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

Lukas Delgado-Prieto†

May 20, 2021

Click here for latest version

Abstract

Venezuela's unprecedented socioeconomic and political deterioration has triggered massive flows 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 Venezuelan immigration on the Colombian labor market. To analyze dynamic treatment effects I implement an event study research design with two different shift-share instruments. For both instruments I find a negative short-run effect on local native wages since 2017, and a delayed negative response, after controlling for preexisting trends, on local native employment. Overall, the labor supply shock is absorbed by the informal labor market, reducing informal wages, but also reducing formal employment. A model of formal and informal labor with different hiring costs can explain such finding.

Keywords: Immigration, Event study, Labor market.

JEL Codes: F22, O15, O17, R23.

*I am grateful to Jan Stuhler for all the detailed comments, useful arguments, in-depth suggestions and close guidance, to Luigi Minale for further remarks and discussions, and to Juanjo Dolado for the extensive revision of the manuscript. I am also thankful to Jesús Fernández-Huertas, Jorge Pérez Pérez, and Sarah Schneider-Strawcynski for helpful feedback. In general, I am indebted to all the comments from participants in the UC3M Applied Reading Group, the Junior Economics of Migration Seminar, Seminario de Economía Aplicada BANREP-UIS-UNAB and the 10th European Meeting of the Urban Economics Association. I thank also all the support from the staff at DANE, especially from Geovanni Portilla. Replication files are available under request. All errors are my own.

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

1 Introduction

Despite the large number of studies over the last three decades, the impact of immigration on native wages and native employment still remains as one of the most relevant, albeit disputed, issues in empirical labor economics (see Dustmann, Schönberg, and Stuhler (2016) and Borjas and Chiswick (2019) for a summary of findings). This paper contributes to this debate by analyzing a recent exogenous and increasing change in the number of Venezuelan immigrants in Colombia to estimate its causal effects on Colombian wages and employment using detailed micro-data on labor and demographic status. The setting I analyze emerges 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) more than 4.6 million people have left Venezuela between 2016 and 2019, 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, both in the short and long-run, a wide range of socio-economic outcomes in the host countries, including employment, health and education.

In this paper, I focus on the labor market impacts of the Venezuelan mass migration in 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) surged 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 low-educated 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. Moreover, the changes in wages could interact with changes in equilibrium values of employment, in line with the widely documented trade-off between these two outcomes in the face of migration shocks (Borjas, 1999).

The empirical strategy I use to evaluate the size and sign of previous effects relies on an event study of calendar time that takes advantage of the intensity of treatment. In effect, while some states (called *Departamentos*) in Colombia received vast inflows of migrants, others hardly got any. For this research design to be valid, the key assumption needed for the causal interpretation of the parameter is the unconditional parallel trends assumption (UPTA). In practice, this assumption might not be fulfilled since migrants can endogenously sort themselves into those areas that offer the best economic conditions, which might lead to differential trends prior to the shock. To deal with the potential sorting and other endogeneity issues, I combine the event study design with a shift-share instrument (IV). Specifically, I construct two plausibly exogenous shares: (*i*) distance between capital cities in the two neighboring countries and (*ii*) historical enclaves or past settlements of Venezuelans. 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 high-quality sources of information that provide a rich set of demographics on natives and migrants from Venezuela. These are the Colombian Labor Force Survey (LFS) and the most recent Population and Housing census from 2018. Both datasets offer rich information on the individual characteristics of the migrant, like the date of entrance to the country, and native population in Colombia, reducing in this way the compositional bias that could arise from aggregating recent immigrants with already settled ones who differ in their observable characteristics. But not completely, as the LFS is a repeated cross-section survey that does not follow the same workers through time.

My estimates show that the inflow of Venezuelans has persistently reduced native (hourly) wages since 2017. In particular, a 1 percentage point (p.p) increase in the share of employed Venezuelans over the employed population in each department reduces local native wages by 1.6%-1.7%. Importantly, these estimates are robust to the two chosen instruments, to different definitions of local labor markets and to different wage sources. Compared to previous literature on native wage response to immigration, my estimates are: (*i*) smaller to those found in the related setup studied by Caruso, Canon, and Mueller (2019), where a 1 p.p 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 p.p 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 p.p increase in the overall employment of Czech workers decreased local native wages by 0.13%. In view of this evidence, one of the main 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 labor market adjustment to the unexpected supply shock.[1]

I argue that a key factor driving previous negative estimates is the lack of downward rigidity of wages in the informal sector, since 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 to the immigration shock since 2018, after controlling for preexisting trends. On aggregate terms, I find that informal employment grew and informal wages decayed, whereas formal employment decrease and formal wages were not affected, precisely by the rigidity of minimum wages. A model with two labor inputs (formal and informal labor) is introduced to explain these results. According to this model, the degree of substitutability between formal and informal workers for the firm relative to the price elasticity of output demand determines the differential labor demand reaction to an informal labor supply shock.

In addition, Colombians appear to be working more hours: a 1 p.p increase in the migration rate increases hours worked per week of natives by 0.9%, partially offsetting the reduction in natives (hourly) wages. In terms of the wage distribution, most affected wages are located at the lower part of the local wage distribution (in the 25th percentile), while wages on the upper part (90th percentile) are almost unaffected. The last set of results of this paper analyzes heterogeneous effects, the reduction on native wages is driven by less-educated workers (with high school or less), while the highest reduction of employment is observed on the youngest workers (from 18 to 25 years). Finally, I complement this analysis studying also the price response. Overall the insignificant estimates on prices indicate a stronger supply effect via lower wages or higher market competition.

Relative to the general literature on migration, the contributions of this paper can be summarized as follows. The characteristics of the immigration shock under study, namely, a large and sudden inflow of migrants driven by the conditions in the sending country, help 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 this paper, I have more than

[1] For instance, Monras (2020) find that low-skilled Mexicans who left their country as a result of the Peso crisis had a high transitory labor market impact on the US, that quickly dissipated across states as time passed and local markets adjusted.

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), child labor (where I find negative point-estimates, though not significant) and prices (I find close to zero estimates).

With respect to previous studies that estimate the impact of the Venezuelan migration on Colombia's labor market (Caruso, Canon, and Mueller, 2019; Lebow, 2020; 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 preexisting trends. Second, the use of this design is also motivated by the fact that the static coefficient reported by Caruso, Canon, and Mueller (2019); Lebow (2020); Morales-Zurita et al. (2020) in 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 heterogenous treatment effects (Goodman-Bacon, 2018; de Chaisemartin and d'Haultfoeuille, 2019). Third, I extend the results found in previous studies by analyzing the effect of immigration not only on the average local wage but across the entire native wage distribution. Fourth, I provide a theoretical model that links the informal and formal sectors of the labor market. This is highly relevant in this setup because Venezuelan migrants are disproportionally employed in the informal sector, but empirical findings shows that the resulting downward pressure in informal wages led to a reallocation from formal to informal labor for the firm. So, while the migration wave decreased wages in the informal sector, its employment effects are felt primarily in the formal sector.

The rest of the paper is structured as follows. Section 2 discusses 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 documents the differential impact of immigration between the formal and informal market and introduces the theoretical model. Section 8 shows heterogenous impacts of immigration and additional outcomes. Section 9 is about the robustness tests performed. Finally, Section 10 discusses and concludes.

2 Related Literature

Early studies on immigration focused on the comparison of local labor markets across different areas (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 UPTA). 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.[2] I use this approach in this paper, following the empirical specification of Dustmann, Schönberg, and Stuhler (2017) that implements an event study design with IV to study a commuting policy on the Bavarian region of Germany, these authors 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 mi-

[2]Nonetheless, as early papers received a lot of scrutinies, 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 scarce, omitting the pre-trends test on whether the PTA holds, also the selection of good control groups was limited. Second, it was hard to correctly estimate the standard errors on few treated clusters using differences-in-differences (DiD) (Bertrand, Duflo, and Mullainathan, 2004). As a result, there have been several critical replies to the first Algerian and Mariel Boatlift papers which have highlighted the controversy around immigration studies. For the Mariel Boatlift, the first reply of Borjas (2017b) focused on the impact on high-school dropouts wages (omitted in Card (1990)), by 1985 they were excluded by 30% (relative to control cities). Also Peri and Yasenov (2019) reexamined this natural experiment using a synthetic control method, building a synthetic control Miami with a composition of cities 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, the author finds a strong decline in wages in counter-view to the original Hunt study. All these replies, at the same time, 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 employing the spatial approach.

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

gration event. For instance, Aksu, Erzan, and Kırdar (2018) find that the influx of Syrians strongly decreases 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 a very large negative effect on wages, a 1 p.p increase in the labor supply of immigrants in Colombia diminishes in 7.6% the hourly wage of workers, a negative effect which is mainly driven 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]

Three more recent unpublished papers 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. Using 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 the 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 the 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 the time of arrival, so that a compositional bias can arise if recent migrants have different characteristics from the settled migrant population. A third paper

[4]The instrument used by Caruso, Canon, and Mueller (2019) is the distance between the 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.

by Lebow (2020) hinges on a more structural approach and estimates the substitutability between natives and immigrants in Colombia to perform, afterwards, simulations of the wage effect in the absence of occupational downgrading.[5]

An empirical issue that can arise in Caruso, Canon, and Mueller (2019), Morales-Zurita et al. (2020) and Lebow (2020) is their use of a static two-way fixed effects (TWFE) regression. 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 and migration flows between 2013-2015, but they still do not analyze if their instrument predicts trajectories of economic outcomes during those years. Finally, Lebow (2020) study pre-trends for several outcomes, to find significant ones on unemployment and participation rates. 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) points out, the coefficient of interest in the static TWFE model consists of a weighted average of treatment effects, where these weights are not necessarily positive when timing of treatment varies. Thus, using this regression could bias the results, since, for instance, TWFE uses 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 different results are likely 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.

[5] In terms of reduced form estimates, wage effects are somewhat smaller to this paper (-1.05% vs. -1.7%). However, since the migration rate is built from GEIH survey, which can increase the measurement error of immigrants, estimates can be attenuated (Aydemir and Borjas, 2011).

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 its 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 in 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 flows 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 (UCAB, 2020).

Therefore, the reasons for 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. In the data section, I show that a substantial amount of Venezuelans enter the country through the main crossing bridges, probably walking without any formal documents to prove their previous education level or work experience.

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, and even smaller depending on the informality definition, than the Colombian ones 1b. From 2015 onwards, the picture changed and more Venezuelan 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).[6] Aimed at fostering legal and more accessible employment for Venezuelans without the need for 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 wages, unemployment or participation in the labor market for natives (Bahar, Rozo, and Ibáńez, 2020).[7] In this sense, the findings of this paper are not confounded by the possible short-run effect of the amnesty policy.[8]

4 Data

I use two main datasets in this paper. The first one is the Labor 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 240.000 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 webpage.

To begin with, DANE implemented a migration module in the GEIH of 2012, then in 2013-II

[6] 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.

[7] 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 of migrants of getting offers to be formally employed.

[8] Compared to Ecuador, Olivieri et al. (2020) find using simulations that the provision of work permits to Venezuelan workers would increase their average earnings.

improved the questionnaire by adding questions on place of birth and, finally, after 2015 DANE removed an initial filter question on the residence.[9] This module contains questions on where the person was born, where the person lived 12 and 60 months ago and reasons for 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. Besides, as the census applies the same migration module, it allows 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 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 the distance instrument I use. 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 the 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, importantly as noted earlier the majority enter the country walking with all their belongings.

4.1 Descriptive Statistics for Natives and Migrants

For the descriptive statistics, I differentiate three main groups of interest. The first one is of native Colombians residing permanently in Colombia. The second one is of Venezuelans who emigrated

[9] In effect, in the beginning, the module was only answered by the people that were born in other cities different from the one they are currently living in, but from 2016 onwards it was answered by all the respondents of the survey.

to Colombia in the last year. Finally, the third group corresponds to Colombians who resided in Venezuela and then returned to Colombia when the crisis started. For the causal inference, I focus only on the group of Colombians who did not migrate in the previous year from Venezuela, as the sample size for the other two groups, specially when you split by region-year cell, turns to be very small. With this in mind, in 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 in Colombia tend to be young, though their age seems to be steadily growing: prior to 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, skills. One relevant takeaway is that arrivals of Venezuelans seem to be more educated in the latest years. In terms of gender composition, there is no unbalance, with the shares of both men and women being similar. Finally, prior 2016 returning Colombians were the main group coming from Venezuela, but afterwards, Venezuelans surpassed greatly this group and became the predominant immigrant group.

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".[10] 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

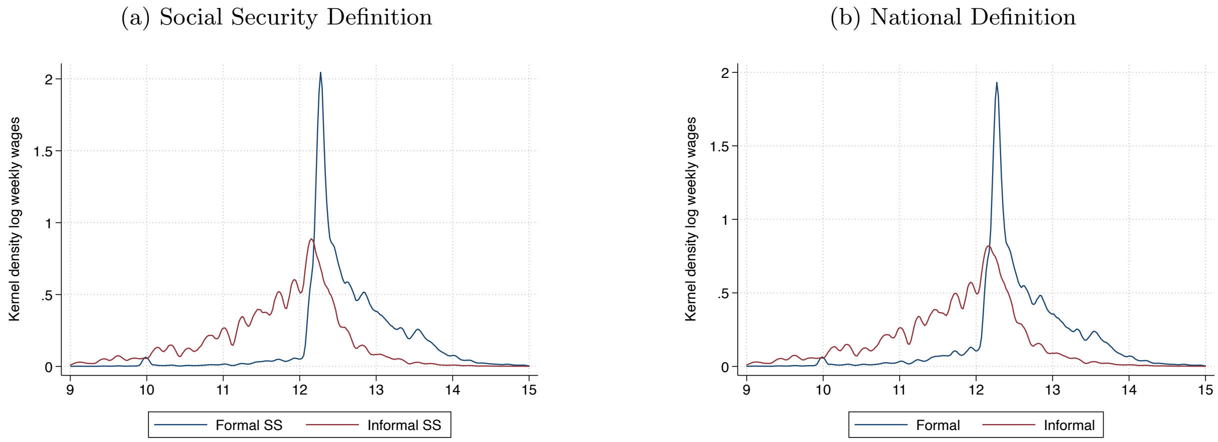
[10] In Colombia, not necessarily firms that employ on the informal sector are illegal (i.e., not paying taxes) or not formally registered in the state agencies. There are crossing definitions of informality depending on the side you focus on (firm or worker). Throughout this paper, I focus on the side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

(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 the two margins is needed. According to Ulyssea (2018) there is a *extensive* margin that represents firms that 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 can describe two main definitions of labor informality. The first one is the national definition based on firm size and occupation, 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.[11] 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 on whether the worker has access to health insurance or contribute to the pension system, 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 workers in the formal sector poses a binding restriction, this "stickiness" helps to explain why there cannot be wage losses for formal workers. Lastly, the Colombian labor market has a substantive portion of non-salaried workers (self-employed), though the wage analysis throughout the paper covers all type of labor income, not just salaried workers wages.

