Febrero 2022

WP2-2022-002
N° de serie

DOCUMENTO DE TRABAJO

'When a Stranger Shall Sojourn with Thee': The Impact of the Venezuelan Exodus on Colombian Labor Markets

Autores

Julieth Santamaria

COLOMBIA CIENTÍFICA
Conocimiento Global para el Desarrollo

ALIANZA EFI
Economía Formal e Inclusiva

‘When a Stranger Shall Sojourn with Thee’: The Impact of the Venezuelan Exodus on Colombian Labor Markets*

Julieth Santamaria†

February 6, 2022

[Download the latest version]

Abstract

This paper analyzes the effect of open-door immigration policies on local labor markets. Using the sharp and unprecedented surge of Venezuelan refugees into Colombia, I study the impact on wages and employment in a context where work permits were granted at scale. To identify which labor markets immigrants are entering, I overcome limitations in official records and generate novel evidence of refugee settlement patterns by tracking the geographical distribution of Internet search terms that Venezuelans but not Colombians use. While official records suggest migrants are concentrated in a few cities, the Internet search index shows migrants are located across the country. Using this index, high-frequency labor market data, and a difference-in-differences design, I find precise null effects on employment and wages in the formal and informal sectors. A machine learning approach that compares counterfactual cities with locations most impacted by immigration yields similar results. All in all, the results suggest that open-door policies do not harm labor markets in the host community.

Keywords: Migration, Employment, Wages, Google searches
JEL: J61, J68, C81

*First and foremost, I would like to express my heartfelt gratitude to Joe Ritter, Paul Glewwe, and Fernando Lino, whose support and continuous encouragement made this paper possible. I am also grateful to Marc Bellemare, Alan Benson, Jason Kerwin, Daniel Valderrama, Jorge Pérez, and Jeff Bloem for their insightful feedback. I wish to thank participants of the Labor Network of LACEA 2019, DIAL 2019, AAEA 2020, the labor seminar at Universidad del Rosario, and the Development Seminar in the Department of Applied Economics at the University of Minnesota. I also appreciate the financial support from the University of Minnesota through its Doctoral Dissertation Fellowship and the technical support from Alianza EFI-Colombia Científica (grant with code 60185 and FP44842-220-2018).

†Inter-American Development Bank, Data and Digital Government Cluster. E-mail: juliethsa@iadb.org

1 Introduction

In light of large waves of displacement worldwide, there is considerable speculation about the effects of migration on host communities. Approximately one in three people around the world believe migrants take jobs that native workers want (Gallup, 2015)[1]. This issue, among other concerns regarding migrants, has been at the center of media attention and political debate for years. With 70.8 million forcibly displaced people worldwide (UNHCR, 2019) and nearly 201.3 million international migrants of working age (IOM, 2020), policymakers worry about the effects of open-door policies on local labor markets, particularly as the numbers keep rising. The impact of the presence of economic migrants has been widely studied, often yielding mixed results (Basso et al., 2019; Brücker and Jahn, 2011; Venturini and Villosio, 2006). However, economic migration is gradual and predictable, and is driven by local economic conditions (Peri and Yasenov, 2019). For this reason, studying sudden and unexpected inflows of displaced populations in contexts of open-door policies may provide a cleaner source of exogenous labor supply shocks.

This study examines an unexpected migration-induced supply shock to evaluate its effects on wages and employment in the host community. In particular, this paper analyzes the recent and unprecedented surge of migration from Venezuela and the open-door policy Colombia has put in place to draw its conclusions. Three political events in Venezuela are associated with the unleashing of the current migratory crisis that started in July 2016 and peaked in mid-2018. Figures reported by the International Organization for Migration (IOM) and the United Nations High Commissioner for Refugees (UNHCR) estimate that over four million Venezuelans have fled the country since the crisis started, 80% of whom have chosen Latin American countries as their main destination. The referred entities also estimate that Colombia has welcomed around 1.8 million immigrants, which represents about 3.6% of Colombia's population and 7.2% of the labor force[2].

One of the main challenges of conducting research using unexpected migration waves is to identify where migrants settle. Most refugees remain undocumented for long periods of time, making it difficult to track them through official records. For example, when the Venezuelan exodus started, the Colombian government put in place registration centers to provide migrants with temporary residence and work permits. However, migrant records are not a reliable source of data to identify where migrants settle. There are several reasons for this lack of reliability. First, the number of Venezuelans in the official registry is about half the amount estimated by IOM and UNHCR. This underestimation can be partially attributed to the voluntary nature of the registration process. Second, the number of records may overestimate the

[1]Analysis using 142 countries (2012-2014)

[2]Calculations by the author using data published by the Ministry of Labor

number of migrants located near the border or in the biggest cities; and, third, registering in one city does not necessarily imply that the migrant will enter the labor market in said location.

To identify where migrants settle, I use geographical variation in the Internet search intensity of keywords that Venezuelans are more likely to use compared to Colombians. Those keywords include 'Venezuelans in' given that new migrants are more likely to look for their fellow-countrymen communities; and 'PTP' or 'PEP', which stand for the residence and work permits granted by the government to allow Venezuelan migrants to enter Colombia. Compared to official records, this index suggests that migrants are not as congregated at the borders or in the main cities. For example, while official records indicate that about one third of the migrants have settled in Bogotá, the Internet search index suggests that merely one fifth have entered that market. A similar analysis using the US data also confirms that this type of text analysis of online interactions is promising at identifying where migrants settle, especially considering undocumented migrants. One of the reasons is that it is relatively easy and low-cost to conduct Internet searches given the easy access to free Wi-Fi zones or internet cafes. In addition, Google searches can identify the density of internet hits even in remote areas with a sufficiently identifiable number of searches.

Combining the time and geographical variation in the Internet search index, individual information on wages and employment from the Colombian Labor Market Survey, and a difference-in-differences design, the findings reveal negligible changes in wages in both formal and informal sectors due to migration-induced supply shocks. If anything, there are mild reductions in wages of natives working in occupational labor markets in which migrants are entering disproportionally more, such as elementary occupations, services, and clerical jobs. The results suggests that, in the worst-case scenario, a one-percentage point increase in the migrant labor supply will result in a decrease of up to 5 cents per dollar. Regressions also suggest a precisely estimated zero change in employment.

The results of the model are key to assess the average effect of migration at the country level. However, one might be concerned about the impact that migrants may have had in cities characterized by relatively stronger migratory flows. To assess the impact of migration at the city level, I construct artificial counterfactual (ArCo) cities using an elastic net model to train the pre-treatment data and predict the post treatment period to make inference [3]. Comparing each metropolitan area that had a relatively large influx of migrants with its respective ArCo city, the results are consistent with the country-level effects. That is, there are mild reductions in wages and null effects on employment. Altogether, the findings support the

[3] This design aggregates data at the regional level to construct a high-dimensional panel time-series dataset. Then, it uses an elastic net to construct the artificial counterfactual following the methodology by Carvalho et al. (2018).

idea that opening borders and allowing immigrants to enter a country freely do not damage the labor prospects of their native population.

This paper makes contributions to two aspects of the literature on migration. On the one hand, although many highlight the advantages of using episodes of forced displacement as a natural experiment to exploit exogenous migration flows (Borjas, 2003; Card, 1990; Del Carpio and Wagner, 2015; Friedberg, 2001; Bahar et al., 2020a), data on the geographical location of new refugees and displaced populations is scarce (Ruiz and Vargas-Silva, 2013). Thus, my first contribution is to introduce the use of text analysis of Internet searches to identify the location of immigrants. This method is especially promising in contexts where surges of immigration are unexpected and where tracking migrants is costly. Similar approaches are currently being developed, such as the use of social media to identify and locate migrants (Martin and Singh, 2018; Palotti et al., 2020). Beyond the purpose of developing better indexes to obtain more precise estimates, these strategies may serve policymakers to improve targeting of aid programs for immigrants and natives.

Second, the literature on migration has largely focused on the effects of south-to-north migration. This type of migration is frequently characterized by the arrival of imperfect substitutes to the local labor force, mainly due to differences in the native languages of immigrants and natives. Peri and Sparber (2009) found that the downward pressure on wages due to the arrival of migrants is partially alleviated by this imperfect substitutability. They find that immigrants specialize in occupations intensive in physical labor, while locals will reallocate to language intensive jobs. But what happens when both population groups share the same language and when substitutability between foreign and native workers is larger? With 85% of the displaced population worldwide being hosted in developing countries (UNHCR, 2019) and 79% settling in neighboring countries that may share the same language, this highly relevant question may be answered by analyzing south-to-south migration. Thus, this study provides evidence that, even when workers are more substitutable, there are still null effects on wages and employment caused by migrants.

The remainder of this paper is as follows: in Section 2, I provide a summary of the unfolding of the events that led to the Venezuelan exodus and briefly describe the characteristics of the migrant population. In Section 3, I describe the data and the identification strategy. In Section 4, I provide a summary of the results. In the last section, I provide some final insights and policy implications of this research.

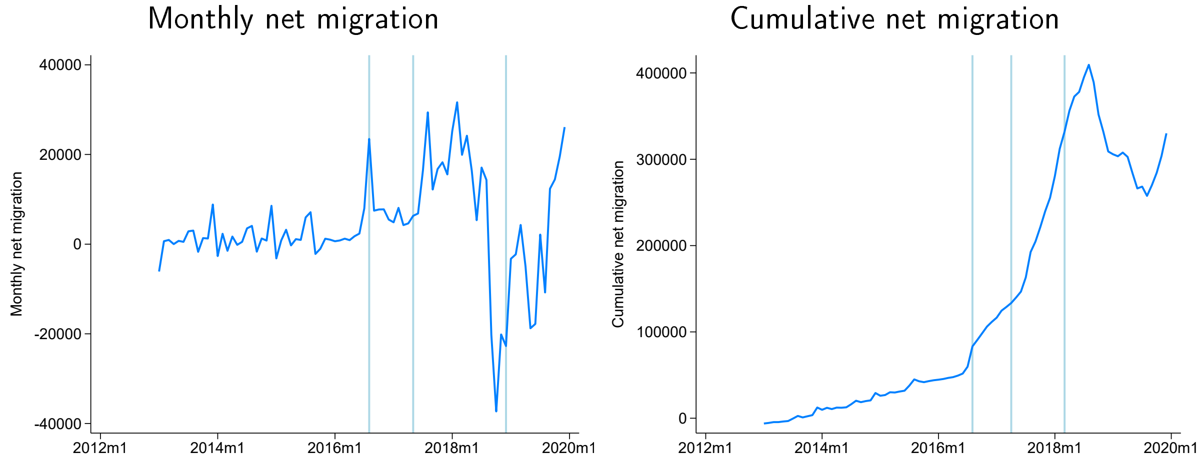
2 The Venezuelan Exodus

Chavez's presidential term ended with his death in 2013, leaving Venezuela without a strong private sector, with a weakened oil industry, with oil exports committed to repaying debts, with limited cash flow (Hernandez et al., 2016), with crises in its diplomatic relations with the neighboring countries (Romero, 2008), and with an opposition party that was increasingly gaining popularity (Lopez and Watts, 2013). In April 2013, Nicolás Maduro assumed power after being elected president by a narrow margin. Once in power, Maduro displayed further abuse of power, and the economy of Venezuela showed signs of a growing economic crisis.

In 2015, a diplomatic tension between Colombia and Venezuela grew due to an alleged presence of a Colombian armed group in Venezuelan territory. In retaliation, Maduro declared a state of exception and announced the closure of borders. Said closure, which was initially put in effect only in the State of Táchira, banned any transaction of goods and migratory flows. The border re-opened a year later, on July 6, 2016, in response to massive protests in Venezuelan border cities. Figure 1 associates the first peak in net migration, defined as inflows minus outflows, to this event.

During the year when the border remained closed, inflation, poverty rates, and food insecurity grew dramatically in Venezuela. In 2016, a Venezuelan household survey reported that 93.3% of households considered their earnings as insufficient to cover food expenditures, and 72.7% of individuals reported weight loss in the last year. Poverty rates were skyrocketing, with 81.8% of households in poverty and 51.5% in extreme poverty. Violence was also problematic for many households: 94% of households declared violence was on the rise. (ENCOVI, 2016). Thus, one can associate the economic situation's worsening, increases in violence, lack of access to services and hyperinflation to the start of a massive migratory wave in 2016. The Venezuelan exodus reached its peak in mid-2018. The controversial re-election of Maduro can be associated with this peak.

Figure 1: Net migration of Venezuelans in Colombia



Notes: Own calculations using data from Migración Colombia. The graph on the left shows monthly net migration, while the graph on the right shows cumulative net migration. As depicted in the graph, the first peak in net migration flows take place in July 2016 after the re-opening of borders. The light blue vertical lines trace the events described in Appendix C1.

