Dynamics of Local Wages and Employment: Evidence from the Venezuelan Immigration in Colombia*

Lukas Delgado-Prieto†

December 15, 2020
Click here for latest version

Abstract

The unprecedented socioeconomic and political deterioration of Venezuela has triggered a massive outflow of people leaving the country since 2016, both in a voluntary and a forced manner. Colombia has been the major receiver country with more than 1.2 million working-age Venezuelans (4.1% of the working-age population living in Colombia) as of 2019. I use this quasi-natural experiment to identify the causal impact of the Venezuelan immigration on the Colombian labor market. To analyze dynamic treatment effects I implement an event-study design with two different shift-share instruments. For both instruments I find that immigration from Venezuela had a highly negative short-run effect on local native wages since 2017, and the impact is mainly suffered by less skilled workers and workers without access to social security. Moreover, wages in lower percentiles of the native local wage distribution are severely more affected compared to those in upper percentiles. In terms of native employment, I find a delayed negative response after controlling for preexisting trends. On aggregate, the supply shock affected mainly the informal labor market with lower wages and higher employment on average.

Keywords: Immigration, Event study, Labor market.

JEL Codes: F22, O15, O17, R23.

*I am grateful to Jan Stuhler for all the detailed comments, in-depth suggestions and close guidance, to Luigi Minale for further remarks, to Juanjo Dolado for the extensive revision, to Jesús Fernández-Huertas, to Jorge Pérez-Pérez and to the participants in the UC3M Applied Reading Group. I thank also all the support from the technical staff at DANE.

†Universidad Carlos III de Madrid, Department of Economics, PhD Candidate. Email: ludelgad@eco.uc3m.es.

1 Introduction

The impact of immigration on native wages and native employment is one of the most relevant, albeit disputed, economic questions (see Dustmann, Schönberg, and Stuhler (2016) for a summary of findings). The growing literature on this topic has surged rapidly in the last thirty years (Borjas and Chiswick, 2019). To contribute to this economic debate, I use a recent exogenous and increasing change in the amount of migrants in Colombia, combined with detailed micro data on labor and demographic status. The setting I study is derived from external conditions, more concretely, from the collapse of the Venezuelan economy, that has led to the largest migration crisis in recent times in Latin America and the Caribbean.

To measure the extent of this migration crisis, according to UNHCR (2019), between 2016 and 2019 more than 4.6 million people have left Venezuela, with a daily outflow of around 4,000 to 5,000 Venezuelans. The main destination countries have been Colombia, Peru and Ecuador. These massive and sudden inflows of individuals can influence different socio-economic outcomes, such as employment, health and education in the receiving countries, both in the short and long-term.

In this paper I focus on the labor market impacts of the Venezuelan mass migration on Colombia, since this is by far the biggest destination country (UNHCR, 2019). In Colombia, the labor supply of migrants, measured as working-age Venezuelans over working-age natives, went up from 0.2% in 2015 to 4.1% in 2019. The standard prediction in a model of factor proportions would be that a large and positive labor supply shock reduces the relative price of labor. The effect will of course depend on the skill composition of migrants which, in this study, mainly corresponds to young-uneducated individuals. Thus if migrants have a high degree of substitutability with respect to less-skilled natives, immigration will trigger lower wages for natives. Yet if they have complementary skills to those held by competing natives, the influx could lead to higher wages. In addition, the change in wages could interact with changes on employment of natives, existing a trade-off response to immigration between these outcomes (Borjas, 1999).

My empirical strategy to test the size and sign of previous effects is a yearly difference-in-difference (DiD) approach that takes advantage of the intensity of treatment, given that some states (or *Departamentos*) in Colombia received vast inflows of migrants while others did not. The key assumption needed for the causal interpretation of the DiD parameter is the unconditional parallel trends assumption. In practice this assumption might not be fulfilled, since migrants can endogenously sort in the states that offer the best economic conditions, meaning differential trends of the outcome on the treated areas. To test for possible violations of this assumption, and to take advantage of the staggered and increasing characteristics of the treatment, I implement an event-study design. In addition, to deal with the potential endogeneity issues of migrants' self-selection into areas with further economic opportunities, I use two distinct instruments: 1) distance between capital cities in the two neighboring countries and 2) historical enclaves or past-settlements of Venezuelans. Therefore, an unconditional parallel trend assumption with Instrumental Variable (IV) is required to identify the causal parameters. In practice, I show that the chosen instruments do not predict the trends in native wages before the migration crisis started, which provides indirect support of the identifying assumption.

As regard data, I use two available official sources that provide a rich set of demographics on natives and migrants from Venezuela, namely, the Labour Force Survey of Colombia and the most recent population and housing census from 2018. Both datasets offer plenty information on the individual characteristics of the migrant and native population in Colombia at different geographical levels, removing possible compositional bias that could emerge with the cumulative arrivals of immigrants.

My estimates show that the inflow of Venezuelans persistently reduced native hourly wages since 2017. More concretely, a 1 percentage point (pp) increase in the share of employed Venezuelans over the employed population in each department reduces local native wages by 1.6%-1.7%. Compared to previous literature, my estimates are: (*i*) smaller to those found in related setup by Caruso, Canon, and Mueller (2019), where a 1 pp increase in the share of Venezuelans in Colombia diminishes wages by 7.6%, (*ii*) similar to Edo (2017) findings for the Algerian inflow in France where a 1 pp increase of repatriates lowered wages of natives between 1.3%-2%, and (*iii*), much larger than the results reported in Dustmann, Schönberg, and Stuhler (2017) for the case of the commuting policy in Germany where a 1 pp increase in the overall employment of Czech workers decreased local native wages by 0.13%. Thus, one of the goals of this paper is to understand why the sizable negative short-run effect of migration on wages, that can help to achieve a fast labour market adjustment to the unexpected supply shock.[1]

[1]For instance, Monras (2020) find that low-skilled Mexicans who left their country as a result from the Peso crisis

In addition, Colombians are working more hours: a 1 pp increase in the migration rate increases hours worked per week by 0.9%. If analyzing heterogeneous effects, the reduction on native wages is concentrated among less-educated workers (with high school or less) and those without access to social security. Moreover, most affected wages are located at the lower part of the local wage distribution (in the 25th percentile). A key explanation that could drive previous negative estimates is the lack of downward rigidity of wages, the majority of workers in Colombia are employed on a contractual labor relation without binding minimum wages or formal contracts. Thus the flexibility of wages in that setting is full.

In terms of native employment, I find a delayed negative response since 2018 after controlling for preexisting trends, yet results are sensitive to the exclusion of survey weights in the estimation procedure, as estimates turn to be insignificant. Taking into account the sensitivity of the employment results, low skilled workers (with high school or less) and young ones (between 18 and 25 years) appear to be the most affected groups. On aggregate terms, I find that informal employment grew and informal wages decayed, while formal wages were unaffected and formal employment decreased only after 2017. Thus the supply shock of Venezuelan immigrants mainly affected the informal labor market. A simple model of homogeneous labor is introduced to explain this finding.

In relation to the general literature on migration, the contributions that this paper makes are the following. The characteristics of the supply shock under study, namely, a large and sudden inflow of migrants driven by the conditions in the sending country, help to identify its impact (not many immigration events follow these characteristics, among which, possibly the best known is the Mariel Boatlift in Florida Card (1990)). Yet, in contrast to the latter, I have more than one treatment area, exactly 24 treated areas. In addition, I use a recent population census which gives the most reliable up-to-date figures of the amount of Venezuelans in Colombia, reducing the extent of measurement error or undercoverage bias and consequently the attenuation bias (Aydemir and Borjas, 2011; Amior, 2020). Finally, I assess the impact of immigration on novel outcomes, such as firm creation (where I find a positive estimate in 2016 followed by a negative one in 2017) and child labor (where I find negative point-estimates, though not significant in the majority of post-years).

With respect to previous studies that estimate the impact of the Venezuelan migration on the

had a high transitory labor market impact on the US, that quickly dissipated across states as time passed and local markets adjusted.

Colombia's labor market (Caruso, Canon, and Mueller, 2019; Morales-Zurita et al., 2020; Santamaria, 2019), my contributions can be summarized as follows. First, the very high negative impact of the Venezuelan migration on Colombian wages found by Caruso, Canon, and Mueller (2019) motivates a detailed empirical assessment. To do so, I go beyond a simple comparison of outcomes before-and-after the immigration shock by implementing an event-study design with continuous treatment while using two different identification strategies that can test for the presence of pre-existing trends. Second, the use of the event-study design is also motivated by the fact that the coefficient in Caruso, Canon, and Mueller (2019); Morales-Zurita et al. (2020) panel IV regression, can be interpreted as a weighted average of treatment effects, where some of these weights can even be negative in the presence of differential timing of treatment with dynamic effects (Goodman-Bacon, 2018; de Chaisemartin and d'Haultfoeuille, 2019). Third, I study the effect of immigration not only on the average local wage, but across the entire native wage distribution.

The structure of the paper is as follows. The next section describes the related literature. Section 3 describes the data used and descriptive statistics of natives and immigrants. Section 4 gives a brief overview of the Venezuelan crisis and the institutional background. Section 5 is about the empirical specification and the identification assumptions needed. Section 6 reports the results for different outcomes. Section 7 introduces a simple theoretical model to explain the empirical findings. Section 8 is about the robustness tests performed. Finally, Section 9 discusses and concludes.

2 Related Literature

Early studies on immigration focused on the comparison of local labor markets across different cities (Grossman, 1982; Card, 1990; Hunt, 1992). The identification strategy relied on the sudden and unexpected inflow of migrants in some specific (treated) areas, that if migrants had not arrived, treated and non-treated areas would follow a similar pattern (i.e., the unconditional Parallel Trend Assumption-PTA). The main finding of this literature is that immigration had a small negative impact on wages (in the case of the Algerian inflow in France), or even insignificant effects (in the case of the Mariel Boatlift). This research design is known as the spatial or area approach and identifies the *overall* effect of immigration, tying an observed shock with a particular outcome. I use this approach in this paper.

Nonetheless, as early papers received a lot of scrutiny, several critiques to the previous methods or data used have been made. To name some of these concerns, the first one was that information on the outcome before the treatment happens was often scare, making it the use of tests on whether the PTA holds and the selection of good control groups. Second, it was hard to correctly estimate the standard errors using DiD (Bertrand, Duflo, and Mullainathan, 2004). As a result, there have been several critical replies to the Algerian and Mariel Boatlift papers which have highlighted the controversy around immigration studies.[2] Yet, at the same time, they have raised interesting conceptual points (i.e., statistical inference with only one treated unit, specification choices and placebo tests) that have rigorously improved the study of immigration by means of the spatial approach.

The most related paper to this one, in terms of the empirical specification, is a recent one by Dustmann, Schönberg, and Stuhler (2017) that implements an event-study design with IV to study a commuting policy on the Bavarian region of Germany, they find a 1-to-1 displacement effect on employment that is persistent on time.[3] One aspect to be noted is that all the papers listed above are based on developed countries (i.e., USA, Germany or France). If reviewing the effects of migration on developing countries the literature is more scare. One case study similar in magnitude to the Colombian one is the recent supply shock of Syrian refugees on the Turkish labor market, several papers have analyzed this migration event. For instance, Aksu, Erzan, and Kırdar (2018) find that the influx of Syrians strongly decrease native wages in the informal sector, particularly on low-educated and younger workers, while upgrading wages and employment of natives in the formal sector, similar to Del Carpio and Wagner (2015) findings. As for the case of the Venezuelan emigration, the only published paper, to the best of my knowledge, that estimates its causal impact on the Colombian labor market is Caruso, Canon, and Mueller (2019). Using a panel IV regression they find an immense negative effect on wages, a 1 pp increase in the labor supply of immigrants

[2]For the Mariel Boatlift, first reply of Borjas (2017b) focused on the impact on high-school dropouts wages (omitted in Card (1990)), by 1985 they were reduced by 30% (relative to control cities). Then Peri and Yasenov (2019) reexamine this natural experiment using a synthetic control method, that builds a synthetic control Miami with a composition of cities totally different in comparison to the original Card study. But they find closely similar results to Card (1990), and argue that the Borjas reply was sensitive to the definition of "low-skill" worker. For the Algerian inflow in France Edo (2017) reexamines this experiment with better wage data that allows the separation of the effect in repatriates and natives, author finds strong decline in wages in counter-view to the original Hunt study.

[3]Authors use the exogenous distance to crossing borders in the Czech Republic as instrument.

in Colombia diminishes in 7.6% the hourly wage of workers, a negative effect which is mainly concentrated among by the informally employed and urban workers. Given the large size of this estimate, a reexamination of the previous findings motivates, in part, this study.[4]

There are two more recent unpublished papers that estimate the labor market effects of the Venezuelan immigration in Colombia. The first one is Santamaria (2019) who uses novel figures of immigration flows from Google trends. By means of a DiD research design, this author finds insignificant effects of immigration on wages, and remarkably, negligible reductions on wages among informally employed workers. While in this DiD setup there is pre-treatment data that can remove the unobserved heterogeneity, the endogenous sorting of immigrants poses some weakness for the chosen identification strategy. As Jaeger (2007) and Borjas (2001) have pointed out, immigrants tend to settle in areas that offer the best economic opportunities for the skills they provide. Furthermore, when the supply shock is persistent and increasing on time, there can be anticipation effects that could lead to prior adjustments in the local markets.

The second paper related to this case study is Morales-Zurita et al. (2020) which focuses on the relationship between immigration and unemployment, both for natives and migrants in Colombia. Using a panel IV regression, where instrument is a shift-share of historical enclaves interacted with economic conditions in Venezuela (measured through lagged quarterly inflation), these authors find a negative effect of immigration on unemployment of migrants and insignificant estimates for native employment and wages. A shortcoming of this paper, however, is that they consider all migrants, without distinguishing by time of arrival, so that a compositional bias can arise if recent migrants have different characteristics from the settled migrant population.

Another empirical issue that also arises in Caruso, Canon, and Mueller (2019) and Morales-Zurita et al. (2020) is their use of a pooled regression, where all time periods get stacked. This estimation procedure might not account for differences in local economic trends before migration occurs ("pre-trends"). Although Caruso, Canon, and Mueller (2019) provide evidence on the lack of correlation between historical immigration rates (in 1973) and more recent ones (in 2005) with current outcomes, more recent pre-treatment data is not analyzed. In Morales-Zurita et al. (2020) they use shorter in time pre-treatment data to find null correlations between historical enclaves

[4] The instrument used by Caruso, Canon, and Mueller (2019) is the distance between port of emigration and port of arrival. In my analysis I use the same instrument, but with newer administrative data that allows a more precise characterization of the location in Venezuela where migrants are coming from.

and migration flows between 2013-2015, but they still do not analyze if their instrument predicts trajectories of economic outcomes during those years. Thus the reported estimates found could be capturing other effects, in the sense that the panel estimator compares areas with differing trends on the outcomes that might confound the true impact of immigration. Moreover, as Goodman-Bacon (2018) and de Chaisemartin and d'Haultfoeuille (2019) point out, the coefficient of interest in the two-way fixed effects (TWFE) model consists of a weighted average of treatment effects, where some of these weights can even be negative, then the use of pooled regressions could bias the results when the timing of treatment varies, as TWFE is using already treated groups as control groups.