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

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



Note: All the data on wages is stacked across periods and departments and then is plotted. The 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 period. The raw data also indicates that almost all the new arrivals of Venezuelans are being employed in 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). 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)

	LFP	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)

	LFP	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: LFP stands for Labor Force Participation. 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 Venezuelans workers have a job (46.9%), while the corresponding share for Colombians workers is nearly 1/3 (≈ 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 between migrants and natives. To do that, I regress the log hourly real wage on a set of control variables including a dummy of birthplace.[12] On average, Colombian workers earn 0.29 log points higher wages than its Venezuelan counterparts

[12] The wage gap is calculated in an unweighted regression of log hourly real wages on the dummy of the place of birth, plus two polynomials of age, schooling, gender, the 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).

(Table 2). Some unobservables that can help to explain the gap between both groups is the missing work experience of migrants in their home country. One key aspect of this immigration event is that 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 Choice of Base Period

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.[13] 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

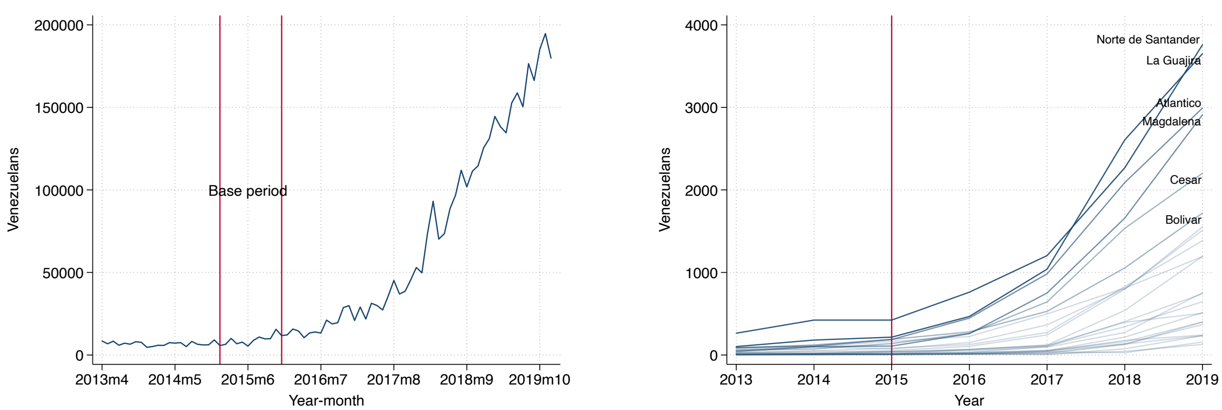
[13]The GEIH survey question asks the respondents where they were born, if in the same municipality they are residing, or in others from Colombia. If it is from another country, they ask in which Country.

simplicity, and to remove seasonal effects, I select as a base period of comparison the year 2015, which was the last year before the huge increase of Venezuelans in Colombia. In this respect, it should be stressed that the number of immigrants does not seem to have reached a peak yet.[14] 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 a small number of Venezuelans in the pre-treatment years, followed by a staggering increase which varies greatly in the 24 departments sample (Figure 2b), motivating the spatial approach undertaken here. Note that, using the static TWFE (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 this setup, to study the effect year-by-year in an event study design.

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.

[14]Nevertheless with the COVID-19 pandemic significant number of Venezuelans are returning 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.[15] 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 of the comparison of groups, within a defined area, before and after 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 there was a not clear inflow effect of natives not coming to the most affected areas (a downward trend is noticeable but only significant in 2019), but more an outflow effect of natives leaving the most immigrant affected areas, relative to the base period (see Figure A.6 in Appendix). However, since this approach is suited for the short-run analysis, when capital is fixed, given mobility responses does not posses a main problem on the identification of the overall effect of immigration.

In terms of empirical specification, my preferred one has the interpretation of an event study one as I select a base year of comparison 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 2015,

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

and $M_{d,2018}$ is a time-invariant treatment variable, constructed from the 2018 census records. Note that by construction $\beta_{2015} = 0$ and all coefficients β_t measure the effect relative to 2015. It

[15] When restricting the analysis to capital cities only, results are slightly similar with higher coefficients (in absolute numbers).

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 remain constant over time, with nearly perfect correlation across different years, while the 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 β_{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

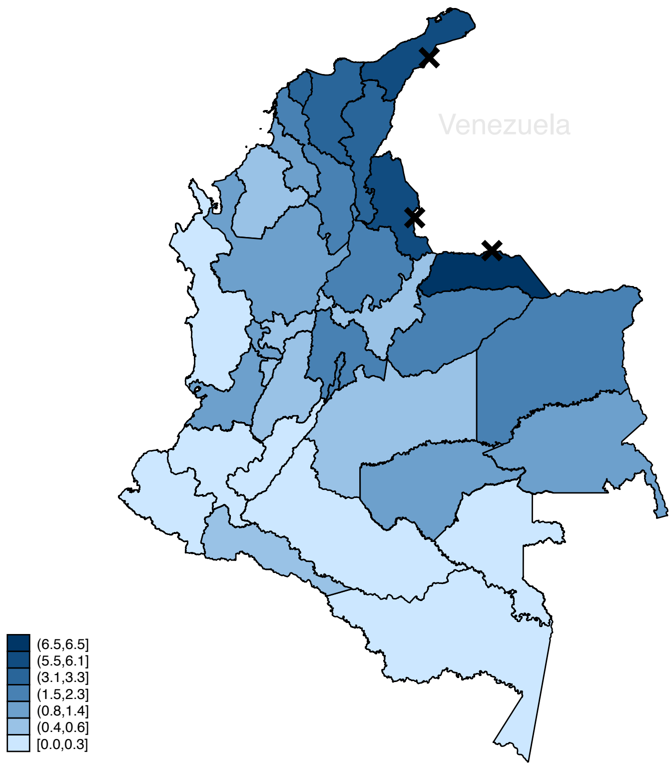
$$M_{d,2018} = \frac{L_{Ven,d,2018} - L_{Ven,d,2015}}{L_{Total,d,2018}} * 100 \tag{2}$$

where the numerator is the total number of employed Venezuelans (between 18 and 64 years) in department d who arrived in Colombia in the previous 5 years, starting from 2018, minus the 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 survey, not from panel data.

Figure 3 plots the Colombian map along with the migration rate $M_{d,2018}$ by departments.[16] Not surprisingly, the highest migration rates are observed in those areas which are closer to the Venezuelan frontier, and especially 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 2.8% of the total population in Colombia according to the 2018 census.

[16]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: Migration rate $M_{d,2018}$ is plotted. 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 setup, I use clustered standard errors as there could be arbitrary serial correlation of the errors within each department. Also, I do not use department sample size weights in the main specification because the heteroskedasticity test proposed by Solon, Haider, and Wooldridge (2015), which consists in regressing the square residuals on the inverse sample size, gives a insignificant estimate.[17] One clear downturn of this setup is the small number of treated units ($N = 24$), which can increase Type I error rate considerably (Pustejovsky and Tipton, 2018). For that reason, I also control for the over-rejection of the null hypothesis implementing the wild cluster bootstrap method of Roodman et al. (2019), and reporting the p-value for the main estimates. Moreover, I build a sample of smaller local labor markets ($N = 51$), to find that results are, if anything, more

[17] Another use of regression weights in applied economics is to identify population average partial effects. However Solon, Haider, and Wooldridge (2015) state that this is not so accurate, pointing arguments on both ways: for the use or not of weights. In any case, when using regression weights I find similar estimates for native employment and higher for native wages, still both negative.

negative compared to the department sample ($N = 24$). Showing that negative results are not driven by the sample size.[18]

5.2 Event Study with Time-varying Migration Rate

Given 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 GEIH survey as a robustness check. Yet, a potential caveat of this approach is the small variability of migration before 2017 (as shown in Figure 2b) which could widen the standard errors. For this and other reasons mentioned above, inference mainly comes from $M_{d,2018}$.

In view of this discussion, equation (1) can be rewritten into a differences regression, where I interchange $M_{d,2018}$ from census with M_{dt} from GEIH survey:

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

Note that when M_{dt} is replaced by $M_{d,2018}$ the coefficient of interest in equation (1) and (3) is identical ($\beta_t = \theta_t$). The definition of the time-varying migration rate is as follows

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

where 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). The denominator in turn is fixed on the base year following the definition in Card and Peri (2016). For instance, if the outcome of interest is $log(wages)$, then the estimated coefficient of interest becomes

$$\hat{\theta}_t = \frac{\hat{Cov}(M_{dt}, \Delta log(w_{dt}))}{\hat{Var}(M_{dt})} \tag{5}$$

where $log(w_{dt}) - log(w_{d,2015}) = \Delta log(w_{dt})$.[19] Then, plugging model (3) in the last expression yields

---

[18] Note that increasing the sample size might be problematic due to measurement error in both sides of equation (1), as GEIH is not representative for those 51 areas, only for the 24 departments used.

[19] The resulting expression of $\hat{\theta}_t$ 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.

$$\hat{\theta}_t = \theta_t + \frac{\hat{Cov}(M_{dt}, \Delta u_{dt})}{\hat{Var}(M_{dt})} \tag{6}$$

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., $E[M_{dt}\Delta u_{dt}] \neq 0$) that change over time. For instance, it could be the case that economic conditions, relative to base period, are correlated with immigration inflows.[20]

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 plausible 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),

$$z_{1,d} = \sum_s \underbrace{(\lambda_s / T_{s,d})}_{share} * \overbrace{M_{2018}}^{shift} \tag{7}$$

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 λ_s is the share of Venezuelans that emigrate from s according to RAMV. Whereas M_{2018} are arrivals of Venezuelans to Colombia in 2018. Note that there can be only one national shift, thus exogeneity of $z_{1,d}$ needs to arise from the shares.

The use of the distance instrument $z_{1,d}$ is motivated by the fact that Colombia and Venezuela share more than 2.000 kilometers of terrestrial borders. Therefore, new arrivals M_{2018} to location d are determined by the travel distance from city x to city y, in the sense that travel distance

[20] Although the coefficient in equation (1) and (3) is the same when using a time-invariant treatment variable, clustered standard errors in (1) are going to be less or equal. This is because regression (1) group observations of units over time -allowing to control for arbitrary serial correlation of errors- while regression (3) gives first difference errors with respect to a specific time interval for every unit. With this in mind, in (1) I use clustered standard errors while in (3) is only possible to use robust standard errors. In any case, the difference in the size of standard errors from (1) and (3) is quite low.

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 less trade, than the counterpart far-located states (violation of the exclusion restriction). For that reason, I show that when including a control for changes in the business cycle in departments (state GDP), the results for wages are slightly smaller but remain significant, still this can be a "bad control" (Angrist and Pischke, 2008, p. 49). In addition, I show that when including trade patterns with Venezuela (measured as the share of total exports in USD to Venezuela over total exports to the world from DANE in 2015) results are higher and significant. Thus the inclusion of previous controls does not alter drastically the coefficient of interest.

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

$$z_{2,d} = \underbrace{\frac{1}{L_{d,2015}} * \frac{Ven_{d,2005}}{Ven_{2005}}}_{share} * \overbrace{M_{2018}}^{shift} \tag{8}$$

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 $L_{d,2015}$ in d at the base period, whereas M_{2018} are Venezuelan arrivals to Colombia in 2018.

The validity of the past settlement instrument $z_{2,d}$ relies on the fact that new arrivals M_{2018} 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, which goes back to 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 and therefore captures previous immigration adjustments (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 past settlement 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 shifts (Borusyak, Hull, and Jaravel, 2018), in my case, the identification can only rely on the exogeneity of the selected shares, not on the aggregate shift, to have valid instruments (i.e., $E[z_{i,d}\Delta u_{dt}] = 0$ $i = 1, 2$).

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,

$$M_{d,2018} = \xi_i + \eta_i z_{i,d} + \upsilon_{i,d} \quad i = 1, 2 \tag{9}$$

where $\upsilon_{i,d}$ 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 49.7%.

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

	(1) $M_{d,2018}$	(2) $M_{d,2018}$
Distance ($z_{1,d}$)	0.00376***	
	(0.000299)	
Past settlement ($z_{2,d}$)		34.32***
		(5.757)
Constant	-1.271***	0.740**
	(0.233)	(0.225)
N	24	24
R^2	0.884	0.497
F st	157.9	35.54

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 instruments are time-invariant, the first stage is the same in all the years analyzed.

