

RESEARCH

Open Access

Looking for a better future: modeling migrant mobility



Isabella Loaiza Saa^{*}, Matej Novak, Alfredo J. Morales  and Alex Pentland 

^{*}Correspondence: isal@mit.edu
MIT Media Lab, Massachusetts
Institute of Technology, 20 Ames
Street, Cambridge, MA, USA

Abstract

Massive migrations have become increasingly prevalent over the last decades. A recent example is the Venezuelan migration crisis across South America, which particularly affects neighboring countries like Colombia. Creating an effective response to the crisis is a challenge for governments and international agencies, given the lack of information about migrants' location, flows and behaviors within and across host countries. For this purpose it is crucial to map and understand geographic patterns of migration, including spatial mobility and dynamics over time. The aim of this paper is to uncover mobility and economic patterns of migrants that left Venezuela and migrated into Colombia due to the effects of the ongoing social, political and economic crisis. We analyze and compare the behavior of two types of migrants: Venezuelan refugees and Colombian nationals who used to live in Venezuela and return to their home country. We adapt the gravity model for human mobility in order to explain migrants' dispersion across Colombia, and analyze patterns of economic integration. This study is a first attempt at analyzing and comparing two kinds of migrant populations in one destination country, providing unique insight into the processes of mobility and integration after migration.

Keywords: Refugees, Migrants, Human mobility

Introduction

The Syrian Refugee crisis (Bakker et al. 2019) and the displacement of millions of Venezuelans across South America (Palotti et al. 2020) are two recent examples of massive migrations taking place around the world. Technological improvements in transportation, unabiding political unrest, armed conflicts and consequences of climate change will only increase the probability, frequency, and scale of such migrations (Pritchett 2003). Hence a better understanding about how they unfold in time and space is vital for their effective management in the near future.

Efforts by governments and international agencies often focus on delivering effective responses to the humanitarian crises that arise from forced migrations (Bryant 2005). While tending to migrants' urgent needs is undeniably important in the short-term, building the foundations for the future of these communities will have more permanent effects on the well-being of both migrants and natives (Alba and Nee 2009). However, policies designed for long-term impact have the potential to either perpetuate or ameliorate the

negative impact of migrations. For instance, the provision of refugee housing could influence how segregated and unequal these communities become over time (Balkan et al. 2018). Therefore, the design of long-term policies to tackle migration crises should be rooted in a deeper understanding of the complexity of these events.

In this paper, we study the behavior of migrants in Colombia coming from Venezuela between 2016 and 2019. We analyze migrants' mobility patterns across the host country and indicators of economic integration. In the analysis we include both Venezuelan migrants and Colombian nationals that used to live in Venezuela and return to their home country due to the effects of the Venezuelan social, economic and political crisis. This work is a first attempt at analyzing and comparing mobility and integration patterns from two different kinds of migrant populations with potentially different social capital in one destination country. In the remainder of the paper, we will refer to the first group as *migrants* and the latter group as *returnees*.

The paper is structured as follows. “**Background**” section contains a revision of the literature about migrations and the Venezuelan crisis. In “**Data**” section we describe the data and present the demographic properties of migrants and returnees. In “**Methods**” section we explain our modeling framework. In “**Stocks**” section we describe the results of migrants' spatial distribution and dynamics. “**Economic**” section is dedicated to economic integration including dynamics of unemployment rate and income differences. Finally, we conclude and summarize the results in “**Conclusion**” section. A Supplemental Information (SI) contains additional details on the methods and results presented in this work.

Background

Recent studies on international migration show the benefits that host countries can reap from the inflow of newcomers (Tabellini 2020). For example, some studies establish a link between migration and innovation (Kerr 2010), investment flows (Docquier and Rapoport 2012), and diversification of skills available in the host country's labor force (Reitz 2005). Other work has suggested that migration can help combat the effects of an aging population (Group and et al. 2016), and expand activities in community life by bringing new traditions, languages, and cuisine. A recent paper shows that the European immigration of the 20th century in the US increased local employment and industrial production without generating losses (Tabellini 2020). However, the authors show that despite the economic benefits, hostile policies and reactions still emerged, perhaps due to cultural differences between immigrants and natives.

Migration does not occur uniformly across the geographical space (Allen and Turner 1996). Migrants commonly cluster in specific cities for various reasons, including the existence of social and support networks of previous migrants (Yang et al. 2018; Blumenstock et al. 2019), closeness to the source of migration (Lu et al. 2012), regional cultural identity (Vermeulen et al. 2019; Aricat and Chib 2014), and economic opportunities. Quantifying the attractiveness of specific cities for internal and international migrants is a current challenge for migration studies (Prieto-Curiel et al. 2018).

The introduction of large scale datasets has enabled the observation of migration patterns across multiple countries over time (Abel and Sander 2014). International migrant flows have been analyzed by different types of population (Azose and Raftery 2019) as well as by gender (Abel 2018). Other covariates like unemployment and corruption have been

included in migration models in order to explain the structure of international flow networks (Malaj and de Rubertis 2017). While these studies provide a comprehensive view of international migration flows, they do not consider the internal displacements that occur when migrants arrive at destination, which are crucial for mapping out humanitarian crises (Risam 2019).

As part of internal displacements, previous studies show how local people change homes within their countries (Stillwell and Thomas 2016) but do not include recently arrived migrants. More recent studies using machine learning achieve accurate predictions of migration flows but lose explicability and hence any further actionable information (Robinson and Dilkina 2018). Applications of human mobility models include the prediction of future migrations due to climate crises (Robinson et al. 2020; Davis et al. 2018; Isaacman et al. 2018). Other models of migration include Markov processes (Constant and Zimmermann 2012), Agent-Based Models (Wrathall et al. 2019) and Decision Theory (Yan and Zhou 2019).

Previous studies on the Venezuelan migration crisis in Colombia use alternative data sources, such as online platforms like Facebook (Palotti et al. 2020; Spyrtatos et al. 2019) or Twitter (Hausmann et al. 2018) in order to estimate the number of migrants in a timely fashion. Other work has focused on quantifying the effects of migration on local economies and wages (Calderón-Mejía and Ibáñez 2016; Peñaloza Pacheco 2019), and the positive impact of incorporating young migrant workers into Colombia's aging workforce (Reina et al. 2018). Upcoming research, on the other hand, aims at understanding the effects on education outcomes (Namen et al. 2019). The spatial mobility of Venezuelan migrants in Colombia is relatively unexplored in this growing body of work.

In terms of migration policies, it is common for countries to restrict the inflow of displaced people and migrants. It has happened with migrants crossing the Mediterranean sea (Euronews 2018) or through the Balkans (Medecins Sans Frontieres 2016), as well as Central American people heading to the US (Medecins Sans Frontieres 2018; United Nations High Commissioner for Refugees 2019). However, the Venezuelan case is different. Colombia has been remarkably supportive to receive Venezuelan migrants since the beginning of the crisis (Baddour 2019; Prieto 2020) perhaps in reciprocity of a large migration of Colombians to Venezuela during the 80s and the 90s (El Tiempo 2018). As part of this support, the Colombian government publicly announced an acceleration of work permits approvals for those who participate in national surveys (Migracion Colombia 2018).

Data

The data used in this study was collected by the Colombian statistical bureau (DANE) as a part of their monthly household survey (GEIH) DANE Great Integrated Household Survey (GEIH) (2019). GEIH canvasses an approximate of 19,000 randomly selected households per month about a variety of topics divided into almost 24 different modules. We use the modules related to workforce, demographics, and migration, which are publicly available since 2006. The migration module only became publicly available recently and dates back to 2012. According to a press release by the Director of DANE, the question used to direct survey-takers to this module changed in 2016. Hence, in order to avoid introducing unknown biases or noise in our analysis we only use data from January 2016 until December 2019. In this time frame, the survey covers a total of 867,889

unique households and 2,878,194 unique individuals. The people in the survey are current residents of Colombia.

DANE provides expansion factors associated with each observation to scale sample statistics to the population level. For example, given the expansion factor of f_{exp} , the total population is given by $total_population = \sum_i f_{exp_i}$.