All in all, when looking at the wage effects of immigration in the above mentioned papers, there is a range of findings (i.e., Caruso, Canon, and Mueller (2019) find an incredible high negative impact while Morales-Zurita et al. (2020) and Santamaria (2019) obtain insignificant estimates), even if they use as a main source of information the same database, namely, the Colombian Labor Force Survey. Since it is likely that different results are driven by the different empirical specifications implemented, or the definition of the migration rate (treatment variable) used, I take these studies as a basis to improve and, convincingly, determine which is the prevailing effect that the Venezuelan immigration has had on the Colombian labor market.

3 Venezuelan Crisis and Institutional Background

3.1 A brief Overview of the Venezuelan Crisis

A recent timeline of the factors that caused the Venezuelan humanitarian crisis can be summarized as follows. When Hugo Chávez died in 2013, his presidential term ended abruptly after more than 14 years as a president of Venezuela. At that time, Venezuela did not have a private sector and their economy was mainly based on the oil industry. In April of 2013, Nicolás Maduro succeeded him after winning, by a narrow margin, the presidential election. After two years of Maduro as a president, in 2015 the economy in Venezuela started to decay as the oil prices almost dropped half, restricting the only source of revenue of the government. This implied reductions on the universal social programs fundings and in the subsidies for basic products, like medicines and food, that Maduro's government had, generating more social discontent. Then, in 2017 the ruling party of

Maduro won the majority of state elections and massive outflows of Venezuelans started leaving the country, fleeing from a growing dictatorship. More recently, in 2018 Venezuela reached a five digit hyperinflation ($\approx 65.000\%$), as well as an extensive economic deterioration in which the GDP decreased by two digits yearly since 2016 and, in 2019 reached an all-time low of -34% (IMF, 2020). An independent survey from three universities measured that in 2019 96.2% of all Venezuelans were poor, and 79.3% were extremely poor (IIES, 2020).

Therefore, the reasons of emigration from the country are several, and include, the political crisis and instability, the lack of a private market and economic opportunities, the inexistent market value of the Bolívar currency, and the common food supply shortages (due to price controls and trade restrictions). Is in this context that the Venezuelan exodus is occurring, both with voluntary and involuntary immigration.

3.2 Regulatory framework for Venezuelans

In terms of work permits, before 2018 Venezuelans needed a special permit granted by a work visa. This visa had a sponsor company and allowed temporary residence. Other work visas were granted if a sufficiently large investment in Colombia was made. One could argue that the regulation implied a higher informal employment rate of Venezuelans than the Colombian counterpart given the difficulties of getting a work visa. However, before 2015 Venezuelan informality rates were similar, an even smaller depending on the informality definition, than the Colombian ones 1b. From 2015 onwards, the picture changed and more Venezuelans workers than native ones were informally employed.

In the second half of 2018, the Colombian administration implemented a change in the work regulation of Venezuelans, providing a new framework to create what was called a Special Permit of Permanence (PEP, by its acronym in Spanish).[5] Aimed at fostering legal and more accessible employment for Venezuelans without the need of sponsor companies or investments, the PEP was initially valid for 90 days and could be renewed for up to two years. This policy was the largest migratory amnesty program offered to undocumented migrants in recent history. A short-term study of this policy indicates insignificant effects on several labor market outcomes, such as monthly

[5] In July 2018, the salient president of Colombia Juan Manuel Santos unexpectedly announced the creation of the special permit to work for all the Venezuelans that were registered in RAMV.

wages, unemployment or participation in the labor market for natives (Bahar, Rozo, and Ibáñez, 2020).[6] In this sense, the findings of this paper are not confounded by the possible short-run effect of the amnesty policy. Compared to Ecuador, Olivieri et al. (2020) find using simulations that the provision of work permits to Venezuelan workers would increase their average earnings.

4 Data

I use two main datasets in this paper. The first one is the the Labour Force Survey of Colombia (GEIH, by its acronym in Spanish) and the second one is the census of Population and Housing done in Colombia between January and October of 2018 (CNPV, by its acronym in Spanish). GEIH is a cross sectional monthly survey that characterizes the main outcomes of the Colombian Labor Market. It covers approximately 232.000 households per year, and is the survey with the most detailed sample coverage in Colombia. Both datasets are administered by the National Statistics Office of Colombia (DANE, by its acronym in Spanish), and are available on their official webpage.

To begin with, DANE implemented a migration module in the GEIH of 2012, then in 2013-II improved the questionnaire by adding questions on place of birth and, finally, after 2015 DANE removed an initial filter question on residence. In effect, at the beginning, the module was only answered by the people that were born in other cities different from the one they are currently living, but from 2016 onwards it was answered by all the respondents of the survey. This module contains questions on where the person was born, where the person lived 12 and 60 months ago and reasons of migration. With this information, I can identify immigration status in the short and long-term using a representative national survey. In my study I use this data from 2013 to 2019. In addition, as the census applies the same migration module, it allows me to completely characterize the native and migrant population in the country, reducing the measurement error of immigrants that can arise in standard surveys or even in the US census (Aydemir and Borjas, 2011; Amior, 2020).

Supplementary databases are used to construct external instruments. The first one is the Administrative Record of Venezuelan Migrants (RAMV, by its acronym in Spanish) that characterizes

[6] Potential explanations that authors argument are several. The first one is that the main target of the program, from a migrant perspective, was to have access to public services, which include health and education, and not to switch jobs, since migrants can perceive no real benefit of switching from informally to formally employed. The second explanation is the impossibility by migrants of getting offers to be formally employed.

the entire population of undocumented Venezuelans in Colombia. Nearly 443,000 individual records were gathered from April 6 to June 8 in 2018 at different frontier points in all the territory. It was an optional and *go to the registration point* kind of survey for undocumented Venezuelans. I take the information from *which* state in Venezuela immigrants are coming to build my distance instrument. The detailed information on origin is an improvement with respect to the distance instrument in Caruso, Canon, and Mueller (2019) that makes use of demographic information from the last census in Venezuela to predict origin of immigrants.

Is worth noting that Colombia and Venezuela have long-lasting relationships of trade with common interactions of businesses and people around the frontier. The main bridges that connect the two countries are three: Simón Bolívar International Bridge (in Norte de Santander), Paraguachón International Bridge (in La Guajira) and Páez Bridge (in Arauca). According to RAMV microdata, more than 2/3 of the Venezuelans in Colombia, until 2018, entered through Paraguachón and Simón Bolívar International Bridge.

4.1 Descriptive Statistics for Natives and Migrants

I differentiate three main groups of interest. The first one is of native Colombians residing permanently in Colombia; throughout the paper I focus on this group of Colombians, who did not migrate in the previous year from Venezuela, for the causal inference. The second one is of Venezuelans who emigrated to Colombia in the last year. Finally, the third group corresponds to Colombians who resided in Venezuela and then returned back to Colombia when the crisis started. In the Appendix (A.1) I present a Table with some descriptive statistics regarding the age profile, level of education and gender composition for the different groups, according to the different years of arrival.

Several stylized facts stand out. First, Venezuelan immigrants arriving to Colombia tend to be young, though their age seems to be steadily growing: prior 2017 the highest share of arrivals was in the range of 0-14 years, and after 2017 it is in the range of 14-28 years. Second, the returning Colombians are more concentrated in older ages: before 2016 the majority was in the range of 15-28 years, whereas after 2016 the predominant range was of 41-64 years. Third, in terms of education levels (taking into account that education rates in the country are low), the three groups have the highest share of individuals in the group with no high school degree. In particular, returning Colombians are the ones with the lowest share of tertiary education, while Venezuelans and Colombians have similar shares of education, and likely, of skills. One relevant takeaway is that arrivals of Venezuelans seems to be more educated in the latest years. Finally, in terms of gender composition there is no unbalance, with the shares of both men and women being similar.

4.2 Labor Market Structure

The structure of the Colombian labor market is embodied by the interdependence of two main categories of employment. The first category, normally occurring at small firms, is one without binding minimum wages and no access to social security (i.e., pension and health system), leading to what is commonly called the "informal sector".[7] The second category is characterized by the existence of a binding minimum wage and access to social security, leading to what is called the "formal sector". A clarification note, not all the workers in the formal (informal) sector are high (low) skilled, there is a combination of both types of skills in each sector, in which the majority of workers of the formal sector are skilled and of the informal sector are unskilled. To have a more general picture of informality, the distinction between two margins is needed. According to Ulyssea (2018) there is a *extensive* margin that represent firms which do not register formally to avoid paying taxes or regulatory costs, and the *intensive* margin that corresponds to formal firms who hire workers "off the books" to avoid complying with the contributions to the social security system.

In this paper, I use two definitions of labor informality. The first one is the national definition based on firm size, in which the firms with less than or equal to five workers, including employer and/or partner, unpaid family workers, domestic workers, day laborers and self-employed workers, are informal.[8] Note that government workers are always formally employed and workers without payment are informally employed, irrespective of firm size. The second definition of informal employment is based whether on the worker has access to health benefits or to a pension, a more general and comparable definition with other countries. Figure 1a and 1b display the density of wages according to the two definitions stated above. As can be observed, there is a bunching around a minimum wage in the formal sector, while not in the informal sector. Thus a large portion of

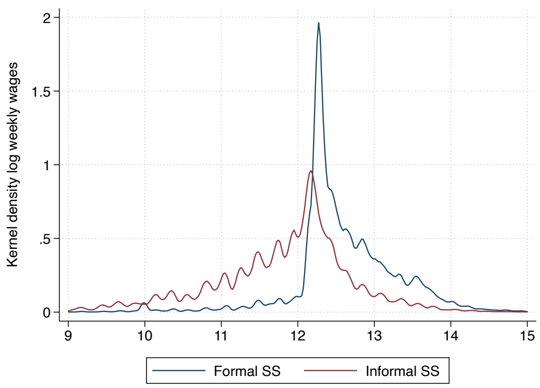
[7]In Colombia, not necessarily firms that employ on the informal sector are illegal (i.e., not paying taxes) or not formally registered in the state agencies. There are crossing definitions of informality depending on the side you focus (firm or worker). Throughout this paper I focus on the side of the worker.

[8]The definition excludes independent professional workers and the owners of the firm that employs 5 workers or less.

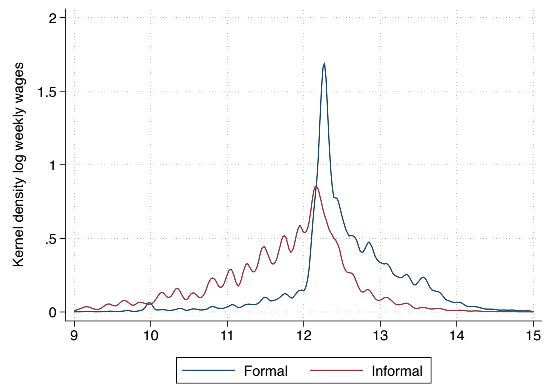
workers in the formal sector poses a binding restriction.

Figure 1: **Density of wages for the two sectors of employment**

(a) Social Security Definition



(b) National Definition



Note: All the data on wages is stacked across periods and departments and then is plotted. Sample is restricted to ages between 18 and 64 years old. Log weekly wages are in real terms using monthly CPI from DANE. No sampling weights are used. Kernel function is epanechnikov. Optimal band-width is used. Source: GEIH 2013 to 2019.

Table 1a and 1b present labor force statistics for Colombians and Venezuelans, for all the sample years. First, the majority of Colombian workers are informally employed, independent of the definition used. However there is a downward trend in the proportion of workers that belong to that sector in the last years with the informal rate going down from 58.2% in 2013 to 51.7% in 2019, using the definition of affiliation to social security. The opposite occurred to Venezuelan workers, where the same rate went from 62.1% to 89.1% in the same time period. The raw data also indicates that almost all the new arrivals of Venezuelans are being employed on the informal sector.

Second, comparing both Venezuelans and Colombians, in 2019 there is a higher labor force participation rate for Venezuelans (81.8% vs. 74%), a higher employment rate (69.4% vs. 65.4%) and a higher unemployment rate (15.1% vs. 11.6%), and that happens in all the years after the base period (see Table 1a and 1b). A higher employability of migrants could be associated with lower reservation wages compared to natives and a more inelastic labor supply (Borjas, 2017a).

Table 1: **Labor force statistics of Colombians and Venezuelans**

(a) Colombians (in rates)

	Labor force participation	Employment	Unemployment	Informal	Informal Social Security	N (15-64)	Population
2013	75.3	67.6	10.2	53.9	58.2	356,597	23,336,824
2014	75.4	67.7	10.2	52.7	56.4	473,671	23,654,284
2015	75.6	68.1	9.9	52.6	55.5	474,871	24,022,452
2016	75.3	67.4	10.5	52.1	54.3	470,835	24,463,513
2017	75.0	67.0	10.7	52.1	53.6	462,484	24,516,791
2018	74.5	66.3	11.0	51.3	52.8	455,238	24,435,938
2019	74.1	65.5	11.6	50.1	51.9	444,442	24,221,821

(b) Venezuelans (in rates)

	Labor force participation	Employment	Unemployment	Informal	Informal Social Security	N (15-64)	Population
2013	80.5	68.8	14.5	49.6	62.1	583	36,364
2014	78.9	68.4	13.4	48.8	59.6	839	43,772
2015	73.6	65.8	10.7	51.2	65.5	959	50,802
2016	77.4	65.8	15.0	58.4	71.8	1,742	94,291
2017	81.5	68.6	15.9	63.3	82.0	4,112	206,427
2018	84.4	71.4	15.4	70.7	88.6	10,751	625,390
2019	81.8	69.4	15.1	69.3	89.1	18,440	1,114,666

Note: The rates are calculated using national sampling weights from GEIH. The sample is restricted to population from ages between 15 and 64 years in urban areas. In (a) are restricted to natives living for more than one year in Colombia. The rate of informal employment is calculated as the proportion of workers that are informally employed, according to both definitions stated on the paper, over total employment. Source: GEIH, 2013-II to 2019.

Table 2 shows in which industries Venezuelans and Colombians workers are occupied. Note that immigrants are overrepresented with respect to natives in two industries. The first one is the *Commerce, hotels and restaurants* industry, where almost half of all the Venezuelans workers have a job (46.9%), while the corresponding share for Colombians workers is nearly 1/3 ($\approx$ 30%). The second one is the *Construction* industry (11.1% vs 7%). Conversely, immigrants are underrepresented relative to natives in two main industries of employment: *Real estate, business and rental activities* (6% vs 9.5%) and *Community, social and personal services* (15.6% vs 23.5%).

The next step is to compute the observed wage gap of migrants and natives. To do that, I regress the log hourly real wage on a set of control variables including a dummy of birth place.[9] On average, Colombian workers earn 0.29 log points higher wages than its Venezuelan counterparts (Table 2). Some unobservables that can help to explain the gap between both groups is the work experience of migrants at their home country. One key aspect of this immigration event is that

[9] The wage gap is calculated in an unweighted regression of log hourly real wages on the dummy of place of birth, plus two polynomials of age, schooling, gender, interaction of department and industry, and fixed effects of year and month. Restricted to workers between 18 and 64 years in urban areas, stacking all periods under analysis (2013 to 2019).

both groups share the same language, thus there can not be disadvantages in communication skills.