6 Results for Natives

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 (1) for the log mean 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.[21] Note that, despite being non significant, the point estimate of OLS indicates that immigrants are going to areas with rising wages, though this selection gets more or less corrected with the IV estimates.[22] The influx of Venezuelan immigrants, as standard economic theory predicts, has a negative effect on native wages for both methods (Figure 4). The OLS estimates are all negative and, even if they could be upward biased due to omitted variables, they do not differ much from the IV ones due to the high R^2 of the First-Stage regression (see Table 3). Moreover, when using the two instruments defined before separately, the results are again negative and significant.[23] Since the outcome is in logarithms I multiply the coefficient times 100. Thus, for 2018, the year I defined the migration rate from the census, a 1 p.p increase in the share of employed Venezuelans decreases the wages of natives by 1.7%, with the past settlement instrument, and by 1.6%, with the distance instrument.[24] Note that both point-estimates are significant using a wild cluster bootstrap method (see Table 4). Moreover, I show in Appendix that, when using residual wages (having controlled for individual characteristics as age, years of schooling and gender) instead of observed wages, the results remain similar.

Scaling up these estimates, the total shock according to the census is about 1.7 p.p of the

[21] A joint F-test cannot reject the null hypothesis of both pre-treatment coefficient equal to zero with p-values around 0.34 and 0.88, depending on the method used.

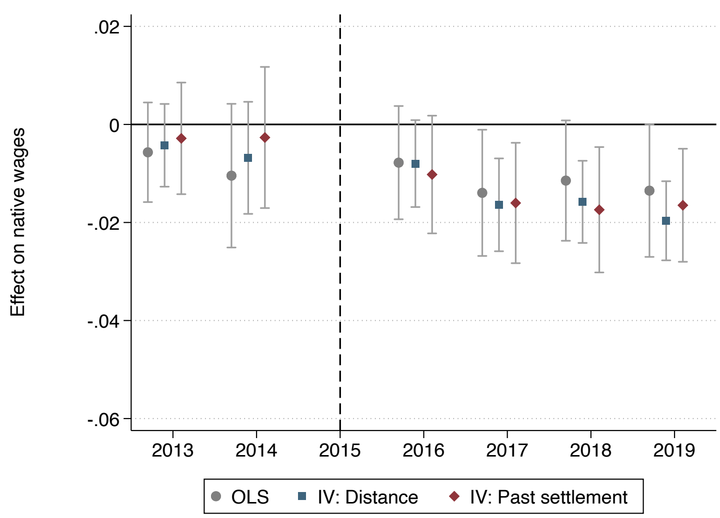
[22] The construction of the wage (or labor income) variable requires some steps to construct it, in Appendix I explain in detail what are the steps required to have a more precise measure of wages. Which covers all types of labor income, including the self-employed wages.

[23] I do not combine both instruments in the first stage since the distance instrument captures all the predictive power of the past settlement instrument, as its coefficient turns to be insignificant.

[24] Caruso, Canon, and Mueller (2019) find a point-estimate of the Venezuelan immigration on hourly wages of 7.6% for a 1 p.p increase in their migration rate using IV panel-data regression. The differences with my estimates mainly arise from 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.

employed population (in absolute numbers ≈ 254,000 employed Venezuelans), and hence the total impact on wages, for 2018, is between 2.7%-3% depending on the instrument selected.[25] To interpret the wage response, since native local wages increased by nearly 2.1% in real terms between 2015 and 2018, the immigration negative effect implies a decline in natives' real wages. Yet, it should be 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.[26]

Figure 4: **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. 95% Confidence Interval. Explanatory variable is $M_{d,2018}$. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9 and for past settlement instrument is 35.54. Hourly wages are in real terms using the monthly CPI from DANE.

Second, I use the time-varying migration rate M_{dt} built from GEIH survey as the explanatory

[25]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 35.54.

[26]I do not include explicitly in the migration rate returning Colombians due to the different observable and unobservable characteristics with the Venezuelan migrants. For instance, returning Colombians are on average less educated and older, besides, they are entitled to work permits, can have stronger network effects or job offers, and probably suffer less wage discrimination. In addition, returning Colombians were concentrated in the years prior 2016 (close to my pre-treatment period), while after 2016 the main group of immigrants, in absolute numbers, were Venezuelans (see Table A.1c).

variable. Interestingly, estimates follow a similar pattern (negative and significant) but different in magnitude from those obtained with the fixed $M_{d,2018}$ from the census. For instance, in 2017 I observe a higher negative effect, but with wider confidence intervals, probably due to the low variability of migration in 2017 vs. 2015 between departments. By contrast, in the census year, estimates of the effect are similar as before, ranging between 1.4%-1.6% (see Figure A.8a in Appendix). Moreover, the correlation of the yearly migration rate M_{dt} from GEIH with other post-treatment years $M_{dt,t\neq 2018}$ is between 0.93-0.98, highlighting the fact that when using the constant $M_{d,2018}$ from census I am still capturing the dynamics for other years, as I find nearly perfect correlation in every t.

The insight for such a high negative finding relies on several factors pointing to high substitutability of natives and migrants. In effect, migrants speak the same language as 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 without certified education or home experience (which implies downgrading of tasks); and finally, wage flexibility in the informal sector can lead to large wage cuts when migrants arrive.

6.2 Employment Responses

Regarding the employment effects of immigration, the first result to highlight is that, opposite to wages, there seems to be significant differences in the employment trends before the migration event happened (Figure 5a).[27] 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 UPTA with IV required for the identification of the causal parameter. The fact that both 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), whereas the outcome variable is the change in log (total native employment) relative to base period. Figure 5b depicts the the pre-treatment coefficients and pre-trends in 2014 are no longer significant for both

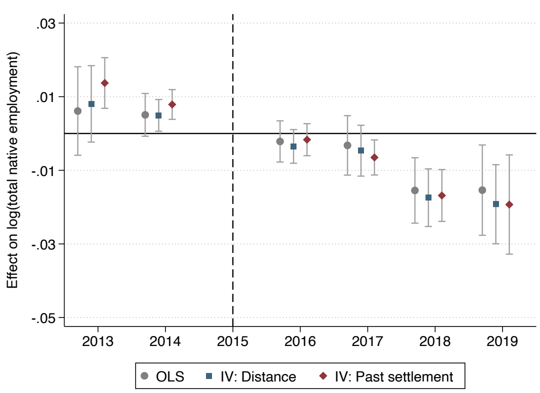
[27] A joint F-test rejects the null hypothesis of both pre-treatment coefficient equal to zero for the distance instrument (p-value=0.08, thus only at 10% significance level) and past settlement instrument (p-value=0.00), while for OLS I cannot reject it (p-value=0.22).

instruments.

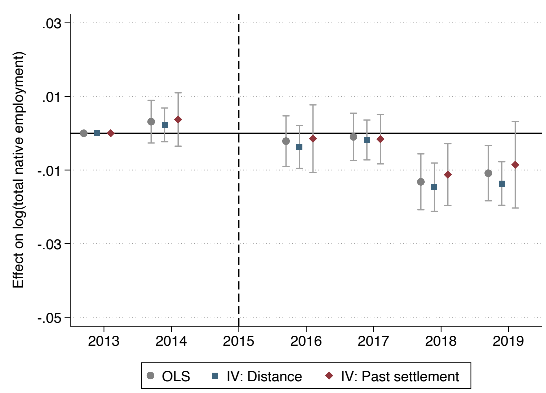
Table 4 shows the results for employment and wages for natives in 2018, relative to the base period[28]. For the interpretation of the employment impact, a 1 p.p increase in the migration rate reduces on average 1.5% local employed natives in 2018 relative to 2015 using past settlement as the instrument and by 1.1% using distance as the instrument (coefficients of Table 4-Column 4). Note that the distance instrument coefficient is the only significant according to the wild cluster bootstrap method.[29]

Figure 5: **Event study estimates on log employed Colombians**

(a) Employment levels



(b) Employment levels with trend-adjustment



Note: Dependent variable is log native employment relative to base period. Departments in the regression are $N = 24$ per year. 95% Confidence Interval. The explanatory variable is $M_{d,2018}$. Trend-adjustment estimates have as a control the growth in employment from 2013 to 2015. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9 and for past settlement instrument is 35.54.

[28] Recall that, since wages did not present pre-trends, I did not use trend-adjusted estimates in the previous section

[29] In Figure A.8b in Appendix I show that when using M_{dt} as the 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
Panel A: OLS				
$M_{d,2018}$	-0.0115	-0.00962	-0.0155***	-0.0132**
	(0.00540)	(0.00595)	(0.00391)	(0.00334)
Wild cluster bootstrap p-value	.147	.222	.01	.028
Panel B: IV				
Distance instrument				
$\hat{M}_{d,2018}$	-0.0158**	-0.0146*	-0.0174***	-0.0147**
	(0.00427)	(0.00465)	(0.00399)	(0.00335)
Wild cluster bootstrap p-value	.03	.052	.008	.02
First stage: F st	157.9	228.3	157.9	152.6
Past settlement instrument				
$\hat{M}_{d,2018}$	-0.0174**	-0.0167	-0.0168	-0.0113
	(0.00653)	(0.00612)	(0.00359)	(0.00430)
Wild cluster bootstrap p-value	.032	.184	.194	.280
First stage: F st	35.54	55.16	35.54	19.14
Trend-adjusted	No	Yes	No	Yes
N	24	24	24	24

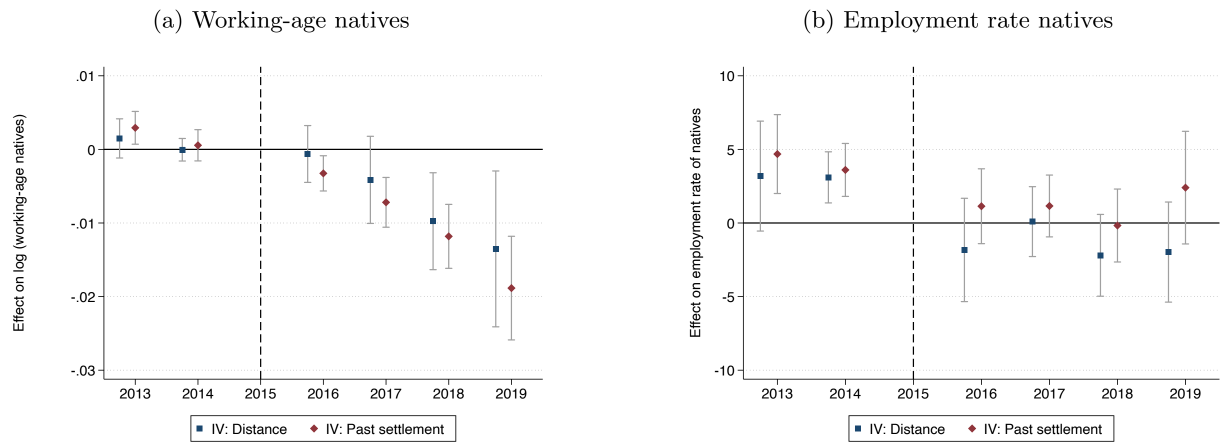
Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. The outcomes are in logarithms, thus the coefficients are interpreted as percentages. Department sampling weights from GEIH are used to construct aggregate outcomes. 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. Wild bootstrap p-values are computed from *boottest* command using 999 bootstrap replications.

The loss in native employment (in levels) is part of the impact on overall working-age natives (i.e., the sum of the unemployed, employed and inactive population). In Figure 6a, the dependent variable is the change in log working-age natives relative to base period. For 2018, the census year, I find that a 1 p.p increase in the migration rate $M_{d,2018}$ decreases on average the number of working-age Colombians in the local area by 1%-1.2% depending on the instrument used (wild cluster bootstrap p-value is 0.042 and 0.022, respectively). Such a significant migration response is an important margin of adjustment of Colombian natives to the Venezuelan immigration. This might dampen the spatial approach analysis, but as Venezuelan arrivals are so sharp, the spatial approach manages to capture the overall effect in the short-run, not equal for the long-run, when capital can adjust. To conclude, if employed and working-age natives are decreasing, the effect on the employment rate should be small (as both the numerator and denominator of it are decreasing).

In Figure 6b, I show the impact on the native employment rate and, accordingly, the effect is insignificant.

Figure 6: **Event study estimates on log working-age and employment rate of Colombians**



Note: Dependent variable in (a) is log working age natives and in (b) is employment rate of natives, both relative to base period. Departments in the regression are $N = 24$ per year. 95% Confidence Interval. The explanatory variable is $M_{d,2018}$. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients.

6.3 Results with a Larger Sample

As mentioned earlier, one potential shortcoming of this setup is the small number of treated areas per year ($N = 24$). To address this problem, I build a larger comparison sample of Functional Urban Areas (FUAs) in Colombia using Sanchez-Serra (2016) methodology. This sample consists of the 53 biggest urban areas in the country defined from population grid data, municipal boundaries and inter-municipal commuting flows. However, since there are not representative surveys or administrative data at this level with yearly information of wages or employment, I use GEIH municipal surveys to construct the outcomes, noting that sampling error can increase considerably as GEIH is not meant to be representative for these areas.[30]

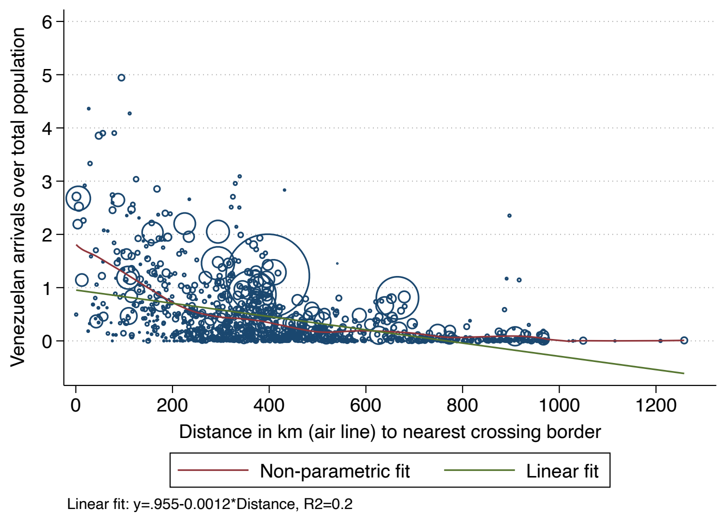
In this exercise, I compare the results for 24 departments and 51 FUAs that are available in GEIH. [31] Moreover, I use a slightly different distance instrument from before since I expand the

[30] This municipal location variable of the survey is not publicly available.