From the migration module we define migrants, returnees, and locals as follows:

- Migrants: Individuals born in Venezuela, that lived in Venezuela 5 years or 1 year ago.
- Returnees: Individuals born in Colombia, that lived in Venezuela 5 years or 1 year ago.
- Locals: Individuals born in Colombia, that lived in Colombia 5 years and 1 year ago.

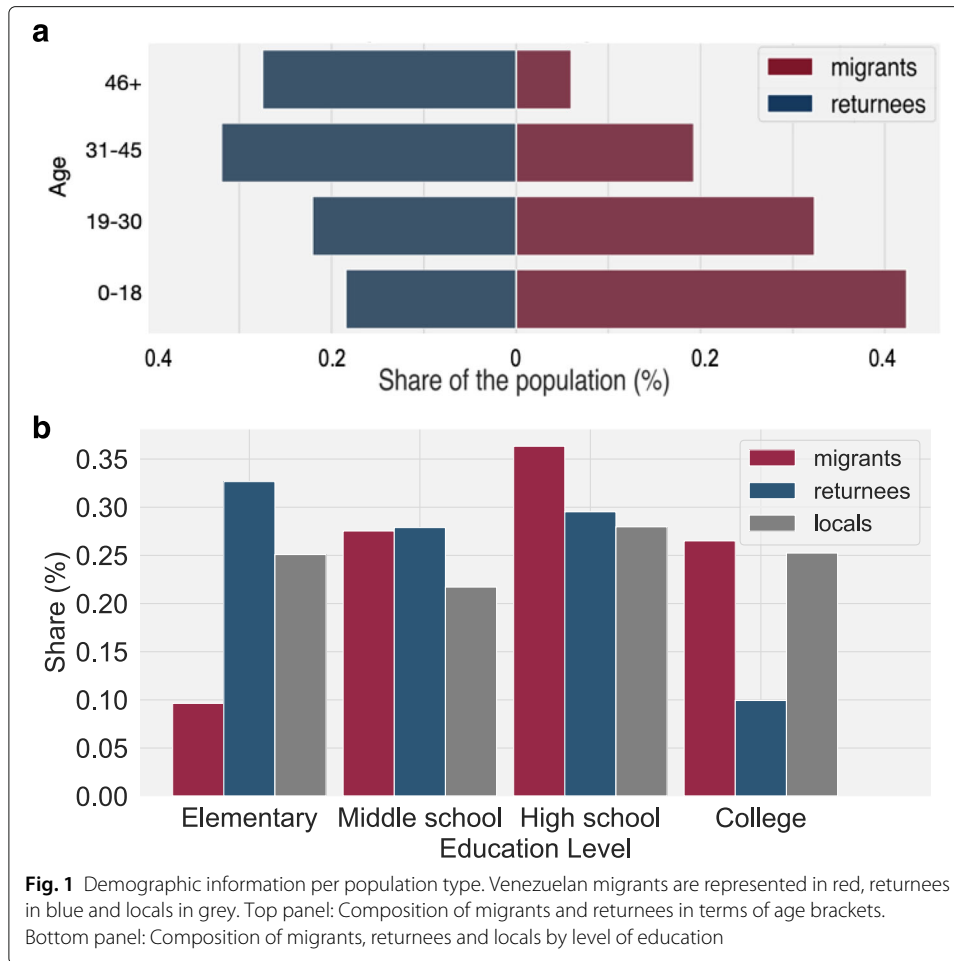
These definitions have some considerations. For example, we don't include foreign-born Colombians or Colombians who have lived in a foreign country different from Venezuela in the last five years. Moreover, we do not consider the floating population between Colombia and Venezuela. That is, we only account for migrants who have moved permanently to Colombia and don't oscillate between the two countries depending on the economic outlook.

The data has some limitations. A particularly relevant one for this work is the spatial granularity. Location is specified at the department level, and not at city level or finer. While regions do have modes in their capital cities, it is not a direct transfer of knowledge to city level insights. Furthermore, the data is only collected for 24 out of the 33 administrative divisions that make up present-day Colombia. Thus, the analysis does not include Amazonas, Arauca, Casanare, Guainía, Guaviare, Putumayo, Vaupé, Vichada, and the Archipelago of San Andres and Providencia. The combined population of the excluded regions represents 3% of the total population in Colombia according to DANE. Moreover, migrants without permanent housing and migrants living in shelters are likely not represented in these data. While this may not be a source of significant bias for the local population, it disproportionately affects the estimates of migrants and returnees who may be living under these conditions. Thus, this data may be underestimating the number of migrants and returnees.

Migrants and returnees differ in demographic features. Figure 1 describes both populations in terms of age (top) and education levels (bottom). Migrants (red) are mostly younger individuals with an average age of 28 years old and standard deviation of 11 years old. Returnees (blue) are an older group with an average age of 39 years old and standard deviation of 16 years old. Moreover, migrants are more educated both in terms of secondary and higher education (bottom panel in Fig. 1) than their Colombian counterparts. This information is consistent with previous reports created from other Colombian nation-wide surveys (Bahar et al. 2018).

Methods

Two prominent models are commonly used to understand human mobility patterns across geographical locations, namely, the gravity model (Lewer and Van den Berg 2008) and the radiation model (Simini et al. 2012). The gravity model describes the attraction between two places as being directly proportional to their respective populations and inversely proportional to the distance between them. The traditional gravity model is



expressed as follows:

$$A_{ij} \propto \frac{M_i^\alpha M_j^\beta}{D_{ij}^\gamma}, \tag{1}$$

where A_{ij} represents the attraction between location i and j (generally measured in terms of the amount of people who travel between locations), M_i and M_j represent the respective population of location i and j to the power of exponents α and β respectively, and D_{ij} represents the distance between both locations to the power of γ . The value of the exponents indicate the importance of each factor for determining the amount of human mobility among locations. The exponent γ is especially important because it describes the cost of distance for human mobility in the context of migration (Westerlund and Wilhelmsson 2011). Mobility decays more rapidly with distance as γ increases. Higher γ values indicate a further confinement of migration in space.

The parameters of the gravity model can be estimated by OLS regression after applying a logarithmic transformation:

$$\log(A_{ij}) = \beta_0 + \beta_1 \log(M_i) + \beta_2 \log(M_j) + \beta_3 \log(D_{ij}) + \epsilon_{ij}, \tag{2}$$

where $\beta_3 < 0$ and ϵ_{ij} represents the prediction errors.

The gravity model has been readily applied to many human endeavors like trade (Martínez-Zarzoso and Nowak-Lehmann 2003), transportation (Erlander and Stewart 1990), communication (Krings et al. 2009), and migration studies (Vermeulen et al. 2019; Karemera et al. 2000). In terms of migrations, a recent study shows that international migrants tend to move within clusters of countries and compute the expected migration flows among these clusters (Messias et al. 2016). Similar work (Belyi et al. 2017) shows that people tend to permanently relocate to a smaller set of countries of interest, while the set of country choices for touristic or other short-term motivations is much broader.

Previous research shows that the gravity model performs better than the radiation model in multiple scenarios (Bouchard and Pyers 1965) including migrations (Hankaew et al. 2019; Poot et al. 2016). The radiation model underestimates the flows across many modes of transportation, specially if destinations have larger populations (Masucci et al. 2013). Moreover, the applicability of the radiation model at larger scales and longer distances has been debated (Beiró et al. 2016). Counter to these arguments, other authors criticize the application of the gravity model for estimating spatial interactions matrices (Stefanouli and Polyzos 2017), such as commuting patterns and county migration. The effects of distance or population can be disproportionately weighted in the case of areas with large population densities and abundant opportunities (Simini et al. 2012). This is not necessarily the case of Latin American countries, where rural regions are generally not dense and underdeveloped. Other limitations of the gravity model include biases created by the logarithmic transformation (Burger et al. 2009) or the relationship between the scaling exponents (Deville et al. 2016).

In order to explain migrant mobility patterns in Colombia, we propose a modification of the gravity model for population flows. As opposed to the original gravity model, we do not consider the attraction between two places but only the *gravitational pull* of the destinations. Destinations are defined as the Colombian departments (analogous to states or provinces in other countries). The model is defined as follows:

$$a_j = \beta_0 + \beta_1 \log(M_j) + \beta_2 \log(D_j) + \beta_3 \log(G_j) + \epsilon_i. \quad (3)$$

The dependent variable a_j represents the attractiveness of the hosting department j and is measured in the number of individuals per type of population. M_j is the population size of department j , D_j is the distance between the department j and Caracas, Venezuela, and G_j represents the gross domestic product (GDP) per capita in department j . We chose GDP per capita in order to avoid issues of co-linearity with population size.

