2022-12
Working paper. Economics
ISSN 2340-5031

INMIGRATION, WAGES, AND EMPLOYMENT UNDER INFORMAL LABOR MARKETS

LUKAS DELGADO PRIETO.

uc3m | Universidad Carlos III de Madrid

Immigration, Wages, and Employment under Informal Labor Markets*

Lukas Delgado-Prieto†

August, 2022

Abstract

This paper studies the labor market impacts of Venezuelan immigrants in Colombia. Exploiting spatial variation in exposure, I find a negative effect on native wages driven by the informal sector (where immigrants are concentrated) and a reduction in native employment in the formal sector (where the minimum wage binds for many workers). To explain this asymmetry, I build a model in which firms substitute formal for informal labor in response to lower informal wages. Consistent with the model's predictions, I document that the increase in informality is driven by small firms that use both labor types in production.

Keywords: Immigration, Event study, Labor market, Informality.

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 for helpful feedback to Jesús Fernández-Huertas, Jorge Pérez Pérez, Lorenzo Aldeco, and Sarah Schneider-Strawcynski. In general, I am indebted to all the comments from participants in the UC3M Applied Reading Group, the Junior Economics of Migration Seminar, the Seminario de Economía Aplicada BANREP-UIS-UNAB, the 10th European Meeting of the Urban Economics Association, the 17th IZA Annual Migration Meeting, the Seminar Series of Banxico, the 14th Migration and Development Conference, and the Seminar at ZEW. I am also grateful for all the support from DANE staff, especially from Geovanni Portilla. This paper previously circulated as: "Dynamics of Local Wages and Employment: Evidence from the Venezuelan Immigration in Colombia". Replication files are available upon request. All errors are my own.

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

1 Introduction

Despite a large number of studies over the last three decades, the impact of immigration on native wages and native employment 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). In Latin America, the number of migrants increased significantly in the last years with the Venezuelan crisis, as more than 4.6 million people left the country between 2016 and 2019 (UNHCR, 2019). These massive and sudden inflows can influence a wide range of socio-economic outcomes in the host countries, both in the short- and long-run.

This paper studies the labor market impacts of Venezuelan immigrants in Colombia, where the supply of migrants surged from 0.2% in 2015 to 4.1% in 2019 (this means around 1.2 million working-age Venezuelans in absolute numbers).[1] 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. This effect, however, might be different in settings with a large informal sector and binding minimum wages in the formal sector. Mainly because informal wages adjust more flexibly while formal wages are more rigid. Thus, the impacts on formal and informal employment might be different (Kleemans and Magruder, 2018).[2]

The main contribution of this paper is to show empirically that immigration affects differently the wage and employment response of the formal and informal sector in Colombia, and then to show theoretically why there exists an asymmetric response between these sectors. First, for the empirical part, I take advantage of varying treatment intensity across areas and time. Certain areas in Colombia received vast inflows of migrants, while others barely received any. This is the main source of variation exploited in the empirical strategy.

The critical assumption needed in this setup is the parallel trends assumption (PTA). In practice, this assumption might not be satisfied as migrants can endogenously sort into the areas that offer the best economic conditions, which might lead to differential trends in the outcomes even in the absence of any impact of immigration. To deal with the potential sorting and other endogeneity issues, I use an event study research design with instrumental variable (IV). Two instruments are constructed: (*i*) distance between capital cities in the two neighboring countries and (*ii*) past

[1] The supply of migrants is measured as working-age Venezuelans over working-age natives.

[2] The trade-off between wages and employment in the face of migration shocks goes back to Borjas (1999).

settlements of Venezuelans. To provide indirect support for the PTA with IV, I show that the chosen instruments do not predict the trends in native wages before the migration crisis started.[3]

Overall, I find that the inflow of Venezuelans reduces native (hourly) wages. A one percentage point (pp) increase in the share of employed Venezuelans over the employed population in a department reduces local native wages by between 1.6% and 1.7%. Importantly, these estimates are robust to the two chosen instruments, to alternative definitions of local labor markets, and to different wage datasets. Compared to other studies analyzing this immigration shock in Colombia, my wage estimate lies in between the large negative coefficient of Caruso, Canon, and Mueller (2021) and the more positive or insignificant ones of Morales-Zurita et al. (2020) and Lebow (2021a).[4] Compared to other migration episodes, I find a wage estimate that is more negative than several recent papers (Aksu, Erzan, and Kırdar, 2018; Dustmann, Schönberg, and Stuhler, 2017; Monras, 2020) except for the wage estimate of Edo (2020), which is similar in magnitude.[5] I then complement the effect on average wages with the distributional impacts of immigration. Exploiting the wage distribution, I find that wage losses are concentrated at the bottom of the distribution, while wages in the upper part are almost unaffected.

In terms of native employment, or the extensive margin, I find a negative response. A one pp increase in the share of employed Venezuelans reduces local employment by between 1.1% and 1.5%. On the intensive margin, natives appear to be working more hours in response to the immigration shock; still, this can be mechanical as part-time workers seem to be crowded out more from employment than full-time workers.

I then argue that a key factor driving the sizable negative wage elasticity arises from the large number of informal workers in Colombia (51.9% of native workers did not contribute to social security as of 2019). In the informal sector, the lack of downward rigidity of wages, without a minimum wage, allow wages to adjust flexibly to a labor supply shock. This is relevant as almost

[3] In addition, to alleviate concerns with the exclusion restriction, I further show that trade shocks arising from the Venezuelan crisis are not so relevant in the post-treatment period, as most of the trade adjustments happened before the analysis period.

[4] In fact, Caruso, Canon, and Mueller (2021) find the most negative wage elasticity (–7.6%) relative to the existing migration literature cited in Dustmann, Schönberg, and Stuhler (2016). In Appendix A, I show that the difference between the wage estimates of this paper and that one is driven by the number of periods they analyze (until 2017) and the empirical specification they choose (panel regression). This motivates, in part, the analysis performed in this paper.

[5] Edo (2020) finds for the Algerian inflow in France that a one pp increase of repatriates lowered native wages by between 1.3% and 2%.

all Venezuelans are being employed by the informal sector. Specifically, for natives, I find an insignificant effect on informal employment and a considerable decrease of informal wages; with a one pp increase in the immigration shock, informal wages decrease by 1.9%.

For formal employment of natives, I find a reduction of 2.2% for a one pp increase in the immigration shock, while for formal wages, I find an insignificant effect around zero. The insignificant impact is determined partly by the downward rigidity imposed by the minimum wage, which is binding for many formal workers in Colombia (in 2015, 31.7% of formal workers earned around the minimum wage). From a policy perspective, the existence of a relatively high minimum wage magnifies the adverse employment effects in the formal sector.

To rationalize these empirical findings, I first show that firms combine formal and informal labor in their production function. Importantly, this combination mainly depends on firm size: when firms are bigger, the share of informal workers decreases. Exploiting this fact, workers are divided by their firm size categories to find that the largest drop in formal employment is in workers employed in the smallest firms, and this finding motivates the theoretical framework. I build a model where a firm can hire two types of inputs, formal and informal labor, with different costs as in Ulyssea (2018) but allowing them to be either complements or substitutes in production. In this model, to hire formal workers, the firm must pay a constant payroll tax, while to hire informal workers, the firm pays a cost that is increasing in the size of informal labor within the firm (the cost of evasion).

Next, I derive the elasticity of formal labor demand when there is a change in informal wages. The sign of this elasticity depends on the elasticity of substitution between informal and formal workers and the price elasticity of the good produced by the firm. To have a negative formal employment response, formal and informal workers must have a high degree of substitutability. When their substitutability is indeed high, the model predicts that firms will respond to lower informal wages by substituting formal for informal workers, especially in the smallest firms, where the immigration shock is more salient. The fact that formal and informal workers are very substitutes in (aggregate) production is a result that goes beyond the specific immigration context, as it shows how is the general functioning of informal labor markets.

In terms of heterogeneous effects, I find that formal employment decreased especially for low-skilled natives (defined as with high school or less) and informal wages decreased in most subgroups (except high-skilled and older natives). I also exploit area-industry variation to find that in the industries with the largest shares of employed Venezuelans, native wages decline the most. Last, to complement the impact on wages and employment, I study the price response on a bundle of goods and services and find insignificant estimates. These findings indicate a more substantial supply effect from lower wages or higher market competition rather than a higher demand from migrant consumption of goods and services.

This paper contributes to different strands of the literature. First, to the literature on the labor market effects of immigration (Edo, 2020; Dustmann, Schönberg, and Stuhler, 2017; Monras, 2020). The characteristics of the immigration shock under study—namely a large and sudden inflow of migrants driven by the conditions of Venezuela—help identify its impact. Such a sharp change in the migration flows can help to alleviate concerns from previous immigration adjustments conflating short-run impacts (Jaeger, Ruist, and Stuhler, 2018). Not many immigration events have these characteristics, among which possibly the best known is the Mariel Boatlift in Florida (Card, 1990).

Second, to the literature on the labor market effects of immigration in settings with a large informal sector. As the majority of evidence between immigration and labor markets comes from developed economies, less is known about how immigration affects wages and employment in developing economies, where informality is substantive. In this respect, most recent studies exploit the recent influx of Syrian refugees in Turkey (Aksu, Erzan, and Kırdar, 2018; Ceritoglu et al., 2017; Del Carpio and Wagner, 2015), a case study with a migration shock of similar magnitude to the Colombian one. For instance, Aksu, Erzan, and Kırdar (2018) find that the influx of Syrians decreases native wages in the informal sector while upgrading the wages and employment of native men in the formal sector. This setup contrasts the Colombian one as Syrians speak a different language than natives, and Turkey did not implement an open border policy with migrants like Colombia. In any case, I provide a theoretical model that links the informal and formal sectors of the labor market to explain the differential responses. This is highly relevant because Venezuelan migrants are disproportionally employed in the informal sector, but the empirical findings show that the resulting downward pressure in informal wages leads the firm to substitute formal labor for informal labor. So, while the migration wave decreased wages in the informal sector, its employment effects are felt primarily in the formal sector.

