International Institute for Applied Systems Analysis
IIASA

Address: IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria
Email: repository@iiasa.ac.at Telephone: +43 (0)2236 807 342

YSSP Report

Young Scientist Summer Program

Estimating and forecasting bilateral migration flows from Europe to South America, 1986-2060

Author: Andrea Aparicio-Castro
Email: andrea.apariciocastro@manchester.ac.uk

Approved by

Supervisor: Dilek Yildiz
Co-Supervisor: Michaela Potancokova
Program: Population and Just Societies Program (POPJUS)
September 30, 2021

This report represents the work completed by the author during the IIASA Young Scientists Summer Program (YSSP) with approval from the YSSP supervisor.

It was finished by September 30, 2021 and has not been altered or revised since.

Supervisor signature:

International Migration Review
():2–27
SAGE

Estimating and forecasting bilateral migration flows from Europe to South America, 1986-2060

Andrea Aparicio-Castro[1,2]

Abstract

Europe has been one of the main destinations for migrants since 1990, hosting more than 30% of the global migrant stocks. Given the increasing importance of Europe as a destination, research usually focuses on migration towards the region. Less attention has been paid to migration from Europe to the rest of the world, neglecting the need of analysing the impact of emigration. Emigration is a widely debatable topic amongst policymakers insofar as it may represent the loss/waste of human capital. South America (SA) has been one of the usual receiving regions of high-skilled migrants arriving from Europe, even producing the change of the direction of Europe-SA flows in some periods, e.g., of economic crisis. The absence of a complete-time series of these flows entangles drawing conclusions about them. Therefore, we aim at (a) estimating a complete and consistent time series of bilateral migration flows from Europe to SA from 1986 to 2019, and (b) quantifying how likely these patterns will remain across time, forecasting these flows until 2060. We use census data as an input to estimate flows since they are the most complete and reliable data. However, using censuses entails (a) working with five-year transition data that are not always comparable as they do not always refer to the same five years, (b) measurement errors caused by measuring migration indirectly, and (c) missingness in the intercensal periods, for which data are not available. We propose a two-part hierarchical Bayesian model, which (a) translates five-year transition census data into one-year values; (b) corrects the measurement errors that using census data involves; (c) imputes flows for the intercensal periods; and, (d) forecasts these flows until 2060. The output is a set of synthetic estimates of bilateral migration flows from Europe to SA from 1986 to 2060 with measures of uncertainty.

Keywords
Estimation, forecast, migration flows, Europe, South America, hierarchical model, Bayesian inference.

1 Introduction

Europe has been one of the main contemporary destinations for migrants around the world (Czaika et al. 2021, pp.4-8). Migrant stocks provide a hint of this aspect. Based on the United Nations (UN) estimates, Europe has hosted more than 30% of the global migrant stocks between 1990 and 2020, representing the highest percentage across all regions and periods (but in $2010 - 2015$ when Asia surpassed Europe) (UNDESA 2020). Due to the increasing importance of Europe as a destination, research usually focuses on migration towards the region. Less attention has been paid to migration from Europe to the rest of the world, neglecting the need of analysing the impact of emigration. Emigration is a widely debatable topic amongst policymakers insofar as it may represent the loss or waste of human capital, especially, of high-skilled migrants (McAuliffe and Khadria 2019, p.176).

Sassone and del Castillo (2014, p.24) hypothesise that South America has been one of the usual receiving regions of high-skilled migrants arriving from Europe. According to Sassone and del Castillo (2014), this has even altered the direction of migration flows between these two regions during specific periods (e.g. in the 2008 economic crisis). However, the absence of a complete-time series of migration flows from Europe to South America entangles drawing conclusions about (a) whether these rearrangements of flows answer specific conjunctural circumstances or respond to structural changes, and (b) future patterns of these flows. Thus, this paper aims at (a) estimating a complete and consistent time series of bilateral international migration flows from Europe to South America from 1986 to 2019, and (b) quantifying how likely these patterns will remain across time, forecasting these flows until 2060.

There are different types of data to analyse migration flows. We use census data because they are the most complete, reliable and comparable information (Bryant and Zhang 2018, p.186; Juran and Snow 2018; Rodríguez-Vignoli and Rowe 2018). In addition, census data are not tied to any country-specific legislation, which enables overcoming cross-national dissimilarities in the definition of migrants due to being collected mostly following commonly shared guidelines (e.g. UNDESA 2017, 2008, 1998).

[1]Social Statistics Department, University of Manchester, UK.
[2]Population and Just Societies Program (POPJUS), International Institute for Applied Systems Analysis (IIASA), Austria.

Corresponding author:
Andrea Aparicio-Castro, Social Statistics Department, Humanities Bridgeford Street, University of Manchester, M15 6FH, UK.
Email: andrea.apariciocastro@manchester.ac.uk

Notwithstanding that census data allow tackling cross-national differences, using censuses to estimate annual migration flows entails three main problems. The first problem is related to having five-year transition data. Censuses usually ask only about residence five years prior to each census. The problem is that censuses refer to unmatched five-year periods, which start and end in dissimilar years. Therefore, a transformation from five- to one-year flows is needed to make census data comparable. The second problem of using census data is that census information leads to measurement errors or biases because a migrant is categorised as such if their previous residence is different from the census place, which means that migration is captured indirectly. The current paper addresses four of the usual sources of errors originated from: (a) the dissimilarities in the number of migrants registered in censuses which assume the *de jure* or the *de facto* approach (Swanson and Tayman 2011, p.7); (b) the omission of infant migrants, which corresponds -in this case- to children aged 0-4 years old; (c) the neglection of non-surviving migrants; and (d) the (im)possibility of identifying migrants by their birthplace or residence. The third problem of using census data to analyse migration flows is the lack of information in the intercensal periods, for which census data are not available.

To create a complete and consistent time series of bilateral migration flows from Europe to South America from 1986 to 2060, we propose a two-part hierarchical Bayesian model. The first part of our model (a) translates five-year transition census data into one-year values, and (b) corrects the measurement errors or biases that using census data involves. The second part of our hierarchical Bayesian model (a) imputes flows for the intercensal periods from 1986 to 2019, and (b) forecasts these flows until 2060. The resulting estimated and forecast values are the true (unobserved) bilateral international migration flows, where a migrant can be defined as a person whose place of residence is different at the beginning and the end of each year.

For imputing and forecasting migration flows, we use the estimates and projections gauged by the International Institute for Applied Systems Analysis (IIASA) as part of the Shared Socioeconomic Pathways (SSP) scenarios (Lutz et al. 2018; KC and Lutz 2017; Cuaresma 2017; van Vuuren et al. 2017; Fricko et al. 2017; Fujimori et al. 2017). These sets of values have been widely employed not only by researchers but also by policymakers, making the current paper useful to assess the versatility of these values.

The rest of this paper is divided into six sections. In Section 2, we present the migration system considered for this research and the respective input data. In Section 3, we specify the two-part hierarchical Bayesian model that aims at estimating and forecasting migration flows from Europe to South America. Section 4 displays the results of our model. In Section 5, we describe the estimated and forecast annual bilateral migration flows. Section 6 shows the sensitivity analysis performed on the resulting estimates. Section 7 corresponds to the concluding remarks.

2 Background

2.1 Data.

There are different types of data to analyse migration flows. We use census data, specifically, microdata of the 10 biggest South American countries that cover the period from 1986 to 2019[*]. Even though there are other types of data (e.g. from administrative sources) that are produced more frequently than census data, they are usually tied to country-specific legislation. This leads to considerable cross-national dissimilarities in the definition of who constitutes a migrant, and therefore, affect cross-national comparability. Amongst all types of data, census data are still the most comparable information, enabling overcoming cross-national differences in the definition of a migrant due to being collected mostly following commonly shared guidelines (e.g. UNDESA 2017, 2008, 1998).

In census data, we categorise migrants as such if their residence five years prior to each census is different from their residence at the census date. This enables the identification of migrants despite their birthplace, and therefore, being able to capture movements from native-born migrants (i.e. returners) (Villa 1991, p.25). Data on country of usual residence when available is used to remove non-usual residents (e.g. seasonal population and daytime population, Swanson and Tayman 2011). The data were extracted using (a) `Redatam7`, which is a software for processing data designed by CELADE (2019), and (b) `Redatam+SP`, which is the online version of `Redatam7`.

2.2 Migration system.

Migration flows can be conveniently arranged in a contingency table, where rows refer to origin i, and columns indicate destination j in a year t[†]. Considering the 32 European countries that conform the EU-27, EEA, Switzerland and UK, and the 10 biggest South American countries (i.e. Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay and Venezuela), this contingency table can be seen as a partitioned matrix with four blocks (see Table (1)). The upper-left block corresponds to the number of migrants from European countries who arrive and leave any other country within Europe. The upper-right block refers to the flows from any European country to South America. The lower-left block alludes to flows from South America to Europe. Finally, the lower-right block values correspond to flows within South America.

