Journal of the Royal Statistical Society Series A: Statistics in Society, 2023, **00**, 1–26
https://doi.org/10.1093/jrsssa/qnad127

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS — Journal of the Statistics Society Series A Statistics in Society

Original Article

A Bayesian approach to estimate annual bilateral migration flows for South America using census data

Andrea Aparicio Castro[1], Arkadiusz Wiśniowski[1] and Francisco Rowe[2]

[1]Social Statistics Department, University of Manchester, Manchester, UK
[2]Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool, Liverpoool, UK

Address for correspondence: Andrea Aparicio Castro, Social Statistics Department, Humanities Bridgeford Street, University of Manchester, Manchester M13 9PL, UK. Emails: andrea.apariciocastro@manchester.ac.uk; alaparicioc@gmail.com

Abstract

Censuses are an important source of international migration flow data. However, their use is limited since they indirectly reflect migration, capturing migrant transitions over long intervals rather than migration events, whilst also underestimating the number of infants and deaths. Censuses also neglect migration of those who are native-born when they only include questions on country of birth, and have sparse temporal availability. We propose a Bayesian hierarchical model to overcome these limitations and produce a set of robust annual migration flow estimates for South American countries. Our model translates five-year transition data from censuses into annual series, corrects biases that arise due to differences in measurement and census data quality across countries, and is grounded in migration theory to impute missing migration data between censuses.

Keywords: Bayesian inference, census data, demographic accounting, international migration, migration flows, South America

1 Introduction

World population growth has decreased monotonically over the last six decades, with South American countries accounting for some of the steepest declines (UNDESA, 2022). Based on the United Nations (UN) estimates, between 1950 and 2021, the average annual growth rate in South America fell from 2.68% to 0.58% (UNDESA, 2022). This was driven by a drop in the regional total fertility rate from 5.6 to 1.81 children per woman and a rise in average life expectancy at birth from 50 to 72.96 years (UNDESA, 2022).

Under this scenario of declining fertility and population ageing, South American governments have acknowledged the importance of international migration as the main factor which attenuates population decline. South American governments have committed to the *Montevideo Consensus on Population and Development* (ECLAC, 2013a) and have continuously emphasised the importance of international migration during the *Regional Conferences on Population and Development in Latin America and the Caribbean* (ECLAC, 2013b, 2015, 2019a, 2022). National governments have also promoted multilateral initiatives (e.g. the *Southern Common Market*—its Spanish acronym is MERCOSUR, 2023), which have sought to stimulate migration as part of broader efforts to promote economic growth.

Received: June 30, 2023. **Accepted:** September 23, 2023

In this context, data on international migration flows are essential. They allow (1) assessing how national policies and multilateral initiatives impact migration; (2) measuring of the evolving migration diasporas and patterns between origins and destinations (Willekens, 2008, p. 119; Rogers et al., 2002, p. 32); (3) understanding to be gained regarding key driving factors and motivations influencing individual migration decisions (De Beer, 2008, p. 292–302; Bilsborrow et al., 1997, p. 36–39); and (4) assessing the impacts of migration on national population size and structure (Willekens et al., 2016, p. 897).

Relatively few studies have attempted to estimate migration flows for South America. Perhaps, the most important endeavour to date has been the *International Migration Research in Latin America* project (IMILA, by its Spanish acronym) led by the *Economic Commission for Latin America and the Caribbean* (ECLAC) of the UN (ECLAC, 2020; Moya, 1993). ECLAC routinely estimates migration flows for South American countries based on census data and focuses on estimating the number of migrants over five-year intervals (labelled as recent migrants, Vargas-Silva, 2013, pp. 142–143).

Three key limitations can be identified in the ECLAC estimates. First, they rely on a deterministic demographic accounting method to calculate migration flows, which involves building contingency tables of recent migrants without accounting for random variation across countries, time or data sources. This means that they do not provide measures of uncertainty of their resulting estimates. Second, ECLAC exclusively uses country of birth as the basis for identifying origins, rather than country of previous residence, thereby neglecting native-born and return migration. Third, ECLAC only accounts for migrants from origin countries whose migrant stocks are more than 500 people, omitting small flows (ECLAC, 2020).

In addition to IMILA, migration estimates for South American countries were included in a set of global estimates produced by Abel and Cohen (2019), who reviewed six methods to gauge five-year bilateral migration flows for 200 countries based on census migration stock data. Abel (2018, p. 821) and Abel (2013) proposed a key method that produces estimates which rely on the assumption that the estimated flows are equal to the minimum number of transitions that match the difference in migrant stocks measured at two points in time. Azose and Raftery (2019, p. 116) used a pseudo-Bayesian method which relaxes the ‘minimum migration’ assumption of Abel (2013, 2018) and included multiple types of migrants in their estimates (e.g. returnees).

These estimates continue to comprise five-year migrant transition data. Unlike event data, which directly capture migrations, transition data capture migrants. Transitions reflect migration indirectly by comparing places of residence at two different points in time, undercounting migrations (events, Willekens, 2008, p. 119). The longer the transition interval, the larger the undercount (Rees et al., 2017, p. 4; Rees et al., 2000; Rowe et al., 2019). Thus, the five-year estimates of Abel and Cohen (2019) and Azose and Raftery (2019) exhibit smaller flows and lower variability than actual changes in flows captured by events data. Estimates that refer to shorter transition intervals, such as one year, should reduce the negative bias of using transition data to estimate flows.

A common feature defining the limitations of previous studies is the use of census data. Censuses are a key source of migration data due to their completeness, reliability, and comparability (Bryant & Zhang, 2018, p. 186; Juran & Snow, 2018; Rodríguez-Vignoli & Rowe, 2018). However, using censuses to estimate annual migration flows leads to (1) using five-year data that result from the most common census question on previous residence, which usually refers to five-year periods, and whose initial and final years do not necessarily match amongst censuses for various countries; (2) measurement errors or biases that arise because a migrant is categorised as such if their previous residence is different from the census place, an indication that migration is captured indirectly; and, (3) missing data for the intercensal periods, for which census data are not available.

This study overcomes the challenges that estimating migration flows raises by relying on five-year flow (transition) census data, rather than data on country of birth (or migrant stock data) that were the main data source in other studies. From a theoretical perspective, information on the place of residence enables the use of a fixed-interval measure that provides additional detail and more consistent measurements of migration (Newbold, 2001).

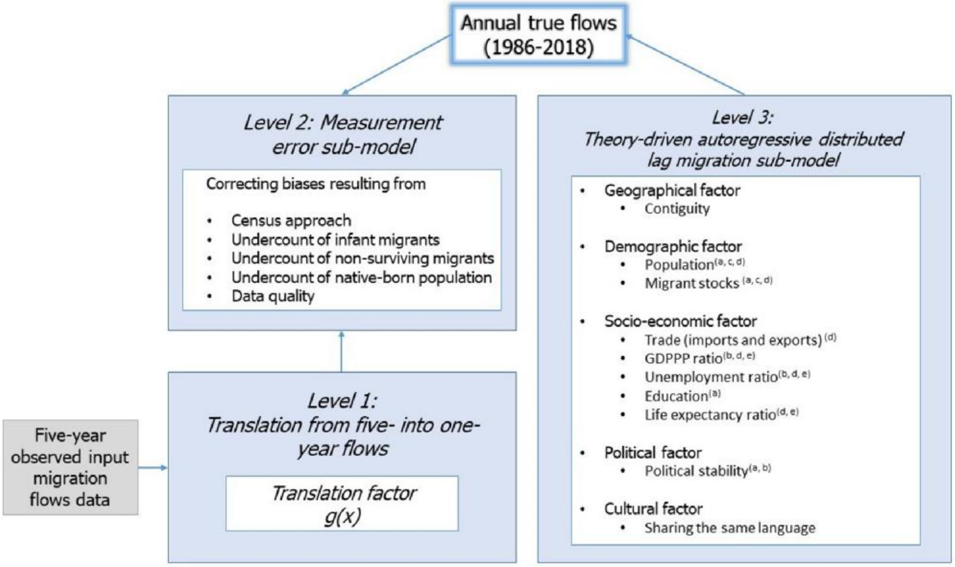

Figure 1. Modelling framework to estimate annual bilateral migration flows from five-year transition data extracted from censuses. Translation factors are based on five-year migrant stocks and one-year number of foreign-born individuals whose first arrival to census places was at most four years prior to each census (i.e. $x = 0, 1, 2, 3, 4$). (a) A variable is defined independently for origins and destinations; (b) both variables at time t and lagged term are included in the model; (c) only the lagged term is included in the model; (d) natural logarithm transformation is used; (e) ratio is calculated as the value in origins over values in destinations. GDPPP (Gross Domestic Product per person). Source: Authors' own work.

We develop a three-level Bayesian hierarchical model to estimate consistent and robust annual bilateral migration flows (see Figure 1) for the 10 largest South American countries, using five-year census transition data.[1] The first level of our Bayesian hierarchical model translates five-year transition census data into one-year transition data. While the resulting one-year estimates may be subject to a translation error, they are valuable for three reasons. First, one-year estimates may reflect changes that cannot be captured in five-year data. For example, five-year estimates could not mirror the pattern of those migrants who left Venezuela between 2015 and 2019 due to the Venezuelan socio-economic crisis but returned to their country in the context of COVID-19 in 2020 (López et al., 2020). Second, one-year estimates can be more valuable than five-year data for assessing the short-term effects of changes to migration policies. A relatively recent example is a resettlement scheme that was created by the UK government for Hong Kong residents that generated unanticipated effects, but could only be analysed by data that referred to short periods (e.g. a year, see Graham-Harrison, 2020). Third, five-year migration flows are not comparable with other demographic and socio-economic data, which are usually reported annually (Kitsul & Philipov, 1980).

The second level of our Bayesian hierarchical model (see Figure 1) is a measurement error sub-model which is similar to the data model in Raymer et al. (2013). In this sub-model, observed data are assumed to map onto the true but unobserved migration flows; that is, the true migration flows with some noise due to errors or biases. In this paper, our measurement error sub-model corrects one-year flows that result from the first level of our model by standardising them to the most common census approach in South America (i.e. *de facto* approach) and by reducing the influence of biases due to the omission of infant migrants,

[1] This paper illustrates the potential to effectively predict annual migration flows from five-year transition data extracted from censuses, capturing corridor-specific variations. South America is the case study used for this purpose. We expect to expand our method with data from other regions. This work has already been undertaken under the FUME (Future Migration Scenarios for Europe) initiative, which is a project funded by the European Union Horizon 2020 (Grant ID 870649), in which country-to-country migration flows between South America and Europe are estimated using our method.

migrant deaths, the migration of native-born populations, as well as differences in census data quality quantified from census availability, coverage and the use of a unique questionnaire.

The third level of our Bayesian hierarchical model in Figure 1 is a theory-driven migration sub-model; analogous to the migration model in Raymer et al. (2013). This sub-model imputes missing migration flows for years during intercensal periods, i.e. when census data are unavailable. The third level of our proposed method corresponds to an Autoregressive Distributed Lag (ADL) sub-model with repeated measures that uses variables highly correlated with migration.

The resulting estimates from our three-level Bayesian hierarchical model are of the true (unobserved) international migration flows amongst countries, in which a migrant is tacitly defined as a person whose country of residence at the date of the (given) census differs from their place of residence at least one year before arriving at the census place. Following Raymer et al. (2013), the term 'true flows' is used in reference to the latent (unobserved) flows, and for which measures of uncertainty are computed.

The remainder of this paper is organised into five sections. In Section 2, we present the input data and the challenges that using census data imply, that is, the five-year/one-year problem, census-specific biases and a lack of information in the intercensal period. In Section 3, we specify the three-level Bayesian hierarchical model, which seeks to tackle the challenges identified in Section 2. Section 4 presents the results of our model. In Section 5, we describe the estimated annual bilateral migration flows and compare them against other estimates. Section 6 provides a summary of the work and reflections on this paper's contribution to the field of migration studies.

2 Background

2.1 Data

We consider a migration system composed of 10 destinations j and 18 origins i. Destinations refer to Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, and Venezuela. Our selection of destinations was limited by the availability of micro-census data. We selected countries with (1) the biggest population sizes; (2) the largest number of native-born living abroad; and (3) the highest number of foreign-born individuals living in their territories. According to UNDESA (2022), the total population of the chosen destinations covers more than 90% of the total South American inhabitants. Additionally, the selected countries represented 96% of the South American international migrant stocks (UNDESA, 2020c) and hosted 98.2% of the international migrants who lived in the region in 2020 (UNDESA, 2020a).[2]

Origins consist of all the destinations plus territories which have had intensive migration exchange with South America, that is, North America (which embodies the USA and Canada) and Spain. Flows from these origins were considered independently of their respective continents and were prioritised over other potential countries since North America and Spain (1) have long-established histories of migrant exchange with South American countries and (2) remain as significant arriving territories for South American migrants (De Haas et al., 2022). According to UNDESA (2020c), North America has been the main destination of South American migrant stocks outside South America since 1990. Indeed, the FEM & IOM (2022, 2023) prioritise the analysis of migration from and to the USA and Canada over other territories in the Americas. In the case of Spain, UNDESA (2020c) shows that Spain has been the main origin of migrant stocks to South America since 2010. The National Institute of Statistics of Spain (Instituto Nacional de Estadística, 2023) confirms this, and identifies South America as a major contributor to the country's international pool of migrants, and especially Colombia, Ecuador, Peru, and Venezuela.