This wave of migrants, which included Venezuelans as well as Colombian returnees, entered Colombia by land through three entry points: the Simón Bolivar International Bridge (Villa del Rosario/Cúcuta), the Páez Bridge (Arauca) and the Paraguachón International Bridge (Maicao), the former being the main entry point. Upon reaching the Colombian border, Venezuelans walked, hitchhiked, or took buses to their destination. Colombian authorities allowed Venezuelans to enter freely[4], even accepting expired passports as proof of identification (Ministerio de Relaciones Exteriores, 2019). Nevertheless, lack of appropriate documentation forced many to enter through irregular paths and become undocumented migrants.

The Colombian government implemented a number of policies to facilitate the integration of migrants. For example, they created a special residence permit, called PEP (Permiso Especial de Permanencia) for the exclusive use of Venezuelan nationals. Application to this permit was simple, with zero application fees involved, and with only minor requirements such as not being a convict and having entered through an official border-control site. This virtually allowed all legalized Venezuelans to work and access the healthcare system (World Bank, 2018). This permit was valid for 90 days and could be renewed every 90 days for up to 2 years. Recent initiatives from the bureau of migration include the expedition of new permits through an online renewal process, since most permits have expired or are about to expire.

Given the large numbers of undocumented migrants entering Colombia through irregular routes, at the peak of the crisis (between April and June 2018), the Colombian government

[4]Venezuelans are allowed to enter Colombia using different types of documents. For example, people living at the border can use a border mobilization card (or Tarjeta de Movilidad Fronteriza) that can be requested online with a fee of about 5 dollars. This card allows Venezuelans to enter the country and stay for up to 7 days. They can also enter as visa holders if they have the sponsorship of a company (Universidad del Rosario, 2020).

set up registration centers at the borders and in large metropolitan areas. The objective was to build an administrative dataset of immigrants (RAMV) that would eventually grant them access to a PEP. Every Venezuelan seeking to temporarily or permanently reside in Colombia was encouraged to register in person with any identification document certifying their country of origin.

Despite these efforts, only one-third of Venezuelans in Colombia are currently documented and hold a residence or work permit, while two-thirds remain undocumented (Migración Colombia, 2020a). In this context, there are two reasons why official records may lack the precision to determine where undocumented migrants settled. First, considering that registration offices put in place for undocumented immigrants were mainly located at the borders and in large cities, the official records may underestimate Venezuelans' presence in other cities of the country. Second, most immigrants requesting residence permits often end up in jobs in the formal sector. Given that most formal jobs are located in metropolitan areas, data coming from official registries might overestimate Venezuelans' congregation in these cities.

3 Data and Descriptive Statistics

3.1 Labor market data

Individual information on wages and employment as well as the characteristics of the native and immigrant population come from the Colombian Great Integrated Household Survey (or GEIH, for its acronym in Spanish) designed by DANE (Departamento Administrativo Nacional de Estadistica). This survey also provides detailed demographic information of Colombian households including age, gender, years of education, municipality or department of residence, type of job, hours worked, labor earnings, and firm size. The GEIH records information across 23 departments, 13 metropolitan areas and 11 intermediate cities [5]. This paper uses the sample of monthly information from January 2013 to December 2019. This repeated cross-sectional dataset reports data on around 52,000 people per month, which, for the time frame of the analysis, sets the number of observations at 3,674,040.

To identify the nonimmigrant population, which is the focus of this study, as well as the characteristics of Venezuelan migrants, this study uses the migration module of the GEIH. Since 2012, this additional module has been applied to all the respondents, recording information of the place of birth, changes in the place of residence in the last year and over the last five years, and the motives for the change of location. Given the relatively recent nature of migration in Colombia, the sample of migrants in this module was not representative of the population before late 2017. This module became publicly available in September 2019

[5] Colombia has 32 departments and 1023 municipalities

after a sizeable sample of immigrants was achieved (DANE, 2019)[6].

Table 1: Descriptive statistics of the nonimmigrant and immigrant sample after July 2016

Variables	(1) Nonimmigrant Mean	(2) Nonimmigrant S.D.	(3) Venezuelan immigrant Mean	(4) Venezuelan immigrant S.D.	(5) Colombian Returnee Mean	(6) Colombian Returnee S.D.
Female (%)	51.24	49.98	51.12	49.99	49.94	50.01
Age	36.78	14.17	29.29	10.31	38.84	13.25
Education (%)						
None	3.61	18.64	1.07	10.28	4.38	20.46
Primary	21.32	40.96	8.87	28.43	32.26	46.75
Secondary	48.13	49.97	63.16	48.24	54.36	49.82
Higher education	26.94	44.37	26.90	44.35	9.01	28.63
Labor market						
Labor Force (%)	80.26	39.81	92.10	26.98	80.32	39.77
Employment (%)	91.47	27.93	81.57	38.77	81.12	39.14
Informal (Firm size - %)	66.50	47.20	88.68	31.69	88.53	31.87
Informal (Social security - %)	59.75	49.04	95.77	20.13	88.10	32.39
Usual hours worked	44.67	16.12	49.61	17.63	45.56	17.47
Hourly earnings (COP)	6148.94	5649.75	4000.97	2791.51	3705.66	2579.73
Observations	2,592,639		16,008		6,320	
% of the sample	99.15%		0.61%		0.24%	

Notes: This table presents descriptive statistics for the sample of analysis. Data come from the Colombian Labor Market Survey - GEIH. The statistics are computing using individuals observed after July 2016, the month where the re-opening of borders took place. A description of the labor market variables is available in Appendix A. The table presents weighted means and standard deviations of selected variables for three groups: Nonimmigrants (columns 1-2), Venezuelan immigrants (Columns 3-4), and Colombian returnees (Columns 5-6).

This paper distinguishes between three groups of population. First, nonimmigrants, that is, individuals who were born in Colombia and who have not changed the municipality of residence in the last five years. This group is the center of analysis in this paper since it is the most susceptible to migration-induced supply shocks. Second, Venezuelan immigrants, who were born and whose place of residence over the last year was Venezuela. Third, Colombian returnees, who were born in Colombia and whose place of residence in the last year was Venezuela. Information on the latter two groups is used to understand the characteristics of the supply shock, while measuring the effects of migration in these groups is beyond the scope of this paper.

Table 1 displays descriptive statistics of nonimmigrants, Venezuelan immigrants and Colombian returnees following the migration crisis that started in July 2016. The table shows that sex ratios of immigrants are not different from the ones of the native population. However,

[6] At the time of publication, the Bureau of Statistics - DANE - warned that data on employment of migrants should be interpreted carefully and only using moving averages of the last year. For that reason, this study focuses exclusively on the effects of the native population and does not use this data to identify the location of immigrants. Also see https://www.facebook.com/DANEColombia/videos/3366521323562668

the average Venezuelan immigrant is 7.5 and 9.5 years younger than nonimmigrants and returnees, respectively. The latter is consistent with Becker's theory that younger individuals are more likely to migrate because their lifetime-expected benefits are larger, given a greater estimated duration of stay in the host county (Becker 1964).

About 26% of immigrants and nonimmigrants have completed a bachelor's degree, indicating that both groups have similar levels of education. This, along with the fact that Venezuelan immigrants have the same native language as Colombians, suggests that labor supplied by these two groups might be highly substitutable. On the other hand, Colombian returnees seem to be slightly less educated, with only 9% having accessed higher education.

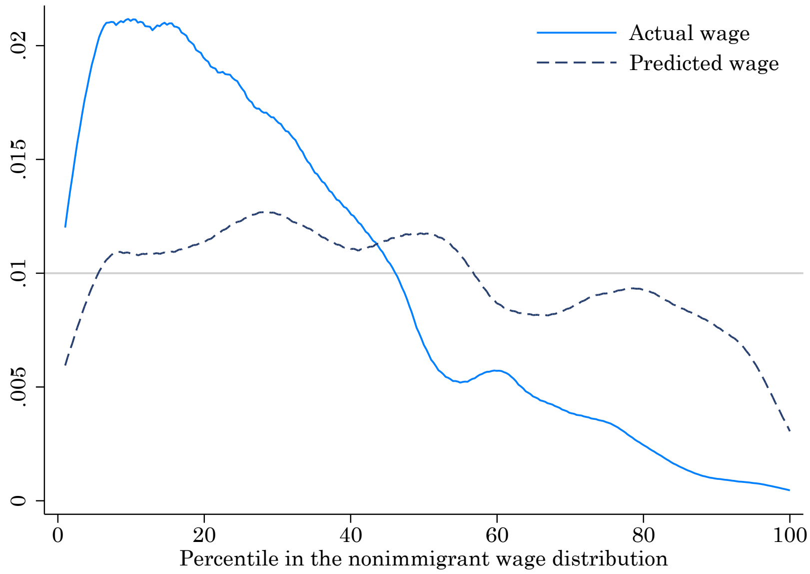
In terms of access to labor markets, relatively more immigrants are in the labor force. That is, among individuals aged 15-65, 92.1% of immigrants are either employed or searching for a job while the equivalent figure is 80.26% for the average Colombian. On the other hand, unemployment among immigrants and returnees is about 9 percentage points higher than unemployment among Colombians. This suggests that, although relatively more immigrants are willing to work, it is more difficult for them to find a job.

Venezuelan immigrants are mostly employed in informal jobs according to two definitions of informality: the Bureau of Statistics classifies a worker as informal if they work for a firm with fewer than 5 employees; the second definition, more broadly used in the literature of informality, classifies a worker as informal if they do not have a healthcare plan through their employer or do not have a retirement plan. Table 1 also suggests that immigrant workers work slightly longer hours for a lower rate per hour. This is indicative of the fact that Venezuelans not only have difficulty finding jobs, but when they do find employment, they access jobs of lower quality.

Overall, the descriptive statistics point toward a downgrade in the returns immigrants receive for their skills. 'Downgrading' occurs when immigrants with the same measured skills as natives, say with the same education and experience, receive significantly lower earnings. Dustmann et al. (2016) and Dustmann et al. (2012) use the cases of Germany and the US to illustrate that downgrading usually takes place in the years immediately after immigrants arrive and fades away in the long run after the assimilation. These authors point to an association between downgrading and the lack of formal requirements or complementary skills, such as fluency in English. In the presence of downgrading, assuming that immigrants compete only with natives in the same observed education-experience cell tends to produce more negative biased estimates.

To show the degree of downgrading of Venezuelan immigrants, Figure 2 depicts kernel estimates of the actual and predicted densities of immigrants in the non-immigrant wage distribution. The graph shows where immigrants are currently located and where we would assign them if they received the same return to their education and potential experience as natives. If there were no downgrading, the light blue line and the dashed dark blue line would overlap.

Figure 2: Downgrading of Venezuelan immigrants



Notes: This graphs shows evidence of the downgrading of Venezuelan immigrants in Colombia by presenting kernel estimates of the actual (light blue) and predicted (dashed dark blue) density of immigrants in the nonimmigrant wage distribution. The prediction can be interpreted as the wages immigrants would receive if their returns to education and experience were the same as that of the nationals. The horizontal line shows as a reference the nonimmigrant wage distribution. The kernel estimates above the horizontal line are where most immigrants are concentrated.

The results are striking. Venezuelan immigrants are predicted to be significantly more concentrated below the 50th percentile and underrepresented in the middle and upper ends of the nonimmigrant wage distribution. Given that Colombians and Venezuelans speak Spanish, we cannot necessarily associate this downgrading with the lack of language skills. However, anecdotal evidence suggests that Venezuelan immigrants face high costs and long waits to validate their past educational credentials (Universidad de Antioquia, 2020). This is perhaps the reason why they are entering more service-related and elementary occupations (see Figure C1).

Downgrading of Venezuelans suggests that they exert more pressure at the low end of the nonimmigrant wage distribution. Thus, using Borjas (2003) approach also known as the national skill-cell approach, which relies on the relative density of immigrants in education-experience cells, is not suitable for this context. Instead, using a pure spatial approach that exploits geographical variation of immigrants to evaluate its impact among low-income native workers would provide better estimates. Section 5 will further explain this empirical approach, which will allow me to evaluate the true impact of the Venezuelan migration on Colombian labor markets.

