesselbaum, D. and Aburn, A. (2019). Gone with the wind: International migration. *Global and Planetary Change* 178: 96–109. doi:10.1016/j.gloplacha.2019.04.008.

Williams, N. (2009). Education, gender, and migration in the context of social change. *Social Science Research* 38(4): 883–896. doi:10.1016/j.ssresearch.2009.04.005.

Williamson, J.G. (2014). World migration in historical perspective: Four big issues. In: Chiswick, B. and Miller-Curtin, P. (eds.). *Handbook of the economics of international migration*. Amsterdam: North-Holland: 89–101.

Wiśniowski, A., Forster, J.J., Smith, P.W., Bijak, J., and Raymer, J. (2016). Integrated modelling of age and sex patterns of European migration. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179(4): 1007–1024. doi:10.1111/rssa.12177.

Wittgenstein Centre for Demography and Global Human Capital (2018). Wittgenstein centre data explorer version 2.0. [online]. http://www.wittgensteincentre.org/dataexplorer.

Yildiz, D. and Abel, G. (2023). Migration flow estimates by age, sex and education. Laxenburg: IIASA (IIASA working paper).

Yildiz, D. and Abel, G. (2024). Migration flows by age, sex and educational attainment. Laxenburg: IIASA (IIASA working paper). https://pure.iiasa.ac.at/id/eprint/19399/1/WP-24-001.pdf.

Yilma, M. and Regassa, N. (2019). Gender differentials in internal migration decisions: The case of Dilla Town, Southern Ethiopia. *KONTAKT Journal of Nursing and Social Sciences Related to Health and Illness* 21(3): 312–319. doi:10.32725/kont.2019.026.

Zagheni, E. and Weber, I. (2012). You are where you e-mail: Using e-mail data to estimate international migration rates. In: *Proceedings of the 4th annual ACM web science conference*. New York: ACM: 348–351. doi:10.1145/2380718.2380764.

Appendix

Calibration

While the bilateral flow data from F1 are a good choice for calibrating the original model (Rikani and Schewe 2021), it cannot directly be used to calibrate the age-dependent model, for the following reason. To use the bilateral flow data for calibration, one needs to collapse the simulated cohorts into a single bilateral migration flow. More specifically, one would evaluate $M^{Sim}_{i\to j} = M_{i,i\to j} - M_{j,i\to j} = \sum_{s,a} (M_{s,a,i,i\to j} - M_{s,a,j,i\to j})$.

Considering the model approaches in Equations (11) and (5), the coefficients σ_a and κ_a scale linearly with $M^{Sim}_{i\to j}$. Since we are aggregating over the age cohorts, the calibration algorithm estimates unrealistic age parameters. We obtain, for example, results where one of the σ_a parameters is estimated to be quite large so that it roughly fits the cohort-independent flows. The other two σ_a parameters are then estimated to be small or even negative so that two of the cohort flows can be interpreted as a correction to the major flows resulting from the first cohort or in a way that two cohorts cancel out each other due to negative coefficients.

Because of this problem, we decide to use the age-specific emigration and immigration data for the calibration of the age-dependent parameters. We would like to point out that this problem does not play such a major role when considering education or gender dynamics since those parameters do not linearly scale with the magnitude of the flows and cannot lead to negative migration flows, we use the cohort aggregation method here.

Table A-1: Summary of datasets used in this paper, including reference and abbreviations

Data	Abbreviation	Origin	Reference
Bilateral stocks	P1	UN Department of Economic and Social Affairs	(UN DESA 2019a)
Total population by age, education, and gender	P2	Wittgenstein Centre for Demography and Global Human Capital	(Wittgenstein Centre for Demography and Global Human Capital 2018)
Bilateral refugee stocks	R1	UN High Commissioner for Refugees	(UN High Commissioner for Refugees 2020)
Historical GDP data	G1	Penn World Tables 8.1	(Feenstra, Inklaar, and Timmer 2015)
Historical GDP data	G2	Penn World Tables 9.0	(Geiger 2018)
Future GDP projections	G3	Koch and Leimbach	(Koch and Leimbach 2022)
Bilateral flows	F1	Bilateral international migration flow estimates for 200 countries	(Abel and Cohen 2022)
Age-dependent immigration and emigration flows	F2	Yildiz and Abel (methods in (Yildiz and Abel 2024))	(Yildiz and Abel 2024)
Bilateral flows	F3	DEMIG C2C (country-to-country) database	(DEMIG 2023)