[*] Guyana and Suriname are excluded from this analysis due to the lack of access to their census microdata. Moreover, dependent territories (i.e. Falkland Islands-UK, South Georgia and South Sandwich Islands-UK) and internal territories (i.e. French Guiana–France) are excluded from the current analysis.

[†] In this study, origins and sending countries are equivalent. Likewise, destinations, census places, census countries and receiving countries are used interchangeably

Table 1. Arrangement of migration flows between Europe and South America (highlighted part=focus of the current research).

Orig\Dest	AT	BE	BG	$\dots$	GB	AR	BO	BR	$\dots$	VE
AT	0	$z_{1,2,t}$	$z_{1,3,t}$	$\cdots$	$z_{1,32,t}$	$z_{1,33,t}$	$z_{1,34,t}$	$z_{1,35,t}$	$\cdots$	$z_{1,42,t}$
BE	$z_{2,1,t}$	0	$z_{2,3,t}$	$\cdots$	$z_{2,32,t}$	$z_{2,33,t}$	$z_{2,34,t}$	$z_{2,35,t}$	$\cdots$	$z_{2,42,t}$
BG	$z_{3,1,t}$	$z_{3,2,t}$	0	$\cdots$	$z_{3,32,t}$	$z_{3,33,t}$	$z_{3,34,t}$	$z_{3,35,t}$	$\cdots$	$z_{3,42,t}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
GB	$z_{32,1,t}$	$z_{32,2,t}$	$z_{32,3,t}$	$\cdots$	0	$z_{32,33,t}$	$z_{32,34,t}$	$z_{32,35,t}$	$\cdots$	$z_{32,42,t}$
AR	$z_{33,1,t}$	$z_{33,2,t}$	$z_{33,3,t}$	$\cdots$	$z_{33,32,t}$	0	$z_{33,34,t}$	$z_{33,35,t}$	$\cdots$	$z_{33,42,t}$
BO	$z_{34,1,t}$	$z_{34,2,t}$	$z_{34,3,t}$	$\cdots$	$z_{34,32,t}$	$z_{34,33,t}$	0	$z_{34,35,t}$	$\cdots$	$z_{34,42,t}$
BR	$z_{35,1,t}$	$z_{35,2,t}$	$z_{35,3,t}$	$\cdots$	$z_{35,32,t}$	$z_{35,33,t}$	$z_{35,34,t}$	0	$\cdots$	$z_{35,42,t}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
VE	$z_{42,1,t}$	$z_{42,2,t}$	$z_{42,3,t}$	$\cdots$	$z_{42,32,t}$	$z_{42,33,t}$	$z_{42,34,t}$	$z_{42,35,t}$	$\cdots$	0

We focus our analysis on the upper-right part of the matrix displayed in Table (1) since censuses capture immigrants better than emigrants. To be able to capture emigrants, censuses require that individuals leaving a region are in the census place at the census date or has a relative who reports his/her movement. Additionally, South American censuses -in particular- do not usually have a question about emigration. Even if they do, questions do not refer to the period in which migrants emigrated (e.g. 1993 Colombian census) or the options are grouped, making the identification of the period of the movement non-viable (e.g. 2010 Brazilian census).

In the case of origins, we also reduce the number of sending countries due to the level of missingness in South American census data. Figure (1) shows the percentage of missing values amongst the years covered by South American censuses. Countries whose missing values do not surpass 10% are defined as origins.

After defining our origins and destinations, our migration system is composed of 14 sending countries and 10 receiving countries. Origins correspond to Denmark, Italy, Spain, Poland, Austria, Belgium, France, Germany, Greece, Netherlands, Portugal, Sweden, Switzerland and the UK. Destinations refer to Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela. Each origin i and destination j pair conforms a migration corridor, where $i \neq j$. In total, there are 140 migration corridors, corresponding to flows from the 14 selected European countries to the 10 biggest South American countries.

From a multilevel perspective, our data can be seen as a two-level dataset. At level 1, we have repeated measures from 1986 to 2060. These repeated measures are nested within each of the defined migration corridors at level 2. Thus, our expected output is a set of 10500 synthetic entries (i.e. 140 corridors$\times 75$

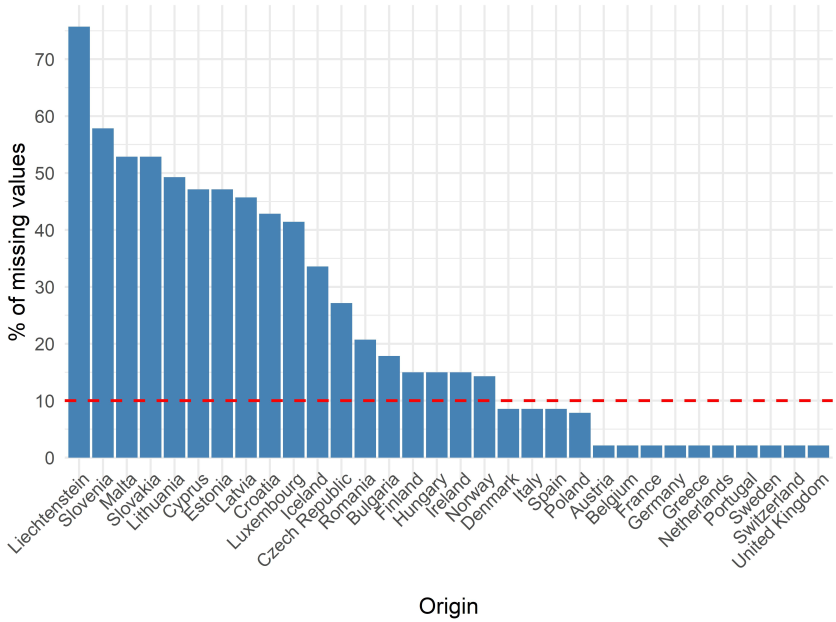

Figure 1. Percentage of missing values in available migration flow data by origins.

years), corresponding to the estimated and forecast bilateral true migration flows from Europe to South America from 1986 to 2060 with measures of uncertainty.

3 Methods: modelling framework

To create a complete time series of bilateral migration flows from the 14 European countries (see Section 2) to South America from 1986 to 2060, we propose a hierarchical Bayesian model composed of two parts, which are modelled simultaneously (see Figure (2)). The first part corresponds to a Poisson-lognormal data sub-model, whose main purposes are (a) to translate five-year transition census data into one-year values, and (b) to correct the measurement errors or biases that using census data entails. The second part of our hierarchical Bayesian model refers to an Autoregressive Distributed Lag (ADL) sub-model of order 1, which aims at (a) imputing migration flows for the intercensal periods from 1986 to 2019, and (b) forecasting these flows until 2060.

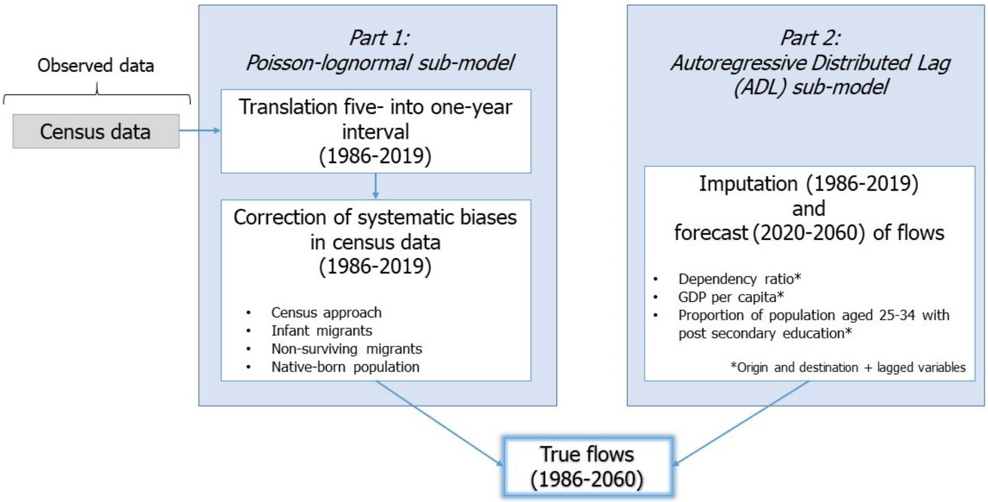

Figure 2. Modelling framework

3.1 Poisson-lognormal data sub-model.

The first part of our Bayesian hierarchical model is the Poisson-lognormal data sub-model that replicates the method implemented by Aparicio-Castro et al. (nd), which is a generalisation of the model proposed by Raymer et al. (2013) for integrating European migration flows. Our Poisson-lognormal data sub-model aims at (a) translating five-year transition census data into one-year values, and (b) correcting the measurement errors or biases that using census data involves.