Table 2: **Distribution of workers by industry and place of birth**

Industry	Colombians	Venezuelans
Agriculture, livestock, hunting, forestry and fishing	3.6	1.4
Mining and quarrying	0.7	0.4
Manufacturing industry	13.5	12.3
Electricity, gas and water supply	0.6	0.2
Construction	7.0	11.1
Commerce, hotels and restaurants	30.1	46.9
Transport, storage and communications	9.6	5.6
Financial intermediation	1.9	0.6
Real estate, business and rental activities	9.5	6.0
Community, social and personal services	23.5	15.6
N (Workers, 18-64)	1,979,144	24,706
Wage gap (Colombians vs Venezuelans)	-0.288	
Standard Error	(0.0198)	

Note: Shares are calculated using national sampling weights from GEIH. Shares across columns should sum up to 100% adding the unknown occupation share. The sample is restricted to all Colombians and Venezuelans from ages between 18 and 64 years in urban areas. The sample aggregates all the periods of analysis from 2013-II to 2019. Colombians are restricted to residing permanently in Colombia. The wage gap is calculated in an unweighted regression of log hourly real wages on the dummy of place of birth, plus two polynomials of age, schooling, gender, interaction of department and industry, and fixed effects of year and month. Source: GEIH, 2013-II to 2019.

4.3 Base period of Analysis

To select the base period for the event-study, first I exploit the monthly information of the amount of Venezuelans in Colombia from GEIH survey, and, second I use the following timeline of the immigration event.[10] In August of 2015 the Venezuelan government, for different reasons, unilaterally closed the national border restricting the exits from their country. As a result, the number of Venezuelans in Colombia remained similar to previous months. A year after, in August of 2016, the Venezuelan government decided to re-open again the border, and, due to the conditions I explained above, there was an increase in the number of immigrants in Colombia, that grew rapidly as the political and economic crisis widened in Venezuela over the succeeding months (see Figure 2a). For simplicity, and to remove seasonal effects, I select as a base period of comparison the year of 2015, which was the last year before the huge increase of Venezuelans in Colombia. In this respect, it

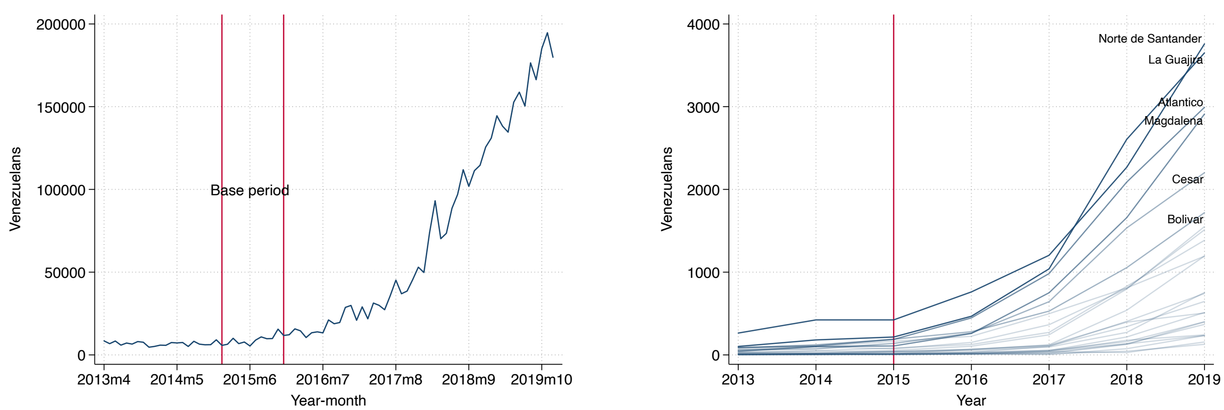
[10]The GEIH survey question asks the respondents where they were born, if in the same municipality they are residing, or in others from Colombia. If it is from another country, they ask in which Country.

should be stressed that the number of immigrants does not seem to have reached a peak yet.[11] Hence, adding new information to the analysis is important. For instance, Caruso, Canon, and Mueller (2019) could only analyze the effect with information up to 2017, but immigration arrivals doubled in 2018 and, again, in 2019.

Next, I analyze the number of Venezuelans by departments -an administrative division in Colombia similar to states in the US- which is the treatment unit for the event-study. Again, there is small number of Venezuelans in the pre-treatment years, followed by a staggered increase which varies greatly in the 24 departments sample (Figure 2b), motivating the spatial approach undertaken here. Note that, using the two-way fixed effects model (i.e., $y_{it} = \alpha_i + \alpha_t + \beta^{DiD} D_{it} + e_{it}$), in the presence of differential timing of the policy might yield biased estimates (Goodman-Bacon, 2018), it is preferable, given the variation in my setup, to study the effect year-by-year in an event-study with treatment intensity.

Figure 2: **Venezuelans in Colombia at the national and department level**

(a) National monthly (b) Department yearly



Note: National sampling weights of GEIH are used in (a), in (b) no weights are used. The base period of the event-study is 2015. In (a) the cumulative amount of Venezuelans in each year is given by the sum of Venezuelans per month in that year. Source: GEIH, 2013-II to 2019.

[11]Nevertheless with the COVID-19 pandemic significant number of Venezuelans are returning back to their home country (Reuters, 2020).

5 Empirical Specification

5.1 Event Study with fixed Migration rate

As already mentioned, my empirical strategy relies on exploiting the intensity and heterogeneity of Venezuelan immigration in the different areas of Colombia. To do so, I focus on the aggregate level of departments to analyze the effect that immigration had. The department unit has a local labor market in its capital city that is interlinked with the surrounding smaller cities, having therefore some degree of independence from the labor markets in other departments. As explained in the literature review, I follow a spatial approach, which is arguably the most common and oldest method used in the migration literature. Basically it consists on the comparison of groups, within a defined area, after and before immigration occurs.

One critique of this approach is that it might not reflect the true immigration effect if there is a mobility response of inputs, say of native workers or capital, from areas more affected by the immigrant supply shock to those less affected (Aydemir and Borjas, 2011). In my setting, I can test this hypothesis to show that neither there was a clear inflow effect of natives not coming to the most affected areas (a downward trend is noticeable but only significant in 2019), nor an outflow effect of natives leaving the most immigrant affected areas (point-estimate is positive in post-treatment years but significant until 2018), relative to the base period (see Figure 15 in the Appendix).

In terms of empirical specification, my preferred one has the interpretation of an event-study one as I select a base year of comparison (2015) and it perform differences between pre and post-treatment periods t, with respect to the base period, for the different departments d. Thus, I estimate the following regression, in which the omitted year is when $k = 0$,

$$Y_{dt} = \alpha + \gamma_d + \gamma_t + \sum_{t=2013}^{2019} \beta_t M_{d,2018} \mathbf{1}\{t - 2015 = k\} + u_{dt} \tag{1}$$

and $M_{d,2018}$ is a time-invariant treatment variable, constructed from the 2018 census records. Note that by construction $\beta_{2015} = 0$. It should be emphasized that, to motivate the event-study research design and to examine its validity, a constant migration rate is required. This is not problematic since the arrivals of migrants between departments remains constant over time, with nearly perfect correlation across different years, while total immigrant population is increasing, as shown previously in Figure 2b. Therefore, it makes sense to select the best measured data for this migration rate, which comes from the census. However, since using a fixed in time migration rate complicates the direct interpretation of the β_t coefficients, in the results I focus mainly on 2018 coefficients, also I present in the next subsection a time-varying migration rate. With this in mind, $M_{d,2018}$ is defined as follows

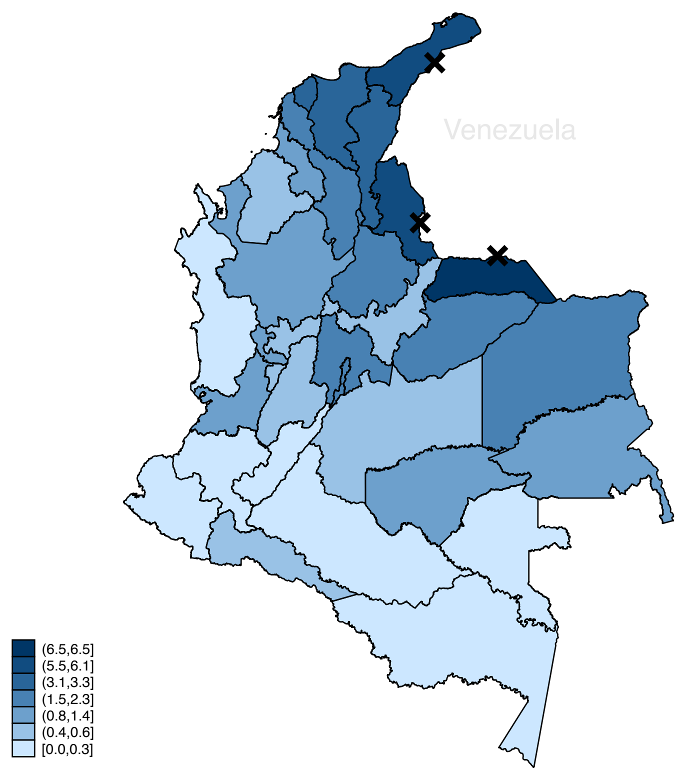
$$M_{d,2018} = \frac{L_{Ven,d,2018} - L_{Ven,d,2015}}{L_{Total,d,2018}} * 100 \tag{2}$$

where the numerator is total number of employed Venezuelans (between 18 and 64 years) in department d who arrived to Colombia in the previous 5 years, starting from 2018, minus total number of employed Venezuelans in d whose year of arrival was 2015 -recall that the census is a static picture that does not take into account movements across space-. Figure 3 depicts this variable. Finally, I add fixed effects of year γ_t and of department γ_d to the regression model. Notice that β_t captures the correlation between immigration and the outcome Y, for period t, and recall that data come from a repeated cross section, not from panel data.

Figure 3 plots the Colombian map along with the migration rate $M_{d,2018}$ by departments.[12] Not surprisingly, the highest migration rates are observed in those areas which are closer to the Venezuelan frontier, and specially to the main crossing bridges discussed above (see the **X** in the Map). Note that information about the outcomes, mostly from GEIH, is mainly available to 24 departments, not to all the 33 in the country. Yet the missing 9 departments, mostly located in the Amazonia and Orinoquia region, only account for approximately 3% of the total population.

[12]Because the census recollection ended on October of 2018 it does not take into account all the possible arrivals in November and December of 2018.

Figure 3: **Spatial distribution of Venezuelans by departments**



Note: To characterize recent Venezuelan migrants, the census asked if the person lived in the last 12 months in Venezuela. Only Venezuelan-born migrants are taken into account in the numerator of the rate. The **X** represent the three main crossing bridges discussed in the Data Section. Source: CNPV 2018.

In this setting, I use clustered standard errors as observations between treated units are independent. One possible downturn of this approach is the small number of units $N = 24$, which increase Type I error rate considerably (Pustejovsky and Tipton, 2018), meaning that I over-reject the null hypothesis when it is true. For that reason, I also implement a wild bootstrap test in the Robustness Checks to correct for the possible increase of the Type I error (Roodman et al., 2019).

5.2 Event Study with time-varying Migration rate

Given the fact that I have yearly information on migration (not just the census one), I can also use a migration rate that varies on post-treatment years from the GEIH survey as an additional explanatory variable, in order to interpret more easily the coefficients for specific years. Notice that a potential caveat of this approach is the low variability of migration before 2017, implying that the changes with respect to the base period are small in those years, which could increase the standard errors.

Considering this, equation (1) can be rewritten as a multiple linear regression changing $M_{d,2018}$ with M_{dt}

$$Y_{dt} - Y_{d,2015} = \delta_t + \beta_t M_{dt} + e_{dt} \tag{3}$$

where e_{dt} can be viewed as $u_{dt} - u_{d,2015}$, and the definition of the time-varying migration rate is

$$M_{dt} = \frac{L_{Ven,d,t} - L_{Ven,d,2015}}{L_{Total,d,2015}} * 100 \tag{4}$$

such that the numerator measures the employed Venezuelans (between 18 and 64 years) in d for all the post-years t from GEIH survey, relative to the base period (2015), while the denominator is fixed on the base year, following the definition in Card and Peri (2016). For instance, if the outcome of interest is $logwages_{dt}$, then the regression analysis yields

$$\hat{\beta}_t = \frac{\hat{Cov}(M_{dt}, \Delta logwages_{dt})}{\hat{Var}(M_{dt})} \tag{5}$$

where $logwages_{dt} - logwages_{d,2015} = \Delta logwages_{dt}$.[13] Then, plugging model (3) in the last expression yields

$$\hat{\beta}_t = \beta_t + \frac{\hat{Cov}(M_{dt}, e_{dt})}{\hat{Var}(M_{dt})} \tag{6}$$

Thus, even if we remove the time-invariant heterogeneity γ_d, a bias can still emerge if migration is driven by unobservables in the departments (i.e., $E[M_{dt}e_{dt}] \neq 0$) that change over time, for instance it could be the case that economic opportunities, relative to base period, could be correlated with migration rates.

[13]The resulting expression of $\hat{\beta}_t$ can be explained as follows, the numerator measures the covariance between the inflow of Venezuelans and the change in wages with respect to base period, while the denominator weights this covariance with the observed dispersion of migration.

5.3 Shift-Share Instrumental Variable

If migrants self-select into areas where the economic outcomes are better, the migration rate $M_{d,2018}$ would become endogenous in the previous empirical specification. To estimate causally the effect of immigration, I instrument migration rate with two available sources of exogeneity: (i) distance between capital cities in the two neighboring countries and (ii) past-settlements of Venezuelans in Colombia. These two shift-share instruments have been used previously in the migration literature.

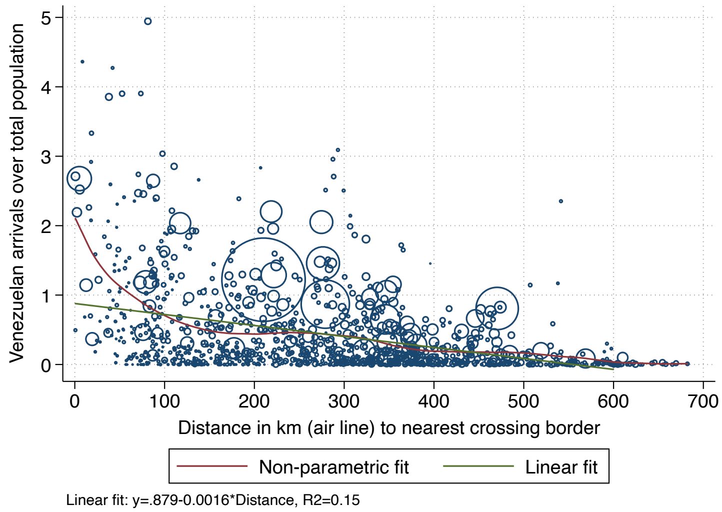
First, the construction of a distance instrument is based on Del Carpio and Wagner (2015) and Caruso, Canon, and Mueller (2019),

$$z_{1,dt} = \sum_s \underbrace{(\theta_s/T_{s,d})}_{share} * \overbrace{M_t}^{shift} \tag{7}$$

where $T_{s,d}$ is the road distance in kilometers from capital city in state s in Venezuela to capital city in department d in Colombia computed with the algorithm in Weber and Péclat (2017), and θ_s is the share of Venezuelans that emigrate from s according to RAMV.