Table A-2: Flow number for Figure 6: Abel and Cohen (F1) 2015–2019 (Total Flows: 91.09). All numbers are given in millions

Origin / Destination	Low	Lower Middle	Upper Middle	High
Low	3.74	2.16	1.38	2.18
Lower Middle	1.64	3.71	4.45	17.03
Upper Middle	1.21	2.73	8.04	12.78
High	0.98	6.17	7.93	8.04

Table A-3: Flow number for Figure 6: Original 2015–2019 (Total Flows: 54.47). All numbers are given in millions

Origin / Destination	Low	Lower Middle	Upper Middle	High
Low	0.41	0.80	1.40	1.44
Lower Middle	0.49	1.96	2.83	13.99
Upper Middle	0.12	1.03	3.17	12.59
High	0.26	1.75	2.68	9.56

Table A-4: Flow number for Figure 6: Full 2015–2019 (Total Flows: 84.31). All numbers are given in millions

Origin / Destination	Low	Lower Middle	Upper Middle	High
Low	0.85	1.67	2.60	3.48
Lower Middle	0.82	3.34	4.99	27.46
Upper Middle	0.14	1.15	4.23	19.23
High	0.26	1.68	2.60	9.82

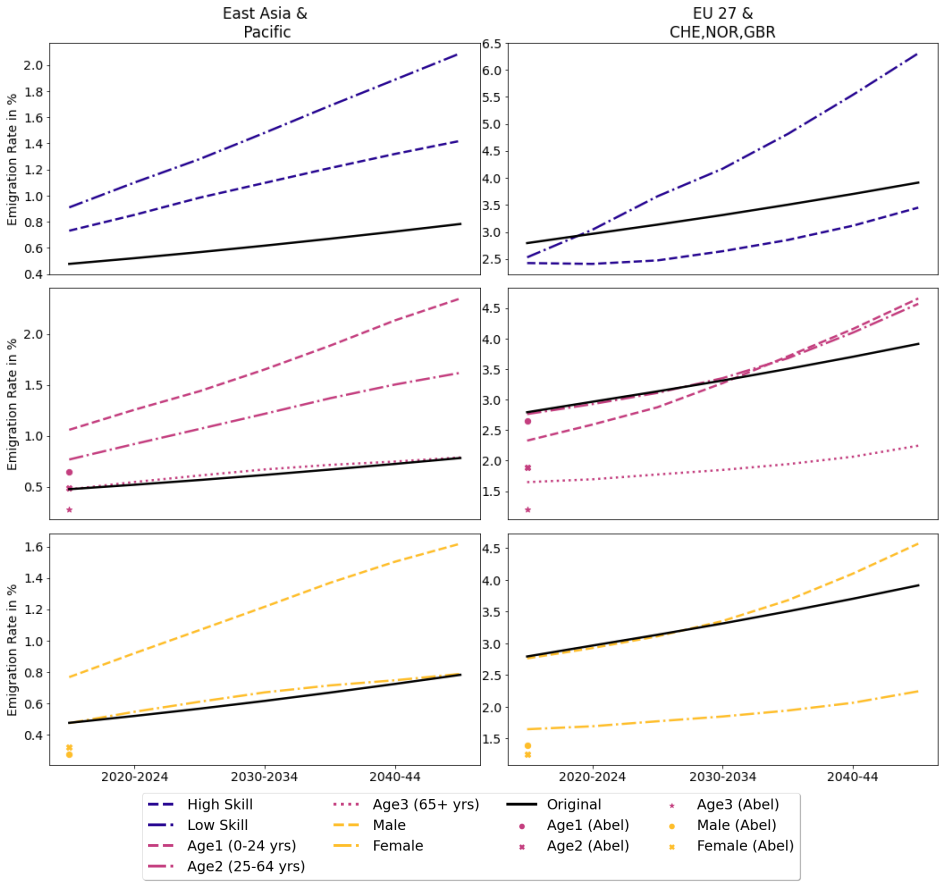
Table A-5: Flow number for Figure 6: Full 2045–2049 (Total Flows: 226.89). All numbers are given in millions