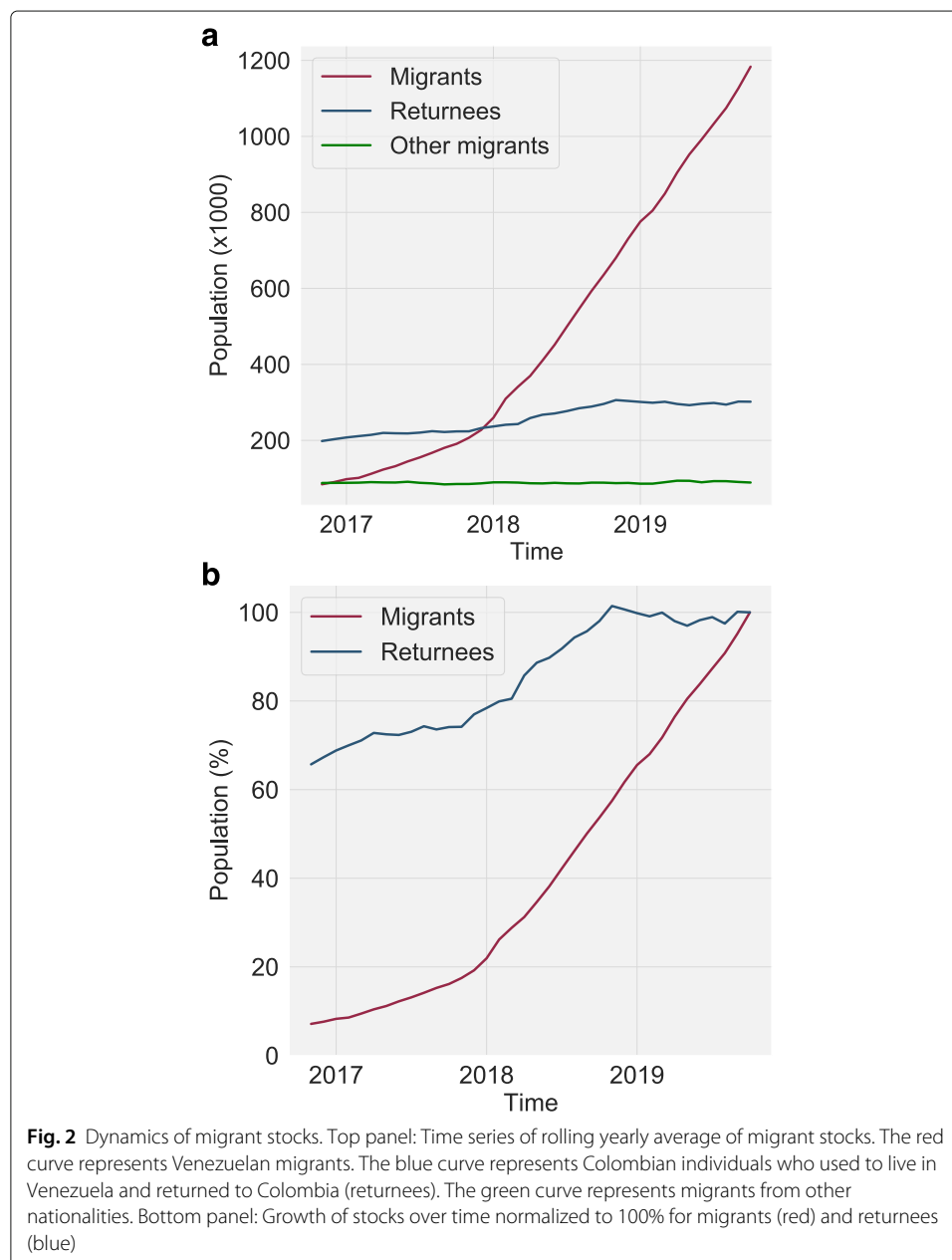
The city of Caracas is defined as the reference point in Venezuela because it is the capital and most populous city in Venezuela, housing approximately 10% of the country's population according to 2015 estimates. It is likely that present-day Caracas houses an even larger share of the Venezuelan population after their political and economic upheaval triggered an internal migration in search for livable conditions. We applied the model using other reference points in Venezuela, such as Maracaibo or Valencia (second and third largest cities respectively), and obtained consistent results as shown in the SI.

Previous research (Burger et al. 2009) shows that there are issues when estimating scarce mobility matrices given the high number of zeros. In our case, we are not trying to estimate mobility matrices but the attractiveness of certain regions based on their population, distance and economic opportunities from a single source, which reduces considerably

the number of zeros, as well as disregards other modeling methods such as the radiation model.

Stocks

The stocks of migrants measure the number of migrants in a department at a particular point in time. Stocks grow as more migrants flow into Colombia, and decrease as they leave the country. Figure 2 (top) depicts the starkly different trends in the growth of migrant and returnee stocks over time. The total number of migrants (red) grows faster than the number of returnees (blue) which shows a more steady pace since 2016. In 2019, the number of Venezuelan migrants in Colombia was over than 15 times larger than the number of Venezuelan migrants in early 2016. The number of returnees in late 2019 was

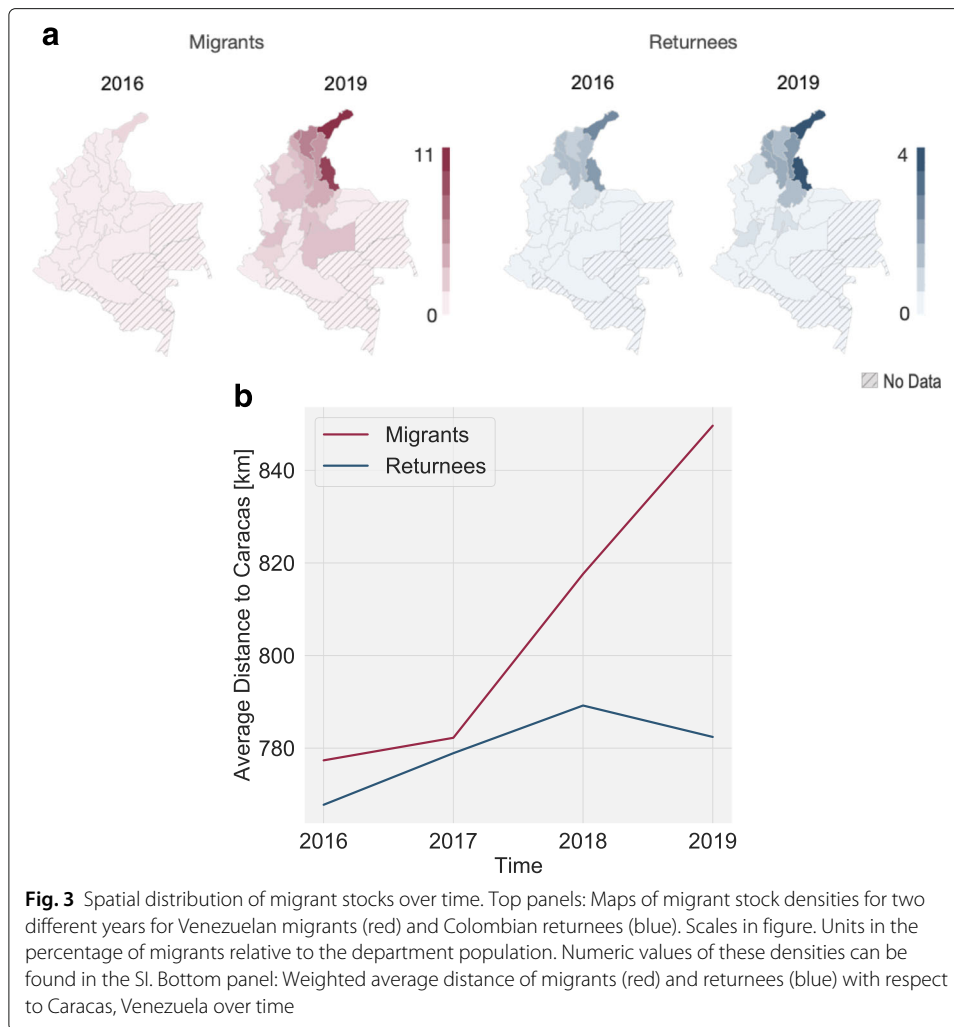


only approximately 1.5 times that of 2016. Migration from other countries to Colombia remained steady for the years in our data oscillating around one hundred thousand people (green curve).