Third, this paper also contributes to the literature on the interaction between immigration and minimum wages. For countries without a large informal sector, Edo and Rapoport (2019) find that minimum wages in the US protect native wages and employment from an immigrant supply shock. But for countries with a large informal sector, Kleemans and Magruder (2018) find that internal migration in Indonesia, where the binding minimum wage is high, creates asymmetric responses in the informal and formal sectors. That is, the negative wage effects are driven by income reductions in the informal sector, while the negative employment effects are driven by employment reductions in the formal sector. These results are similar to ones found in this paper, but I study international (not internal) migration. Overall, it is suggestive how binding minimum wages, interacted with the size of the informal sector, can determine the wage and employment response for natives in the presence of a labor supply shock.

Last, this paper makes several contributions to the studies that estimate the impact of the Venezuelan immigration on Colombia's labor market (Caruso, Canon, and Mueller, 2021; Lebow, 2021a; Morales-Zurita et al., 2020; Santamaria, 2020; Rozo and Vargas, 2021). To start, I use an event study design with continuous treatment while using two different instruments that can test for the presence of preexisting trends.[6] The use of this research design is important because pre-treatment coefficients can show if a trade or any related shock from Venezuela is affecting the outcomes of treated areas differently before immigrants arrive.[7] In addition, the dynamic specification of the event study has the advantage of being precise about the timing and total impact of immigration. Since migrants' arrivals are substantially increasing over time, this specification helps to distinguish between short-run adjustments with partial adjustments to long-run impacts.[8] Finally, I use a population census that measures Venezuelans in Colombia with the greatest detail,

[6]The event study design is motivated by the fact that the static coefficient in most of previous studies comes from a panel IV regression, and can be interpreted as a weighted average of treatment effects. The issue is that these weights can even be negative, for instance, because the timing of treatment is varying across groups or because the treatment is continuous (De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). Moreover, this specification needs the PTA, among other assumptions, to yield consistent estimates. Caruso, Canon, and Mueller (2021) provide evidence on the lack of correlation between migration rates in 1973 and migration rates in 2005 with present outcomes, yet the authors do not analyze pre-treatment data before the immigration shock. In that sense, Morales-Zurita et al. (2020) use recent pre-treatment data to find null correlations between past settlements and migration flows between 2013–2015, but they do not study if their instrument predicts economic outcomes during the same years. Finally, Lebow (2021a) study pre-trends for different outcomes between 2013–2015 to find significant differences in unemployment and labor force participation rates before the immigration shock. Therefore, these studies need to account for possible differences in pre-trends, for some outcomes.

[7]For instance, Caruso, Canon, and Mueller (2021), Morales-Zurita et al. (2020), and Lebow (2021a) include in their post-treatment period the years before 2015. But these years of low immigration rates coincide with the massive drop in cross-border trade with Venezuela (as shown later in Figure 3a), resulting in a source of bias in their estimates.

[8]It also helps to distinguish the possible effects of the change in the legal status of Venezuelans during the post-treatment years.

reducing the extent of measurement error in standard surveys and consequently any attenuation bias (Aydemir and Borjas, 2011).

The rest of the paper is structured as follows. Section 2 gives a brief overview of the Venezuelan crisis and the institutional background. Section 3 describes the data used and descriptive statistics of natives and immigrants. Section 4 details the empirical specification and the instrumental variables. Section 5 reports the baseline results for wages and employment. Section 6 shows the differential impact of immigration between the formal and informal sectors and introduces the theoretical model. Section 7 shows the heterogeneous impacts of immigration, Section 8 provides robustness checks, and Section 9 concludes.

2 Venezuelan Crisis and Institutional Background

2.1 A Brief Overview of the Venezuelan Crisis

In 1998, when Hugo Chávez was elected president of Venezuela, the country gradually moved to a state-based economy characterized by a small private sector and a large oil industry. In 2013, after more than 14 years as president, Hugo Chávez died and Nicolás Maduro succeeded him. With Maduro as president, in 2015, the Venezuelan economy suffered a sizable negative shock as oil prices dropped by almost 50%. With the primary source of government revenue reduced, the government reduced funding for social programs and subsidies for essential products, like medicines and food. The social discontent exploded in 2017 when the Maduro party won the majority of state elections with apparent signs of fraud. At this point, massive flows of Venezuelans started leaving the country due to the political and economic conditions there.

To give a sense of the economic crisis in Venezuela, in 2018, it reached five-digit hyperinflation ($\approx$ 65,000%) accompanied by an extensive economic deterioration. The gross domestic product (GDP) decreased by two digits yearly from 2016 and, in 2019, reached an all-time low of –34% (IMF, 2020). A 2019 independent survey from three universities measured that 96.2% of all Venezuelans were poor and 79.3% were extremely poor (UCAB, 2020). In this context, the Venezuelan exodus began, both with voluntary and involuntary immigration. As of 2019, more than 1.1 million working-age Venezuelans were in Colombia, according to the GEIH survey (see Table 1b). Most of these Venezuelans plan to stay in Colombia for more than one year (RAMV, 2018). It is also important to note that Venezuelans speak the same language as Colombians, and most of them enter the country through the terrestrial frontiers, without formal documentation of their previous education level or experience (RAMV, 2018).

2.2 Regulatory Framework for Venezuelans

Before 2018, all Venezuelans needed a special visa to work legally. This visa needed a sponsor company to grant a temporary residence. Other work visas were also granted when a sufficiently large investment was made. However, in the second half of 2018, the Colombian government implemented a substantive change in the work regulation of Venezuelans that provided a new framework to work, called a Special Permit of Permanence (PEP, by its acronym in Spanish).[9] The PEP's goal was to foster legal and more accessible employment for Venezuelans without the need for sponsor companies or investments.[10] This policy was the most extensive migratory amnesty program offered to undocumented migrants in recent history and highlights the importance of open border policies to the Colombian government. A short-term study of this policy indicates insignificant effects on several labor market outcomes, such as monthly wages, unemployment, and participation in the labor market for natives (Bahar, Ibáñez, and Rozo, 2021).[11] Even if the short-run impacts of this policy seem to be insignificant, the policy is part of the overall effect of the Venezuelan immigration on natives' labor market outcomes. Still, with the dynamic specification proposed, I find significant wage effects even before this legalization happened in 2018.

3 Data

Two primary datasets are used in this paper. The first is the GEIH survey and the second is the Colombian Census of Population and Housing (CNPV, by its acronym in Spanish) done in Colombia between January and October of 2018. GEIH is a monthly cross-sectional survey that characterizes 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

[9] 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 who were registered in the Administrative Record of Venezuelan Migrants (RAMV, by its acronym in Spanish).

[10] The PEP was initially valid for 90 days and could be renewed for up to two years.

[11] In Ecuador, Olivieri et al. (2020) find that providing work permits to Venezuelan workers would increase their average earnings.

by the National Statistics Office of Colombia (DANE, by its acronym in Spanish) and are available on their webpage.

In 2013, DANE implemented a migration module that contained questions on where the person was born, where they lived 12 and 60 months ago, and the reasons for migrating. In this study, I use data from 2013–2019 to differentiate natives from migrants for the main outcomes (wages and employment).[12] Furthermore, for the treatment variable (immigration rates), the CNPV is used to reduce the extent of measurement error of immigrants, and the bias that can arise in traditional surveys or even in the US censuses (Aydemir and Borjas, 2011; Amior, 2020). Finally, for robustness I show wage and employment results using the static measure of migration from the census and the time-varying migration measure from GEIH. Overall, I prefer the census data because the GEIH survey is built to represent natives only, not migrants. In addition, its weights are built using the 2005 census adjusted by population projections, and since the survey weights have not been updated with the 2018 census, there is additional uncertainty with the survey.

Supplementary databases are used to construct external instruments and additional outcomes. The first one is the RAMV survey, which characterized the population of undocumented Venezuelans in Colombia.[13] 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 (2021) that uses demographic information from the last census in Venezuela to predict immigrants' origin. I also use information from chambers of commerce (Confecamaras), which collects all the new registered firms in Colombia with tax records, to construct the information of newly registered firms. Last, I use price data from DANE; the prices are representative of 23 capital cities in Colombia, with a monthly collection at the store level.

3.1 Descriptive Statistics for Natives and Migrants

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

[12] Moreover, 2011 and 2012 are also included to have additional pre-treatment periods, assuming all survey respondents in those years were Colombian.

[13] Nearly 443,000 individual records were gathered from April 6 to June 8 in 2018 at different points in all the territory. It was an optional and go-to-the-registration-point kind of survey for undocumented Venezuelans.

Colombia in the last year, and the third corresponds to Colombians who lived in Venezuela and then returned to Colombia when the crisis started. For the causal inference, I focus only on Colombians who did not migrate in the previous year from Venezuela. This is because the sample size for the Venezuelan migrants and Colombian returnees, especially when split by region-year cell, turns to be very small. With this in mind, Appendix A presents a table with descriptive statistics regarding the age profile, level of education, and gender composition for the different groups, according to the different years of arrival.

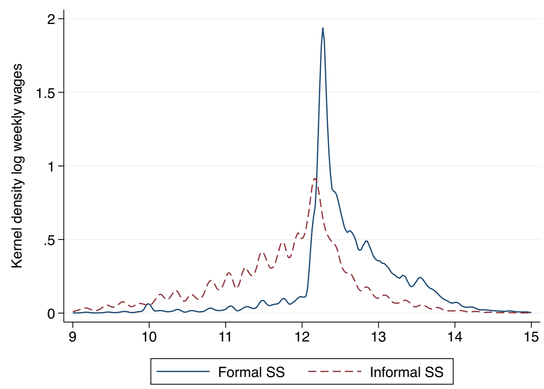
Several stylized facts stand out. First, Venezuelan immigrants arriving in Colombia tend to be young, although their average age at arrival seems to be growing. Before 2017, the highest share of arrivals was in the range of 0–14 years, and after 2017, the range changed to 14–28 years. Second, returning Colombians are more concentrated in older ages; before 2016, the majority were 15–28 years old. After 2016, the predominant age group was between 41–64 years. Third, in terms of education, the three groups have the highest share of individuals with no high school degree. In particular, returning Colombians have the lowest share of tertiary education, while Venezuelans and Colombians have similar shares of education and, likely, of skills. Suggesting that migrants can be substitutes, more than complements, of natives. Another relevant takeaway is that the Venezuelan arrivals seem to be more educated in the latest years. Fourth, in terms of gender composition, the shares of men and women are similar. Finally, before 2017, returning Colombians were the main group coming from Venezuela, but afterward, Venezuelans greatly surpassed this group and became the predominant immigrant group.