Origin / Destination	Low	Lower Middle	Upper Middle	High
Low	2.98	6.10	6.80	27.92
Lower Middle	2.16	7.93	11.55	88.01
Upper Middle	0.27	1.18	5.21	26.81
High	3.33	11.30	5.36	19.99

Table A-6: Countries (ISO 3166-1 alpha-3) included in the calibration and projection (indicated with the respective region) and those that are missing

Region	Countries
EAP (East Asia and Pacific)	**CHN**: China, **FJI**: Fiji, **HKG**: Hong Kong, **IDN**: Indonesia, **JPN**: Japan, **KHM**: Cambodia, **KOR**: South Korea, **LAO**: Laos, **MAC**: Macao, **MMR**: Myanmar (Burma), **MNG**: Mongolia, **MYS**: Malaysia, **NZL**: New Zealand, **PHL**: Philippines, **SGP**: Singapore, **SLB**: Solomon Islands, **THA**: Thailand, **TLS**: Timor-Leste (East Timor), **VNM**: Vietnam, **VUT**: Vanuatu, **WSM**: Samoa
EU (EU 27, Great Britain, Switzerland and Norway)	**ALB**: Albania, **AUT**: Austria, **BEL**: Belgium, **BGR**: Bulgaria, **CHE**: Switzerland, **CYP**: Cyprus, **CZE**: Czech Republic, **DEU**: Germany, **DNK**: Denmark, **ESP**: Spain, **EST**: Estonia, **FIN**: Finland, **FRA**: France, **GBR**: United Kingdom, **GRC**: Greece, **HRV**: Croatia, **HUN**: Hungary, **IRL**: Ireland, **ISL**: Iceland, **ITA**: Italy, **LTU**: Lithuania, **LUX**: Luxembourg, **LVA**: Latvia, **MLT**: Malta, **NLD**: Netherlands, **NOR**: Norway, **POL**: Poland, **PRT**: Portugal, **ROU**: Romania, **SVK**: Slovakia, **SVN**: Slovenia, **SWE**: Sweden
EECA (East Europe and Central Asia)	**ARM**: Armenia, **AZE**: Azerbaijan, **BIH**: Bosnia and Herzegovina, **BLR**: Belarus, **GEO**: Georgia, **KAZ**: Kazakhstan, **KGZ**: Kyrgyzstan, **MDA**: Moldova, **MKD**: North Macedonia, **MNE**: Montenegro, **RUS**: Russia, **SRB**: Serbia, **TJK**: Tajikistan, **TKM**: Turkmenistan, **UKR**: Ukraine
LAC (Latin America and Caribbean)	**ARG**: Argentina, **BHS**: Bahamas, **BLZ**: Belize, **BOL**: Bolivia, **BRA**: Brazil, **CHL**: Chile, **COL**: Colombia, **CRI**: Costa Rica, **DOM**: Dominican Republic, **ECU**: Ecuador, **GTM**: Guatemala, **GUY**: Guyana, **HND**: Honduras, **HTI**: Haiti, **JAM**: Jamaica, **LCA**: Saint Lucia, **MEX**: Mexico, **NIC**: Nicaragua, **PAN**: Panama, **PER**: Peru, **PRI**: Puerto Rico, **PRY**: Paraguay, **SLV**: El Salvador, **SUR**: Suriname, **TTO**: Trinidad and Tobago, **URY**: Uruguay, **VCT**: Saint Vincent and the Grenadines, **VEN**: Venezuela