3.1.1 Translating five-year data into one-year flows: The first purpose of our Poisson-lognormal data sub-model is to translate from five- to one-year flows. This is needed, given that most of the South American censuses only provide information about residence five years prior to each census. All five-year periods allude to five years, whose starting and ending points in time do not match. Thus, we require to transform five-year data into one-year flows in order to make these data comparable. We implement the same method as in the first level of the model developed by Aparicio-Castro et al. (nd), where a deflation factor for translating from five- into one-year flows is used (see Equation (1)).

$$z_{ijt-x}^{(1)} \sim \text{Poisson}\left(\mu_{ijt}^{(1)} = \mu_{ijt}^{(5)} \cdot \frac{S_{ijt-x}^{(1)}}{S_{ijt}^{(5)}}\right) \text{for } x = 0, 1, 2, 3, 4. \tag{1}$$

Equation (1) displays that the annual number of migrants $z^{(1)}_{ijt-x}$ from origin i to destination j at a given year t follows a Poisson distribution with an expected value $\mu^{(1)}_{ijt}$. The term $\mu^{(1)}_{ijt}$ results from multiplying the expected value of the number of migrants arriving at census place j over five-year periods registered in South American censuses $\left(\mu^{(5)}_{ijt}\right)$ and the deflation factor $\left(\dfrac{S^{(1)}_{ijt-x}}{S^{(5)}_{ijt}}\right)$. The term $S^{(1)}_{ijt-x}$ refers to the annual migrant stocks born in country i and living in census country j at time t, whose first arrival to j was at most four years prior to each census (i.e. $x = 0, 1, 2, 3, 4$). The denominator of the deflation factor $S^{(5)}_{ijt}$ is the stocks of migrants over those five years. These data can be also extracted from the South American censuses. Specifically, data on birthplace and year of the first arrival asked the foreign-born individuals in a receiving country j are employed as input for translating from five- to one-year flows.

3.1.2 Correcting measurement errors and biases of census data: The second aim of the first part of our Bayesian hierarchical model is to correct the measurement error or biases that using censuses involves. This part is analogous to the second level of the model developed by Aparicio-Castro et al. (nd). The biases are produced by the fact that a migrant is categorised as such if their country of previous residence or country of birth is different from the census place, which implies that migration is captured indirectly. In other words, the movement of the migrant is not recorded when it happens.

This research tackles four of the main sources of errors in census-based estimates. The first source of error surges from the differences in the number of migrants that are registered in *de jure* and *de facto* censuses. While *de jure* censuses count only legal residents, *de facto* censuses reports every individual who is present at the census place and census time (Swanson and Tayman 2011, p.7; Siegel and Swanson 2004, p.49; Shryock et al. 1976, p.49). There are 18 *de facto* censuses out of 28 South American censuses taken in the 1990s, 2000s and 2010s. Therefore, we determine a variable whose group of reference is *de facto* censuses in order to standardise flows to the most prevalent census approach in South America.

The second source of error in census-based estimates is related to the omission of infant migrants, who are those persons whose birth occurred in the country of origin and reached their destination within the transition interval of interest (Rees et al. 2000, pp.208). Bearing in mind that we use the census question about residence five years prior to each South American census time, children under five years old are defined as infant migrants. Following Rogers and Jordan (2004, p.42) and Rogers et al. (2003, p.56), we assume that the number of foreign-born children aged 0-4 equals the migration flow in the same age group.

The third source of error in census-based estimates is associated with the omission of non-surviving migrants. Hinde (2014) and Rees (1977, p.251) define a non-surviving migrant as a person who migrated from origin i to destination j and died in his/her destination without being registered in the respective census. The UN Demographic books 1986-2018 provide the total number of deaths in each census place.

We deduce the number of migrant deaths by multiplying the total of deaths in census country in a given year by the proportion of the number of migrants in the same year $z^{(1)}_{ijt-1}$ (see Equation (1) in Section 3.1.1) out of the total population in destination.

The fourth source of error in census-based estimates is caused by the difference in the number of migrants identified by previous residence and birthplace. Even though migrants can be classified as such by their residence five years prior to each census in most of the South American censuses, some of the censuses do not include the question on residence or they do not have the appropriate codes that enable labelling origins. As a result, migrants can only be identified by their birthplace. This occurs in the 1990 Venezuelan and 2010 Argentinean censuses.

Based on the previous sources of errors in census-based estimates, we take the one-year estimates obtained in Section 3.1.1 and correct them by (1) standardising flows to the most common census approach in South America (i.e. *de facto* approach), (2) removing biases due to the omission of infant migrants, (3) correcting errors due to migrant deaths and (4) adjusting estimates by the fact that some censuses only can identify migrants by their birthplace. The corrected flows are the true (unobserved) flows which refer to the number of migrants whose residence five years prior to each census is different from the census country. Equation (2) presents the relationship between the expected value of the non-corrected one-year flows $\mu^{(1)}_{ijt}$ and the true flows y_{ijt}:

$$\ln(\mu^{(1)}_{ijt}) \sim \text{Normal}\left(\ln(y_{ijt}) + \theta X_{1jt} + \omega X_{2ijt} + \kappa X_{3ijt} + \nu X_{4jt}, \sigma^2\right), \tag{2}$$

The term y_{ijt} corresponds to the one-year true migration flows deduced from censuses that assume the *de facto* approach, corrected for the omission of infant migrants, migrant deaths and when migrants can be identified by their previous residence instead of their birthplace. The variable $X_{1jt} = 1$ when the census is *de jure* and $X_{1jt} = 0$ otherwise. While X_{2ijt} refers to the number of children aged $0 - 4$ moving from origin i to census place j at year t, X_{3ijt} is the number of non-surviving migrants who moved from origin i to the destination j.The variable $X_{4jt} = 1$ when birthplace is used to categorise a migrant as such and $X_{4jt} = 0$ otherwise.

3.2 Autoregressive Distributed Lag (ADL) sub-model.

The second part of our Bayesian hierarchical model corresponds to an ADL sub-model. This part of our model aims at (a) imputing migration flows for the intercensal periods from 1986 to 2019, and (b) forecasting these flows until 2060.

3.2.1 Defining highly correlated variables to estimate and forecast migrant flows - the Shared Socioeconomic Pathways scenarios: When there is no observed census data (e.g. for the intercensal periods), there is the need to use auxiliary variables to estimate/forecast migration flows. To do this, previous studies have used variables that are highly correlated to migration. These variables are mainly associated with demographic and economic factors. In the current research, we consider two predictors to account for the demographic factor. The first one is dependency ratios (Lutz et al. 2018; KC and Lutz 2017). This covariate is selected based on Castro and Rogers (1984), who indicates that the number of migrants is highly sensitive to the fluctuations in dependency levels of the population at risk of migrating. We investigate the effect of the difference between the dependency ratio in the year when the transition between countries can be identified and its lagged version. We also analyse the effect of this difference in sending and receiving countries.

The second variable related to the demographic factor is the proportion of the population aged 25-34 with post-secondary education attainment (Lutz et al. 2018; KC and Lutz 2017). We define this variable to account for two matters. The first one is related to Sassone and del Castillo (2014)'s statement, in which the authors characterise migrants from Europe to South America as high-skilled individuals. Therefore, we consider the proportion of the population with post-secondary education attainment amongst those in the most mobile age groups. The second aspect that we seek to capture with this predictor is the fact that young adults are those who are more likely to migrate (Castro and Rogers 1984; Rogers and Castro 1981). Moreover, to account for the relative sizes of the population of each country, we decide on working with proportions instead of counts. Furthermore, we look for the effect of any increase or decrease in these proportions, for which we define the difference between the proportion of people aged 25-34 years old with post-secondary education and the lagged variable of it as a covariate. We define this difference for origins and destinations separately.

Regarding the economic factor, we establish the Gross Domestic Product per capita (GDPPP) as a variable (Cuaresma 2017). Income can be seen as a proxy of the economic possibility of migrating for people in the countries of origin and the potential earnings that a migrant might achieve in the countries of destination (Ortega and Peri 2013; Cornia 2011; Mayda 2010; Girma and Yu 2002). Specifically, we use the logarithmic transformation of the ratio of GDPPP in origins over the GDPPP in destinations. While the logarithmic transformation reduces the skewness of the GDPPP variable (Changyong et al. 2014), working with the ratio of GDPP enables capturing the relative benefits or constraints that a migrant requires to assess when decides on migrating to another country. In this regard, we include not only a covariate in the same year when the movement occurs but also a lagged predictor of this variable. This enables capturing also the effect of the time that the migrant needs to make his/her decision.

The variables of dependency ratios, GDPPP and the proportion of individuals with post-secondary education attainment are provided by IIASA and were calculated as part of the SSP scenarios (Lutz et al. 2014). The values of the SSP scenarios have been implemented not only to produce global analyses, but also regional and more local studies that address a broad range of demographic themes (e.g. from general topics such as population growth to more specific matters such as migration Reimann et al. 2021; Jiang et al. 2020; Frame et al. 2018). In addition to research, policymakers have been also widely using IIASA projections to inform their agendas (e.g. for the Intergovernmental Panel on Climate Change). So, employing the SSP values makes the current paper an opportunity to assess the versatility of these estimates.