[31] Cities from the Amazonia or Orinoquia region are not taken into account in this paper. Recall, that the population of those regions is around 2.8% of the total population in Colombia.

sample. I now use the second polynomial of the distance from given area to the nearest crossing bridge with Venezuela. As a preamble, I find a strong positive relationship between immigrant arrivals in municipalities and distance to nearest crossing border using the census immigration inflows, that captures all the municipalities in the country ($N = 1122$) -see Figure 7-. In terms of native results, when focusing on more detailed local labor markets (the FUAs) coefficients tend to be more negative, specially for wages. For the 24 departments I find a wage estimate of -1.4% while for the 51 FUAs the corresponding estimate is -2.1%.[32] Overall, the estimates I found are more conservative, but the main results hold when increasing the sample of analysis.[33] Thus, as standard errors are quite similar in both cases and the sampling error is much higher for FUAs, I focus on the sample of the 24 representative departments in what follows.

Figure 7: **Arrivals in preceding year from Venezuela by municipalities vs distance to crossing bridge**



Note: Municipalities are $N = 1121$ as one municipality coordinates are not available. 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.

[32] The result of departments is slightly different than before because I am using a distinct instrument with the migration rate from GEIH, not the census, to have a better comparison with the FUAs sample.

[33] The employment increase in the larger sample can be explained by the fact that FUAs can capture more detailed geographic movements from natives, which can be masked in the department analysis, as documented for the US in **?**.

Table 5: **Native wages and employment estimates for different samples, 2015-2018**

	(1) Native Wages	(2) Native Employment
Panel A: Departments		
$M_{d,2018}$ (IV: Distance to nearest crossing)	-0.0142**	-0.0141***
	(0.00547)	(0.00369)
F st	16.27	14.93
N	24	24
Panel B: FUAs		
$M_{d,2018}$ (IV: Distance to nearest crossing)	-0.0211***	-0.0185**
	(0.00540)	(0.00620)
F st	14.31	13.30
N	51	51
Trend-adjusted	No	Yes

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 predicted migration rate from GEIH survey with the distance instrument. The instrument is the second polynomial of the distance to nearest crossing border with Venezuela. The outcome is the difference in 2018 with the base period. The 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. The variables are in logarithms, thus the coefficients are interpreted as percentages. Hourly wages are in real terms using the monthly CPI from DANE.

7 Labor Market Linkages Between Formal and Informal Sector

In this section, I first document the differential response of wages and employment in the informal and formal sector in the presence of a labor supply shock, as those responses are mirror images of each other. Then I show some results that motivate how these two sectors can interact. Lastly, I introduce a theoretical model to interpret these results.

7.1 Impact on Natives

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 an immigration shock. In this section I report results using the distance instrument for expositional clarity but, before that, I to test for the reliability of the survey data by comparing its results with the administrative data from the Colombian Social Security Records (PILA, by its acronym in Spanish). In Appendix Figure A.1, I show that in terms of total employment the two sources are very similar. The same happens in terms of results (see

Figure A.2a and Figure A.2b). Reassuringly, the results coming from GEIH survey are robust.[34]

Figure 8a plots the estimated effect of immigration for formally and informally employed native wages.[35] Not surprisingly, informally employed workers suffer the largest wage losses (2018 coefficient is -1.9% with wild bootstrap p-value of 0.05), while formally employed ones have insignificant estimates.[36] These findings differ from those found for Turkey, a country with high informal levels of employment like Colombia, where 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. Now, with regard to the effect on employment for the Colombian context, formal and informal workers suffer a negative significant effect after 2017, yet the impact is more severe for those employed in the the formal sector. The role of the minimum wage can help to explain these findings, since formal workers have a high probability mass to the right of the minimum wage (see Figure 1a) there cannot be further wage drops, and this translates into higher employment losses, since informal wages can freely adjust and formal/informal workers are substitutes.

[34]Still, due to measurement issues of PILA, that I explain with more details in the Appendix A, I only use this dataset for comparison purposes.

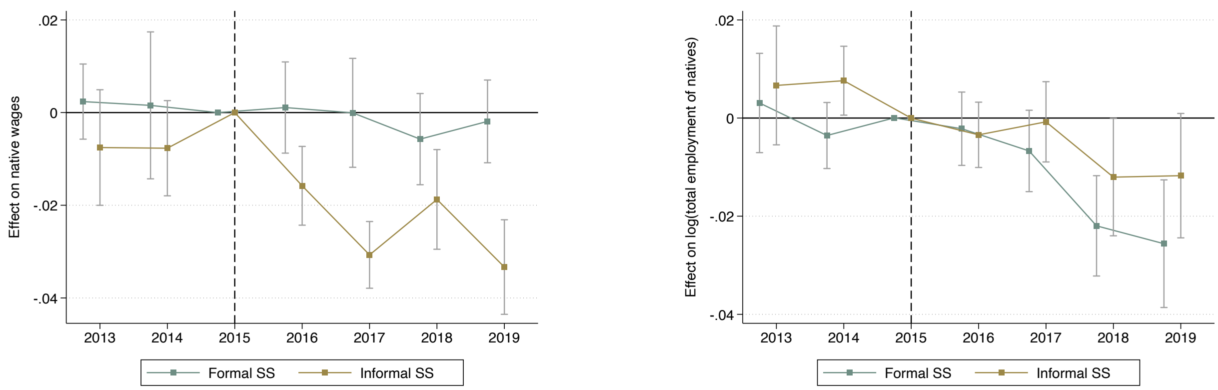
[35]I use the definition of informality according to the affiliation to social security (using the national definition of informality yields closely the same results).

[36]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.

Figure 8: **Event study estimates by affiliation to Social Security (SS)**

(a) IV (distance) estimates on native wages

(b) IV (distance) estimates on native employment

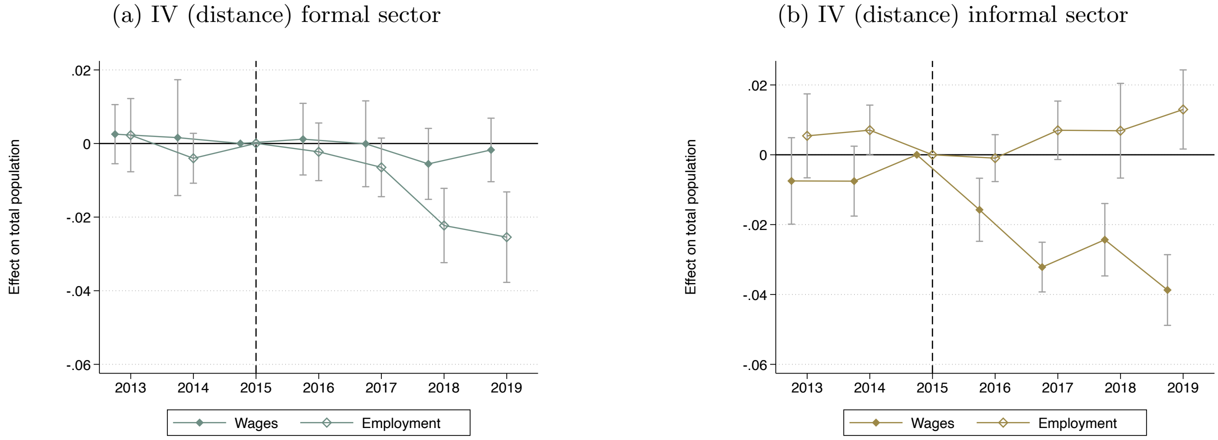


Note: Dependent variable is log native employment relative to base period. Departments in the regression are $N = 24$ per year. 95% Confidence Interval. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. I use the definition of affiliation to Social Security (SS). In (b) no controls for pre-trends are used. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

7.2 Overall Market Responses

Next, I document that the labor supply shock affected mainly the informal labor market (see Figure 9a and 9b). To do so, I focus on overall employment and total wages using the definition of informality by affiliation to social security. The main takeaway is that I find insignificant effects for formal wages while for informal wages I find negative effects. In addition, formal employment decreased significantly only after 2017 while informal employment increased. In the next subsection, I develop a model to explain these contrasting findings on wages and employment for these two markets. In a few words, suppose that the informal labor market behaves 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. By contrast, the minimum wage acts as a downwardly wage rigidity in the formal sector, which can induce reductions on formal employment, given the relatively cheaper informal labor.

Figure 9: **Employment and wage estimates for total population by sector**



Note: Departments in the regression are $N = 24$ per year. Informality definition is given by affiliation to social security. I do not explicitly control for pre-trends. 95% Confidence Interval. The explanatory variable is $M_{d,2018}$. Department sampling weights from GEIH are used to construct aggregate outcomes. The outcome is the difference in 2018 with the base period. The sample is restricted to respondents between 18 and 64 years in urban areas. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

7.3 Interaction of Formal and Informal Employment

To explain why formal employment can be reduced in the presence of a labor supply shock I exploit some firm characteristics from GEIH survey. First, in Table 6 I show suggestive evidence that workers without social security are employed mainly on the smallest firms, still not all workers in a given firm size are informal, there are also formal ones too. In principle, there could be a combination of formal and informal labor inputs for a given firm production function. This combination of labor depend mainly on its size, as documented for Brazil (Ulyssea, 2018) or Mexico **?**. Unfortunately, due to the nature of the GEIH survey, it is not possible to identify the individual firm the employee works, thus the information is at the aggregate level. In any case, the results in Table 6 show that, of all workers in the GEIH survey that state to work alone, 89.8% were informal, while among all workers who declare to work in a firm with more than 100 workers only 5.6% were informal.[37]

[37] Again, informal according to the definition of affiliation to social security.

Table 6: **Share of workers without access to social security by firm size**

Firm Size	Share of Informality (%)
1 worker	0.898
2 to 3 workers	0.857
4 to 5 workers	0.723
6 to 10 workers	0.550
11 to 19 workers	0.389
20 to 30 workers	0.250
31 to 50 workers	0.167
51 to 100 workers	0.116
101 or more workers	0.056

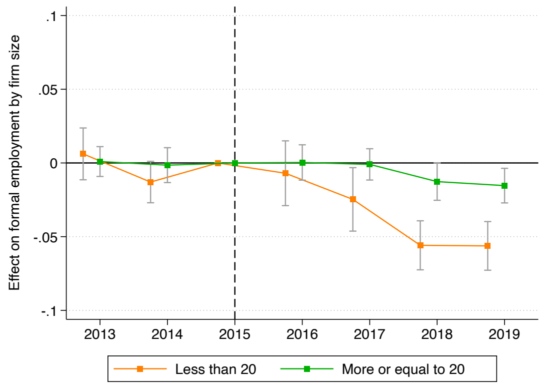
Note: Shares are calculated using national sampling weights from GEIH. The sample is restricted to all workers from ages between 18 and 64 years in urban areas. Source: GEIH, 2015.

With this information in mind, I aggregate in two categories of firm size: less than 20 workers (smaller firms) and more or equal to 20 workers (bigger firms). As shown in Figure 10a, which plots the change in log (total formal) employment with respect to base period (2015), smaller firms significantly reduce formal employment in response to the immigration shock. Moreover, in Figure 10b I analyze the change in the share of informal workers for a given firm size with the base period, to show that in these smaller firms the share of informal workers increase more pronouncedly than in bigger firms.[38] Therefore, smaller firms seems to be replacing some of their (expensive) formal workers with (now cheaper) informal workers.

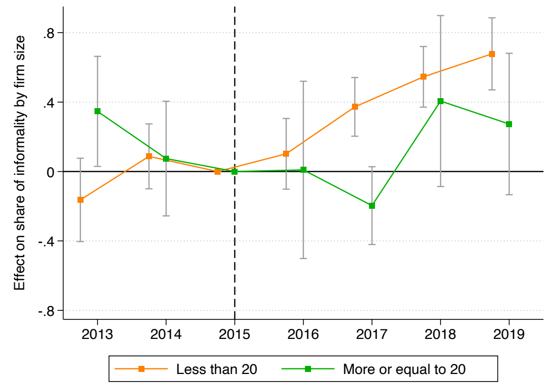
[38]Note that informal employment significantly increase on smaller firms but not as much to counteract the formal employment decrease, so the increase in the share of informality is driven by a reduction of formal workers.

Figure 10: **Event study estimates for formal employment and share of informal workers by firm size**

(a) IV (distance) formal sector



(b) IV (distance) share of informality



Note: Departments in the regression are $N = 24$ per year. Informality definition is given by affiliation to social security. I do not explicitly control for pre-trends. 95% Confidence Interval. The explanatory variable is $M_{d,2018}$. Department sampling weights from GEIH are used to construct aggregate outcomes. The sample is restricted to respondents between 18 and 64 years in urban areas. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9.

Next, I check whether the previous results are driven by the entry of new firms or by a rise in the number of workers in existing firms.[39] First, as regards firm entry, in Figure 11a I find a positive effect on firm creation in 2016, that becomes insignificant in 2017 and 2018.[40] Thus, I conclude that employment changes by firm size are less driven by the entry of new firms (registered on tax records).[41] For the second one, I construct a proxy of the average number of workers that report to work in a given firm size. In this case, I find that for smaller firms there seems to be a decrease in the average number of workers while in larger firms it appears to increase. Overall, these findings support the fact that employment changes are mainly due to changes in the workforce of smaller and larger firms (see Figure 11b).

