DOCUMENTOS DE TRABAJO SOBRE

ECONOMÍA REGIONAL Y URBANA

Locked up? The development and internal migration nexus in Colombia

By: Karina Acosta
Hengyu Gu

No. 304
January, 2022

BANCO DE LA REPÚBLICA

CENTRO DE ESTUDIOS ECONÓMICOS REGIONALES (CEER) - CARTAGENA

Locked up? The development and internal migration nexus in Colombia*

Karina Acosta† **Hengyu Gu**‡

Abstract

Although a sizable number of studies have been exploring the migration development nexus in international settings, there is still a reduced number on internal contexts in recent years. This research aims to estimate the causal effect of origin economic conditions on internal population migration using a time series of the Colombian states between 2012 and 2019. This analysis provides a macro perspective of associations and causation between population dynamics and development in the current changes observed using spatial interaction models. Likewise, it analyses the current portray of internal migration in Colombia (defined by five-years and one-year flows). The evidence shows that the migration hump depends on the scale at which it is analyzed. At an aggregated scale, initial economic conditions are negatively associated with migration until a threshold where this relationship is reversed. The opposite is observed in the rural migrants subsample.

Keywords: internal migration, development, migration hump, spatial interaction models, Colombia.

JEL classification: O15, O1, C21.

---

* The authors thank earlier discussions and revisions from Ravi Kanbur, Filiz Garip, Kieran Donaghy, and Patrick Sullivan. We are also grateful for helpful comments from Constanza Martínez, Jhorland Ayala, Javier Pérez, Jaime Bonet y Andrea Otero, and the participants at the 67th Annual North American Meetings of the Regional Science Association International (NARSC) and at the XVI internal seminar at the Central Bank of Colombia. All errors are authorial.
† Banco de la República, Cartagena, Colombia. Email: kacostor@banrep.gov.co
‡ The Chinese University of Hong Kong. Email: hengyu@uchicago.edu

El nexo entre desarrollo y migración interna en Colombia[§]

Karina Acosta[**] **Hengyu Gu** [††]

Resumen

Aunque un número considerable de estudios ha estado explorando el nexo entre la migración y el desarrollo en entornos internacionales, todavía hay un número reducido de estos estudios en contextos internos en los últimos años. Esta investigación tiene como objetivo principal estimar el efecto causal de las condiciones económicas de origen sobre la migración interna de la población colombiana utilizando una serie de tiempo de departamentos entre 2012 y 2019. Este análisis proporciona una perspectiva macro de asociaciones y causalidad entre la dinámica poblacional y el desarrollo en el país utilizando modelos de interacción espacial. Asimismo, analiza el retrato actual de los migrantes internos en Colombia (definida por flujos de cinco y un año). La evidencia en este caso específico muestra que la llamada U invertida de la migración depende de la escala utilizada. Por departamentos agregados, las condiciones económicas iniciales se asocian negativamente con la migración hasta un umbral en el que esta relación se invierte. Un comportamiento inverso se encuentra en la submuestra de migrantes rurales.

Palabras clave: migración interna, desarrollo, curva de migración, modelos de interacción espacial, Colombia.

Clasificación JEL: O15, O1, C21.

[§] Los autores están profundamente agradecidos por discusiones y revisiones iniciales por parte de Ravi Kanbur, Filiz Garip, Kieran Donaghy y Patrick Sullivan. También agradecemos comentarios de Constanza Martínez, Jhorland Ayala, Javier Pérez, Jaime Bonet, Andrea Otero y a los participantes de 67[th] Annual North American Meetings of the Regional Science Association International (NARSC) y el XVI seminario interno del Banco de la República. Todos los posibles errores son propios.
[**] Banco de la República, Cartagena, Colombia. Email: kacostor@banrep.gov.co
[††] The Chinese University of Hong Kong. Email: hengyu@uchicago.edu

1. Introduction

The world is experiencing a fascinating and challenging movement of its population. It is even faster in medium- and low-income countries, where a large number of people is moving not just towards high-income countries but within countries, mainly from rural to urban areas and from small and medium urban areas to large cities. According to UNDP (2009), in conservative terms, around 740 million people were moving within countries in the earlier 21st century; which accounts for around four times the movements across countries' borders. Nonetheless, international migration has experienced an important increasing trend, growing from 2.3% of the global population in 1970 to 3.5% in 2019 (McAuliffe & Khadria, 2019). This important shifting context should not be neglected in social sciences disciplines, because the spatial redistribution of the population can have important socioeconomic consequences at origins and destinations.

The increase of human and economic spatial mobilization and its further intensification of agglomeration have become the core of public debates, because of their tight associations with development. Population re-settlements can be both cause and effect of development. Moreover, one of the major concerns of the current developing world is the lack of inclusion of economic progress. In this scenario, the rapid change of population distribution has been considered a tool and a challenge to tackle persistent poverty and inequality and it will, therefore, be a feature of development (Christiaensen & Kanbur, 2018). Drawing on the theoretical migration models of Todaro-type and further recent enhances, it has been concluded that the relocation of the population from primarily rural/disperse areas to more densely populated areas can have a differentiated effect on the aggregated poverty, incomes, and/or inequality based on the population size of the destination (Christiaensen et al., 2017). This would be understood as an argumentative unidirectional effect from migration to development.

In the same vein, another body of literature has been exploiting the other direction: from development to migration. This is, how initial development conditions can generate incentives to migrate and whether such relocation decisions produce different outcomes for different groups (Clemens, 2014; Dao et al., 2018). Although multiple studies on the development of migration association have been done internationally, there is no agreement on the predominant theory that takes place, nor concur on the empirical evidence about the nature of such association (Clemens, 2014; Dao et al., 2018). A survey of the strand of literature on this issue shows that such theories are still to be tested with empirical data in recent migration trends, and most of the literature concerned with this association has been limited by looking at relationships rather than causality, especially when the units of analysis are spatial. Moreover, a survey on this issue also shows an increasing bias towards international migration studies. Earlier research and data on migration were primarily from internal sources, but there is an increasing concentration in international studies.

Low economic development and poverty have been recognized as the utmost triggers of population reallocation (Bencek and Schneiderheinze, 2020). Indeed, one of the most well-known migration models and evidence points towards an inverted U-curve relationship between migration and development (Haas, 2010; Clemens, 2014; Dao, 2018). These conclusions have had important consequences in the public policy discussions, given that it indicates that policies directed to assist poorer areas and their development boosts would generate larger movements towards the most well-off places, which is seen as a thread for hosting areas (Clemens and Postel, 2018).

Provided the importance of such a relationship, our primary goal is to estimate, or disentangle, the causal effect of origin economic development, using different metrics, on internal population migration (between states) in Colombia. In other words, this study aims to test the hypothesis of a so-called migration hump at a local level and under what conditions it holds. This analysis will provide a macro perspective of associations and causation between population dynamics and well-being in the current changes observed in Colombia from 2012 to 2019. This period was selected due to data availability restrictions. Furthermore, we aim to study recent internal migration patterns and changes within Colombia, from which there was no information while this study was conducted. Lastly, this study attempts to add to the literature on the methodological front, expanding gravity models proposed by LeSage & Pace (2008) and Chun (2008), by accounting for the actual causal effect of our variable of interest (development) which is not commonly used in the literature.

This document is organized as follows. The next section summarizes the literature on migration and development. It is followed by a description of the theories and models encountered in the literature to explore spatial interaction models and shows the methodology used. The fourth section presents the data. The fifth section describes the migration flows within Colombia and provides a descriptive exploration of the migration hump hypothesis. Sections six and seven present the results and robustness checks, correspondingly. Lastly, section eight concludes.

2. Migration theories and the migration-development nexus: an overview

Former models/theories of migration, whose underground concept is still used in recent theories, started in the nineteenth century by the geographer Ravenstein (1885). Ravenstein's work formulated the basis of the push-pull theories. Analogous to Newton's law of gravity, the decision of someone to move is understood as the consequence of the positive difference between two masses: the mass or attraction factors of the destination place (pull) subtracted from the mass or repulsion factors of the origin place (push). The set of push/pull factors integrate environmental, economic, and demographic characteristics. Some push factors commonly include congestion, population growth, political repression, lack of economic opportunities, or environmental changes. On the other hand, some pull factors are labor demand, access to better life quality, goods diversity, among others. Hence, a person decides to migrate if the aggregation of the forces at places of destination outweighs the aggregation of the forces that attach them to their places of origin.

Among other theories of internal migration using the Ravenstein (1885) concept are the gravity model and the importance of relative inequalities of opportunity by Stouffer (1960) and the bimodal model proposed by Lee (1966). Some of the critiques of the theory of push-pull factors are the lack of a clear explanatory system that disentangle the effect of the independent forces on migration, its difficulty explaining simultaneous movements of migration and return migrants, the assumption of population growth as a push factor (although it can be a pull factor) and the short-sighted selection of characteristics such as the incapability of some population to move, precisely because of their socioeconomic condition (Castles et al., 2014).

One of the most well-known theories of migration began in the economic theory. Particularly, the first migration theorists were concerned with the understanding of labor migration. As an alternative to unravel one of the components of the push-pull theory, emerged a macroeconomic theory of migration within the economic literature. Grounded on the seminal document by Lewis (1954), Harris & Todaro (1970) proposed an economic-based explanation of rural-urban migration, which is based on the hypothesis of the relative labor supply and demand across different geographical units.

The fundamental idea of the Harris and Todaro theory (and its further extensions) is that a differential excess of labor demand relative to supply in urban centers creates a positive wage gap between urban and rural areas (Harris & Todaro, 1970). Subsequently, population moves towards urban areas are generated until an optimal equilibrium is reached: the expected wages of urban locations are equal to earnings in rural areas. Within the Harris-Todaro model, the expected wages in urban areas are the probability of employment times the wages paid in these areas. Meanwhile, the reference in the rural counterpart is the marginal product of production. Even though the Harris-Todaro model was initially appealing and used to explain internal migration (at a time when typical competitive models were not able to explain), it has also been extended to international migration. Yet, this theory has several limitations, specifically its assumptions, that make it incompatible with several facts observed in the real world. The strongest assumption is that peoples' movements are fundamentally wage-driven.

Parallel to the neoclassical macroeconomic view of migration is the neoclassical microeconomic theory introduced by Todaro (1969), which was further extended by Borjas (1989) to international settings. In this theory, migration decisions are a consequence of individuals' income maximization decisions. In other words, it is a strategy based on an individual cost-benefit analysis. Within this framework, people are assumed to be rational individuals that migrate not just to places where they can earn higher wages (due to larger labor productivity), but also if "destination bonuses" allow them to mitigate the costs associated with migration: cost of traveling, new costs of living conditions, cost of adaptation to new location/labor market, psychological costs, among other. As opposed to the macroeconomic theory, individuals migrate when their characteristics increase their probability of employment in potential destinations. Also, migration is more likely to happen if individual characteristics, linking social conditions or technologies lower the cost of migration (e.g., reduction of transportation costs, household assets, etc). Further extensions have been attempted to improve the individualistic agent of Todaro (1969), such as Nakosteen & Zimmer (1980), which incorporates endogenous self-selection in the models of migration and income. Though Todaro's model and its extensions overcome some of the limitations of the macro views of Harris-Todaro, their fundamental weakness is that people are assumed to be rational individual actors (islands), without considering their affiliations to a group (e.g., families, cultural groups, and so forth).

As these challenges appear in the previous theories, a new theory related to the labor market emerged: the New Economics of Labor Migration. This theory was an effort to incorporate the social factors underneath migration decisions in the individualistic view of the Todaro model. In the particular case of the New Economics of Migration (NEM), the social affiliation that was taken into account was family. According to this theory, profit-maximizing decisions and functions are done at a household level (as opposed to individuals). Particularly, models within the NEM consider that households diversify their income risks or not-smooth consumption by allocating their labor force across different locations. This idea was primarily presented by Stark & Bloom (1985) who suggested that some of the main issues of the previous theories were not only limited by the individualistic view but also because they lack the incorporation of relativeness of well-being.

Although there have been three core streams of analysis of migration, as shown before, one from an individual point of view and others at a macro level, this study will be centered on the latter. As such, our units of analysis are geographical units, instead of individuals. This was our preferred approach because of data limitations at a micro-level. For instance, we can observe the current income of migrants, but there is no information on previous income at origins.

2.1. The development and migration nexus

The nexus between development and migration can be studied from three different lenses. This is, unraveling the bidirectional nature of these two notions: (1) the effect of migration on the development indicators of migrants themselves and hosting communities, (2) the impact of migration on sending communities (e.g., through remittances, brain drain, amongst other), and (3) the effect of development indicators in the propensity to migrate (development as a push factor). (1) and (2) can also be grouped as the unidirectional effect from migration to development and (3) could be separated as the unidirectional effect from development to migration. However, some authors, like Haas (2010), claim that these two concepts could be even conceived as non-separated. Instead, Haas (2010) considers the decision of migration an integral part of the concept of development. In his view, the decision to move is part of the concept of development within Sen's view, where the decision to migrate is understood as freedom or capability of an individual.

But what do we understand by (human) development? And specifically, what kind of development has been explored in the migration literature? Although the concept of development has evolved from a monetary perspective to a multidimensional perspective that involves a holistic view of well-being, most of the research in migration has been devoted to the monetary proxies of development. The hypothesis that will be shown in the following part primarily evokes the concept of development as the first theorists and policies suggested: an economic view. At the macro level, this is associated with macroeconomic indicators such as GDP, and monetary-base poverty. At the micro-level, this refers to individuals/families' total incomes and human capital assets (e.g., knowledge, skills, technological diffusion). It is worth highlighting that those recent studies of development are largely concerned with development in a holistic perspective as proposed by Seers (1979) and Sen (1999). However, both monetary/economic visions of development go hand-in-hand with multidimensional development. Furthermore, it was not possible to explore a multidimensional measurement of development due to data restrictions in our case study. The survey that collects migration information is different from the national survey that targets multidimensional poverty.

From the summary of the empirical studies of the development and migration relationships, we can derive three major conclusions. First, this literature has been recently focused on international settings and also has been centered on the effect of net receiving (migrants) societies and less on net sending societies. Perhaps, this international bias is due to the circumstances and the popularity of what Castles et al. (2014) have called the age of (international) migration and globalization, which have overemphasized international migration because hosting areas are mostly high income-capital countries and also because of the rise of unwanted immigration that has been taking place all over the world. Second, minimum literature is found in the role of development as a causal factor of migration at macro scales. Specifically, how socioeconomic conditions in origins can act as propellants and/or barriers to decide to move. A relatively larger proportion of the studies have concentrated on the study of migration as a driver of development (Bencek & Schneiderheinze, 2020). Third, there is not a homogeneous and widely accepted relationship between migration and development. Due to the focus of this research, we will draw attention to the one-directional analysis: from development to migration.