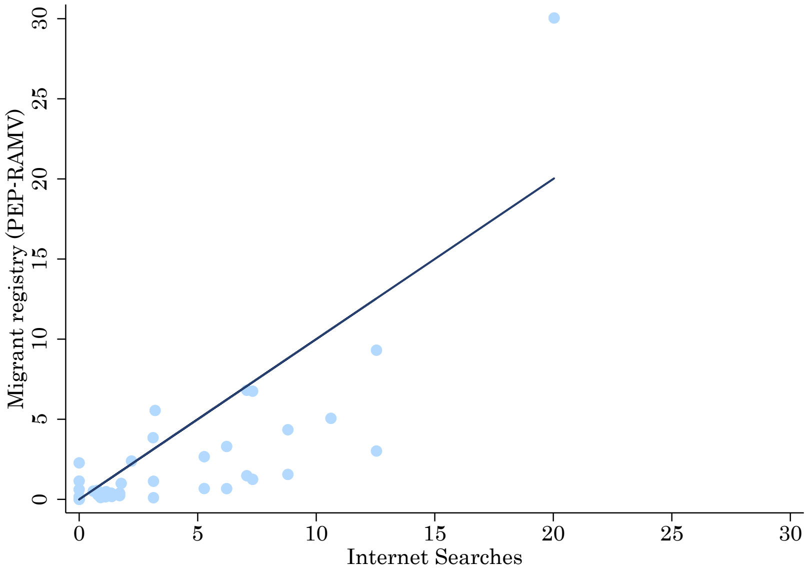
3.2 Geographical density of immigrants

New unexpected influxes of migrants are often thought of in the literature as exogenous shocks to the labor supply, however, tracking the location of new immigrants is challenging. For Colombia, which barely has any experience receiving large waves of immigrants, tracking Venezuelans was the case. As mentioned in Section 2, two initiatives were implemented to this end. On the one hand, the registry of residence permits (*Permiso Especial de Permanencia* - PEP), which collects individual information on those who voluntarily requested the permit. On the other hand, the Census of immigrants (RAMV), which was conducted through encouraging every Venezuelan seeking to temporarily or permanently reside in Colombia to register in person at any registration center. Compared to the former, registration to the RAMV targeted mostly undocumented immigrants who did not have a PEP.

By May 2020, these administrative registries have gathered data on about 784,234 Venezuelans out of 1,809,782 estimated to have settled in the country (Migración Colombia, 2020b). There are multiple reasons why these official records may lack the precision to determine where immigrants settled. First, official records may overestimate immigrants' density at the borders and in large metropolitan areas given that registration offices were mostly located in these regions. Second, given the registration's voluntary nature, only those who had a higher likelihood of getting a formal job might have registered at a higher rate. This may also point to an overestimation of immigrants' density in large cities, where most formal jobs are located. Third, having a record of the place of registration does not necessarily mean the immigrant will enter that geographical labor market. For these reasons, an alternative measure is used in this paper.

To identify where immigrants settle, this paper uses monthly geographical variation in the Internet search intensity of keywords that Venezuelans are more likely to use. Those keywords include 'Venezuelans in' given that new migrants are more likely to look for communities from their own country; or 'PEP', which is the residence and work permit created exclusively for Venezuelan immigrants. For each month after July 2016, when the first peak of positive net migration was recorded, this index assigns 100 to the location with most keyword searches. Every other area is assigned a number between 0 and 100 depending on the geographical density of hits for that keyword. Then, the index is transformed to reflect the share of searches that took place in region r. Thus, each month, the transformed index adds up to one and reflects the percentage of immigrants entering the labor market r at a point in time t. Using the national estimate of Venezuelan immigrants age 15-65 in Colombia, one can compute the number of immigrants entering each labor market. Finally, for each region at each point in time, I compute the size of the immigrant population as percentage of the local population in the labor force[7].

Figure 3: Comparison of official records and Internet searches



Notes: This graph compares the geographical distribution of immigrants obtained from two sources: official records (PEP-RAMV) and Internet searches. Official records are computed using publicly available data on PEP-RAMV. I aggregate individual records to obtain number of people in each region. Then, I compute the contribution of that region to the pool of immigrants. In the case of the Internet Search index, the geographical distribution of immigrants is computed using the procedure explained in Appendix B. If both sources of data provided a similar distribution of immigrants, all the light blue dots would be placed along the dark blue line. The kernel estimates above the horizontal line are where most immigrants are concentrated.

Figure 3 compares immigrants' geographical density drawn from the official registry with the one based on Internet searches. If the geographical distribution of immigrants according to both measures was the same, all dots would be located along the dark blue line. However, this is not the case. According to official records, Bogota hosts nearly one-third of the immigrants. In contrast, the Internet search index shows that about one-fifth are located in the capital city. Such a large difference has substantial implications in determining the share of immigrants that other regions in the country may host. While official records put a greater weight on large metropolitan areas and border cities (see C5), the internet search index indicates that immigrants are located in other areas too (see C4).

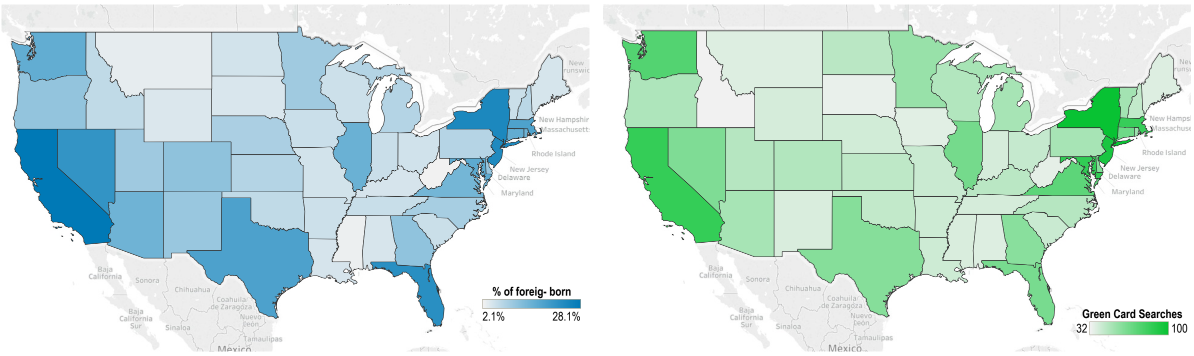
To determine whether the geographical intensity of Internet searches is a good measure to approximate the density of immigrants, we can apply the same algorithm in another context. The long history of immigration to the United States, along with their experience studying and collecting data on immigrants, makes this country the right subject for a robustness

[7] See Appendix 7 for more details regarding this transformation

check. Figure 4 compares the number of foreign-born individuals in US states as a share of the total population (in blue) and the Internet search intensity index for the pair of words 'Green Card' (in green). The geographical distribution of immigrants according to both, the official records in the US and the search index is very similar. The correlation between these two measures is 0.87. Thus, in a country with relatively more accurate data on the geographical density of immigrants, the Internet search measure seems to capture most of the variation.

Figure 4: Application of the Internet search index to the US case

Share of foreign born | Internet searches



Notes: This graph compares the regional distribution of two measures. The graph on the left (in blue), uses information from IPUMS international to compute the share of foreign-born population in each state. The graph on the right shows the distribution of the Internet search intensity for the pair of words 'Green Card'. The correlation of these two distributions is 0.87, which imply, a very good overlap between both distributions.

Thus, the Internet search index for keywords that immigrants are more likely use has applications beyond the case of Colombia. This measure can be particularly useful in contexts of forced displacement or unexpected migration waves where tracking immigrants becomes a challenging and costly task. The development of this type of measure might allow policymakers to get timely updates on the movements of displaced populations and allocate humanitarian aid accordingly.

4 Empirical Strategy

Dustmann et al. (2016) classify studies of migration into three categories. First, the national skill-cell approach, which exploits the relative density of immigrants in school-experience cells. Given that this approach often includes education fixed effects, it usually estimates the impact of immigration on more experienced local workers relative to workers with less experience. The main problem with the approach is the assumption that immigrants enter markets appropriate to their school-experience classification. In other words, it assumes no downgrading of immigrants' skills set. The second design is a pure spatial approach that exploits geographical variation on the density of immigrants. In the presence of downgrading, this approach estimates the total effect of immigration for local workers in certain education-experience group. For this reason, this model is more appropriate to study new waves of migration. Finally, the third approach is a mixture approach that exploits the relative density of immigrants in school-experience-location cells. This approach shares the disadvantages of skill cell-approach.

This paper studies the sudden and unprecedented nature of the Venezuelan exodus to evaluate its effect on the labor markets of the host community using a pure-spatial approach. Using individual data on wages and employment, a difference-in-differences design exploits the geographical variation in exposure to migration-induced supply shocks and the timing of events [8]. The analysis focuses on the host community that has secondary completed or below. The evidence of downgrading of immigrant credentials shown in Figure 2 along with the low share of the host and immigrant community accessing higher education, support the idea that this segment of the population is more likely to be affected. The model is as follows:

$$Y_{i,rt} = \beta M_{Ven,rt} + \eta_1 M_{Ret,rt} + \eta_1 M_{Int,rt} + \phi' X_{\{i\}} + \alpha_r + \tau_t + \kappa t + u_{i,rt} \quad (1)$$

where $Y_{i,rt}$ is the outcome of analysis for person i, observed only once in location r at time t; T_0 is a dummy that takes the value of 1 after July 2016, i.e., after the reopening of borders[9]; $m_{Ven,rt}$ is the size of the migration shock in r at time t; $m_{Ret,rt}$ controls for the migration shock caused by Colombian returnees; $m_{Int,rt}$ controls for migration shocks that the entry/exit of locals might generate in certain regions; $X_{\{i\}}$ is a vector of individual characteristics that include potential experience, potential experience squared, years of schooling, gender, a dummy for whether the individual is a part-time worker, and the logarithm of the aggregate labor force.; τ_t is a set of monthly dummies that control for seasonal patterns, α_r are regional fixed effects, and t is a time trend. The coefficient of interest, β_1, is an estimate of the average impact that Venezuelan immigrants generate in Colombia.

The use of a difference-in-differences model imposes the assumption of *parallel trends*. That is, had not the migration shock happened, trends on wages and employment would have been similar across regions. Using an event study[10], I evaluate whether there are differences

[8] There are 30 departments collected in the data plus 13 metropolitan areas. From here on, the word 'regions' will refer to both, departments and metropolitan areas. Thus, 43 regions are having different levels of exposure to the supply shock.

[9] Recall that I referred to this event as the first wave of migration.

[10] To analyze the dynamic impact of migration and the parallel trends assumption, an event study is conducted using the following specification:

$$Y_{i,rt} = \alpha_r + \tau_t + \sum_{t=2013q1}^{2019q4} \gamma_t \ \tau_t \mathbb{1}[M_{Ven,r} \geq \tilde{M}] + u_{i,rt}$$

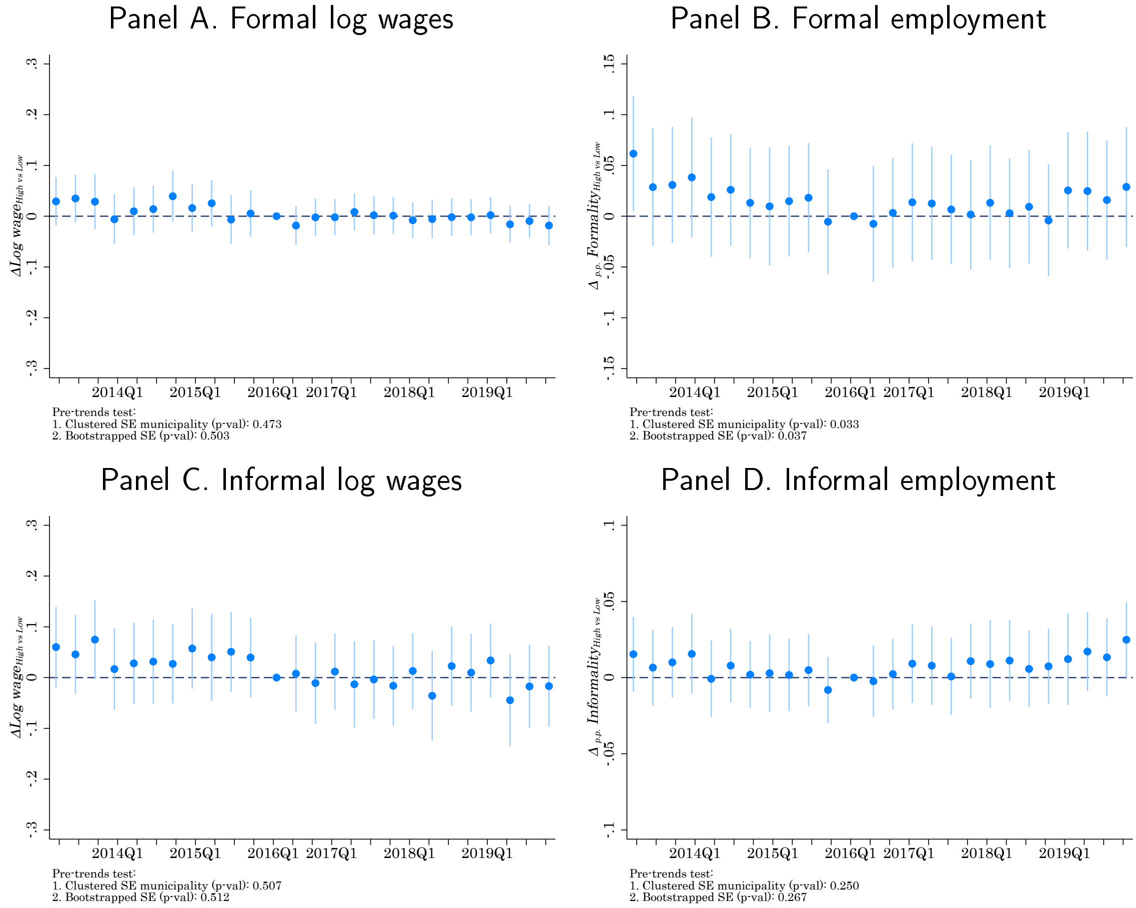
, where $\mathbb{1}[M_{Ven,r}|t \geq \tilde{M}]$ represents a dummy that takes the value of 1 if the average migration shock in the in wages and employment between regions with a high influx of immigrants and those that received relatively few. In particular, I classify all regions into two groups: those that received an immigration shock above the median and those below the median. The event study shows the differences between those two groups every quarter. The omitted category is the second quarter of 2016, right before the reopening of borders.