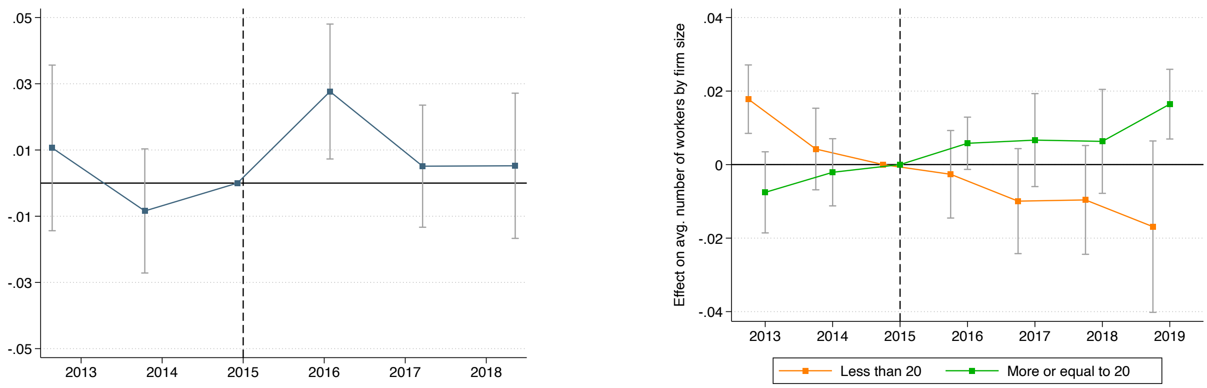
[39]For the first one, I use information from *Confecamaras*, which collects all the new registered firms within Colombia, then I construct the log cumulative number of newly registered firms for every year and department.

[40]Recall that this effect only pertains to registered firms that have a tax record, restricting highly the universe of new firms, especially in a country with high levels of informal businesses (i.e., without legal registration) like Colombia.

[41]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 11: **Event study estimates on cumulative firm creation and average category of workers by firm size**

(a) IV (distance) on firm creation (b) IV (distance) on avg. category of workers



Note: Departments in the regression are $N = 24$ per year. The sample is restricted to registered, or formal, new firms that are in the databases of the corresponding state agencies. Outcome for (a) is the logarithm of new firms relative to new firms in the base period. Outcome (b) is the average change in category of firm size. The explanatory variable is $M_{d,2018}$. 95% Confidence Interval. The base period is 2015. β_t from equation (1) are the plotted coefficients. F-statistic for distance instrument is 157.9. Source: (a) Confecamaras, 2013-2018. (b) GEIH, 2013-2019.

7.4 Theoretical Framework

To aid the interpretation of the empirical findings I propose a model inspired by Ulyssea (2018) with two labor inputs where the key distinction arises from the differential cost of labor. More concretely, a firm can hire workers L_f paying the official payroll taxes, but also can hire workers L_i "off the books" to avoid complying with the contributions to the social security system.[42]

The profit function of the firm can be written as

$$\max_{L_i, L_f} \pi = pF(L_i, L_f) - \tau(L_i) w_i L_i - (1 + \tau_f) w_f L_f \tag{10}$$

where $\tau(L_i)$ represents s a convex cost that is increasing on informal labor size within the firm (i.e., $\tau'(L_i), \tau''(L_i) > 0$). In particular, it is assumed that $\tau(L_i) = L_i^{\eta}$ with $\eta = 0, 1..., N$, which captures the cost of evasion related to law enforcement exerted by the government. Whereas τ_f

[42] In this model, I abstract from the extensive margin followed in Ulyssea (2018), to take into account only the labor choices of a given formal firm (the intensive margin). That means I do not take into account the decision of the firm to register or not in the tax records (become a formal firm), as my goal is to model informality through the worker side and analyze changes within a given firm. Moreover, this setup could have firms that hire only informal or formal labor, but then the interest will be in the between-firms effect, not the within-firm effect.

represents the payroll taxes that the firm has to enroll when paying for formal workers, similar to Ulyssea (2018) cost function.

Specifically, the production function has a Constant Elasticity of Substitution (CES) form,

$$F(L_i, L_f) = Q = (\alpha_i L_i^{\rho} + \alpha_f L_f^{\rho})^{\frac{1}{\rho}} \tag{11}$$

where $\sigma = \frac{1}{1-p}$ (with $\rho \leq 1$) is the elasticity of substitution between formal and informal workers, and productivity parameters are standardized such that $\alpha_i + \alpha_f = 1$. Moreover, firms set the aggregate output price p according to an (inverse) demand function: $p = C^{1-\epsilon}Q^{-(1-\epsilon)}$ as in Borjas (2013), where C is the number of consumers, and $\epsilon^D = 1/(1-\epsilon)$ is the price elasticity of demand in absolute value.[43]

In this setup, the above maximization problem imply that the market wages satisfy:

$$(\tau'(L_i)L_i + \tau(L_i))w_i = C^{1-\epsilon}\epsilon\alpha_i L_i^{\rho-1}(Q)^{\epsilon-\rho} \tag{12}$$

$$(1 + \tau_f)w_f = C^{1-\epsilon}\epsilon\alpha_f L_f^{\rho-1}(Q)^{\epsilon-\rho} \tag{13}$$

Under the assumptions that the firm is competitive in the labor market, and therefore takes prices w_i and w_f as given. I then analyze the change in optimal labor choices within the firm when a lower wage for informal workers takes place as a result of the migration inflow. After some algebraic derivations in Appendix A, the elasticities of informal and formal labor with respect to informal wages (ε_{L_i,w_i} and ε_{L_f,w_i}) are equal to:[44]

$$\varepsilon_{L_i,w_i} = -\frac{(1 - \epsilon s_f - \rho s_i)}{(1-\epsilon)(1-\rho) + \eta(1 - \epsilon s_f - \rho s_i)} \tag{14}$$

$$\varepsilon_{L_f,w_i} = -\frac{s_i(\epsilon - \rho)}{(1-\epsilon)(1-\rho) + \eta(1 - \epsilon s_f - \rho s_i)} \tag{15}$$

[43]In this case, I assume for simplicity, that the number of consumers grows at the same rate as the workforce, that is, what Borjas (2013) defines as *product market neutrality*.

[44]Another mechanism to be studied is the effect on firm formation that are not registered on tax records (informal firms), I have shown previously the empirical effect on firm creation of formal firms and it seems quite low. It might be that informal firms are growing in higher immigration areas.

First, note that the denominator in both expressions is always non negative as $\epsilon \leq 1$, $\rho \leq 1$ and $\eta = 0, 1, ..., N$. Where η is the integer power of cost function ($\tau(L_i) = L_i^{\eta}$) and s_g is the labor share in production ($g = i, f$). Thus, in the short-run, if $\epsilon < 1$ or $\rho < 1$, the sign of ε_{L_i,w_i} will always be negative. By contrast, the sign of ε_{L_f,w_i} depends on the interaction between the substitution parameter ρ and the product demand parameter ϵ, which determines the sign of its numerator. Two cases arise:

1. If the product market is competitive, so that the price of the good p is fixed (i.e., $\epsilon = 1$) and formal-informal workers are imperfect substitutes ($\rho < 1$), there will be *scale effects* and firms increase employment as a response to lower informal wages (firms will hire more formal and informal labor in this case, and in general when $\epsilon > \rho$).

2. If price elasticity of demand is smaller to labor elasticity of substitution ($\epsilon^D < \sigma$), there are *substitution effects* and firms response to lower informal wages will be to hire less formal and more informal labor, with final size depending on η and s_g.[45]

In graphical terms, Figure 12 depicts the equilibrium outcomes in both markets (f and i), with the characteristic that f has a binding minimum wage (or price-ceiling) which distorts the equilibrium in that market.[46] Thus, the change in equilibrium employment in each sector can be calculated as follows:

$$\frac{\Delta L_i}{L_i} = \frac{\Delta w_i}{w_i} * \varepsilon_{L_i,w_i} \tag{16}$$

$$\frac{\Delta L_f}{L_f} = \frac{\Delta w_i}{w_i} * \varepsilon_{L_f,w_i} \tag{17}$$

Therefore, the labor supply shock lowers market wages of informal labor and increases its employment L_i (see supply-demand graph 12), consistent with empirical findings in Figure 9b. On the one hand, note also there is a decline in the number of natives who work in the informal market from L_i to the intersection of w_{i2} at S_1. On the other hand, since in Figure 9a I find insignificant effects on formal wages and negative ones of formal employment, this evidence would

[45]Note that Ulyssea (2018) assumes $\rho = 1$, therefore formal and informal workers in his setup are perfect substitutes.

[46]In practice, there is evidence of bunching of formal workers around the minimum wage (see Figure 1a).

be consistent with $\rho > \epsilon$. This corresponds to the case where formal and informal workers are strong substitutes relative to the price elasticity of demand or, according to Borjas (2013) two-good economy framework, relative to the consumer substitution among available goods.[47]

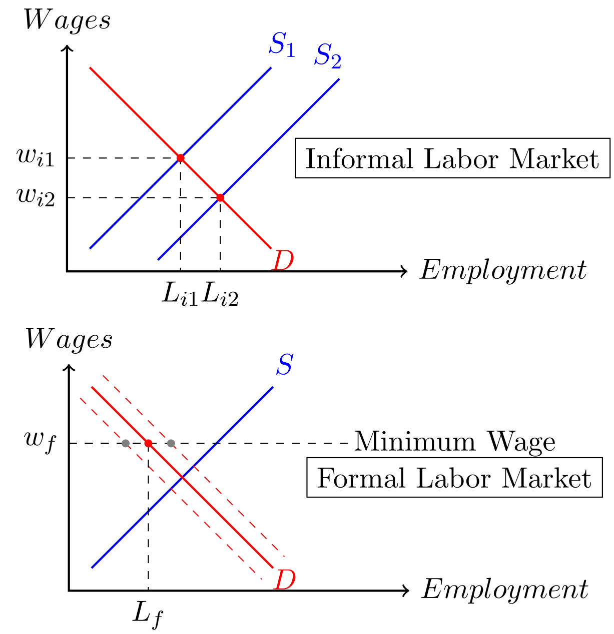
The next step is to use previous estimates to recover values of own-elasticity and cross-elasticity, where the latter is, to the best of my knowledge, the first time to be estimated in the migration literature. First, using the ratio between the overall change of informal employment and the native informal wage change for 2018, I find that own-elasticity of labor demand is: $\varepsilon_{L_i,w_i} = \frac{0.69\%}{-1.87\%} = -0.37$.[48] This estimate lies in the range of values of labor demand elasticities previously found in the literature (Lichter, Peichl, and Siegloch, 2015; Hamermesh, 1996). [49] Finally, I use the ratio between the overall change of formal employment and native informal wage change for 2018 to find that: $\varepsilon_{L_f,w_i} = \frac{-2.22\%}{-1.87\%} = 1.19$, suggesting that formal and informal workers might be close to perfect substitutes in production.

[47] Borjas (2013) states a quasi-linear utility function for the consumer in terms of a locally produced good and an imported good, it can be assumed that ϵ matches the substitution parameter for these two goods.

[48] I use native wage change and not overall wage change to remove possible compositional bias from immigrants lower wages.

[49] If we restrict to informal labor 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.

Figure 12: **Local market responses to a supply shift when immigrants and natives are perfect substitutes**



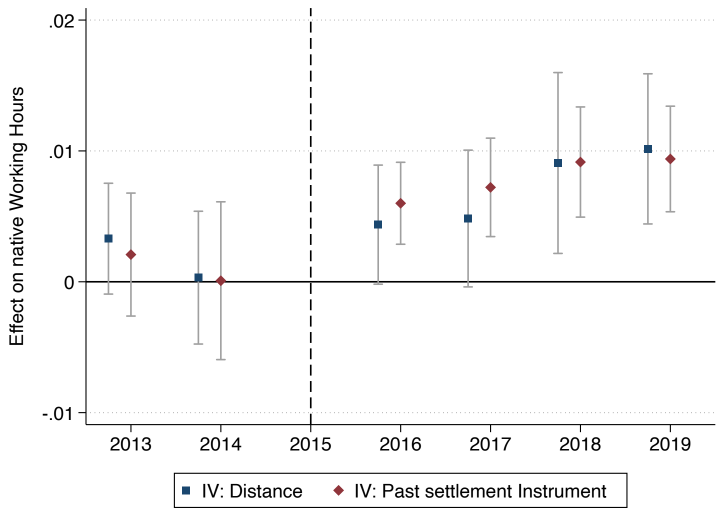
Note: Only the informal labor market suffer a positive supply shock. Wages for formal workers are downwardly rigid. There can happen two spillover effects on formal labor demand given the reduction in the cost of informal hiring: *i*) a reduction or *ii*) an increase, this will depend on the degree of substitution between formal and informal workers relative to the price elasticity of demand. Note that there can be several definitions on informality, explicitly for this case I focus on the worker side, abstracting from informality at the firm level. Moreover, I define informal workers through labor markets but can also be sectors or types of employment.

8 Additional Outcomes and Heterogeneous Impacts

8.1 Intensive Margin Effect

Given that natives' hourly wages on average are being reduced by the immigration event, a natural question arising is what is the effect of this fall in wages on Colombians labor supply decisions: are they working longer hours or are they enjoying more leisure time? To answer these questions, I make use of GEIH survey questionnaire which asks how many (weekly) working hours usually the worker works. The outcome variable in Figure 13 is log (working hours) relative to the base period. I find 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. Note, however, that an important caveat could arise from the repeated cross-section type of GEIH survey, since it is not possible to rule out entirely compositional effects (i.e., part-time workers moving relatively more than full-time workers out of employment/labor force). In any case, when estimating the wild cluster bootstrap p-value for the 2018 coefficients I do not reject the null hypothesis of $\beta_{2018} = 0$, as for distance instrument I find $p = 0.268$ and for past settlement instrument I find $p = 0.124$.