3.2 Labor Market Structure

The Colombian labor market's structure is characterized by the interdependence of two main employment categories. The first category, mainly happening at small firms, is one without binding minimum wages or contributions to social security (i.e., the pension and health system). This category of employment commonly belongs to the *informal sector*. The second category is characterized by the existence of a binding minimum wage and contributions to social security. This category belongs to the *formal sector*. Not all the workers in the formal (informal) sector are high (low) skilled: there is a combination of both types of skills in each sector, in which most of the workers in the formal sector are skilled and most in the informal sector are unskilled.

In this paper, I define informal employment based on whether the worker contributes to either the health or pension system. Figure 1 plots the wage density for the two sectors. The figure shows strong bunching around the minimum wage in the formal sector, which is much less pronounced in the informal sector. Thus a large portion of workers in the formal sector has a binding restriction; this "stickiness" helps explain why there cannot be wage losses for formal workers in response to a labor supply shock.

Figure 1: **Wage density for the two sectors of employment**



Note: Informality is defined according to social security (SS). In this figure all the data on wages are stacked across periods and departments. In addition, the sample is restricted to ages between 18 and 64 years old, and no survey weights are used. Log weekly wages are in real terms using monthly CPI from DANE. The kernel function is Epanechnikov, and I use the optimal bandwidth. Source: GEIH 2013 to 2019.

Tables 1a and 1b show labor force statistics for Colombians and Venezuelans, considering jointly males and females. Two important findings stand out. First, most Colombian workers are informally employed. There has been a downward trend in the informality rate in the last years, decreasing from 58.2% in 2013 to 51.7% in 2019. The opposite occurs for Venezuelan workers, where the same rate increases from 62.1% to 89.1% in the same period. Raw data indicates that almost all new arrivals of Venezuelans got employment in the informal sector.

Second, comparing both Venezuelans and Colombians, in 2019 Venezuelans have a higher labor force participation rate (81.8% versus 74%), a higher employment rate (69.4% versus 65.4%), and a higher unemployment rate (15.1% versus 11.6%); this occurs every years after 2015 (see Tables 1a and 1b). The higher employability of migrants could be associated with lower reservation wages compared to natives and a more inelastic labor supply (Borjas, 2017). Last, the Colombian labor market has a substantive portion of self-employed workers (44.2% of all native workers in 2019 were self-employed) and relative to Venezuelans the share is slightly higher (48.4% in 2019). Importantly, the wage analysis throughout the paper covers all types of labor income, not just salaried workers' wages.

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

(a) Colombians (in percentages)

	LFP	Employment	Unemployment	Informality	Self-employment	N (15-64)	Population
2013	75.3	67.6	10.2	58.2	47.3	356,597	23,336,824
2015	75.6	68.1	9.9	55.5	45.4	474,871	24,022,452
2017	75.0	67.0	10.7	53.6	45.3	462,484	24,516,791
2019	74.1	65.5	11.6	51.9	44.2	444,442	24,221,821

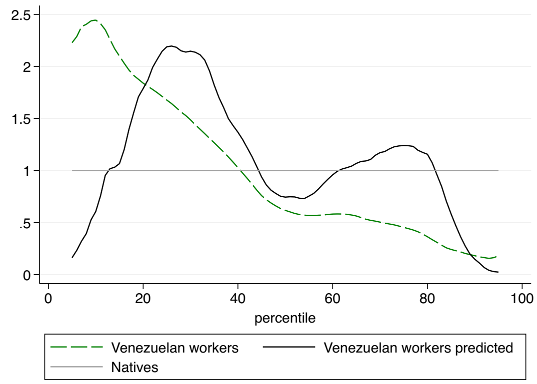
(b) Venezuelans (in percentages)

	LFP	Employment	Unemployment	Informality	Self-employment	N (15-64)	Population
2013	80.5	68.8	14.5	62.1	46	583	36,364
2015	73.6	65.8	10.7	65.5	49.2	959	50,802
2017	81.5	68.6	15.9	82.0	51.2	4,112	206,427
2019	81.8	69.4	15.1	89.1	48.4	18,440	1,114,666

Note: LFP stands for labor force participation. In this table, the rates are calculated using national survey weights from GEIH. In addition, the sample is restricted to the population ages 15–64 years in urban areas. Panel (a) is restricted to natives living in Colombia for more than one year. Informal employment rate's is calculated as the proportion of workers who are informally employed over total employment. Source: GEIH, 2013-II to 2019.

Last, Figure 2 compute the extent of the migrants' occupational downgrading, comparing their actual and predicted wages in terms of the native wage distribution. The results shows that migrants are overrepresented in the bottom of the distribution, meaning they are more likely to earn lower wages compared to natives and are underrepresented in the upper part of the distribution. Figure 2 also compares Venezuelan predicted wages, according to the returns to education and experience of natives, with their actual wages. This measure of migrant downgrading indicates that migrants consistently downgrade (if the solid and dashed line coincide over the distribution, this would mean no downgrading).

Figure 2: **Position of Venezuelans in the native wage distribution**



Note: This figure plots the actual wages and predicted wages for Venezuelan immigrants. The log wage equation is constructed separately for men and women, including schooling years, age, age squared, potential experience, and dummies of month as controls. I restrict the sample to natives between 18 and 64 years old in urban areas. Source: GEIH-2018.

3.3 Choice of Base Period

For the event study analysis, the year 2015 is selected to compare pre- and post-treatment periods. To choose this period, I use the timeline of employed Venezuelans in Colombia and political incidents around the border of Colombia and Venezuela. Starting in August of 2015, the Venezuelan government unilaterally closed the national border, restricting exits from its country. One year later, in August of 2016, it decided to re-open the border again. Due to the conditions explained above, the number of immigrants in Colombia suddenly increased due to the political and economic crisis in Venezuela that had worsened over the succeeding years (see Figure 3a). For simplicity and to remove seasonal effects, I select 2015 as the base year of comparison, which was the last year before the massive increase of Venezuelans in Colombia.[14] Using data until 2019 is an improvement concerning Caruso, Canon, and Mueller (2021). The authors analyzed this migration event using data up to 2017, but migrants' arrivals doubled in 2018 and again in 2019. Appendix A shows that this limitation is critical to explain the sizable negative wage effect in their study.[15] Figure 3a also shows that exports to Venezuela were following a decreasing pattern before the migration crisis and that imports from Venezuela were small in all periods of analysis, reducing the concern of trade as a potential confounder in the analysis.

[14]Until 2019, the number of immigrants does not seem to have yet reached a peak. Nevertheless, with the COVID-19 pandemic, a significant number of Venezuelans are returning to their home country (Reuters, 2020).

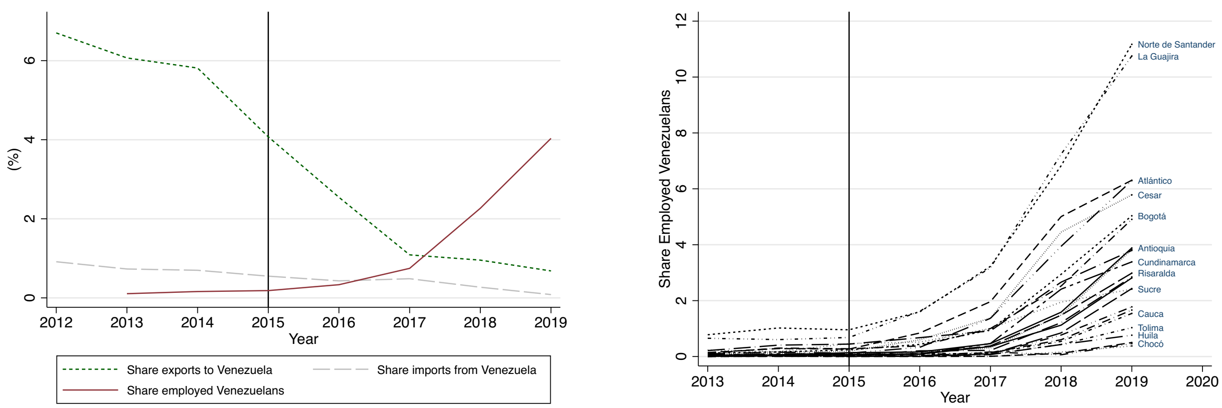
[15]A recent paper by Lebow (2021b) summarizes the differences in the wage effects of immigration in Colombia from available studies.

Next, I analyze the share of employed Venezuelans by departments, an administrative division in Colombia similar to states in the US. Again, there is a small number of Venezuelans in the pre-treatment years, followed by an increase that varies greatly across the 24 departments (see Figure 3b), motivating the spatial approach undertaken here. Last, using the static TWFE (i.e., $y_{it} = \alpha_i + \alpha_t + \beta^{DiD} D_{it} + e_{it}$) in the presence of the differential timing of a discrete policy and heterogeneous treatment effects might yield biased estimates (Goodman-Bacon, 2021). Hence, given the variation in this setup, the effect is studied year by year in an event study design.

Figure 3: **Share of Venezuelans in Colombia at the national and department level**

(a) Migration and trade with Venezuela

(b) Migration by departments



Note: Panel (a) is at the national level. The share of exports is constructed as total exports to Venezuela over total exports to the world. The share of those employed is constructed as employed Venezuelans over all the employed population, both between 18 and 64 years. For clarity, I only include the names of a few departments. Source: GEIH and *Exportaciones*, 2013-II to 2019.