The results of the event study, depicted in Figure 5, show that there are no significant differences in trends between the regions that received a high influx of immigrants with those that did not. Joint coefficient tests with clustered as well as bootstrapped standard errors confirm that differences between regions are not significantly different from zero before the re-opening of borders for both, wages and employment in the formal and informal sectors. Thus, the events study provide evidence supporting that trends in the growth of wages and employment were similar regardless of the shock the region received.

---

post-treatment period in region r is above the median $\tilde{M}$, and 0 otherwise. α_r and τ_t are region and time fixed effects, respectively.

Figure 5: Event study comparing regions with migration shocks above and below the median



Notes: These set of graphs present quarterly estimates coming from an event study that compares the respective outcome in regions with high and low inflows of immigrants. Each point corresponds to one coefficient that measures this difference at each quarter. The vertical line is a 95% confidence interval. The black horizontal dash line is set at 0. The point of reference, T_0, is where the blue dot without a confidence interval is located: in the second quarter of 2016, right before the re-opening of borders. Equality in pre-treatment trends are tested using a joint coefficient test of the coefficients before the second quarter of 2016. The first tests uses clustered standard errors at the region-quarter level, while the second tests uses bootstrapped standard errors. Panels A and C display the results of the event study in formal and informal wages, respectively. Similarly, Panels B and D depict estimated coefficients in formal and informal employment, respectively.

The pure-spatial approach estimates the national average of the local wage and employment elasticities to immigration. But one might also be interested in computing the size of the effect in region r. In particular, one would be interested in knowing whether the region with the highest influx of immigrants had effects consistent with the national impact. The main challenge is that we do not observe what would have happened with region r in the absence of the migration-induced supply shock.

To compute city-level effects, I construct artificial counterfactual cities for each of the 5 metropolitan areas with the highest influx of immigrants following the methodology by Carvalho et al. (2018). Each of these cities is compared with its artificial counterfactual in the post-treatment period to make inference. This methodology requires high dimensionality in the data. For this reason, data is aggregated at the regional level to construct a time-series panel with monthly frequency. The dataset used in this methodology has $r_{low} + 1$ regions, the donors and the treated unit, observed T times.

The Artificial Counterfactual (ArCo) methodology follows two steps:

1. Suppose y_t is the outcome we are evaluating, say wages or employment. For the unit of analysis r, we do not observe what would have happened had it now received a high influx of immigrants, that is, we do not observe y_t^0 for $t \geq T_0$, where T_0 is the time the unit started to receive the migration-induced labor shock. However, we can predict that scenario using the first $T_0 - 1$ observations of the time series and the following model: $y_t^0 = \mathcal{M}_t + \varepsilon_t$, where $\mathcal{M}$ is an elastic net in this paper but could be any measurable mapping. In other words, this step computes an artificial treated region in the pre-treatment period using the pre-treatment observations of the donor units.

2. Once we have computed the first step, we can compute the counterfactual in the post treatment period $\hat{\mathcal{M}}_t$ for $T_0, ..., T$.

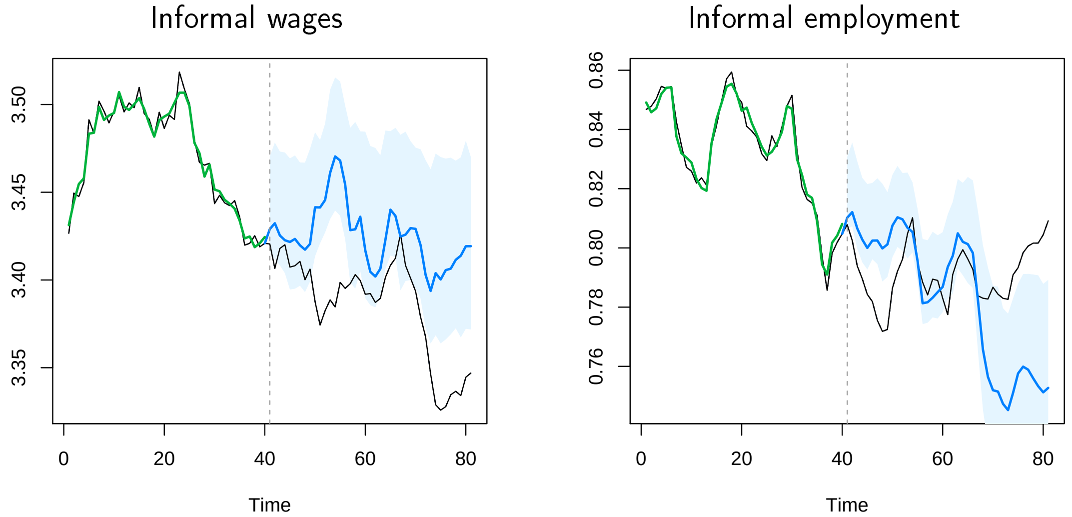
The ArCo estimator is then

$$\hat{\Delta}_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} (y_t - \hat{y}_t^0) \tag{2}$$

Although this methodology is similar to the Synthetic Control methodology pioneered by Abadie and Gardeazabal (2003), the ArCo method offers two key advantages for the study of labor outcomes. First, it does not rely on a convex combination of donor observations to construct the counterfactual, which can lead to biased estimators (Ferman and Pinto, 2016). Second, it exploits time variation of outcomes, which increases the number of observations from which the counterfactual is build. On the contrary, the synthetic control usually removes the time-series dynamics because it uses time averages of the observed donors. Thus, the number of observations used to compute weights in a pure cross-sectional setting is very small.

Figure 6 illustrates how this methodology fits the data in Cúcuta, the largest entry point from Venezuela to Colombia. The figure depicts the actual data in black for both, informal wages (on the left) and informal employment (on the right). The green lines portray the fitting of the elastic net on the training data. Note that the training data is that of the pre-treatment period or the monthly observations before the re-opening of borders. The vertical dashed line visually marks this point in time. The figure shows that the elastic net model is able to fit the data almost perfectly. Using the results of the training data, the counterfactual is constructed and depicted in light blue along with its confidence interval at the 95%. The average difference between the light blue line and the dark blue line in the post-treatment period is the average treatment effect ($\hat{\Delta}_T$). Preliminary results from the visual inspection of this figure point towards a decrease in informal wages and an increase in informal employment of nationals in the city of Cúcuta.

Figure 6: Fitting of the artificial counterfactual to the data in Cúcuta, the main entry city.



Notes: Both graphs display the fitting of the ArCo methodology to the data using as example the results for Cúcuta, the main border city. The graph on the left shows the results in informal wages and the one on the right in informal employment. The black line depicts the observed data. The green line shows how the elastic net fits the observed data in the pre-treatment period. The light blue line depicts the prediction of the model in the post-treatment period along with a 95% confidence interval. The vertical dashed line signals the moment in time when borders re-opened. Results for other regions are shown in Figures C2 and C3 in Appendix C.

5 Results

Table 2 shows the main results for the *difference-in-differences* model described in equation 1. Panel A presents effects on wages and Panel B on employment. Each column presents results coming from a linear regression that uses the sample of Colombians who have not changed their place of residence in the last year. $M_{Ven,rt}$ measures the supply shock from the influx of Venezuelan immigrants while $M_{Ret,rt}$ measures the equivalent shock from Colombian returnees. Although Table C4 presents evidence that the arrival of immigrants is not instigating Colombians to out-migrate to other cities within the country[11], this and subsequent estimations in this paper control for $M_{Int,rt}$. The literature shows that failing to control for this effect would bias the estimates towards zero (Card, 2001; Borjas, 2003).

[11] Table C4 shows estimates of a linear regression of the flow of Venezuelans into city r at time t on the flow of internal migrants. The regression includes region and time fixed effects. The regression shows that for an increase in 1 percentage point in the number of Venezuelan immigrants as a share of the labor force, there is a displacement of 0.01 percentage points. This effect is small and not significant.

Table 2: Effects on wages and employment

	(1) Formal	(2) Informal	(3) Both
Panel A. Wages			
$M_{Ven,rt}$	-0.000	-0.006***	-0.005***
	(0.001)	(0.002)	(0.002)
Observations	276,332	689,062	965,394
R-squared	0.049	0.091	0.133
Panel B. Employment			
$M_{Ven,rt}$	0.002*	0.001	0.001*
	(0.001)	(0.000)	(0.000)
Observations	404,347	927,731	1,223,911
R-squared	0.143	0.053	0.038

Notes: This table presents the results of the estimation of Equation 1 using a difference-in-differences design. These regressions use the sample of Colombians who have not changed their city of residence in the past year and who have completed secondary or less. Panel A displays the effects on wages while Panel B shows the effects on employment. A description of the labor market outcomes is available in Appendix A. Column (1) presents the results in the formal sector, Column (2) in the informal sector, and Column (3) in both. Standard errors, in parentheses, are clustered at the region-month/year level. *** p<0.01, ** p<0.05, * p<0.1

There is a negligible reduction in wages of nationals cause by the arrival of Venezuelan Immigrants. If the number of immigrants as a share of the labor force were to increase one percentage point, the reduction would be of 0.6% in the informal sector. Colombian returnees on the other hand generate a larger impact. There is a reduction of 5.3% in wages in the informal sector of the nonimmigrant population if the share of returnees were to increase 1% percentage point. There are no effects of immigration on formal wages. The latter can be partially attributed to the fact that immigrants enter at a lower rate into formal markets, in spite of the relatively easy access to work permit. Thus, the downward pressure on wages is larger in informal than formal markers.

These close to zero effects are not abnormal in the literature of immigration that use a pure-spatial approach. Card (2001); Boustan et al. (2010); Dustmann et al. (2012)provide examples with similar effects in the United States and the United Kingdom. Part of the rational for the lack of effects on wages is related to rigidities in the short term. The presence of institutional constraints such as minimum wages and unions makes it challenging for wages to be elastic downwards. This is especially true in the low tail of the nonimmigrant wage distribution, which is the focus of this paper.

Turning to the effects on employment, the results suggest null effects. If anything, for every increase of one percentage point of the immigrant population there is an increase in formal employment of 0.2 percentage points. However, this effect is only significant at the 10% level of confidence. There is no significant effect on informal employment. The large number of observations imply this is a very precise null effect. This effect is not uncommon in the literature either. On the contrary, papers that find adverse effects on employment caused by migration-induced supply shocks usually fail to define the appropriate labor market that immigrants enter (for instance, those that use the pure national education-experience cell approach).

Two main reasons might be driving the results on employment: First, there is a relatively sizable segment in the literature of migration that shows immigrants are risk takers. This characteristic makes them more prone to be entrepreneurs and innovators (Bahar et al., 2020b; Bernstein et al., 2018). If immigrants are becoming entrepreneurs upon arrival or becoming self-employed, employment of the nationals would remain unaffected. Second, the time frame of the Venezuelan exodus is not long enough to let the market react to the shock. This is a plausible explanation given that the peak was achieved in May 2018, and this paper uses data until December 2019. Future research to evaluate the medium and long-term effects of this migration wave is thus needed.

The city-level effects, which compares selected cities with their own *Artificial Counterfactual city*, draw similar results. Figures C2 and C3 depict the fitting of the elastic net in 6 selected cities/regions on the training data (in green), and the prediction on the post-treatment period (in blue). It shows that each artificial city fits the data perfectly in the pre-treatment period.

The comparison between the observed data and the artificial counterfactual identifies the treatment effect from the migration-induced supply shock. Table 3 summarizes the results. Columns (1) and (4) show the size of the impact on wages and employment while columns (2)-(3) and (5)-(6) display their confidence intervals at the 5% level of confidence. Panel A displays results on wages while panel B on employment. In terms of wages, the cities most hit by the immigration wave are the cities of Cúcuta and Bucaramanga in the informal sector, and the department of Atlántico in the formal sector. Although statistically significant, these effects are still small. In Cúcuta, the most important border city, an increase in the immigrant share of 1 percentage point leads to a decrease in hourly earnings of up to 4%. For all other cities the effect is also small or nonsignificant.