2.2. Development as a push/constraint factor to migrate

"The world distribution of opportunities is extremely unequal. This inequality is a key driver of human movement and thus implies that movement has a huge potential for improving human development" (UNDP, 2019, p. 8). This statement that captures those unequal opportunities across the space are drivers of migration and it is also the fundamental hypothesis of the proposed research. The importance of inequality of opportunities as a motorist of migration is not a recent topic. It was an idea introduced by Stouffer (1960). According to this author, the mass of people moving within a given distance is proportional to the number of opportunities. Stated differently, the higher the opportunities within areas or origins, the lower the propensity to move.

Although a large literature has been trying to understand that relatively low wages encourage movements from rural to urban areas, this does not consider that poor conditions itself can retain population, despite their willingness to move to more prosperous areas. Seeing from the historical perspective theory, underdeveloped conditions do not necessarily lead to migration. From a purely economic analysis, unfavorable socioeconomic conditions lead to migration if destinations are better-off than sending places. The main hypothesis found in the literature is that human movements from relatively unprivileged areas occur insofar as families/individuals can overcome the initial investments necessary to move. In other words, if they have the minimum socioeconomic conditions necessary to do so, especially in the case of internal migration. It has also been found that further distance movements are more expensive in monetary and social differences terms (Christiaensen et al., 2013, 2017). Therefore, poorer families are assumed to have a limited number of options to move, compared to relatively richer families/areas.

A summary of the effect of development on migration can be found in Clemens (2014). Based on the summary provided by this author, although the theory on migration suggests that increasing incomes in the least developed areas would reduce migration, recent evidence has shown the opposite. A phenomenon that was initially recognized as “mobility transition” was initially proposed by Zelinsky (1971). Particularly, at the macro empiric level, the studies surveyed, starting from research as early as Vanderkamp (1971) and Lucas (1975) to later studies such as Hatton & Williamson (2011) and Carlos (2002) find an inconsistent pattern in the relationship between origin incomes and migration. Some of them follow a standard inverted-U relationship, while others show a negative association. These contradictory findings are hypothetically explained by the length of the period of the study analyzed used by the authors as well as the type of data (cross-sectional versus time-series). Nonetheless, we suggest that it is important to explore such phenomena in the short term as they shape the current socio-economic distribution of the population, which in turn, are critical for the development of accurate public policies at local/regional scales.

3. Model

3.1. The core: Gravity Models

Gravity models have a long-standing trajectory in the literature concerned with population, goods, and services flows. Early developments date back to Ravenstein (1885) and Tinbergen (1962) who pioneered the empirical steps towards the incorporation of gravity models foundations in the understanding of migration patterns and trade flows, respectively. Traditional migration models are inspired by Newton’s law of gravity, which states that the attraction of two objects depends on their relative mass and their distance. Analogously, the largest movement of population from one place to another is controlled by forces or characteristics of each place, as well as their distance. As described by Anderson (2011), gravity models were until recently orphan tools unconnected to economic theory; although an effective device in empirical work. Primarily, because the concept of distance was removed from the equation of mainstream theories (Leamer & Levinsohn, 1995). While gravity models have been better integrated team economic theory, by incorporating distance components to it (Anderson, 2011), and some more progress of models specifications have been faced in the Regional Science and geography literature (Patuelli et al., 2016), there are yet some limitations to confront and further explore regarding these models.

Due to its compelling potential in explaining a plethora of socioeconomic phenomena, gravity models have become notorious instruments to understand movements of services and objects across space. Although it emerged around a century ago, it is still prevailing. A massive literature has used these models to explain commuting behaviors, retail trade, access to services, commodity flows, among others (Murdock et al., 1978). Nonetheless, its earliest developments are grounded on economics, geography, and demography, and are encountered in the work of Carey (1859), Zipf (1946), and Stewart (1941) (Murdock et al., 1978). Although Ravenstein (1876, 1885) did not develop an explicit formulation of these models, his pioneer work on internal migration provided insights to the careful analysis of migration patterns and their close dependence on distances between places. Ravenstein's legacy on the field of migration and, particularly, his "laws of migration" molded the field of spatial interaction models as understood nowadays (Greenwood, 2019).

Since the first sketches of gravity models, a myriad of further complexities has been added to these models. In the area of migration, such models have been supplemented with multiple socioeconomic patterns that might affect flows between regions/places. The basic and general form of a gravity model of inter-regional flows as described by LeSage & Thomas-Agnan (2015) and Sen & Smith (1995) is as follows:

$$\mu_{(i,j)} = CX_d(i)X_o(j)H(i,j) \qquad (1)$$

where $\mu_{(i,j)}$ represents the flow from region j = 1, ..., n to region i = 1, ..., n, thus, $\mu_{(i,j)}$ symbolizes an N (N-1) vector of interactions across space. $X_d(i)$ and $X_o(j)$ denote a set of variables characterizing the mass of destination and origin, respectively. C denotes a constant. H(*i,j*) characterizes the impediment/deterrent component between i and j flows and it is commonly represented by distance. In the gravity models literature, such deterrence can be represented by distance, neighborhood adjacency, or cultural links which might be influencing the interaction between a pair of areas (Egger & Pfaffermayr, 2016; Guo & Lu, 2015).

Despite the ability of early developments of econometric gravity models to explain flows well and their simplified expressions (using few parameters whereas consistent with equilibrium theory), there are multiple limitations to those early developments (Egger & Pfaffermayr, 2016). The main issue encountered in early flows models is that they disregarded the interconnectivity component of streams. Such models assume the OLS basic assumption of independence of disturbances. Moreover, early literature runs into the issue of specifications, such as the inclusion of the dependence of a bilateral flow to a third spatial unit (Beenstock & Felsenstein, 2016). These two absent characteristics in early literature lead to inefficient, biased, and inconsistent results; which are based on assumptions inclined to be breached in the real world (Beenstock & Felsenstein, 2016; Egger & Pfaffermayr, 2016). Another commonality that is not often confronted in the literature is the decomposition of independent causal effects which are complex in the context of gravity models. Most of the literature encountered is limited to the understating of overall patterns behaviors and does not consider causality in their specifications. Indeed, those decompositions in the context of gravity models are empirically challenging due to multicollinearity, measurement errors matters (Egger & Pfaffermayr, 2016), and endogeneity, which are produced by a conspicuous dual causality in migration models in the case of the reverse causation between migration and development defined.

3.2. Spatial interaction models

As opposed to standard spatial econometrics theories, spatial interaction data is concerned with links of pairs in geographic space instead of discrete locations as economic agents targeted for analysis (Patuelli et al., 2016). Although such gaps were early recognized by Griffith & Jones (1980), was not only until Bolduc et al. (1992) and Fischer & Griffith (2008) that they incorporated spatial lags disturbances to practical models, and LeSage & Pace (2008) who extended gravity models by incorporating spatial lagged dependent variables in the context of spatial interaction.

The latter was the first to underscore the issue of biases arising in the absence of acknowledging flows interaction.

We will draw attention to LeSage & Pace's (2008) proposal as the basis of this study. It is a general form of origin-destination flows model, which considers origin-to-destination flows dependence. These authors propose multiple specifications to integrate such dependency into gravity models. Particularly, LeSage & Pace (2008) propose a spatially lagged dependent variable and all the gravity residuals to be spatially correlated. Also, it only considers the cases where origin spatial units equal destination spatial units. It is described by the log-linear general form specification:

$$Ay = \alpha w \zeta_N + \beta_d X_d + \beta_o X_o + \gamma g + \epsilon \quad (2)$$

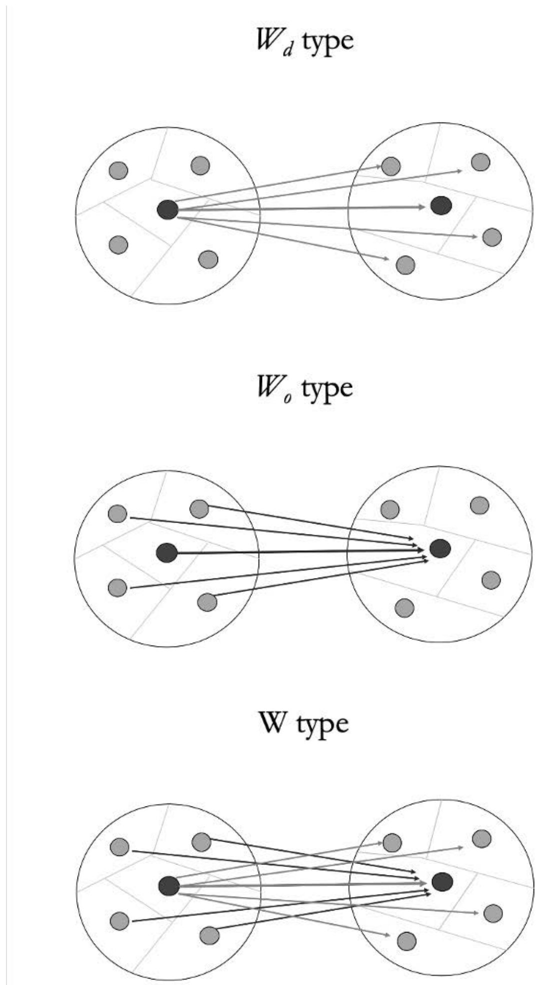
$$n^{-1} Y \iota_n = y$$
$$A = I_N - \varrho_d W_d - \varrho_o W_o + \varrho_w W_w$$
$$W_d = I_N \otimes W$$
$$W_o = W \otimes I_N$$
$$W_w = W_o \otimes W_d = W_d \otimes W_o = W \otimes W$$

In equation (2), y represents a vector N x 1, as described above, compiling all possible flows from an n-by-n flows matrix Y. Such vector can be stacked as an origin centric vector or as a destination centric vector. As an origin centric, y arranges flows from each origin first to all its corresponding possible destinations. As destination centric, y arranges flows from each destination first to each possible origin.[7] g represents the deterrent factor between pair of observations, typically spatial distance in practice. ϵ denotes the disturbance term. $symbolize$ symbolizes a spatial and row standardized matrix of size N by N which leads to destination-based and origin-based flows interaction, respectively. These matrices are produced by Kronecker products, corresponding to the respective interaction between the identity matrix I_N and the neighbors' matrix W reflects

[7] A detailed description can be found in Le Sage and Pace (2008).

destination or origins (as it may apply). Graphically, the possible interactions as described in Figure 1.

Figure 1. Spatial interaction neighbors' structures.



Source: a modified version of Chun (2008). Note: each polygon within circles represents a spatial unit and the small circles represent the centroid or unit reference (for example, major city) for each polygon.

In Figure 1, the representation of models of destination type W_d ($\varrho_o = \varrho_w = 0$) reflect the linkage between the flow from a j^* origin to an i^* destination and the flows from j^* to neighboring spatial units of destination i^*. We would define this sort of effect spillover, as it can be interpreted as the effect on flows to one location on the spread to other near destinations. Analogously, models of origin type W_o ($\varrho_d = \varrho_w = 0$) expose the interconnectivity between flows from origin j^* to destination i^* and neighbors of j^*. We would define this type of effect peer effect, as it can be interpreted as ones' decision to move to another location is assumed to affect movements decisions of contiguous locations. Lastly, type W (ϱ_d= ϱ_o= 0) shows the dependency between neighbors of destination and origin. Notice that the specification of spatial interaction model, as described in (2), encompasses a general form that can take different forms based on the restrictions imposed. For instance, a basic gravity model (assuming independence of flows) is retrieved setting ϱ_o, ϱ_d and ϱ_w to 0, or non-spatial dependence assumption. Moreover, these authors also proposed five additional models which can be obtained from different other specifications of ϱ_o, ϱ_d and ϱ_w, and although valid, they might be, in our opinion, overly restrictive in practical terms.

Despite manifold virtues of LeSage & Pace (2008) approach, there are some potential issues related to their approach associated with efficiency of computation to exploit the complexity of some spatial structures, the characteristics of the spatial units defined (symmetry of origin and destinations, N=n^2) or closed systems, and some basic assumptions of the general equilibrium approach considered by these authors (which restricts their estimations to cross-sectional data).

As an attempt to relax the assumption of temporal equilibrium and estimations limitations provided by LeSage & Pace (2008), some further extensions as space-time specifications have been considered. One example is found in Chun & Griffith (2011), who take advantage of Eigen spatial filtering techniques to understand migration patterns of network panel data with spatial and temporal random effects. The principle behind spatial filtering is to break down variables into spatial and non-spatial components with the assistance of eigenfunctions (Griffith, 2008; Tiefelsdorf & Griffith, 2007). Spatial filters enter into linear equations as an independent variable of a set of selected eigenvectors, which in the case of Chun & Griffith (2011) is derived from matrix (3).

$$\left(I_N - 11^{T}/_{n}\right)C\left(I_N - 11^{T}/_{n}\right) \qquad (3)$$

Where I_N and 1 are a standard identity matrix and a vector of 1s, respectively. C represents the neighbors' matrices, as described in Figure 1. Nonetheless, Chun & Griffith (2011) added yet another W type defined by a Kronecker sum instead of a Kronecker product: $W_o = W \oplus W = W \otimes I_N + I_N \otimes W$. Models using spatial filtering are of the form:

$$y_{it} = \beta x_i + \gamma E_i + b_i + \varepsilon_{it} \qquad (4)$$

Where i and t are the observations (dyads) and time indicators, respectively. E_i symbolizes the spatial filters that are random autocorrelation spatial dependence. The remaining elements in equation (4) denote the standard characteristics of models, as described before.

Although spatial filtering techniques can assist to efficiently estimate complex linear mixed models (Griffith, 2008; Tiefelsdorf & Griffith, 2007), where spatial network interaction and time interconnect, it works at the expense of regression coefficients interpretation. It is noticeable in Chun & Griffith (2011) that their analysis is centered on signs direction and model fit rather than coefficient interpretation. One of the advantages of Le Sage and Pace (2008, 2009) is its potential interpretability. An important task, particularly in the case of analysis of this study is the interpretation of coefficients, which requires the interpretation of elasticity of migration to initial development conditions. In the context of cross-sectional spatial interaction models, LeSage & Thomas-Agnan (2015) present a strategy to extract independent spillover and network effect using partial derivatives from the total effect observed.