The initial set of values used in the current research corresponds to the SSP2 scenario (Fricko et al. 2017). This scenario theorises that fertility, mortality and migration will continue having the same current medium trend (Lutz et al. 2018, 2014, pp.598). Later, in Section 6.3, we assess the impact of the SSP2 values on migration flows. To do this, we replace SSP2 data in the model explained in Section 3.2.2 with the SSP1 and SSP3 values. The SSP1 scenario assumes that there will be low mortality rates, leading by an improvement of health conditions and education attainment. In terms of fertility, the SSP1 values reflect the hypothesis about medium fertility rates in the richest countries of the Organisation for Economic Co-operation and Development (OECD) and high fertility rates in the rest of the countries, producing an acceleration of the demographic transition. Migration is assumed to be medium for all countries (van Vuuren et al. 2017; Lutz et al. 2014, pp.596-597). By contrast, the SSP3 scenario considers a world where mortality rates are still high, whereas fertility is low in the richest OECD countries and high in the rest of the globe. This scenario also assumes low education attainments and migration rates, but high mortality (Fujimori et al. 2017; Lutz et al. 2014, pp.598). Figure 3 shows the distribution of dependency ratios, GDPPP and the proportion of the population size aged 25-34 with post-secondary education attainment in each of the SSP scenarios. As it can be seen, the SSP1 values have a higher variance than the values of the SSP2 and SSP3 scenarios.

3.2.2 ADL model specification: Many studies have developed and assessed different methods to forecast migration (Bijak et al. 2019; Bijak 2010; Bijak and Wiśniowski 2010; Wiśniowski et al. 2015; Disney et al. 2015). Amongst these methods, Bijak et al. (2019) highlights the results obtained from using ADL models. According to this author, ADL models produce reasonable biases and calibration errors (i.e. $90\% - 100\%$ of observed values were within the 80% intervals estimated by using ADL models). The author also emphasises that this type of model can deal with not only stationarity time series but also non-stationarity data. This means that ADL models are useful for capturing steady changes and/or deterministic tendencies. Furthermore, ADL models allow the analysis of the effect of different covariates on migration occurring over time.

Based on the advantages that ADL models provide, we define an ADL model to impute and forecast migration flows from Europe to South America. Equation (3) displays the respective model specification, in which we extend the ADL models proposed by Bijak et al. (2019) and Cappelen et al. (2015), accounting for the variables described in Section 3.2.1.

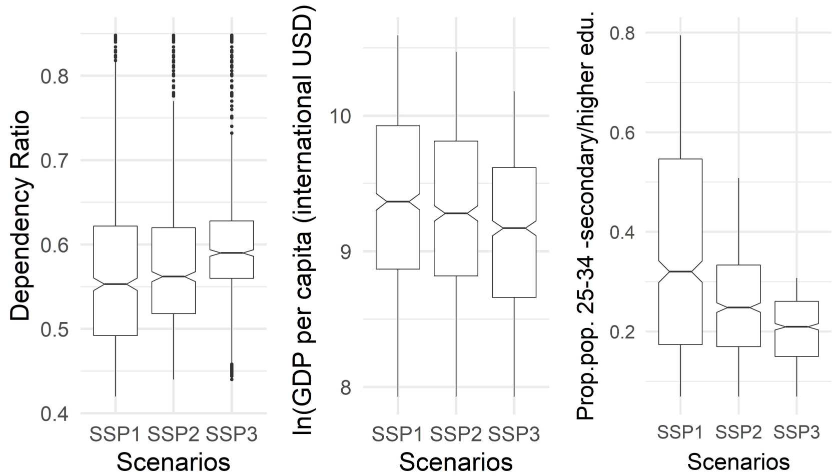

Figure 3. Box plots of dependency ratios, GDPPP and the proportion of the population size aged 25-34 with post-secondary education attainment per Shared Socioeconomic Pathways (SSP) scenario.

$$\begin{aligned}\ln(y_{ijt}) \sim \mathrm{Normal}\,(& u + b\, y_{ijt-1} + \\ & \beta_1(DR_{jt} - DR_{jt-1}) + \beta_2(DR_{it} - DR_{it-1}) + \\ & \beta_3(P_{25-34jt} - P_{25-34jt-1}) + \beta_4(P_{25-34it} - P_{25-34it-1}) + \\ & \beta_5\,\ln\left(\frac{G_{it}}{G_{jt}}\right) + \beta_6\,\ln\left(\frac{G_{it-1}}{G_{jt-1}}\right), \sigma_y^2\end{aligned} \tag{3}$$

While u is a constant, b is the parameter related to the use of the immediate previous flows $t-1$ to regress values at time t. The term β_1 refers to the difference between the dependency ratio and the lagged variable of it in destinations; β_2 captures the divergences between the dependency ratio and the lagged variable of it in origins; β_3 corresponds to the discrepancies between the proportion of people aged 25-34 years old with post-secondary education and the lagged variable of it in origins; β_4 is associated to the differences between the proportion of people aged 25-34 years old with post-secondary education and the lagged variable of it in destinations; β_5 relates to the natural logarithmic transformation of the ratio of the GDPPP in origins over the GDPPP in destinations; and β_6 concerns the lagged variable of the previous predictor. Whereas the variables based on demographic items (i.e. dependency ratios and proportions of the population size aged 25-34 with post-secondary studies) contribute to reducing the uncertainty of forecasting due to not being time-sensitive (Azose and Raftery 2015), the covariate regarding income can quantify abrupt changes (Kim and Cohen 2010). The term σ_y^2 indicates the variance of the true flows.

4 Results

4.1 Poisson-lognormal data sub-model.

4.1.1 Translating five-year data into one-year flows: Figure 4 compares the common logarithmic transformation of the observed one-year migration flows for censuses which provide this information (i.e. the 1991, 2000 and 2010 Brazilian, the 2005 and 2018 Colombian census, and the 2011 Uruguayan census) and the common logarithmic transformation of the translated one-year flows resulted from Section 3.1.1 with 95% Credible Intervals (CIs). If all observed flows were exactly mirrored by the translated flows, all values would fall on the 45-degree line in Figure 4. Nonetheless, not all of the means of the translated flows coincide with the observed data, but the 95% CIs of the first ones do cover the observed information. Uruguay resulted with the smallest differences between the observed and translated flows. These small discrepancies might be explained by the fact that migration from European-born individuals has constituted a big share of the Uruguayan inflows (Koolhaas et al. 2017). Therefore, assuming the same ratio of migration for foreign- and native-born is sensible.

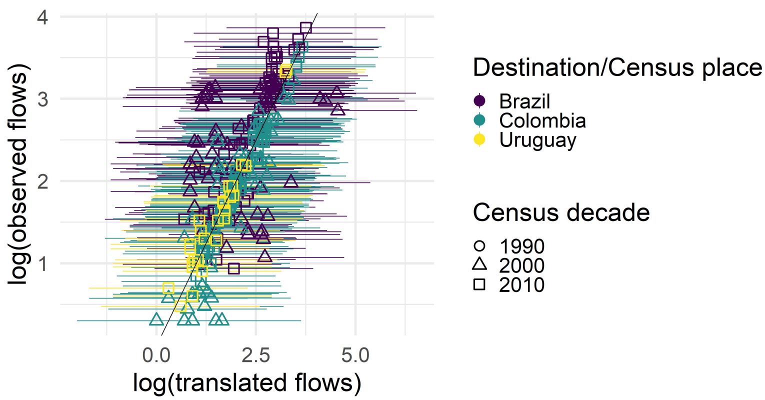

Figure 4. Common logarithmic transformation of the observed one-year migration flows for censuses which provide this information (i.e. the 1991, 2000 and 2010 Brazilian, the 2005 and 2018 Colombian census, and the 2011 Uruguayan census) and common logarithmic transformation of the translated one-year flows with 95% Credible Intervals (CIs).

By contrast, Brazil obtained the biggest differences between the observed and translated flows. In particular, the greatest gap in the translated flows resulted from the values based on the 2000 census, mostly from flows coming from Germany and Portugal. In the case of Germany, the predicted estimates are higher than the observed ones. Conversely, the translated flows from Portugal are smaller than the observed data. This variation amongst origins could be attributed to the composition of flows from Europe

Table 2. Parameters related to the correction of measurement errors in census data.

Parameter	Mean	SD	$q2.5\%$	Median	$q97.5\%$	psrf[1]	ESS[2]	MCSE[3]
θ (de jure)	-0.05	0.04	-0.17	-0.04	-0.002	1.00	180	0.003
ω (infant migrants)	-0.0007	0.0007	-0.0024	-0.0004	-0.00002	1.01	180	0.000
κ (non-surviving migrants)	-0.0005	0.0005	-0.0015	-0.0003	-0.00001	1.00	289	0.000
ν (native-born migrants)	-0.04	0.04	-0.16	-0.03	-0.003	1.04	180	0.003
σ^2	0.37	0.02	0.32	0.37	0.41	1.01	147	0.002

[1] Potential scale reduction factor.
[2] Effective Sample Size.
[3] Monte Carlo standard error $\left(MCSE = \frac{SD}{\sqrt{ESS}}\right)$ (Kruschke 2014, p.187).

to Brazil. While flows from Portugal may be due to foreign-born migrants, migration from Germany may result from returners.