Table 3: Effects on regions with high immigration shocks using artificial counterfactuals

	Informal sector			Formal Sector		
	(1)	(2)	(3)	(4)	(5)	(6)
		Lower	Upper		Lower	Upper
	$\hat{\Delta}_T$	Bound	Bound	$\hat{\Delta}_T$	Bound	Bound
Panel A. Wages						
Antioquia	0.050	-0.043	0.144	0.025	-0.043	0.094
Atlántico	0.038	0.014	0.062	-0.067	-0.087	-0.047
Barranquilla	0.002	-0.031	0.036	0.072	0.054	0.091
Bogotá	0.029	0.018	0.040	0.004	-0.044	0.052
Bucaramanga	-0.074	-0.092	-0.055	-0.040	-0.094	0.014
Cali	0.000	-0.019	0.019	0.043	-0.069	0.155
Cartagena	-0.012	-0.074	0.050	0.020	-0.008	0.048
Cúcuta	-0.040	-0.066	-0.013	-0.039	-0.117	0.039
Medellin	0.053	0.036	0.071	0.041	0.032	0.049
Pereira	0.033	0.004	0.062	0.015	-0.024	0.054
Villavicencio	-0.016	-0.030	-0.002	0.036	0.006	0.065
Panel B. Employment						
Antioquia	-0.025	-0.031	-0.019	-0.028	-0.041	-0.014
Atlántico	0.012	0.002	0.021	-0.009	-0.034	0.015
Barranquilla	0.004	-0.008	0.017	0.007	-0.002	0.015
Bogotá	-0.007	-0.105	0.090	-0.012	-0.046	0.021
Bucaramanga	-0.006	-0.010	-0.001	-0.006	-0.016	0.004
Cali	-0.028	-0.033	-0.023	0.005	-0.002	0.012
Cartagena	0.017	-0.064	0.098	-0.004	-0.025	0.018
Cúcuta	0.008	-0.100	0.115	-0.007	-0.712	0.697
Medellin	-0.023	-0.029	-0.017	-0.015	-0.031	0.001
Pereira	0.024	-0.003	0.052	0.040	0.016	0.065
Villavicencio	-0.008	-0.023	0.008	-0.005	-0.011	0.001

Notes: This table presents the results of the estimation of Equation 2 using the ArCo methodology. These models use a panel time series constructed by aggregating observations at the regional level. Panel A displays the effects on wages while Panel B shows the effects on employment. A description of the labor market variables is available in Appendix A. Columns (1) and (4) display the average treatment effects for each city/region. Columns (2)-(3) and (5)-(6) show a 95% confidence interval.

6 Heterogeneous Effects

Although the national and city-level effects are negligible, one might think that the arrival of migrants has affected specific groups of the population. To test this hypothesis, I first assess whether women are more affected by this shock. Using a variation to the model of Equation 1 in which a triple interaction is included, I find that formal and informal wages are relatively less affected in women (see Table C2). Remembering that the average effect is zero, this result reinforces the tendency to zero. Table C3 shows that the effects on employment are not different for women and men in the informal sector. However, the results in the formal sector indicate that women benefit relatively more from the migration shock by increasing their holding of formal jobs.

One can also think that younger generations are the most affected by the influx of immigrants due to their shorter experience and lack of credentials. Using a triple difference model for categories of variables, evidence is found that, compared to workers over 55 years of age, the youth are slightly more affected in terms of wages. The size of this effect is, however, small. For an increase in 1 percentage point on the share of immigrants, individuals aged 15-24 perceive a reduction in informal wages of 0.7% while an increase in formal wages of 0.2%.

As can be seen in figure C1, the majority of Venezuelan immigrants enter service and sales occupations, administrative jobs, and elementary occupations. For this reason, tables C3 and C2 present results that focus on these sectors. Each row of panel C comes from a linear model that distributes Venezuelans in the national territory according to the distribution shown in Figure C1. The final result is a regression that assumes there is a migrant-induced supply shock for each occupation-region-month cell. The results show a significant 5.3% reduction in hourly wages for those Colombians working in elementary occupations in the informal sector. The drop in wages is smaller for sales, service and clerical support workers. In terms of employment, the estimates suggest that migration has no impact on these sectors. In other words, labor demand is capable of absorbing the shock quickly.

7 Final remarks

Large immigration waves cause concern among policymakers and nationals. According to Gallup (2012-2014), 1 in 3 people think immigrants take jobs that nationals want. Colombia has not been an exception to this rule. The recent influx of Venezuelans to Colombia has been very substantial for a country the size of Colombia.

In this paper, I studied the effects on labor markets of a massive and unprecedented immigration flow from Venezuela to Colombia. In this context, I take advantage of the timing of events, the size of the event, and an open-borders policy to evaluate how the surge of immigration has affected wages and employment in the formal and informal sectors for Colombians. But evaluating if there is an effect of immigration on wages has been difficult in the past for two main reasons: first, it is difficult to find large and unexpected (exogenous) waves of immigration, luckily, the venezuelan exodus enters into this category; and two, it is difficult to identify where migrants settle once they cross the border in the presence of documented and undocumented migrants.

To address the latter, I use the Internet search intensity of keywords that Venezuelans, not Colombians, are more likely to use. The results show that the Internet search index is more parsimonious than official records in contexts where immigration is relatively a new phenomenon. While the official records overestimate the presence of immigrants in large and border cities, the Internet search index shows immigrants are spread across the national territory. This measure seems to work in other contexts too. For example, I show evidence of the high correlation between the Internet searches of the words 'Green Card' and the percentage of foreign-born in the US States.

Combining the time and geographical variation in the Internet search index, individual information on wages and employment from the Colombian Labor Market Survey, and a difference-in-differences design, the findings reveal negligible changes in wages in both formal and informal sectors due to migration-induced supply shocks. If anything, there are mild reductions in wages of natives working in occupational labor markets in which migrants are entering disproportionally more, such as elementary occupations, services, and clerical jobs. The results suggests that, in the worst-case scenario, a one-percentage point increase in the migrant labor supply will result in a decrease of up to 5 cents per dollar. Regressions also suggest a precisely estimated zero change in employment.

Using artificial counterfactual (ArCo) cities to compute city-level effects, the results are consistent with those of the difference-in-differences model for most cities. That is, there are mild reductions in wages and null effects on employment. Altogether, the findings support the idea that opening borders and allowing immigrants to enter a country freely do not damage the labor prospects of their native population. For the main entry point, Cúcuta, there are slightly larger adverse effects.

In summary, this paper finds that the open borders policy that Colombia implemented has not generated adverse effects on the labor outcomes of Colombians. Other upcoming papers point in the same direction (see for example Bahar et al. (2020a); Tribín-Uribe et al. (2020); Morales-Zurita et al. (2020)). This paper adds to a growing wave of literature that is accumulating evidence supporting open-door policies. This branch of literature has also found that, contrary to the believe of some, immigrants do not generate increases in crime or victimize nationals (Tribín-Uribe and Knight, 2020). On the contrary, in the long term, the literature has found evidence of the benefits that the arrival of immigrants has in terms of entrepreneurship and innovation. Newly produced research have also found that places where immigrants go reduce unemployment because they cause increases in housing value and increases in consumption (Howard, 2019; Tribín-Uribe et al., 2020).

References

Abadie, Alberto and Javier Gardeazabal, "The economic costs of conflict: A case study of the Basque Country," *American economic review*, 2003, *93* (1), 113–132.

Arango, Diana, Lucia Hanmer, Eliana Rubiano, and Julieth Santamaria, "How does poverty differ among refugees Taking an age and gender lens to the data," *World Bank Publications, forthcoming*, 2018.

Bahar, Dany, Ana Ibanez, and Sandra Rozo, "Give Me Your Tired and Your Poor: Impact of a Large-Scale Amnesty Program for Undocumented Refugees," 2020.

_ , Prithwiraj Choudhury, and Hillel Rapoport, "Migrant inventors and the technological advantage of nations," *Research Policy*, 2020, p. 103947.

_ , T Piccone, and Harold Trinkunas, "Venezuela: a path out of misery," *Brookings. edu*, 2018.

Basso, Gaetano, Francesco D'Amuri, and Giovanni Peri, "Immigrants, labor market dynamics and adjustment to shocks in the Euro Area," *IMF Economic Review*, 2019, *67* (3), 528–572.

Bernstein, Shai, Rebecca Diamond, Timothy McQuade, Beatriz Pousada et al., "The contribution of high-skilled immigrants to innovation in the United States," Technical Report 2018.

Borjas, George J, "Self-selection and the earnings of immigrants," Technical Report, National Bureau of Economic Research 1987.

_ , "The economics of immigration," *Journal of economic literature*, 1994, *32* (4), 1667–1717.

_ , "The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market," *The quarterly journal of economics*, 2003, *118* (4), 1335–1374.

Boustan, Leah Platt, Price V Fishback, and Shawn Kantor, "The effect of internal migration on local labor markets: American cities during the Great Depression," *Journal of Labor Economics*, 2010, *28* (4), 719–746.

Brücker, Herbert and Elke J Jahn, "Migration and wage-setting: reassessing the labor market effects of migration," *Scandinavian Journal of Economics*, 2011, *113* (2), 286–317.

Cameron, Maxwell A. and Flavie Major, "Venezuela's Hugo Chavez: Savior or Threat to Democracy," *Latin American Research Review*, 2001, *36* (3), 255–266.

Card, David, "The impact of the Mariel boatlift on the Miami labor market," *ILR Review*, 1990, *43* (2), 245–257.

_ , "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Carpio, Ximena V Del and Mathis Wagner, *The impact of Syrians refugees on the Turkish labor market*, The World Bank, 2015.

Caruso, German, Christian Gomez Canon, and Valerie Mueller, "Spillover effects of the Venezuelan crisis: migration impacts in Colombia," *Oxford Economic Papers*, 2019.

Carvalho, Carlos, Ricardo Masini, and Marcelo C. Medeiros, "ArCo: An artificial counterfactual approach for high-dimensional panel time-series data," *Journal of Econometrics*, 2018, *207* (2), 352 – 380.

Ceritoglu, Evren, H. Burcu Gurcihan Yunculer, Huzeyfe Torun, and Semih Tumen, "The impact of Syrian refugees on natives' labor market outcomes in Turkey: evidence from a quasi-experimental design," *IZA Journal of Labor Policy*, Jun 2017, *6* (1), 5.

DANE, "Rueda de prensa. Resultados Mercado Laboral," https://www.facebook.com/watch/live/?v=3366521323562668&ref=watch_permalink 2019.

Dekker, Rianne and Godfried Engbersen, "How social media transform migrant networks and facilitate migration," *Global Networks*, 2014, *14* (4), 401–418.

Dinero, "Maduro aumenta 300% el mínimo en Venezuela," https://www.dinero.com/internacional/articulo/maduro-aumenta-300-salario-minimo-en-venezuela/266002 2019. Accessed: 6.30.2019.

Dustmann, Christian, Tommaso Frattini, and Ian P. Preston, "The Effect of Immigration along the Distribution of Wages," *The Review of Economic Studies*, 04 2012, *80* (1), 145–173.

_ , Uta Schönberg, and Jan Stuhler, "The impact of immigration: Why do studies reach such different results?," *Journal of Economic Perspectives*, 2016, *30* (4), 31–56.

El Comercio, "Salario mínimo en Venezuela: Qué se puede comprar con 18 mil bolívares (20,9 dólares)?," https://elcomercio.pe/mundo/venezuela/salario-minimo-venezuela-comprar-18000-bolivares-pais-nicolas-maduro-nnda-nnlt-noticia-597 2019. Accessed: 6.30.2019.

Escalona, Alejandro, "Venezuela: cronologia de los ultimos 20 anos," 7 2017.

Ferman, Bruno and Cristine Pinto, "Revisiting the synthetic control estimator," *Working paper*, 2016.

Foged, Mette and Giovanni Peri, "Immigrants' effect on native workers: New analysis on longitudinal data," *American Economic Journal: Applied Economics*, 2016, *8* (2), 1–34.

Fonseca, Yuri R, Ricardo M Prates, Marcelo C Medeiros, and Gabriel FR Vasconcelos, "ArCo: An R package to Estimate Artificial Counterfactuals.," *R Journal*, 2018, *10* (1), 91.

Freitez, Anitza, "La emigracion desde Venezuela durante la ultima decada.," *Revista temas de coyuntura*, 2011, (63).

Friedberg, Rachel M, "The impact of mass migration on the Israeli labor market," *The Quarterly Journal of Economics*, 2001, *116* (4), 1373–1408.

Gallup, "World Split on Whether Immigrants Take Jobs Away," https://news.gallup.com/poll/186425/world-split-whether-immigrants-jobs-away.aspx 2015.

Hernandez, Igor, Francisco Monaldi et al., "Weathering Collapse: An Assessment of the Financial and Operational Situation of the Venezuelan Oil Industry," Technical Report, Center for International Development at Harvard University 2016.

Howard, Greg, "The migration accelerator: Labor mobility, housing, and demand," 2019.

IOM, "Determinants of Internet Use Among Migrants in South-East Asia: A case study of internal migrants in Thailand and Viet Nam," *Migration Research Series*, 2019, (58).

— , "World Migration Report 2020," 2020.

Lebow, Jeremy, Jonathan Moreno Medina, and Horacio Coral, "Immigration and Trust: The Case of Venezuelans in Colombia," *Available at SSRN 3660641*, 2020.