3.3. Causal effects in spatial-temporal interaction models

Despite the advantages of the so-called spatial interaction models explained, there are some limitations. One of them, as aforementioned, is the assumption of a closed system, where the out-migration geographical units match in-migration geographical units. In multiple applications, it is not the case. For instance, international studies, usually ana out-migration from developing countries to host developed countries, leading to unbalanced origin-destination models. In the specific case of Colombia, information related to migration is not collected for states in the Amazon region. Therefore, this region has outflow internal information (pulled out from hosts reports) but not inflows, producing asymmetric flow matrices. Moreover, most of the spatial interaction models are concentrated on cross-sectional data. Even though spatial lags are understood as long-term equilibrium effects, they rule out temporal changes affecting the dependent variable. In the case explored in this study, it might be possible that an important external shock could have affected internal migration such as the international migration from Venezuela, for example. Lastly, most of the literature in multiple areas concerned with gravity models has studied overall patterns of explanatory forces explaining migration, but less room has been devoted to filtering independent causal effects that lead to the phenomena of migration.

The usage of spatial filter models can assist with the complex space-time interaction as in Chun & Griffith (2011). Nonetheless, this is not our preferred approach, due to lack of coefficient interpretability as discussed before, and because it cannot conclude about the changes across time or the variables of interest. In order to incorporate time and causal effects to the discussion, this study will extend the cross-sectional spatial interaction model to panel data. In addition, we try to disentangle the actual causal effect of poverty and income at origins on the propensity to migrate. Remarkably, a vast literature in social sciences has been motivated by the understanding of causal effects using common models' auxiliaries such as instrumental variables, disregarding spatial interactions which might lead to biased estimations (Betz et al., 2019). The opposite has been the case in the surveyed spatial interaction literature. Accounting for spatial interaction has nourished models to enhance the overall understanding of flows forces while dealing with spatial endogeneity, but it has not broadly considered non-spatial endogeneity such as the bidirectional causality inherent in socioeconomic phenomena, including the dual development-migration linkage.

3.4. Empirical model

Our model follows a pooled (constant effect) log-linear specification. The log-linear specification observes the standard gravity models, where logs are used instead of actual values of flows.[8] If level data (e.g., migration/trade flows) is used instead of logs, a heteroskedasticity problem occurs due to different mass sizes across space, usual on gravity models (Baltagi et al., 2015). Moreover,

[8] It was not necessary to correct for zero-inflated models given that the number of zeros in the data was minimum. Zero values were substituted for a small value that has no significant power to affect results, following a strategy used in estimations of logarithmic measures of inequality.

although there are multiple panel data specifications and potential issues, and advantages, to each one, we chose a simple panel structure due to two reasons. Firstly, when the cross-section size exceeds time series (N >>t), which is the common scenario and also the case for the Colombian data, the cross-sectional spatial component dominates time effects (Anselin et al., 2008). Second, we want to draw attention to the non-spatial endogeneity in the context of spatial interaction models.

As shown in Betz et al. (2019), provided a standard notation, an error term endogeneity problem can be a compound of a spatial and non-spatial component as follows:

$$y = \beta x + e \qquad (5)$$

$$e = \rho W y + u \qquad (6)$$

$$plim_{n\to\infty}\widehat{\beta_{ols}} = \beta + \frac{cov(x,u)}{var(x)} + \rho\frac{cov(x,Wy)}{var(x)} \qquad (7)$$

Equation (7) encompasses the two main sources of endogeneity that might arise from the error term. The former defines the non-spatial bias as a consequence of the potential reverse causality in our case between dependent (migration) and a relevant predictor variable (development). The latter bias describes the spatial bias arising from spatial interaction dependence. The approach suggested in the literature to control such simultaneity is using spatial two-stage least square (Anselin et al., 2008; Betz et al., 2019; Kelejian & Prucha, 2004). This approach can be used in this study as an extension of the cross-section method proposed by Le Sage & Pace (2008) and as a special case of Chun (2008).

We are going to follow the two types of models described. A cross sectional and a panel type. The former is of the form:

$$Ay = \alpha\zeta_N + \beta_d X_d + \beta_o X_o + \gamma g + \delta + \overline{z_\iota}\eta + \vartheta_{i1} \qquad (8)$$

$$\vartheta_{it1} = a_{i1} + \epsilon_{it1}$$

The spatial filtering panel is of the form:

$$y_{it} = \beta x_i + \gamma E_i + \overline{z_\iota}\eta + b_i + \epsilon \qquad (9)$$

There is one main difference between equations (8) and (9) and the original specifications, it is instrumentation, defined by $\overline{z_\iota}$. The main variables of interest in our case are the migration flows from state-to-state within one and five years (y). Moreover, among the critical explanatory variables, we are particularly interested in the understanding of the effect of the measurement of development, for which household mean income, GDP per capita, and poverty at origin are used (X'_{ot}). An additional set of information on characteristics observed are included as control variables, which are based on migration's main theories and the Colombian specific context. For instance, it is well-known that Colombian internal migration has been dependent on violent period throughout history. Among the controls, we incorporate population size, proportion of working age group, sex structure, stock of migrants (as a proxy of social network), homicide rates, and proportion of households with legal property titles both at origin and destination.

The set of instrumental variables selected are based on the evidence encountered in the literature of development economics and political institutions. The major hypothesis held by this stream is that historical institutions have shaped the current spatial distribution of economic performance (Acemoglu et al., 2002). Using multiple examples from different parts of the world, Acemoglu et al. (2012) state the thesis of the existence of two types of institutions: extractive and inclusive institutions, which triggered different outcomes leading to the current distribution of wealth. Fergusson (2017) and Meisel-Roca (2012) further explore such hypotheses for the particular case of Colombia. These authors identified a few institutional proxies explaining the current spatial inequality in Colombia. This study takes advantage of some potential variables from these studies including: minister participation of each state of Colombia between 1900 and 2000, Minister of Finance participation, and slaves rates from the Colombian Census 1843. Additionally, we use historical information of land Gini coefficient from 1997. Acemoglu et al. (2012) and Fergusson et al. (2017) have provided significant evidence of potential causality between current poverty/underdevelopment of multiple regions in Colombia and the presence of extractive institutions. It can serve as satisfactory instruments in attempts of causal effects extraction. This argumentation provides some evidence that the potential instruments used to meet their criteria: exclusion restriction and relevance, given that they are correlated with the current state of development and there are not enough arguments to claim that those instruments cause *y*, other than through development.

Given the non-linear relationship between development, particularly income per capita and GDP per capita, encountered in the literature and the data targeted, we additionally incorporated a second instrument. The instrument was suggested by Wooldridge (2000) and consists of obtaining a predicted value from the first stage of the standards 2SLS, of the development variable. This predicted value is squared and used as the second instrument in the estimations of the observed development indicators and its squared.[9]

4. Data

The main variables (migration, poverty, and income) come from Gran Encuesta Integrada de Hogares de Colombia -Integrated Household Survey- (IHS), a national survey representative at departamento (state) level in Colombia The IHS is collected on a monthly basis every year since 2006. Nonetheless, the module on migration has been available since 2012, hence this study is limited to the period from 2012 to 2019. During the data collection, there are thirty-two official political states in Colombia and one district, Bogotá D.C. The survey is representative of twenty-three out of the total states of Colombia, which further limited our analysis to the data available for those areas. The states excluded from the survey collection are those belonging to the Amazon-Orinoquía region (created with the constitution of 1991) and San Andrés: Guainía, Guaviare, Vaupés, Amazonas, Putumayo, Caquetá, Arauca, Casanare and San Andrés y Providencia. According to the population information of Colombia, these states had around 2.9% of the total population through the period of analysis. Therefore, we are confident about the representativeness of the results for Colombia. The final set of departments and their neighborhood structure, based on a queen approach, is plotted in Figure 2.[10] The techniques used for the collection of the information are cluster sampling technique, multistage and stratified.[11] The universe is all the civil population resident in Colombia, and it is representative at the national and state scale.

[9] The correlation between our main instrument and the instrumented variables is fairly stable through the short period explored. Its changes are negligible. Therefore, we assume that a historical instrument is suitable for its target in this context.
[10] The Queen criteria establishes that a polygon is considered a neighbor of another if they share either sides or vertices.
[11] More information about the details of the collection of the information can be found in DANE (2009).

Figure 2. Neighborhood structure

The IHS attempts to achieve the following goals: (1) Estimation of the main labor market indicators and their variation over time. These include general characteristics of employment, unemployment, and inactivity. (2) Classify the population according to the general guidelines established by the International Labour Organization (ILO). (3) Gather information about the socioeconomic characteristics of the population: sex, age, marital status, education, etc., (4) Collect general characteristics of the dwelling, utility access, and government and private benefits and beneficiaries. The IHS survey is comprised of eighteen modules. This study uses the modules related to general characteristics, dwelling conditions, labor market, and migration. The latter was not publicly available until recently.[12]

[12] The microdata can be found at:
http://microdatos.dane.gov.co/index.php/catalog/ MICRODATOS/about_collection/23/1

The questions used from the migration module are the following: where you were born, where did you live five years previous to the survey, and where did you live one year previous to the survey. All years have the same set of questions, except for 2012. The 2012 survey questions the number of years a person has lived in the state where is surveyed. We recovered five\- and one\-year migrants pool from the number of years reported. For the purposes of this study, we limited our analysis to the definition of migrant based on the last two questions. As will be shown in the next section, this study uses different migration definitions based on these reported questions, which are mostly disregarded in the migration literature.

In addition to the IHS, this research uses supplementary information to enrich the explanatory variables from multiple sources. Information on homicide rates was obtained from the Terridata portal from the National Planning Department. Terridata's original source of this information corresponds to the Ministry of National Defense and DANE. Information on population size, age, and sex rates come from the state-level reports of the National Bureau of Statistics. Furthermore, different instruments and a combination of them are used. All of the selected instruments are based on the literature that relates current socioeconomic conditions with historical political and economic institutions, as aforementioned (Acemoglu et al., 2012): historical Gini coefficient of land (1984, 1996/7 and 2002) from Instituto Geográfico Agustín Codazzi (2012), slaves rates 1843 (DANE, n.d.. Historical Census of Colombia), (4) Historical participation of states in the Senate of the Republic and Ministry of Finance of Colombia from (Meisel, 2014; Meisel-Roca, 2012).

4.1. Descriptive statistics

The IHS is collected every month in about 62,000 households/month and around 20,000 dwellings. The final sample size of individuals, based on the migration data, used for the estimations shown in this document is 5,911,922.[13] This study aggregates the sample at the State level and was restricted to the period 2012 to 2019, given that the collection of information on migration started in 2012. Moreover, the information was averaged by year/month. That is, the information of each year represents the mean estimated for each month of each year. The aggregated levels estimated

[13] The yearly/monthly distribution of such sample has been roughly stable. The sample by year is: 2012: 684,064, 2013: 587,792, 2014: 788,101, 2015: 787,044, 2016: 778,238, 2017: 7678,67, 2018: 762,753, 2019: 756,063.

from individual surveys consider the expansion factor provided, which attaches a value to each sample unit according to its weight in the universe (Colombian population) and acknowledges the sampling design. The calculation of sampling weights considers the strata the observed unit belongs to, and it is also adjusted by the non-answered surveys. More details about its estimation can be found in (DANE, 2009). A summary of the main variables and the number of observations are found in Table 1.

Table 1. Summary statistics of main variables.

Variable	Mean	SD	Min	Max
Flows 1 year	2,641.7	4,505.5	0	71,451.5
Flows 5 years	5,744.3	11,664.7	0	197,131.8
Income per capita (COP)	559869.9	203373.5	200479.3	1331511
GDP per capita (COP millions)	14.1	7.2	5.6	45.6
Monetary poverty (%)	35.83	13.28	10.1	68
Prop. pop without land title property	0.057	0.099	0.00	0.47
Population	1,961,746	1,817,345	459,515	8,281,030
Homicide rate (Cases per 100,000 population)	27.3	13.2	5.94	79.83
Prop. Male population	0.497	0.007	0.483	0.507
Prop. Population in working age (15-64)	0.643	0.028	0.564	0.695
Prop. cumulative stock migrants	0.430	0.129	0.158	0.739
Observations per year		529		
Total number of observations		4232		

Source: estimations based on GEIH. Note: the number of observations for each year represents the total number of interactions out-state to in-state of a balanced set of geographical units.