The use of the distance instrument z_1 is motivated from the fact that Colombia and Venezuela share more than 2.000 kms of terrestrial borders. In practice, with the census immigration flows can be measured at a more granular level, I find a strong positive relationship between immigrant arrivals in municipalities and distance to nearest frontier (see Figure 4). Therefore, new arrivals M_t to location d are determined by the travel distance from city x to city y, in the sense that travel distance poses a time and economic restriction to new immigrants. A threat to this identification strategy arises if the border states suffer more, in terms of economic shocks such as trade, than the counterpart far-located states. For that reason, I show that when including a control for changes in the business cycle in departments (with departmental GDP from DANE), with the caution that it can be a "bad control" (Angrist and Pischke, 2008), results for wages are slightly higher but remain significant (Figure 19). In addition, I show that when including trade patterns with Venezuela (measured as share of total exports in USD to Venezuela over total exports to the world from DANE in 2015) results are again higher and significant. Thus the inclusion of previous controls do not alter drastically the coefficient of interest.

Figure 4: **Arrivals in preceding year from Venezuela by municipalities vs distance to frontier**



Note: Municipalities are $N = 1117$. To characterize recent Venezuelan migrants, census question asked if the person lived in the last 12 months in Venezuela. Only Venezuelan-born migrants are taken into account in the numerator of the rate. Municipalities are weighted with native population according to the census. Kernel regression is used for the non-parametric fit. Source: CNPV-2018.

Second, the construction of a past-settlement instrument is based on Altonji and Card (1991) and Card (2001),

$$z_{2,dt} = \underbrace{\frac{1}{L_{d,t-1}} * \frac{Ven_{d,2005}}{Ven_{2005}}}_{share} * \overbrace{M_t}^{shift} \tag{8}$$

where the second term is the share of Venezuelans in every department d in Colombia (according to the 2005 population census), normalized by the working age population $L_{d,t-1}$ in d at year $t-1$, whereas M_t are arrivals of Venezuelans in t measured by GEIH. Both instruments vary on time and location.

The validity of the past-settlement instrument z_2 relies on the fact that new arrivals M_t to department d are attracted by the network effects in that location, while current economic trends in d are unlikely to be systematically related to lagged immigration shares (if those shares are lagged sufficiently). If this holds, then the instrument is valid, in the sense that lagged immigrant location is related to new arrivals (relevance) but not related to current economic conditions (exogeneity). However, this assumption might fail if local economic trends are highly serially correlated, such that labor demand shifts that attracted immigrants in the past are still correlated with contemporaneous demand shifts. This issue can be reduced by selecting sufficiently lagged shares, that goes back until 1973, to show that when using more distant in time shares, that, supposedly, are less correlated with current economic conditions, the results are not significantly altered.

Next, concerns with shift-share instruments are important, for instance in dynamic settings it tends to be serially correlated, as the share is constant (Jaeger, Ruist, and Stuhler, 2018). In my study because arrivals surge rapidly after 2016, it is possible to break the serial correlation and have a valid instrument. Lastly, notice a discussion about the exclusion restriction of Bartik-type instruments that lies between the exogenous shares (Goldsmith-Pinkham, Sorkin, and Swift, 2018) or the exogeneity of the aggregate-level shift (Borusyak, Hull, and Jaravel, 2018), in my case, I can argue that the selected shares are exogenous to have a valid instrument.[14]

Empirically, I use both instruments separately to predict $\hat{M}_{d,2018}$ from the census in a first-stage regression, and then run again regression (3) using the predicted migration rate. Therefore, the first stage regression for both instruments is the following,

$$M_{d,2018} = \varphi + \eta z_{i,dt} + \upsilon_d \quad i = 1, 2 \tag{9}$$

where υ_d is capturing the endogenous component of $M_{d,2018}$. The results of this stage are presented in Table 3, the distance instrument explains 88.4% of the variation of the migration rate, while the past-settlement instrument explains 48.8%.

[14]As a preliminary, to evaluate exogeneity of instruments, I use a Sargan-Hansen J test for over-identifying restrictions, where I do not reject the null hypothesis ($p - value = 0.315$). The IV-regression is for 2018 predicting $M_{d,2018}$ using both instruments simultaneously.

Table 3: **First Stage: The Inflow of Venezuelans and the two instruments**

	(1) $M_{d,2018}$	(2) $M_{d,2018}$
Distance (z_1)	0.00376***	
	(0.000299)	
Past-settlement (z_2)		34.75***
		(5.657)
Constant	-1.271***	0.751**
	(0.233)	(0.231)
N	24	24
R^2	0.884	0.488
F st	157.9	37.73

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The Table reports the coefficients of the first stage regression of the instruments with the migration rate $M_{d,2018}$. Since the migration rate $M_{d,2018}$ from the census and the shares of the two instruments are time-invariant, the first stage is the same in all the years analyzed.

6 Results

6.1 Wage Responses

One of the main advantages of the event-study design is the possibility to test for previous trends in the outcome (pre-trends) and, eventually, control for them, if they exist. This is important as departments can exhibit differing local economic tendencies before the immigration event occurred, contaminating the true effect of immigration.

With this in mind, I first regress equation (3) for real log hourly wages of natives, under two methods (OLS and IV) with the explanatory variable $M_{d,2018}$, an important finding is that pre-trends are not significant.[15] The influx of Venezuelan immigrants, as standard economic theory predicts, had a negative effect on native wages for both methods (Figure 5). The OLS estimates follow a negative pattern, arguably the estimates are upward biased by omitted variables, but not so different from the IV ones due to the high R^2 of the First-Stage (see Table 3). Moreover, when using the two instruments defined before separately, the results are again negative and significant. For 2018, the year I defined the migration rate from the census, a 1 pp increase in the share of employed

[15]The construction of the wage variable requires some steps to construct it, in the Appendix I explain in detail what are the steps required to have a more precise measure of wages.

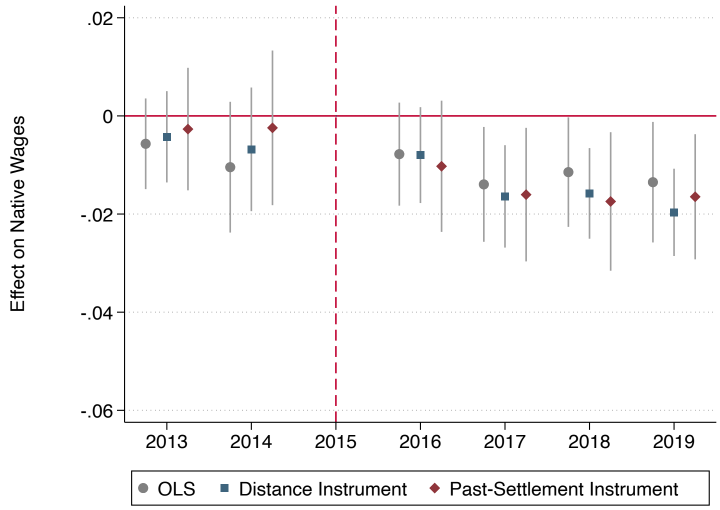
Venezuelans decreases the wages of natives by 1.7%, with the past-settlement instrument, and by 1.6%, with the distance instrument.[16] Later on I show that when using residual wages, retrieved from individual characteristics as age, years of schooling and gender, instead of observed wages, previous results are similar.

Scaling up these estimates, the total shock according to census is about 1.7 pp of the employed population (in absolute numbers $\approx$ 254,000 employed Venezuelans), and hence the total impact on wages, for 2018, is between 2.7%-3% depending on the instrument selected.[17] To interpret the wage response, the average native local wage increase by nearly 2.1% in real terms (comparing 2018 vs 2015), thus the immigration negative effect imply a decline in natives' real wages. It should be, however, noticed that the shock can be understated (and the effect overstated) because the census ended in October of 2018, omitting the arrivals of Venezuelans in November-December of that year, and also because the measurement of the shock only considers employed Venezuelans but not Colombians returning from Venezuela too.

[16] Caruso, Canon, and Mueller (2019) find a point-estimate of the Venezuelan immigration on hourly wages of 7.6% for a 1 pp increase in their migration rate using IV panel-data regression. The differences with my estimates mainly arise by the specification used and the difference in the period analyzed (they only have data until 2017). When adding the information of 2018 to their empirical specification the coefficients are almost halved.

[17] The relevance condition of the instrument is measured through the F-statistic, a rule-of-thumb for a good instrument is a F-statistic higher than 10. However, recently Lee et al. (2020) argued for a higher number (104.7). In this case, the distance instrument statistic is 157.9 and the past-settlement instrument is 37.73.

Figure 5: **Event Study estimates on log hourly wages of Colombians**



Note: Dependent variable is log hourly wages relative to base period. Departments in the regression are $N = 24$ per year. The Confidence Intervals correspond to $\alpha = 0.95$. Explanatory variable is $M_{d,2018}$. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. F-statistic for Distance Instrument is 157.9 and for Past-Settlement Instrument is 37.73. Hourly wages are in real terms using the monthly CPI from DANE.

Second, I use the varying on time migration rate M_{dt} built with GEIH survey as explanatory variable. Interestingly, estimates follow a similar pattern (negative and significant) but differently in magnitude from those obtained with the fixed $M_{d,2018}$ from census. For instance, in 2017 I observe a much higher negative effect, with wider confidence intervals probably due to the low variability of migration between 2017-2015. By contrast, in 2018, my reference year, the estimates of the effect are similar as before, ranging between 1.4%-1.6% (see Figure 16a in the Appendix).

The insight for such a high negative finding relies on several factors pointing to a high substitutability of natives and migrants. In effect, migrants speak the same language than natives, overcoming communication skills problems; they share cultural traits, which can reduce wage discrimination; the majority come as forced migrants (which implies a relatively low reservation wage); and finally wage flexibility in the informal sector can lead to large wage cuts when migrants arrive.

6.2 Employment Responses

When analyzing the employment effects of immigration, the first result to highlight is that, opposite to wages, there are significant differences in the employment trends before the migration event happened (Figure 6a). This indicates that distance to Venezuela and historical enclaves of migrants would predict local native employment in the pre-policy period, suggesting a violation of the PTA with IV required for the identification of the causal parameter. If instruments predict employment trends before the migration crisis, but not wages, could be in part for differences in amenities or housing prices in the different areas. To address this problem, I explicitly control for the pre-trend in the regression for all the years to get the trend-adjusted estimates (the control is the change in log employment between 2015 relative to 2013). In Figure 6b are plotted the pre-treatment coefficients and pre-trends in 2014 are no longer significant for both instruments.

Table 4 shows the results for employment and wages for natives in 2018, relative to the base period (recall that, since wages did not present pre-trends, I did not use trend-adjusted estimates in previous section). For the interpretation of the employment impact, it is useful to keep in mind that a 1 pp increase in the migration rate reduces on average 1.5% local employed natives in 2018 relative to 2015 using past-settlement as instrument and by 1.1% using distance as instrument (coefficients of Table 4-Column 4).[18] However, I show in the Appendix that employment results are sensitive to the exclusion of survey weights in the estimation procedure, as estimates turn to be insignificant (see Figure 20b).

[18] In Figure 16b in the Appendix I show that when using M_{dt} as explanatory variable, instead of the fixed $M_{d,2018}$ from the census, results of native employment are similar, but with wider confidence intervals in 2017 due to the low variability of migration before 2018.

Table 4: **Wages and Employment estimates for Colombians, 2015-2018**

	(1) Wages	(2) Wages	(3) Employment	(4) Employment
$M_{d,2018}$ (OLS)	-0.0115*	-0.00962	-0.0155***	-0.0132***
	(0.00540)	(0.00595)	(0.00391)	(0.00334)
$M_{d,2018}$ (IV: Distance)	-0.0158***	-0.0146**	-0.0174***	-0.0147***
	(0.00427)	(0.00465)	(0.00399)	(0.00335)
F st	157.9	228.3	157.9	152.6
$M_{d,2018}$ (IV: Past-settlement)	-0.0174**	-0.0168**	-0.0170***	-0.0113**
	(0.00653)	(0.00614)	(0.00358)	(0.00434)
F st	37.73	59.05	37.73	19.89
Trend-adjusted	No	Yes	No	Yes
N	24	24	24	24

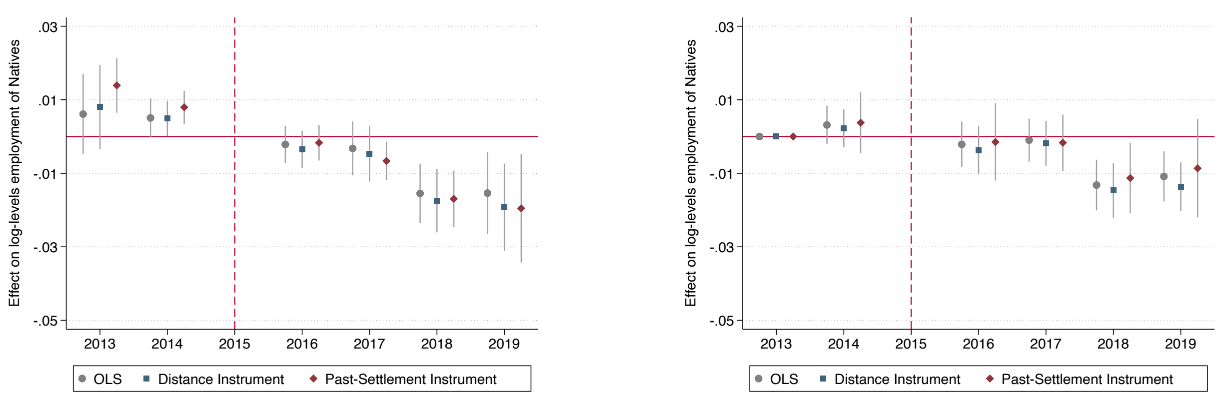
Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The Table reports the coefficients of the second stage regression of the instruments with the migration rate $M_{d,2018}$. The outcome is the difference in 2018 with the base period. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Standard errors are clustered at the department level. The variables are in logarithms, thus the coefficients are interpreted as percentages. Department sampling weights from GEIH are used. Trend-adjustment estimates have as a control in the regression the growth in employment from 2013 to 2015. Hourly wages are in real terms using the monthly CPI from DANE.

Figure 6: **Event Study estimates on log employed Colombians**

(a) IV estimates

(b) IV estimates with trend-adjustment

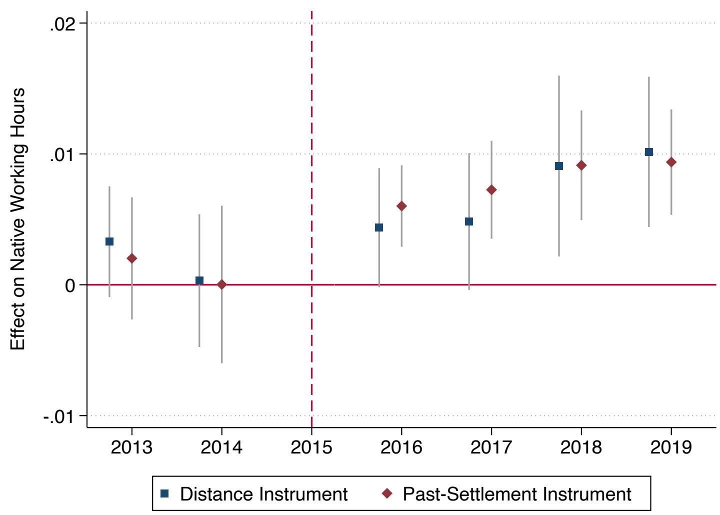


Note: Dependent variable is log native employment relative to base period. Departments in the regression are $N = 24$ per year. The Confidence Intervals correspond to $\alpha = 0.95$. The explanatory variable is $M_{d,2018}$. Trend-adjustment estimates have as a control the growth in employment from 2013 to 2015. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients.