4 Empirical Specification

4.1 Event Study

The empirical strategy exploits the differential intensity of Venezuelan arrivals across departments in Colombia. Departments have a local labor market in its capital city that is connected with the surrounding smaller cities, with some degree of independence from labor markets in other departments.[16] One critique of this empirical strategy is that it might not reflect the true impact of immigration if there exists mobility responses of inputs, say of native workers or capital, from

[16]When restricting the analysis to capital cities only, wage estimates are slightly similar with more negative coefficients.

areas more affected by the labor supply shock to areas less affected (Aydemir and Borjas, 2011). In my setting, I can test this hypothesis to show that there are not fewer inflows of natives coming to the most affected areas (a downward trend is noticeable, only significant in 2019), but more of an outflow effect of natives leaving the most immigrant affected areas (see Appendix Figure A.4). If this is the case, the results on wages might be attenuated by this out-migration response.

In terms of the empirical specification, I use an event study regression that performs differences between pre- and post-treatment years t, with respect to a base period, for the different departments d. Thus, I estimate the following equation, in which the omitted year is 2015:

$$Y_{dt} = \gamma_d + \gamma_t + \sum_{t=2011, t\neq 2015}^{2019} \beta_t \Delta M_{d,2018} + u_{dt}. \tag{1}$$

$\Delta M_{d,2018}$ is a time-invariant treatment variable, constructed from the 2018 census records.[17] I use this constant immigration rate to motivate the event study research design and to examine its validity, as it has the advantage of integrating the placebo tests on pre-trends easily in the analysis. This constant rate is not problematic since the arrivals of migrants between departments remain constant over time, with a nearly perfect correlation over the years. Therefore, selecting the best possible measure for this immigration rate makes sense. In terms of interpretation, for $2015 < t < 2018$, β_t measures the impact for outcome periods that are less than complete, as defined in Autor, Dorn, and Hanson (2021). This reveals how long it takes for the immigration shock to affect the outcomes. For $t > 2018$, β_t measures the impact on outcomes after the immigration shock is realized. This reveals whether the shock attenuate over time. With this in mind, $\Delta M_{d,2018}$ is defined as follows:

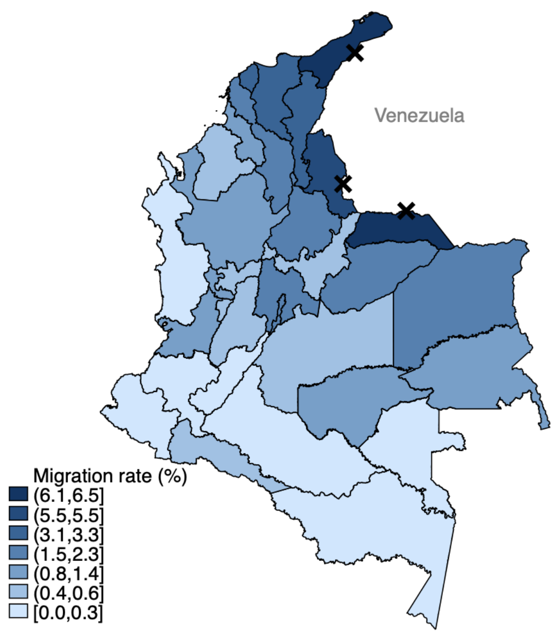
$$\Delta M_{d,2018} = \frac{L_{Ven,d,2018} - L_{Ven,d,2015}}{L_{Total,d,2018}}, \tag{2}$$

where the numerator is the stock 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 stock of employed Venezuelans in d whose year of arrival was 2015. The denominator $L_{Total,d,2018}$ is the total employed population in the department. Finally, the regression has fixed effects of year γ_t and of department γ_d.

[17] By construction, $\beta_{2015} = 0$ and all coefficients $\beta_t \in t = 2011, ..., 2019$ measure the effect relative to 2015.

Figure 4 plots the Colombian map with the immigration rate $\Delta M_{d,2018}$ by departments. The highest immigration rates are observed in the departments closer to the Venezuelan border, especially near the main crossing bridges.[18] According to RAMV, more than two-thirds of the undocumented Venezuelans in Colombia entered through the Paraguachón and Simón Bolívar bridges (see the **X** in Figure 4).

Figure 4: **Spatial distribution of Venezuelans by departments**



Note: This map plots the immigration rate $\Delta M_{d,2018}$. To characterize recent Venezuelan migrants, the census asked if the person lived in the last 12 months in Venezuela. I take only Venezuelan-born migrants in the rate's numerator. The **X** represents the three main crossing bridges that are: Simón Bolívar International Bridge in Norte de Santander, Paraguachón International Bridge in La Guajira, and Páez Bridge in Arauca. Source: CNPV 2018.

In this setup, the standard errors are clustered at the department level as there could be an arbitrary serial correlation of the errors within each department. Also, sample weights of department are not used in the main specification because the heteroskedasticity test proposed by Solon, Haider, and Wooldridge (2015), which consists of regressing the square residuals on the inverse sample size, gives an insignificant estimate.[19] One downside of this setup is the small number of

[18] The data for the outcomes are available to 24 departments, not to all the 33 in the country. The missing nine 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.

[19] Regression weights can also be used to estimate population average partial effects. However Solon, Haider, and Wooldridge (2015) state that this is not straightforward, bringing up arguments for using or not using the weights. In any case, when using regression weights I find similar estimates for native employment and more negative for native wages.

treated units ($N = 24$), which can increase type I error rates (Pustejovsky and Tipton, 2018). For this reason, to control for the over-rejection of the null hypothesis, the wild cluster bootstrap method of Roodman et al. (2019) is implemented, reporting the p-value for the main estimates. Moreover, an alternative definition of local labor markets ($N = 51$) is built from GEIH to find that the estimates are, if anything, more negative compared to the department sample ($N = 24$), showing the number of treated areas does not drive the negative wage and employment coefficients.

4.2 Instrumental Variable

If migrants self-select into the departments where the economic conditions are better, $\Delta M_{d,2018}$ is going to be endogenous and the OLS estimates downward biased.[20] To address this, I estimate equation (1), instrumenting it with two exogenous variables: (i) the distance between capital cities in the two neighboring countries and (ii) past settlements of Venezuelans in Colombia.

The distance instrument is based on Del Carpio and Wagner (2015) and Caruso, Canon, and Mueller (2021), and is constructed as follows:

$$z_{1,d} = \sum_{s} \frac{\lambda_s}{T_{s,d}} * M_{2018}, \tag{3}$$

where $T_{s,d}$ is the road distance in kilometers from the capital city of state s in Venezuela to the capital city of 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. M_{2018} are the arrivals of Venezuelans to Colombia in 2018.[21]

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 a city in d to a city in o, in the sense that travel distance poses a time and economic restriction to new immigrants. A threat to this identification strategy

[20] Jaeger (2007) and Borjas (2001) have pointed out that immigrants tend to settle in areas that offer the best economic opportunities for the skills they provide.

[21] As a robustness check, I develop the test proposed in Goldsmith-Pinkham, Sorkin, and Swift (2020) to calculate the weights of the distance instrument, namely the *Rotemberg* weights, of the overall coefficient. The instrument is decomposed into 15 shares arising from the different origin cities of migrants in Venezuela to determine which of them gets more weight in the overall estimate. This exercise yields that Maracaibo and San Cristobal concentrate 0.7 of the weights, which sums up to 1. In fact, those cities are closer to the border with Colombia; hence, effectively, the instrument is comparing cities closer to the border with ones further away. To name some characteristics of these migrants, Maracaibo is an industrial city focusing on oil extraction, while San Cristobal relies more on the service and commerce sector.

arises if the border states suffer more economic shocks, such as less trade than the counterpart far-located states (violation of the exclusion restriction). To deal with these concerns, Figure 3a shows that the trade shock with Venezuela started in the years before the immigration shock. In the post-treatment years, especially after 2016, the share of exports with Venezuela is regularly around zero.

Later, I show that the wage and employment estimates are more negative and still significant when including export patterns with Venezuela as a control (measured as the share of total exports in USD to Venezuela over total exports in 2015 for every department), besides, the control coefficient is insignificant. I also include a control of department GDP (a proxy of business cycles) to show that estimates for wages and employment are similar, yet this can be a "bad control" (Angrist and Pischke, 2008, p. 49). Last, the dynamic specification does placebo tests on differences in pre-treatment outcomes and I find insignificant pre-trends for wages and employment.[22] Hence, the main concerns of the exclusion restriction of this instrument are alleviated.

The past settlement instrument is based on Altonji and Card (1991) and Card (2001), and is constructed as follows:

$$z_{2,d} = \left(\frac{Ven_{d,2005}}{Ven_{2005}} * M_{2018} \right) / L_{d,2015}, \tag{4}$$

where the first 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 as in Card (2001), 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 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 because the lagged immigrant location is related to new arrivals (relevance) but is not related to current economic conditions (exogeneity). However, this assumption might fail if local economic trends are highly serially correlated, such that the labor demand shifts that attracted immigrants in the past are still correlated with contemporaneous demand shifts. In Appendix A I show the results for wages and employment of natives using shares from the census of 1973, and they are not significantly altered. Last, in dynamic settings,

[22]For employment, the pre-trends are insignificant only for the distance instrument.

past settlement instruments tend to be serially correlated, capturing previous immigration adjustments (Jaeger, Ruist, and Stuhler, 2018). In my study, it is possible to break the serial correlation because arrivals surge rapidly.

Formally, the exclusion restriction of the instruments used in this study can be expressed as $E[z_{i,d}\Delta u_{dt}] = 0$ for $i = 1, 2$.[23] With this in mind, the first-stage regression for both instruments is the following:

$$\Delta M_{d,2018} = \xi_i + \eta_i z_{i,d} + \upsilon_{i,d} \quad i = 1, 2, \tag{5}$$

where $\upsilon_{i,d}$ captures the endogenous component of $\Delta M_{d,2018}$. The results of this regression are presented in Table 2. The distance instrument explains 88.4% in the immigration rate's variation, while the past settlement instrument explains 49.7%. The positive coefficient of the distance instrument indicates that cities closer to the border receive more migrants than cities located farther away.[24]

Table 2: **First stage: The Inflow of Venezuelans and the two instruments**

	(1) $\Delta M_{d,2018}$	(2) $\Delta M_{d,2018}$
Distance ($z_{1,d}/100$)	0.376***	
	(0.030)	
Past settlement ($z_{2,d}$)		34.321***
		(5.757)
Constant	-1.271***	0.740**
	(0.233)	(0.225)
N	24	24
R^2	0.884	0.497
F st	157.9	35.5

Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This table reports the coefficients of the first-stage regression of the instruments with the immigration rate $\Delta M_{d,2018}$*100. Since the immigration rate $\Delta M_{d,2018}$ from the census and the instruments are time invariant, the first stage is the same in all the years analyzed.