MENA (Middle East and North Africa)	**ARE**: United Arab Emirates, **BHR**: Bahrain, **DZA**: Algeria, **EGY**: Egypt, **IRN**: Iran, **IRQ**: Iraq, **ISR**: Israel, **JOR**: Jordan, **KWT**: Kuwait, **LBN**: Lebanon, **MAR**: Morocco, **OMN**: Oman, **PSE**: Palestine, **QAT**: Qatar, **SAU**: Saudi Arabia, **SYR**: Syria, **TUN**: Tunisia, **TUR**: Turkey, **YEM**: Yemen
NAM (North America)	**CAN**: Canada, **USA**: United States of America
SAS (South Asia)	**AFG**: Afghanistan, **BGD**: Bangladesh, **BTN**: Bhutan, **IND**: India, **LKA**: Sri Lanka, **MDV**: Maldives, **NPL**: Nepal, **PAK**: Pakistan
SSA (sub-Saharan Africa)	**AGO**: Angola, **BDI**: Burundi, **BEN**: Benin, **BFA**: Burkina Faso, **BWA**: Botswana, **CAF**: Central African Republic, **CIV**: Ivory Coast (Côte d'Ivoire), **CMR**: Cameroon, **COD**: Democratic Republic of the Congo, **COG**: Republic of the Congo, **COM**: Comoros, **CPV**: Cape Verde, **ETH**: Ethiopia, **GAB**: Gabon, **GHA**: Ghana, **GIN**: Guinea, **GMB**: Gambia, **GNB**: Guinea-Bissau, **GNQ**: Equatorial Guinea, **KEN**: Kenya, **LBR**: Liberia, **LSO**: Lesotho, **MDG**: Madagascar, **MLI**: Mali, **MOZ**: Mozambique, **MUS**: Mauritius, **MWI**: Malawi, **NAM**: Namibia, **NER**: Niger, **NGA**: Nigeria, **RWA**: Rwanda, **SDN**: Sudan, **SEN**: Senegal, **SLE**: Sierra Leone, **SOM**: Somalia, **STP**: São Tomé and Príncipe, **SWZ**: Eswatini (formerly Swaziland), **TGO**: Togo, **TZA**: Tanzania, **UGA**: Uganda, **ZAF**: South Africa, **ZMB**: Zambia, **ZWE**: Zimbabwe
Missing	**ATG**: Antigua and Barbuda, **BRB**: Barbados, **BRN**: Brunei, **DJI**: Djibouti, **DMA**: Dominica, **ERI**: Eritrea, **FSM**: Federated States of Micronesia, **GRD**: Grenada, **KNA**: Saint Kitts and Nevis, **KIR**: Kiribati, **LBY**: Libya, **MHL**: Marshall Islands, **MRT**: Mauritania, **NRU**: Nauru, **PLW**: Palau, **PNG**: Papua New Guinea, **SMR**: San Marino, **SSD**: South Sudan, **SYC**: Seychelles, **TCD**: Chad, **TWN**: Taiwan, **UZB**: Uzbekistan