In the case of Colombia, the translated flows replicate very well the observed values. The major differences occur in the estimates resulting from the 2005 census. In particular, the disagreements are in the values from Denmark, Poland and Sweden. However, the results suggest that assuming the same ratio for foreign and native-born flows from Europe to Colombia is well-founded.

4.1.2 Correction of measurement errors and biases in census data: Table 2 presents the parameters resulted from the corrections of the measurement errors of census data. The posterior of θ indicates that the *de facto* censuses reports, on average, $-(exp(-0.05) - 1) \cdot 100\% = 5.1\%$ less flows than *de jure* censuses, which means that true flows are smaller by the same percentage. Weeks (2016, p. 111) mentions that a possible explanation for this is the fact that these countries have few foreign workers coming from Europe.

The parameter ω expresses that for every 20% increase in the number of children aged 0-4 not captured by censuses, the annual flows estimates resulted from Section 4.1.1) are $(1.20^{-0007}-1) \cdot 100\% = -0.012\%$ smaller than the true flows. Similarly, the term κ denotes that the one-year translated flows are $(1.20^{-0.0005}-1) \cdot 100\% = -0.009\%$ smaller than the true flows for every 20% increase in the omission of non-surviving migrants in South American censuses. These results reveal that the error due to the omission of infant and non-surviving migrants in the one-year translated flows is marginal respecting the true flows. This agrees with some studies which have been characterising European migrants as single highly-skilled middle-aged individuals.

The parameter ν implies that censuses, in which migrants could be categorised as such by their residence (i.e. when both foreign- and native-born migrants could be defined), report $(exp(-(-0.04)) - 1) \cdot 100\% = 4.44\%$ more migrants than censuses, in which migrants could be identified only by their birthplace. These results provide a hint about the existing gap between the residence-based estimates and

Table 3. Autoregressive Distributed Lag (ADL) model parameters.

Parameter	Mean	SD	$q2.5\%$	Median	$q97.5\%$	psrf[1]	ESS[2]	MCSE[3]
β_1 (dependency ratio in j)	-0.56	0.88	-2.39	-0.38	0.85	1.03	210	0.05
β_2 (dependency ratio in i)	-0.42	0.91	-2.41	-0.38	1.03	1.06	245	0.05
β_3 (prop. 25-34 aged pop. with post-secondary education in j)	2.78	0.97	0.82	2.86	4.57	1.08	260	0.05
β_4 (prop. 25-34 aged pop. with post-secondary education in i)	1.26	0.74	-0.01	1.24	2.64	1.00	260	0.04
β_5 (GDPPP ratio in t)	-1.92	0.20	-2.33	-1.89	-1.62	1.07	174	0.02
β_6 (GDPPP ratio in $t-1$)	1.84	0.20	1.53	1.80	2.25	1.06	186	0.02
b	0.95	0.01	0.92	0.95	0.97	1.06	137	0.00
u	0.40	0.10	0.24	0.39	0.61	1.09	190	0.02
σ_y^2	0.54	0.02	0.51	0.54	0.58	1.09	260	0.00

[1] Potential scale reduction factor.
[2] Effective Sample Size.
[3] Monte Carlo standard error $\left(MCSE = \frac{SD}{\sqrt{ESS}}\right)$ (Kruschke 2014, p.187).

birthplace-based values. In the case of flows from Europe to South America, the effect of this difference may be considered negligible. However, the same type of divergences may be greater depending on the region of origin and destination (see Aparicio-Castro et al. (nd)).

4.2 Autoregressive Distributed Lag sub-model results.

Table 3 presents the posterior of the ADL model parameters. In absolute terms, the parameter β_3 holds the largest effect, which is associated with the difference of the proportion of people aged 25-34 years old with post-secondary education and the lagged variable of this predictor in origins. This covariate contributes to predicting a boost of $(1.2^{2.78} - 1) \cdot 100\% = 66.07\%$ in the true flows for every 20% increase in this difference.

The parameter β_5 relates to the natural logarithmic transformation of the ratio of the GDPPP in origins over the GDPPP in destinations. The mean of the posterior of $\beta_5 = 2.78$, indicating that an increment of 20% in this covariate predicts a decrease in the true flows of $(1.2^{-1.92} - 1) \cdot 100\% = -29.57\%$. This is sensible, considering that 99.95% of the ratios of the GDPPP in origins over the GDPPP in destinations are greater than 1. This implies that the GDPPP in origins are bigger than the GDPPP in destinations. This suggests that migrants see small financial benefits from moving to South America (Ratha and Shaw 2007, pp.17-19).

The term β_6 concerns to the lagged variable of the natural logarithmic transformation of the ratio of the GDPPP in origins over the GDPPP in destinations. It shows the effect of change of this covariate on estimating migration flows. In particular, a growth of 20% signifies that this variable predicts an increase in the true flows of $(1.2^{-1.84} - 1) \cdot 100\% = 39.88\%$.

In the case of the terms β_1, β_2 and β_4, the CIs contains 0. This means that neither the difference of the dependency ratio and the lagged variable of it in destinations and origins nor the growth of the proportion of people aged 25-34 years old with post-secondary education and the lagged terms of this variable in destinations makes a difference when imputing and forecasting migration flows.

5 The true flows

Figure 5 portrays the posterior means and the 95% CIs of the estimated and forecast corridor-specific migration flows from the 14 European countries selected in Section 2.1 to the 10 biggest South American countries.

5.1 The estimated true flows.

The estimated true flows comprehend the period from 1986 to 2019. The 95% CIs of these estimates contain approximately 84% of the non-corrected one-year migration flows derived in Section 3.1.1. The total number of migrants who moved from the selected European countries to South America across the period from 1986 to 2019 is around 4.6 million people. The trend of the true flows gradually increase from $86k$ migrants to a peak of $122k$ migrants in 1990. From 1991 to 1996, flows range between $94k$ and $104k$ migrants. In 1997, a jump happens, reaching $138k$ migrants. Then, the number of migrants plummet, hitting the lowest point of $83k$ migration flows in 2000. After 2000, flows started climbing gradually until 2008. At this point, the tendency of flows rocket from $101k$ to $237k$ in 2017 (the highest point of the true flows). After this, flows continue increasing moderately until 2019 when the estimates fall to $218k$.

The most active corridor from 1986 to 2019 is the Portugal-Brazil one with more than one million migrants. This agrees with the facts that (a) Brazil is the main destination of flows coming from Europe, hosting approximately 33% of them, and (b) Portugal is the country which contributes the biggest share of migrants towards South America (about 42% of migrants). Venezuela, Argentina and Chile have the second, third and fourth-largest proportions of all inflows from 1986 to 2019: 15.17%, 15%, and 13.22%, respectively. Bolivia is the least common destination. Only 1.63% of migrants arriving from Europe moved to Bolivia. This coincides with the fact that Bolivia integrates half of the 10% of the least active country-specific corridors from 1986 to 2019. In terms of origins, the UK contributes with the smallest proportion of migrants towards South America (1.26%). Sweden and Switzerland provide the second and third smallest percentages of migrants (1.46% and 1.40%, correspondingly).