Lopez, Virginia and Jonathan Watts, "Nicolas Maduro narrowly wins Venezuelan presidential election," 04 2013.

Martin, S and Lisa Singh, "Data analytics and displacement: Using big data to forecast mass movements of people," *Digital lifeline? ICTs for refugees and displaced persons*, 2018, pp. 185–206.

Migración Colombia, "Todo lo que quiere saber sobre la migración venezolana y no se lo han contado," 2019.

— , "Más de 1 millón 825 mil venezolanos estarían radicados en Colombia," 2020.

— , "Número de venezolanos radicados en Colombia desciende por primera vez en 5 años," 2020.

Ministerio de Relaciones Exteriores, "Resolución 0872," 2019.

Morales-Zurita, Leonardo Fabio, Leonardo Bonilla-Mejía, Didier Hermida, Luz Adriana Flórez, Leonardo Bonilla-Mejía, Leonardo Fabio Morales, Didier Hermida-Giraldo, and Luz A Flórez, "The Labor Market of Immigrants and Non-Immigrants Evidence from the Venezuelan Refugee Crisis," *Borradores de Economía; No. 1119*, 2020.

Palotti, Joao, Natalia Adler, Alfredo Morales-Guzman, Jeffrey Villaveces, Vedran Sekara, Manuel Garcia Herranz, Musa Al-Asad, and Ingmar Weber, "Monitoring of the Venezuelan exodus through Facebook's advertising platform," *Plos one*, 2020, *15* (2), e0229175.

Peri, Giovanni and Chad Sparber, "Task specialization, immigration, and wages," *American Economic Journal: Applied Economics*, 2009, *1* (3), 135–69.

— **and Vasil Yasenov**, "The labor market effects of a refugee wave synthetic control method meets the mariel boatlift," *Journal of Human Resources*, 2019, *54* (2), 267–309.

Romero, Simon, "Crisis at Colombia Border Spills Into Diplomatic Realm," *New York Times*, 2008, *4.*

Ruiz, Isabel and Carlos Vargas-Silva, "The economics of forced migration," *The Journal of Development Studies*, 2013, *49* (6), 772–784.

Tribín-Uribe, Ana María, Adhvaryu Achyuta, Cesar Anzola, Oscar Ávila-Montealegre, Leonardo Bonilla-Mejía, Juan Carlos Castro-Fernández, Luz Adriana Flórez, Ánderson Grajales-Olarte, Alexander Guarín-López, Franz Hamann-Salcedo et al., "Migración desde Venezuela en Colombia: caracterización del fenómeno y análisis de los efectos macroeconómicos," *Revista Ensayos Sobre Política Económica; No. 97, octubre 2020. Pág.: 1-74*, 2020.

_ **and Brian Knight**, "Immigration and Violent Crime: Evidence from the Colombia-Venezuela Border," *Borradores de Economía; No. 1121*, 2020.

Tumen, Semih, "The use of natural experiments in migration research," *IZA World of Labor*, 2015.

UNHCR, "As Venezuelans flee throughout Latin America, UNHCR issues new protection guidance," 2018.

_ , "Figures at Glance," https://www.unhcr.org/figures-at-a-glance.html 2019.

Universidad de Antioquia, "La carrera de los venezolanos para convalidar sus títulos," *Academia Sociedad*, 2020.

Universidad del Rosario, "Guía de derechos y deberes para venezolanos en Colombia," *Consultorio Jurídico y Centro de Conciliación*, 2020.

Venturini, Alessandra and Claudia Villosio, "Labour market effects of immigration into Italy: An empirical analysis 1," *International Labour Review*, 2006, *145* (1-2), 91–118.

World Bank, "Migración desde Venezuela a Colombia: Impactos y estrategia de respuesta en el corto y mediano plazo," 2018.

Appendix A: Definitions

Experience: It is potential experience. That is the age reported by the individual minus the years of schooling.

Hourly real wages: They are defined as the labor earnings received last month, adjusted by a monthly price index, and divided by the number of hours that an individual usually works per month. The latter is defined as four (weeks) times the number of hours that a person usually works per week. Wages in the formal sector were also adjusted by the contribution employers make to health insurance and retirement plants of their employees. In Colombia, employers contribute 8.5% of their wages to their employees' health insurance plan and 12% to their retirement plans. I also dropped observations below the 1% and above the 99% of the wage distribution.

Labor force: It is defined as the economically active. That is, those people who are 12+ years old and, who are either working or looking for employment[12].

Informality: This paper uses two definitions of informality. The fist measure defines an informal workers as the individuals who do not have access to a health insurance through their employer or does not have a retirement plan. The main results of this paper use this measure. an alternative measure, used by DANE, defines informality according to the size of the firm. If the worker is employed by a firm of 5 or less employees, he is classified as informal. Self-employed also enter in this category. Only if specified in the footnote, usually on appendix tables, this measure will be used.

[12]Although the official definition from the National Bureau of Statistics in Colombia takes into account people who are 10+ years old in rural areas, and those who are 12+ years old in urban areas, I dropped those people aged 10-11 years old from the sample.

Appendix B: Google searches index

Google Trends queries return a Google search index for the location and time range specified. In this paper, I use a query for key terms in Colombia for the past 4 years (2015-2018). The result of this query is a dataset with a unique index per city. The google search index for citi i using this type of query would be:

$$SI_i = \frac{hits_i}{\max_{j \in N}\{hits_j\}} \tag{3}$$

, where SI_i is the search intensity measure that one gets from the query, $hits_i$ is the number of hits of searches of key words in geographical area i during the period 2015-2018. The denominator of the equation returns the number of hits of the city where most of the hits took place. Thus, the Google search intensity index returns, per city, a number that ranges from 0 to 100. The city with most of the searches will be assigned a value of 100 and every other city in the country will be assigned a value relative to that value. Cities with very low search are automatically assigned a value of zero. Google Search also removes multiple searches made by the same person within a specific time period. [13]

Let's calculate a composite index that is the summation of the city indexes

$$\sum_{i=1}^{N} SI_i = \frac{1}{\max_{j \in N}\{hits_j\}} \sum_{i=1}^{N} hits_i \tag{4}$$

Thus, one can calculate the share of hits in location i in the following manner:

$$m_i^2 = \frac{SI_i}{\sum_{i=1}^{N} SI_i} \tag{5}$$

$$= \frac{\frac{hits_i}{\max_{j \in N}\{hits_j\}}}{\frac{1}{\max_{j \in N}\{hits_j\}} \sum_{i=1}^{N} hits_i} \tag{6}$$

$$= \frac{hits_i}{\sum_{i=1}^{N} hits_i} \tag{7}$$

m_i^2 is a proxy for the number of migrants in location i as a share of the total number of migrants.

[13] See https://support.google.com/trends/answer/4365533?hl=en

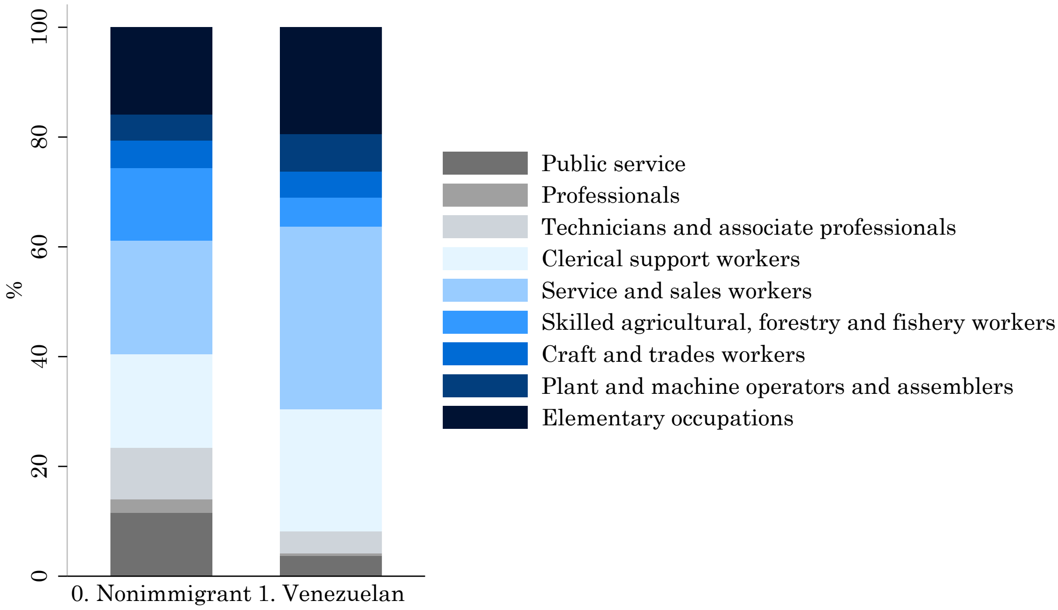
Appendix C: Additional figures and tables

Table C1: Timing of the events

August 19, 2015 •	Maduro orders to close the border crossing in Táchira
First wave •	
August 13, 2016 •	Venezuela-Colombia border is reopened after almost one year
Second wave •	
July 16, 2017 •	Venezuelan Referendum takes place
October 15,2017 •	Maduro's political party wins 17 out of 22 Governorships
Third wave •	
May 20, 2018 •	Presidential elections take place. Maduro is re-elected.

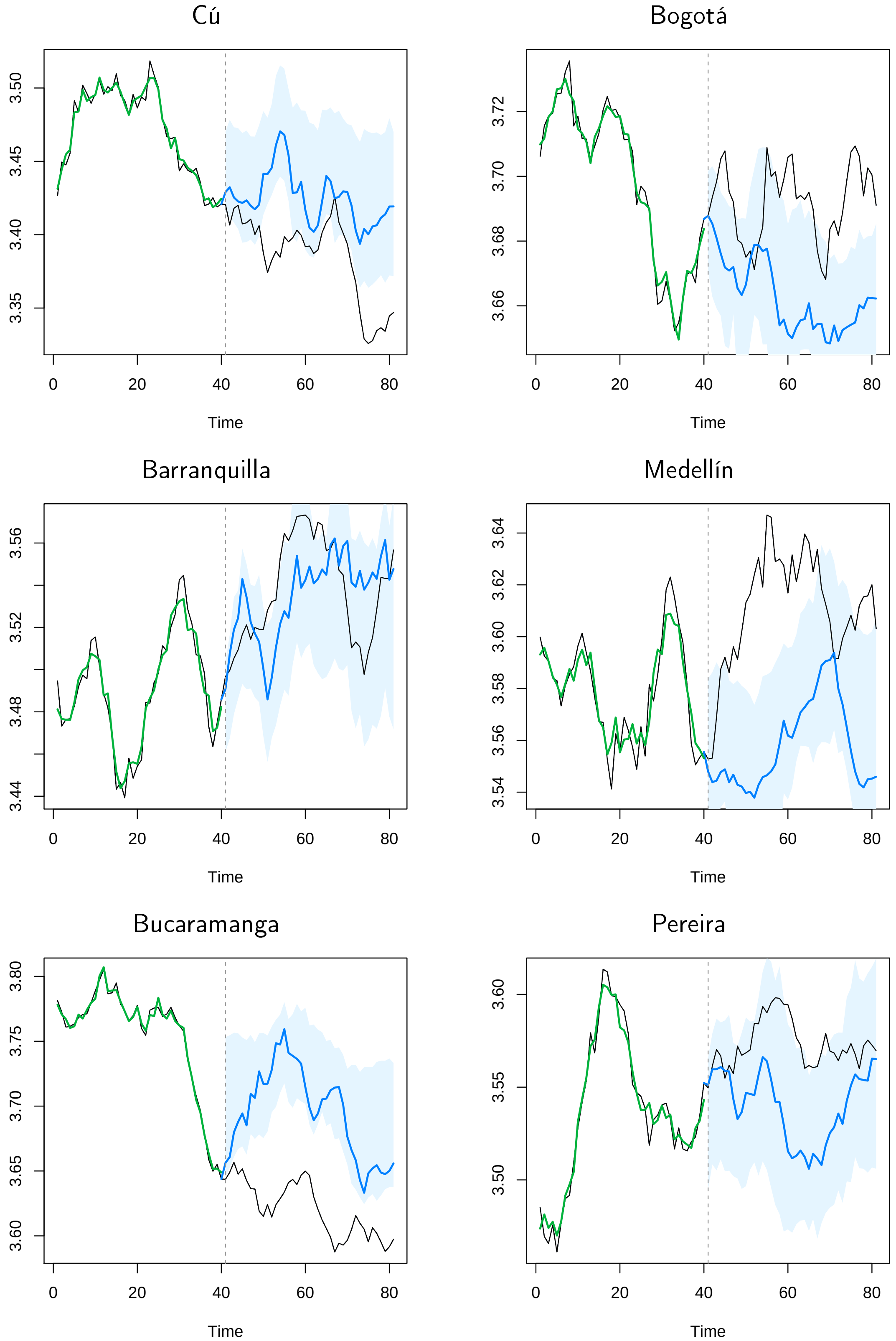
Notes: This table summarizes three key events that unfolded the Venezuelan exodus. Accuracy of the data was cross-checked using multiple journals cited on the references section.