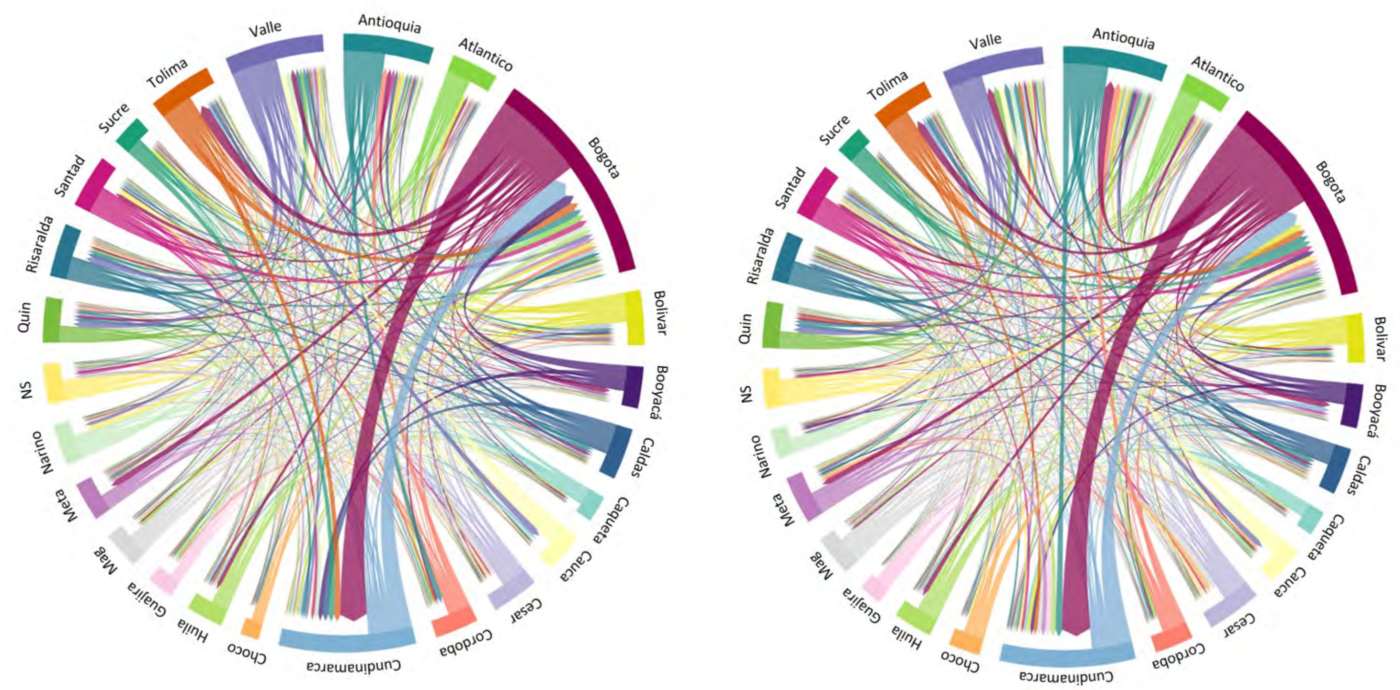
5. Migration patterns in Colombia

The concept of migrant can vary by the case of study, international or national. Irrespectively, two critical decisions ought to be made in migration studies: delimitation of the geographical units' span, and the length of period considered for a person to be classified as a migrant. In our case, the former is bounded by data restrictions, given that the IHS can be disaggregated, at most, at the state level. Regarding the decision of migrant classification, a person is typically classified as migrant based on three-time ranges: those who move to their current place within five years, twelve months, or lifetime migration, which classifies someone as migrant if their current place differs from the place where they were born (Molloy et al., 2011). Each classification has some measurement error due to potential multiple movements and/or short periods of movements within which we measure migration. Such measurement error increases as the time frame expands, hence it is more likely to occur in the born-based definition (Molloy et al., 2011). For the purpose of this study, we will explore the two types of internal migration in Colombia, a one- and five-years pool of migrants. Nonetheless, due to the relatively low flow within one year, our analysis is primarily focused on 5-years migrants, and 1-year migration estimates are used as a robustness exercise.

Although much attention has been given to migration in the national context of Colombia, it has been concentrated on the understanding of forced migrants (Ibáñez & Velez, 2008; Saldarriaga & Hua, 2019), due to historical internal conflicts. A recent new branch of local migration studies has redirected attention to the international migration due to the crisis in Venezuela (Fernández & Orozco, 2018; Tribín-Uribe et al., 2020; World Bank Group, 2018). Though these current and historical phenomena are relevant for the Colombian case, we argue that focusing on the relatively large flow of internal migration can provide additional tools to understand and explain the current geography of development in Colombia, a discussion that has been mostly disregarded.

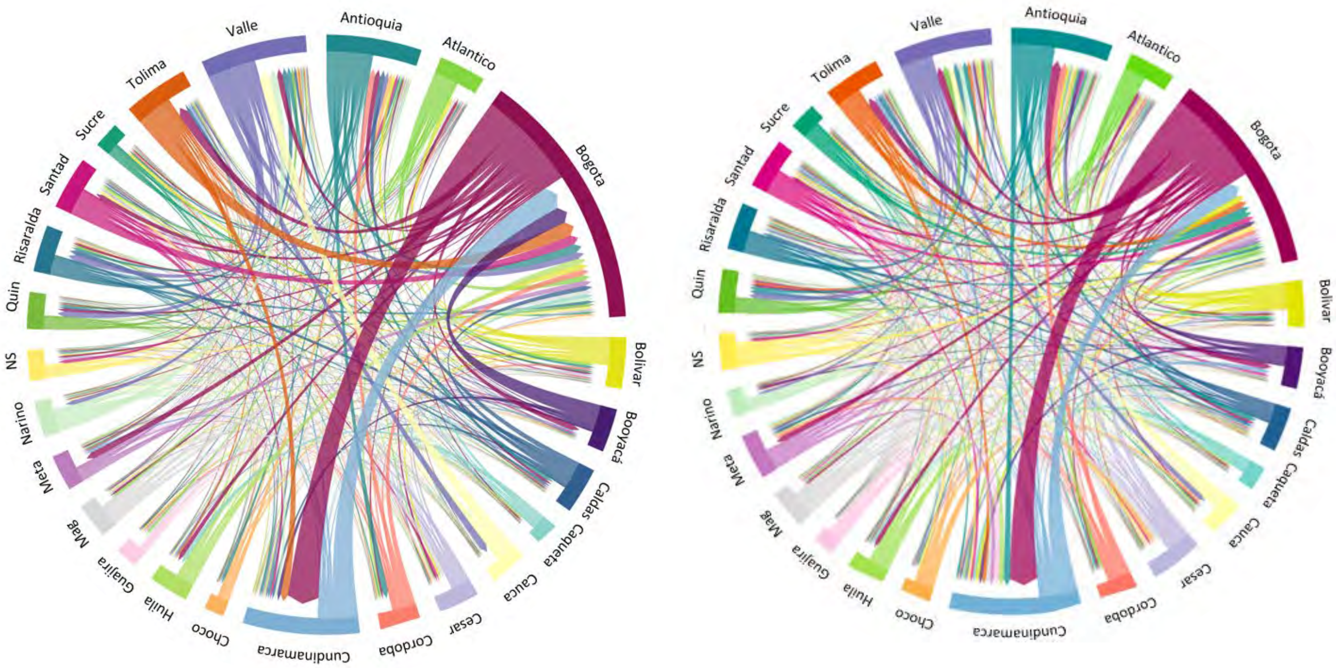
Among the few studies encountered in the literature which discuss overall internal migration in Colombia are Martine (1975), Galvis-Aponte (2002), Jaramillo (1999), Martinez & Rincon (1997), and Romero-Prieto (2011). Galvis-Aponte (2002) and Romero-Prieto (2011) are the closest to this study as both works use a standard gravity model to understand internal migration in Colombia. Both of them are concerned with multiple determinants of internal migration. Galvis (2002), for instance, shows that according to the census explored, net migration positively correlates with GDP per capita, particularly true within the period 1988-1993. That is, departments with higher GPD per capita tend to have higher net migration rates. This study claims that migration helps with the reduction of the income disparities between regions, leading to economic geographic convergence. According to Romero (2011) some regions have been benefitted from internal migration primarily through labor productivity, which has affected net-hosts incomes. Martine (1975) also provides some insights into the characteristics between migration and development. This author emphasizes that, in 1964, the largest mass of human flows come primarily from young and unmarried people, which essentially tend to have lower income levels, and that the most equipped group was absorbed by the most economically attractive area, Bogotá. A more recent study by Galvis-Arias (2019) at a micro level maintains that food insecurity, which can be considered as one component of development, is also a force of outmigration in Colombia.

Figure 3. One-year migration flows charts 2012 (left) and 2019 (right)



Note: own elaboration based on IHS 2012-2019.

Figure 4. Five-years migration flows charts 2012 (left) and 2019 (right)



Note: own elaboration based on IHS 2012-2019.

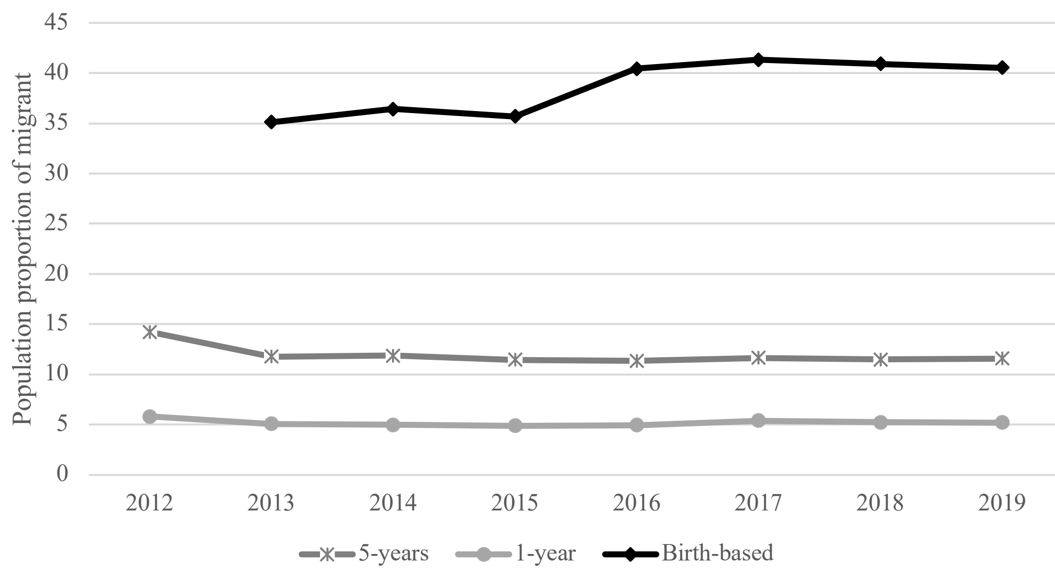
Opposed to the previous literature, this document emphasizes the importance of acknowledging migration networks to understand population flows in the context of spatial dependency. Moreover, there is a gap in the migration literature of Colombia as, to the best of our knowledge, there are no studies on recent overall migration patterns within the country, which are relevant for understanding current and future directions of public policies related to development and population dynamics and might affect the future distribution of development indicators. Furthermore, this research expands the dates of analysis to more recent years (from 2012 to 2019) and takes advantage of recent survey data, while previous studies have primarily used census data. At the time of the collection of the information for this research, the last available Census for Colombia was 2005, which followed 1993. Such period limits the time span of analysis. Lastly, as explained, this analysis attempts to unravel the causal effect of development on migration.

The migration flows for the years 2012 and 2019 are presented in Figures 3 and 4. The figures show, as expected, that the largest migrations flows are linked to Bogotá. Bogotá is the major hubs of migration attraction. The patterns of flows are stable through time. Nonetheless, a closer look at the five-years migration flows shows some important variations. Bogotá still receives a comparatively large number of migrants in 2019, but it has diminished when compared to 2012. In contrast, Cundinamarca, adjacent to Bogotá, expanded its relative inflow from other states, including Bogotá itself. This result is align with the findings of Aguilera-Díaz et al. (2020) who find that the population growth of Bogotá has decelerated while it has rapidly increased since 1985 around the municipalities close to Bogotá, which belong to Cundinamarca. There is also a noticeable expansion of migrants to Valle between the two periods. Likewise, among the least well-off departments, there was a fairly large increase of outmigration between the two periods in Chocó, La Guajira and Norte de Santander. Furthermore, as suggested by the interaction models, there seems to exist a dependency between flows, given that there is an apparent high interchange of flows between nearby states. For instance, the largest interchange of flows come from near-to Bogotá areas: Bogotá-Tolima-Cundinamarca-Boyacá. There is also an apparent interconnectivity between other nearby flows such as Valle-Quindío-Cauca. These descriptive characteristics provide elements that suggest the hypothesis aforementioned.

*Inter*state migration *rates* reveals a different story. Among the states with the largest push force are Chocó, Caquetá and Tolima (Chocó occupied the first place in 2019). The states with the lowest outmigration rates in 2019 are Antioquia, Atlántico, and Valle, where Antioquia has invariantly remained the least ejector, followed by Atlántico for most of the period (Appendix, Figure A1). During the time interval analyzed, the states that escalated more rapidly in the ranks of ejectors were Sucre, La Guajira and Meta. On the other hand, the states with the largest receiver rates were, invariantly, Quindío and Cundinamarca. In contrast, the ones with the least receivers rates have varied, but in recent years are Bolívar and Córdoba (Appendix, Figure A2). Interestingly, among the bottom-ten puller states are those belonging to the Caribbean coast, except for Cesar. This indicates that the Caribbean coast is not among the most desired destinations of migrants within Colombia.

The data also indicates that migration using different definitions in Colombia has been relatively stable through the period of analysis (Figure 5). The rates of migration using different classifications is similar to those observed in other countries such as the United States (Molloy et al., 2011). As expected, the largest rates are associated to the birth-based definition, followed by 5-years migrants and 1-year migrants. The 5-years and 1-year migrants rates have remained fairly stable in around 12 and 5, respectively. Bell et al. (2015) estimate that Colombia is middle-ranked among a selected group of countries in terms of Aggregate Crude Migration Intensity (ACMI) for 1-year migrants, a proxy of migration rate. According to this study, ACMI in Colombia is very close to Sudan, the Netherlands, Germany, and Austria. Hence, internal migration within Colombia is not negligible.

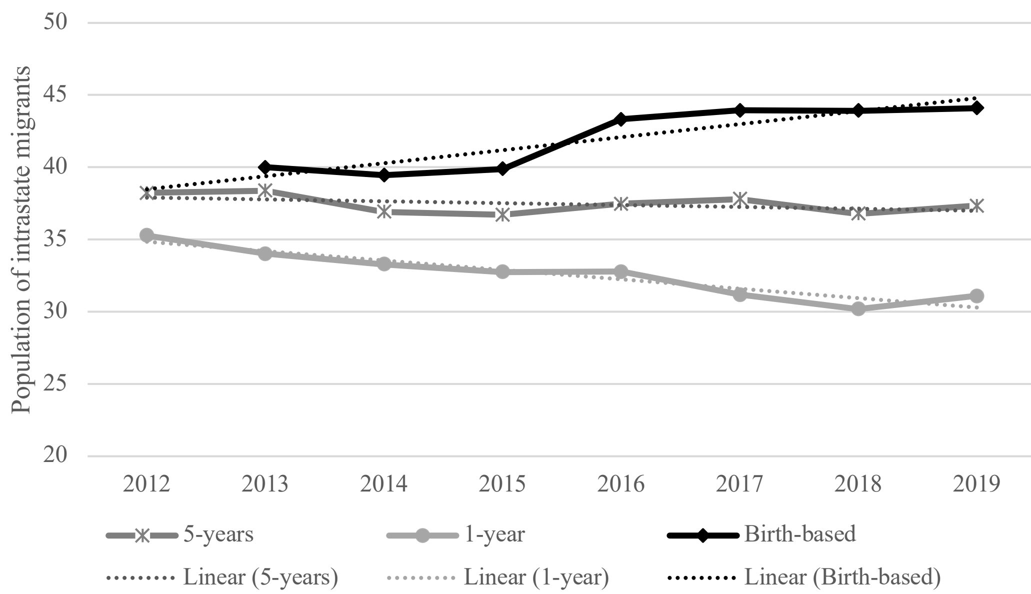
Figure 5. Trends of migrants rates by definitions



Notes: [1] birth-based migration corresponds to all people that live in a place other than their birthplace, [2] migration includes all migrants sampled (interstate and intrastate migrants), [3] source: IHS 2012-2019.