We grouped all other origin countries or the rest of the world into continents: America, Africa, Asia, Europe, and Oceania, and assumed a closed population system (Carmichael, 2016, p. 21). Migrants from the rest of the world to South America represented 29.6% of the total recent

[2] For the rest of the paper, the term *South America* refers to the set of the 10 chosen countries. Guyana, Suriname, dependent territories (i.e. the Falkland Islands–UK, South Georgia and South Sandwich Islands–UK) and internal territories (i.e. French Guiana–France) were excluded from the analysis because of two reasons. First, their census data either lacked the question of place of residence and/or the labels of origins were not available or accessible at the moment of the extraction of the data. This, in turn, made it impossible to identify the origins of migration flows. Second, flows from and to these territories can be considered to be negligible. This appreciation is based on their share of stocks (UNDESA, 2020b).

Table 1. Year and type of censuses in South America

Country	Decade					
	1990s	Type of census	2000s	Type of census	2010s	Type of census
Argentina	1991	De facto	2001	De facto	2010	De facto
Bolivia	1992	De facto	2001	De facto	2012	De facto
Brazil	1991	De jure	2000	De jure	2010	De jure
Chile	1992	De facto	2002	De facto	2017	De facto
Colombia	1993	De jure	2005	De jure	2018	De jure
Ecuador	1990	De facto	2001	De facto	2010	De facto
Paraguay	1992	De facto	2002	De facto	–	–
Peru	1993	De facto[1]	2007	De facto	2017	De facto
Uruguay	1996	De facto	–	–	2011	De jure
Venezuela	1990	De jure	2001	De jure	2011	De jure

[1] Despite that this census is *de facto*, neither the census questionnaire nor the database contain a question regarding usual residences, impeding the identification of usual residents. Source: Authors' own work.

migrants reported in the census data in our destination countries. We do not consider migration amongst the rest of the world's continents.

Migration flow data can be seen as two-level hierarchical data. We have migration corridors on the top level (level 2). A *migration corridor* comprises an origin i and a destination j, where $i \neq j$. In total, there are 170 migration corridors ($(18-1)\times 10$).[3] We have repeated measures per each migration corridor in the bottom level (i.e. level 1) of the two-level flow data structure. We drew on census micro-data of the destinations covering the period from 1986 to 2018 (33 years), from 28 censuses (see Table 1). Data availability constrained the analysis to this time frame, and yielded a data set of 5, 610 entries (i.e. 170 corridors×33 years). See Table 1 and Supplementary Material (SM) A for details about the data. The data were extracted using `Redatam7` (ECLAC, 2019b) and `Redatam+SP` (ECLAC, 2019c).[4]

We identify migrants by comparing their country of residence at the census date as well as that five years prior to the census (so-called transition flows), and captured origins and destinations as well as different types of migrants (e.g. returnees, see Villa, 1991, p. 25). With these data, migrants can be identified regardless of their country of birth (Vargas-Silva, 2013, pp. 142–143). Information on their usual place of residence, when available (see Table 1), was used to remove non-usual residents, e.g. visitor population. Additionally, information on countries of birth and years of arrival was used to translate five- into one-year transition data (as detailed in Section 2.2).

2.2 Translating from five- to one-year transition data

Census data are a key source of migration data because of their completeness, reliability and comparability (Bryant & Zhang, 2018, p. 186; Juran & Snow, 2018; Rodríguez-Vignoli & Rowe, 2018). Nonetheless, most censuses only provide information on recent migrants, i.e. the number of people whose place of residence at the census date was different from their residence five years prior to the census. As mentioned in Section 1, it is more convenient to work with one-year data than five-year flows due to the higher value of the former.

It follows that converting five- into one-year flows not only improves the usefulness of the census-based estimates but also enables the differences between five- and one-year data to be quantified; thereby providing a better understanding of migration and the effects of how these demographic

[3] Destinations, census places, census countries, and receiving countries are used interchangeably in this document. Additionally, origins and sending countries are equivalent in this study.

[4] `Redatam7` is software which enables the extraction of data from censuses, national household surveys, and administrative records of Latin America and the Caribbean countries (ECLAC, 2019b). `Redatam+SP` is the online version of `Redatam7` (ECLAC, 2019c).

events are reflected and measured in migrant data (Nowok, 2010; Nowok & Willekens, 2011). However, translating from five-year to one-year interval data is a challenging task. This is because dissimilarities between the interval data are driven by various factors (Kitsul & Philipov, 1980; Rees, 1977; Rogers et al., 2010; Rogers, Raymer & Newbold, 2003), with return and onward migration being the key sources of differences (Dyrting, 2018; Rogerson, 1990).

SM B shows the proportion of first-time movers, returnees, and onward migrants estimated by using data extracted from censuses for which both five- and one-year information is available: Brazil 1991, Brazil 2000, Brazil 2010, Colombia 2005, Colombia 2018, Uruguay 2011. First-time movers are migrants whose country of birth and residence one or five years prior to the (given) census were the same, but different from the given census (and residence) country. Returnees are migrants whose country of birth differed from their residence country one or five years before the census, but was the same as the census country at the census date. Finally, onward migrants are migrants whose country of birth and residence one or five years prior to the census and census place at the census date were all dissimilar.

In general, differences in the numbers of first-time movers and onward migrants for one- and five-year intervals were small, with the latter larger than the foremost. Out of the total number of migrants registered in the Brazilian, Colombian and Uruguayan censuses, the percentage of first-time movers captured via the question on residence one year prior to a census was 34.4%, close to the 37.1% of first-move five-year migrants. Onward one- and five-year migrants' shares were 5.95% and 6.89% in the censuses, respectively. In contrast, the divergences in numbers of five- and one-year returnees tended to be larger. Returnees represented 61.6% of the one-year migrants and 58% of the five-year movers. For some specific origins, the differences were larger. As an illustration, the 1990 Brazilian census shows that returnees from Ecuador to Brazil between 1986 and 1990 were more than five times greater than returnees between 1989 and 1990 from the same origin.

Rogers et al. (2010) assessed the influence of age profiles who, conditional on survivorship, shaped differences in five- and one-year migration data. They showed, for example, that five-year migration age schedules displayed a consistent and rapid decline in the propensity of migration at around age 10. By contrast, one-year migration age schedules exhibited a more gradual fall during the 20s. These disparities may respond to the different motives that individuals have to migrate to their destinations. It is likely that a five-year migrant had moved between countries due to a family reunion, whereas a one-year migrant had changed his/her country of residence due to work. These differences are not reflected in South American data. For instance, the 2010 Brazilian census shows that there was not much disparity in the ages of five- and one-year migrants, although the counts of migrants did vary depending on the transition interval. This suggests that the motives captured by both one-year and five-year migrants in South America may not differ as much as in other regions.

Discrepancies between five- and one-year migration can also be seen spatially (Kitsul & Philipov, 1980; Rogers et al., 2010; Rogers, Raymer & Newbold, 2003). In the case of South American countries, the origins that contribute the largest and the smallest number of migrants to destinations are similar for the five- and one-year intervals. For example, the 2018 Colombian census shows that the first four origins with the highest number of migrants residing outside Colombia one or five years prior to the respective census date were, in order, Venezuela, the USA, Spain, and Ecuador. In the one- and five-year flows, the share of migrants was approximately 87% from Venezuela, 2.2% from the USA, and 1.8% from Spain. In the case of Ecuador, the percentage changed from 1.4% in the five-year flows to 1.01% in the one-year flows. From the fifth main origin, the rank of other sending countries and/or their size do vary. While five-year migrants showed -in order-Argentina, Chile, Mexico, Brazil, and Panama as the next five important sending countries, the order of one-year migrants' origins was: Chile, Argentina, Mexico, Brazil, and Panama.

All previous studies have established and exploited the relationship that exists between five- and one-year data to translate five-year data into one-year estimates. We could not establish this relationship for all South American countries. Indeed, this was only possible for six out of 28 South American censuses, because not all censuses contain questions about residence for both five and one year prior to a census date. We, therefore, implemented a different approach, explained in Section 3.1.

2.3 Migration measurement errors in censuses

The fact that censuses do not measure migration directly, i.e. by counting events, results in multiple measurement errors. We use the concept of *error* in this paper to refer to the (unknown) differences that exist between the (non-corrected) one-year flows and true (unobserved) flows. Errors include not only systematic biases but also random terms (Buonaccorsi, 2010; Gustafson, 2003).

A first source of error in census-based estimates arises from differences in migrant counts based on *de jure* and *de facto* censuses. *De jure* censuses count legal residents, and may imply the capturing of individuals who are not present on census day as well as the exclusion of unauthorised migrants (Swanson & Tayman, 2011, p. 7). In contrast, *de facto* censuses count every person present on census day, which may result in not enumerating usual residents who are absent from their country of residence (Siegel & Swanson, 2004, p. 49), and may also include non-resident population such as visitor population, seasonal population, and daytime population (Swanson & Tayman, 2011). In practice, the latter is less problematic as *de facto* censuses usually contain a question on the actual country of usual residence. Table 1 reveals predominantly *de facto* censuses in our sample of South American countries (18 *de facto* censuses vs. 10 *de jure* censuses).

Existing literature on measuring the effects of *de facto* and *de jure* census approaches on migration has focused on three aspects: (1) subnational populations (Swanson & Tayman, 2012, pp. 105–114, 313–330), (2) specific groups of migrants (e.g. foreign-born, Judson & Swanson, 2011); and, (3) the effect of only one of the two approaches on population estimates (Rayer, 2015; Swanson & Tayman, 2011). In our study, we measure the effects of the difference between the *de jure* and *de facto* approaches at a national level.

A second source of error is the omission of infant and non-surviving migrants. Rees et al. (2000, p. 208) define infant migrants as those individuals who were born in a different country than the census place and have migrated there within the transition interval of interest. Since the question about residence five years prior to each census was used to capture migrants in this study, infant migrants refer to children under five years old. Typically, infant migrants are estimated based on one-year interval data (Bell et al., 1999; Raymer & Rogers, 2007; Rees et al., 2000). However, such data are not available for all South American countries. To take into account infant migrants, we assumed that migrant stocks of persons aged 0–4 were equal to the migration flow in the same age group, i.e. infant migrants were all children aged 0–4 whose country of birth differed from the census place (Rogers & Jordan, 2004, p. 42; Rogers, Raymer & Willekens, 2003, p. 56). We thus imply and assume that all children aged 0–4 have migrated only once during their lifetime.

As previously stated, transition census data also neglect non-surviving migrants, i.e. individuals who were alive and migrated to a census place during the transition interval of interest but died before the census date (Hinde, 2014; Rees, 1977, p. 251). Non-surviving migrants have been incorporated into our flows estimates by using data on deaths from different sources (e.g. population registers), and demographic accounting methods to control for deaths in migration flows (e.g. Abel & Cohen, 2019; Azose & Raftery, 2019; White & Van der Knaap, 1985). We correct flows for non-surviving migrants by using annual data on migrant deaths deduced from the Demographic Books 1986–2018 published by the UN Demographic Yearbook System (UNSD, 2020). The total number of migrant deaths in census country j from origin i at year t, denoted by $d_{ijt}^{(1)}$, is equal to the total deaths $D_{jt}^{(1)}$ in the destination j and year t times the proportion resulting from the number of migration flows per origin, $z_{ijt}^{(1)}$ divided by the total population $P_{jt}^{(1)}$, that is:

$$d_{ijt}^{(1)} = D_{jt}^{(1)} \cdot \frac{z_{ijt}^{(1)}}{P_{jt}^{(1)}}. \tag{1}$$

The superscript '(1)' indicates that these values are annual information. Notice that we assumed that the mortality rate of migrants arriving from origin i is the same in all age groups, given that we do not disaggregate flows by age. The resulting data on both infant migrants and migrant deaths has been standardised by dividing by two standard deviations (SD). Following Gelman (2008, p. 2871), these variables are first logged and then standardised to put them on a common scale and to enable the interpretation of the coefficients as low or high values.