Figure 13: **Event study estimates on log working hours of Colombians**

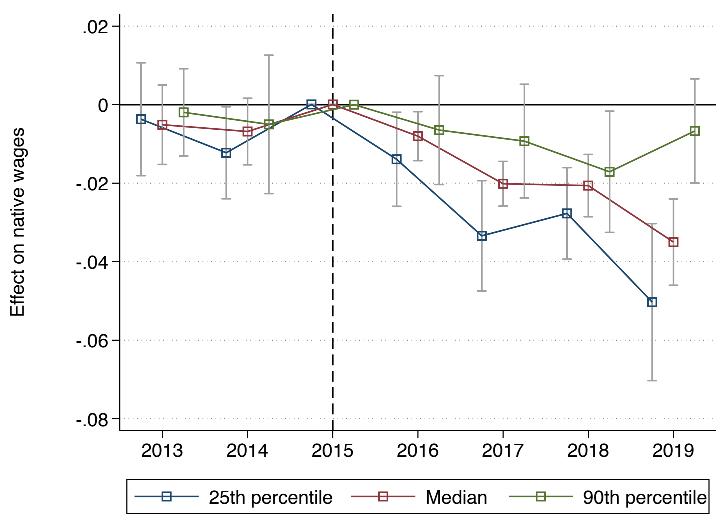


Note: Departments in the regression are $N = 24$ per year. 95% Confidence Interval. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9 and for past settlement instrument is 35.54.

8.2 Distributional Effects of Immigration

Immigrant arrivals are likely to have differential impacts across the distribution of wages. For instance, the effect of the immigration shock on the mean wage can be different from the effect on wages located on lower or higher percentiles of the wage distribution. Using data from GEIH, which provides the most detailed sample coverage in Colombia with information on wages, I am able analyze whether there are heterogenous effects at selected quantiles 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 (where the outcome variable is the log wage in the selected quantile in t relative to the log wage for the same quantile in 2015) are plotted in Figure 14. 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. At the median, which is a more robust estimate to outliers and censored data in my sample, I find an estimate of -2.1% for a 1 p.p increase in the migration rate $M_{d,2018}$, which is higher in absolute terms compared to the mean estimate (-1.6%) obtained when distance to Venezuela is the chosen instrument. Comparing these results with the UK, Dustmann, Frattini, and Preston (2013) document that immigration depresses native 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 more educated than natives.

Figure 14: **Event study estimates on log hourly wages of Colombians by percentiles**



Note: Departments are $N = 24$ per year. 95% Confidence Interval. For the outcome, instead of the average wage in each department, I use the value at given quantiles of the local wage distribution as outcome. instrument used is past settlement. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

8.3 Effects by Education, Type of Job, Gender and Age for Natives

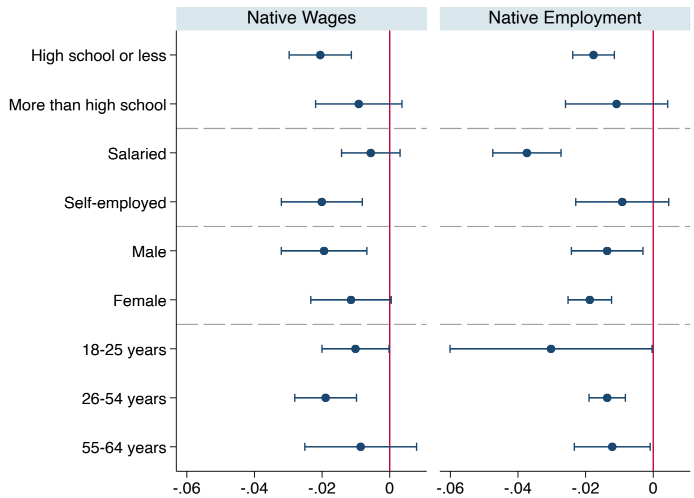
Next, I test for heterogeneous effects across different native subpopulations. Noting that because instruments did not predict local trajectories of wages in the pre-treatment period, while for employment it does, I use trend-adjusted estimates for employment not for wages, with the caveat that trend-adjusted estimates for wages differ little (as shown in Table 4). First, by educational achievement or skill group, I find similar results for 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 (defined as with high school or less). On the contrary, native skilled workers (defined as with more than high school) are non significant (see Figure 15).

Second, a substantive portion of Colombian workers are self-employed or without fixed salaries (with their labor income mainly comes from sales of goods and services). In this subpopulation (which are mainly informal workers), I find a negative wage impact, while for salaried workers (mainly formal employees) the effect is insignificant. Interestingly, the opposite happens for employment responses where salaried native workers experience the largest negative effect (see Figure 15).

Third, in terms of gender impacts (Venezuelan men and women arrived to Colombia in similar magnitudes), males suffer the highest reduction in wages, in contrast to females for whom no significant effects are found. Once more, this finding contrasts with the employment effects by gender, where females experience a higher reduction than males, possibly due to their more elastic labor supply (Blau and Kahn, 2017).

Fourth, I study the immigration 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 terms of wage reductions. In terms of employment responses, the most affected native group is the younger one although with wide confidence intervals, while the estimates for the medium and older group are much smaller (see Figure 15). Recall that Venezuelan immigrants are concentrated in younger ages.

Figure 15: **Native wages and employment estimates by subpopulations, 2015-2018**

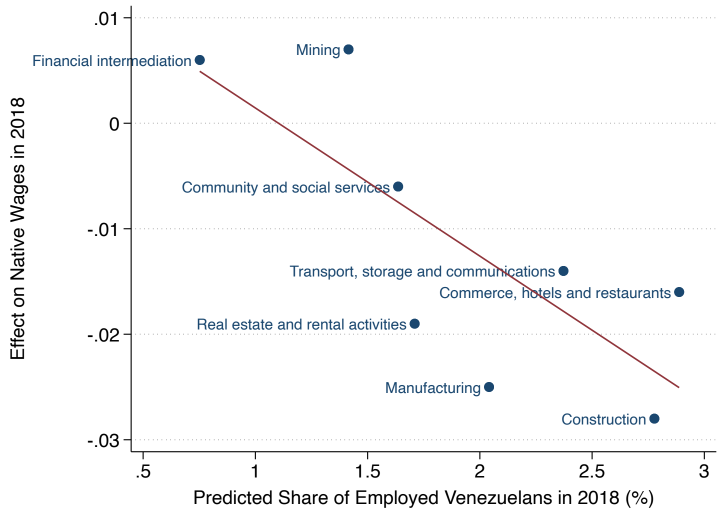


Note: The Figure reports the coefficients of the second stage regression of the distance instrument with the migration rate $M_{d,2018}$. The outcome is the difference in log hourly wages or log employment in 2018 with the base period. Standard errors are clustered at the department level. The variables are in logarithms, thus the coefficients are interpreted as percentages (*100). Department sampling weights from GEIH are used to construct aggregate outcomes. For employment I use trend-adjusted estimates, which are controlled by the growth in native employment from 2015 compared to 2013 for each subpopulation. F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

Finally, I estimate separate effects of immigration by sector of 2018, distinguishing eight big industries in the Colombian economy. Figure 16 plots the estimated effects against the predicted share of employed Venezuelans in those industries.[50] Importantly the industries with larger shares of employed Venezuelans experienced the largest declines in native wages.

[50]The predicted share is calculated regressing the actual share of Venezuelan workers in each industry for 2018 against the share of informal workers in each industry in 2015.

Figure 16: **Native Wage Effects by Industry**



Note: Outcome variable is the difference between 2018 and 2015. 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. The predicted share is calculated regressing the actual share of Venezuelan workers in each industry for 2018 against the share of informal workers in each industry in 2015. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. Past settlement instrument is used, the first stage F-statistic is 35.54. Hourly wages are in real terms using the monthly CPI from DANE.

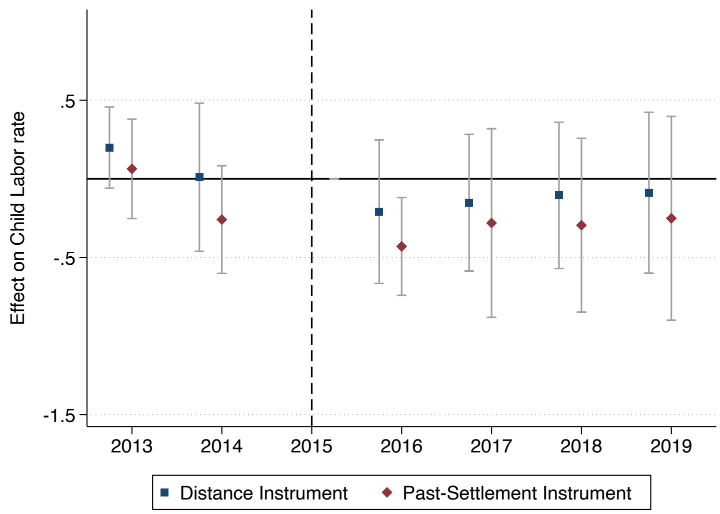
8.4 Impact on Child Labor

The reduction of the relative price of labor can also have spillover effects on 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 17 plots the estimated effect of immigration on the CLR, where I find a negative point-estimate insignificant for all the years using the distance instrument.[51]. Similarly, when past settlement is used as the instrument, a significant negative effect only appears in 2016. The coefficient for this year can be interpreted as follows: a 1 p.p increase in the migration rate decreases the CLR by 0.43 p.p, which is a sizable reduction but then gets attenuated afterwards. This attenuation effect could be driven by the outmigration response of working-age natives seen in 6a. One way to rationalize this suggestive negative effect would be to think that the opportunity cost for parents of working kids increases as a result of the reduction in the relative price of labor in the informal sector, where children are more likely

[51]The CLR data is only available for capital cities, thus I only have 23 units from the GEIH survey.

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 17: **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 (1) are the plotted coefficients. 95% Confidence Interval. F-statistic for distance instrument is 157.9 and for past settlement instrument is 35.54.

8.5 Impact on Prices

The wage and employment responses from the Venezuelan immigration in Colombia are not the only channels of adjustment in the presence of a supply shock. To complement my previous analysis, I also analyze the price response of a specific bundle of goods and services. In this context, there are two opposing mechanisms that can affect prices, the first one is the demand effect that arises from the relatively higher consumption of goods and services from migrants. The second one is a negative supply effect from relatively lower wages that can alter production costs. Moreover, there is also a search channel as migrants might be more sensitive to prices, which can spur more competition in given markets which drives prices down. Therefore, the final impact on prices is ambigous. Previous studies have found that: (i) low-skilled immigration in the US reduces prices of non-tradable goods and services where migrants are more likely to compete with natives, via lower wages (Cortes, 2008), (ii) the mass inflow of Soviet Union immigrants to Israel in the 1990s lowered the price of goods, via higher price elasticity and lower search costs of immigrants (Lach, 2007), and (*iii*) immigration in the UK decreases prices of non-tradable goods and services with low-wage labor intensive, via lower production costs, while prices of tradable goods increase, via the demand effect (Frattini, 2008).

Price data is obtained from DANE official webpage. This data covers 23 capital cities of Colombia with a monthly recollection at the store level, for the regressions I use quarterly information. Noting that the Consumer Price Index (CPI) outcome period is from 2013 to 2018 with base year 2008 ($index = 100$) .[52] In the specification I use, equation (12), I regress log(price index) on the predicted migration rate $\hat{M}_{d,2018}$ using the distance instrument. In Figure A.4a in Appendix, I plot the event study estimates from this regression for different price indexes. Importantly, there is an estimate close to zero on the overall CPI, indicating a stronger supply effect of immigration. For that reason, when I compare the impact on nominal or real wages the effect is strikingly similar. Finally, when focusing on prices of health and education goods and related services I find a negative trend for the former, while for the latter I find an increasing pattern on prices (see Figure A.5b). Negative health prices might arise due to positive changes in the provision of public and private health services in the high-immigration areas.

9 Robustness Checks

The first robustness check of the previous empirical results relates to the exclusion of areas in the Venezuelan border from the analysis. The goal is to check if the direction and significance of previous estimated effects hold when removing departments that are more geographically close to Venezuela. Note that these departments could be more affected by Venezuelan crisis through fewer trade links or lower business interactions. The main results of this exercise, displayed in Figure A.7a and A.7b, yields no significant variations in the estimate, irrespective of the instrument being used. In addition, I also restrict the sample to natives residing in current department for more than one year, removing possible compositional bias of natives reallocating between departments due to the Venezuelan immigration. This is a relevant check since, according to Figure 6a there is an out-migration response of natives after the arrival of migrants. Importantly, when imposing this sample

[52] A reform to the CPI in 2018, which includes more cities and a different composition of the main division of goods and services, makes the comparability with previous periods unfeasible.

restriction, estimates tend to be higher, implying that the results obtained with the unconstrained sample could be viewed as conservative (see Table A.2). Finally, to account for unobservable shocks related to proximity with the Venezuelan border, I explicitly control with quintiles of distance to the nearest crossing bridge with Venezuela. I find that the coefficients are all smaller, specially for wages with the distance instrument (-1.3% vs. -1.6%), This small difference, however, is plausible as there can be a correlation between the instrument and the quintiles build. Likewise, the estimates obtained with the past settlement instrument are now insignificant, a result that could be expected since distance is a better predictor of Venezuelan arrivals than previous enclaves, as seen from the first-stage regression.

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 IPUMS (2019). Figure A.9 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 the change in real log GDP 2018-2015 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, reported in Table A.2, confirm the statistical significance of these effects (with a lower point-estimate for GDP and a higher one for Trade).

Finally, there is the issue that in the 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. [53] As shown in Table A.2, which shows the coefficients of immigration taking as dependent variable the observed wage and the residual wage, the residual wage has a higher estimate

[53] 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.

(in absolute terms) for the distance instrument and smaller one for the past settlement instrument. Thus, there is no much gain in regressing previous individual characteristics in previous analysis.

9.1 More pre-treatment years

When analyzing event studies, pre-treatment information is crucial to assess the violation of the common trend assumption. In practice, the UPTA 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.[54] Figure A.10a 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 instruments, 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 fluctuation but were not on persistently different growth paths, as the coefficient for 2011 turns to be close to zero (Figure A.10b).