The data is also conclusive about the trends of participation of *intra*state versus *inter*state migrants. Figure 6 displays a downward trend of the relative participation of 5-years *intra*state migrants through the period of analysis. This trend is mimicked by 1-year flows, with a salient negative slope. Put differently, *inter*state migration is growing in importance in most recent years, notably in the one-year migration participations. This data is more likely to manifest the most recent trends as it does not have a time-frame smoothness effect resulting from larger periods to define a migrant, more apparent in other two definitions. The opposite is distinct in the birth-place migration.

Figure 6. Trends of intrastate participation of migrants



Notes: [1] birth-based migration corresponds to all people that live in a place other than their birthplace, [2] migration participation was calculated as the reason between intrastate migrants and the total number of internal migrants (interstate and intrastate migrants), [3] source: IHS 2012-2019.

Although additional migration determinants are not the main concerned of this study, it might be reflected that violence is the main source of migration in Colombia, because of the country´s history (Romero-Prieto & Meisel-Roca, 2019). The survey has included a question on the reason to migrate to those that migrated within the last 12 months. Until 2016, this question was limited to four possible answers: labor, studies, health, and others. Since 2016 it has expanded the options, among which it has been explicitly included threat or risk to lives or physical integrity of a person caused by armed conflict (or not). The first years show that the main answer provided was "others", closely followed by labor. In recent years, where there are ten options, including violence, the largest weight of reasons to migrate is "to accompany other members of the family" (around 40%), followed by labor, which has around 10 p.p. less. According to the last two surveys (2018 and 2019) for which this information is available, the proportion of the population that attributes migration to armed conflict or other type of violence thread was 9.83% and 6.35%, respectively.[14]

[14] This was estimated adding the two types of possible violence-related reasons to migrate.

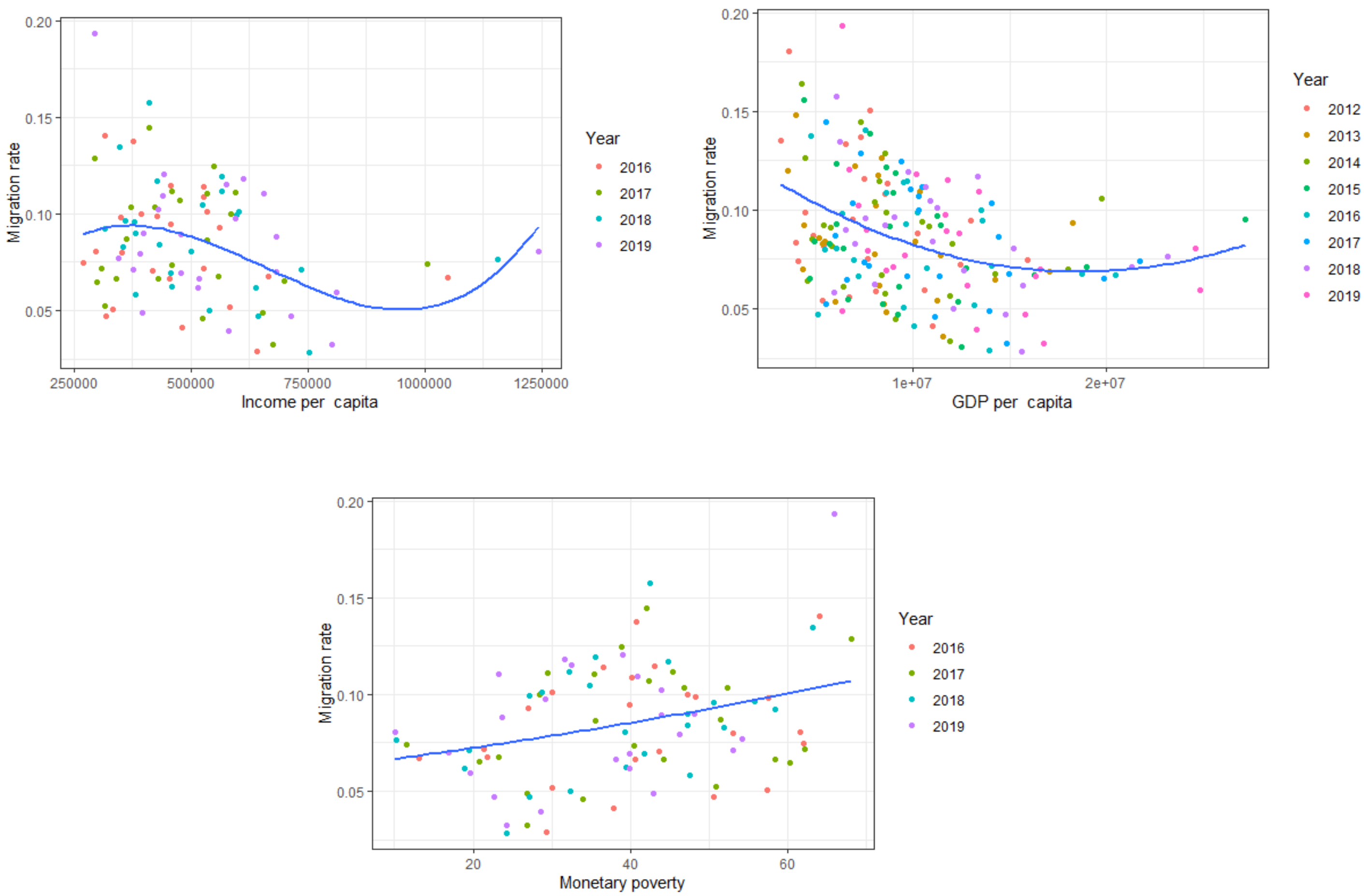
Therefore, we can conclude that, based on recent data, violence is not the main trigger of internal migration in Colombia. We want to warn that important changes in recent years such as the peace agreement in 2016 could have had impacted the reasons to migrate and, thus, we cannot posit unambiguous conclusions about time variations due to data limitations, as we cannot sub-classify the answer “others”.

5.1. Migration (inverse) hump evidence

In the international literature, it has been concluded the existence of a development migration hump. In other words, development stimulus at origin places are likely to produce out-migration up to a certain level of development (M. A. Clemens & Postel, 2018). Although most of the work on that topic has been based on correlations instead of causation, it is indirectly assumed that it is the case (Bencek & Schneiderheinze, 2020). To test such hypothesis, this section explores the association between development and migration in a local context, and the next section deals with causal interpretations.

To explore the development-migration relationship we proceed to plot such a relation for the years available and the two types of migration rates estimated. Figure 7 displays the association between migration rates (5 years) and three different proxies of development, namely, reported household income per capita, GDP per capita, and monetary poverty rates. The results for 1-year migration rates are similar and reported in the Appendix. Figure 7 shows some evidence of a U-curve, conflicting with the expected results based on the migration hump theory and the international evidence of an inverted U-curve shape. Poverty and migration rates graphs lead to the conclusion that poorer states tend to migrate more. The correlation between migration rates and income per capita, GDP per capita, and poverty rates were -0.27, -0.16, 0.15, respectively. Intriguingly, these correlations have varied through the time span targeted. We find that the correlation between outmigration rates and poverty has significantly increased between 2012 and 2019, with a positive trend. While at the beginning of the period this association surrounded 0.1, in recent years it soared to around 0.4. Concurrently, the correlation between income per capita and outmigration rate has significantly shrunk, from around 0.2 to 0.4 with a negative trend, indicating a stronger correlation.

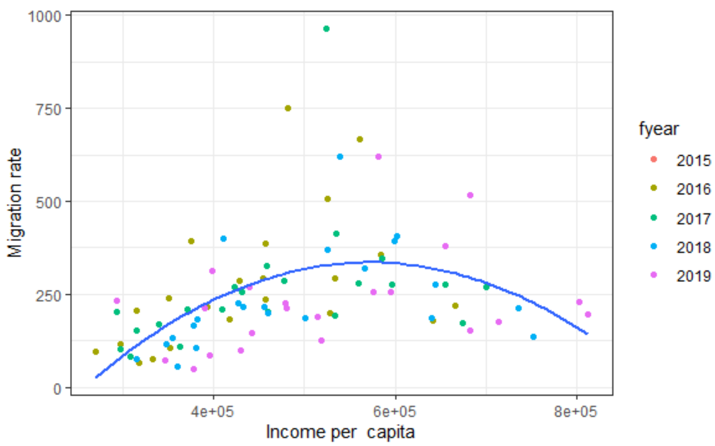
Figure 7. Development and out-migration association (5-years migration flows)

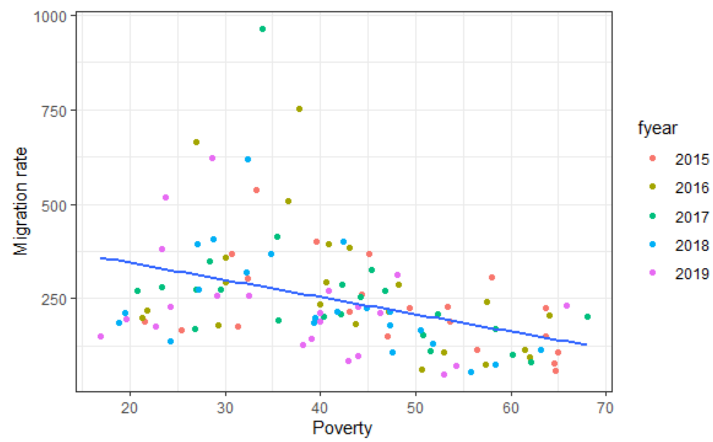


Source: own estimations based on IHS and population projections.

To verify if such a relationship holds at different scales, we also estimated the migration rates by states based on rural migrants. To do so, we identified the set of people that reported moves within the intervals considered, but also reported whether they lived in a rural area before. The denominator for these rural migrants' rates was the projected rural population by states, as published by DANE. Due to the size and potential representativeness issues of our sample, we limited this analysis to 5-years migrants. The results in Figure 8 interestingly illustrate a different association between development and migration when compared to the estimates resulting from the inclusion of all migrants. It shows some support of the Harris-Todaro model, which was primarily inspired by the rural-urban migration. Monetary poverty rates are negatively correlated to out-migration rates (-0.39) and the reported income per capita displays evidence of a migration hump, as encountered in the international literature (a correlation of 0.46).

Figure 8. Development and out-migration association





Source: own estimations based on IHS and population projections.

Deceptively, the subset of the sample of analysis leads to differing results about the curve shape of the relationship of interest. We hypothesize that such difference could be explained by the income and potential income gap (or living conditions) between sending areas and hosting areas, as well as the initial condition. For instance, the propensity to out-migrate from rural areas within countries can be more comparable to international settings, where there is evidence of large income gaps between sending countries and positive net immigrant countries. Hence, it is expected that our rural subset follows a pattern like common observations in international (cross-sectional) settings studies concerned with this nexus. Still, the gaps between urban areas, where most of the Colombian migration takes place, are less dissimilar than the gaps between rural to urban, the main destination of the out-rural migration in Colombia. Interestingly, this finding fits the schematic representation of demographic transition (proxy of development transition) and rural-urban migration postulated by Zelinsky, where the downturn relationship after an income is reached is explained by “counterubanization” (Skeldon, 2012).

Although this evidence is not enough to conclude that more development causes less migration in the case of Colombia, it shows some signs of this statement at the scale of aggregation used for the econometric estimations (complete states).

6. Models results

This section presents two sets of results. In the first place, it shows least squares estimates and temporal-spatial filter models. The second set of results shows cross-sectional outputs, as suggested by Le Sage and Pace (2008).

The results of standard OLS estimations and spatial filter models using different instruments are displayed in Tables 2, and 3. 122 eigenvectors for such models were produced using the interdependence W type, as suggested by Fischer and Griffith (2008). One of the advantages of spatial filtering models is their capacity to filter spatial endogeneity, which comes at a cost of coefficients interpretation. Due to this limitation, this first analysis is reduced to significance, signs interpretation, and relative effect. The candidates of eigenvectors were restricted to the minimum number of vectors necessary to control spatial effects (Tiefelsdorf & Griffith, 2007).

The analysis of results is centered on the 5-years migrants. There are some noticeable overall characteristics of the parameters estimated. In the first place, the OLS results displayed in Table 2 are particularly distinctive from the corrected spatial panel models. Thus, our discussion is focused on the latter. Results of spatial-IV models report fairly similar results, either using historical ministers' participation per region or Gini Land. Our preferred instrument is the minister's participation because there is already some empirical evidence of their power to explain the current spatial distribution of wealth in Colombia and because there is minimal evidence it could be associated with migration. Moreover, the Gini land coefficient might show some bias as it was corrected for the case of Bogotá, due to the absence of information for this city-state.

Our main attention is the potential effect of development on the propensity to migrate in Colombia. Table 3 shows that there is indeed negative causation between average income at origin on 5-years migration flows. In other words, after controlling for other push factors at origin and pull factors at the destination, there is evidence that relatively richer states tend to migrate less, following non-linear causation. Remarkably, the corrected models show a larger negative coefficient compared to the standard gravity OLS, regardless of the development indicator used. This indicates that conclusions extracted from the standard gravity models regarding particular variables can be highly biased, as suggested by the literature. In addition, such a relationship shows some signal of non-linearity between the income-related variables and migration flows. On the other hand, poverty indicates a positive effect on migration, which is aligned with the evidence encountered on the other two models (income-based development) for flows occurring within a five-year time frame. The empirical evidence corroborates, therefore, that the U-curve concluded from the descriptive statistics does take place within Colombia. Moreover, the results also suggest that such U-shape could be tighter, or more sensitive to income variations when conclusions are driven from corrected models versus classical gravity models.