For the 1990 Venezuelan and 2010 Argentinian censuses, the origins of recent migrants cannot be specified. Instead, we used the country of birth as their origins as in ECLAC (2020). This implies

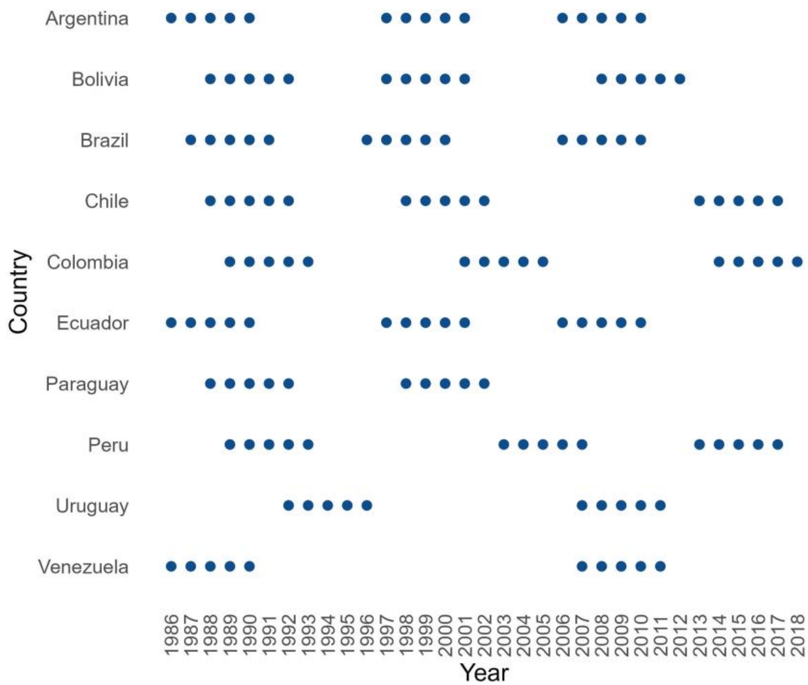

Figure 2. Five-year periods to which the South American census questions on residence five years prior to each census refer. Source: Authors' own work.

that, for these two censuses, only foreign-born recent migrants could be identified, but not the native-born. For the rest of the South American censuses, the number of both foreign-born and native-born recent migrants could be established.

Migration flow estimates can also be biased due to differences in census data quality (Jensen, 2013, p. 11–12). We assessed the quality of census data based on the UN recommendations on availability, coverage, and the use of a unique questionnaire (UNDESA, 1998, 2008, 2017). Details of the criteria and assessment method are reported in SM C. Data quality was evaluated for each country. Bolivia, Ecuador, and Venezuela were classified as having very good census data quality. Argentina, Brazil, Chile, and Peru were categorised with good quality, whereas Colombia, Paraguay, and Uruguay were assessed as being fair quality. Following previous research (Del Fava et al., 2020; Kupiszewska & Wiśniowski, 2009; Raymer et al., 2013; Wiśniowski et al., 2016), we used census quality as a classification parameter in the model that determines the accuracy of the data (see Section 3.2).

2.4 Migration flows in the intercensal periods

UNDESA (1998, 2008, 2017) recommends carrying out censuses every 10 years to balance the need for census data and the complex logistics involved in collecting them. This gap leads to missing data in the intercensal periods. Figure 2 shows the years for which information on migration in census data is not available for our South American countries. The percentages of missing annual data vary between 54.5% and 69.7%. Given this, the imputation of missing data is necessary, and we do so by using a theory-driven migration sub-model (see Section 3.3) that corresponded to an ADL specification with repeated measures that use variables highly correlated with migration.

3 Methods

3.1 Translating five-year intervals into a one-year interval sub-model

The first level of our hierarchical model translates five-year census data into one-year transition data. To achieve this, we followed a similar approach to the method proposed by Rogers, Raymer and Newbold (2003, p. 593), which converted data on interregional migration flows for the USA and Canada. Rogers, Raymer and Newbold (2003) assumed that five-year interval migrants $z_{ijt}^{(5)}$ from origin i registered in destination j in year t equate to one-year interval migrants $z_{ijt}^{(1)}$ multiplied by an inflation factor (aka re-scaling factor), $f(\cdot)$:

$$z_{ijt}^{(5)} = z_{ijt}^{(1)} \cdot f(\cdot). \tag{2}$$

Unlike Rogers, Raymer and Newbold (2003), we do not have observed data on flows for $z_{ijt}^{(1)}$ or for building a re-scaling factor $f(\cdot)$. Instead, South American censuses provide observed five-year flows $z_{ijt}^{(5)}$ and observed stocks of migrants S_{ijt} born in country i and living in destination (census country) j. Resultantly, we assumed an analogous re-scaling (or translation) factor $g(x)$, the expected value of which is based on the ratio of the observed total number of five-year foreign-born migrants registered at the census place $S_{ijt}^{(5)}$ and the number of foreign-born individuals $S_{ijt-x}^{(1)}$, whose first arrival to j was at most four years prior to each census (i.e. at time $t-x$ where $x = 0, 1, 2, 3, 4$). From a theoretical perspective, and following Bengochea and Saucedo (2018, p. 55), the year of the arrival of foreigners enables knowing the entries of migrants, and is a good proxy for migration flows. As presented in Equation 3, we defined that our translation factor $g(x)$ follows a Normal distribution truncated to negative values, and its corresponding precision is defined as $\tau_g = \frac{1}{\sigma_g^2}$, where $\sigma_g^2 \sim \Gamma(10, 2)$.

$$g(x) \sim \text{Normal}_+\left(\frac{S_{ijt}^{(5)}}{S_{ijt-x}^{(1)}}, \tau_g\right) \qquad \text{for} \quad x = 0, 1, 2, 3, 4. \tag{3}$$

We assumed that our observed five-year flows $z_{ijt}^{(5)}$ follow a Poisson distribution with an expected and unobserved value $\mu_{ijt}^{(5)}$ (see Equation 4). The Poisson distribution allows estimating (1) non-negative figures and (2) right-skewed values with high-density mass for small counts but very low flows at the positive end of the distribution.

$$\begin{gathered} z_{ijt}^{(5)} \sim \text{Poisson}\left(\mu_{ijt}^{(5)}\right) \\ z_{ijt}^{(5)} \sim \text{Poisson}\left(\mu_{ijt-x}^{(1)} \cdot g(x)\right) \qquad \text{for} \quad x = 0, 1, 2, 3, 4. \end{gathered} \tag{4}$$

The expression $\mu_{ijt}^{(5)}$ is the expected value of the observed five-year flows $z_{ijt}^{(5)}$. The term $\mu_{ijt-x}^{(1)}$ —or $\mu_{ijt}^{(1)}$ as referred to in the rest of this paper—alludes to the translated and uncorrected flows, which are estimated and are the input for our measurement error sub-model (see Section 3.2). The term $\mu_{ijt-x}^{(1)}$ is the expected value of the unobserved $z_{ijt}^{(1)}$ and reflects the fact that the translated flows come from time-specific means instead of a common mean. The expression $g(x)$ is our translation factor based on migrant stocks, in which x is the number of years since the migration occurred. For $g(x)$, data on stocks were used and extracted from South American censuses, which ask foreign-born individuals for their year of arrival in the country of destination. For countries that do not have data on migrant stocks or years of arrival, we followed Newell (1988, p. 150). That is, we assumed the ratios of counter flow. For example, suppose that the census of Argentina does not have data on the year of arrival for migrants whose country of birth is Bolivia. In such an instance, the same translation factor as for the Bolivia–Argentina flow was used for the Argentina–Bolivia flow. SM D illustrates our derivation of translation factors with synthetic data.

3.2 Measurement error sub-model

This section presents the second level of our Bayesian hierarchical model. This level corresponds to a measurement error sub-model, and is analogous to the one proposed by Raymer et al. (2013). Our measurement error sub-model corrects for discrepancies in censuses in measuring migration (Section 2.3) by standardising values to the most common census approach used in South America (i.e. *de facto* perspective), and removes biases due to the omission of infant migrants, migrant deaths, and flows of native-born, as well as quantifying differences in census data quality.

We assumed a lognormal sub-model for the expected one-year interval counts of migrants, $\mu_{ijt}^{(1)}$ (Equation 5), to estimate the true (unobserved) migration flows y_{ijt} from origin i to destination j in a given year t. The lognormal distribution ensures non-negative migration flow estimates and allows the overdispersion in the variance of the estimated migrant counts (Willekens, 2008, p. 124). The relationship between one-year interval migration flows and true flows is then

$$\ln\left(\mu_{ijt}^{(1)}\right) \sim \text{Normal}\left(\ln(y_{ijt}) + \theta X_{1jt} + \omega X_{2ijt} + \kappa X_{3ijt} + \nu X_{4jt},\ \tau_{\beta_{q(j)}}\right). \tag{5}$$

We assumed that the expected value of the log-transformed translated flows ($\ln(\mu_{ijt}^{(1)})$) corresponds to the log-transformed true migration flows y_{ijt}, corrected for errors or biases, as explained in Section 2.3. That is, y_{ijt} refers to one-year translated migration flows inferred from *de facto* data, and corrected for the omission of infant migrants, migrant deaths, or native-born flows; represented by the other additive components in Equation 5. We seek to estimate y_{ijt}, and provide corresponding measures of uncertainty for them.

Variable $X_{1jt} = 1$ when a census in the destination assumes a *de jure* approach; $X_{1jt} = 0$ otherwise. X_{2ijt} and X_{3ijt} represent the standardised log-transformed annual counts of infant and non-surviving migrants, respectively. $X_{4jt} = 1$ when the values observed in the census neglect native-born population flows (i.e. in the 1990 Venezuelan and the 2010 Argentinian censuses); $X_{4jt} = 0$ otherwise. The term $\tau_{\beta_{q(j)}}$, where $q(j) = 1, 2, 3$, refers to the precision of the errors of each data quality group (very good quality, good, and fair quality, respectively).

SM E presents assumptions made on prior distributions for each parameter in Equation 5 and discusses how they were defined. Since there is no 'gold standard' benchmark data used in the sub-model, we ensured the parameters' identification and convergence by truncating some of them. The truncation was based on the belief that certain biases can produce that translated flows $\mu_{ijt}^{(1)}$ are smaller than the true flows y_{ijt}. For example, translated flows were expected to neglect infant migrants and, therefore, were smaller than the true flows. Consequently, the truncation, for example, for ω, was placed in the positive numbers of the distribution.

3.3 Theory-driven migration sub-model

As stated in Section 2.4, censuses capture data on migration over a fixed interval from a specific point in the past to the census date. Therefore, data gaps exist between census years. To create a complete time series of data points, the third level of our Bayesian hierarchical model (see Figure 1) imputes missing data. In so doing, we employed a theory-driven migration sub-model that corresponded to an ADL model with repeated measures of order 1 that uses variables highly correlated with migration. This sub-model is analogous to the migration model in Raymer et al. (2013). SM F.1 indicates the data sources. We specify our migration sub-model as:

$$\begin{aligned}
\ln(y_{ijt}) \sim \text{Normal}(&u_{0ij} + u_{1ij}y_{ijt-1} + \alpha_1 B_{ijt} + \alpha_2 \ln(P_{jt-1}) + \alpha_3 \ln(P_{it-1}) \\
&+ \alpha_4 \ln(S_{ijt-1}) + \alpha_5 \ln(S_{jit-1}) + \alpha_6 \ln(Im_{ijt}) \\
&+ \alpha_7 \ln(Ex_{jit}) + \alpha_8 \ln\left(\frac{G_{it}}{G_{jt}}\right) + \alpha_9 \ln\left(\frac{G_{it-1}}{G_{jt-1}}\right) \\
&+ \alpha_{10} \ln\left(\frac{U_{it}}{U_{jt}}\right) + \alpha_{11} \ln\left(\frac{U_{it-1}}{U_{jt-1}}\right) \\
&+ \alpha_{12} E_{jt} + \alpha_{13} E_{it} + \alpha_{14} \ln\left(\frac{LE_{it}}{LE_{jt}}\right) \\
&+ \alpha_{15} PSV_{jt} + \alpha_{16} PSV_{jt-1} + \alpha_{17} PSV_{it} + \alpha_{18} PSV_{it-1} \\
&+ \alpha_{19} L_{ijt},\ \tau_y),
\end{aligned} \tag{6}$$

where y_{ijt} refers to the true migration flows as in Equation 5, and τ_y is the precision of it, for which $\tau_y = \frac{1}{\sigma_y^2}$, where $\sigma_y \sim \Gamma(3, 2)$. The terms $u_{0ij} \sim \text{Normal}(0, 12)$ and $u_{1ij} \sim \text{Normal}(0, 12)$ reproduce a two-level data structure that enables corridor-specific intercepts and autoregressive terms, respectively. As a supra level or at level 2, we have 170 corridors, which correspond to the number of all origin-destination pairs ij where $i \neq j$. For each corridor, we had repeated measures, i.e. the true flows were estimated across 33 years from 1987 to 2018 with 1986 being the baseline year, that is, $t = 0, 1, 2, 3, \ldots, 32$. In total, we had 5,610 observations in our level 1: 33 years ×170 corridors.

The regression coefficients $\alpha_1, \ldots, \alpha_{19}$ are regression coefficients of our predictors, and include geographical, demographic, socio-economic, political, and cultural factors. The geographical factor is the contiguity between pairs of countries (or belonging to the same continent for grouped origins) B_{ijt}. Contiguity is considered to be a proxy for 'distance' between countries, which is one of the key terms in the gravity models (Bertoli & Moraga, 2017, p. 72; Sen & Smith, 1995). Given that gravity models are sensitive to the assumed measure of distance (Alimi et al., 2015), contiguity was used as a variable to reflect the fact that intraregional flows represent a large proportion of the South American flows that occur between border cities/towns (Pagnotta, 2014). In addition, we wanted to avoid highly sensitive outcomes resulting from the adoption of a specific distance measure (Alimi et al., 2015). For example, if distances between capital cities in South America were taken into account, then flows from Brazil/Chile to Venezuela may be affected. The distance between Brasilia (Brazil) and Caracas (Venezuela) is greater than the distance between Brasilia and Santiago de Chile (Chile, see Mayer & Zignago, 2011). Nonetheless, Brazil, and Chile have higher intensities of migration. Whilst Brazil and Chile do not have a border in common, Brazil, and Venezuela share a border of 2,199 km (Mangabeira & Montilla, 1928; Pereira Leal & Sanojo, 1859).