Figure C1: Occupation of nationals and immigrants in Colombia



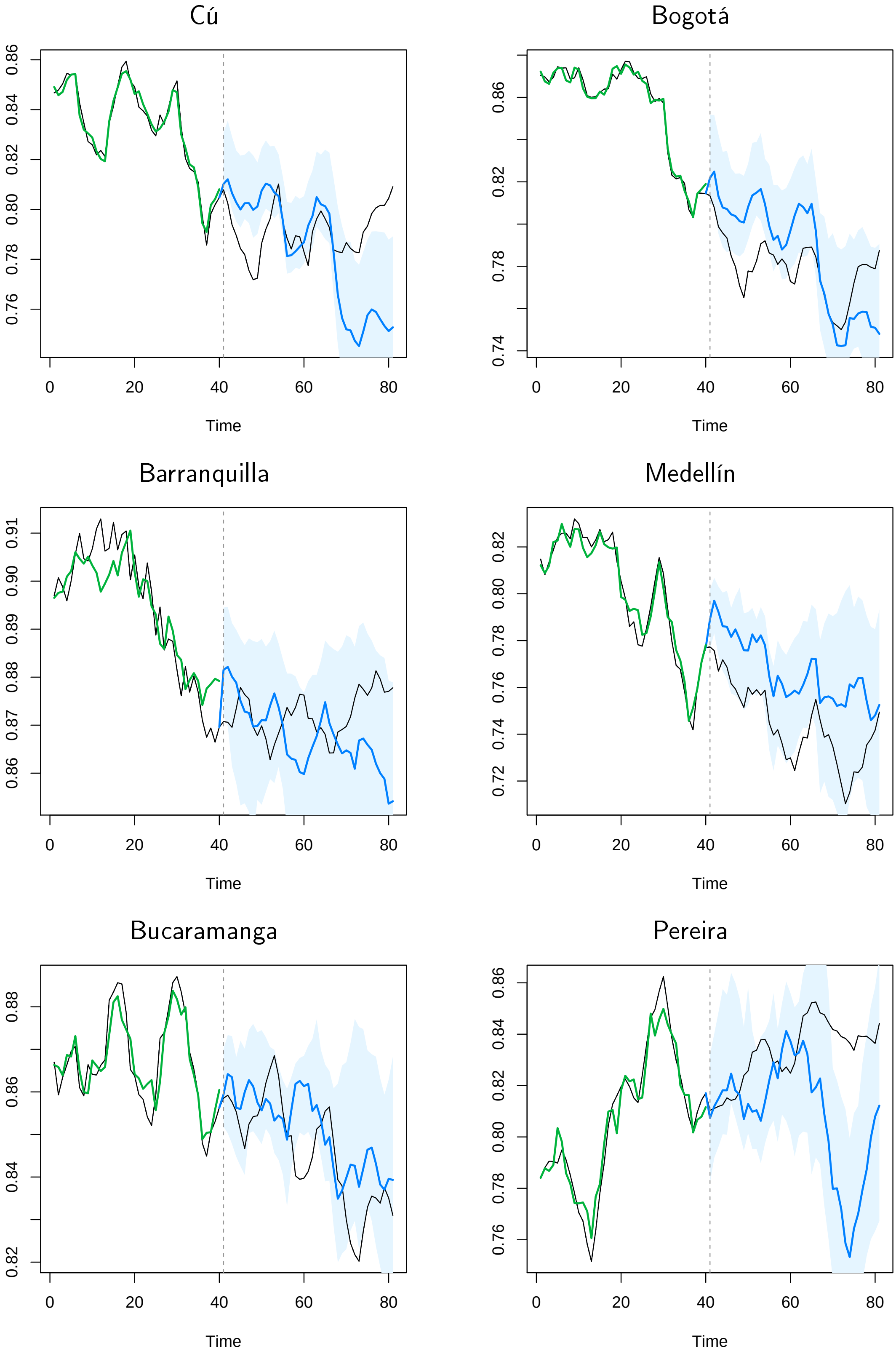
Notes: This graph displays the distribution of nonimmigrants, Venezuelan immigrants and Colombian returnees in the GEIH sample. The distribution is computed using all individuals observed after July 2016, i.e., after the re-opening of borders.

Figure C2: ArCo results in informal wages



Notes: This set of graphs depict the fitting of the ArCo methodology in 6 regions that had a relatively higher influx of immigrants. Al the graphs depict results in informal wages. The black line presents the observed data. The green line shows how the elastic net fits the observed data in the pre-treatment period. The light blue line depicts the prediction of the model in the post-treatment period along with a 95% confidence interval. The vertical dashed line signals the moment in time when borders re-opened.

Figure C3: ArCo results in informal employment



Notes: This set of graphs depict the fitting of the ArCo methodology in 6 regions that had a relatively higher influx of immigrants. Al the graphs depict results in informal employment. The black line presents the observed data. The green line shows how the elastic net fits the observed data in the pre-treatment period. The light blue line depicts the prediction of the model in the post-treatment period along with a 95% confidence interval. The vertical dashed line signals the moment in time when borders re-opened.

Table C2: Heterogeneous effects on wages

	(1) Formal	(2) Informal	(3) Both
Panel A. By gender			
$M_{Ven,rt} \times Female$	0.002*	0.005***	0.007***
	(0.001)	(0.001)	(0.001)
Observations	276,332	689,062	965,394
R-squared	0.058	0.097	0.151
Panel B. By age category			
$M_{Ven,rt} \times 15-24$	0.004**	-0.007***	0.001
	(0.002)	(0.002)	(0.002)
$M_{Ven,rt} \times 25-34$	-0.004**	-0.003*	0.003*
	(0.002)	(0.002)	(0.002)
$M_{Ven,rt} \times 35-44$	-0.006***	-0.006***	-0.001
	(0.002)	(0.002)	(0.001)
$M_{Ven,rt} \times 45-54$	-0.006***	-0.002	0.000
	(0.002)	(0.001)	(0.001)
Observations	276,332	689,062	965,394
R-squared	0.058	0.097	0.151
Panel C. By occupation			
Clerical support workers	0.005	-0.016*	-0.016*
	(0.010)	(0.010)	(0.009)
Observations	33,676	141,766	169,272
Service workers	-0.004	-0.025***	-0.017***
	(0.004)	(0.004)	(0.004)
Observations	76,659	168,195	236,751
Elementary occupations	-0.013	-0.053***	-0.045***
	(0.009)	(0.010)	(0.010)
Observations	67,019	140,107	200,540

Notes: This table summarizes the findings of heterogeneous effects of the migration-induced supply shock on wages. These regressions use the sample of Colombians who have not changed their city of residence in the past year and who have completed secondary or less. Panel A and Panel B show the results of a model of triple differences that tests heterogeneous effects by gender and age category, respectively. Each subpanel from Panel C comes from a different regression that uses the subsample of workers in the corresponding occupation. In this case, each regression uses a migration shock at the region-month-occupation level given that the household survey allows to identify which occupations immigrants enter. Column (1) presents the results in the formal sector, Column (2) in the informal sector, and Column (3) in both. Standard errors, in parentheses, are clustered at the region-month/year level. *** p<0.01, ** p<0.05, * p<0.1

Table C3: Heterogeneous effects on employment

	(1) Formal	(2) Informal	(3) Both
Panel A. By gender			
$M_{Ven,rt} \times Female$	0.009***	0.000	0.001
	(0.001)	(0.001)	(0.000)
Observations	404,347	927,731	1,223,911
R-squared	0.160	0.054	0.038
Panel B. By age category			
$M_{Ven,rt} \times 15-24$	0.001	-0.005***	-0.004***
	(0.002)	(0.001)	(0.001)
$M_{Ven,rt} \times 25-34$	0.001	-0.004***	-0.002***
	(0.002)	(0.001)	(0.001)
$M_{Ven,rt} \times 35-44$	0.001	-0.002***	-0.001**
	(0.002)	(0.001)	(0.001)
$M_{Ven,rt} \times 45-54$	0.004***	-0.000	0.000
	(0.001)	(0.001)	(0.000)
Observations	380,171	820,690	1,103,035
R-squared	0.159	0.055	0.039
Panel C. By occupation			
Clerical support workers	0.006	0.003	0.002
	(0.008)	(0.003)	(0.003)
Observations	59,215	196,722	234,447
Service workers	0.007*	0.004	0.004**
	(0.004)	(0.003)	(0.002)
Observations	115,555	214,779	295,385
Elementary occupations	0.007	0.006	0.004
	(0.008)	(0.004)	(0.003)
Observations	92,674	175,457	245,705

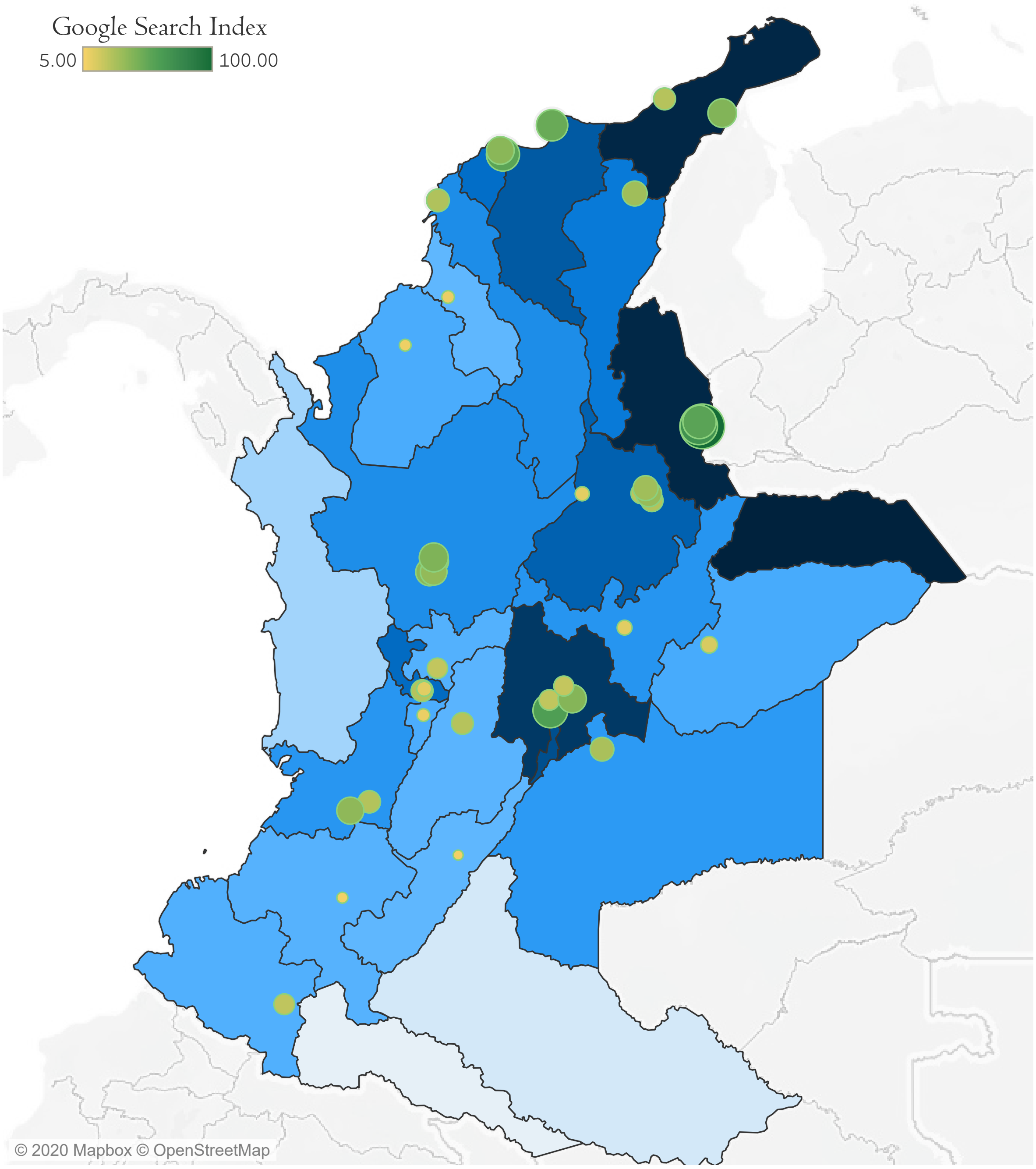
Notes: This table summarizes the findings of heterogeneous effects of the migration-induced supply shock on employment. These regressions use the sample of Colombians who have not changed their city of residence in the past year and who have completed secondary or less. Panel A and Panel B show the results of a model of triple differences that tests heterogeneous effects by gender and age category, respectively. Each subpanel from Panel C comes from a different regression that uses the subsample of workers in the corresponding occupation. In this case, each regression uses a migration shock at the region-month-occupation level given that the household survey allows to identify which occupations immigrants enter. Column (1) presents the results in the formal sector, Column (2) in the informal sector, and Column (3) in both. Standard errors, in parentheses, are clustered at the region-month/year level. *** p<0.01, ** p<0.05, * p<0.1