Table 2. Pooled Least-Squares estimates. Dependent variable: 5 years migration flows

	Log Income pc		**Log GDP pc**		**Poverty**	
Explanatory variables	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
O-Development	-14.85	0.001	-16.37	0	-0.019	0
O-Development sqr	0.60	0.001	0.51	0		
O-Prop Males	0.24	0	-0.11	0.048	0.09	0.07
O-Prop working	-0.08	0	-0.05	0	-0.09	0
O-Population	0.75	0	0.85	0	0.83	0
O-Prop. no title	0.01	0.001	0.01	0.009	0.01	0.002
O-Homicide rate	0.27	0	0.07	0.206	0.23	0
D- Development	0.96	0	0.10	0.128	-0.02	0
D-Prop Males	0.02	0.114	-0.07	0.121	0.02	0.738
D-Prop working	-0.02	0.095	0.004	0.788	-0.038	0.016
D-prop Migrants	0.02	0	0.03	0	0.02	0
D-Population	0.70	0	0.73	0	0.71	0
D-Homicide rate	0.21	0	0.11	0.041	0.21	0
Distance	-1.30	0	-1.33	0	-1.32	0
Constant	59.9	0	137.0	0	-4.51	0.282
Observations	2,208		2,208		2,208	
R-squared	0.6205		0.6174		0.6161	
Year fixed effects						

Notes: [1] Variable names description= O: origin variables, D: destination variables, Development: Log of Income per capita or GDP per capita, Development sqr: Square of Development, Prop Males: proportion of males in the state, Prop working: proportion of working age population (15 to 64 years old), Population: Log of the population in State, Prop. no title: proportion of the population without housing title, Homicide rate: homicide rate per 100,000 population, Prop Migrants: proportion of the population that were not born in the state. [2] Covariables lagged 5 years

The direction of the correction yielded from both strategies, Le Sage and Pace (2008) and Chun (2008), are consistent. Both of them advise about a potential downward correction of the association of interest. Additionally, it is also noticeable from Table 4 results, an augmentation through the years of the sensitivity of migration flows to changes in income, evident when 2016 and 2019 results are compared. Lastly, the opportunity to estimate the elasticity of migration flows to income per capita changes, as described, shows that population flows are largely sensitive to origin changes on income (-1.27) and relatively less to destination changes (0.12), and a positive network influence (0.40).

Provided the descriptive and model-based evidence, we can conclude that some changes of the interstate migration pattern are occurring in Colombia in the last years. Also, we hypothesize that contrary to the international evidence, where high migration costs and other non-classical theories dominate the effect of development on migration, it seems that the classical theories are dominating in the case of Colombia (seen as a whole) to an extent. The case of Colombia does not perfectly fit neoclassical theories of migration, given that it would exhibit a linear downward trend, while our finding is non-linear. We observe that after an income threshold is crossed, the expected neoclassical outcome is reversed. The linearity only exists in the case of poverty. This is, when units of analysis are aggregated states, the relative outmigration occurs insofar precarious conditions at their places of origin overshadows those at potential destinations. This can be an important area of study of regional planning as migration between states, adding pressure to net-hosting areas.

Table 3. IV and spatial filter LMM panel estimates. Dependent variable: 5 years migration flows

	Log Income pc		**Log GDP pc**		**Poverty**	
Explanatory variables	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
O-Development	-30.683	0.001	-33.305	0.001	0.128	0.068
O-Development sqr	1.174	0	1.094	0.001		
O-Prop Males	0.062	0.684	0.150	0.581	-0.986	0.027
O-Prop working	-0.011	0.795	-0.187	0.073	0.297	0.117
O-Population	0.667	0	0.790	0	0.708	0
O-Prop. no title	0.006	0.069	0.015	0.05	-0.007	0.285
O-Homicide rate	0.238	0.018	0.394	0.212	-0.110	0.65
D- Development	1.072	0	0.026	0.817	-0.016	0.004
D-Prop Males	0.131	0.043	-0.251	0.074	-0.156	0.03
D-Prop working	-0.032	0.033	0.006	0.737	-0.024	0.236
D-prop Migrants	0.019	0	0.021	0	0.014	0
D-Population	0.727	0	0.678	0	0.671	0
D-Homicide rate	0.124	0.036	0.214	0.129	0.119	0.17
Distance	-1.285	0	-1.406	0	-1.530	0
Constant	173.282	0.01	260.432	0.001	31.829	0.005
Observations	2,208		2,208		2,208	
R-squared	0.6187		0.3106		0.4093	
Year fixed effects						

[1] IVs: Ministers participation and squared predicted income.
[2] Variable names description in Table 2.
[3] Covariables lagged 5 years.

Although it is not the main concentration of this study, it is worth noting that some covariables do not show the expected signs. For instance, while on the push factor side homicide rates has a positive sign, as expected, this does not have the expected sign on the pull side. Particularly, larger homicide rates are likely to produce more out-migration. In contrast, there is a contradictory positive effect of destination homicide rates on migration. This finding suggests that there might be multicollinearity issues in our models due to correlations between explanatory variables (Kutner et al., 2005). Nonetheless, our *vif* test values do not exceed 10 in any case, which indicate that there are not important multicollinearity issues.[15] Certainly, however, the existence of relatively

[15] This holds true for all variables except for the development squared and development for obvious reasons.

high correlations between covariates is a good example of why gravity models outputs ought not be interpreted without appropriate adjustments, specially of the main variables. In our case, this does not represent a thread due to the model correction. Moreover, the high correlation expected between the development variable and the development squared has proved to have no significant effect on models' outcomes. On the other hand, it is also perceptible a relevant variable among the pull factors, namely, proportion of migrants as a rough proxy of network effect. It is positively associated with out-migration to those destinations, which is compatible with sociological network migration theories.

Table 4. IV and Spatial Autoregressive Interaction models 2016 and 2019. Dependent variable: 5 years migration flows

	Year 2019			Year 2016		
Covars	**Coef.**	**Lower 0.05**	**Lower 0.95**	**Coef.**	**Lower 0.05**	**Lower 0.95**
ρ_d	0.04	-0.04	0.13	-0.04	-0.12	0.04
ρ_o	0.09	0.02	0.16	0.07	0.00	0.15
ρ_w	-0.01	-0.14	0.13	0.05	-0.09	0.18
O-PC Income	-70.64	-140.00	-0.49	-40.22	-73.83	-6.16
O-PC Income sqr	2.55	0.10	4.97	1.54	0.31	2.76
O-Prop Males	-0.30	-0.99	0.42	0.18	-0.24	0.60
O-Prop working	0.23	-0.22	0.67	-0.01	-0.13	0.11
O-Population	0.69	0.47	0.90	0.75	0.57	0.92
O-Prop. no title	-0.01	-0.04	0.02	0.01	0.00	0.02
O-Homicide rate	0.15	-1.11	1.39	0.00	-0.01	0.01
D-PC Income	0.55	-0.25	1.33	1.42	0.80	2.06
D-Prop Males	0.24	0.04	0.45	0.23	0.03	0.43
D-Prop working	0.02	-0.04	0.08	-0.04	-0.09	0.01
D-prop Migrants	0.03	0.02	0.04	0.01	0.00	0.02
D-Population	0.75	0.58	0.93	0.63	0.49	0.78
D-Homicide rate	0.005	0.00	0.01	0.009	0.00	0.01
Distance	-1.33	-1.48	-1.18	-1.35	-1.48	-1.21
Constant	460.49	-52.12	967.42	221.01	-32.01	468.76
Observations	2208					

[1] IVs: Ministers' participation and squared predicted income.
[2] Variable names description in Table 2.
[3] Covariables lagged 5 years.
[4] The SAR models corresponds to model 9 from Le Sage and Pace (2008).
[5] ρ_o, ρ_d, and ρ_w represent the unrestricted model parameters which identify origin, destination, and network effect dependence.

7. Robustness checks

As a robustness check we estimated the validity of the instruments used in this model. An instrument is defined as valid if it is relevant, exogenous, and only affects the dependent variable through the instrumented covariable. To test the relevance of the instrument we performed an F-test on the coefficients of the instruments of an OLS of the endogenous variable (income per capita) on the instruments and the exogenous variables of the model. The results are reported in Table 5. They indicate that both instruments are significant, but it is also important to check if the instruments are strongly correlated with the endogenous variable using the rule of thumb. It specifies that an instrument is not weak if the F statistic of the coefficients of the instruments is higher than 10. This exercise produces an F-statistic of 720.26 for the minister participation IV and 656.9 for the Gini land IV - df of (15, 2,192)-. Therefore, we can conclude that the instruments used are indeed relevant. Because we only use one instrument at a time, these models do not run into potential overidentification and there is evidently no need for additional weak instruments tests. Due to the similarities between the two instruments explored and for parsimonious purposes, the models were limited to the use of ministers participation. In addition, we estimate a placebo test by randomizing the instruments of interest. These results further confirm the effectiveness of our instruments.

Table 5. First stage regression of endogenous variable and instruments

		Minister participation		Gini land	
	Dep: Income per capita	Coef.	Std. Err.	Coef.	Std. Err.
O	Prop. Males	-13.38	0.63	-16.54	0.61
O	Prop. working age pop.	5.73	0.15	5.38	0.15
O	Log Population	-0.08	0.01	0.00	0.01
O	Log Homicide rate	-0.11	0.01	-0.09	0.01
O	Log PerCapita Income	-0.04	0.02	-0.04	0.02
D	Prop. Males	-0.58	0.72	-0.64	0.75
D	Prop. working age pop.	0.21	0.19	0.20	0.20
D	Prop. Migrants	-0.01	0.04	0.00	0.04
D	Log Population	0.00	0.01	0.00	0.01
D	Log Homicide rate	0.00	0.01	0.00	0.01
	Distance	-0.01	0.00	-0.01	0.00
	Intersect	18.20	0.70	19.29	0.72

As a second strategy robustness check, we performed the models for different definitions of migration, namely 1-year migrants. In this case, conclusions remain unaffected. A similar pattern is found in the 1-year spatial interaction models. However, in contrast to the 5-years, in the 1-year models poverty at origin does seem to encourage migration. Appendix table show that there is indeed negative causation between average income at origin and GDP per capita on migration flow of 1 year. Such contrast indicates dissimilarities in push factors at different periods, which is worth further exploration in future research.

In our view, the exploration of three different measures of development, in addition to the variation of the migration timeframe, indicates that our conclusions are robust and conclusive about the existence of a U-type causation between development (measured by income), and the population movements in Colombia, and linear and negative between poverty and migration. It also indicates that population within Colombia stay in their places insofar positive potential gains of income are such that greatly surpasses potential gains in the actual place of residence.

8. Conclusions

Does development push migration? Based on our results, the answer to this question is: it depends. Our raw exploratory analysis is consistent with the spatial interaction models' results that attempt to disentangle such causality. They illustrate that, in the case of Colombia, the migration hump observed in most of the international literature does not hold. In contrast, a U-curve is found between 2012 and 2019, using states as units of analysis. The models' results also suggest that there is indeed a causal non-linear effect. An additional observation of the exploratory analysis advises about the possibility of this curve taking the hump shape when the sample is restricted to out-migration from rural areas. This is not the concentration of this study, but it recommends that more research at different scales or subgroups are necessary to draw conclusions about the conditions required for the migration-development U-curve to hold. These findings also call for caution as they show trends that should not be oversimplified. Moreover, individuals' migration decisions cannot be derived from these findings, as this would lead to an ecological fallacy.

The second set of conclusions derived from this study are related to the overall behavior of the internal migration in Colombia. We find that the largest participation of internal migration within our targeted period happens between states rather than within states when a migrant is defined as someone that changes residence within the last 12 months or 5 years. The internal migration patterns are also similar to recent rates reported for the US. This research estimates that around 12% and 5% of population are classified as internal five-years and one-year migrants, respectively. These rates have remained virtually invariant between 2013 and 2019. Additionally, as expected, Bogotá is the major migration hub. Interestingly, the relative inflow to Bogotá has significantly diminished through the period of analysis, while the relative inflow to Cundinamarca has remarkedly increased. Although the largest flows of populations are identified between the pairs clustered by Bogotá-Cundinamarca-Tolima-Boyacá, the largest outmigration *rates* are primarily identified in Chocó, Caquetá, and Quindío, while Antioquia, Atlántico and Valle are among the largest retainers of their population, primarily Antioquia. Interestingly, most of the Caribbean coast states are grouped within the least attractive migration destinations, based on their immigration rates. Lastly, we find that although it could be thought that violence is the main reason to migrate within Colombia, it is not the case according to the data available for the last years.

Studies on the understanding of migration are far from been exhaustive. In fact, many scholars agree that due to the complexity of this phenomena, recent trends and the importance of several context-related characteristics, we cannot derive a unique theory that answers all the causes and consequences of migration in international and national settings. Moreover, even empirical studies do not coincide when they are applied to different settings or with different empirical techniques. In our view, due to the contemporary large size and speed of internal migration (virtually interchangeable with urbanization nowadays) in developing countries and its further development consequences, this topic ought to be in recent development research agendas. Countries like Colombia increased its urban population at a speedy rate in recent decades. These changes can be associated with overcrowding and negative impact on the well-being of its population and with positive economic growth that are essential for public policy allocations.

We propose that additional research is necessary within specific contexts to understand what the effect of a growing population agglomeration is and if net-receiving areas are "catching-up" with the additional costs associated to it on the well-being front. An additional gap that this research highlights and could be further explored is if the development effects are heterogeneous across different classifications of territories. Lastly, although this study shows that poorer areas have greater sending migrants' rates, it is necessary to understand at the micro level who is migrating and whether hosting communities are "absorbing" such migrants while providing better socioeconomic conditions. This is also a discussion that goes deeper into the needs and peculiarities of those areas. People migrate because the places of residence are not offering necessary opportunities, suggesting a lack of spatial equilibria across space.

Henderson (2010, p. 529) succinctly stated that: "most of this research is new and there is a long way to go to establish solid findings to properly inform policy debate about urbanization in developing countries." It would be added that there is still a long path to pursue in order to explain how a migratory process is shaping development and how development has shaped migration, specially within developing settings where data is limited. Such studies are even more challenging in Today's world due to the fast movement of population.