Figure 2 (top) shows that returnees started moving to Colombia before the large influx of Venezuelan migrants. During 2016 and a fraction of 2017 the number of returnees is larger than the number of Venezuelan migrants, and only around mid-2017 does the number of Venezuelan migrants starts surpassing the number of returnees. Notice that the economic and political situation in Venezuela started deteriorating progressively since long before our observation period and that could have triggered a previous displacement of Colombian returnees. Economic difficulties have been previously reported as causes for the return of migrants to their home countries in other cases (Bandiera et al. 2013). In order to further understand the different migration dynamics of migrants and returnees, we normalize both series such that they reach 100% at the end of the observation period. The normalized results are presented in Fig. 2 (bottom). The curves show that returnees (blue) had saturated and reached 100% before migrants, which by the end of the observation period continue to grow steadily.

The difference between the flows of migrants and returnees is also present in their spatial mobility patterns. Figure 3 (top) shows maps with the spatial distribution of migrants (red) and returnees (blue) in 2016 and 2019 respectively. The distribution of migrants across the country changes more than the distribution of returnees. In 2019, migrants are more present in inner Colombian regions than returnees which remain clustered near the Venezuelan border in the north, and about as spatially constrained as in 2016. The values of the both population densities are shown in the SI. We quantify changes of the spatial distribution of individuals by measuring the weighted average of the distance of migrants and returnees to Caracas, Venezuela every year. The results are presented in Fig. 3 (bottom). Over time, the average distance to Caracas increases for migrants, while remains bounded for returnees. This happens as migrants head towards the inner-most parts of Colombia where big cities such as Bogotá, Medellín or Cali are located.

In Fig. 4 we present the relative number of migrants (x -axis) versus returnees (y -axis) per department (dots) in 2019. The dot size is proportional to the hosting department's population. The dot color indicates the distance to Caracas, Venezuela after respectively subtracting the average distance and dividing by the standard deviation across all departments. Blue dots represent departments closer to Venezuela, while green and red dots represent departments that are farther from the border. The vertical and horizontal dashed lines respectively show the expected relative number of migrants or returnees if the distributions were uniform across all departments. The diagonal line is the identity ($x=y$). The relative number of both migrants and returnees is higher in either departments hosting big cities (Cali, Medellín or Bogotá) or departments that are closer to Caracas (blue dots). There are differences between the two populations. The relative number of migrants in larger cities is larger than the relative number of returnees (annotated dots below the diagonal). In turn, the relative number of returnees is larger in departments closer to Venezuela than the relative number of migrants (blue dots above the diagonal). The exact relative numbers of migrants and returnees per department are presented in the SI.



Model

Migrants and returnees have different mobility patterns, suggesting that different mechanisms drive their spatial trajectories and spatial behaviors. We applied the model presented in Eq. 3 to the data of migrant stocks per department. We analyze the effects of each parameter of the model by progressively controlling for additional variables and evaluating changes in the magnitude of the regression coefficients. The results are presented in Table 1. Tables showing consistent results using other Venezuelan cities as reference points are included in the SI.

We first analyzed the effects of population size shown in the top four rows of Table 1. Across all years, migrants and returnees mobility is positively influenced by departmental population. The evolution of the magnitude of the coefficients and explained variance over time indicates that the importance of population size in the hosting department increases as migrants and returnees spend more time in Colombia. The coefficients also show that the effect of population on migrant mobility is larger for Venezuelan migrants than for returnees.

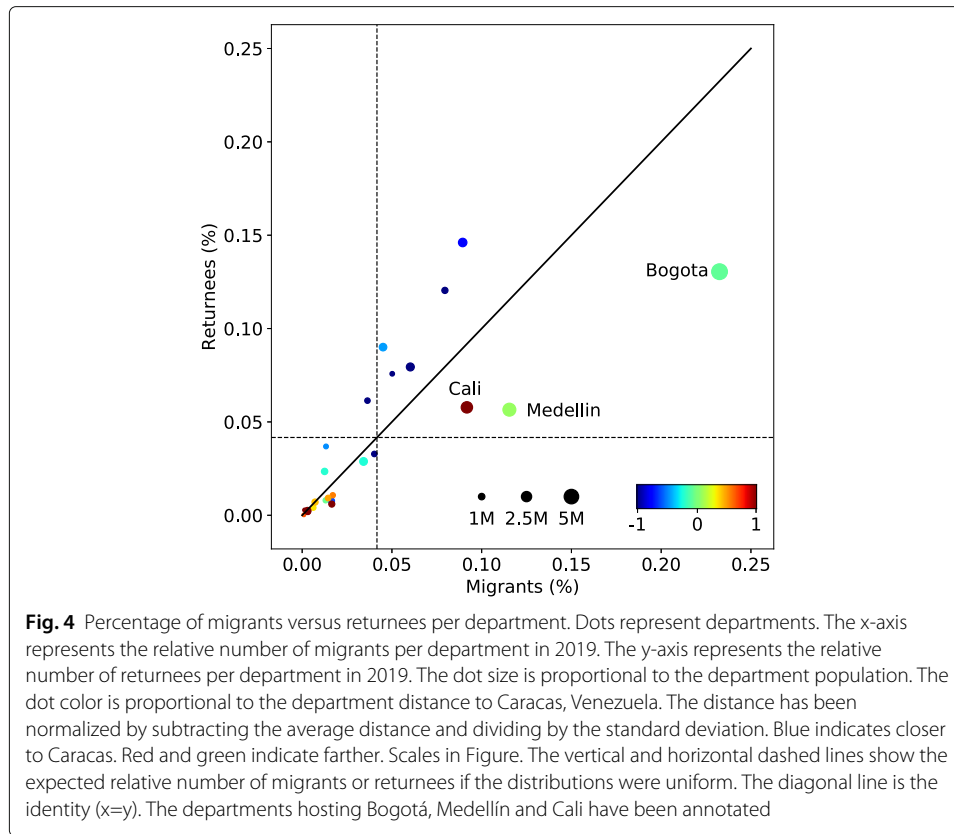


Table 1 Gravity model parameters after fitting the data at multiple years (rows) for both migrants (left columns) and returnees (right columns)

Year	Migrants				Returnees			
	Pop.	Dist.	GDPpc	R ²	Pop.	Dist.	GDPpc	R ²
2016	1.864**			0.268	1.070*			0.195
2017	1.698**			0.291	1.277*			0.260
2018	1.601***			0.452	1.096*			0.259
2019	1.499***			0.545	1.389**			0.349
2016	0.523***	-0.619***		0.652	0.448***	-0.756***		0.767
2017	0.542***	-0.709***		0.794	0.512***	-0.693***		0.740
2018	0.677***	-0.644***		0.866	0.515***	-0.723***		0.781
2019	0.741***	-0.568***		0.868	0.594***	-0.669***		0.796
2016	0.633***	-0.653***	-0.182	0.672	0.518***	-0.777***	-0.115	0.775
2017	0.532***	-0.705***	0.017	0.794	0.557***	-0.709***	-0.075	0.743
2018	0.665***	-0.639***	0.021	0.866	0.575***	-0.745***	-0.102	0.788
2019	0.720***	-0.561***	0.035	0.869	0.658***	-0.692***	-0.108	0.803