[23] The first step to evaluate the exogeneity of the two instruments and possible heterogeneous effects is to perform a Hansen J test for over-identifying restrictions. In this case, both instruments are used in the first-stage regression 5 to find that the null hypothesis of the instruments being exogenous is not rejected ($p - value = 0.85$).

[24] Recently, Lee et al. (2020) argues for a higher F-statistic in the first stage, exactly a value of around 104.7. In this case, the distance instrument's F-statistic is 157.9, and the past settlement instrument's F-statistic is 35.5.

5 Baseline Estimates

In this section, I report the baseline estimates of the overall effect of migration on wages and employment of natives. I then demonstrate that these baseline estimates hide an interesting asymmetry in the effects across the informal and formal sectors of the labor market.

5.1 Wage Responses

I first regress equation (1) for log hourly wages of natives under two methods (OLS and IV) on the explanatory variable $\Delta M_{d,2018}$, which measures the share of employed Venezuelans in Colombia according to the 2018 census. One of the main advantages of the event study design, compared to the panel regression in Caruso, Canon, and Mueller (2021) and Lebow (2021a), is the possibility to test for previous trends in the outcome and eventually control for them if they exist. The first finding is that pre-trends for wages are not statistically significant (see Figure 5a).[25] Note that despite being non-significant, the OLS point estimates indicate that immigrants are going to areas with rising wages, though this selection gets more or less corrected with the IV estimates.[26]

Figure 5a shows that 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 2). When using the two instruments separately, the wage estimates are again negative and significant.[27] A one pp 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. Both point estimates are significant using a wild cluster bootstrap method (see Table 3). I also show in the Appendix that when using residual wages (having controlled for individual characteristics such 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 a 1.7 pp increase in the employed population in Colombia (in absolute numbers, this means around 254,000 more employed

[25] Instruments predict a high pre-treatment coefficient in 2011, which does not seem to be a big problem since trends have been relatively stable after 2011. A joint F-test for coefficients from 2012 to 2014 yields a p-value of between 0.27 and 0.42 depending on the instrument.

[26] The construction of the wage (or labor income) variable requires some steps that are shown in Appendix A. Importantly, it covers all types of labor income, including self-employed earnings.

[27] I do not combine both instruments in the first stage as the distance instrument captures all the predictive power of the past settlement instrument.

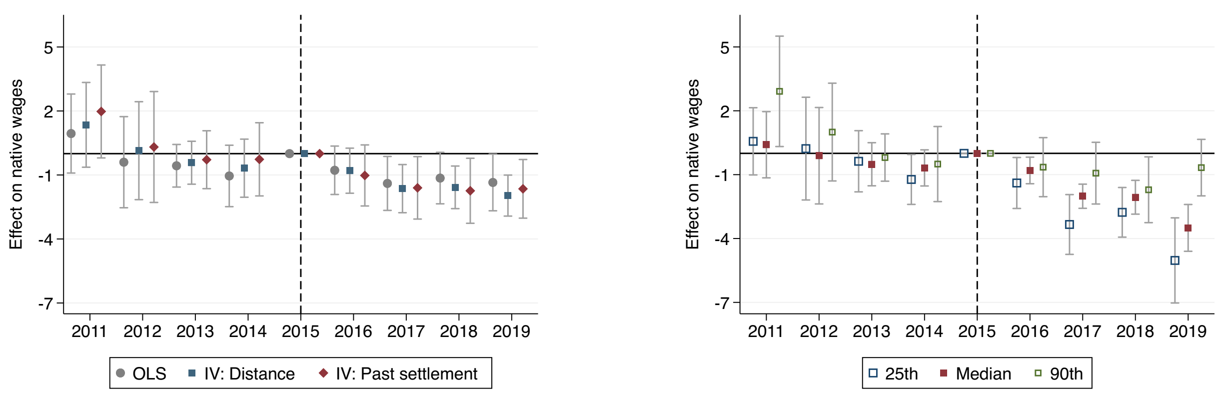
Venezuelans relative to 2015). Hence, the total impact on wages for 2018 is between –2.7% and –3%, depending on the instrument selected. Note that the shock can be understated (and the effect overstated) because the census recollection ended in October of 2018, omitting the arrivals of Venezuelans in November and December of that year.[28]

Next, I analyze whether there are distributional effects of migration by taking as outcome certain percentiles of the distribution of wages. The results of this exercise, where the outcome variable is the log wage at different percentiles in year t, are plotted in Figure 5b. In effect, the native wages at the bottom part of the distribution are most affected by immigration. At the median, which is a more robust estimate to outliers and censored data, the coefficient is of –2.1% for a one pp increase in the immigration rate $\Delta M_{d,2018}$. Comparing these results with the UK, Dustmann, Frattini, and Preston (2013) document that immigration depresses native wages below the 20th percentile and contributes to wage growth above the 40th percentile. However, over their analysis period, immigrants in the UK are more educated than natives.

Figure 5: **Event study estimates on log hourly wages of natives**

(a) Mean wages

(b) Percentiles of wages



Note: In a first step, I use department survey weights from GEIH to construct department level outcomes from individual data of natives between 18 and 64 years in urban areas. In a second step, I use the department level data to estimate β_t from equation (1) and use clustered standard errors at the department level. I use a 95% confidence interval. In 2011 and 2012, I assume all survey respondents are Colombian. In panel (b), instead of the average wage in each department, I use the value at given percentiles of the local wage distribution as outcome. The F-statistic for the distance instrument is 157.9 and for past settlements instrument is 35.5. In panel (b) the instrument used is distance. Hourly wages are in real terms using the monthly CPI from DANE.

[28]Returning Colombians are not included in the immigration rate as they arrived mainly in the years before 2016, while after 2016, most of the immigrants were Venezuelans (see Appendix Table A.1c). Thus, when taking differences of immigrants between 2018 and 2015 in the immigration rate for my specification, the variation stems from Venezuelan immigrants, not returning Colombians. This is conceptually different from the panel regression specification of Lebow (2021a) that uses a five-year immigration rate.

I next use the time-varying immigration rate ΔM_{dt} built from the GEIH survey as the explanatory variable (see Appendix Figure A.7a). Interestingly, the estimates follow a similar pattern (negative and significant) but are different in magnitude from those obtained with the fixed $\Delta M_{d,2018}$ from the census. In 2018, the census year, I find estimates that are slightly more positive with the survey than with the census, in part due to the measurement error of migrants and, consequently, due to the attenuation bias. In 2019, the year with the highest immigration rates, I find similar estimates to 2018, thus there is not a substantially larger effect in 2019, even if more migrants are observed than in 2018. Last, the correlation of the yearly immigration rate ΔM_{dt} from GEIH with other post-treatment years is between 0.93 and 0.98, highlighting the fact that using the constant $\Delta M_{d,2018}$ from the census still captures the dynamics of the treatment.

The explanation for such a negative finding hinges on several factors pointing to the high substitutability of natives and migrants. In effect, migrants speak the same language like natives, overcoming communication skills problems. In addition, they share cultural traits, which can reduce wage discrimination, and most come as forced migrants (which implies a relatively low reservation wage) and without certified education or home experience (which implies downgrading of tasks). The government has also pushed an open policy border, facilitating work permits. Finally, wage flexibility in the informal sector can lead to significant wage cuts when migrants arrive.

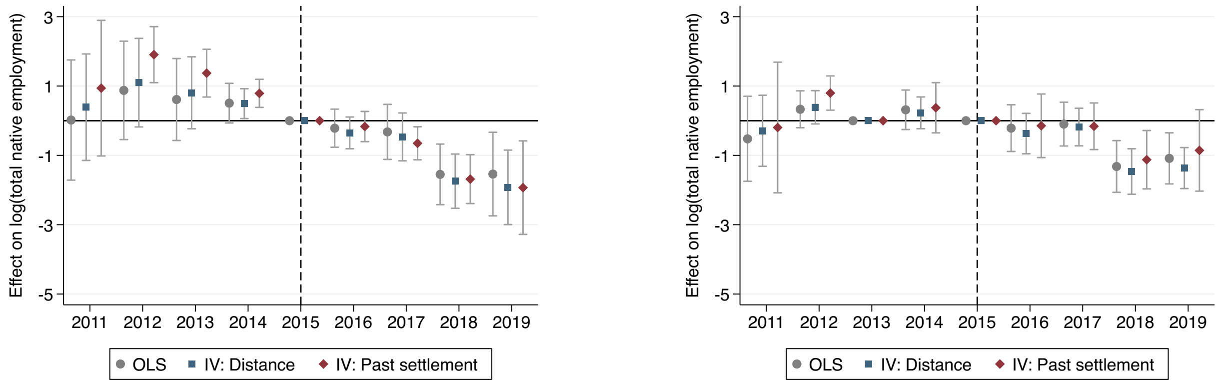
5.2 Employment Responses

I then regress equation (1) for log native employment using OLS and the two instruments. Figure 6b shows that, in contrast to wages, for the past settlement instrument there are significant differences in the trends before the migration event happened (see Figure 6a).[29] This indicates that past settlements of migrants, but not distance to Venezuela, would predict local native employment in the pre-policy period, suggesting a violation of the PTA with IV required for identifying the causal parameter. To address this problem, I focus on the distance instrument in the next sections, but also control for the pre-trends 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). By construction, the coefficient of 2013 is now zero ($\beta_{2013} = 0$).

[29] A joint F-test do not reject the null hypothesis of pre-treatment coefficients equal to zero for the distance instrument (p-value $= 0.1338$). However, for the past settlement instrument I do reject it (p-value $= 0.00$).

Figure 6: **Event study estimates on log employed natives**

(a) Employment (b) Employment with trend adjustment



Note: In a first step, I use department survey weights from GEIH to construct department level outcomes from individual data of natives between 18 and 64 years in urban areas. In a second step, I use the department level data to estimate β_t from equation (1) and use clustered standard errors at the department level. I use a 95% confidence interval. Trend-adjusted estimates have, as a control, the growth in employment from 2013 to 2015. The F-statistic for distance instrument is 157.9 and for past settlements instrument is 35.5.