Table C4: Effect of Venezuelans immigrants on internal migration

	(1)
Variable	$M_{Int,mr}$
$M_{Ven,mr}$	-0.011
	(0.016)
Observations	1,418
R-squared	0.771

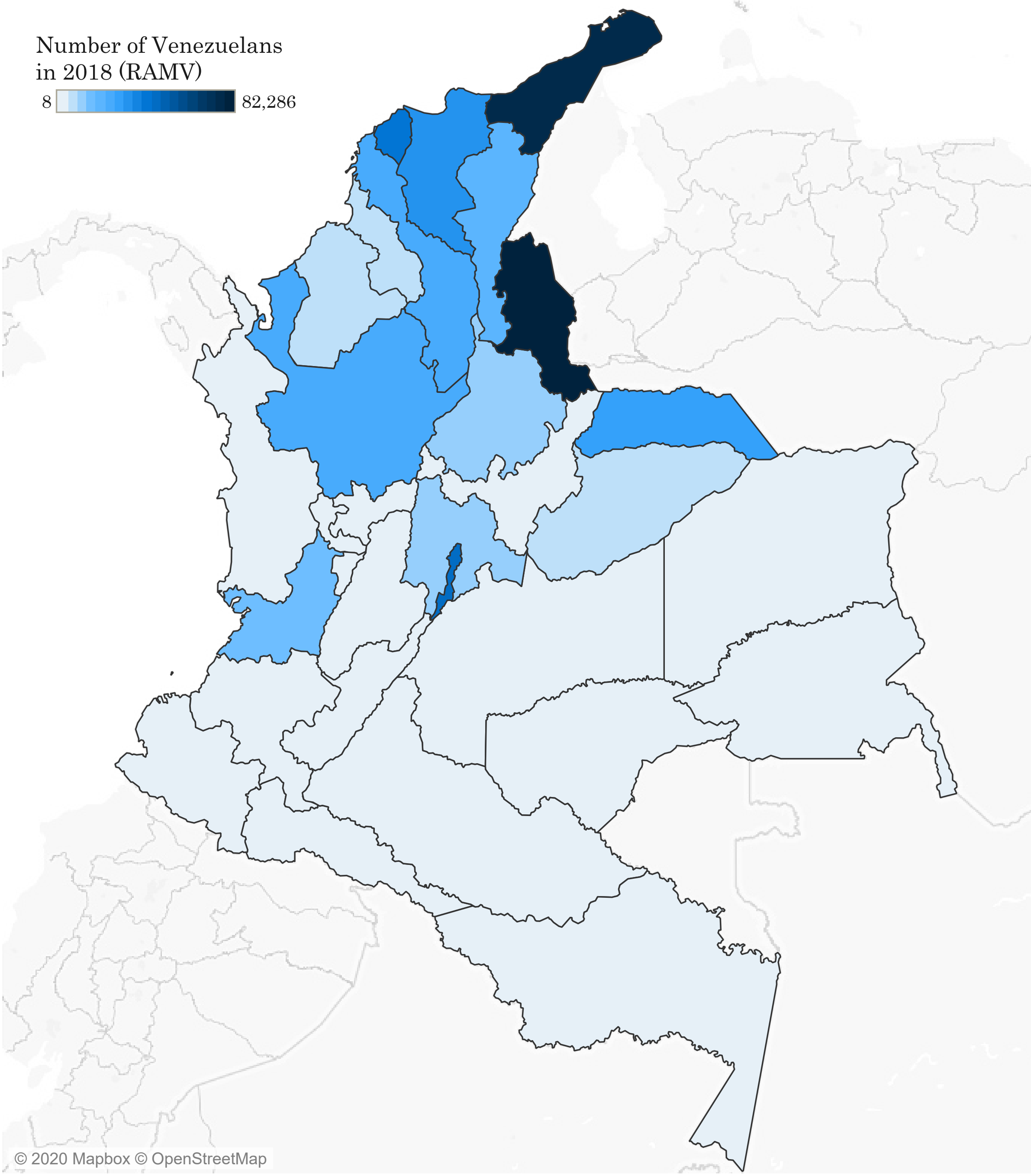
Notes: This table regresses net migration of Venezuelan immigrants on the net migration of internal migrants. The results are obtained from a dataset that contains time and regional variation in the outcomes of interest. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure C4: Geographical distribution of the raw Internet search index



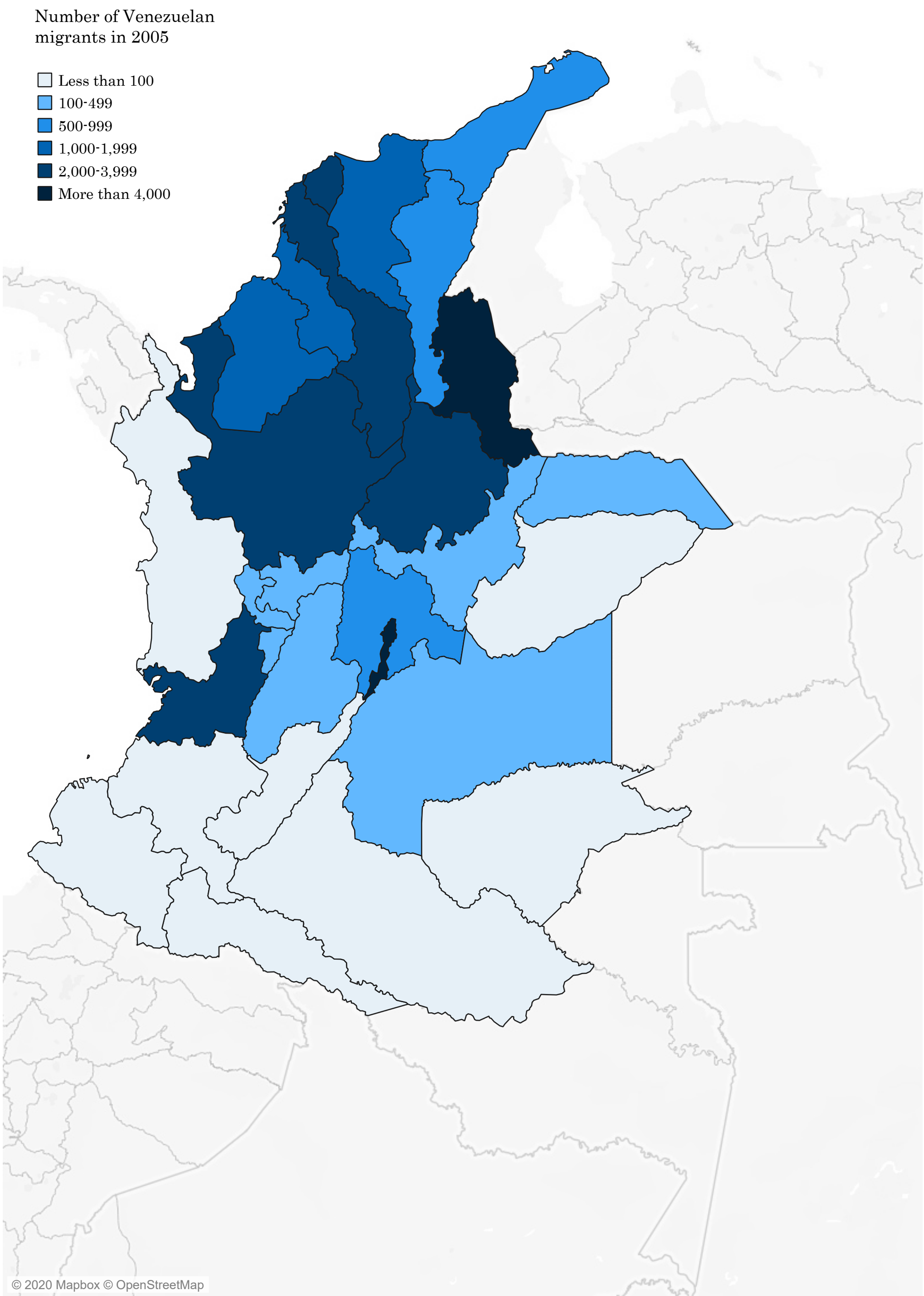
Note: This map depicts the geographical distribution of the number of immigrants according to the Internet search index. Two overlapping distributions are shown. In a scale of blues, the geographical density at the Department level, while in green the geographical density of selected cities. The maximum value of the raw index is 100.

Figure C5: Venezuelan migrant counts according to official records



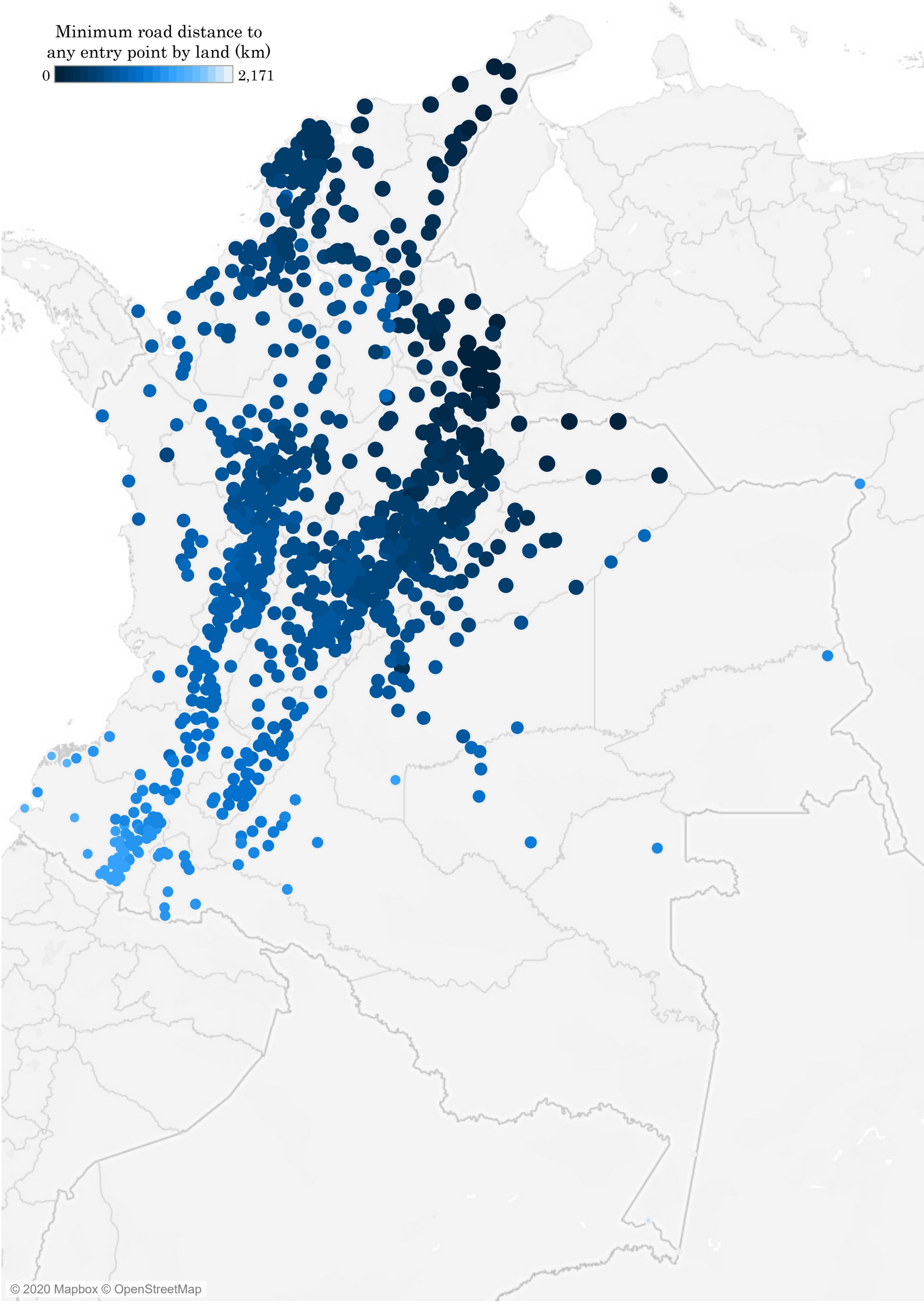
Note: This map depicts the geographical distribution of the number of immigrants by Department using Data from administratiive records (PEP-RAMV). Data was collected from extractions of Tableau dashboards produced by Migración Colombia.

Figure C6: Venezuelan migrants in 2005



Note: Produced by the author using information from IPUMS international 2005. Number of immigrants in 2005 correspond to the weighted sum of the number of people born in Venezuela surveyed in the 10% Census sample.

Figure C7: Road distance to the border



Note: This map depicts the road distance between the closest entry point and the target city. Three official entry points are used: Cucuta, Maicao and Arauca. Notice that the road will not necessarily match linear distance between two points.