Figure A-1: Projected emigration rates disaggregated by education and age, East Asia and Pacific and EU 27, CHE, NOR, GBR



Notes: Black solid lines show the aggregate projections from the original model; other lines show the education-specific, age-specific, and gender-specific projections from the full model. Single markers indicate the age-specific data for the 2015–2019 period from F2 and the gender-specific data for the 2015–2019 period from F1. Regional figures include both flows between countries within a region and flows between regions.

Figure A-2: Projected emigration rates disaggregated by education and age, East Europe and Central Asia and Latin America and Caribbean

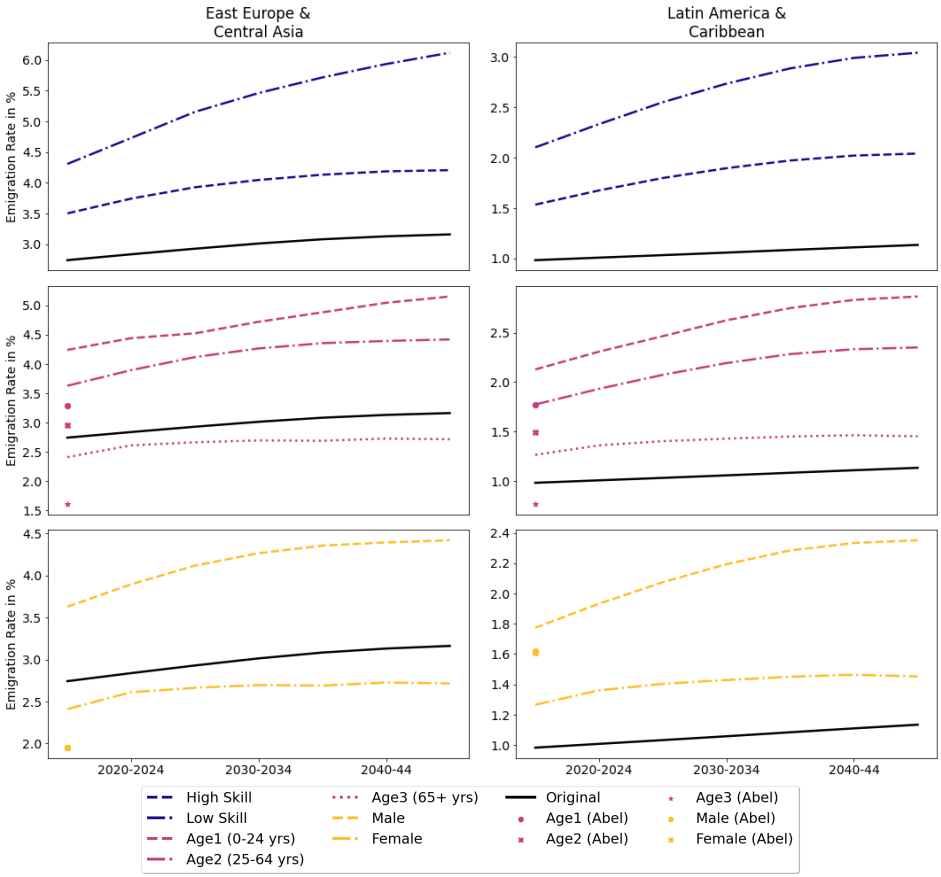


Figure A-3: Projected emigration rates disaggregated by education and age, Middle East and North Africa and North America