6.3 Intensive Margin Effect

Given that natives hourly wages on average are being reduced by the immigration event, a natural question to make is: ¿What is the decision of Colombians on working hours, in particular are they working more or having more leisure time? To answer previous questions, GEIH survey asks how many working hours usually the worker works in the previous week. Figure 7 shows that natives are working more hours in all the post-treatment years (an upward trend on the estimates is noticeable). In my reference period, 2018, the coefficient is around 0.9% for both instruments. Thus, if the migration event increased the employed migrant population in 2018 by 1.7 pp compared with 2015, the hours worked per week by Colombians went up by 1.5% in that year.

Figure 7: **Event Study estimates on log working hours of Colombians**

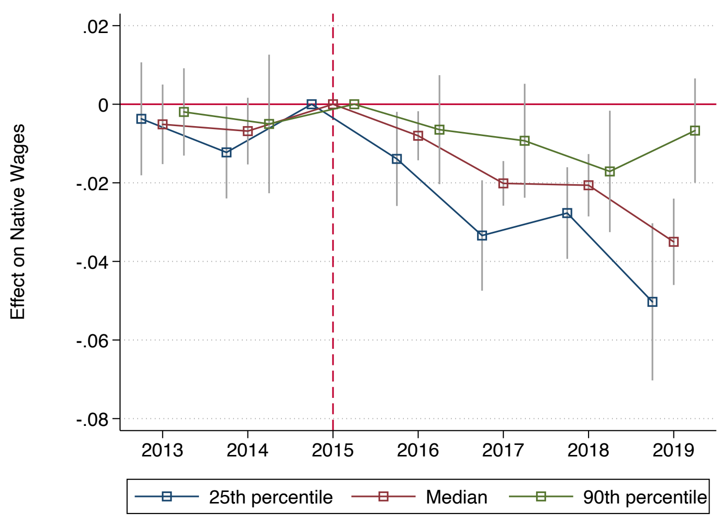


Note: Departments in the regression are $N = 24$ per year. The Confidence Intervals correspond to $\alpha = 0.95$. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. F-statistic for Distance Instrument is 157.9 and for Past-Settlement Instrument is 37.73.

6.4 Distributional Effects of Immigration

Immigrant arrivals can have differential impacts depending on the percentile selected across the distribution of wages. For instance, the effect on the mean can be different to the effect on lower or higher percentiles, as the impact is not homogeneous. By using data from GEIH, which provides the most detailed sample coverage in Colombia with information on wages, I can analyze what is happening at other points of the distribution (i.e., 25th percentile, the median or 90th percentile) to understand the heterogeneous effects of immigration across the local wage distribution, that is normally aggregated in the mean analysis. The results of this exercise, plotted in Figure 8 point to a differential effect depending on the percentile. In effect the native wages at the lower part of the distribution are the most affected by immigration, if workers are sorted by their local wages. Analyzing the median, which is a more robust estimate to outliers and censored data in my sample, yields a coefficient of -2.1% for a 1 pp increase in the migration rate $M_{d,2018}$, which is higher in absolute terms compared to the mean estimate (-1.6%) -using as IV the distance to Venezuela-. Comparing these results with the UK, Dustmann, Frattini, and Preston (2013) find that immigration depresses wages below the 20th percentile, while it contributes to wage growth above the 40th percentile. However, one should notice that, over their period of analysis, immigrants in the UK are much more educated than natives.

Figure 8: **Event Study estimates on log hourly wages of Colombians by percentiles**



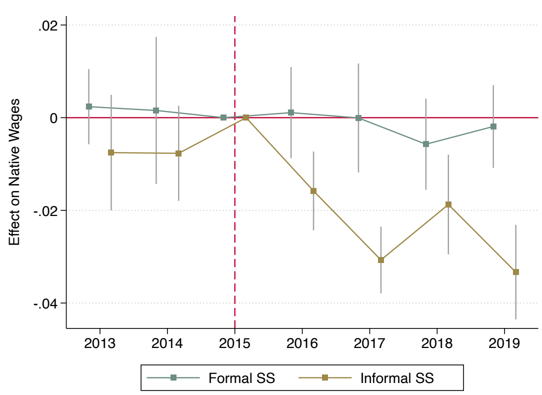
Note: Departments are $N = 24$ per year. The Confidence Intervals correspond to $\alpha = 0.95$. Instrument used is past-settlement. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. F-statistic for Past-settlement Instrument is 37.73. Hourly wages are in real terms using the monthly CPI from DANE.

6.5 Heterogeneous Effects

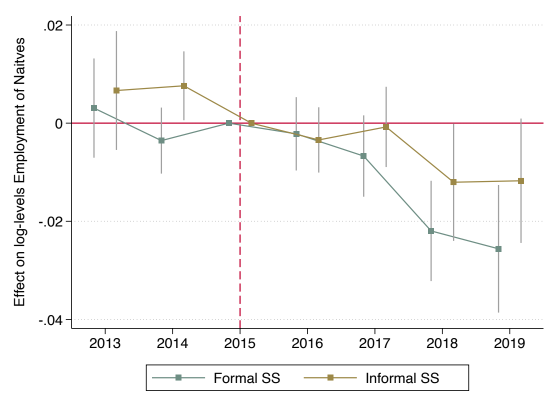
Next, as explained above, the majority of Colombians are informally employed with no binding minimum wages or written contracts that can protect their wages. Arguably, informal workers are in theory the most vulnerable group in the presence of a supply shock. Figure 9a plots the estimated effect of immigration for formally and informally employed native wages.[19] Not surprisingly, informally employed workers suffer the largest wage losses, while formally employed ones are unaffected.[20] Compared to Turkey, a country with high informal levels of employment like Colombia, Del Carpio and Wagner (2015) find that the inflow of Syrian refugees increased formal wages of natives, by occupational upgrading, while they have an insignificant effect on informally employed workers wages. With regard to the effect on employment, formal and informal workers suffer a negative effect after 2017, yet the impact is more severe on formal employment. In the next section I develop a simple model to explain the findings on wages and employment for these two markets.

Figure 9: **Event Study estimates by affiliation to Social Security**

(a) IV (Distance) estimates on Native Wages



(b) IV (Distance) estimates on Native Employment



Note: Dependent variable is log native employment relative to base period. Departments in the regression are $N = 24$ per year. The Confidence Intervals correspond to $\alpha = 0.95$. Sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. I use the definition of affiliation to Social Security (SS). In (b) no controls for pre-trends are used. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. F-statistic for Distance Instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

In a next step, I study heterogeneous effects by level of education. Noting that because instruments did not predict local trajectories of wages in the pre-treatment period, while for employment it does, I use trend-adjusted estimates for employment not for wages, with the caveat that trend-adjusted estimates for wages differ little (as shown in Table 4). First, I find similar results for

[19] I use the definition of informality according to the affiliation to social security (using the national definition of informality yields closely the same results).

[20] Similar to the findings in Caruso, Canon, and Mueller (2019) where authors find an insignificant effect on wages on the formal sector and a negative effect on wages on the informal sector in Colombia.

wages and employment as predicted by the standard factor proportions model when migrants are mostly unskilled. Most affected native workers in terms of wage and employment impacts are the unskilled (with high school or less), with significant negative effects for both instruments. While native skilled workers (with more than high school) seems to be less affected by the immigration event (see Panel A Table 5). Second, in terms of gender impacts, males present the highest reduction in wages, in contrast to females where I do not find a significant effect. The opposite to what happens for employment, where females present the highest reduction. Take into account that systematically Men and Women migrants are arriving similarly in magnitude to Colombia.

Then, I study the effect by ages, aggregated in three groups: (*i*) a younger group between 18 and 25 years old, (*ii*) a medium-aged group between 26 and 54 years old, and (*iii*) an older group between 55 and 64 years old. The main finding is that the medium group of workers appear to be the most affected in term of wages, suffering a significant negative effect with both instruments, for the older group I find insignificant effects. In terms of employment responses, most affected native group is the younger one, while the medium and older one present much smaller estimates (see Panel C Table 5).

Table 5: **Native Wages and Employment estimates by subpopulations, 2015-2018**

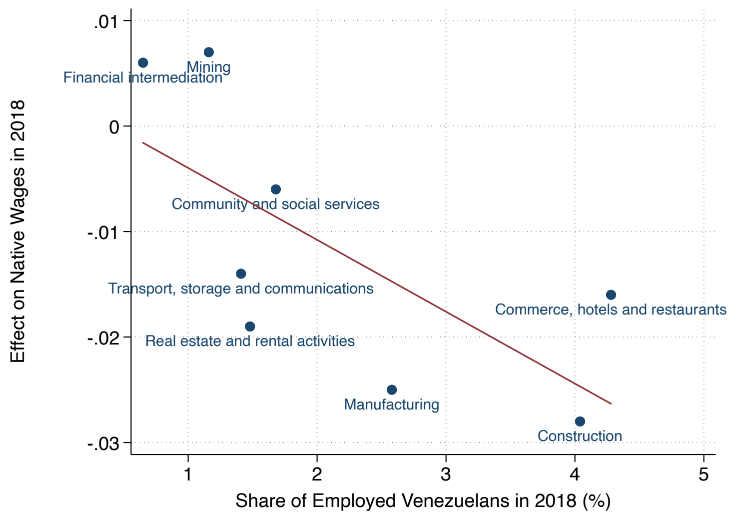
	(1) Wages	(2) Employment
Panel A: Skill-group		
High school or less (Distance)	-0.0205***	-0.0177***
	(0.00469)	(0.00313)
High school or less (Past-settlement)	-0.0186***	-0.0113***
	(0.00477)	(0.00333)
More than high school (Distance)	-0.00915	-0.0108
	(0.00651)	(0.00771)
More than high school (Past-settlement)	-0.00688	-0.0150
	(0.00917)	(0.00871)
Panel B: Gender		
Male (Distance)	-0.0194**	-0.0136*
	(0.00645)	(0.00540)
Male (Past-settlement)	-0.0260***	-0.0132*
	(0.00789)	(0.00638)
Female (Distance)	-0.0114	-0.0188***
	(0.00607)	(0.00328)
Female (Past-settlement)	-0.00585	-0.0152***
	(0.00765)	(0.00332)
Panel C: Age-group		
18-25 Years (Distance)	-0.0101*	-0.0302*
	(0.00506)	(0.0152)
18-25 Years (Past-settlement)	-0.00959	-0.0410*
	(0.00536)	(0.0166)
26-54 Years (Distance)	-0.0189***	-0.0136***
	(0.00465)	(0.00274)
29-54 Years (Past-settlement)	-0.0210**	-0.00978*
	(0.00673)	(0.00437)
55-64 Years (Distance)	-0.00858	-0.0121*
	(0.00843)	(0.00572)
55-64 Years (Past-settlement)	-0.0133	-0.00422
	(0.00929)	(0.00838)
Trend-adjusted	No	Yes
N	24	24

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The table reports the coefficients of the second stage regression of the instruments with the migration rate $M_{d,2018}$. The outcome is the difference in 2018 with the base period. Standard errors are clustered at the department level. The variables are in logarithms, thus the coefficients are interpreted as percentages. Department sampling weights from GEIH are used. Trend-adjusted estimates are controlled only by the growth in employment from 2015 compared to 2013 for each subpopulation. F-statistic for Distance Instrument is 157.9 and for Past-Settlement Instrument is 37.73. Hourly wages are in real terms using the monthly CPI from DANE.

Finally, I estimate separately the effect of the Venezuelan immigration for eight big branches of economic activities or industries for 2018. In Figure 10 I plot these effects against the share of employed Venezuelans in those industries. Importantly the industries with larger shares of employed Venezuelans experienced the largest declines in native wages.

Figure 10: **Native Wage Effects by Industry**



Note: Outcome variable is the difference between 2018 and 2015. Past-settlement instrument is used. Agricultural industries are removed as the analysis is restricted to the urban population. Electricity, gas and water supply is also removed due to the small sample available. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. F-statistic for Past-settlement Instrument is 37.73. Hourly wages are in real terms using the monthly CPI from DANE.

6.6 Overall Impacts

Next, the effect of immigration on the total population, not just native, can identify effects of the supply shock on equilibrium quantities of the labor market. First, I analyze jointly the change in (logged) total employment and total wages between 2018 and 2015. Table 6 shows the impact of immigration on aggregate employment and log hourly wages, the coefficient of the former is not significant, while for the latter I found a negative significant effect, which is higher than the coefficient on just natives and is driven by migrants' lower wages. Therefore, the labor supply shock reduces on average local wages in the economy, while the impact on overall employment is insignificant.

Table 6: **Aggregate Wages and Employment estimates, 2015-2018**

	(1) Total Wages	(2) Overall Employment
$M_{d,2018}$ (IV: Distance)	-0.0229***	-0.000126
	(0.00425)	(0.00328)
F st	157.9	165.0
$M_{d,2018}$ (IV: Past-settlement)	-0.0236***	0.00247
	(0.00642)	(0.00430)
F st	37.73	20.91
Trend-adjusted	No	Yes
N	24	24

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

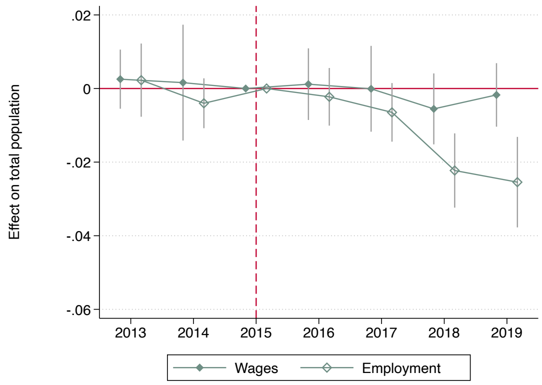
Note: The table reports the coefficients of the second stage regression of the instruments with the migration rate $M_{d,2018}$. The outcome is the difference in 2018 with the base period. Standard errors are clustered at the department level. Sample is restricted to respondents between 18 and 64 years in urban areas. Trend-adjusted estimates are controlled only by the growth in employment from 2015 compared to 2013 for each subpopulation. Department sampling weights from GEIH are used. The variables are in logarithms, thus the coefficients are interpreted as percentages. Hourly wages are in real terms using the monthly CPI from DANE.

In particular, I can analyze if the supply shock affected more the informal or formal labor market, as it is expected that migrants compete directly with natives on arrival for informal jobs. Figure 11a and 11b plots the estimated effect of immigration on overall employment and total wages using the definition of informality by affiliation to social security. A differential impact is noticeable, while formal wages are unaffected by the supply shock and formal employment decreased only after 2017, informal wages were negatively affected and informal employment positively affected.[21] In the following section I develop a simple model to guide these findings, intuitively suppose that the informal labor market behave as a standard competitive market with a downward sloping demand, then a positive supply shock puts downward pressure on wages and increases total employment, with new equilibrium values subject to the elasticities of labor supply and demand.