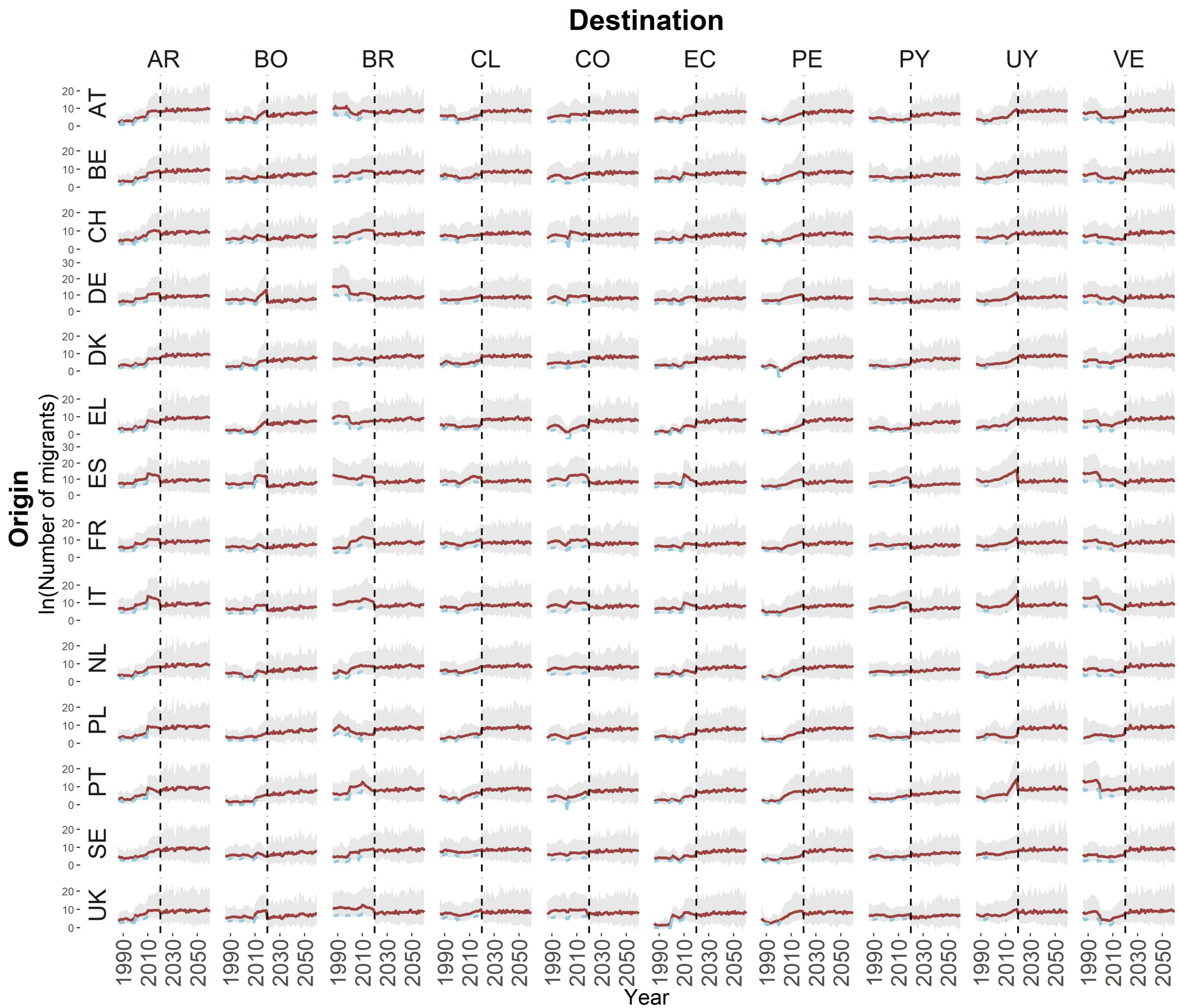

Figure 5. Natural logarithmic transformation of the estimated and forecast one-year (true) migration flows **(solid red line)** with 95% Credible Intervals (CIs) (gray area) from 14 selected European countries to South America from 1986 to 2060. **Blue solid line** represents the observed data. **Dashed vertical lines** refers to 2019, when the forecast migration flows start.

5.2 The forecast true flows.

The forecast true flows comprehend the period from 2020 to 2060. As it was expected, the 95% CIs of the forecast true flows are wider than the 95% CIs of the estimated values. On average, around 3.43 million ($3.23 - 4.8$ million) are expected to migrate from Europe to South America between 2020 and 2060. It is anticipated that, annually, the number of migrants oscillates between $48k$ and 1.3 million of individuals. The lowest number of persons moving from Europe to South America will be recorded in 2034 with $48k$ individuals, whilst the highest number of movers is forecast to be in 2045 with $130k$ migrants.

The inflows in Argentina, Venezuela and Chile (18.67%, 17.89%, and 15.20%, respectively) are expected to be highest. By contrast, Bolivia, Peru and Ecuador are likely to have the smallest percentages of migrants coming from Europe (2.12%, 2.31%, and 6.17%, correspondingly). In terms of origins, Poland, Germany and Denmark will contribute with more than 35% of migrants to South American countries. Belgium will provide the smallest share of flows (only 3.53%).

6 Sensitivity analysis

6.1 Assessment of estimated flows: comparative analysis

Since there is not a set of gold standard estimates to assess how sensible our resulting true flows are, we compare our estimated flows against the number of emigrants by country of next residence reported by Eurostat (see Figure (6)). We reduce the period of comparison from $1986 - 2019$ to $1998 - 2019$, for which Eurostat information is available. With a Pearson correlation coefficient of 0.73, we can say that both sets of values have a strong relationship, suggesting that they share similar trends. It must be highlighted that around 55% of Eurostat values are higher than the mean of our estimated flows. This can be attributed to the fact that data extracted from Eurostat correspond mainly to administrative data, which, most of the time, employs nationality as a criterion to define who is a migrant (Eurostat nd). By contrast, our estimates use transition data extracted from censuses, in which migrants are identified by their residence.

6.2 Assessment of forecast flows: partial removal data.

To assess the variation of the forecast values, we remove partially some input data and set a validation model with them. We remove data for the most recent censuses. They correspond to the 2017 Chilean, 2018 Colombian and 2017 Peruvian censuses (i.e. censuses after 2015). Then, we forecast the removed values and compare them with the estimates obtained from the original model whose input data was the full data set. Figure (7) portrays the scatter plot between the natural logarithmic of the true flows calculated with complete data and data with partially removed values. As it can be seen, not all the

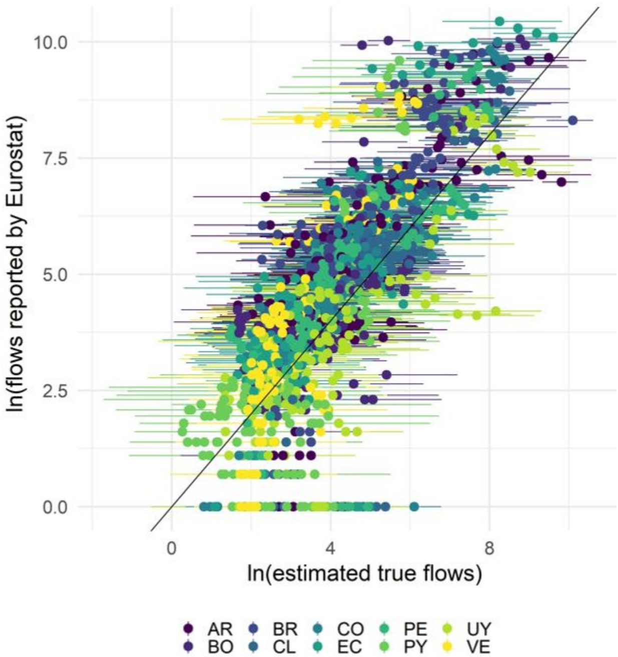

Figure 6. Natural logarithmic transformation of the number of emigrants by country of next residence reported by Eurostat and the estimated true flows with 95% Credible Intervals (CIs) per destination.

estimates obtained from the latter coincide with the values resulted from using the complete data set. Nevertheless, the interquartile ranges of both sets of values overlap and fall over the 45-degree line of the plot, indicating that their difference is not statistically significant.

6.3 Assessment of impact of Shared Socioeconomic pathways on estimated and forecast true flows.

The original model of the present paper uses the SSP2 values as input data. This scenario foresees that fertility, mortality and migration follow a steady path (Lutz et al. 2018, 2014). We assess the impact of employing the SSP2 values on the estimated and forecast true migration flows, replacing the SSP2 input data in our model with the SSP1 and SSP3 values. While the SSP1 scenario presupposes a rapid development of fertility and education, as well as low mortality rates (van Vuuren et al. 2017), the SSP3 scenario hypothesises a stalled demographic transition, in which the richest OECD countries would have population decrements and the developing countries would have an increase on their population sizes.

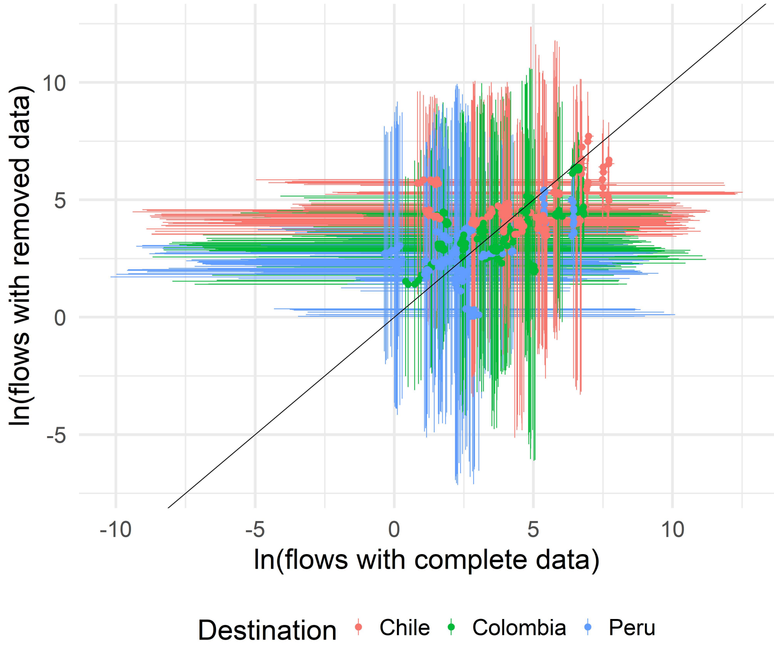

Figure 7. Natural logarithmic of the estimated and forecast true flows calculated with complete data and data with partially removed values. We delete data from the 2017 Chilean, 2018 Colombian and 2017 Peruvian censuses. The horizontal lines refer to the interquartile ranges for estimates gauged from the complete data. The vertical lines indicate the interquartile ranges for estimates computed from data with partially removed values. The intersect of the interquartile ranges is the mean of both sets of lines.

Education and migration are assumed to be low for all countries in the SSP3 scenario (van Vuuren et al. 2017; Fujimori et al. 2017; Lutz et al. 2014).