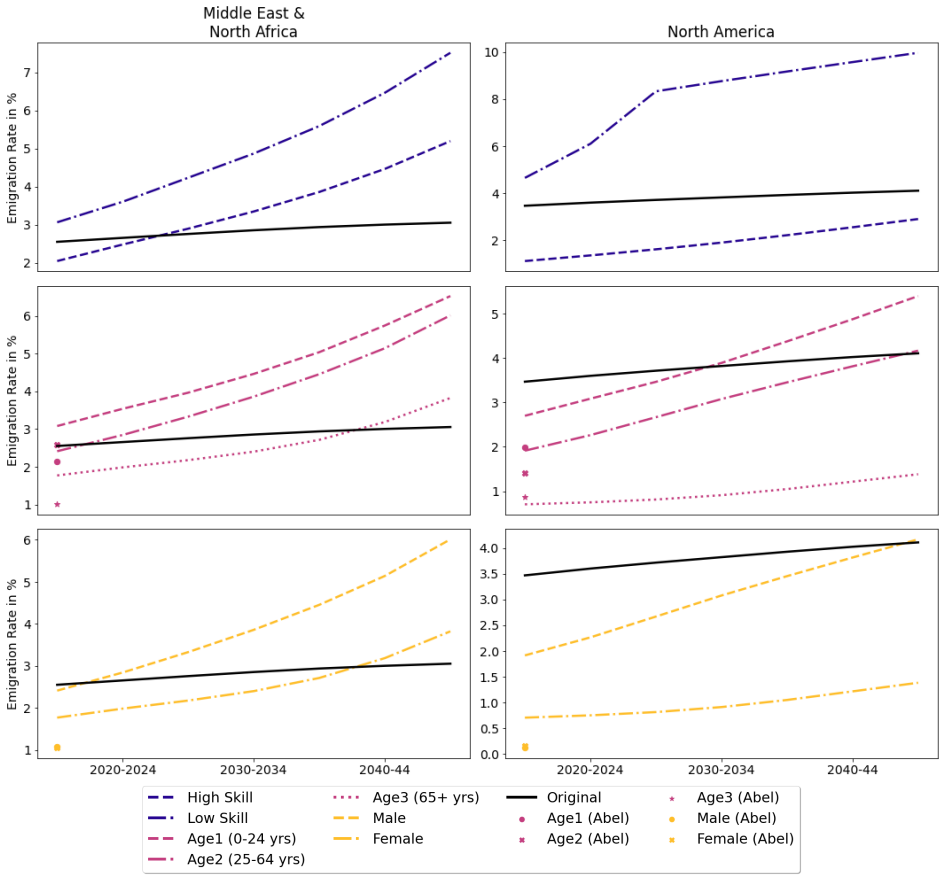


Figure A-4: Projected emigration rates disaggregated by education and age, South Asia and sub-Saharan Africa

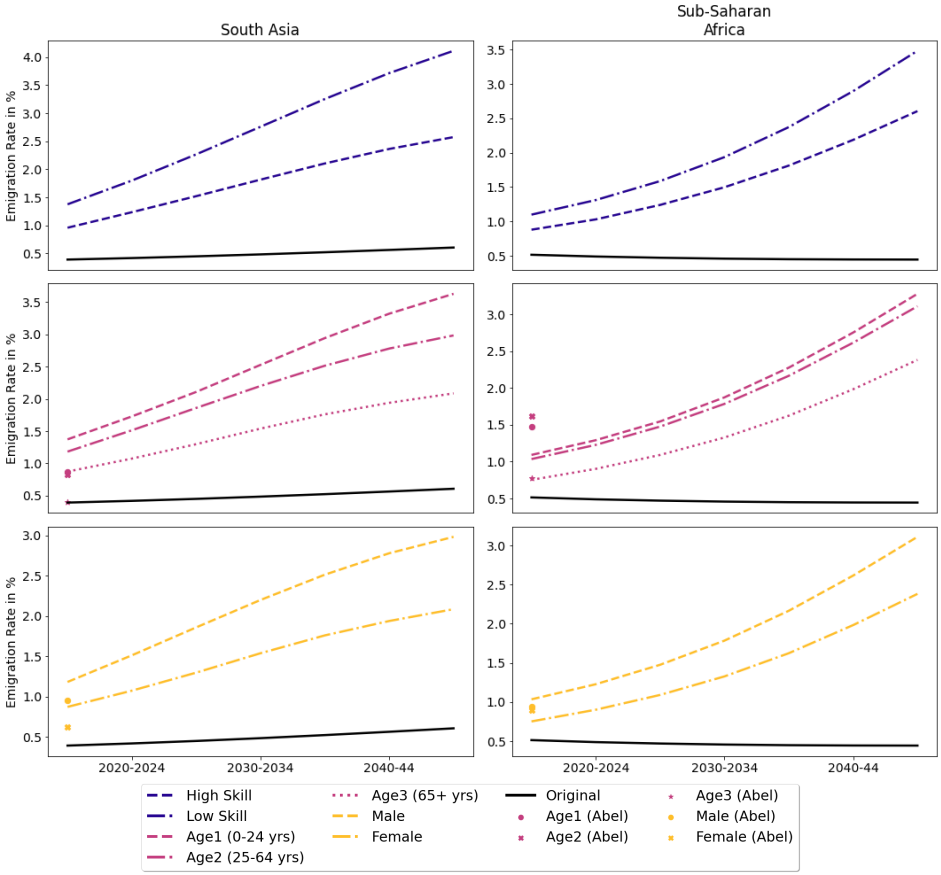


Figure A-5: Projected immigration rates disaggregated by education and age, East Europe and Central Asia and Latin America and Caribbean