10 Conclusion

This paper analyzes the impact of the Venezuelan mass migration on the Colombian labor market. Exploiting the differential intensity of Venezuelan arrivals within Colombian departments, I implement an event study design with two separate instruments to assess the causal validity of local impacts. For both instruments I find a negative effect on native hourly wages of around 1.6%-1.7% for a 1 p.p increase in the share of employed Venezuelans over the total employed population. This negative wage response is in between the high estimates reported by Caruso, Canon, and Mueller (2019) and the small and insignificant estimates found in Morales-Zurita et al. (2020) and Santamaria (2019) for this specific case study. The differences between my findings and those in these studies can be explained by the empirical specification implemented, the definition of migration rate used and the number of periods analyzed.

When analyzing heterogeneous wage effects, negative wage impact of the immigration shock mainly affects workers with an informal contractual labor arrangements, the self-employed or less-

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

educated ones. When I study the effect along the wage distribution, I find that native wages located in lower percentiles (25th) are more affected by the Venezuelan immigration compared to those wages located in the upper percentiles (90th). Furthermore, the impact on native employment appears to have a delayed negative response compared to wage response, which is more pronounced on younger individuals (from 18 to 25 years) and salaried workers. The analysis on prices also suggest that the higher demand for Colombian goods and services stemming from a larger number of immigrants has been small, probably the higher demand effect has been offset by lower production costs or lower search costs of immigrants. Overall, the influx of Venezuelans immigrants increased informal employment and decreased informal wages in Colombia due to higher wage flexibility in this sector. By contrast, it has had an insignificant effect on formal wages and negative effects on formal employment because wages in the formal sector are downward rigid. I therefore conclude that the labour supply shock under study affected mainly informal sector wages whereas the employment effects are primarily felt in the formal sector.

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.

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.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, 55(3):789–865.

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. (2013). The analytics of the wage effect of immigration. *IZA Journal of Migration*, 2(1):1–25.

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*.

Cortes, P. (2008). The effect of low-skilled immigration on us prices: evidence from cpi data. *Journal of political Economy*, 116(3):381–422.

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*.

Frattini, T. (2008). Immigration and prices in the uk. *University College London, mimeo*.

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.

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).

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

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.

Lach, S. (2007). Immigration and prices. *Journal of political Economy*, 115(4):548–587.

Lebow, J. (2020). Refugees in the colombian labor market: The consequences of occupational downgrading. *Available at SSRN 3778377*.

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.

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.

Sanchez-Serra, D. (2016). Functional urban areas in colombia.

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.

UCAB (2020). Encuesta nacional de condiciones de vida. Technical report, Universidad Católica Andrés Bello. Available on: https://assets.website-files.com/5d14c6a5c4ad42a4e794d0f7/5f03875cac6fc11b6d67a8a5_Presentaci%C3%B3n%20%20ENCOVI%202019-Pobreza_compressed.pdf (last accessed 7th of July 2020).

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.

A Appendix

A. Descriptive Statistics for Natives and Migrants

Table A.1: 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

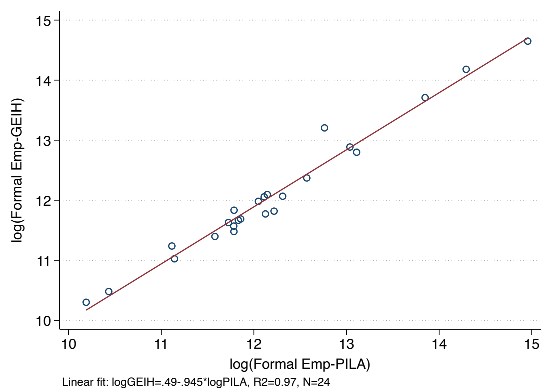
Note: NHS stands for No High School and HS stands for High School. The shares are calculated using national sampling weights from GEIH. College aggregates all the technical levels of education after high school. (b) and (c) are restricted to population that in the survey responded that they were living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

B. Comparison of Administrative and Survey data

I use the Colombian Social Security Records (PILA) from the Health Ministry to complement the analysis using GEIH survey's outcomes. Importantly, PILA covers only formal employment, so the comparison population from GEIH should be workers that contribute to the health system. First, I test if the number of formally employed workers in aggregate numbers is the same in both sources, for that comparison, I aggregate outcomes for GEIH using department sampling weights (the same as for previous results). In A.1 I show the scatter plot of the two measures, with the corresponding fitted values of the linear regression. Each dot matches very closely the values in both sources, giving reliability and robustness to the survey data utilized.

After testing for the robustness of the information, I build the outcomes from the two datasets to compare the results. Before I show the results I want to highlight that PILA, specially between 2018-2019, suffers from a measurement error of the location variable, as the DIVIPOLA codes (location identifier from DANE) are not captured correctly in the individuals observation of the data. Still, is not possible to quantify this misinformation as I only have access to the aggregate data. For that reason, the results might not be the first-best option, as one could think a priori, and can be used more as a complement to GEIH. With this in mind, I turn to the results to show that for both wages and employment in PILA I find very similar results as for GEIH, with formal wages having insignificant estimates.[55] While for formal employment I find negative coefficients on both, yet not significant for PILA, mainly due to the measurement problem of the location variable which gives wider confidence intervals for this outcome.

Figure A.1: Log formal employment for GEIH and PILA

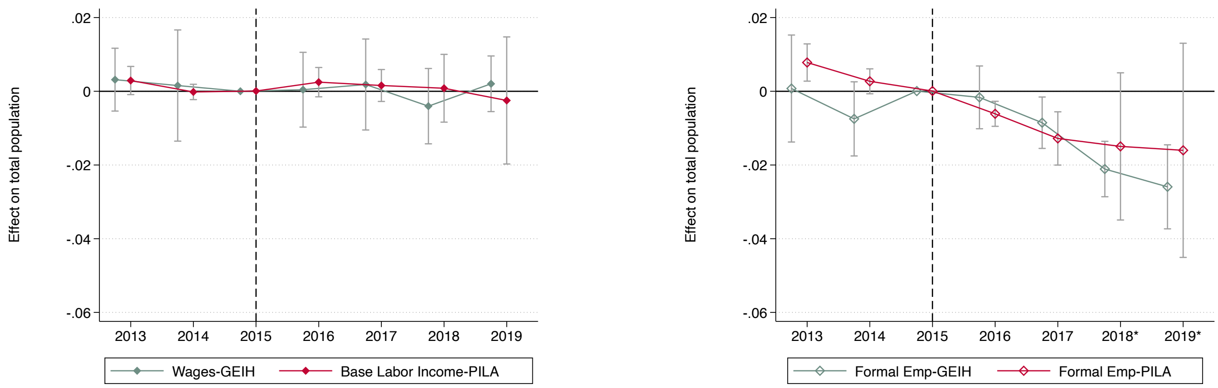


Note: Departments are $N = 24$ for 2015. The sample is not restricted and covers the universe of formal workers in the department. Department sampling weights from GEIH are used to construct aggregate outcomes. Source: GEIH-2015 and PILA-2015.

[55] In PILA is called basic income and refers to the amount of labor income you want to declare for the contributions to the social security system. Information accessed the 7th of May of 2021.

Figure A.2: Event study estimates on log hourly wages and log employed with GEIH and PILA

(a) IV estimates on Wages (b) IV estimates on Employment



*Data for those years suffer from measurement error as authorities did not verify the location variable in PILA. Note: Departments in the regression are $N = 24$ per year. 95% Confidence Interval. The sample is not restricted and covers the universe of formal workers in the department. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE. Source: GEIH:2013-2019 and PILA:2013-2019.

C. 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 regression, where the base period is the third quarter of 2015,

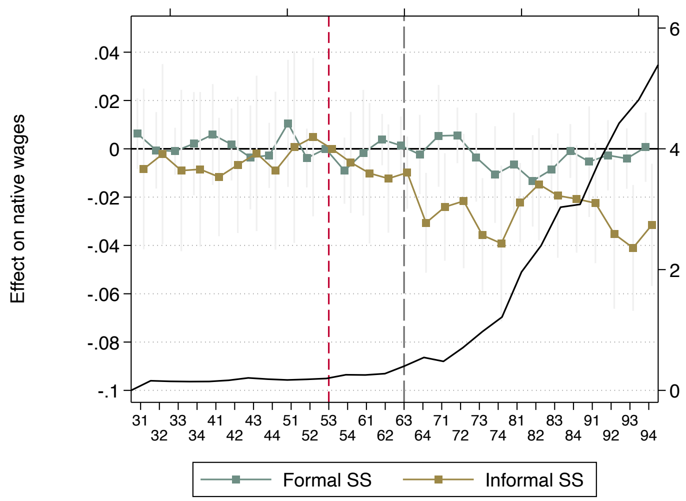
$$Y_{dq} = \gamma_q + \gamma_d + \sum_{q=2013-1, q\neq 2015-3}^{2019-4} \beta_q \hat{M}_{d,2018} + u_{dt} \tag{18}$$

and $\hat{M}_{d,2018}$ is the predicted migration rate using the distance instrument. First, I complement this analysis for the wage response by worker informality and then I do it for prices.

Native Wages

First, note that the red dashed line in Figure A.3 represents the base period of analysis of the event study regression, whereas the grey long-dashed line represents the re-opening of the border between Colombia and Venezuela, and finally the thick black line is the quarterly migration rate build from GEIH survey. With this in mind, β_q estimates on native wages after the grey line (re-opening of the border) are more pronounced for workers without access to social security, while for the formal ones seems to be unaffected. Interestingly, the relation between the effect on informal wages and the migration rate seems to be non-monotonic, as increases in the thick black line (migration rate) does not necessarily translate into higher reductions on informal wages.

Figure A.3: 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 defined as before. 95% Confidence Interval. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015 third quarter. β_q from equation (12) are the plotted coefficients, standard errors are clustered at the department level. F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

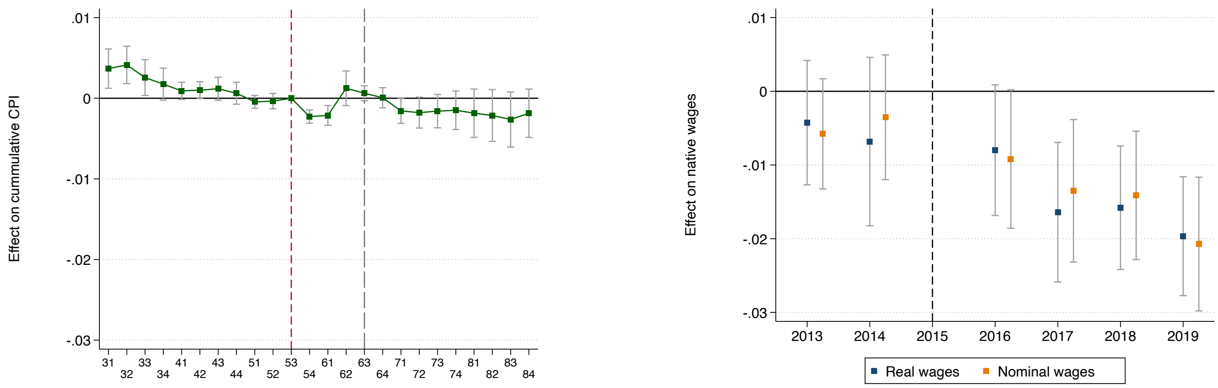
Prices

Note that the red dashed line represents the base period of analysis, while the grey long-dashed line represents the re-opening of the border between Colombia and Venezuela. As explained above, the results on prices indicate that lower production costs or higher search costs from immigrants might offset the demand effect of more migrants consuming goods and services.

Figure A.4: Event study estimates on CPI and wages

(a) CPI

(b) Real vs. Nominal

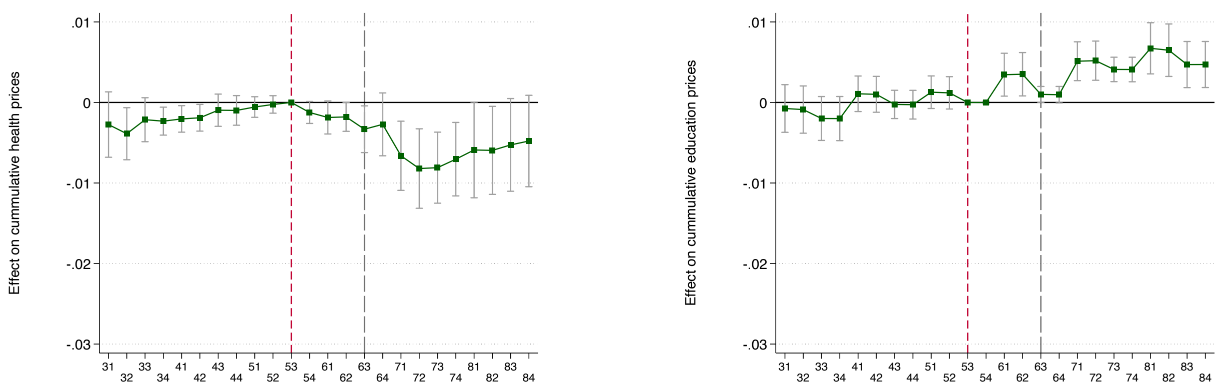


Note: Capital cities in the regression are $N = 23$ per year. 95% Confidence Interval. The base period is 2015 third quarter. β_q from equation (12) are the plotted coefficients in (a), β_t from equation (1) are the plotted coefficients in (b), standard errors are clustered at the department level. F-statistic for distance instrument is 157.9.

Figure A.5: Event study estimates on log price indexes

(a) Health prices

(b) Education prices

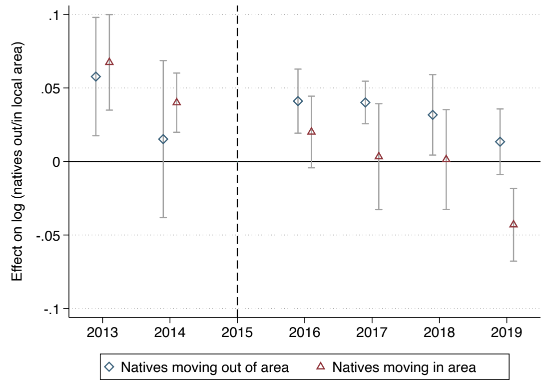


Note: Capital cities in the regression are $N = 23$ per year. 95% Confidence Interval. The base period is 2015 third quarter. β_q from equation (12) are the plotted coefficients, standard errors are clustered at the department level. F-statistic for distance instrument is 157.9.