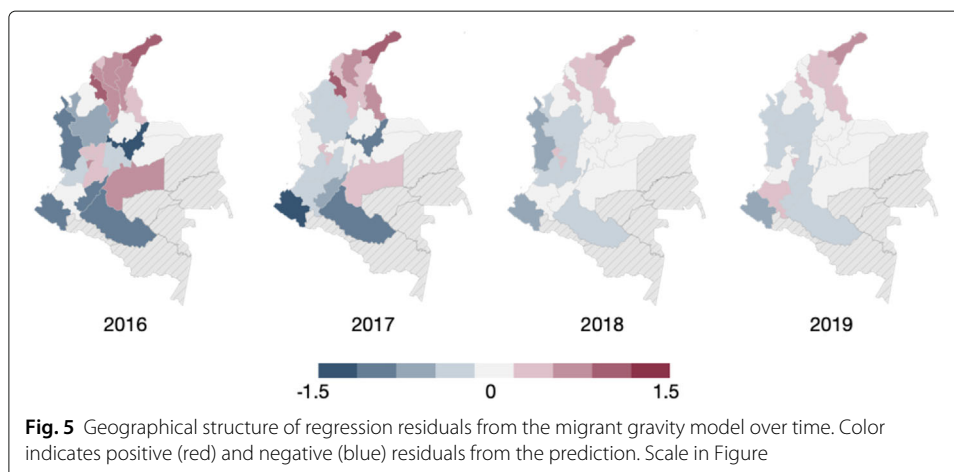
The dependent variable is the number of individuals per type of population (migrant or returnee) at each year for each Colombian department. Independent variables include (i) Population of the hosting department (Pop.), (ii) Distance of the hosting department to Caracas, Venezuela (Dist.) and (iii) the hosting department's GDP per capita (GDPpc). Independent variables have been added progressively. The explained variance (R²) is shown in the table for each model realization. The significance levels are indicated next to the coefficients: *p < 0.05, **p < 0.01 and ***p < 0.001. Tables showing results using other Venezuelan cities as reference points are included in the SI

We then add the parameter representing the distance to Caracas, Venezuela. The results are shown in the middle rows of Table 1. The sign of the distance coefficient is consistently negative across all cases. Because we apply regression over logarithmic values, a negative coefficient indicates that the number of migrants or returnees is inversely proportional to the distance between the hosting department and Venezuela. For migrants, the importance of distance as a constraint for their mobility across Colombia decreases over time, while for returnees the distance coefficient remains more or less constant. The lack of significant changes in the distance exponent for returnees suggests that the effect of distance on returnee mobility remains constant in time rather than being unimportant.

Finally, we add the third parameter representing the department-level GDP per capita. The results are shown in the bottom rows of Table 1. GDP per capita is not an important descriptor of migrant mobility. It is not statistically significant for any of the years considered and either increases the model explained variance. This suggests that other factors, such as social networks, might be accounting of the remaining variance for determining where migrants will move to. A previous report on international migration shows that social networks are one of the strongest predictors for choosing the destination country independently of the income level of the origin country (Migali and et al. 2018). The same report shows that the geographical distance between the origin and the destination is more important for low income countries than for middle or high income countries, and that the destination GDP per capita growth is not a determining factor for choosing where to migrate.

The prediction of migrant stocks improves over time. The three versions of the model increase their explained variance in approximately 20% from 2016 to 2019. This is evident when plotting model residuals (Fig. 5). The residuals are spatially correlated across the whole observation period but their magnitude significantly decreases over time. We believe it takes a couple of years for the model to become better at predicting mobility because migrants in Colombia tend to move on foot and make many stops before reaching their final destination. Thus, during 2016 and 2017 most had recently arrived in Colombia and were making plans to move towards the inner-part of the country, but it took them months and even a year to gather the resources to do so.

One limitation of our model is that it under estimates the number of migrants in departments near the border and tends to overestimate the number of migrants in areas that



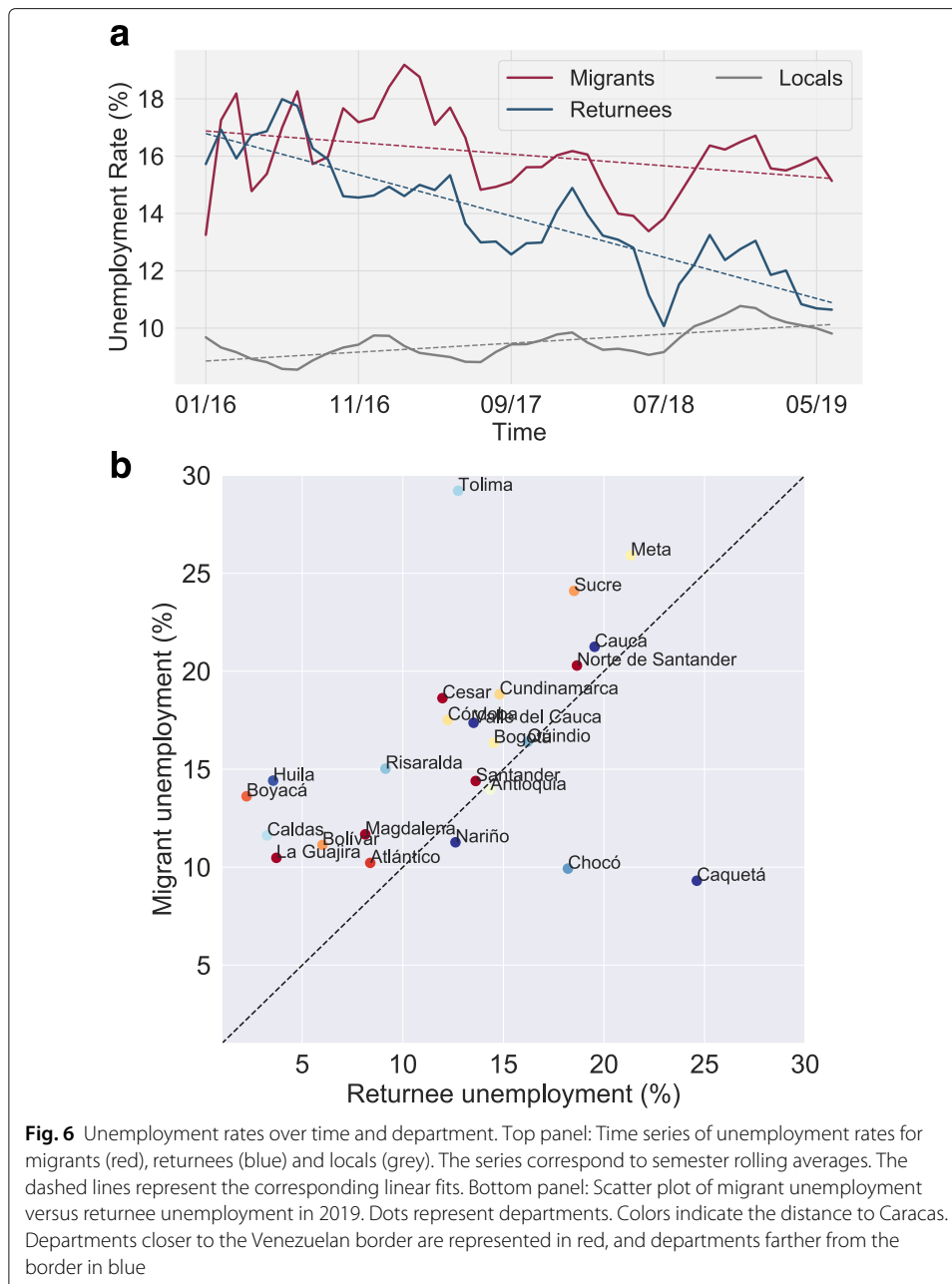
are further from Venezuela (see the residuals in Fig. 5). This is part of a more general limitation of analyses based on linear regression that produce single coefficients for each feature that average the effects of the different variables across all samples. The biases however are reduced as time advances and the magnitude of the residuals decrease across the whole territory.

Besides the factors included in Eq. 3, other variables such as politics could also explain particularities of massive migrations. On a larger scale, a recent study shows that Venezuelan migrants choose their country of destination taking into account the ideology of the ruling party (Palotti et al. 2020). The number of migrants in countries whose Presidents are known to strongly support the Venezuelan regime represents a tiny fraction of those living in countries that openly oppose them. Such is the case of Bolivia and Nicaragua in comparison to Peru or Panama. These countries respectively show similar profiles in terms of geographical location, territory, and population, and mainly differ in the ideology of their leaders.