Demographic factors include the lagged terms of population size at origins P_{it-1} and destinations P_{jt-1}, bilateral migrant stocks S_{ijt-1} born in i and living in j, and migrant stocks S_{jit-1} born in j and living in i. The size of the population at the origins captures the population at risk of migration (Hinde, 2014, p. 193), while the size of the population at the destinations captures the combined effect of increased emigration and return flows leading to high gross migration (Kim & Cohen, 2010, p. 902; Newell, 1988, p. 84). Migrant stocks served as a proxy for migrant networks in destinations and a source of migrants returning from their origins (Mayda, 2010, p. 1253). We used lagged terms for variables related to populations to reduce endogeneity (Abdallah et al., 2015; Arestis et al., 2007).

Socio-economic factors are represented by imports Im_{ijt} and exports Ex_{jit} at census countries j; the log ratio of the Gross Domestic Product per person (GDPPP) at sending countries G_{it} over GDPPP in receiving countries G_{jt}, and its lagged term; the log ratios of the unemployment rate at origins U_{it} over the same rate at destinations U_{jt}, and the immediately previous value; mean years of schooling at sending E_{it} and receiving countries E_{jt}; the log ratio of life expectancy at origins LE_{it} over destinations LE_{jt} (for rationale see, e.g. Hatzigeorgiou, 2010, p. 13; Girma & Yu, 2002, pp. 123–124; Ortega & Peri, 2013, pp. 59–62; and Mayda, 2010, pp. 1268–1270). We used ratios of GDPPP, unemployment rates and life expectancy to capture the assessment that migrants undertake when comparing their conditions in their origins and destinations. We also included lagged terms of predictors associated with GDPPP and mean years of schooling to account for delayed effect of the changes in these covariates.

Political factors include the Political Stability and Absence of Violence (PSV) Index for origins PSV_{it} and destinations PSV_{jt}. Hatton (2009), Melander and Öberg (2006, pp. 145–146) and Neumayer (2005, pp. 394–396) show evidence of increased violence being linked to higher emigration. By contrast, inflows, e.g. of displaced persons, tend to increase in countries recording political stability and low levels of violence. We also added the lagged term of these variables to consider the deferred effect of violence on migration.

To measure the influence of cultural factors, we included a binary variable L_{ijt}, which captures whether the origin and destination countries share their respective first official language (Adserà & Pytliková, 2015, pp. 71–72; Ginsburgh & Weber, 2016, p. 343; Özden & Schiff, 2007, p. 42).

Continuous predictors are standardised by centring and dividing by two SDs as suggested by Gelman (2008). For continents or grouped origins, we used the sum of the population in sending countries i and receiving countries j, migrant stocks born in i and living in j, migrant stocks born in

j and living in i, imports and exports. For the rest of the continuous variables, we used their mean. We imputed missing information on predictors, given that the migration sub-model required complete information (see SM F.2). The prior for $\alpha_p \sim \text{Normal}(0, 0.5)$ where $p = 1, \ldots, 19$. Terms $u_{0ij} \sim \text{Normal}(0, 12)$ and $u_{1ij} \sim \text{Normal}(0, 12)$ (see SM G for sensitivity analysis to the assumed prior distributions).

Unlike other studies such as Sorichetta et al. (2016), we decided to use a wide range of variables to impute our true migration flows. We have employed non-time-sensitive (e.g. population sizes) and time-sensitive variables (e.g. income) to impute our migration flows. While the first group reduces the uncertainty of estimating flows (Azose & Raftery, 2015), the second set of covariates can quantify abrupt changes (Bijak et al., 2019; Kim & Cohen, 2010). Similar to this study, existing research has used a similar, or wider set of factors to model migration flows (e.g. Abel, 2010; Jennissen, 2004; Raymer et al., 2013), since migration is a complex and volatile process. SM G assesses the number of covariates in the theory-driven Autoregressive Distributed lag model.

4 Model results

This section presents the results of our three-level Bayesian model, the parts of which were fitted jointly in order to tackle and address the problems of the census-based estimates described in Sections 2.2, 2.3, and 2.4. The posterior characteristics of the parameters of the proposed hierarchical model were obtained by using MCMC implemented in JAGS software, which is a cross-platform program for Bayesian analysis (Plummer, 2003). JAGS code and data on the estimated true migration flows can be found in Aparicio-Castro et al. (2023). Convergence was assessed visually and numerically, using the diagnostics available in the package *coda* (Plummer et al., 2006, 2020). Further in this section, we discuss the results of all three sub-models.

4.1 Translating from five- to one-year migration sub-model results

Figure 3 compares the natural logarithmic transformation of country-to-country one-year observed migration flows z_{ijt} obtained from censuses which provide one-year flows (i.e. the 1991

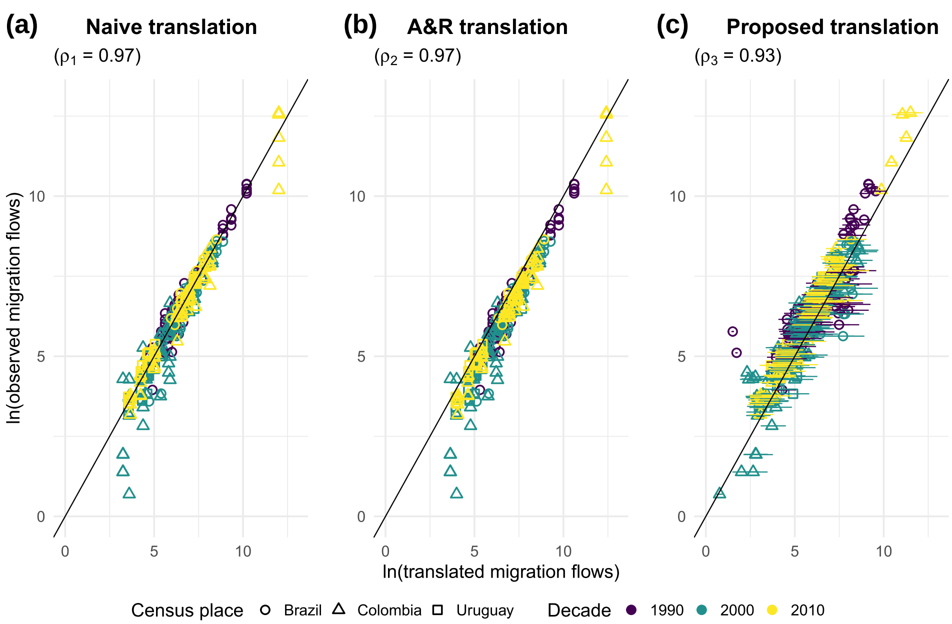

Figure 3. Comparison between the natural logarithmic transformation of one-year observed migration flows z_{ijt} obtained from censuses which provide one-year flows (i.e. the 1991 Brazilian, 2000 Brazilian, 2010 Brazilian, 2005 Colombian, 2018 Colombian and 2011 Uruguayan censuses), and the natural logarithmic transformation of the one-year migration flows calculated based on (**a**) a naive translation (i.e. assuming that $z^{(1)}_{ijt} = z^{(5)}_{ijt}/5$), (**b**) the translation approximation proposed by *A&R* (Azose & Raftery, 2019, i.e. assuming a Long–Boertlein index equal to 1.5), and (**c**) the mean resulted from implementing our proposed translation. Source: Authors' own work.

Brazilian, 2000 Brazilian, 2010 Brazilian, 2005 Colombian, 2018 Colombian, 2011 Uruguayan censuses), and the natural logarithmic transformation of the one-year migration flows calculated based on three methods. In Figure 3a, we show the migration flows resulting from a naive translation, i.e. assuming that five-year migration data are five times higher than annual data ($z_{ijt}^{(1)} = z_{ijt}^{(5)}/5$). Figure 3b depicts the migration flows obtained from assuming a Long–Boertlein index equal to 1.5 as in *A&R* (or Azose & Raftery, 2019, see Supporting Information in pp. 9–10) to account for the difference between one- and five-year transitions. Figure 3c displays the translated flows calculated from our proposed method.

Following Abel and Cohen (2019), the comparison of the logarithmic transformation enables exploration of whether the patterns of each pair of sets of estimates are similar to the trend of the observed flows. If the translated flows coincided exactly with the observed flows, then all values would fall on the 45-degree lines in Figure 3a, b, and c. Overall, the trend of the one-year ratios based on the year of the arrival of foreigners follows the pattern of the observed one-year migration flows obtained from the censuses, which have five- and one-year data (Brazil 1991, Brazil 2000, Brazil 2010, Colombia 2005, Colombia 2018 and Uruguay 2011). However, the correlation coefficient for our proposed translation is the lowest Pearson coefficient ($\rho_3 = 0.93$). Despite ρ_3 being the smallest coefficient, it still shows a strong correlation with the observed data.

In addition to correlation coefficients, we computed the difference between the input data and the resulting flows using the naive, Azose and Raftery (2019), and our proposed translation. On average, flows obtained from the naive translation show 53 fewer migrants ($SD = 13,685$) than the observed values. Migration flows gauged by adopting Azose & Raftery's (2019) assumption resulted in a mean difference of 2, 026 fewer migrants ($SD = 17,197$) than the input data. The mean of our translated flows shows, on average, 2, 079 more migrants ($SD = 16,779$) than censuses that provide one- and five-year flows. Our calculated 95% Credible Intervals (or $CIs_{95\%}$) captured 73% of the 330 observed census values we use for the current assessment of our proposed translation. We obtained a mean precision $\tau_g = 1.89$ ($SD = 0.65$) with $CIs_{95\%}(1.04, 3.49)$ that reached convergence. SM G presents the sensitivity analysis to the assumed prior distributions.

Generally, this paper's approach is more accurate when it comes to predicting extreme values. This is beneficial as it prevents the smoothing out of sudden changes caused by unforeseen events, such as the political and economic crises that may lead to migration. This information is useful for those who use estimates of migration flows, for example, it may enable policy makers to make more informed decisions. As an example, the proposed method could more accurately estimate the flow of people from Venezuela to Colombia (as seen by the yellow triangles in Figure 3c) and provide a better indication of the impact of migrants on Colombian territory (ICG, 2022).

The greatest disparities are seen in the data from earlier censuses, such as the number of migrants from Canada recorded in the 1991 Brazilian census (the two outermost circles on the left side of the 45-degree line in Figure 3c). These data are likely to be of lower quality than more recent censuses, as census data collection processes have improved over time.

The response rate for questions regarding the year of first arrival was a factor in determining the quality of the translation method. Countries that provide five- and one-year migration flows have higher response rates. All Brazilian censuses had a perfect response rate of 100%. The Colombian census response rate improved from 91% in 2005 to 95.5% in 2018, while the 2011 Uruguayan census had a response rate of 98%. Nevertheless, the average response rate among the 28 South American censuses during the study period was 79.4% ($SD = 17.4\%$).

Despite the discrepancies between the observed and the resulting mean estimated one-year migration flows, the proposed method is an improvement over existing methods for three reasons. First, it maximises the utility of one of the most reliable and common data available worldwide, transition census data, by converting five-year flows into annual values. This reduces the effect of long time intervals in transition data on the estimation of migration flows (Kupiszewska & Wiśniowski, 2009; Weeks, 2016). Second, the proposed method can capture country-specific fluctuations in flows within each five-year period rather than averaging them. This allows for the capture of extreme flows, such as those moving from Venezuela to Colombia during the 2013–2018 period, and provides measures of uncertainty. Third, the proposed translation method is based on the dynamics of international migration rather than internal migration, as in Azose and Raftery (2019), Rogers, Raymer and Newbold (2003), and Newbold (2001).

Table 2. Measurement error sub-model parameters using census data

Parameter	Mean	*SD*	*q*2.5%	Median	*q*97.5%	*psrf*[1]	ESS[2]	MCSE[3]
θ (de jure)	−0.56	0.16	−0.87	−0.56	−0.25	1.00	3265.30	0.00
ω (infants)	−0.003	0.003	−0.010	−0.002	−0.0001	1.00	5580.44	0.00
κ (deaths)	−0.002	0.002	−0.006	−0.001	−0.00004	1.00	6000.00	0.00
ν (native-born migrants)	0.94	0.13	0.69	0.94	1.19	1.00	6000.00	0.00
τ_{β_1} (very good quality)	5.96	1.12	4.26	5.80	8.54	1.00	3041.12	0.02
τ_{β_2} (good quality)	3.54	0.39	2.86	3.51	4.38	1.00	3227.84	0.01
τ_{β_3} (fair quality)	3.83	0.47	3.02	3.78	4.90	1.00	4090.25	0.01

[1] Potential scale reduction factor. [2] Effective Sample Size. [3] Monte Carlo standard Error defined as $MCSE = \frac{SD}{\sqrt{ESS}}$ in Kruschke (2014, p. 187). Source: Authors' own work.