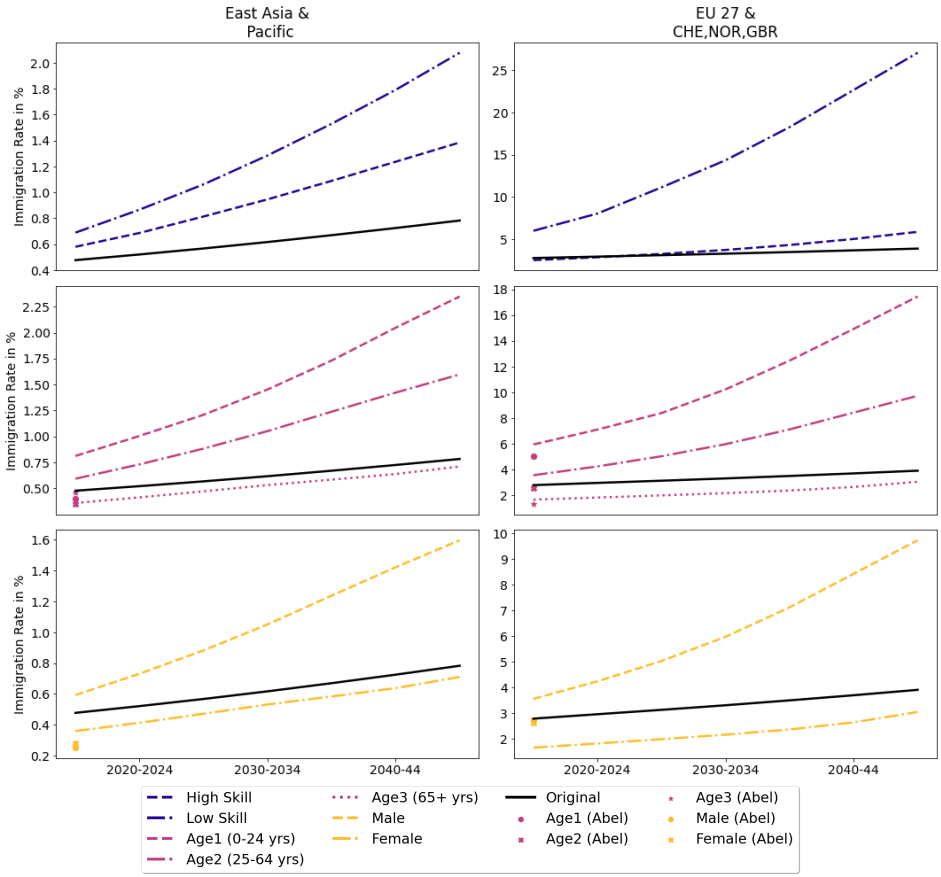


Figure A-6: Projected immigration rates disaggregated by education and age, East Europe and Central Asia and Latin America and Caribbean