Economic

We investigate the economic integration of migrants and returnees by measuring unemployment rates and differences in their monthly average and median income. Unemployment rate is calculated for each type of migrant as the relative number of unemployed individuals relative to the total economically active population in each cohort. Figure 6 (top) shows the national unemployment rate for migrants (red), returnees (blue) and locals (grey) over time. In 2016 both migrants and returnees showed similar unemployment rates around 17%. Since then a change of regime happened. Returnees decreased their unemployment rates at a faster pace than migrants (slopes of -1.04% vs -0.4% monthly unemployment decrease respectively). As a result, the gap between returnees and locals unemployment rates is drastically reduced in a period of four years, while the gap between migrants and locals remained significantly higher. A potential explanation for this trend includes the rapid growth of the migrant population after 2016 (see Fig. 2). However, in the SI we show that the ratio between employed migrants and the active migrant population is consistently lower than the ratio between employed returnees and the active returnee population across the whole observation period.

In order to further understand the patterns of unemployment between migrants and returnees we disaggregated the rates by department. Figure 6 (bottom) compares the unemployment rates for migrants (y -axis) and returnees (x -axis) in 2019 at different locations. Departments are represented by dots and colored by their distance to Caracas, Venezuela. Red dots show departments closer to the Venezuelan border and blue dots show departments farther from the Venezuelan border. Most departments are above the diagonal line which means that the rate of migrant unemployment is higher than the one for returnees in these locations. In departments closer to Venezuela, where the returnee population is dense (see Fig. 3), the unemployment rate of migrants remarkably exceeds the one for returnees (see La Guajira, Boyaca, Cesar or Sucre). On the other hand, in departments with large cities such as Bogotá, Antioquia (Medellín), or Valle del Cauca (Cali), the rates of unemployment for the two types of population are similar to each other. This result suggests that migrants could be flowing toward large cities looking for economic opportunities since jobs near the border, which are less developed areas, could be taken by returnees.



We also measured differences in median monthly income among the three populations across the whole observation period using propensity score matching (Becker and Ichino 2002). This method aims at approximating a randomized experiment using observational data. It divides the population into treatment and control groups, while controlling for possible confounding variables such as demographics. Thus, here it helps reducing estimation biases of the average effect of being a migrant. We define migrants and returnees as the treatment group and locals as the control group. We matched pairs of individuals from both groups based on age, gender, year of appearance in the data, department, and level of education. To check the quality of the matches, we used computed a chi-squared test of independence for these variables before and after matching and confirmed that all

variables are balanced after matching. Then, we bootstrapped 100 samples made up of 60% of the data and applied a two-sided t-test. See the SI for more information on the matching process.

Table 2 shows the relative difference in value of the median monthly income between migrants and locals, returnees and locals, as well as migrants and returnees. Migrants and returnees show significantly lower median monthly incomes than their local counterparts. Moreover, the gap between returnees and locals is smaller than the gap between migrants and locals. This shows that returnees consistently show economic advantages over migrants and should be considered as a different type of population in terms of migration and integration policies.

According to recent reports (Bahar et al. 2018) and our data, Venezuelan migrants in Colombia tend to be more educated than locals and returnees (see “Data” section). Therefore, the differences in unemployment rate and income should not be explained by lacking training or skills but more probable because of returnees’ pre-existing social networks (which are initially inaccessible for Venezuelans). However, the data also shows that Venezuelan migrants are generally younger individuals. In the US, immigrants who arrive before the age of 25 considerably reduce the time required to equal the economic conditions of natives (Hatton 1997). Nevertheless, despite taking longer for certain groups, the full migrant population may reduce and eliminate income gaps over time (Minns 2000; Abramitzky et al. 2014).

Discussion

According to previous reports, international migrations are mainly explained by the GDP of the country of origin, social networks in the destination country and demographic changes (Migali and et al. 2018). These reports also note that armed conflicts, state fragility and consequent poverty are drivers for migrants seeking asylum. Venezuela’s fragility index was the most-worsened one in 2019 (Messner and et al. 2019) and Venezuelans represent the second highest nationality seeking asylum in 2018 across the whole world (Alto Comisionado de las Naciones Unidas para los Refugiados (UNHCR) 2018). The destination chosen by Venezuelans varies by socio-economic status (Palotti et al. 2020). Nearby countries, such as Colombia, are preferred by low-income migrants who travel by land, while far-flung and more economically developed countries are chosen by higher-income individuals. Despite exploding recently, the Venezuelan migration has been developing over multiple years, which may have influenced a prior development of social networks.

The choice of Colombia as the main destination country for Venezuelans has several elements that characterize it with respect to other migrational cases. On the one hand,

Table 2 Median income difference among migrants, returnees and locals

Treatment	Control	Median Difference
Migrants	Locals	-10%*
Returnees	Locals	-5%*
Migrants	Returnees	-8%*

The table shows the relative difference of the treatment group (first column) against the control group (second column). The p -values of the median comparison result from Mood’s tests. *These differences are statistically significant ($p < 0.001$)

Colombia is the neighboring country, and the choice of nearby destinations is especially important for lower-income populations (Migali and et al. 2018). On the other hand, Colombia has very strong ties with Venezuela. Both countries share many cultural affinities, such as language, common history, and a long history of trade and social exchange. This trend is contrary to the international context, where cultural distance does not always play an important role for choosing a destination. However, Brazil or Guyana also share borders with Venezuela and do not account for a comparable number of migrants. Other destination countries, such as Spain, Panama and the United States, show that social networks are more important than distance, especially for people with more resources.

Our results show that the internal distribution of migrants and returnees in Colombia follows patterns that are consistent with international migration. The choice of destination department is not influenced by GDP per capita, but rather by population, distance from Venezuela and by possibly the social network available to the migrant. In the case of returnees, we see that departments near Venezuela are more attractive than big cities. Our data do not have information on the origin of returnees before their original migration to Venezuela. However, previous reports show that the Colombian migration in Venezuela generally came from departments near the border such as Atlantico, Bolivar and Norte de Santander (Santana Rivas 2009). A previous study also finds that in other cases returnees go back to regions they are from (Hausmann and Nedelkoska 2018).

Previous studies covering the economic consequences of migrations for returnees and locals quantify the difference in wages between return migrants and non-migrants (Nekby 2006; De Coulon and Piracha 2005; Biavaschi 2016), or the effect on wages that return migration has on home job markets (Hausmann and Nedelkoska 2018; Zhao 2002). They find that returnees usually earn more (Hausmann and Nedelkoska 2018; Biavaschi 2016) or come back with more or different skills (Riccardo and et al. 2017). In our case, returnees make less money than locals, but more than Venezuelan migrants. This is probably explained by the complementarity of skills described in (Hausmann and Nedelkoska 2018). In this paper we compare the difference in income, unemployment and employment rates by between locals, foreign and return migrants. Thus adding an additional population to the analysis, which is not common for return migration studies. Moreover, the return migrants and migrants in our data are more comparable than other studies since they are coming around the same time and from the same country. Other studies are focused in estimating optimal period of migration (Dustmann 2003) and motives for migration (Dustmann and Weiss 2007) or return (Anghel and et al. 2016; De Haas and Fokkema 2011; Riccardo and et al. 2017).

One main limitation of our study is that we have only access to the number of migrants and returnees per department at multiple points in time and cannot measure social relationships among the population. We do not have either access to detailed trajectories followed by individuals across departments. Access to data from electronic communication such as social media or mobile phones could enable a further understanding of the structure and dynamics of social networks and their contribution for the agglomeration and integration of migrants in the hosting departments.

Conclusion

In this paper we explained and compared the mobility patterns of two different migrant populations in Colombia: Venezuelan migrants and Colombian nationals who return to their home country from Venezuela. Findings suggest that migrants and returnees have different drivers behind their mobility patterns and destination choices. While migrants are less constrained by distance and more attracted to larger cities, returnees seem to stay in places where they are from and have social networks.

We also analyzed the economic integration of these individuals and found that migrants are less integrated than returnees despite having slightly higher levels of education. We hypothesize that such difference in economic integration is due to the advantages of having pre-existing social networks. The resolution of the available data does not allow a deeper investigation of the effects of social networks on the evolution of spatial and economic integration in migrants. However, given the challenges that hosting societies face in the context of sudden migrations, these findings highlight their importance for future research.

Acknowledgements

The authors would like to thank Morgan Frank for the helpful comments as well as Pearl Li for helping with data preparation procedures. Finally, we thank DANE for providing the data used in this analysis.