4.2 Measurement error sub-model results

Table 2 presents the results of the measurement error sub-model. The posterior of θ indicates that censuses that assume the *de facto* concept registers, on average, $-(exp(-0.56)-1)\times 100\% =$ 43.0% smaller flows than censuses that adopt the *de jure* notion. This means that the true flows are smaller than the translated flows, and implies that the *de jure* population in the South American countries is larger than the *de facto* population. In terms of migration, this suggests that the number of legal South American residents is greater than the actual number of dwellers in these countries, confirming that emigration from the region is very common. Furthermore, the result of θ signals that unauthorised migrants do not constitute a large proportion of South American flows, which could be an outcome of multilateral agreements (e.g. MERCOSUR) which enable free movement and the possibility of residing and working in countries of the region that are different from individual's own nationality (Weeks, 2016).

Our results suggest that the concept used to conduct a census does have an impact on the estimation of migrants in South America. This is in contrast to Siegel and Swanson (2004), who argued that the differences between *de jure* and *de facto* concepts are insignificant at the national level. This could be because most South American censuses use the *de facto* notion, which is different from the *de jure* view that is more common in developed countries (Swanson & Tayman, 2012; UNDESA, 2007).

The parameter ω suggests that the log-transformed annual flows derived from the reported five-year flows (see Section 2.2) are $-(1.20^{-0.003}-1)\cdot 100\% = 0.05\%$ smaller than the true flows for every $20\%\cdot 2SD$ increase in the number of infant migrants. Likewise, the parameter κ expresses that the log-transformed and translated annual flows are $-(1.20^{-0.002}-1)\cdot 100\% = 0.03\%$ smaller than the true flows per $20\%\cdot 2SD$ increase in the number of migrant deaths. These results reveal that the error in the one-year flows estimated in Section 2.2 due to the omission of infant migrants is of the same magnitude (i.e. marginal) as the error arising from neglecting migrant deaths. The difference between the parameters related to infant migrants was expected to be larger because infant migrants represent 7.2% ($SD = 6.6\%$) of the total flows (that is, children under five years of age + migrant deaths + estimated one-year flows). In contrast, a small bias for non-surviving migrants was foreseen since the percentage of migrant deaths out of the total flows is 0.5% ($SD = 0.2\%$).

The parameter ν suggests that estimates based on country of birth, which only accounts for foreign-born migrants, are $-(exp(-0.94)-1)\times 100\% = 60.9\%$ lower than flows based on residence or true flows, in which both foreign-born and native-born migrants can be distinguished. This outcome was anticipated since native-born migrants constitute a significant share of the observed flows—around 55% of the total migrants registered in South American censuses.

The higher the posterior means of τ in Equation 5, the more precise the true flows are (that is, they have high accuracy). As expected, the mean posterior precision of censuses classified as having very good data quality (τ_{β_1}) was the highest. However, censuses determined to have good quality data (τ_{β_2}) resulted in smaller mean precision than censuses classified with fair quality data (τ_{β_3}), although the 95% CIs overlap. This could be due to aspects that this paper did not consider in the data quality assessment detailed in SM C, or because the difference in the evaluation of the census quality of 'good' and 'fair' is negligible.

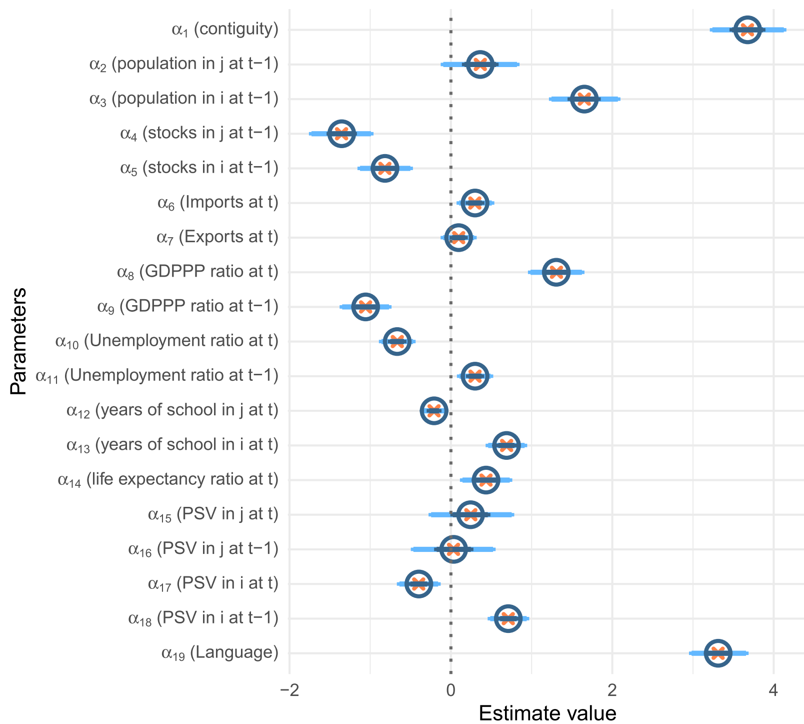

Figure 4. Posterior means (circle) and medians (cross mark) of migration sub-model parameters with 60% (darker segments) and 95% (lighter outer lines) Credible Intervals; *i* = origins and *j* = destinations. Source: Authors' own work.

4.3 Migration sub-model results

Figure 4 reports the posterior characteristics of migration sub-model parameters and SM G displays the results of our sensitivity analysis to the assumed prior distributions. In absolute terms, the most important predictors are contiguity ($|\alpha_1| = 3.68$), language ($|\alpha_{19}| = 3.31$), population in origins ($|\alpha_3| = 1.65$), migrants stocks born in i living in j ($|\alpha_4| = 1.35$) and GDPPP ratio at time t ($|\alpha_8| = 1.31$). The effect of contiguity is consistent with the findings in Pagnotta (2014) related to the importance of adjoining cities/towns in intraregional flows. Similarly, and as Adserà and Pytliková (2015), Ginsburgh and Weber (2016), and Özden and Schiff (2007) indicate, cultural factors, such as sharing a common language, play a significant role in South American migration.

In the case of migrant stocks, our results show the existence of a negative effect of migrant stocks on flows, a finding which is in line with the downward trend of the inverse U-shaped relationship between flows and stocks described by Bauer et al. (2009) for the Mexican–USA migration. According to Bauer et al. (2009), the initial positive relationship between flows and stocks is due to the fact that migrant stocks attract more migrants. However, when the concentration of individuals with similar skills and backgrounds reaches a certain point, it has a negative effect on wages and, consequently, on the number of migrants whose main motivation is their destination income.

Surprisingly, the population in origins obtained a higher effect than other factors (e.g. migrant stocks). This could be seen as evidence of the aspirations-capabilities framework of De Haas (2021). According to De Haas (2021) and the empirical evidence of Clemens (2014), the realisation of a migration is based on the aspirations and capabilities of an individual to undertake an international movement. If a country of origin has a large population and, thus, a high number of potential migrants, migration flows may increase if the country's overall income rises sufficiently to allow their residents to afford an international movement (i.e. the number of potential migrants could be translated into actual migrants). In the case of South America, the effect of the population of origin may be amplified by the fact that South American income has increased over at least the last couple of decades (OECD, CAF, ECLAC & European Commission, 2019).

Associated with the population in origins, the GDPPP ratio parameter signs a positive result, corroborating the suggestions that individuals have to acquire the resources needed to migrate. Following Ortega and Peri (2013), Cornia (2011), Mayda (2010), and Girma and Yu (2002), and the fact that OECD, CAF, ECLAC & European Commission (2019) has indicated that South American income is likely to continue to rise, it would not be surprising if the region increases its importance as an origin of migrants for other areas, for example, Europe.

In absolute terms, the mean years of schooling at the destinations ($|\alpha_{12}| = 0.21$) had the weakest association with the estimated true migrant flows. This variable was included in the migration submodel to reflect whether migrants' value human capital investment and the subsequent return that they can acquire from migrating to, and arriving in a specific receiving country as Dustmann and Glitz (2011) suggest. This study's results show that certain forms of migration that focus more on acquiring new skills are not determinant in intraregional South American flows. Our results also suggest that quality of life (seen in $|\alpha_{14}| = 0.44$) may not differ much between sending and receiving countries, not affecting migration heavily in South America.

5 The estimated true migration flows

5.1 Description of results

Figure 5 displays a set of posterior means and 95% CIs for the natural logarithmic transformation of the estimated true corridor-specific migration flows among the South American countries in our sample. SM H presents the natural logarithmic transformation of the estimated true migration flows from the rest of the world to South America.

The mean total number of migrants to South America was determined by adding up the resulting estimates of this research for all migration corridors for all years, and equates to 9.4 million people with $CI_{60\%}(5.6, 16.2)$. To identify the main historical destinations in the region, the total number of migrants to each destination was divided by the total number of migrants across all South American countries during the 1986–2018 period. Brazil, Colombia, and Chile recorded the highest proportions: 32.8%, 16.6%, and 12.1%, respectively. In contrast, Ecuador, Uruguay, and Bolivia had the smallest inflows; 2.9%, 3.0%, and 3.4%, correspondingly.

This paper also identifies the primary historical origins of South American migrants by dividing the total number of movers from a given sending country by the total number of migrants in all of

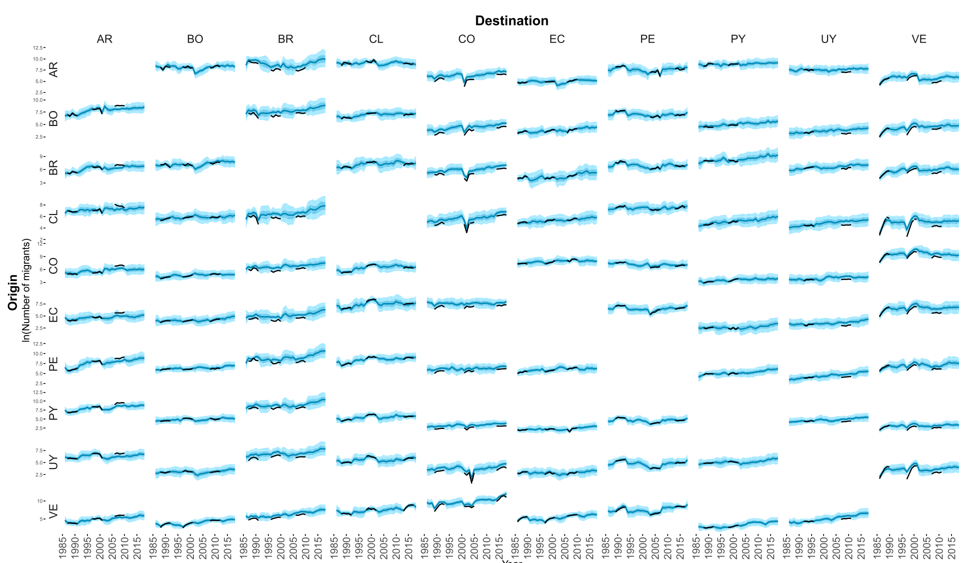

Figure 5. Natural logarithmic transformation of the translated one-year flows (black solid line) and estimated one-year (true) migration flows (blue solid line) with 60% and 95% Credible Intervals (CIs) amongst South American countries, i.e. Argentina (AR), Bolivia (BO), Brazil (BR), Chile (CL), Colombia (CO), Ecuador (EC), Paraguay (PY), Peru (PE), Uruguay (UY) and Venezuela (VE), from 1986 to 2018. Source: Authors' own work.

South America from 1986 to 2018. Analysis of the data reveals that the countries of origin were Venezuela (12.4%), Argentina (12.0%), and Colombia (8.1%). Europe was the region outside of South America with the highest number of migrants, with 1.2 million people, which is 13.0% of the total South American immigration. In contrast, Oceania, Canada, Uruguay, and Africa had the lowest number of migrants, each contributing less than 1.0% of total flows.

The resulting estimates show considerable variation in the main migration routes in the region throughout the years of analysis. From 1986 to 1999, the estimated average number of migrants was 2.8 million, with 12.9% of them travelling between Colombia and Venezuela. The third largest number of migrants was from Europe to Brazil; 6.0% of the total. From 2000 to 2009, the estimated mean was 2.7 million, with more than 10% of the inflows coming from Venezuela to Colombia. Additionally, migrants from Europe and Asia to Brazil accounted for 5.1% and 2.9%, respectively. From 2010 to 2018, the estimated number of migrants was 3.9 million, with the highest proportions of total flows from the region being from Venezuela to Colombia (23.1%) and from Europe to Brazil (21.0%).

5.2 Comparative analysis

Conducting a comparative analysis of the resulting true migration flows was challenging because there is no set of gold standard estimates for South America that could be used. Therefore, this paper compared the resulting synthetic data with three sets of existing migration flows. The first set of flows was calculated by Abel and Cohen (2019), and who indicated that one of the methods that produces consistent estimates is the Demographic Account Minimisation closed method (DAMC). The second set of calculations is that gauged by Azose and Raftery (2019), who implemented the Demographic Account Pseudo Bayesian closed method (DAPBC). This method relaxes the assumption of 'minimum migration' of the DAMC estimates. According to Abel and Cohen (2019), both the DAPBC and DAMC methods produce consistent estimates. Finally, the last set of values used in this comparative analysis corresponds to the recent migrants reported by ECLAC (2020); referred to as IMILA values.