In Table 3, I show results for native employment and wages, with and without the adjustment for pre-trends, while testing for the significance with the wild cluster bootstrap method. After adjusting for pre-trends, a one pp increase in the immigration rate reduces, on average, 1.5% local native employment using distance as the instrument and by 1.1% using past settlement as instrument (coefficients of Table 3, column 4). The coefficient of the distance instrument is the only significant according to the wild cluster bootstrap method.[30]

[30] Appendix Figure A.7b shows that when using ΔM_{dt} as the explanatory variable, instead of the fixed $\Delta M_{d,2018}$ from the census, the results of native employment are similar but with wider confidence intervals in 2017.

Table 3: **Wages and employment estimates for natives, 2015–2018**

	(1)	(2)	(3)	(4)
	Wages		Employment	
Panel A: OLS				
	-1.146	-0.962	-1.548**	-1.320*
	(0.540)	(0.595)	(0.391)	(0.334)
Wild cluster bootstrap p-value	0.147	0.222	0.01	0.028
Panel B: IV				
Distance instrument				
	-1.579*	-1.464	-1.744**	-1.466*
	(0.446)	(0.497)	(0.417)	(0.358)
Wild cluster bootstrap p-value	0.03	0.052	0.008	0.02
First stage: F-st	157.9	228.3	157.9	152.6
Past settlement instrument				
	-1.741*	-1.669	-1.685	-1.125
	(0.682)	(0.655)	(0.375)	(0.459)
Wild cluster bootstrap p-value	0.032	0.184	0.194	0.280
First stage: F-st	35.5	55.2	35.5	19.1
Trend-adjusted	No	Yes	No	Yes
N	24	24	24	24

Standard errors are in parentheses.

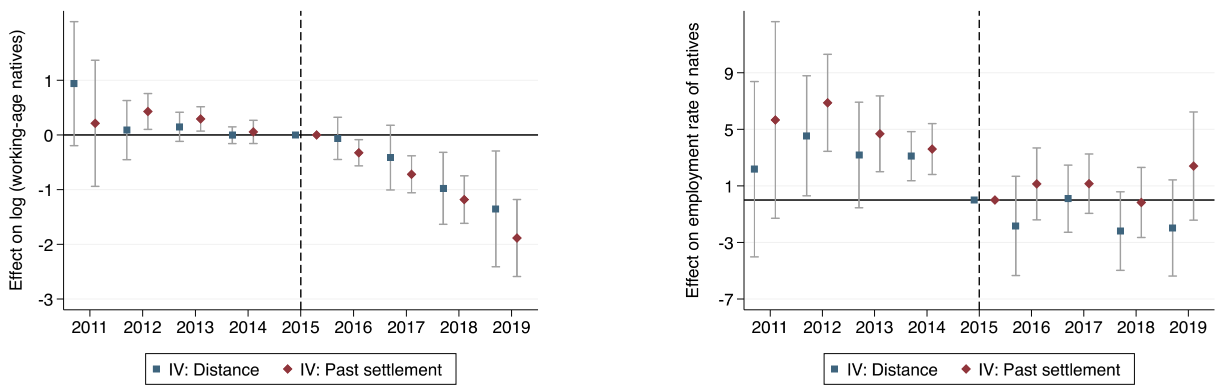
Stars according to bootstrap p-value * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This table reports the coefficients of the second-stage regression of the instruments with the immigration rate $\Delta M_{d,2018}$. The coefficient measures the effect in 2018 relative to 2015. In a first step, I use department survey weights from GEIH to construct department level outcomes from individual data of natives between 18 and 64 years in urban areas. In a second step, I use the department level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Trend-adjusted 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. I compute wild bootstrap p-values from the `boottest` command in Stata using 999 bootstrap replications (Roodman et al., 2019). The stars of significance are according to these p-values and not according to the Standard errors are in parentheses.

The drop in native employment (in levels) is part of the impact on working-age natives (i.e., the sum of the unemployed, employed, and inactive population). In Figure 7a, the dependent variable is the change in log working-age natives relative to the base period. This Figure shows that for 2018, a one pp increase in the immigration rate $\Delta M_{d,2018}$ decreases, on average, the number of working-age natives in the local area by 1% and 1.2% (wild cluster bootstrap p-value is 0.042 and 0.022, respectively). Such a significant out-migration response of Colombian natives is an important margin of adjustment to the Venezuelan immigration. This might dampen the spatial approach undertaken here, but as Venezuelan arrivals are so sharp, this approach manages to capture the overall effect in the short run. This is different when analyzing long-run impacts when firms can adjust flexibly. To conclude, if there are negative impacts on native employment and working-age natives, the effect on the employment rate should be small (as both the numerator and denominator are decreasing). Reassuringly, Figure 7b shows that the impact on the native employment rate and the effect is insignificant.

Figure 7: **Event study estimates on working-age population and employment rate for natives**

(a) Working-age natives (b) Employment rate natives



Note: In panel (a) dependent variable is log working-age natives and in panel (b), it is employment rate of natives (in %), both relative to the base period. In a first step, I use department survey weights from GEIH to construct department level outcomes from individual data of natives between 18 and 64 years in urban areas. In a second step, I use the department level data to estimate β_t from equation (1) and use clustered standard errors at the department level. I use a 95% confidence interval. In 2011 and 2012, I assume all survey respondents are Colombian.

5.3 Results with Alternative Definition of Local Labor Markets

As mentioned earlier, one potential shortcoming of this setup is the small number of treated areas per year ($N = 24$). To address this, I build a larger comparison sample defined as functional urban areas (FUAs) in Colombia using the methodology of Sanchez-Serra (2016). This sample consists of the 53 biggest urban areas in the country defined from population grid data, municipal boundaries, and inter-municipal commuting flows. I use geocoded surveys to construct the outcomes, noting that sampling errors can increase considerably as the GEIH does not represent these areas.[31]

This exercise compares the estimates for 24 departments and 51 FUAs constructed with the GEIH. A slightly different distance instrument is used since the sample is expanded. The instrument is the second polynomial of the distance from a given area to the nearest crossing bridge with Venezuela, similar to the instrument used in Dustmann, Schönberg, and Stuhler (2017). Regarding

[31] This municipal location variable of the survey is not publicly available. I had access to a remote computer with the location variable to retrieve results.

natives, when focusing on the FUAs, the coefficients tend to be more negative, especially for wages. For the departments, the wage estimate is –1.4%, while for the FUAs, the corresponding estimate is –2.1%.[32] Overall, the estimates I find 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.

Table 4: **Native wages and employment estimates for different samples, 2015–2018**

	(1) Native Wages	(2) Native Employment
Panel A: Departments		
IV: Distance to nearest crossing bridge	-1.42**	-1.41***
	(0.570)	(0.390)
F-st	16.27	14.93
N	24	24
Panel B: FUAs		
IV: Distance to nearest crossing bridge	-2.11***	-1.85**
	(0.540)	(0.620)
F-st	14.31	13.30
N	51	51
Trend-adjusted	No	Yes

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

Note: This table reports the coefficients of the second-stage regression of the predicted immigration rate from the GEIH survey with the distance instrument. The instrument is the second polynomial of the distance to the nearest crossing border with Venezuela. The coefficient measures the difference in 2018 relative to the base period. I restrict the sample to natives 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, and thus the coefficients are interpreted as percentages. Hourly wages are in real terms using the monthly CPI from DANE.

6 Labor Market Linkages between the Formal and Informal Sector

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

[32] The result of departments is slightly different from before because I am using a distinct instrument with the immigration 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 Borjas (2006).

6.1 Overall Impacts on the Informal and Formal Sector

In this section, I report results for wages and employment at the regional-level using the distance instrument because the pre-trends for native employment above were insignificant. Figure 8a plots the estimated effect of immigration on wages of formally and informally employed natives.[34] To begin, informally employed natives suffer the largest wage losses (2018 coefficient is –1.9% with wild bootstrap p-value of 0.05), while the coefficient for regional formal wages is insignificant.[35] There is no occupational upgrading, or wage increase, for natives in the formal labor market as found for Turkey (Aksu, Erzan, and Kırdar, 2018; Del Carpio and Wagner, 2015). The language is a crucial difference for Syrians in Turkey, while it is an advantage for Venezuelans as they speak the same language as Colombians.

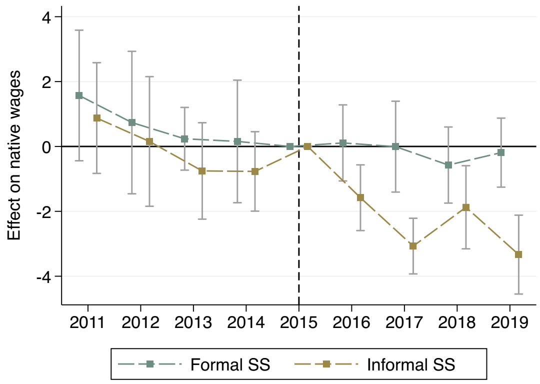
Across the two sectors, there is a drop of native formal employment (2018 coefficient is –2.2% with wild bootstrap p-value of 0.02), while native informal employment presents insignificant estimates. The role of the minimum wage can help to explain these findings since formal workers have a high probability mass to the right of the minimum wage (see Figure 1), and thus there cannot be further wage drops. This might translate into higher employment losses as informal wages are freely adjusted and formal-informal workers can be substitutable for the firm. As a robustness check, the reliability of the survey data is tested by comparing it with the social security records of Colombia. The Appendix shows that using both sources yields fairly similar wage and employment estimates (see Figures A.1a and A.1b).

[34] Because the results are aggregated, this effect is capturing possible spillover effects between the two sectors. That is, workers moving from formal to informal employment or *vice versa*.