Figure (8) compares the natural logarithmic transformation of the estimated and forecast true flows based on the SSP1 and SSP2 assumptions. The vertical lines indicate the interquartile ranges for the SSP1-based estimates. The horizontal lines refer to the interquartile ranges of the original model that employs the SSP2 information. The intersect of the interquartile ranges is the mean of both sets of lines. The fact that most of the mean estimates are above the 45-degree line in Figure (8) implies that the SSP1-based results are greater than those obtained holding the SSP2 assumptions. Nonetheless, it must be highlighted that the CIs of both sets of estimates overlap and fall over the 45-degree line, which implies that the difference between them is negligible. In addition, the 95% CIs obtained from using the SSP1 data are wider than the SSP2-based CIs. This may be reflecting the fact that the SSP1 variables have a larger variance than the SSP2-scenario data (see Figure (3)). The Pearson correlation coefficient between the SSP1- and SSP2-based results is 0.70, suggesting a strong association between both sets of values.

Figure (9) presents the scatter plot between the natural logarithmic transformation of the estimated and forecast true flows based on the SSP1 and SSP2 with their respective 95% CIs. Similar to the previous case, the vertical lines indicate the interquartile ranges for the SSP3-based estimates, whilst the horizontal lines refer to the interquartile ranges of the original model. The correlation between the SSP2-based estimates and the SSP3-estimates is weaker ($\rho - 0.75$) than the correlation between the SSP2- and the SSP1-based values. This may be explain by the fact that the SSP3 scenario assumes that migration is low as opposed to the SSP1 and SSP2, in which migration maintain a medium trend.

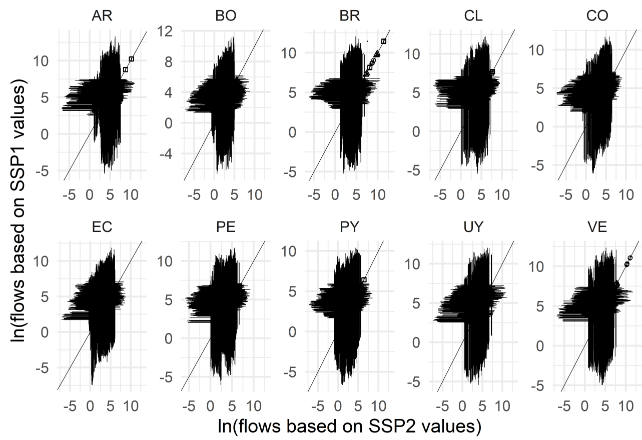

Figure 8. Natural logarithmic transformation of the estimated and forecast true flows based on Shared Socioeconomic Pathways (SSP) 2 and 1 scenario values with 95% Credible Intervals (CIs) per destination. The horizontal lines refer to the interquartile ranges of the SSP2-based resulted values. The vertical lines indicate the interquartile ranges for the SSP1-based estimates. The intersect of the interquartile ranges is the mean of both sets of lines.

We select the estimated and forecast true flows from Italy to Argentina from 1986 to 2060 to see in more detail the impact of the SSP scenarios on our results. In the case of the estimated true flows until 2019, Figure (10) shows that using any of the set of the SSP values does not affect much either the estimated flows or the respective 95% CIs. By contrast, the forecast values do have a variation depending on the set of projections used. While the average true flows do not differ from each other, the mean of the forecast values based on the SSP3 values is slightly smaller than the other sets of values. Additionally, the 95% CIs do seem to be affected by the set of SSP values used. The resulting estimates based on the SSP1 scenario have a higher variance than the values of the SSP2 and SSP3 scenarios. This may be

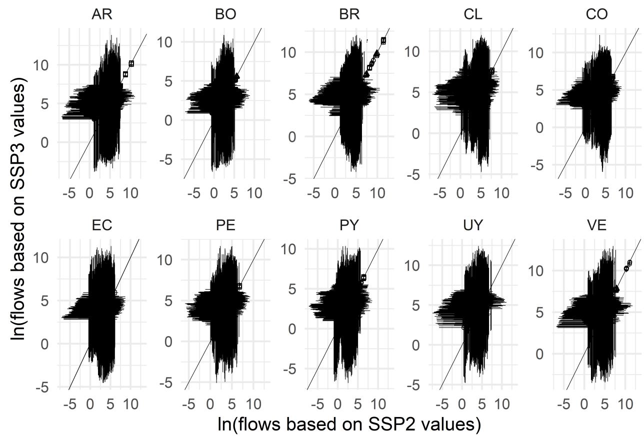

Figure 9. Natural logarithmic transformation of the estimated and forecast true flows based on Shared Socioeconomic Pathways (SSP) 2 and 3 scenario values with 95% Credible Intervals (CIs) per destination. The horizontal lines refer to the interquartile ranges of the SSP2-based resulted values. The vertical lines indicate the interquartile ranges for the SSP3-based estimates. The intersect of the interquartile ranges is the mean of both sets of lines.

explained by the fact that the SSP1 data have a higher variation than the one presented in the SSP2 and SSP3 scenarios (see Figure (3)).

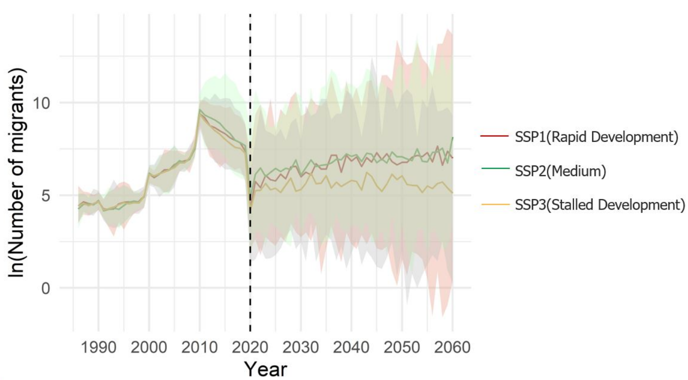

Figure 10. Natural logarithmic transformation of estimated and forecast true flows from Italy to Argentina from 1986 to 2060.

7 Concluding remarks

7.1 Contribution

We estimated a complete and consistent time series of annual migration flows from 14 selected European countries to the 10 biggest South American countries from 1986 to 2060. As opposed to UNDESA (2015) or Bijak et al. (2019), who gauged net migration or gross migration, we calculated bilateral migration flows. This was done by using census data, which enables overcoming cross-national differences in the definition of who constitutes a migrant, given that census data collection follow commonly shared guidelines.

Building a complete and consistent set of values was possible by implementing a two-part hierarchical Bayesian model. The first part of our model was able to translate five-year transition data into one-year flows, making five-year data with unmatched starting and ending years comparable. This first part of our proposed model also constitutes an alternative to deal with the five-/one-year problem that several studies have attempted to solve (e.g. Dyrting 2018; Rogers et al. 2010; Rogerson 1990; Kitsul and Philipov 1981; Rees 1977), providing measures of uncertainty.

Additionally, this first part of our model tackled four of the usual sources of errors in census-based estimates. In particular, our model addressed the biases produced by (a) the differences in the number of migrants that are registered in the *de jure* and *de facto* censuses, (b) the omission of infant migrants, (c) the unableness of censuses for capturing non-surviving migrants, and (d) the divergences in the number of migrants caused by the possibility of identifying them by their previous residence and birthplace.

The second part of our proposed hierarchical Bayesian model research (a) extends the ADL models used by Bijak et al. (2019) and Cappelen et al. (2015) for forecasting Europe-South America migration flows, evidencing that this type of model can handle not only stationarity time series but also non-stationarity data. Furthermore, the second part of our model used projections gauged by IIASA as part of the SSP scenarios (Lutz et al. 2018; KC and Lutz 2017; Cuaresma 2017; van Vuuren et al. 2017; Fricko et al. 2017; Fujimori et al. 2017), proving the versatility of these values.

The resulting true migration flows and the sensitivity analysis performed as part of this research indicates that our hierarchical Bayesian model produces sensible and consistent results. Besides this, the outcome of our sensitivity analysis provides a hint of the magnitude of the difference between estimates computed based on different types of data, in which migrants are identified, e.g., by their residence and/or birthplace.

7.2 Limitations and future work

The present research is mostly limited by the type of data used to estimate migration flows. Future work should integrate census migration data with other sources (e.g. administrative data) to counteract the biases produced by using census data. Moreover, the current study has only estimated the total flows. A further study could disaggregate flows by age, sex, education, and types of migrants (e.g. returners). This would allow the characterisation of who is leaving Europe towards South America, enabling the discussion of statements such as the Sassone and del Castillo (2014)'s one, who hypothesise that migrants from Europe to South America are mostly high-skilled individuals. Finally, further research could elicit expert opinion on specific aspects of the model, such as model parameters and inclusion of corridor-specific variables. This might improve the resulting estimates and forecast values.