We assess the correlation between our true migration flows and the DAMC, DAPBC, and IMILA estimates were assessed by computing Pearson correlation coefficients (see Figure 6). This paper compared counts and their natural logarithmic transformation, and proportions, as suggested by Abel and Cohen (2019). Five censuses were the focus of this comparison: the 2010 Argentinian census, the 2000 and 2010 Brazilian censuses, the 2005 Colombian census, and the 2010 Ecuadorian census (see Figure 6). These censuses were chosen because the estimates from DAMC, DAPBC, and IMILA are based on five-year intervals, and they are a suitable temporal match for the observed data. Only country-to-country flows were compared, and excluded grouped origins (i.e. America, Africa, Asia, Europe, and Oceania) as they might encompass different countries.

The comparative analysis assessed the temporal and spatial consistency that exists between the resulting estimates of this research and the DAMC, DAPBC, and IMILA values. Perfect pairwise correlations were not sought because DAMC, DAPBC, and IMILA estimates refer to five-year intervals, and are based on migrant stock data. In contrast, the resulting true flows are annual values and drew on the number of recent migrants. The Pearson coefficient was highest when comparing the true number of migration flows and the DAPBC values, while the smallest number was the result of contrasting the resulting estimates of this research and IMILA. Figure 6 also reveals that the correlation coefficients obtained by comparing the natural logarithmic transformations of the resulting true flows and the values of DAMC, DAPBC, and IMILA were higher than those resulting from the analysis of the relationship between sets of counts.

We also compute the proportion of bilateral flows over the total number of migrants. The resulting Pearson coefficients indicated that the proportions based on true flows and those derived from DAMC, DAPBC, and IMILA values had a stronger linear relationship than the correlation with the respective counts and logarithmic transformations. IMILA values were higher than those obtained from the resulting true flows, with the largest discrepancy being for the flows from Colombia to Ecuador in the 2005–2010 period. The proportions calculated based on the resulting true flows matched the DAPBC-based proportions in origin-destination-specific migration flows more closely than the other values.

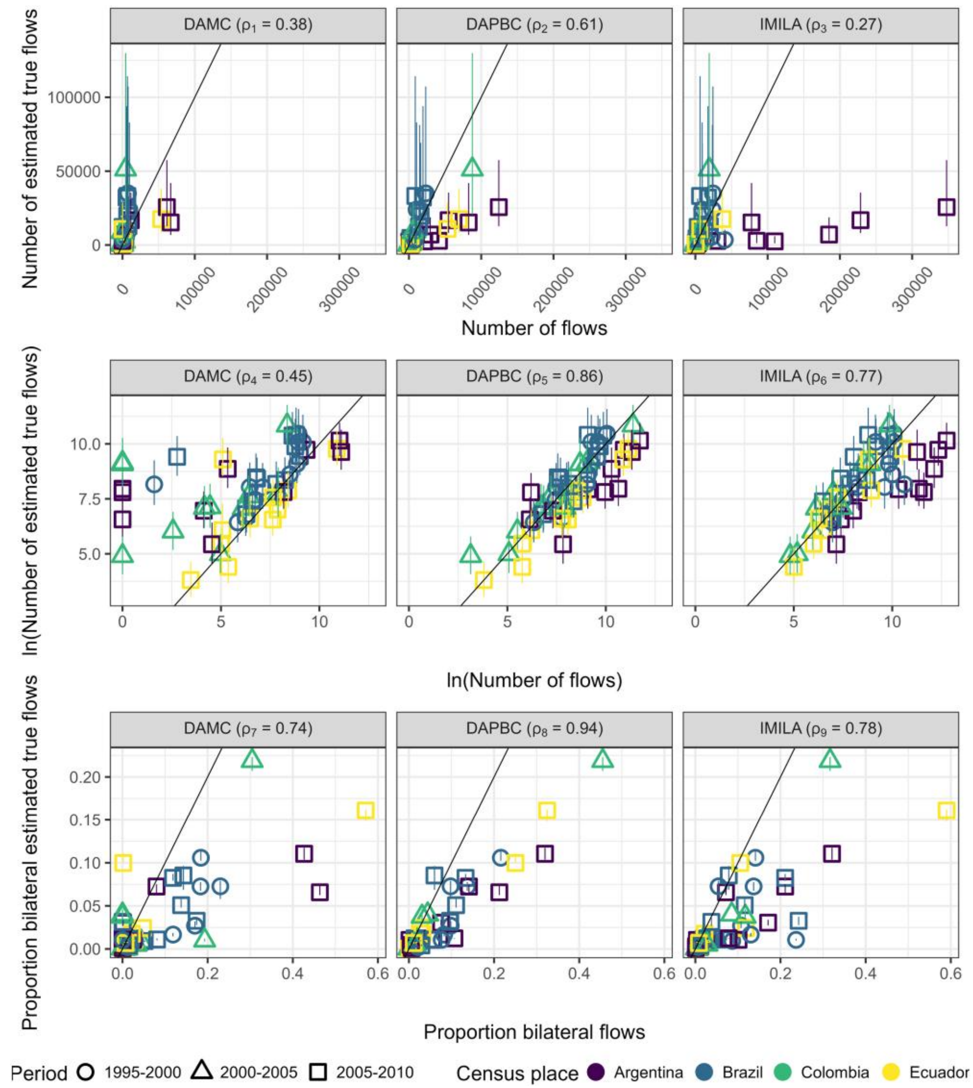

Figure 6. Comparison between the number of estimated true flows, the natural logarithmic transformation of true flows and the proportions of the estimated true migration flows, and the Demographic Account Minimisation closed method, Demographic Account Pseudo Bayesian closed method and International Migration Research in Latin America project estimates for five selected censuses: the 2010 Argentinian, 2000 and 2010 Brazilian, 2005 Colombian and 2010 Ecuadorian censuses. Source: Authors' own work.

Overall, our migration estimates were more consistent with the DAPBC values than those generated using either the DAMC or IMILA techniques. This implies that both sets of values also produce similar spatial patterns, and encompass all counts and most of the log-transformed values. It is reasonable that the resulting flows were more similar to the DAPBC results than the DAMC estimates since the first set of values did not include the assumption of 'minimum migration'. This assumption required that flows correspond to variations in migrant stocks, and meant that flow estimates were strongly connected to foreign-born migration, and disregarded the effect of native-born movements.

In addition, we notice that a high degree of consistency between pairs of sets occurs as the quality of the census is better. The four sets of estimates showed a similar number of migrants for most of the inflows to Ecuador when the census data was classified as very good. However, the estimates of the various methods varied significantly when the quality of the data was poor. The inflows to

Argentina and Colombia had the most significant discrepancies, and their census migration data were classified as 'good' and 'fair', respectively.

5.3 Sensitivity analysis

A sensitivity analysis was conducted to determine how the true flows vary when certain input data were removed. SM I presents the detailed results of the study. Three validation models were created by partially removing input data and then this paper compared the resulting values with the estimates of the original model that were generated using the full data set. The first validation model excluded data from the 1991 Brazilian, the 1990 Ecuadorian, and the 1990 Venezuelan censuses, which provided most of the information before 1990. The second model excluded data from the 2002 Argentinian, the 2001 Bolivian, and the 2002 Paraguayan censuses. The third and final validation model excluded data from all censuses after 2015, such as the 2017 Chilean, the 2018 Colombian, and the 2017 Peruvian censuses.

In general, the removal of data from censuses that provide the majority of information before 1990 had the greatest impact on the estimates for Brazil, for which the original model outcomes were moderately larger than those produced by the validation model. This is reasonable, as the 1991 Brazilian information corresponds to a destination with one of the highest flows in South America. Although the $CIs_{95\%}$ encompassed the values in the validation model, the changes in the means may have been due to the fact that the covariates in the theory-driven sub-model were unable to account for large flows and all the abrupt changes in them.

In the second validation model, the estimates from Argentina and other countries were slightly impacted by the partial removal of the data. The mean estimates of the original model and their 95% confidence intervals were higher than the values obtained by this validation model. It was anticipated that the removal of data from one of the Paraguayan censuses would have a more significant effect on its estimates since Paraguay only had two censuses in 1986 and 2018 compared to the other countries which had three censuses. Nevertheless, the theory-driven migration sub-model could more accurately impute data in the Paraguayan case, which suggests that small flows can be better accounted for. Lastly, the third validation model mainly affected the estimates for Colombia, which were moderately affected by the partial removal of data from the censuses after 2015.

6 Concluding remarks

6.1 Contributions

In this article, we proposed and implemented a novel method, which refers to a bespoke model to estimate complete, reliable, comparable, and consistent annual bilateral migration flows for South American countries from 1986 to 2018 (i.e. a 33-year period) with measures of uncertainty. The proposed method makes six substantive contributions to research on migration flow estimation.

To begin with, the proposed method transformed five-year census data into one-year transition data, maximising the utility of censuses, the most commonly available source of migration data. This paper expanded the utility of migration estimates derived from censuses by converting five-year into annual information. Annual migration flows can evidence short-term changes which cannot be captured in five-year data. Annual migration flows can also facilitate monitoring short-term changes due to migration policy shifts. They can also enable integration with other annually recorded demographic and socio-economic data; and comparability with data recorded over non-overlapping five-year periods.

An additional advantage of the proposed method is the possibility of incorporating origin-destination-specific variation. The proposed method thus offers the potential to generate annual migration estimates for other geographical settings outside South America, dealing with challenges that European or US data may not have due to their data having benefited from the relative enforcement of the harmonisation of migration statistics taking place after the implementation of, e.g. Regulation No. 862/2007.

Moreover, the proposed method corrected differences between censuses and inadequacies in census data to consider cross-national differences in how migration data is collected. Specifically, it measured and corrected differences in *de jure* and *de facto* measurement, as well as the omission of infant and non-surviving migrant populations, and decomposed and incorporated the overall migration flows of foreign-born and native-born migrants. In addition, the proposed method enabled the quantification of the omission of native-born migrants, who are often excluded from migrant count estimates. This is a relevant contribution because native-born migrants comprise a large share of the total flows for specific migration corridors in South America.

To identify any systematic bias or errors in censuses, an understanding of the available data on migration and their sources is required. Much existent literature in migration studies has focused on assessing quality and sources of data on migration for countries such as those belonging to the EU-27, EFTA, and the UK (e.g. Disney et al., 2015; Lemaitre, 2005; Mooyaart et al., 2021; Nowok & Kupiszewska, 2005). This paper systematically reviewed South American censuses and developed a method for quantitatively evaluating census data quality used to measure migration. Previous efforts to do so for South America have mainly been qualitative (e.g. Tacla Chamy, 2006). Including data quality metrics in the migration model specification proved especially relevant when comparing the resulting true migration flows to existing estimates (Abel & Cohen, 2019; Azose & Raftery, 2019; ECLAC, 2020).

Finally, the proposed method included a theory-driven migration sub-model to impute missing migration flows, and consistently filled data gaps between censuses. This study is the first analysis to integrate this type of imputation with the transformation of five-year migration flows into one-year interval data and the correction of estimates for different biases in a coherent methodological framework for South America. This integration included the quantification of uncertainty and improved existing calculations based on migrant stock data.

Alongside the proposed method, the resulting estimates contribute to addressing an existing gap of complete, reliable, comparable, and consistent annual bilateral migration flows for South America. Focusing on the 10 biggest South American countries and since information on migration is usually incomplete and incomparable, the resulting estimates are useful for global partnerships, governmental, non-governmental, and intergovernmental organisations (e.g. the UN and their dependencies like the Commission for Latin America and the Caribbean—ECLAC, and/or the United Nations Department of Economic and Social Affairs—UNDESA) as well as academic and research institutions. Specifically, the resulting estimates help in the study of migration itself as well as the generation of other estimates, e.g. reconstructing populations, for the 10 biggest South American countries, and account for rapid changes and measures pertaining of the uncertainty of migration flows that other analysis has lacked (e.g. in IMILA values). This is not a minor contribution, taking into account that UNDESA (2022) has acknowledged the lack of reliable annual demographic data for more than 75% of the countries worldwide.

6.2 Limitations and future work

We highlight critical and major areas for future improvement. First, the resulting migration estimates could benefit from assuming prior distributions based on eliciting expert opinion on model parameters and selecting predictors. These could improve the estimation of flows of specific corridors whose censuses have fair-quality data. Expert opinion has been used successfully to estimate European migration flows (e.g. in Raymer et al., 2013; Wiśniowski et al., 2013). Yet, surveying experts remains challenging because collecting their opinion requires additional resources and access to experts with relevant knowledge.

Second, future work could combine census migration data with other sources, such as administrative and new forms of data, e.g. derived from social media, to compensate for the limitations of census data (e.g. biases due to the omission of infant migrants and migrant deaths). This would help better understand the relationship between the definitions of migration used in different sources. Finally, the uncertainty of imputing predictors in the migration model could be incorporated into the resulting migration flows.