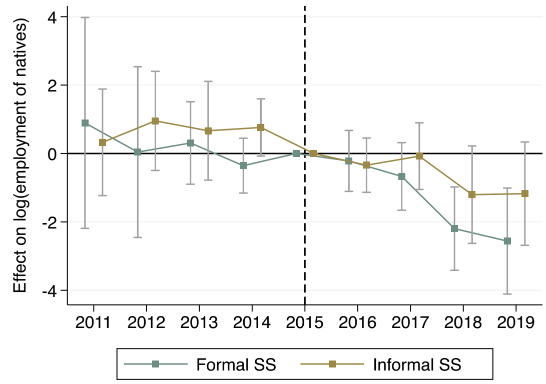
[35] Appendix Figure A.2 shows similar wage estimates using more frequent time windows (quarters instead of years). Also, pre-trends are not significant in this specification.

Figure 8: **Event study estimates by sector for natives**

(a) IV (distance) estimates on native wages



(b) IV (distance) estimates on native employment

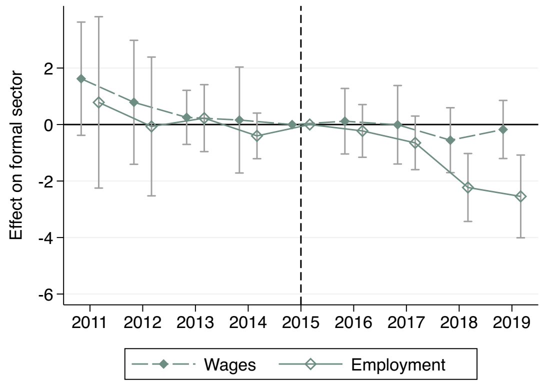


Note: In a first step, I use department survey weights from GEIH to construct department level outcomes from individual data of natives between 18 and 64 years in urban areas. In a second step, I use the department level data to estimate β_t from equation (1) and use clustered standard errors at the department level. I use a 95% confidence interval. In panel (b) I do not use controls for pre-trends. The F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

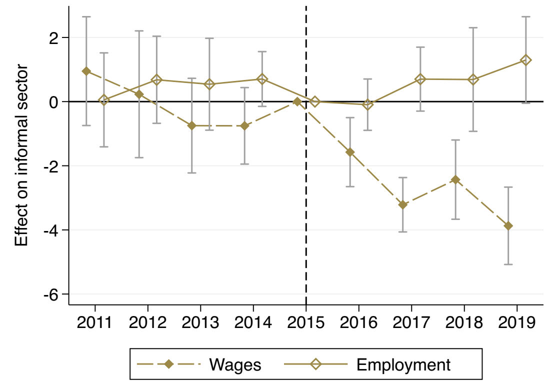
Next, the focus is on overall employment (natives plus migrants) and total wages (wages of both natives and migrants) across the informal and formal sectors. The main takeaway is that there are insignificant effects for formal wages, supporting the hypothesis of a binding minimum wage, while for informal wages I find more negative wage estimates (see Figure 9a and 9b). In addition, formal workers get similarly crowded out after 2017, while informal employment now increases when adding migrants, yet not significantly.

Figure 9: **Employment and wage estimates by sector for natives plus migrants**

(a) IV (distance) formal sector



(b) IV (distance) informal sector



Note: In a first step, I use department survey weights from GEIH to construct department level outcomes from individual data of natives between 18 and 64 years in urban areas. In a second step, I use the department level data to estimate β_t from equation (1) and use clustered standard errors at the department level. I do not explicitly control for pre-trends. I use a 95% confidence interval. In 2011 and 2012, I assume all survey respondents are Colombian.The explanatory variable is $\Delta M_{d,2018}$. The F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

6.2 Interaction of Formal and Informal Employment

To show that firms combine formal and informal labor in production, I first use a cross-sectional survey from 2019 that distinguishes between informality at the firm or worker level (EMICRON, by its acronym in Spanish). This exercise aims to describe how different the workforce composition of firms can be, in terms of formal and informal labor, depending on their size. On average, all informal firms (that do not pay any taxes) hire mostly informal workers (who do not contribute to the social security system). This contrast with formal firms, where formal and informal labor are combined in their production function. Interestingly, the proportion of formal workers increases as the size of the firm gets larger (see Table 6). This combination of formal and informal labor thus depends mainly on firm size, as documented for Brazil (Ulyssea, 2018) or Mexico (Samaniego de la Parra, Bujanda et al., 2020).

Table 5: **Composition of workers by firm type and size**

	Informal firms	**Formal firms**
Workers	Share of informal workers (%)	Share of informal workers (%)
1	98.7	87.9
2	98.7	80.3
3	98.5	75.0
4	97.8	75.3
5	99.1	62.3
6	94.4	59.7
7 to 9	98.4	47.8

Note: In this table, formal firms pay taxes, while informal firms do not pay taxes and are not publicly registered. Also, informal workers do not contribute to the health or pension system. The shares are calculated using survey weights from the EMICRON survey. Only owners of firms with less than 10 workers are surveyed. Firms with 7 to 9 workers are aggregated due to the small sample. Source: EMICRON-DANE, 2019.

Next, I investigate what is the typical firm size of employed immigrants. In 2018, nearly 90% of Venezuelan immigrants were employed in firms with less than twenty workers. Clearly, immigrants are overrepresented in the smallest firms and this finding is consistent with other contexts. For instance, Arellano-Bover and San (2020) find that former Soviet Union Jews arriving in Israel in the 1990s were employed in the smallest firms.

Table 6: **Share of employed migrants by firm size**

Firm Size	Distribution of migrants (%)
1 worker	38.0
2 to 3 workers	20.7
4 to 5 workers	14.1
6 to 10 workers	10.7
11 to 19 workers	4.7
20 to 30 workers	3.2
31 to 50 workers	1.7
51 to 100 workers	1.1
101 or more workers	5.7

Note: In this table the shares are calculated using national survey weights from GEIH. The category of one worker refers to unipersonal firms or self-employed workers. I restrict the sample to all Venezuelan formal or informal workers from ages between 18 and 64 years in urban areas. Source: GEIH, 2018.

Last, focusing only on formal firms, I show how binding the minimum wage can be depending on firms' size. For the smallest formal firms, the share of formal workers who earn the minimum wage is 60.9%, while for the largest firms, the percentage is 9.3% (see Table 7). These three findings indicate that the migration shock should be more salient in smaller firms, where formal and informal labor is combined, most immigrants are employed and most formal workers earn a binding minimum wage.

Table 7: **Composition of workers by firm size**

	Formal firms
Workers	Share of formal workers earning the MW (%)
1-4	60.9
5-9	50.0
10-19	41.2
20-49	32.7
50-99	27.1
100-999	20.6
1,000 and more	9.3

Note: The share is constructed as the total number of formal workers earning the minimum wage over total formal employment for every firm size. MW refers to minimum wage. Informal workers are not captured with this administrative data; only workers who contribute as employees are taken into account. Source: PILA, 2015-August.

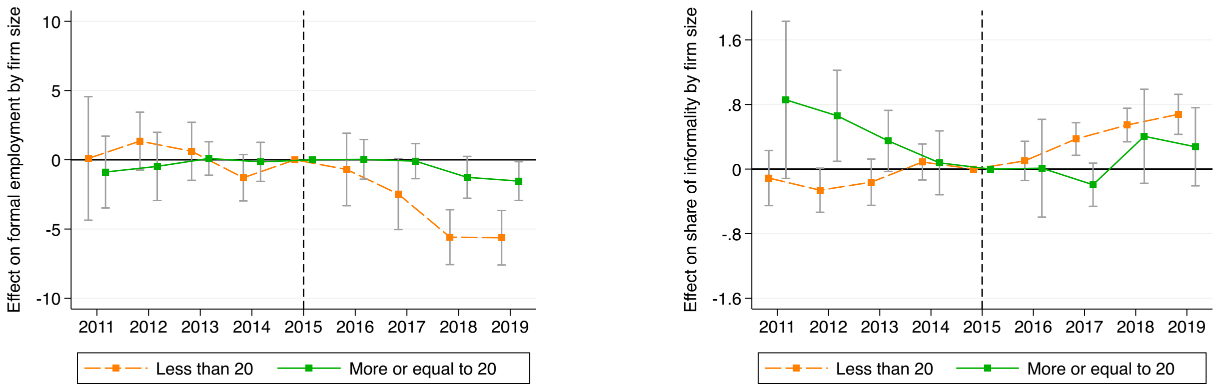
Exploiting these facts, workers are aggregated in two firm size categories: less than 20 workers (smaller firms) and 20 workers or more (bigger firms). As shown in Figure 10a, where the dependent variable is log formal employment, smaller firms significantly reduce formal employment in response to the immigration shock. Moreover, Figure 10b shows that the share of informal workers in smaller firms increases more pronouncedly compared to bigger firms.[36] This suggestive evidence indicates a possible change in the composition of the firms' workforce, with relatively more informal workers than formal workers.

[36]Informal employment significantly increases for 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: In a first step, I use department survey weights from GEIH to construct aggregate outcomes from individual data of natives between 18 and 64 years in urban areas. In a second step, I use the department level data to estimate β_t from equation (1) and I use clustered standard errors at the department level. I do not explicitly control for pre-trends. I use a 95% confidence interval. In 2011 and 2012, I assume all survey respondents are Colombian. The F-statistic for distance instrument is 157.9.

For robustness, I check if the previous results are driven by the entry of formal firms or by an increase in the size of existing firms. First, for firm entry, there is a positive effect on firm creation in 2016, which becomes insignificant in 2017 and 2018 (see Appendix Figure A.5a). Thus, employment changes by firm size are less driven by the entry of new formal firms.[37] For the second one, in the smaller firms, there seems to be a decrease in the average number of workers, while in larger firms, the opposite happens (see Appendix Figure A.5b).[38] Overall, these findings support that employment changes are mainly due to changes in the workforce composition of smaller and larger firms.[39]

6.3 Model

To help interpret the empirical findings, I build a partial equilibrium model inspired by Ulyssea (2018) with two labor inputs, formal and informal labor. Different from Ulyssea (2018), I flexibly allow for a degree of substitution between these inputs that will be key to rationalize the findings.

In this model, the crucial distinction between the labor inputs arises from their differential

[37] In comparison to Turkey, Altındağ, Bakış, and Rozo (2020) find that the large refugee shock of Syrians boosted firm creation in the country, especially for those with foreign partnerships.