Bibliography

Acemoglu, D., García-Jimeno, C., & Robinson, J. A. (2012). Finding Eldorado: Slavery and long-run development in Colombia. *Journal of Comparative Economics*, *40*(4), 534–564. https://doi.org/10.1016/j.jce.2012.07.003

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics*, *117*(4), 1231–1294. https://doi.org/10.1162/003355302320935025

Aguilera-Díaz, M., Mora-Quiñones, R., & Ramírez-Rodríguez, C. (2020). Bogotá y área de influencia: Tendencias económicas en las primeras dos décadas del siglo XXI. *Documento de Trabajo sobre Economía Regional y Urbana; No. 291*. https://doi.org/10.32468/dtseru.291

Anderson, J. E. (2011). The Gravity Model. *Annual Review of Economics*, *3*(1), 133–160. https://doi.org/10.1146/annurev-economics-111809-125114

Anselin, L., Gallo, J. L., & Jayet, H. (2008). Spatial Panel Econometrics. In L. Matyas & P. Sevestre (Eds.), *The Econometrics of Panel Data* (3rd Edition, pp. 625–660). Springer Verlag.

Baltagi, B. H., Baltagi, B. H., Egger, P., & Pfaffermayr, M. (2015). Panel Data Gravity Models of International Trade. In B. H. Baltagi (Ed.), *The Oxford Handbook of Panel Data*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199940042.013.0020

Beenstock, M., & Felsenstein, D. (2016). Double Spatial Dependence in Gravity Models: Migration from the European Neighborhood to the European Union. In R. Patuelli & G. Arbia (Eds.), *Spatial Econometric Interaction Modelling* (pp. 225–251). Springer International Publishing. https://doi.org/10.1007/978-3-319-30196-9_11

Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M., & Kupiszewska, D. (2015). Internal Migration and Development: Comparing Migration Intensities Around the World. *Population and Development Review*, *41*(1), 33–58.

Bencek, D., & Schneiderheinze, C. (2020). *Higher economic growth in poor countries, lower migration flows to the OECD: Revisiting the migration hump with panel data* (No. 2145; Kiel Working Papers). Kiel Institute for the World Economy (IfW). https://ideas.repec.org/p/zbw/ifwkwp/2145.html

Betz, T., Cook, S. J., & Hollenbach, F. M. (2019). Spatial interdependence and instrumental variable models. *Political Science Research and Methods*, *8*(4), 646–661. https://doi.org/10.1017/psrm.2018.61

Bolduc, D., Laferrière, R., & Santarossa, G. (1992). Spatial autoregressive error components in travel flow models. *Regional Science and Urban Economics*, *22*(3), 371–385. https://doi.org/10.1016/0166-0462(92)90035-Y

Borjas, G. J. (1989). Economic Theory and International Migration. *International Migration Review*, *23*(3), 457–485. https://doi.org/10.1177/019791838902300304

Carlos, R. D. (2002). On the Determinants of International Migration in the Philippines: An Empirical Analysis1. *International Migration Review*, *36*(1), 81–102. https://doi.org/10.1111/j.1747-7379.2002.tb00072.x

Castles, S., Haas, H. de, & Miller, M. J. (2014). *The age of migration: International population movements in the modern world* (5. ed). Palgrave Macmillan.

Christiaensen, L., De Weerdt, J., & Kanbur, R. (2017). Cities, Towns, and Poverty: Migration Equilibrium and Income Distribution in a Todaro-Type Model with Multiple Destinations. *IZA Discussion Papers*, *10692*. https://doi.org/10.2139/ssrn.2961720

Christiaensen, L., De Weerdt, J., & Todo, Y. (2013). Urbanization and poverty reduction: The role of rural diversification and secondary towns. *Agricultural Economics*, *44*(4–5), 435–447. https://doi.org/10.1111/agec.12028

Christiaensen, L., & Kanbur, R. (2018). Secondary towns, jobs and poverty reduction. *World Development*, *108*, 219–220. https://doi.org/10.1016/j.worlddev.2018.04.010

Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, *10*(4), 317–344. https://doi.org/10.1007/s10109-008-0068-2

Chun, Y., & Griffith, D. A. (2011). Modeling Network Autocorrelation in Space–Time Migration Flow Data: An Eigenvector Spatial Filtering Approach. *Annals of the Association of American Geographers*, *101*(3), 523–536. https://doi.org/10.1080/00045608.2011.561070

Clemens, M. (2014). Does development reduce migration? In R. E. Lucas (Ed.), *International Handbook on Migration and Economic Development* (pp. 152–185). Edward Elgar Publishing.

Clemens, M. A., & Postel, H. M. (2018). Deterring Emigration with Foreign Aid: An Overview of Evidence from Low-Income Countries. *Population and Development Review*, *44*(4), 667–693.

DANE. (2009). *Metodología Encuesta Nacional de Calidad de Vida*. Departamento Administrativo Nacional de Estadística. https://www.dane.gov.co/files/investigaciones/fichas/ECV.pdf

Dao, T. H., Docquier, F., Parsons, C., & Peri, G. (2018). Migration and development: Dissecting the anatomy of the mobility transition. *Journal of Development Economics*, *132*(C), 88–101.

Egger, P., & Pfaffermayr, M. (2016). A generalized spatial error components model for gravity equations. *Empirical Economics*, *50*(1), 177–195. https://doi.org/10.1007/s00181-015-0980-5

Fergusson, L. (2017). *Who wants violence? The political economy of conflict and state building in Colombia* (No. 015890; Documentos CEDE). Universidad de los Andes - CEDE. https://ideas.repec.org/p/col/000089/015890.html

Fergusson, L., Molina, C., Robinson, J. A., & Vargas, J. F. (2017). The Long Shadow of the Past: Political Economy of Regional Inequality in Colombia. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2932228

Fernández, J. A., & Orozco, K. L. (2018). Migración venezolana en Colombia: Retos en Salud Pública. *Salud UIS*, *50*(1), 6–7.

Fischer, M. M., & Griffith, D. A. (2008). Modeling spatial autocorrelation in spatial interaction data: An application to patent citation data in the European Union. *Journal of Regional Science*, *48*(5), 969–989. https://doi.org/10.1111/j.1467-9787.2008.00572.x

Flórez, C. E. (2003). *Migration and the Urban Informal sector in Colombia*. Conference on African Migration in Comparative Perspective.

Galvis-Aponte, L. A. (2002). Determinantes de la migración interdepartamental en Colombia: 1988-1993. *Revista de Economia del Rosario*, *5*(1). https://doi.org/10.32468/dtseru.29

Galvis-Arias, N. (2019). Migración y seguridad alimentaria en Colombia. *Universidad de Los Andes, Escuela de Gobierno Alberto Lleras Camargo*.

Greenwood, M. J. (2019). The migration legacy of E. G. Ravenstein. *Migration Studies*, *7*(2), 269–278. https://doi.org/10.1093/migration/mny043

Griffith, D. A. (2008). Spatial-Filtering-Based Contributions to a Critique of Geographically Weighted Regression (GWR). *Environment and Planning A: Economy and Space*, *40*(11), 2751–2769. https://doi.org/10.1068/a38218

Griffith, D. A., & Jones, K. G. (1980). Explorations into the Relationship between Spatial Structure and Spatial Interaction. *Environment and Planning A: Economy and Space*, *12*(2), 187–201. https://doi.org/10.1068/a120187

Haas, H. D. (2010). Migration and Development: A Theoretical Perspective1. *International Migration Review*, *44*(1), 227–264. https://doi.org/10.1111/j.1747-7379.2009.00804.x

Harris, J. R., & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review*, *60*(1), 126–142.

Hatton, T., & Williamson, J. G. (2011). Are Third World Emigration Forces Abating? *World Development*, *39*(1), 20–32.

Henderson, J. V. (2010). Cities and Development. *Journal of Regional Science*, *50*(1), 515–540. https://doi.org/10.1111/j.1467-9787.2009.00636.x

Ibáñez, A., & Velez, C. (2008). Civil Conflict and Forced Migration: The Micro Determinants and Welfare Losses of Displacement in Colombia. *World Development*, *36*(4), 659–676.

Jaramillo, S. (1999). Migraciones e interacción regional en Colombia. *Territorios*, *1*, 95–117.

Kelejian, H. H., & Prucha, I. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, *118*(1–2), 27–50.

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied linear statistical models* (Fifth). McGraw-Hill.

Leamer, E., & Levinsohn, J. (1995). *International trade theory: The evidence* (pp. 1339–1394) [Handbook of International Economics]. Elsevier. https://econpapers.repec.org/bookchap/eeeintchp/3-26.htm

Lee, E. S. (1966). A Theory of Migration. *Demography*, *3*(1), 47–57. https://doi.org/10.2307/2060063

LeSage, J. P., & Pace, R. K. (2008). Spatial Econometric Modeling of Origin-Destination Flows. *Journal of Regional Science*, *48*(5), 941–967. https://doi.org/10.1111/j.1467-9787.2008.00573.x

LeSage, J. P., & Thomas-Agnan, C. (2015). Interpreting Spatial Econometric Origin-Destination Flow Models. *Journal of Regional Science*, *55*(2), 188–208. https://doi.org/10.1111/jors.12114

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, *22*(2), 139–191. https://doi.org/10.1111/j.1467-9957.1954.tb00021.x

Lucas, R. E. (1975). An Equilibrium Model of the Business Cycle. *Journal of Political Economy*, *83*(6), 1113–1144.

Martine, G. (1975). Volume, characteristics and consequences of internal migration in Colombia. *Demography*, *12*(2), 193–208. https://doi.org/10.2307/2060760

Martinez, C., & Rincon, M. (1997). Tendencias recientes de las migraciones in-ternas en Colombia. *Desarrollo Urbano En Cifras*, *2*.

McAuliffe, M., & Khadria, B. (2019). *World migration report 2020*. IOM.

Meisel, A. (2014). No Reversal of Fortune in the Long Run: Geography and Spatial Persistence of Prosperity in Colombia, 1500-2005. *Revista de Historia Económica; Montevideo*, *32*(3), 411–428. http://dx.doi.org.proxy.uchicago.edu/10.1017/S0212610914000147

Meisel-Roca, A. (2012). *¿Quién manda aquí?: Poder regional y participación de la Costa Caribe en los gabinetes ministeriales, 1900-2000*. Banco de la República. https://doi.org/10.32468/chee.31

Molloy, R., Smith, C. L., & Wozniak, A. (2011). Internal Migration in the United States. *Journal of Economic Perspectives*, *25*(3), 173–196. https://doi.org/10.1257/jep.25.3.173

Murdock, S. H., Wieland, J. S., & Leistritz, F. L. (1978). An Assessment of the Validity of the Gravity Model for Predicting Community Settlement Patterns in Rural Energy-Impacted Areas in the West. *Land Economics*, *54*(4), 461–471. https://doi.org/10.2307/3146171

Nakosteen, R. A., & Zimmer, M. (1980). Migration and Income: The Question of Self-Selection. *Southern Economic Journal*, *46*(3), 840–851. https://doi.org/10.2307/1057152

Patuelli, R., Linders, G.-J. M., Metulini, R., & Griffith, D. A. (2016). The Space of Gravity: Spatially Filtered Estimation of a Gravity Model for Bilateral Trade. In R. Patuelli & G. Arbia (Eds.), *Spatial Econometric Interaction Modelling* (pp. 145–169). Springer International Publishing. https://doi.org/10.1007/978-3-319-30196-9_8

Ravenstein, E. G. (1876). *Census of the British Isles, 1871: The birthplaces of the people and the laws of migration*. Trübner & Co. http://catalog.hathitrust.org/api/volumes/oclc/12175941.html

Ravenstein, E. G. (1885). The Laws of Migration. *Journal of the Statistical Society of London*, *48*(2), 167–235. https://doi.org/10.2307/2979181

Romero-Prieto, J. E. (2011). El éxito económico de los costeños en Bogotá: Migración interna y capital humano. In L. Bonilla (Ed.), *Dimensión regional de las desigualdades en Colombia* (pp. 255–283). Banco de la Republica de Colombia. https://ideas.repec.org/h/bdr/bdrcap/2011-12-255-283.html

Romero-Prieto, J. E., & Meisel-Roca, A. (2019). Análisis demográfico de la Violencia en Colombia. *Cuadernos de Historia Económica; No.50*. https://doi.org/10.32468/chee.50

Saldarriaga, J. F., & Hua, Y. (2019). A gravity model analysis of forced displacement in Colombia. *Cities*, *95*, 102407. https://doi.org/10.1016/j.cities.2019.102407

Seers, D. (1979). The Birth, Life and Death of Development Economics. *Development and Change*, *10*(4), 707–719. https://doi.org/10.1111/j.1467-7660.1979.tb00063.x

Sen, A. (1999). *Development as Freedom*. Alfred Knopf.

Sen, A., & Smith, T. E. (1995). Gravity Models: An Overview. In A. Sen & T. E. Smith (Eds.), *Gravity Models of Spatial Interaction Behavior* (pp. 49–152). Springer. https://doi.org/10.1007/978-3-642-79880-1_3

Skeldon, R. (2012). Migration Transitions Revisited: Their Continued Relevance for The Development of Migration Theory. *Population, Space and Place*, *18*(2), 154–166. https://doi.org/10.1002/psp.667

Stark, O., & Bloom, D. E. (1985). The New Economics of Labor Migration. *The American Economic Review*, *75*(2), 173–178.

Stouffer, S. A. (1960). Intervening Opportunities and Competing Migrants. *Journal of Regional Science*, *2*(1), 1–26. https://doi.org/10.1111/j.1467-9787.1960.tb00832.x

Tiefelsdorf, M., & Griffith, D. A. (2007). Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach. *Environment and Planning A: Economy and Space*, *39*(5), 1193–1221. https://doi.org/10.1068/a37378

Tinbergen, J. (1962). *Shaping the World Economy; Suggestions for an International Economic Policy*. Twentieth Century Fund. https://repub.eur.nl/pub/16826

Todaro, M. P. (1969). A Model of Labor Migration and Urban Unemployment in Less Developed Countries. *The American Economic Review*, *59*(1), 138–148.

Tribín-Uribe, A. M., Adhvaryu, A., Anzola-Bravo, C., Ávila-Montealegre, O., Bonilla-Mejía, L., Castro-Fernández, J. C., Flórez, L. A., Grajales-Olarte, Á., Guarín, A., Hamann, F., Hermida-Giraldo, D., Khanna, G., Lasso-Valderrama, F., Medina, C., Melo-Becerra, L. A., Mendez-Vizcaino, J. C., Morales, L. F., Nyshadam, A., Ospina-Tejeiro, J. J., … Velásquez, S. (2020). Migración desde Venezuela en Colombia: Caracterización del fenómeno y análisis de los efectos macroeconómicos. *Revista ESPE - Ensayos Sobre Política Económica*, 1–74.