D. Robustness Checks

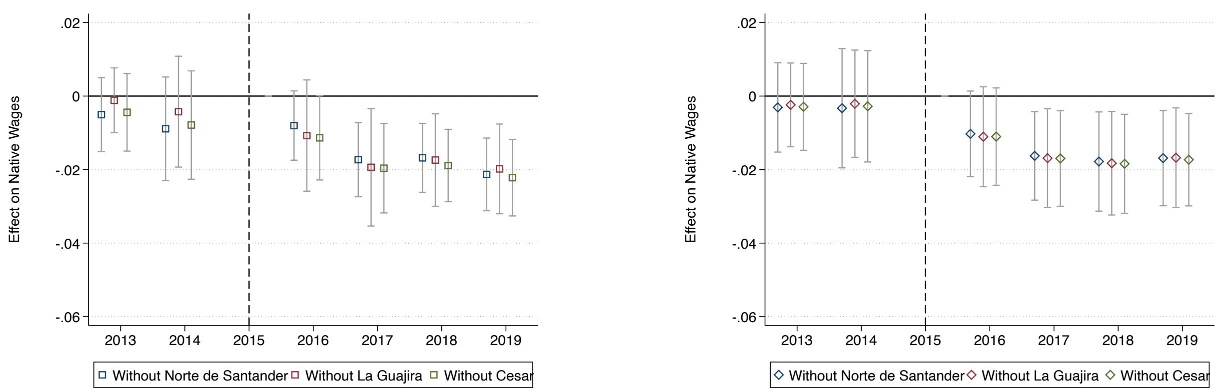
In here I include all the robustness analysis performed throughout paper. Most of the checks are event study estimates for different specifications, changes in the treated group and threats to the identifying assumptions.

Figure A.6: Event study estimates on movements across geographical areas



Note: Departments in the regression are $N = 24$ per year. 95% Confidence Interval. Measures of geographical movements come from GEIH migration module. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. θ_t from equation (3) are the plotted coefficients with explanatory variable $M_{d,2018}$.

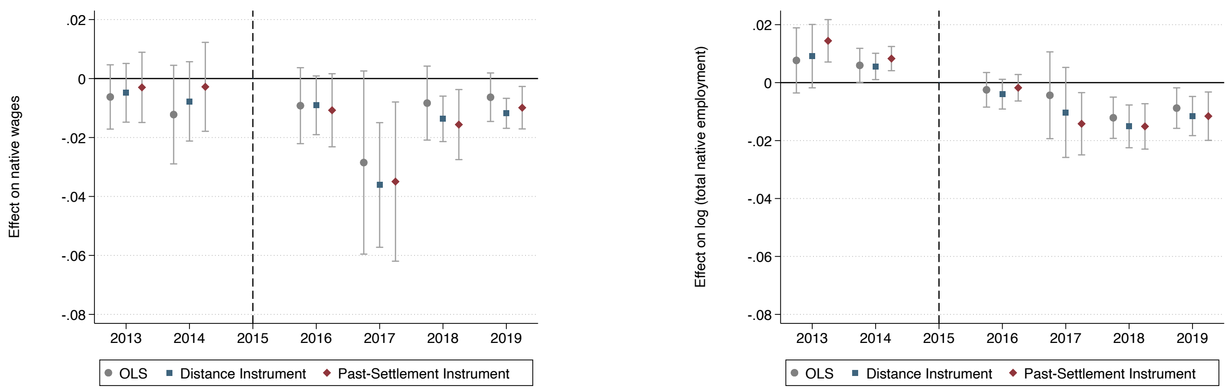
Figure A.7: Event study estimates excluding border departments for native wages



Note: Departments in the regression are $N = 23$. 95% Confidence Interval. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. Hourly wages are in real terms using the monthly CPI from DANE.

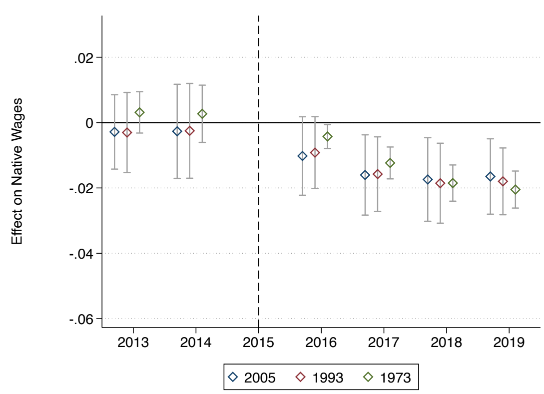
Figure A.8: Event study estimates using M_{dt} from GEIH survey as explanatory variable

(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}. 95% Confidence Interval. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. θ_t from equation (3) are the plotted coefficients. Hourly wages are in real terms using the monthly CPI from DANE.

Figure A.9: Event study estimates using different historical shares for the construction of past settlement instrument

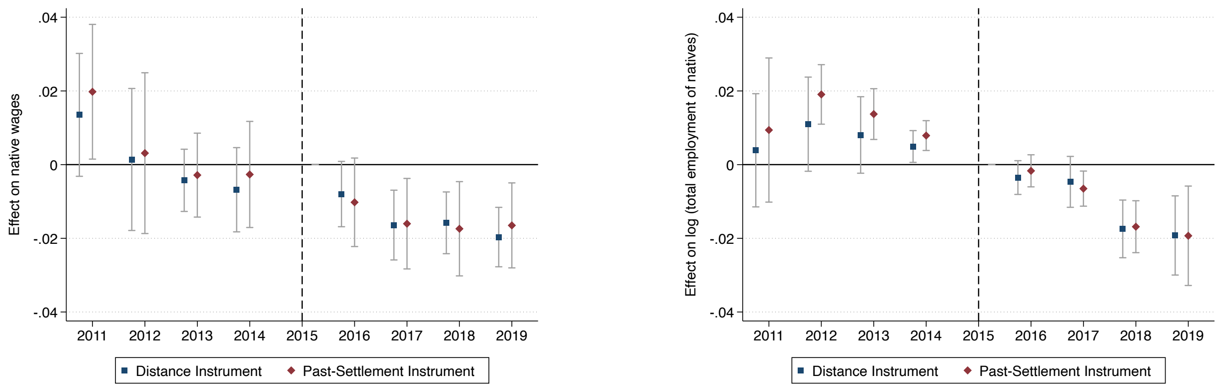


Note: Departments in the regression are $N = 24$ for 2005 and 1993, and $N = 22$ for 1973. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. 95% Confidence Interval. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for past settlement instrument with shares of 2005 is 35.54, with shares of 1993 is 34.39 and with shares of 1973 is 43.02. Source: IPUMS for 1993 and 1973 and DANE for 2005.

Figure A.10: Event study estimates on log hourly wages and log employed Colombians with more pre-treatment years

(a) IV estimates on Wages

(b) IV estimates on Employment



Note: Departments in the regression are $N = 24$ per year. 95% Confidence Interval. In 2011 and 2012 I assume all respondents are Colombians. The sample is restricted to Colombian permanent residents between 18 and 64 years in urban areas. Department sampling weights from GEIH are used to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients, by construction $\beta_{2015} = 0$, and standard errors are clustered at the department level. F-statistic for distance instrument is 157.9 and for past settlement instrument is 35.54. Hourly wages are in real terms using the monthly CPI from DANE.

Table A.2: **Robustness checks: wages and employment estimates for Colombians, 2015-2018**

	(1) Wages	(2) Employment	(3) Wages	(4) Employment
instrument	**Distance**		**Past settlement**	
Main results⋆	-0.0158***	-0.0147***	-0.0174**	-0.0113**
	(0.00427)	(0.00335)	(0.00653)	(0.00430)
Control change GDP 2018-2015	-0.0152**	-0.0146***		
	(0.00460)	(0.00335)		
Control trade with Venezuela 2015	-0.0192*	-0.0153**		
	(0.00640)	(0.00377)		
Control quintiles of distance to Venezuela	-0.0134*	-0.00891*	-0.0171	-0.00609
	(0.00651)	(0.00354)	(0.0103)	(0.00551)
Residualized wages	-0.0192***		-0.0155***	
	(0.00449)		(0.00483)	
Excluding natives that moved in $t-1$	-0.0169***	-0.0177***	-0.0184**	-0.0185***
	(0.00445)	(0.00429)	(0.00667)	(0.00358)
N	24	24	24	24
Trend-adjusted	No	Yes	No	Yes

Standard errors in parentheses. ⋆With significance from standard p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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. The sample is restricted to Colombian 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 to construct aggregate outcomes. 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. Constant prices GDP at the department level are constructed by DANE. Control of distance is constructed as quintiles of the distance to the nearest crossing bridge with Venezuela. 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. When excluding the natives that changed department of residence I do not control for pre-trends in employment as the sample restriction already controls this trend.

E. Derivations of the Theoretical Framework

To derive equations 14 and 15 from the main text, first, I combine the profit function with the price function. So the maximization problem turns to be:

$$\max_{L_i,L_f} \pi = C^{1-\epsilon}Q^{\epsilon} - \tau(L_i)w_iL_i - (1+\tau_f)w_fL_f \tag{19}$$

Then rearranging market wages (from equation 12 and 13) to leave in terms of labor we get that:

$$L_i^{1-\rho}(\tau'(L_i)L_i + \tau(L_i)) = \left(\frac{C^{1-\epsilon}\epsilon\alpha_i}{w_i}\right)(\alpha_iL_i^{\rho} + \alpha_fL_f^{\rho})^{\frac{\epsilon-\rho}{\rho}} \tag{20}$$

$$L_f^{1-\rho} = \left(\frac{C^{1-\epsilon}\epsilon\alpha_f}{w_f(1+\tau_f)}\right)(\alpha_iL_i^{\rho} + \alpha_fL_f^{\rho})^{\frac{\epsilon-\rho}{\rho}} \tag{21}$$

To have a tractable solution, the informal labor cost can be specified as $\tau(L_i) = L_i^2$ or as linear one $\tau(L_i) = L_i$. To be more general, assume $\tau(L_i) = L_i^\eta$ where $\eta = 0, 1, ..., N$.

Then taking logs of last expressions:

$$logL_i = \frac{1}{1+\eta-\rho}((1-\epsilon)logC + log\epsilon\alpha_i - logw_i - log(1+\eta)) + \frac{\epsilon-\rho}{\rho(1+\eta-\rho)}log(\alpha_i L_i^\rho + \alpha_f L_f^\rho) \quad (22)$$

$$logL_f = \frac{1}{1-\rho}((1-\epsilon)logC + log\epsilon\alpha_f - logw_f - log(1+\tau_f)) + \frac{\epsilon-\rho}{\rho(1-\rho)}log(\alpha_i L_i^\rho + \alpha_f L_f^\rho) \quad (23)$$

I differentiate previous equations with respect to informal wages. Noting that formal wages are taken as fixed in the short-run (as formal wages are downwardly rigid by the minimum wage, and I find insignificant changes in the reduced form estimates, this is not problematic). Then these expressions are equal to:

$$\frac{dlogL_i}{dw_i}|_{dw_f=0} * w_i = -\frac{1}{w_i(1+\eta-\rho)}w_i + \frac{\epsilon-\rho}{\rho(1+\eta-\rho)}\left(\frac{\rho(\alpha_f L_f^{\rho-1}\frac{dL_f}{dw_i} + \alpha_i L_i^{\rho-1}\frac{dL_i}{dw_i})}{\alpha_f L_f^\rho + \alpha_i L_i^\rho}\right)w_i \quad (24)$$

Simplifying I get that:

$$\frac{dlogL_i}{dw_i}|_{dw_f=0} * w_i = -\frac{1}{1+\eta-\rho} + \frac{\epsilon-\rho}{1+\eta-\rho}(\alpha_f(L_f/Q)^\rho\varepsilon_{L_f,w_i} + \alpha_i(L_i/Q)^\rho\varepsilon_{L_i,w_i}) \quad (25)$$

Where $\varepsilon_{L_g,w_i} = \frac{dL_g}{dw_i}\frac{w_i}{L_g}$ is the elasticity of labor g with respect to informal wages. Finally we can rewrite last expression as:

$$\varepsilon_{L_i,w_i} = -\frac{1}{1+\eta-\rho} + \frac{\epsilon-\rho}{1+\eta-\rho}(s_f\epsilon_{L_f,w_i} + s_i\epsilon_{L_i,w_i}) \quad (26)$$

Where $s_f = \alpha_f(L_f/Q)^\rho$ and $s_i = \alpha_i(L_i/Q)^\rho$ are the formal and informal labor shares in production and $s_f + s_i = 1$. Then, when I differentiate formal labor with respect to informal wages I get the following:

$$\varepsilon_{L_f,w_i} = \frac{\epsilon-\rho}{1-\rho}(s_f\epsilon_{L_f,w_i} + s_i\epsilon_{L_i,w_i}) \quad (27)$$

Combining the two last expressions I find that:

$$\varepsilon_{L_i,w_i} = -\frac{1}{1+\eta-\rho} + \frac{1-\rho}{1+\eta-\rho}\varepsilon_{L_f,w_i} \quad (28)$$

Using these two last equations to solve in terms of ε_{L_i,w_i} and ε_{L_f,w_i} yields the same equations 14 and 15 in the main text.

F. 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[56]). Is worth noting that this variable is capturing any type of labor income payment, as non-salaried workers are part of my sample, so I use interchangeably wages with labor income. Following up, I transform this nominal labor income variable into a real one using monthly CPI at the national level.[57] 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 usually at this job in the 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 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.

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

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