Acknowledgements

Part of the research was developed during the Young Scientists Summer Program (YSSP) at the International Institute for Applied Systems Analysis (IIASA), Laxenburg (Austria) with financial support from the Economic and Social Research Council (ESRC) through the North West Social Science Doctoral Training Partnership (NWSSDTP). In particular, thanks to Dilek Yildiz, Michaela Potancokova and Arkadiusz Wiśniowski, who provided valuable inputs for this research and acted as my supervisors during the entire Program.

Funding

This work was supported by the Economic and Social Research Council through the North West Social Science Doctoral Training Partnership (NWSSDTP) grant number ES/P000665/1. This research would not have been possible either without the funding provided by the NWSSDTP to undertake an Overseas Institutional Visit (OIV) at the International Institute for Applied Systems Analysis (IIASA) in Austria.

References

Aparicio-Castro, A., A. Wiśniowski, and F. Rowe (n.d.). A bayesian approach to estimate annual bilateral migration flows for south america using census data. *Submitted to Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Azose, J. J. and A. E. Raftery (2015). Bayesian probabilistic projection of international migration. *Demography 52*(5), 1627–1650.

Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view*, Volume 24. Springer Science & Business Media.

Bijak, J., G. Disney, A. M. Findlay, J. J. Forster, P. W. Smith, and A. Wiśniowski (2019). Assessing time series models for forecasting international migration: Lessons from the united kingdom. *Journal of Forecasting 38*(5), 470–487.

Bijak, J. and A. Wiśniowski (2010). Bayesian forecasting of immigration to selected european countries by using expert knowledge. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 173*(4), 775–796.

Bryant, J. and J. L. Zhang (2018). *Bayesian demographic estimation and forecasting*. Chapman and Hall/CRC.

Cappelen, Å., T. Skjerpen, and M. Tønnessen (2015). Forecasting immigration in official population projections using an econometric model. *International Migration Review 49*(4), 945–980.

Castro, L. J. and A. Rogers (1984). What the age composition of migrants can tell us. *Population Bulletin of the United Nations 1983*.

CELADE (2019). Redatam7.

Changyong, F., W. Hongyue, L. Naiji, C. Tian, H. Hua, L. Ying, et al. (2014). Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry 26*(2), 105.

Cornia, G. A. (2011). Economic integration, inequality and growth: Latin america versus the european economies in transition. *Review of Economics and Institutions 2*(2).

Cuaresma, J. C. (2017, jan). Income projections for climate change research: A framework based on human capital dynamics. *Global Environmental Change 42*, 226–236.

Czaika, M., H. Bohnet, and A. Soto-Nishimura (2021). Spatial and categorical dependence of european migration flows.

Disney, G., A. Wiśniowski, J. J. Forster, P. W. Smith, and J. Bijak (2015). Evaluation of existing migration forecasting methods and models.

Dyrting, S. (2018). A framework for translating between one-year and five-year migration probabilities.

Eurostat, T. S. O. o. t. E. U. (n.d.). International migration statistics. reference metadata in euro sdmx metadata structure (esms).

Frame, B., J. Lawrence, A.-G. Ausseil, A. Reisinger, and A. Daigneault (2018). Adapting global shared socio-economic pathways for national and local scenarios. *Climate Risk Management 21*, 39–51.

Fricko, O., P. Havlik, J. Rogelj, Z. Klimont, M. Gusti, N. Johnson, P. Kolp, M. Strubegger, H. Valin, M. Amann, T. Ermolieva, N. Forsell, M. Herrero, C. Heyes, G. Kindermann, V. Krey, D. L. McCollum, M. Obersteiner, S. Pachauri, S. Rao, E. Schmid, W. Schoepp, and K. Riahi (2017, jan). The marker quantification of the shared socioeconomic pathway 2: A middle-of-the-road scenario for the 21st century. *Global Environmental Change 42*, 251–267.

Fujimori, S., T. Hasegawa, T. Masui, K. Takahashi, D. S. Herran, H. Dai, Y. Hijioka, and M. Kainuma (2017, jan). SSP3: AIM implementation of shared socioeconomic pathways. *Global Environmental Change 42*, 268–283.

Girma, S. and Z. Yu (2002). The link between immigration and trade: Evidence from the united kingdom. *Weltwirtschaftliches Archiv 138*(1), 115–130.

Hinde, A. (2014). *Demographic methods*. Routledge.

Jiang, L., B. C. O'Neill, H. Zoraghein, and S. Dahlke (2020). Population scenarios for us states consistent with shared socioeconomic pathways. *Environmental Research Letters 15*(9), 094097.

Juran, S. and R. C. Snow (2018). The potential of population and housing censuses for international migrant analysis. *Statistical Journal of the IAOS 34*(2), 203–213.

KC, S. and W. Lutz (2017, jan). 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.

Kim, K. and J. E. Cohen (2010). Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity. *International migration review 44*(4), 899–932.

Kitsul, P. and D. Philipov (1981). The one-year/five-year migration problem. In *Advances in multiregional demography*, Chapter 1, pp. 1–33. RR-81-006.

Koolhaas, M., M. Nathan, et al. (2017). Inmigrantes internacionales y retornados en uruguay.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Lutz, W., W. P. Butz, and K. e. Samir (2014). *World population and human capital in the twenty-first century*. OUP Oxford.

Lutz, W., A. Goujon, S. Kc, M. Stonawski, and N. Stilianakis (2018). *Demographic and human capital scenarios for the 21st century: 2018 assessment for 201 countries*. Publications Office of the European Union.

Mayda, A. M. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics 23*(4), 1249–1274.

McAuliffe, M. and B. Khadria (2019). World migration report 2020.

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

Ratha, D. and W. Shaw (2007). *South-South migration and remittances*. The World Bank.

Raymer, J., A. Wiśniowski, J. Forster, P. Smith, and J. Bijak (2013). Integrated modeling of european migration. *Journal of the American Statistical Association 108*(503), 801–819.

Rees, P., M. Bell, O. Duke-Williams, and M. Blake (2000). Problems and solutions in the measurement of migration intensities: Australia and britain compared. *Population Studies 54*(2), 207–222.

Rees, P. H. (1977). The measurement of migration, from census data and other sources. *Environment and Planning A 9*(3), 247–272.

Reimann, L., B. Jones, T. Nikoletopoulos, and A. T. Vafeidis (2021). Accounting for internal migration in spatial population projections—a gravity-based modeling approach using the shared socioeconomic pathways. *Environmental Research Letters 16*(7), 074025.

Rodríguez-Vignoli, J. and F. Rowe (2018). How is internal migration reshaping metropolitan populations in Latin America? A new method and new evidence. *Population Studies 72*(2), 253–273.

Rogers, A. and L. J. Castro (1981). Model migration schedules. *International Institute for Applied Systems Analysis*.

Rogers, A. and L. Jordan (2004). Estimating migration flows from birthplace-specific population stocks of infants. *Geographical Analysis 36*(1), 38–53.

Rogers, A., J. Little, and J. Raymer (2010). *The indirect estimation of migration: Methods for dealing with irregular, inadequate, and missing data*, Volume 26. Springer Science & Business Media.

Rogers, A., J. Raymer, and F. Willekens (2003). Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer 55*(1), 56–69.

Rogerson, P. A. (1990). Migration analysis using data with time intervals of differing widths. In *papers of the Regional Science Association*, Volume 68(1), pp. 97–106. Springer.

Sassone, S. M. and I. Y. del Castillo (2014). Crisis global y el sistema migratorio europa-américa latina/global crisis and the europe-latin america migration system. *revista cidob d'afers internacionals*, 13–38.

Shryock, H., J. Siegel, et al. (1976). The methods and materials of demography. *Condensed Edition by Stockwell EG*.

Siegel, J. and D. Swanson (2004). The methods and materials of demography.

Swanson, D. A. and J. Tayman (2011). On estimating a de facto population and its components. *Review of Economics and Finance 7*.

UNDESA (1998). Recommendations of statistics of international migration, revision 1. *Statistical Papers M*(58).

UNDESA (2008). Principles and recommendations for population and housing censuses, revision 2. *Statistical Papers M*(67).

UNDESA (2015). *International Migration Flows to and from Selected Countries: The 2015 Revision*. UN.

UNDESA (2017). Principles and recommendations for population and housing censuses, revision 3. *Statistical Papers M*(67).

UNDESA (2020). International migrant stock 2020: Destination.

van Vuuren, D. P., E. Stehfest, D. E. Gernaat, J. C. Doelman, M. van den Berg, M. Harmsen, H. S. de Boer, L. F. Bouwman, V. Daioglou, O. Y. Edelenbosch, B. Girod, T. Kram, L. Lassaletta, P. L. Lucas, H. van Meijl, C. Müller, B. J. van Ruijven, S. van der Sluis, and A. Tabeau (2017, jan). Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm. *Global Environmental Change 42*, 237–250.

Villa, M. (1991). Introducción al análisis de la migración: apuntes de clase: notas preliminares.

Weeks, J. R. (2016). *Population: An introduction to concepts and issues* (12th ed.). Cengage Learning.

Wiśniowski, A., P. W. Smith, J. Bijak, J. Raymer, and J. J. Forster (2015). Bayesian population forecasting: extending the lee-carter method. *Demography 52*(3), 1035–1059.