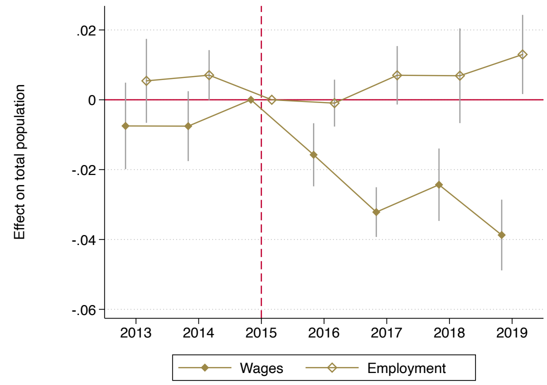
[21] With the caveat that employment results should be taken with caution as they are sensitive to the inclusion of survey weights in the estimation procedure.

Figure 11: **Employment and Wage estimates for total population by employment**

(a) IV (Distance) formal sector estimates



(b) IV (Distance) informal sector estimates



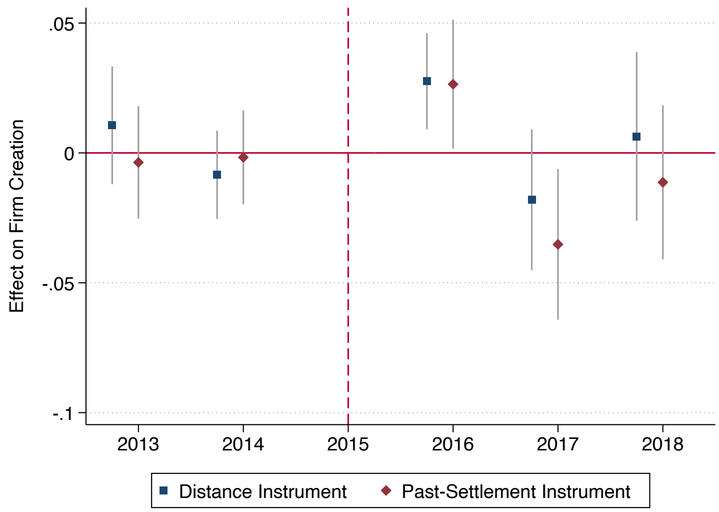
Note: Departments in the regression are $N = 24$ per year. Informality definition is given by affiliation to social security. I do not explicitly ontrol for pre-trends. The Confidence Intervals correspond to $\alpha = 0.95$. The explanatory variable is $M_{d,2018}$. Department sampling weights from GEIH are used. The outcome is the difference in 2018 with the base period. Sample is restricted to respondents between 18 and 64 years in urban areas. β_t from equation (3) are the plotted coefficients.

6.7 Impact on Firm Creation

The negative impact of immigration on wages can generate spillover effects in the form of firm entry. To check this, I analyze if higher immigration flows relate to higher flows of new firms, with respect to the base period of analysis. Using information from *Confecamaras*, that collects all the new registered firms within the country, I construct the (logged) number of newly registered firms for every year and department. In comparison to the base period, I find a time-varying differential effect in the event study, with a positive effect on firm creation in 2016, that is counteracted by a negative effect on 2017 and a close to zero point-estimate in 2018 (see Figure 12).

To interpret this finding, I argue that there could be some anticipation effects in the first post-treatment year as people foresee that the migratory event might increase and be persistent. This pushes natives (or recent migrants) to pursue independent earnings that could come from establishing new firms. Yet, recall that previous effect only pertains to registered firms that have a tax record, restricting highly the universe of new firms, specially in a country with high levels of informal firms like Colombia. In comparison to Turkey, Altindag, Bakis, and Rozo (2019) find that the large refugee shock of Syrians boosted firm creation in the country, especially for those with foreign partnership.

Figure 12: **Event Study estimates on (logged) firm creation**

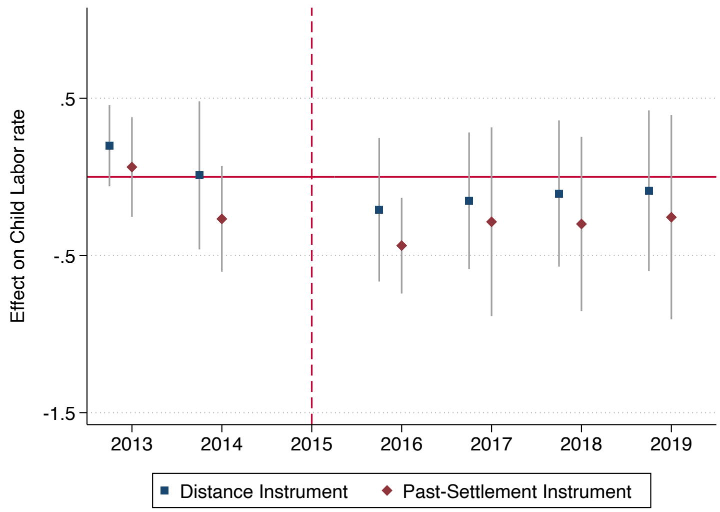


Note: Departments in the regression are $N = 24$ per year. Sample is restricted to registered, or formal, new firms that are in the databases of the corresponding state agencies. Outcome is the logarithm of new firms relative to new firms in the base period. The explanatory variable is $M_{d,2018}$. The Confidence Intervals correspond to $\alpha = 0.95$. The base period is 2015. β_t from equation (3) are the plotted coefficients. Source: Confecamaras, 2013-2018.

6.8 Impact on Child Labor

The reduction of the relative price of labor can have spillover effects on the household decisions, such as the ones regarding child work. In Colombia, nearly 600,000 kids (between 5-17 years) work according to DANE (2020). Thus, if the total amount of kids in Colombia is around 10.9 million, this yields a Child Labor Rate (CLR) of 5.4% for 2019. Figure 13 plots the estimated effect of immigration on the CLR for capital cities, where I find a negative point-estimate insignificant for all the years using distance as instrument. When past-settlement is used as instrument, significant negative effect only appears in 2016. The coefficient for this year can be interpreted (using M_{dt}) as follows: a 1 pp increase in the migration rate decreases the CLR by 0.44 pp, which is a sizable reduction. One way to rationalize this negative effect would be to think that the opportunity cost for the parents of child work increases as a result of the reduction in the relative price of labor in the informal sector, where children are more likely to work. In a related study on this issue for the US, Smith (2012) find that a 10% increase in employment of less educated immigrants reduces teen (16-17 years old) employment rate by 3% while, for adults, the reduction is of 1%, in part because youth labor supply is more elastic.

Figure 13: **Event Study estimates on Child Labor rate**



Note: Capital cities are $N = 23$. The child labor rate is calculated as $CLR = \frac{KidsWork}{TotalKids}$ and kids are defined between 5 and 17 years. The explanatory variable is $M_{d,2018}$.The base period is 2015. β_t from equation (3) are the plotted coefficients. The Confidence Intervals correspond to $\alpha = 0.95$. F-statistic for Distance Instrument is 157.9 and for Past-settlement instrument is 37.73.

7 A simple model with homogeneous labor

To guide the empirical results I introduce a simple model of homogeneous labor with two distinct labor markets or sectors, the main difference between the markets is that in one workers are formally employed (F) while in the other workers are informally employed (I). The aggregate positive supply shock of migrants is affecting only the informal labor market.[22]. There is not, in principle, a spillover demand shock from this supply shift.[23] The output of the representative firm Q_j in every market $j = F, I$ has two inputs: labor L and capital K, thus the output is given by $Q_j = f_j(K, L)$. Therefore, firm Q_i in the informal market hire workers "off the books" to avoid complying with the contributions to the social security system, while firm Q_f pay to the workers it hires a full salary. In addition, there can be a combination of formal and informal hiring to reduce the tax burden from the firms that formally hire as in Ulyssea (2018), and all workers are perfect

[22]Around 90% of immigrants in 2019 were informally employed, see Table 1b.

[23]The consumption of goods and services by migrants could shift to the right the demand curve, but for this analysis I assume not.

substitutes in production ($L = M + N$). Lastly, in the short-run the supply of capital is perfectly inelastic (or fixed), and the elasticities of labor supply and demand are different between the two markets, which are given by $\epsilon_d^I, \epsilon_s^I, \epsilon_d^F, \epsilon_s^F$.

Initially both markets F and I are in equilibrium as seen in Figure 14, with the characteristic that F has a binding minimum wage (or price-ceiling) which distorts the equilibrium in that market. In practice, there is evidence of a bunching around the minimum wage (see Figure 1a). The supply shift is driven entirely by migrants (M). With this information, the change in equilibrium wages in the informal market is given by:

$$\frac{\Delta w^I}{w^I} = \frac{M}{L^I} * \frac{1}{\epsilon_d^I + \epsilon_s^I} \tag{10}$$

Where the first term is the percentage change in the labor supply and the second term adjusts the shift by the elasticities of supply and demand in absolute terms. Then, the change in equilibrium employment in the informal sector is the following:

$$\frac{\Delta L^I}{L^I} = \frac{\Delta w^I}{w^I} * \epsilon_d^I \tag{11}$$

Where the equilibrium change in wages times the elasticity of labor demand gives the new change in equilibrium employment, as the supply moves along the demand curve. Therefore, the supply shift M lowers the market wages and increases the employment L^I (see supply-demand graph 14), consistent with what I find empirically in Figure 11b. Note that, also, there is a decline in the number of natives who work in the informal market from L_1^I to the intersection of w_2^I with S_1. On the other hand, I find in Figure 11a insignificant effects on formal wages and negative ones of formal employment, thus if there exists a reduction on employment is only through a reduction in the demand of formal workers. If the cost of informal hiring decreases this could lead to a readjustment of labor choices from firms (see graph 14 of the formal market).

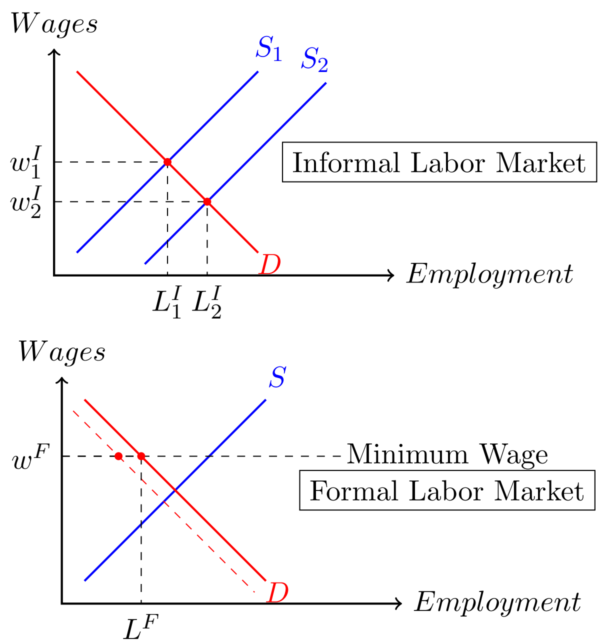
The next step is to use previous estimates to recover the values of the elasticity of labor demand and supply in the informal market. First, using the overall change of informal employment over native informal wage change for 2018, I find that the elasticity of labor demand is given by equation (11): $\epsilon_d^I = \frac{0.69\%}{-1.87\%} = -0.37$.[24] Which is on the range of estimates of labor demand elasticities

[24] I use the change in native wages in 2018 on this sector given in Figure 9a and not the total change in wages to

previously found (Lichter, Peichl, and Siegloch, 2015; Hamermesh, 1996). [25]

As regard the elasticity of labor supply of native workers, it is given by equation (10): $\epsilon_s^I = \frac{1.72\%}{1.87\%} - 0.37 = 0.55$.[26] Note that this is a smaller elasticity compared to Albouy et al. (2019) findings for US local labor markets, yet consistent with a high value of elasticity of labor supply at the local level in which workers can move between regions or can enter/exit the labor force as explained in Dustmann, Schönberg, and Stuhler (2017).

Figure 14: **Market responses to a supply shift when immigrants and natives are perfect substitutes**



Note: Only the informal labor market suffer a positive supply shock. Wages for firms in the formal market are downwardly rigid. The red dashed line shows a possible reduction in the demand of formal labor given the reduction in the cost of informal hiring. All the slopes in the graphs are given by its corresponding elasticities, but graphically I plot them as equal in both markets.

remove possible compositional effects with the arrival of migrants.

[25]If we restrict to informal markets, Guriev, Speciale, and Tuccio (2019) find for Italy an elasticity of labor demand of around -1, meaning a more elastic demand in this sector and close to the long-run one, when it is possible to adjust capital.

[26]Where the numerator of the first term is the supply migrant shift, measured as a 1 pp increase in the share of employed Venezuelans between 2018 relative to 2015 over total employment in the informal sector in 2018, both according to census, and the denominator is the native wage change, finally I subtract by the recovered labor demand elasticity.

8 Robustness Checks

The first robustness check of the empirical results to be implemented relates to the exclusion of border departments from the analysis. The goal is to check if the direction and significance of those results hold when removing departments that are more geographically close to Venezuela, since they could be more affected by the crisis in that country through less trade links or lower businesses interactions. The main results of this exercise are displayed in Figure 17a, using distance as the instrument yields no significant variations in the estimate, also when the instrument is past-settlement coefficients appear to be robust to the exclusion of border departments (see Figure 17b).

The second exercise relates to the validity of the identifying assumption of the instrument based on past-settlements, due to an existing correlation between the distribution of Venezuelans in 2005 and current economic trends.To test the robustness of these results to this instrument, I select two farther historically lagged (1993 and 1973) census shares of Venezuelans in Colombia from Minnesota Population Center (2019). Figure 18 displays the coefficients using that instrument with the three distinct shares (2005, 1993, and 1973). As can be observed, there are no significant differences between them. In the case of the distance instrument, a threat to the identifying assumption might be that trade or business patterns, derived from the Venezuelan crisis, could affect more severely geographically closer departments in which the instrument predicts more migrants. To check that possibility, I use time-varying real log-GDP and share of total exports in USD to Venezuela with respect to the world fixed in 2015, to capture the behavior of exports with Venezuela before the migration crisis started. The corresponding estimates of their effect on native wages are plotted in Figure 19, confirming that all the estimates remain significant (with a lower point-estimate for GDP and a higher one for Trade).

Third, when surveys are used for causal studies, survey weights become problematic and it is unclear whether one should use them or not in the main specification. Solon, Haider, and Wooldridge (2015) argue *always* for the use of weights for descriptive analysis, however for causal analysis the choice is more complicated. In this paper, though I do not use weights in the main specification, I use department weights to compute the means of outcome variables at the department level. Figure 20a presents the causal effect of immigration on wages without using weights for means at department level. This procedure provides higher point-estimates (in absolute terms) and amplifies the negative effect in 2019, but well in the range of previous findings. In terms of native employment, when I exclude the survey weights in the preceding step in order to count the total native employed population in each area, impact on employment now is insignificant for all post-treatment years (Figure 20b), highlighting the sensitivity of this result.

Finally, when using event-study regressions, outcomes are calculated at the department level without taking into account individual information. To check whether this matters, I compute residual wages in a preliminary stage using the individual characteristics of the respondent. [27] As shown in Figure 21, which plots the coefficients of immigration taking as dependent variable the observed wage and the residual wage, the residual wage has a similar yet higher estimate (in absolute terms), in 2019 I find the largest difference. Still there is no much gain in regressing previous individual characteristics in my analysis.