Acknowledgments

In particular, thanks to Mark Brown, who provided valuable input on the content and structure of this paper. We are also grateful to Professor Jakub Bijak and Dr Simon Rudkin for their comments on the content of this document, which they revised as part of Andrea Aparicio-Castro PhD's thesis. This research would not have been possible without the funding provided by the NWSSDTP to undertake an internship in the Latin American and Caribbean Demographic Center (CELADE, by its Spanish acronym) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the UN. The internship enabled the processing of the micro census data in situ. In addition, ECLAC experts, especially Jorge Rodriguez, Jorge Martinez, Helena Cruz Castanheira, Guiomar Bay, and Mario Acuña, offered valuable advice on specific aspects of data processing and analysis.

Conflict of interests: None declared

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 and the University of Manchester. Arkadiusz Wiśniowski acknowledges funding received from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870649, project Future Migration Scenarios for Europe (FUME).

Data and code availability

JAGS and R code, simulations of input data and model output, including estimated true migration flows, can be found in Aparicio-Castro et al. (2023).

Supplementary material

Supplementary material are available at Journal of the *Royal Statistical Society: Series A* online.

References

Abdallah W., Goergen M., & O'Sullivan N. (2015). Endogeneity: How failure to correct for it can cause wrong inferences and some remedies. *British Journal of Management*, *26*(4), 791–804. https://doi.org/10.1111/1467-8551.12113

Abel G. J. (2010). Estimation of international migration flow tables in Europe. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(4), 797–825. https://doi.org/10.1111/j.1467-985X.2009.00636.x

Abel G. J. (2013). Estimating global migration flow tables using place of birth data. *Demographic Research*, *28*(18), 505–546. https://doi.org/10.4054/DemRes.2013.28.18

Abel G. J. (2018). Estimates of global bilateral migration flows by gender between 1960 and 2015. *International Migration Review*, *52*(3), 809–852. https://doi.org/10.1111/imre.12327

Abel G. J., & Cohen J. E. (2019). Bilateral international migration flow estimates for 200 countries. *Scientific Data*, *6*(1), 82. https://doi.org/10.1038/s41597-019-0089-3

Adserà A., & Pytliková M. (2015). The role of language in shaping international migration. *The Economic Journal*, *125*(586), F49–F81. https://doi.org/10.1111/ecoj.12231

Alimi O., Mare D., & Poot J. (2015). Does distance still matter for internal migration and, if so, how? Evidence from 1986 to 2006. *Labour, Employment and Work in New Zealand*. https://doi.org/10.26686/lew.v0i0.3897.

Aparicio-Castro A., Wiśniowski A., & Rowe F. (2023). *Estimating annual flows from censuses*, v.2.0.0, Github repository. https://github.com/alaparicioc/Estimating˙annual˙flows˙from˙censuses.git.

Arestis P., Baddeley M., & McCombie J. S. (2007). *Economic growth: New directions in theory and policy*. Edward Elgar Publishing.

Azose J. J., & Raftery A. E. (2015). Bayesian probabilistic projection of international migration. *Demography*, *52*(5), 1627–1650. https://doi.org/10.1007/s13524-015-0415-0

Azose J. J., & Raftery A. E. (2019). Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences (PNAS)*, *116*(1), 116–122. https://doi.org/10.1073/pnas.1722334116

Bauer T., Epstein G. S., & Gang I. N. (2009). Measuring ethnic linkages among migrants. *International Journal of Manpower*, *30*(1/2), 56–69. https://doi.org/10.1108/01437720910948393

Bell M., Rees P., Blake M., & Duke-Williams O. (1999). *An age-period-cohort database of inter-regional migration in Australia and Britain, 1976–96* (Technical report). Working Paper, School of Geography, University of Leeds. https://eprints.whiterose.ac.uk/5030/1/99-2.pdf.

Bengochea J., & Saucedo S. E. G. (2018). Retos metodológicos para el estudio de la migración intrarregional en américa del sur. In R. Baeninger, L. Machado Bógus, J. Bertino Moreira, L. R. Vedovato, D. Magalhães Fernandes, M. R. De Souza, C. Siqueira Baltar, R. Guimarães Peres, T. Chang Waldman, & L. F. Aires

Magalhães (Eds.), *Migrações Sul-Sul* (pp. 54–65). https://nempsic.paginas.ufsc.br/files/2015/02/LIVRO-MIGRA%C3%87%C3%95ES-SUL-SUL.pdf.

Bertoli S., & Moraga J. F.-H. (2017). Gravity models in the migration and development nexus. *Revue d'économie du Développement*, *25*(3/4), 69–91. http://dx.doi.org/10.3917/edd.313.0069

Bijak J., Disney G., Findlay A. M., Forster J. J., Smith P. W., & Wiśniowski A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, *38*(5), 470–487. https://doi.org/10.1002/for.2576

Bilsborrow R. E., Hugo G., Oberai A. S., & Zlotnik H. (1997). *International migration statistics: Guidelines for improving data collection systems*. International Labour Organization. http://www.ilo.org/public/libdoc/ilo/1997/97B09˙286˙engl.pdf.

Bryant J., & Zhang J. L. (2018). *Bayesian demographic estimation and forecasting*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429452987.

Buonaccorsi J. P. (2010). *Measurement error: Models, methods, and applications*. CRC Press. https://doi.org/10.1201/9781420066586.

Carmichael G. A. (2016). *Fundamentals of demographic analysis: Concepts, measures and methods*. Springer series on demographic methods and population analysis 38, Springer. https://doi.org/10.1007/978-3-319-23255-3.

Clemens M. (2014). Does development reduce migration?. In R. Lucas, (ed.), *International handbook on migration and economic development* (pp. 152–185). Edward Elgar Publishing, chapter 6. https://EconPapers.repec.org/RePEc:elg:eechap:15465˙6.

Cornia G. A. (2011). Economic integration, inequality and growth: Latin America versus the European economies in transition. *Review of Economics and Institutions*, *2*(2), 1–37. https://doi.org/10.5202/rei.v2i2.29

Council of European Parliament. (2007). *Regulation (EC) No 862/2007 of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection and repealing Council Regulation (EEC) No 311/76 on the compilation of statistics on foreign workers (Text with EEA relevance)*. http://data.europa.eu/eli/reg/2007/862/oj.

De Beer J. (2008). Forecasting international migration: times series projections vs. argument-based forecasts. In J. Raymer, & F. Willekens (Eds.), *International migration in Europe: Data, models and estimates* (pp. 283–306). John Wiley & Sons.

De Haas H. (2021). A theory of migration: The aspirations-capabilities framework. *Comparative Migration Studies*, *9*(1), 1–35. https://doi.org/10.1186/s40878-020-00210-4

De Haas H., Castles S., & Miller M. J. (2022). *The age of migration. International population movements in the modern world* (6th ed.), Bloomsbury Academic.

Del Fava E., Wiśniowski A., & Zagheni E. (2020). *Modelling international migration flows by integrating multiple data sources*. doi: 10.31235/osf.io/cma5h.

Disney G., Wiśniowski A., Forster J. J., Smith P. W., & Bijak J. (2015). *Evaluation of existing migration forecasting methods and models*. Economic and Social Research Council (ESRC), Centre for Population Change, University of Southampton. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467405/Migration˙Forecasting˙report.pdf.

Dustmann C., & Glitz A. (2011). Chapter 4 - migration and education. In E. A. Hanushek, S. Machin, & L. Woessmann, (Eds.), *Handbook of the economics of education* (Vol. 4, pp. 327–439). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780444534446000043.

Dyrting S. (2018). *A framework for translating between one-year and five-year migration probabilities*. doi:10.13140/RG.2.2.24261.91362.

ECLAC. (2013a). *Consenso de Montevideo sobre población y desarrollo* (Technical report). The Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://repositorio.cepal.org/bitstream/handle/11362/21835/S20131037˙es.pdf?sequence=4\&isAllowed=y.

ECLAC. (2013b). *Proposed regional agenda on population and development for Latin America and the Caribbean beyond 2014* (Technical report). The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). http://repositorio.cepal.org/bitstream/handle/11362/3123/1/S2013417˙en.pdf.

ECLAC. (2015). *Operational guide for implementation and follow-up of the Montevideo Consensus on Population and Development* (Technical report). The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://repositorio.cepal.org/bitstream/handle/11362/38937/6/S1500859˙en.pdf.

ECLAC. (2019a). *First regional report on the implementation of the Montevideo Consensus on Population and Development* (Technical report), The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://repositorio.cepal.org/bitstream/handle/11362/44458/6/S1801011˙en.pdf.

ECLAC. (2019b). *Redatam7*. The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://redatam.org/en.

ECLAC. (2019c). *Redatam+SP*. The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://redatam.org/en.

ECLAC. (2020). *Investigación de la Migración Internacional en Latinoamérica (IMILA)*. The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). celade.cepal.org/bdcelade/imila/.

ECLAC. (2022). *The sociodemographic impacts of the COVID-19 pandemic in Latin America and the Caribbean* Technical report, The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://repositorio.cepal.org/bitstream/handle/11362/47923/S2200158˙en.pdf.

FEM & IOM. (2022). *Movimientos Migratorios Recientes en América del Sur. Informe Annual 2021* (Technical report). Foro Especializado Migratorio del Mercado Común del Sur (MERCOSUR) y Estados Asociados (FEM), y la Organización Internacional para las Migrationes (OIM or IOM, by its English acronym). https://robuenosaires.iom.int/sites/g/files/tmzbdl626/files/documents/OIM˙FEM˙Informe˙anual˙2021˙0˙0.pdf.

FEM & IOM. (2023). *Movimientos Migratorios Recientes en Amórica del Sur. Informe Annual 2022* (Technical report). Foro Especializado Migratorio del Mercado Común del Sur (MERCOSUR) y Estados Asociados (FEM), y la Organización Internacional para las Migrationes (OIM or IOM, by its English acronym). https://www.argentina.gob.ar/sites/default/files/oim˙fem˙informe˙anual˙2022.pdf.

Gelman A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*(15), 2865–2873. https://doi.org/10.1002/(ISSN)1097-0258

Ginsburgh V., & Weber S. (2016). *The Palgrave handbook of economics and language*. Springer. https://doi.org/10.1007/978-1-137-32505-1.

Girma S., & Yu Z. (2002). The link between immigration and trade: Evidence from the United Kingdom. *Weltwirtschaftliches Archiv*, *138*(1), 115–130. https://doi.org/10.1007/BF02707326

Graham-Harrison E. (2020). *UK government has underestimated take up for Hong Kong resettlement scheme*. The Guardian. https://www.theguardian.com/world/2020/dec/12/uk-government-underestimates-takeup-hong-kong-resettlement.

Gustafson P. (2003). *Measurement error and misclassification in statistics and epidemiology. Impacts and Bayesian adjustments* (1st ed.), Chapman & Hall, CRC Press. https://doi.org/10.1201/9780203502761.

Hatton T. J. (2009). The rise and fall of asylum: What happened and why?. *The Economic Journal*, *119*(535), F183–F213. https://doi.org/10.1111/j.1468-0297.2008.02228.x

Hatzigeorgiou A. (2010). Migration as trade facilitation: Assessing the links between international trade and migration. *The B.E. Journal of Economic Analysis & Policy*, *10*(1), 1–33. https://doi.org/10.2202/1935-1682.2100

Hinde A. (2014). *Demographic methods* (1st ed.), Routledge.

ICG. (2022). *Hard times in a safe haven: Protecting venezuelan migrants in colombia* (Technical Report 94). International crisis group (ICG), Latin American report. https://icg-prod.s3.amazonaws.com/094-protecting-venezuelans-in-colombia˙0.pdf.

Instituto Nacional de Estadística. (2023). *Migraciones exteriores. Resultados Nacionales. Flujo de inmigración procedente del extranjero por año, país de origen y nacionalidad (española/extranjera)*. Instituto Nacional de Estadística (INE), Spain. Accessed: September 9, 2023. https://www.ine.es/jaxiT3/Tabla.htm?t=24295.

Jennissen R. P. W. (2004). *Macro-economic determinants of international migration in Europe* [PhD thesis]. https://research.rug.nl/en/publications/macro-economic-determinants-of-international-migration-in-europe.

Jensen E. B. (2013). *A Review of Methods for Estimating Emigration*. Working Paper Number POP-WP101, US Census Bureau, Population Division. https://www.census.gov/content/dam/Census/library/working-papers/2013/demo/POP-twps0101.pdf.

Judson D. H., & Swanson D. A. (2011). *Estimating characteristics of the foreign-born by legal status. An evaluation of data and methods*. Springer Link. https://doi.org/10.1007/978-94-007-1272-0.

Juran S., & Snow R. C. (2018). The potential of population and housing censuses for international migrant analysis. *Statistical Journal of the IAOS*, *34*(2), 203–213. https://doi.org/10.3233/SJI-170359

Kim K., & Cohen J. E. (2010). Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity. *International Migration Review*, *44*(4), 899–932. https://doi.org/10.1111/j.1747-7379.2010.00830.x

Kitsul P. I., & Philipov D. (1980). *The one-year five-year migration problem*. International Institute for Applied Systems Analysis (IIASA) Working Paper WP-80-081. https://core.ac.uk/download/pdf/33892859.pdf.