UNDP (Ed.). (2009). *Overcoming barriers: Human mobility and development*. Palgrave Macmillan.

UNDP. (2019). *Human Development Report Cambodia 2019. Sutaining natural resources for all*. United Nations Development Programme.

Vanderkamp, J. (1971). Migration Flows, Their Determinants and the Effects of Return Migration. *Journal of Political Economy*, *79*(5), 1012–1031.

Wooldridge, J. M. (2000). *Introductory Econometrics: A Modern Approach*. South-Western College.

World Bank Group. (2018). *Migration from Venezuela to Colombia: Short- and Medium-Term Impact and Response Strategy*. World Bank. ttps://openknowledge.worldbank.org/handle/10986/30651 License: CC BY 3.0 IGO

Zelinsky, W. (1971). The Hypothesis of the Mobility Transition. *Geographical Review*, *61*(2), 219–249. https://doi.org/10.2307/213996

Appendix.

Figure 1A. Out-migration rates (five-years) ranks 2012-19

State	2012	2013	2014	2015	2016	2017	2018	2019	Δ 2019-12
Chocó	4	4	4	3	1	2	2	1	3
Caquetá	1	1	1	1	2	1	1	2	-1
Quindío	5	3	2	2	5	3	3	3	2
Tolima	3	5	5	4	3	7	6	4	-1
Meta	11	8	6	8	6	5	8	5	6
Cesar	8	6	11	7	7	8	4	6	2
Sucre	14	9	13	12	13	11	12	7	7
Caldas	2	2	3	6	4	6	5	8	-6
Magdalena	9	10	12	9	9	9	10	9	0
Huila	10	14	7	11	8	4	13	10	0
Risaralda	6	7	8	5	11	10	7	11	-5
Bogotá	17	17	16	15	18	13	16	12	5
Córdoba	15	13	10	13	12	20	14	13	2
Cauca	13	12	15	14	14	15	11	14	-1
La Guajira	22	22	21	22	21	17	9	15	7
Cundinamarca	12	15	14	16	15	16	20	16	-4
NSantander	23	21	20	20	19	14	19	17	6
Boyacá	7	11	9	10	10	12	15	18	-11
Bolívar	16	19	18	19	16	18	18	19	-3
Santander	19	18	17	17	17	19	17	20	-1
Nariño	18	16	19	18	22	21	21	21	-3
Valle del Cauca	20	20	22	21	20	22	23	22	-2
Atlántico	21	23	23	23	23	23	22	23	-2
Antioquia	24	24	24	24	24	24	24	24	0

Note: own estimates based on IHS.

Figure 2A. In-migration rates (five-years) ranks 2012-19

State	2012	2013	2014	2015	2016	2017	2018	2019	Δ 2019-12
Quindío	1	1	1	1	1	1	1	1	0
Cundinamarca	2	2	2	2	2	2	2	2	0
Risaralda	3	4	4	5	11	5	5	3	0
Caldas	10	14	10	9	6	3	3	4	6
Caquetá	12	12	5	3	4	6	4	5	7
Boyacá	18	13	13	11	7	11	8	6	12
Tolima	7	5	7	6	5	4	6	7	0
Meta	4	3	3	4	3	7	7	8	-4
Cesar	5	8	9	7	14	8	9	9	-4
Chocó	13	10	18	12	8	9	15	10	3
Huila	8	9	11	10	12	10	10	11	-3
Cauca	17	17	14	16	15	17	13	12	5
Bogotá	6	6	6	8	9	12	11	13	-7
Valle del Cauca	15	16	15	15	17	13	12	14	1
Santander	16	11	16	14	16	15	14	15	1
Sucre	11	15	17	18	13	14	17	16	-5
Nariño	23	21	23	23	20	21	23	17	6
Magdalena	20	23	20	19	21	16	18	18	2
Atlántico	22	19	22	20	18	18	19	19	3
Antioquia	24	22	21	22	23	22	21	20	4
La Guajira	9	7	8	13	10	19	16	21	-12
NSantander	14	18	12	21	19	20	20	22	-8
Córdoba	21	24	24	24	24	24	24	23	-2
Bolívar	19	20	19	17	22	23	22	24	-5

Note: own estimates based on IHS.

Figure 3A. Association between one-year migration rates and development

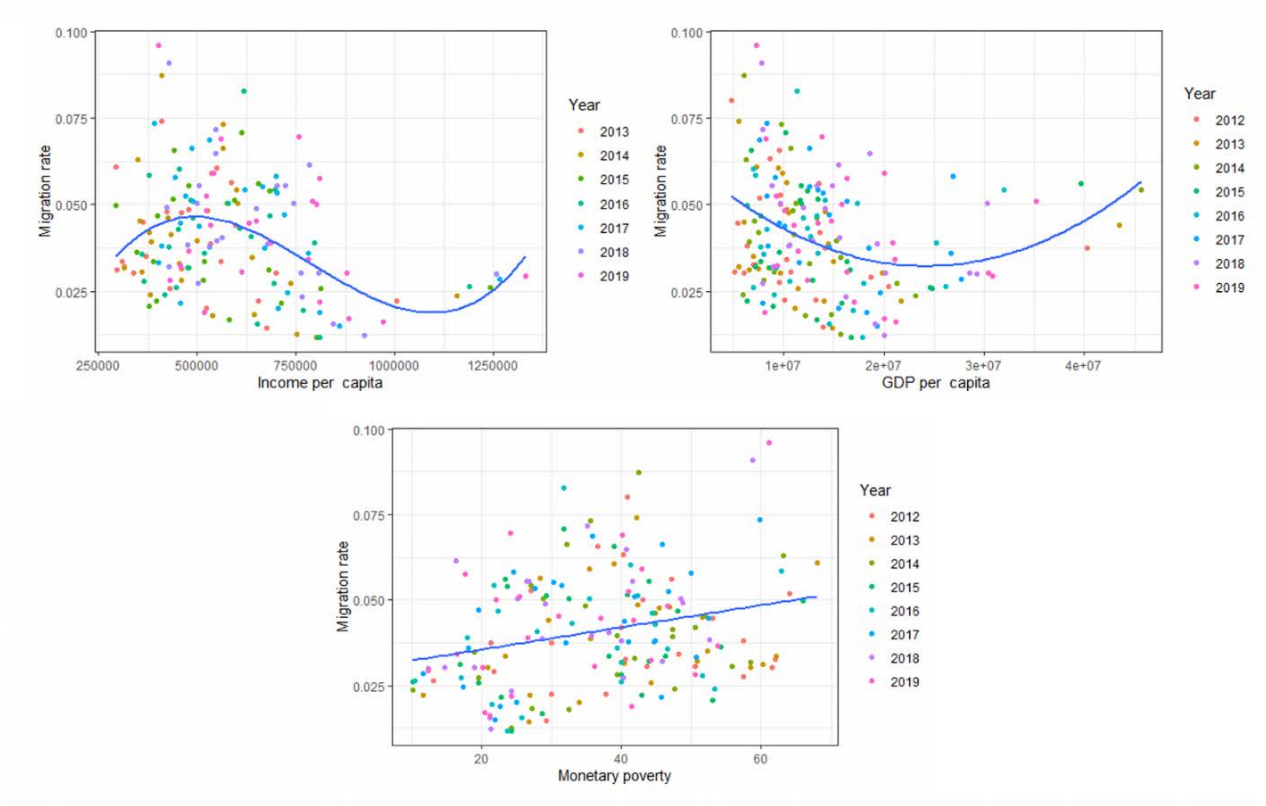


Figure 4A. Association between income and ministers participation

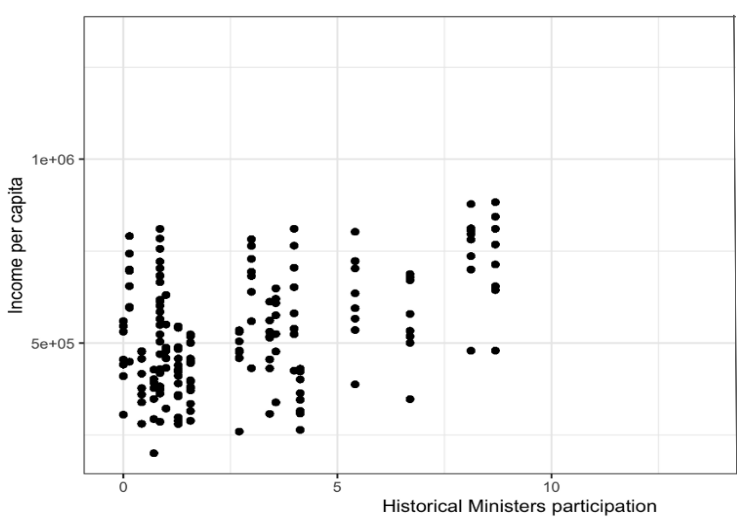


Table 1A. 2-SLS and spatial filter LMM panel estimates (IV: Gini Land). Dependent variable: 5-years migration flows

	Log Income pc		Log GDP pc		Poverty	
Explanatory variables	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
O-Development	-77.879	0.032	-53.761	0	0.012	0.62
O-Development sqr	2.814	0.027	1.604	0		
O-Prop Males	-0.706	0.187	-0.543	0	-0.254	0.104
O-Prop working	0.221	0.183	-0.004	0.841	-0.017	0.798
O-Population	0.554	0	0.980	0	0.752	0
O-Prop. no title	-0.002	0.801	0.002	0.474	0.002	0.533
O-Homicide rate	-0.226	0.493	-0.426	0.006	0.278	0.004
D- Development	1.217	0	0.119	0.102	-0.017	0
D-Prop Males	0.244	0.023	-0.042	0.438	-0.120	0.027
D-Prop working	-0.039	0.039	0.001	0.957	-0.030	0.06
D-prop Migrants	0.020	0	0.025	0	0.014	0
D-Population	0.759	0	0.759	0	0.695	0
D-Homicide rate	0.066	0.415	0.023	0.71	0.198	0.001
Distance	-1.285	0	-1.331	0	-1.464	0
Constant	528.901	0.05	468.231	0	15.381	0.005
Observations	2,208		2,208		2,208	
R-squared	0.446		0.534		0.6355	
Year fixed effects						

[1] covariables variables lagged 5 years.

Table 2A. Pooled Least-Squares estimates. Dependent variable: Log 1 year migration flows

	Log Income pc		Log GDP pc		Poverty	
Explanatory variables	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
O-Development	-1.79	1	-13.63	0	-0.02	0
O-Development sqr	0.12	0.504	0.42	0		
O-Prop Males	0.27	0	-0.02	0.717	0.11	0.022
O-Prop working	-0.08	0	-0.04	0.005	-0.08	0
O-Population	0.79	0	0.81	0	0.84	0
O-Prop. no title	0.01	0.008	0.01	0.013	0.01	0.007
O-Homicide rate	0.24	0	0.10	0.073	0.18	0.001
D- Development	1.09	0	0.12	0.099	-0.02	0
D-Prop Males	0.10	0.101	-0.09	0.041	0.00	0.933
D-Prop working	-0.045	0.002	-0.015	0.276	-0.060	0
D-prop Migrants	0.02	0	0.03	0	0.02	0
D-Population	0.80	0	0.85	0	0.84	0
D-Homicide rate	0.33	0	0.20	0	0.31	0
Distance	-1.33	0	-1.35	0	-1.34	0
Constant	-31.4	0	105.1	0	-7.1	0
Observations	4,416		4,416		4,416	
R-squared	0.4716		0.6174		0.6161	
Year fixed effects						

[1] covariables variables lagged 1 year.

Table 3A. IV and spatial filter LMM panel estimates. Dependent variable: 1 year migration flows

	Log Income pc		Log GDP pc		Poverty	
Explanatory variables	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
O-Development	-23.86	0	-16.72	0	0.01	0.854
O-Development sqr	0.87	0.001	0.48	0.005		
O-Prop Males	-0.17	0.271	-0.22	0.338	-0.10	0.756
O-Prop working	0.04	0.274	0.03	0.721	0.01	0.947
O-Population	0.71	0	0.84	0	0.78	0
O-Prop. no title	0.00	0.489	0.00	0.982	0.00	0.808
O-Homicide rate	0.05	0.671	-0.09	0.76	0.09	0.767
D- Development	1.15	0	0.18	0.025	-0.02	0
D-Prop Males	0.15	0.02	-0.04	0.626	0.04	0.559
D-Prop working	-0.050	0.001	-0.020	0.159	-0.066	0
D-prop Migrants	0.02	0	0.03	0	0.02	0
D-Population	0.83	0	0.88	0	0.87	0
D-Homicide rate	0.22	0	0.08	0.343	0.21	0.004
Distance	-1.32	0	-1.33	0	-1.31	0
Constant	140.7	0	142.6	0	-3.8	1
Observations	4,416		4,416		4,416	
R-squared	0.5306		0.4406		0.465	
Year fixed effects						

[1] independent variables lagged 1 year.

Table 4A. 2-SLS and spatial filter LMM panel estimates (IV: Gini Land). Dependent variable: one-year migration flows

	Log Income pc		**Log GDP pc**		**Poverty**	
Explanatory variables	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
O-Development	-41.415	0.006	-31.038	0	0.031	0.26
O-Development sqr	1.463	0.005	0.915	0		
O-Prop Males	-0.556	0.078	-0.285	0	-0.239	0.193
O-Prop working	0.143	0.082	0.033	0.101	0.074	0.363
O-Population	0.677	0	0.856	0	0.757	0
O-Prop. no title	0.000	0.937	0.000	0.969	-0.001	0.794
O-Homicide rate	-0.207	0.33	-0.171	0.077	-0.051	0.778
D- Development	1.170	0	0.165	0.03	-0.023	0
D-Prop Males	0.187	0.01	-0.047	0.35	0.053	0.353
D-Prop working	-0.050	0.002	-0.019	0.169	-0.065	0
D-prop Migrants	0.023	0	0.028	0	0.022	0
D-Population	0.840	0	0.881	0	0.876	0
D-Homicide rate	0.196	0.004	0.094	0.105	0.192	0.004
Distance	-1.307	0	-1.345	0	-1.294	0
Constant	280.669	0.017	264.858	0	-1.885	0.696
Observations	4,416		4,416		4,416	
R-squared	0.4053		0.4322		0.4488	
Year fixed effects						

[1] covariables variables lagged 1 year.