Author's contributions

ILS, AJM and ASP conceptualized the study. ILS and MN cleaned and processed the data and implemented the models. ILS, MN and AJM visualized the data and wrote and edited the draft. All authors reviewed the final manuscript. All authors read and approved the final manuscript.

Funding

Funding comes from MIT Connection Science research alliance. The authors declare that no outside body impacted the contents of this study.

Availability of data and materials

All data analysed comes from DANE Colombia. General GEIH modules can be found under DANE microdatos service at http://microdatos.dane.gov.co/index.php/catalog/MICRODATOS/about_collection/23/1. Migration-specific GEIH modules can be found under http://microdatos.dane.gov.co/index.php/catalog/MICRODATOS/about_collection/42/1.

Competing interests

The authors declare that they have no competing interests.

Received: 17 March 2020 Accepted: 16 August 2020

Published online: 22 September 2020

References

- Abel GJ (2018) Estimates of global bilateral migration flows by gender between 1960 and 2015. *Int Migr Rev* 52(3):809–852. <https://doi.org/10.1111/imre.12327>
- Abel GJ, Sander N (2014) Quantifying global international migration flows. *Science* 343(6178):1520–1522
- Abramitzky R, Boustan LP, Eriksson K (2014) A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *J Polit Econ* 122(3):467–506
- Alba RD, Nee V (2009) *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Harvard University Press
- Allen JP, Turner E (1996) Spatial patterns of immigrant assimilation*. *Prof Geogr* 48(2):140–155. <https://doi.org/10.1111/j.0033-0124.1996.00140.x>
- Alto Comisionado de las Naciones Unidas para los Refugiados (UNHCR) (2018) *Tendencias Globales Desplazamiento Forzado en 2018*. <https://www.eltiempo.com/mundo/venezuela/anteriormente-la-ola-migratoria-era-de-colombianos-hacia-venezuela-181258>. Accessed 20 Apr 2020
- Anghel R, et al. (2016) International migration, return migration, and their effects: a comprehensive review on the Romanian case. *IZA Discussion Papers* 1(10445):51
- Aricat RG, Chib A (2014) Social integration of male migrant workers in Singapore: The enabling and constraining roles of mobile phones. In: *Proceedings Annual Workshop of the AIS Special Interest Group for ICT in Global Development*. Paper 9
- Azose JJ, Raftery AE (2019) Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proc Natl Acad Sci* 116(1):116–122. <https://doi.org/10.1073/pnas.1722334116>
- Baddour D (2019) *The Atlantic Colombia Radical Plan to Welcome Millions of Venezuelan Migrants*. <https://www.theatlantic.com/international/archive/2019/01/colombia-welcomes-millions-venezuelans-maduro-guaido/581647/>. Accessed 12 Mar 2020

- Bahar D, Dooley M, Huang C (2018) Integración de los venezolanos en el mercado laboral colombiano. mitigando costos y maximizando beneficios. Brookings Global Economy and Development, Washington, DC
- Bakker MA, Piracha DA, Lu PJ, Bejgo K, Bahrami M, Leng Y, Balsa-Barreiro J, Ricard J, Morales AJ, Singh VK, et al. (2019) Measuring fine-grained multidimensional integration using mobile phone metadata: the case of Syrian refugees in Turkey. In: Guide to Mobile Data Analytics in Refugee Scenarios. Springer. pp 123–140
- Balkan B, Tok E, Torun H, Tumen S (2018) Immigration, housing rents, and residential segregation: evidence from Syrian refugees in Turkey. IZA Discussion Paper. <https://www.msf.org/greecebalkans-thousands-stranded-new-arbitrary-border-restrictions-expose-refugees-violence>
- Bandiera O, Rasul I, Viarengo M (2013) The making of modern America: Migratory flows in the age of mass migration. *J Dev Econ* 102(C):23–47
- Becker S, Ichino A (2002) Estimation of average treatment effects based on propensity scores. *Stata J* 2(4):358–377
- Beiró MG, Panisson A, Tizzoni M, Cattuto C (2016) Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Sci* 5(1):30
- Belyi A, Bojic I, Sobolevsky S, Sitko I, Hawelka B, Rudikova L, Kurbatski A, Ratti C (2017) Global multi-layer network of human mobility. *Int J Geogr Inf Sci* 31(7):1381–1402
- Biavaschi C (2016) Recovering the counterfactual wage distribution with selective return migration. *Labour Econ* 38:59–80
- Blumenstock JE, Chi G, Tan X (2019) Migration and the value of social networks. CEPR Discussion Paper No. DP13611. CEPR Discussion Papers
- Bouchard RJ, Pyers CE (1965) Use of gravity model for describing urban travel. *Highw Res Rec* 88:1–43
- Bryant J (2005) Children of international migrants in Indonesia, Thailand and the Philippines: A review of evidence and policies. Innocenti Working Papers, No. 2005/05, UN, New York. <https://doi.org/10.18356/83031a60-en>
- Burger M, Van Oort F, Linders G-J (2009) On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spat Econ Anal* 4(2):167–190
- Calderón-Mejía V, Ibáñez AM (2016) Labour market effects of migration-related supply shocks: evidence from internal refugees in Colombia. *J Econ Geogr* 16(3):695–713
- Constant AF, Zimmermann KF (2012) The dynamics of repeat migration: A markov chain analysis. *Int Migr Rev* 46(2):362–388
- DANE Great Integrated Household Survey (GEIH) (2019). http://microdatos.dane.gov.co/index.php/catalog/MICRODATOS/about_collection/23/1
- Davis KF, Bhattachan A, D'Odorico P, Suweis S (2018) A universal model for predicting human migration under climate change: examining future sea level rise in Bangladesh. *Environ Res Lett* 13(6):064030
- Deville P, Song C, Eagle N, Blondel VD, Barabási A-L, Wang D (2016) Scaling identity connects human mobility and social interactions. *Proc Natl Acad Sci* 113(26):7047–7052. <https://doi.org/10.1073/pnas.1525443113>
- De Coulon A, Piracha M (2005) Self-selection and the performance of return migrants: the source country perspective. *J Popul Econ* 18(4):779–807
- De Haas H, Fokkema T (2011) The effects of integration and transnational ties on international return migration intentions. *Demogr Res* 25:755–782
- Director of DANE Press Release Question. <https://www.facebook.com/DANEColombia/videos/3366521323562668>. Accessed 17 Sept 2019
- Director of DANE Press Release Question (2019). <https://surveyinsights.org/?p=6255>
- Docquier F, Rapoport H (2012) Globalization, brain drain, and development. *J Econ Lit* 50(3):681–730
- Dustmann C (2003) Return migration, wage differentials, and the optimal migration duration. *Eur Econ Rev* 47(2):353–369
- Dustmann C, Weiss Y (2007) Return migration: theory and empirical evidence from the UK. *Br J Ind Relat* 45(2):236–256
- El Tiempo (2018) Así se vivía cuando la ola migratoria era de Colombia hacia Venezuela. <https://www.eltiempo.com/mundo/venezuela/anteriormente-la-ola-migratoria-era-de-colombianos-hacia-venezuela-181258>
- Erlander S, Stewart NF (1990) The Gravity Model in Transportation Analysis: Theory and Extensions, vol 3. VSP, Utrecht
- Euronews (2018) This ship is only going to see Italy by postcard. <https://www.euronews.com/2018/06/22/-this-ship-is-going-to-see-italy-only-by-postcard-italian-minister-s-comments-spark-new-mi>. Accessed 12 Apr 2020
- Group WB, et al. (2016) Global monitoring report 2015/2016: Development goals in an era of demographic change. World Bank Washington, DC
- Hankaew S, Phithakitnukoon S, Demissie MG, Kattan L, Smoreda Z, Ratti C (2019) Inferring and modeling migration flows using mobile phone network data. *IEEE Access* 7:164746–164758
- Hatton T (1997) The Immigrant Assimilation Puzzle in Late Nineteenth-Century America. *J Econ Hist* 57(1):34–62
- Hausmann R, Hinz J, Yildirim MA (2018) Measuring Venezuelan emigration with Twitter. Kiel Working Paper 2106, Harvard University, Kiel
- Hausmann R, Nedelkoska L (2018) Welcome home in a crisis: Effects of return migration on the non-migrants' wages and employment. *Eur Econ Rev* 101:101–132
- Isaacman S, Frias-Martinez V, Frias-Martinez E (2018) Modeling human migration patterns during drought conditions in la Guajira, Colombia. In: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. Association for Computing Machinery, New York. pp 1–9
- Karemera D, Oguledo VI, Davis B (2000) A gravity model analysis of international migration to North America. *Appl Econ* 32(13):1745–1755. <https://doi.org/10.1080/000368400421093>
- Kerr WR (2010) Breakthrough inventions and migrating clusters of innovation. *J Urban Econ* 67(1):46–60
- Krings G, Calabrese F, Ratti C, Blondel VD (2009) Urban gravity: a model for inter-city telecommunication flows. *J Stat Mech Theory Exp* 2009(07):07003. <https://doi.org/10.1088/1742-5468/2009/07/07003>
- Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 Haiti earthquake. *Proc Natl Acad Sci* 109(29):11576–11581
- Lewer JJ, Van den Berg H (2008) A gravity model of immigration. *Econ Lett* 99(1):164–167
- Malaj V, de Rubertis S (2017) Determinants of migration and the gravity model of migration – application on western Balkan emigration flows. *Migr Lett* 14(2):204–220. <https://doi.org/10.33182/ml.v14i2.327>
- Martínez-Zarzoso I, Nowak-Lehmann F (2003) Augmented gravity model: An empirical application to Mercosur-European union trade flows. *J Appl Econ* 6(2):291–316