8.1 More pre-treatment years

When analyzing event studies pre-treatment information is crucial to assess the violation of the common trend assumption. Sometimes in practice the PTA is not rejected in recent periods, yet if you go farther in time there can be differing tendencies. To examine this issue, I check whether the previous results hold when more pre-treatment data is used.[28] Figure 22a show that, when adding two more years of data before treatment, instruments predict local wage trends only in 2011 (even if not significant for both instrument, the point-estimate is quite large) which does not seem to be a big problem since trends have been rather stable after 2011. As for employment, adding more periods shows that department local markets exhibited quite a bit of fluctuations but were not on persistently different growth paths, as the coefficient for 2011 turns to be close to zero (Figure 22b).

8.2 Bootstrap Standard Errors

Given the small number of cluster areas, over-rejection of the null hypothesis when it is true can be problematic (Pustejovsky and Tipton, 2018). Wild bootstrap in this small-sample setups, where

[27] Residual wages are retrieved from an unweighted regression of log hourly wages on two polynomials of age, years of schooling, gender, and fixed effects of department, year and month.

[28] Taking into account that the migration module was implemented in 2013, so that I cannot differentiate between natives and migrants before that year, I assume all the respondents were Colombians in 2011 and 2012.

asymptotic assumptions do not hold, becomes a useful practice. In particular, *boottest* yields a direct p-value and Confidence Intervals (CIs) given α from the bootstrap replications of the empirical distribution (Roodman et al., 2019). Taking into account that symmetry does not carry over in this setup, such a procedure gives a lower or higher bound on the estimate. Figure 23a plots the estimated effect of immigration on native wages with bootstrap standard errors (SEs), showing that, despite the CIs being not symmetric and that the estimates are less precise than before (when I use the empirical SEs), the estimates remain significant in all the post-years after 2016. As for the effect of immigration on (logged) native employment, first the pre-trends are no longer significant when wild bootstrap SEs are used and second the effect on post-years remain significant after 2017.

9 Conclusion

This paper analyzes the impact of the Venezuelan mass migration on the Colombian labor market. Exploiting the treatment intensity in the different departments of Colombia, I implement an event-study design with two separate instruments to assess the validity of the findings. Combining micro evidence from the labour force survey with a recent population census, I estimate a negative effect on native local wages of between 1.6%-1.7% for a 1 pp increase in the share of employed Venezuelans over total employed population. This impact is mainly concentrated among workers with an informal contractual labor relation, less-educated ones and males. When analyzing the effect along the wage distribution, wages located in lower percentiles (25th percentile) are severely more affected by the Venezuelan immigration compared to those wages located in the upper percentiles (90th percentile).

Furthermore, native employment appears to have a delayed negative response compared to wage response, that is more pronounced on younger workers (from 18 to 25 years). On aggregate, the influx of Venezuelans increased informal employment and decreased informal wages, while having an insignificant effect on formal wages. Thus the supply shock affected mainly the informal labor market.

References

Aksu, E., Erzan, R., and Kırdar, M. G. (2018). The impact of mass migration of syrians on the turkish labor market. Technical report, Working Paper.

Albouy, D., Chernoff, A., Lutz, C., and Warman, C. (2019). Local labor markets in canada and the united states. *Journal of Labor Economics*, 37(S2):S533–S594.

Altindag, O., Bakis, O., and Rozo, S. (2019). Blessing or burden? the impact of refugees on businesses and the informal economy. *The Impact of Refugees on Businesses and the Informal Economy (May 15, 2019).*

Altonji, J. G. and Card, D. (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, trade, and the labor market*, pages 201–234. University of Chicago Press.

Amior, M. (2020). Immigration, local crowd-out and undercoverage bias. Technical report, Centre for Economic Performance, LSE.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Aydemir, A. and Borjas, G. J. (2011). Attenuation bias in measuring the wage impact of immigration. *Journal of Labor Economics*, 29(1):69–112.

Bahar, D., Rozo, S., and Ibáñez, A. M. (2020). Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees?

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Borjas, G. J. (1999). The economic analysis of immigration. In *Handbook of labor economics*, volume 3, pages 1697–1760. Elsevier.

Borjas, G. J. (2001). Does immigration grease the wheels of the labor market? *Brookings papers on economic activity*, 2001(1):69–133.

Borjas, G. J. (2017a). The labor supply of undocumented immigrants. *Labour Economics*, 46:1–13.

Borjas, G. J. (2017b). The wage impact of the marielitos: A reappraisal. *ILR Review*, 70(5):1077–1110.

Borjas, G. J. and Chiswick, B. R. (2019). *Foundations of Migration Economics.* Oxford University Press.

Borusyak, K., Hull, P., and Jaravel, X. (2018). Quasi-experimental shift-share research designs. Technical report, National Bureau of Economic Research.

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *ILR Review*, 43(2):245–257.

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

Card, D. and Peri, G. (2016). Immigration economics by george j. borjas: a review essay. *Journal of Economic Literature*, 54(4):1333–49.

Caruso, G., Canon, C. G., and Mueller, V. (2019). Spillover effects of the venezuelan crisis: migration impacts in colombia. *Oxford Economic Papers.*

DANE (2020). Module of child work. Technical report, Departamento Administrativo Nacional de Estadística. Available on: https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/trabajo-infantil (last accessed 6th of July 2020).

de Chaisemartin, C. and d'Haultfoeuille, X. (2019). Two-way fixed effects estimators with heterogeneous treatment effects. Technical report, National Bureau of Economic Research.

Del Carpio, X. V. and Wagner, M. (2015). *The impact of Syrians refugees on the Turkish labor market.* The World Bank.

Dustmann, C., Frattini, T., and Preston, I. P. (2013). The effect of immigration along the distribution of wages. *Review of Economic Studies*, 80(1):145–173.

Dustmann, C., Schönberg, U., and Stuhler, J. (2016). The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives*, 30(4):31–56.

Dustmann, C., Schönberg, U., and Stuhler, J. (2017). Labor supply shocks, native wages, and the adjustment of local employment. *The Quarterly Journal of Economics*, 132(1):435–483.

Edo, A. (2017). The impact of immigration on wage dynamics: Evidence from the algerian independence war. *Journal of the European Economic Association*.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018). Bartik instruments: What, when, why, and how. Technical report, National Bureau of Economic Research.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Technical report, National Bureau of Economic Research.

Grossman, J. B. (1982). The substitutability of natives and immigrants in production. *The review of economics and statistics*, pages 596–603.

Guriev, S., Speciale, B., and Tuccio, M. (2019). How do regulated and unregulated labor markets respond to shocks? evidence from immigrants during the great recession. *The Journal of Law, Economics, and Organization*, 35(1):37–76.

Hamermesh, D. S. (1996). *Labor demand*. princeton University press.

Hunt, J. (1992). The impact of the 1962 repatriates from algeria on the french labor market. *ILR Review*, 45(3):556–572.

IIES (2020). Encuesta nacional de condiciones de vida. Technical report, Instituto de Investigaciones Económicas y Sociales. Available on: https://assets.website-files.com/5d14c6a5c4ad42a4e794d0f7/5f03875cac6fc11b6d67a8a5_Presentaci%C3%B3n%20%20ENCOVI%202019-Pobreza_compressed.pdf (last accessed 7th of July 2020).

IMF (2020). World economic outlook. Technical report, International Monetary Fund. Available on: https://www.imf.org/external/datamapper/profile/VEN/WEO (last accessed 20th of June 2020).

Jaeger, D. (2007). Green cards and the location choices of immigrants in the united states, 1971-2000. *Research in Labor Economics*, 27:131–183.

Jaeger, D. A., Ruist, J., and Stuhler, J. (2018). Shift-share instruments and the impact of immigration. Technical report, National Bureau of Economic Research.

Lee, D. L., McCrary, J., Moreira, M. J., and Porter, J. (2020). Valid t-ratio inference for iv.

Lichter, A., Peichl, A., and Siegloch, S. (2015). The own-wage elasticity of labor demand: A meta-regression analysis. *European Economic Review*, 80:94–119.

Minnesota Population Center (2019). Integrated public use microdata series, international: Version 7.2 [dataset]. https://doi.org/10.18128/D020.V7.2.

Monras, J. (2020). Immigration and wage dynamics: Evidence from the mexican peso crisis. *Journal of Political Economy*, 128(8):000–000.

Morales-Zurita, L. F., Bonilla-Mejía, L., Hermida, D., Flórez, L. A., Bonilla-Mejía, L., Morales, L. F., Hermida-Giraldo, D., and Flórez, L. A. (2020). The labor market of immigrants and non-immigrants evidence from the venezuelan refugee crisis. *Borradores de Economía; No. 1119*.

Olivieri, S., Ortega, F., Rivadeneira, A., and Carranza, E. (2020). Shoring up economic refugees.

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

Pustejovsky, J. E. and Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4):672–683.

Reuters (2020). Colombia says over 52,000 venezuelans return home, cites lockdown. Technical report. Available on: https://www.reuters.com/article/us-health-coronavirus-colombia-migrants/colombia-says-over-52000-venezuelans-return-home-cites-lockdown-idUSKBN22O2Q6 (last accessed 11th of July 2020).

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., and Webb, M. D. (2019). Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal*, 19(1):4–60.

Santamaria, J. (2019). Venezuelan exodus: the effect of mass migration on labor market outcomes.

Smith, C. L. (2012). The impact of low-skilled immigration on the youth labor market. *Journal of Labor Economics*, 30(1):55–89.

Solon, G., Haider, S. J., and Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human resources*, 50(2):301–316.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, 108(8):2015–47.

UNHCR (2019). Venezuela situation. Technical report, Office of the United Nations High Commissioner for Refugees. Available on: https://www.unhcr.org/venezuela-emergency.html (last accessed 13th of June 2020).

Weber, S. and Péclat, M. (2017). A simple command to calculate travel distance and travel time. *The Stata Journal*, 17(4):962–971.

Appendix

A.1 Descriptive Statistics for Natives and Migrants

Table 7: Descriptive statistics for permanently residing Colombians, recent arrivals of Venezuelans and returning Colombians

(a) Colombians residing permanently in Colombia

Year	Age					Gender	Schooling			Sample	
	(%) 0-14	(%) 15-28	(%) 29-40	(%) 41-64	(%) 65+	(%) Men	(%) NHS	(%) HS	(%) College	N	Population
2013	0.275	0.240	0.168	0.242	0.074	0.493	0.599	0.180	0.171	595,847	45,693,877
2014	0.272	0.239	0.168	0.245	0.076	0.493	0.591	0.181	0.177	785,695	46,140,214
2015	0.267	0.239	0.170	0.246	0.078	0.493	0.583	0.191	0.176	783,888	46,627,550
2016	0.263	0.238	0.170	0.247	0.082	0.493	0.570	0.199	0.181	773,524	47,044,882
2017	0.260	0.238	0.171	0.248	0.084	0.493	0.563	0.207	0.183	761,148	47,456,897
2018	0.256	0.236	0.172	0.250	0.086	0.493	0.552	0.212	0.188	750,973	47,590,415
2019	0.253	0.233	0.174	0.250	0.089	0.493	0.544	0.221	0.189	743,301	48,017,793

(b) Venezuelans that arrived in the preceding year to Colombia

	(%) 0-14	(%) 15-28	(%) 29-40	(%) 41-64	(%) 65+	(%) Men	(%) NHS	(%) HS	(%) College	N	Population
2013	0.354	0.214	0.280	0.134	0.019	0.512	0.675	0.052	0.210	119	9,047
2014	0.556	0.333	0.071	0.038	0.003	0.542	0.651	0.170	0.083	205	12,712
2015	0.558	0.269	0.111	0.062	0.000	0.486	0.508	0.121	0.146	475	28,667
2016	0.462	0.326	0.164	0.045	0.003	0.518	0.548	0.171	0.174	1,421	86,128
2017	0.401	0.362	0.179	0.056	0.001	0.510	0.493	0.211	0.201	3,577	205,277
2018	0.325	0.382	0.200	0.089	0.004	0.510	0.477	0.262	0.192	8,543	576,647
2019	0.338	0.361	0.181	0.111	0.009	0.483	0.514	0.260	0.155	10,123	719,121

(c) Colombians that lived in Venezuela in the preceding year and returned back to Colombia

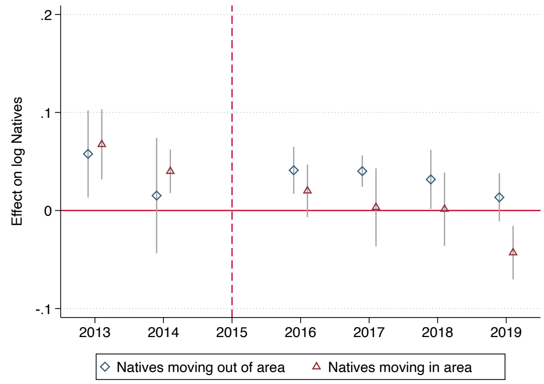
	(%) 0-14	(%) 15-28	(%) 29-40	(%) 41-64	(%) 65+	(%) Men	(%) NHS	(%) HS	(%) College	N	Population
2013	0.156	0.322	0.222	0.237	0.064	0.522	0.650	0.240	0.104	379	25,500
2014	0.162	0.328	0.261	0.218	0.030	0.544	0.635	0.240	0.093	678	33,043
2015	0.165	0.312	0.285	0.225	0.012	0.540	0.629	0.292	0.071	1,062	51,436
2016	0.161	0.278	0.276	0.258	0.026	0.518	0.696	0.232	0.069	1,586	85,716
2017	0.151	0.198	0.283	0.305	0.063	0.488	0.670	0.256	0.072	1,504	84,412
2018	0.056	0.190	0.236	0.419	0.099	0.484	0.710	0.212	0.073	1,591	100,749
2019	0.087	0.169	0.206	0.406	0.132	0.513	0.653	0.222	0.100	846	47,357

Note: NHS stands for No High School and HS stands for High School. The shares are calculated using national sampling weights from GEIH. College aggregates all the technical levels of education after high school. (b) and (c) are restricted to population that in the survey responded that they were living in Venezuela in the last year. Source: GEIH, 2013-II to 2019.

A.2 Robustness Checks

In here I include all the robustness analysis performed throughout paper. Most of the checks are event study estimates for different specifications, changes in the treated group and threats to the identifying assumptions.

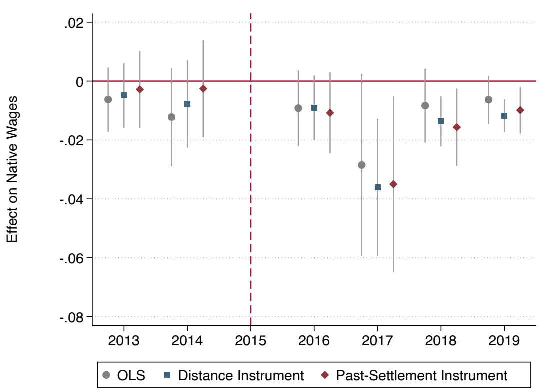
Figure 15: Event Study estimates on movements across geographical areas



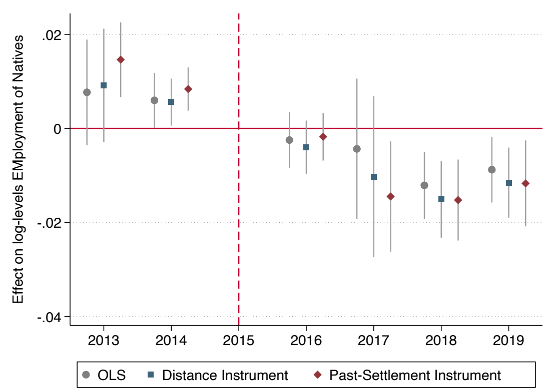
Note: Departments in the regression are $N = 24$ per year. The Confidence Intervals correspond to $\alpha = 0.95$. Measures of geographical movements come from GEIH migration module. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients.