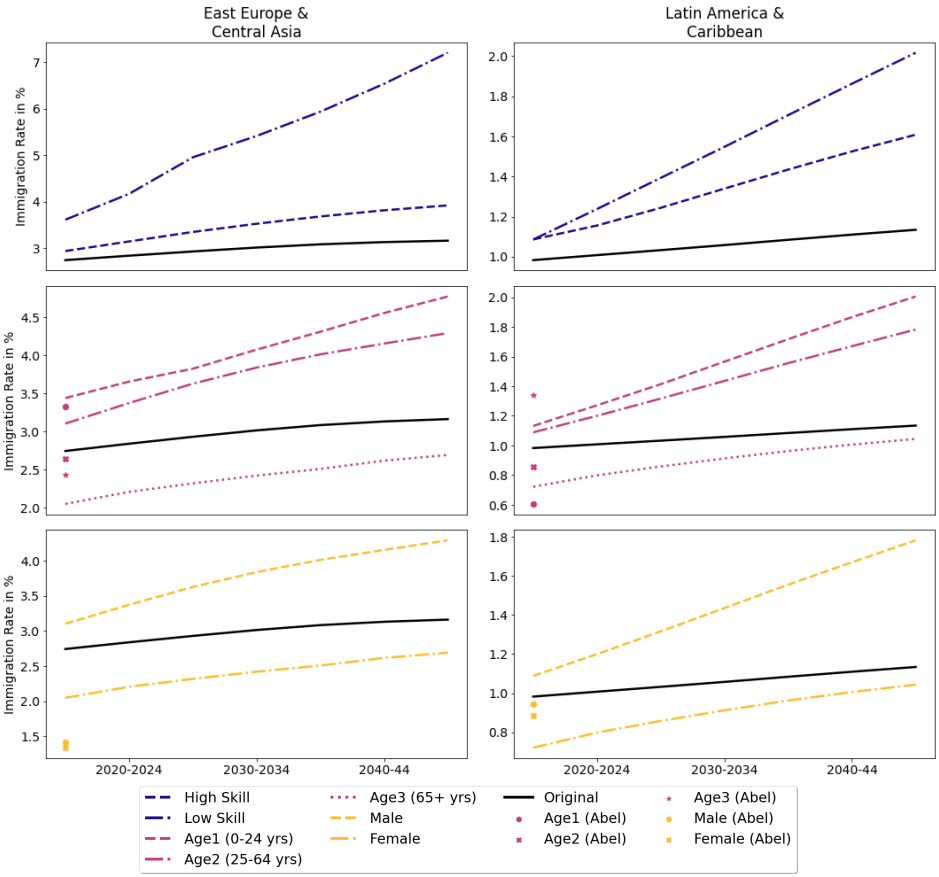


Figure A-7: Projected immigration rates disaggregated by education and age, Middle East and North Africa and North America

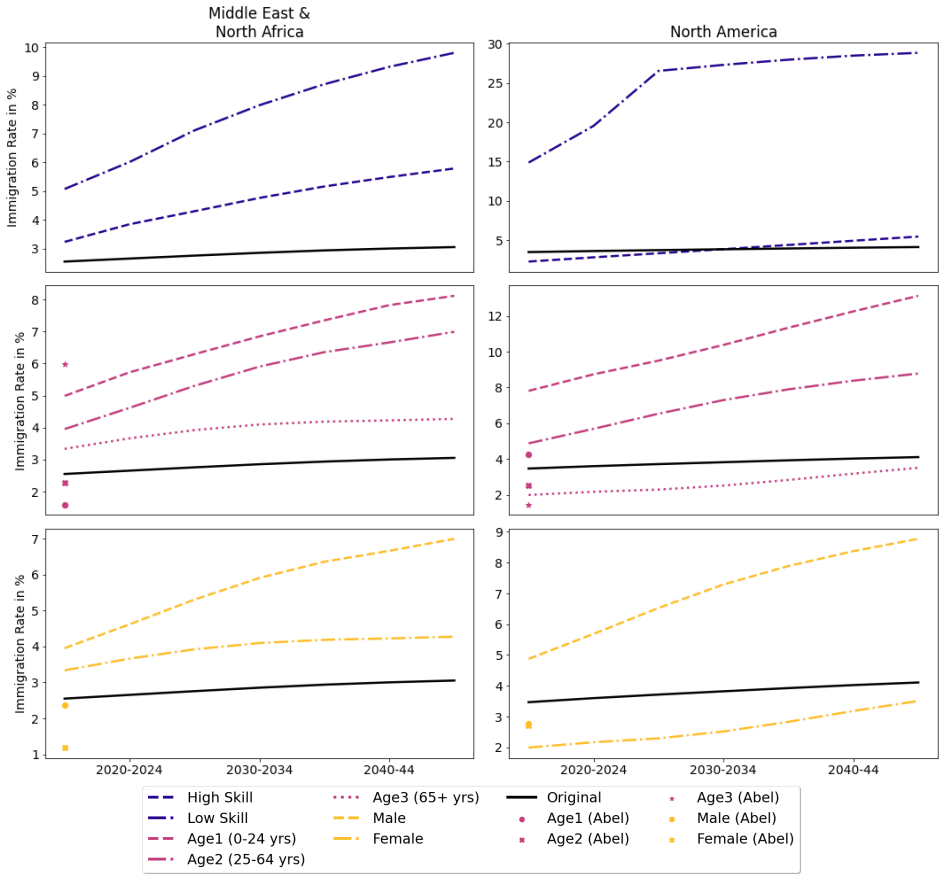


Figure A-8: Projected immigration rates disaggregated by education and age, South Asia and sub-Saharan Africa