- Masucci AP, Serras J, Johansson A, Batty M (2013) Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Phys Rev E* 88(2):022812
- Medecins Sans Frontieres (2016) Thousands stranded as new arbitrary border restrictions expose refugees to violence. <https://www.msf.org/greecebalkans-thousands-stranded-new-arbitrary-border-restrictions-expose-refugees-violence>. Accessed 26 June 2020
- Medecins Sans Frontieres (2018). <https://www.msf.org/us-asylum-restrictions-are-deepening-mexican-border-crisis-migration-refugees>. Accessed 27 June 2020
- Messias J, Benevenuto F, Weber I, Zagheni E (2016) From migration corridors to clusters: The value of google+ data for migration studies. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp 421–428
- Messner J., et al. (2019) Fragile States Index 2019: The Book. Washington, DC: Fund for Peace. <https://fundforpeace.org/2019/04/10/fragile-states-index-2019/#:~:text=The%20Fragile%20States%20Index%2C%20produced,towards%20the%20brink%20of%20failure>. Accessed 9 Mar 2019
- Migali S, et al. (2018) International Migration Drivers. EUR 29333 EN, Publications Office of the European Union. <https://doi.org/10.2760/63833>
- Migracion Colombia (2018) 442.462 Venezolanos identificados en registro RAMV recibirán regularización temporal. <http://es.presidencia.gov.co/noticia/180613-442462-venezolanos-identificados-en-registro-RAMV-recibiran-regularizacion-temporal>
- Minns C (2000) Income, cohort effects, and occupational mobility: a new look at immigration to the United States at the turn of the 20th century. *Explor Econ Hist* 37(4):326–350
- Namen O, Prem M, Roza S, Vargas JF (2019) Effects of Venezuelan migration on education outcomes in Colombia. Research Prosals Inter-American Bank of Development
- Nekby L (2006) The emigration of immigrants, return vs onward migration: evidence from Sweden. *Journal of Population Economics* 19(2):197–226
- Palotti J, Adler N, Morales-Guzman A, Villaveces J, Sekara V, Garcia Herranz M, Al-Asad M, Weber I (2020) Monitoring of the Venezuelan exodus through Facebook's advertising platform. *PLoS ONE* 15(2):1–15. <https://doi.org/10.1371/journal.pone.0229175>
- Peñalosa Pacheco L (2019) Living with the neighbors: the effect of Venezuelan forced migration on wages in Colombia. No. 248. Documentos Trabajo CEDLAS
- Poot J, Alimi O, Cameron MP, Maré DC (2016) The gravity model of migration: the successful comeback of an ageing superstar in regional science. *Investig Regionales* 63:86
- Prieto N (2020) Proyecto Migracion Venezuela Colombia pide apoyo internacional para atender migrantes venezolanos ante pandemia. <https://migravenezuela.com/web/articulo/colombia-pide-apoyo-internacional-para-atender-migrantes-venezolanos-ante-pandemia/1816>
- Prieto-Curiel R, Pappalardo L, Gabrielli L, Bishop SR (2018) Gravity and scaling laws of city to city migration. *PLoS ONE* 13(7):1–19. <https://doi.org/10.1371/journal.pone.0199892>
- Pritchett L (2003) The Future of Migration: Irresistible Forces meet Immovable Ideas. In: Paper presented to the conference The Future of Globalization: Explorations in Light of the Recent Turbulence, Yale University, Center for the Study of Globalization. http://ksghome.harvard.edu/~lpritch.academic.ksg/future_talk.doc. Accessed 15 Feb 2020
- Reina M, Mesa CA, Ramírez T, et al. (2018) Elementos para una política pública frente a la crisis de Venezuela. Cuadernos de Fedesarrollo, Bogotá
- Reitz JG (2005) Tapping immigrants' skills: New directions for Canadian immigration policy in the knowledge economy. *Law Bus Rev Am* 409:432
- Riccardo C, et al. (2017) Why do they return? Beyond the economic drivers of graduate return migration. *Ann Reg Sci* 59(3):603–627
- Risam R (2019) Beyond the migrant "problem": Visualizing global migration. *Telev New Media* 20(6):566–580. <https://doi.org/10.1177/1527476419857679>
- Robinson C, Dilkina B (2018) A machine learning approach to modeling human migration. In: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. COMPASS '18. Association for Computing Machinery, New York. <https://doi.org/10.1145/3209811.3209868>
- Robinson C, Dilkina B, Moreno-Cruz J (2020) Modeling migration patterns in the USA under sea level rise. *PLoS ONE* 15(1):1–15. <https://doi.org/10.1371/journal.pone.0227436>
- Santana Rivas D (2009) Geografía de la inmigración venezolana en Colombia entre 1993 y 2008. *Aracne. Revista Electrónica de Recursos en Internet sobre Geografía y Ciencias Sociales* 124
- Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96–100
- Spyratos S, Vespe M, Natale F, Weber I, Zagheni E, Rango M (2019) Quantifying international human mobility patterns using Facebook network data. *PLoS ONE* 14(10):1–22. <https://doi.org/10.1371/journal.pone.0224134>
- Stefanouli M, Polyzos S (2017) Gravity vs radiation model: two approaches on commuting in Greece. *Transp Res Proc* 24(Supplement C):65–72
- Stillwell J, Thomas M (2016) How far do internal migrants really move? demonstrating a new method for the estimation of intra-zonal distance. *Reg Stud Reg Sci* 3(1):28–47. <https://doi.org/10.1080/21681376.2015.1109473>
- Tabellini M (2020) Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *Rev Econ Stud* 87(1):454–486
- United Nations High Commissioner for Refugees (2019) UNHCR deeply concerned about new US asylum restrictions. <https://www.unhcr.org/news/press/2019/7/5d2cdf114/unhcr-deeply-concerned-new-asylum-restrictions.html>
- Vermeulen W, Roy D, Quax R (2019) Modelling the influence of regional identity on human migration 3:78. <https://doi.org/10.3390/urbansci3030078>
- Wrathall D, Mueller V, Clark PU, Bell A, Oppenheimer M, Hauer M, Kulp S, Gilmore E, Adams H, Kopp R, et al (2019) Meeting the looming policy challenge of sea-level change and human migration. *Nat Clim Chang* 9(12):898–901

- Westerlund J, Wilhelmsson F (2011) Estimating the gravity model without gravity using panel data. *Appl Econ* 43(6):641–649. <https://doi.org/10.1080/00036840802599784>
- Yan X-Y, Zhou T (2019) Destination choice game: A spatial interaction theory on human mobility. *Sci Rep* 9(1):1–9
- Yang Y, Liu Z, Tan C, Wu F, Zhuang Y, Li Y (2018) To stay or to leave: Churn prediction for urban migrants in the initial period. In: *Proceedings of the 2018 World Wide Web Conference*. pp 967–976
- Zhao Y (2002) Causes and consequences of return migration: recent evidence from China. *J Comp Econ* 30(2):376–394

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