Kruschke J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.), Academic Press, Elsevier. https://nyu-cdsc.github.io/learningr/assets/kruschke˙bayesian˙in˙R.pdf.

Kupiszewska D., & Wiśniowski A. (2009). *Availability of statistical data on migration and migrant population and potential supplementary sources for data estimation Report for MIMOSA project* (Technical report). Migrant Information Management System Next Generation Framework (MIMOSA) Deliverable 9.1 A Report, Netherlands Interdisciplinary Demographic Institute. http://mimosa.cytise.be/Documents/Mimosa˙2009.pdf.

Lemaitre G. (2005). The comparability of international migration statistics. Problems and prospects. *Statistics Brief*, *9*, 1–8. https://www.oecd.org/migration/49215740.pdf.

López M. F., Marín L., Romero E., & Smolansky D. (2020). *Situation of Venezuelans who have returned and are trying to return to their country in the context of COVID-19* (Technical report). Organization of American States (OAS). https://reliefweb.int/sites/reliefweb.int/files/resources/OAS˙Venezuelan-Returnees˙ENG.pdf.

Mangabeira O., & Montilla J. A. (1928). *Protocolo. Brazil/Venezuela (24 Julho 1928)*. http://www.info.lncc.br/ve1928.html.

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

Mayer T., & Zignago S. (2011). *Notes on CEPII's distances measures: The GeoDist database*. Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) Working Paper No. 2011-25. https://dx.doi.org/10.2139/ssrn.1994531.

Melander E., & Öberg M. (2006). Time to go? Duration dependence in forced migration. *International Interactions*, *32*(2), 129–152. https://doi.org/10.1080/03050620600574873

MERCOSUR. (2023). *Objectives of MERCOSUR*. Southern Common Market (MERCOSUR , by its Spanish acronym). Accessed: October 4, 2023. https://www.mercosur.int/quienes-somos/objetivos-del-mercosur.

Mooyaart J., Dańko M. J., Costa R., & Boissonneault M. (2021). *Quality assessment of European migration data* (Technical report). Deliverable 6.2, Quantifying Migration (QuanMig) project, EU Horizon 2020. http://quantmig.eu/res/files/QuantMig˙Deliverable%206.2%20vf.pdf.

Moya Ó (1993). *Proyecto IMILA: investigación de la migración internacional en Latinoamérica* (Technical report). Historic Series, Serie A, No. 284, The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://hdl.handle.net/11362/9507.

Neumayer E. (2005). Bogus refugees? The determinants of asylum migration to Western Europe. *International Studies Quarterly*, *49*(3), 389–410. https://doi.org/10.2139/ssrn.482645

Newbold K. B. (2001). Counting migrants and migrations: Comparing lifetime and fixed-interval return and onward migration. *Economic Geography*, *77*(1), 23–40. https://doi.org/10.2307/3594085

Newell C. (1988). *Methods and models in demography*. Guilford Press.

Nowok B. (2010). *Harmonization by simulation. A contribution to comparable international migration statistics in Europe* [PhD thesis]. https://research.rug.nl/en/publications/harmonization-by-simulation-a-contribution-to-comparable-internat.

Nowok B., & Kupiszewska D. (2005). *Official European statistics on international migration flows: availability, sources and coverage*. Central European Forum for Migration Research (CEFMR) Working paper 5/2005. http://www.cefmr.pan.pl/docs/cefmr˙wp˙2005-05.pdf.

Nowok B., & Willekens F. (2011). A probabilistic framework for harmonisation of migration statistics. *Population, Space and Place*, *17*(5), 521–533. https://doi.org/10.1002/psp.624

OECD, CAF, ECLAC & European Commission. (2019). *Latin American Economic Outlook 2019. Development in Transition* (Technical report). Organisation for Economic Co-operation and Development (OECD), Development Bank of Latin America (CAF, by its Spanish acronym), Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://repositorio.cepal.org/bitstream/handle/11362/44515/S1900181˙en.pdf?sequence=4\&isAllowed=y.

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

Özden C., & Schiff M. (2007). *International migration, economic development & policy*. World Bank and Palgrave Macmillan. https://doi.org/10.1596/978-0-8213-6935-7.

Pagnotta C. (2014). Latin American migration within the continent between 1980 and 2000: A historical perspective. *Nuevo Mundo Mundos Nuevos*. https://doi.org/10.4000/nuevomundo.67174.

Pereira Leal F. J., & Sanojo L. (1859). *Tratado de Limites e Navegaçao. Brasil / Venezuela (5 Maio 1859)*. http://www.info.lncc.br/ve1859.html.

Plummer M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003), Vienna, Austria*. https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.

Plummer M., Best N., Cowles K., & Vines K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11. https://oro.open.ac.uk/22547/.

Plummer M., Best N., Cowles K., Vines K., Sarkar D., Bates D., Almond R., & Magnusson A. (2020). *Package 'coda'. Output analysis and diagnostics for MCMC*. CRAN Repository. https://cran.r-project.org/web/packages/coda/coda.pdf.

Rayer S. (2015). Demographic techniques: Small-area estimates and projections. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 162–169). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780080970868310157.

Raymer J., & Rogers A. (2007). Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, *44*(2), 199–223. https://doi.org/10.1353/dem.2007.0016

Raymer J., Wiśniowski A., Forster J. J., Smith P. W., & Bijak J. (2013). Integrated modeling of European migration. *Journal of the American Statistical Association*, *108*(503), 801–819. https://doi.org/10.1080/01621459.2013.789435

Rees P., Bell M., Duke-Williams O., & Blake M. (2000). Problems and solutions in the measurement of migration intensities: Australia and Britain compared. *Population Studies*, *54*(2), 207–222. https://doi.org/10.1080/713779082

Rees P., Bell M., Kupiszewski M., Kupiszewska D., Ueffing P., Bernard A., Charles-Edwards E., & Stillwell J. (2017). The impact of internal migration on population redistribution: An international comparison. *Population, Space and Place*, *23*(6), e2036. https://doi.org/10.1002/psp.2036

Rees P. H. (1977). The measurement of migration, from census data and other sources. *Environment and Planning A: Economy and Space*, *9*(3), 247–272. https://doi.org/10.1068/a090247

Rodríguez-Vignoli J., & Rowe F. (2018). How is internal migration reshaping metropolitan populations in Latin America? A new method and new evidence. *Population Studies*, *72*(2), 253–273. https://doi.org/10.1080/00324728.2017.1416155

Rogers A., & Jordan L. (2004). Estimating migration flows from birthplace-specific population stocks of infants. *Geographical Analysis*, *36*(1), 38–53. https://doi.org/10.1111/j.1538-4632.2004.tb01122.x.

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

Rogers A., Raymer J., & Newbold B. K. (2003). Reconciling and translating migration data collected over time intervals of differing widths. *The Annals of Regional Science*, *37*(4), 581–601. https://doi.org/10.1007/s00168-003-0128-y

Rogers A., Raymer J., & Willekens F. (2003). Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer*, *55*(1), 56–69. https://doi.org/10.1111/0033-0124.01052

Rogers A., Willekens F., Little J., & Raymer J. (2002). Describing migration spatial structure. *Papers in Regional Science*, *81*(81), 29–48. https://doi.org/10.1007/s101100100090

Rogerson P. A. (1990). Migration analysis using data with time intervals of differing widths. *Papers in Regional Science*, *68*(1), 97–106. https://doi.org/10.1111/j.1435-5597.1990.tb01196.x.

Rowe F., Bell M., Bernard A., Charles-Edwards E., & Ueffing P. (2019). Impact of internal migration on population redistribution in Europe: Urbanisation, counterurbanisation or spatial equilibrium? *Comparative Population Studies*, *44*(44), 201–234. https://doi.org/10.12765/CPoS-2019-18

Sen A., & Smith T. E. (1995). *Gravity models of spatial interaction behavior*. Springer Science & Business Media. https://doi.org/10.1007/978-3-642-79880-1.

Siegel J. S., & Swanson D. A. (2004). *The Methods and Materials of Demography* (2nd ed.). Elsevier Academic Press. https://demographybook.weebly.com/uploads/2/7/2/5/27251849/david˙a.˙swanson˙jacob˙s.˙siegel˙the˙methods˙and˙materials˙of˙demography˙second˙edition˙˙2004.pdf.

Sorichetta A., Bird T. J., Ruktanonchai N. W., Erbach-Schoenberg E. Z., Pezzulo C., Tejedor N., Waldock I. C., Sadler J. D., Garcia A. J., Sedda L., & Tatem A. J. (2016). Mapping internal connectivity through human migration in malaria endemic countries. *Scientific Data*, *3*(1), 1–16. https://doi.org/10.1038/sdata.2016.66

Swanson D. A., & Tayman J. (2011). On estimating a de facto population and its components. *Review of Economics & Finance*, *1*, 17–31. https://ideas.repec.org/a/bap/journl/110502.html.

Swanson D. A., & Tayman J. (2012). *Subnational population estimates*. The Springer Series on Demographic Methods and Population Analysis 31, Springer. https://doi.org/10.1007/978-90-481-8954-0.

Tacla Chamy O. (2006). *La omisión censal en América Latina, 1950–2000* (Technical report). http://digitallibrary.un.org/record/573391.

UNDESA. (1998). *Recommendations of statistics of international migration, revision 1* (Technical report). Statistical Papers, Series M, No. 58, Department of Economic and Social Affairs, Statistics Division, United Nations (UNDESA). https://unstats.un.org/unsd/publication/seriesm/seriesm˙58rev1e.pdf.

UNDESA. (2007). *Part two. Measuring international migration through population censuses* (Technical report). Department of Economic and Social Affairs, Statistics Division, United Nations (UNDESA). https://unstats.un.org/unsd/demographic/meetings/egm/migrationegmsep07/TECHREP2˙Part%20Two.pdf.

UNDESA. (2008). *Principles and recommendations for population and housing censuses, revision 2* (Technical report). Statistical Papers, Series M, No. 67, Department of Economic and Social Affairs, Statistics

Division, United Nations (UNDESA). https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles˙and˙Recommendations/Population-and-Housing-Censuses/Series˙M67Rev2-E.pdf.
UNDESA. (2017). *Principles and recommendations for population and housing censuses, revision 3* (Technical report). Statistical Papers, Series M, No. 67, Department of Economic and Social Affairs, Statistics Division, United Nations (UNDESA). https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles˙and˙Recommendations/Population-and-Housing-Censuses/Series˙M67rev3-E.pdf.
UNDESA. (2020a). *International migrant stock 2020: Destination*. Department of Economic and Social Affairs, Statistics Division, United Nations (UNDESA). POP/DB/MIG/Stock/Rev.2020.
UNDESA. (2020b). *International migrant stock 2020: Destination and origin*. Department of Economic and Social Affairs, Statistics Division, United Nations (UNDESA). POP/DB/MIG/Stock/Rev.2020.
UNDESA. (2020c). *International migrant stock 2020: Origin*. Department of Economic and Social Affairs, Statistics Division, United Nations (UNDESA). POP/DB/MIG/Stock/Rev.2020.
UNDESA. (2022). *World population prospects 2022*. Online Edition, Department of Economic and Social Affairs, Statistics Division, United Nations (UNDESA). https://population.un.org/wpp/Download/Standard/Population/.
UNSD. (2020). *Demographic yearbook system*. United Nations Statistics Division (UNSD). Accessed: January 14, 2020. https://unstats.un.org/unsd/demographic-social/products/dyb/index.cshtml.
Vargas-Silva C. (2013). *Handbook of research methods in migration*. Edward Elgar Publishing.
Villa M. (1991). *Introducción al Análisis de la Migración: Apuntes de clase. Notas preliminares* (Technical report). The Latin American and Caribbean Demographic Centre (CELADE, by its acronym in Spanish) of the Economic Commission for Latin America and the Caribbean (ECLAC) of the United Nations (UN). https://repositorio.cepal.org/bitstream/handle/11362/9271/S9100017˙es.pdf?sequence=1\&isAllowed=y.
Weeks J. R. (2016). *Population: An introduction to concepts and issues* (12th ed.), Cengage Learning.
White P. E., & Van der Knaap B. (1985). *Contemporary studies of migration: Proceedings of the second British-Dutch Symposium on Population Geography, Soesterberg, The Netherlands, 14–16 September 1983*. International symposia series, Geo Books.
Willekens F. (2008). Models of migration: observations and judgement. In J. Raymer, & F. Willekens (Eds.), *International migration in Europe: Data, models and estimates*. John Wiley & Sons, pp. 117–148.
Willekens F., Massey D., Raymer J., & Beauchemin C. (2016). International migration under the microscope. *Science*, *352*(6288), 897–899. https://doi.org/10.1126/science.aaf6545
Wiśniowski A., Bijak J., Christiansen S., Forster J. J., Keilman N., Raymer J., & Smith P. W. (2013). Utilising expert opinion to improve the measurement of international migration in Europe. *Journal of Official Statistics*, *29*(4), 583–607. https://doi.org/10.2478/jos-2013-0041
Wiśniowski A., Forster J. J., Smith P. W., Bijak J., & 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. https://doi.org/10.1111/rssa.12177