Figure 16: Event Study estimates using M_{dt} as explanatory variable

(a) IV native wages



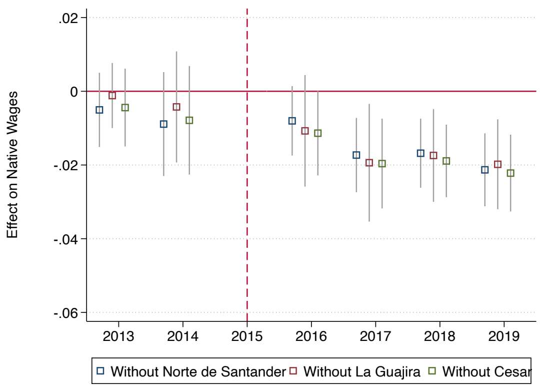
(b) IV native employment



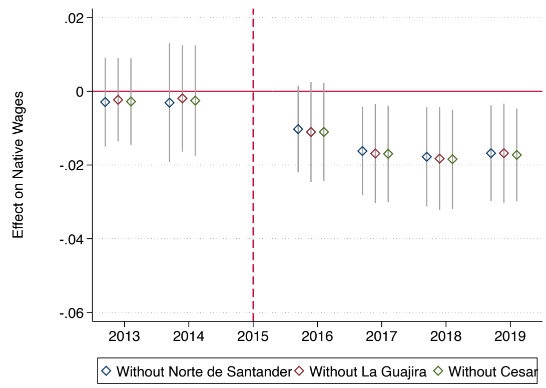
Note: Departments in the regression are $N = 24$ per year. The explanatory variable is $M_{d,2018}$ before 2017, and after is M_{dt}. The Confidence Intervals correspond to $\alpha = 0.95$. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. Hourly wages are in real terms using the monthly CPI from DANE.

Figure 17: Event Study estimates excluding border departments for native Wages

(a) IV (distance) estimates

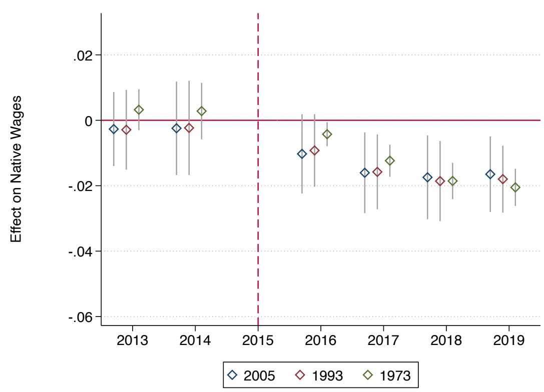


(b) IV (past-settlement) estimates



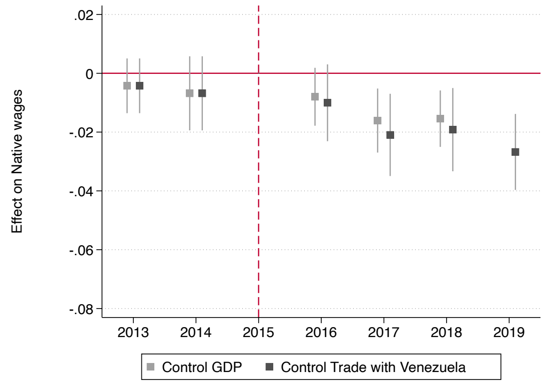
Note: Departments in the regression are $N = 23$. The Confidence Intervals correspond to $\alpha = 0.95$. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. The base period is 2015. β_t from equation (3) are the plotted coefficients.Hourly wages are in real terms using the monthly CPI from DANE.

Figure 18: Event Study estimates using different historical shares for the construction of past-settlement instrument



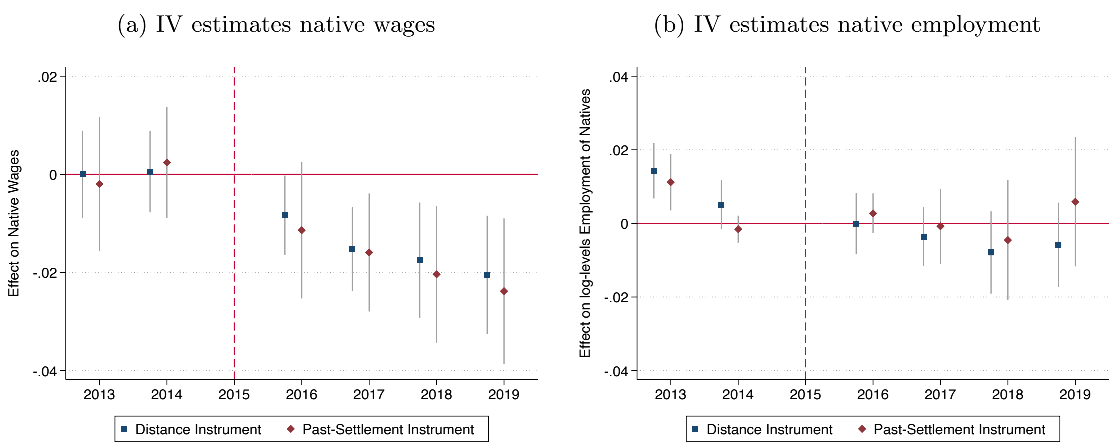
Note: Departments in the regression are $N = 24$ for 2005 and 1993, and $N = 22$ for 1973. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. The Confidence Intervals correspond to $\alpha = 0.95$. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. F-statistic for Past-Settlement Instrument with shares of 2005 is 37.73, with shares of 1993 is 36.24 and with shares of 1973 is 37.73. Source: IPUMS for 1993 and 1973 and DANE for 2005.

Figure 19: Event Study estimates on log hourly wages of Colombians with additional controls



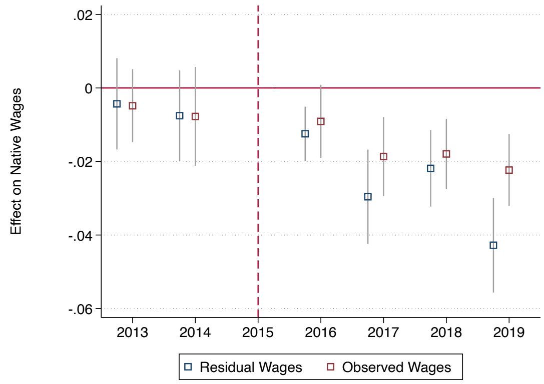
Note: Departments in the regression are $N = 24$ per year. The Confidence Intervals correspond to $\alpha = 0.95$. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. Constant prices GDP at the department level are constructed by DANE.

Figure 20: Event Study estimates without using survey weights for log hourly wages and log employment



Note: Departments in the regression are $N = 24$ per year. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. The base period is 2015. β_t from equation (3) are the plotted coefficients. The Confidence Intervals correspond to $\alpha = 0.95$. F-statistic for Distance Instrument is 157.9 and for Past-Settlement Instrument is 37.73. Hourly wages are in real terms using the monthly CPI from DANE.

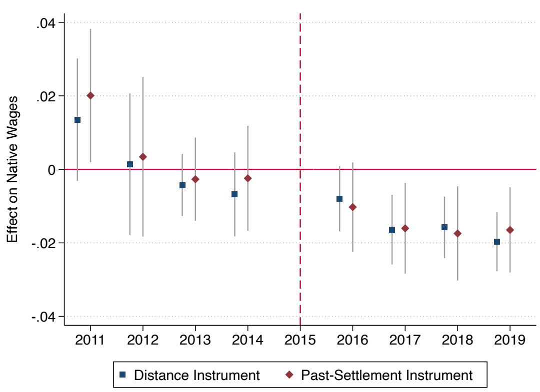
Figure 21: Event Study estimates of residual versus vs observed wages



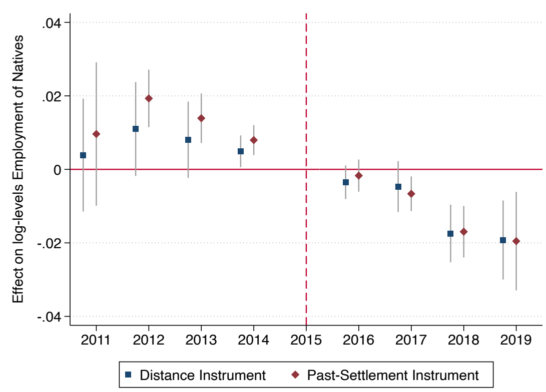
Note: Departments in the regression are $N = 24$ per year. Residual wages come from an unweighted regression of hourly wages on two polynomials of age, years of schooling, gender, and fixed effects of department, year and month. Distance instrument is used. The Confidence Intervals correspond to $\alpha = 0.95$. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. F-statistic for distance instrument is 157.9.

Figure 22: Event Study estimates on log hourly wages and log employed Colombians

(a) IV estimates on Wages



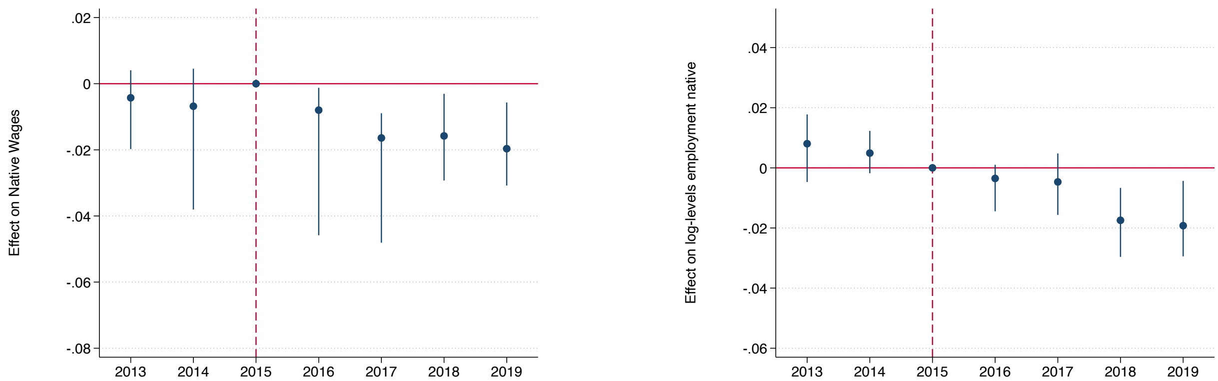
(b) IV estimates on Employment



Note: Departments in the regression are $N = 24$ per year. The Confidence Intervals correspond to $\alpha = 0.95$. In 2011 and 2012 I assume all respondents are Colombians. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. F-statistic for Distance Instrument is 157.9 and for Past-settlement instrument is 37.73. Hourly wages are in real terms using the monthly CPI from DANE.

Figure 23: Event Study estimates on two different outcomes using Bootstrap Standard Errors

(a) IV (distance) estimates (b) IV (distance) estimates



Note: Departments in the regression are $N = 24$ per year. *boottest* command is used. Bootstrap replications are 999. The Confidence Intervals correspond to $\alpha = 0.95$. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015. β_t from equation (3) are the plotted coefficients. F-statistic for Distance Instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

A.3 Definition of Variables

Log hourly real wages. The variable is constructed as follows. First, I use *inglabo* variable from GEIH survey that captures basic pay, pay in-kind and income of second activity, I subtract income of second activity and add allowances for food and transportation (According to ILO definition of wages[29]). Following up, I transform this nominal wage variable into a real one using monthly CPI at the national level.[30] The base of the index (=100) is December of 2018.

$$RealWage_{imy} = \frac{NominalWage_{imy}}{CPI_{my}} * 100$$

Where i stands for individual, m for month and y for year. Then I divide real wages by four to have a weekly wage. Then I divide by the number of working hours that the respondent reported to work in the previous week. In a next step, I only consider positive values of wages and top code wages above the 99% threshold of the wage distribution in each department-year. Finally, I take the weighted averages (with department weights) and use the logarithm transformation of the final expression.

Log employed Colombians. I take as employed all the Colombians between 18 and 64 years in urban areas that reported to work at least one hour in the previous week, paid or unpaid for cash or in-kind from GEIH survey. I count (with department weights) all individuals in each

[29] https://www.ilo.org/wcmsp5/groups/public/---africa/---ro-abidjan/---ilo-pretoria/documents/publication/wcms_413782.pdf

[30] Information was taken from here https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc

department-year and then I take logarithms of that expression.

Employed definition according to Census. Census does not have all the questions of a standard labor force survey regarding occupation. It only had one question which asked what the respondent did last week, if it selects work for a compensated income for at least 1 hour I treat them as occupied, and not otherwise.

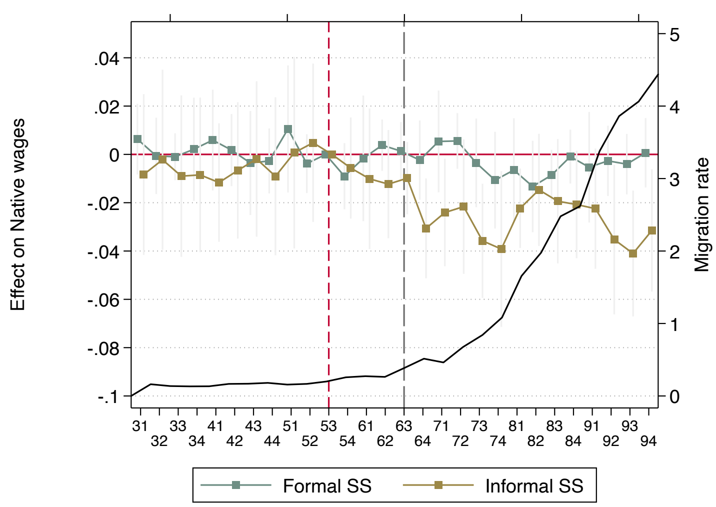
A.4 Event-study with quarterly information

To perform the event-study at a more detailed time frequency, I select as a base period of comparison, instead of the entire 2015, just the 2015-3 quarter, which is when the Venezuelan government closed the border. The empirical specification is the same as before, changing the yearly subscript y with the quarterly one q. Thus, I estimate the following, the base period is 2015-3 when $k = 0$,

$$Y_{dq} = \alpha + \gamma_q + \gamma_d + \sum_{q=2013-2}^{2019-4} \beta_q \tilde{M}_{d,2018} \mathbf{1}\{q - \textit{2015-3} = k\} + u_{dt} \tag{12}$$

and $\tilde{M}_{d,2018}$ is the predicted migration rate using the two separate instruments. Note that the red dashed line in Figure 24 represents the base period of analysis, the grey long-dashed line represents the re-opening of the border between Colombia and Venezuela, and the black line is the quarterly migration rate from GEIH survey (I only use the distance instrument). After the grey line it is more pronounced the effect on native wages on workers without access to social security, while for the formal ones wages are unaffected.

Figure 24: Event Study estimates on log hourly wages by quarters and affiliation to social security



Note: Departments in the regression are $N = 24$ per year. The black line is the quarterly migration rate from GEIH survey. The Confidence Intervals correspond to $\alpha = 0.95$. Sample is restricted to Colombians permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used. The base period is 2015 third quarter. β_t from equation (3) are the plotted coefficients. F-statistic for Distance Instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.