[38] I measure the average number of workers who report to work in a firm size category using the GEIH survey.

[39] Another mechanism to be studied is the effect on the creation of informal firms not registered in tax records.

cost. 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.[40] Moreover, the firm can be a self-employed worker who chooses to be a formal or informal worker. The profit function of the firm can be written as

$$\max_{L_i, L_f} \pi = pF(L_i, L_f) - \tau(L_i)w_i L_i - (1+\tau_f)w_f L_f, \tag{6}$$

where $\tau(L_i)$ represents 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 $\eta \geq 1$, which captures the cost of evasion related to law enforcement exerted by the government.[41] This function also admits the possibility of no informal labor cost (i..e, $\eta = 0$), whereas τ_f represents the payroll taxes that the firm must enroll when paying for formal workers.

Specifically, the production function has a constant elasticity of substitution (CES) form:

$$F(L_i, L_f) = Q = (\alpha_i L_i^{\rho} + \alpha_f L_f^{\rho})^{\frac{1}{\rho}}, \tag{7}$$

where $\sigma = \frac{1}{1-p}$ (with $\rho \leq 1$) is the elasticity of substitution between formal and informal workers, that is allowed to be different than one like in , 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.[42]

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

$$(\tau'(L_i)L_i + \tau(L_i))w_i = C^{1-\epsilon}\epsilon\alpha_i L_i^{\rho-1}(Q)^{\epsilon-\rho}, \tag{8}$$

$$(1+\tau_f)w_f = C^{1-\epsilon}\epsilon\alpha_f L_f^{\rho-1}(Q)^{\epsilon-\rho}. \tag{9}$$

[40]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). This means the model does not account for the firm's decision to register in the tax records (become a formal firm), as the goal is to model informality through the worker side and analyze changes within a given firm.

[41]Fines for hiring workers informally can go up to 500 minimum wages and are enforced by the Ministry of Labor in Colombia.

[42]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.*

Assuming that firms are competitive in the labor market and therefore take prices w_i and w_f as given. I then analyze changes in optimal labor choices within the firm when wages for informal workers decrease as a result of the immigration shock (I take formal wages as fixed in the short run due to the minimum wage).[43]

After some algebraic derivations in Appendix A, I derive the elasticities of informal and formal labor with respect to informal wages (ε_{L_i,w_i} and ε_{L_f,w_i}) as

$$\varepsilon_{L_i,w_i} = -\frac{(1-\epsilon s_f - \rho s_i)}{(1-\epsilon)(1-\rho)+\eta(1-\epsilon s_f - \rho s_i)}, \tag{10}$$

$$\varepsilon_{L_f,w_i} = -\frac{s_i(\epsilon-\rho)}{(1-\epsilon)(1-\rho)+\eta(1-\epsilon s_f - \rho s_i)}. \tag{11}$$

First, note that the denominator in both expressions is always non-negative as $\epsilon \leq 1$, $\rho \leq 1$, and $\eta \geq 1$, 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 elasticity of labor substitution σ (function of ρ) and the price elasticity of demand ϵ^D (function of ϵ), which determines the sign of its numerator. Two special cases arise:

1. If $\epsilon^D > \sigma$, there are *scale effects* and firms' response to lower informal wages will be to hire more formal and informal labor. For instance, when the product market is competitive (i.e., $\epsilon = 1$) and formal-informal workers are imperfect substitutes ($\rho < 1$), firms increase production using both labor inputs without affecting p.

2. If $\epsilon^D < \sigma$, there are *substitution effects* and firms' response to lower informal wages will be to hire less formal and more informal labor, depending on the values of η and s_g.[44]

In graphical terms, Figure 11 depicts the equilibrium in both labor markets (f and i), with the characteristic that f has a binding minimum wage (or price ceiling) that distorts the equilibrium in

[43] A model proposed by Kleemans and Magruder (2018) distinguishes between types of skills for formal and informal workers. The main findings are that if migration is highly skilled, there can be a crowd out effect of existing formal low-skilled workers (with formal wages staying constant) and that migration (of any skill type) will (weakly) decrease wages in the informal sector.

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

that market. Thus, if formal wages are fixed, the change in equilibrium employment in each sector can be calculated as follows:

$$\frac{\Delta L_i}{L_i} = \frac{\Delta w_i}{w_i} * \varepsilon_{L_i,w_i}, \tag{12}$$

$$\frac{\Delta L_f}{L_f} = \frac{\Delta w_i}{w_i} * \varepsilon_{L_f,w_i}. \tag{13}$$

With a labor supply shock that increases informal labor supply, wages of informal labor w_i will decrease while total informal employment L_i will increase (see Figure 11). In Figure 9a, I find a negative effect on formal employment; this finding is consistent with *substitution effects* (i.e., $\epsilon^D < \sigma$). This corresponds to the case where formal wages are fixed and formal-informal workers are strong substitutes relative to the price elasticity of demand or, according to Borjas (2013), the two-good economy framework, relative to the consumer substitution among available goods.[45]

After analyzing the mechanisms and comparative statics of the model, I use reduced-form estimates to recover the values of the own-elasticity of informal labor demand and the cross-elasticity of formal labor demand. First, the ratio between the change of total informal employment and the native informal wage change yields the elasticity of labor demand $\varepsilon_{L_i,w_i} = \frac{0.69\%}{-1.87\%} = -0.37$.[46] This estimate is measuring the slope of the demand curve (see Figure 11), and, reassuringly, it lies in the range of values of labor demand elasticities previously found in the literature (Lichter, Peichl, and Siegloch, 2015; Hamermesh, 1996).[47]

Next, the ratio between the total change of formal employment and native informal wage change yields $\varepsilon_{L_f,w_i} = \frac{-2.22\%}{-1.87\%} = 1.19$, suggesting formal and informal workers might be close to perfect substitutes in production. Finally, I extend this theoretical framework with a general equilibrium model in Appendix A. This model includes capital responses, labor supply elasticities of migrants and natives, and joint employment/wage responses to immigration. Importantly, to match the empirical findings in this model, there must exist heterogeneous labor supply elasticities for formal

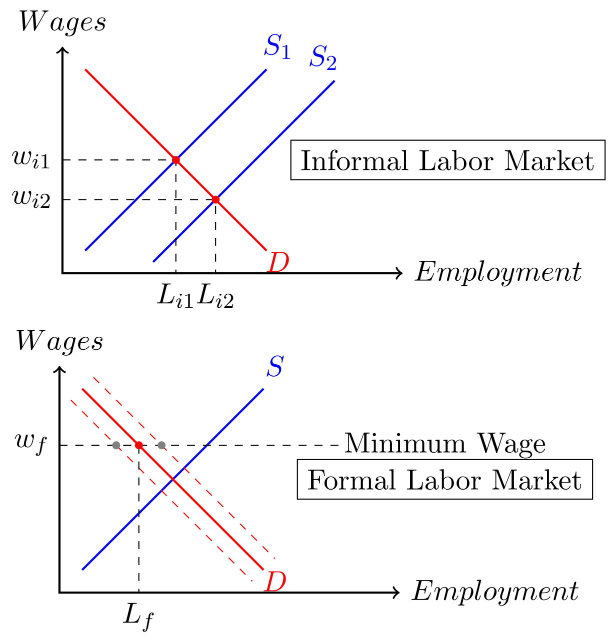
[45] Borjas (2013) states a quasi-linear utility function for the consumer in terms of a locally produced good and an imported good. It can be assumed that ϵ matches the substitution parameter for these two goods.

[46] I use native wage change in 2018 and not overall wage change to remove possible compositional bias from immigrants' lower wages. If, instead of the total change in employment, I include in the numerator the change in native employment, I would be estimating the labor supply elasticity.

[47] In Italy, Guriev, Speciale, and Tuccio (2019) find an elasticity of labor demand in the informal market 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.

and informal workers and a high degree of substitutability between them.

Figure 11: **Local market responses to a supply shift when immigrants and natives are perfect substitutes**



Note: Employment in each market refers to natives and migrants employed. There are two possible 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.

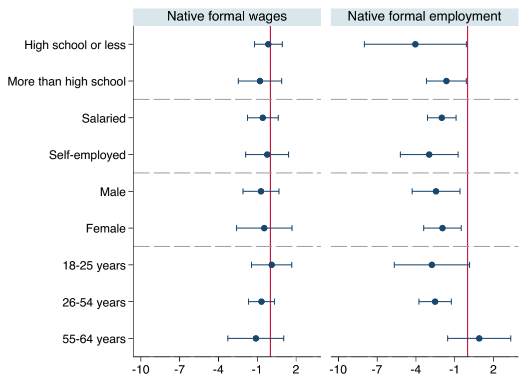
7 Heterogeneous Impacts and Further Adjustment Channels

7.1 Effects by Education, Job Type, Gender, and Age for Natives

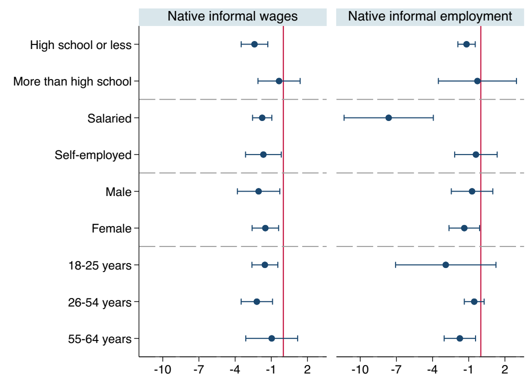
This subsection describes the heterogeneous effects of immigration across different native subpopulations. Because the immigration shock causes asymmetric employment and wage effects across the informal and formal sectors, I first focus on the heterogeneous analysis by these two sectors. Figure 12a shows that wage effects in the formal sector are consistently around zero for the different subgroups, while the employment effects are negative in most of them, especially for low-skilled natives (with high school or less). On the other hand, Figure 12b shows that wage effects in the informal sector are mostly negative and significant (only natives that are high-skilled and older present an insignificant estimate). For the employment effects in the informal sector, salaried natives have the most negative coefficient, while for self-employed natives I find an insignificant effect close to zero.

Figure 12: **Native wages and employment estimates by subpopulations and by sector, 2015–2018**

(a) IV (distance) formal sector



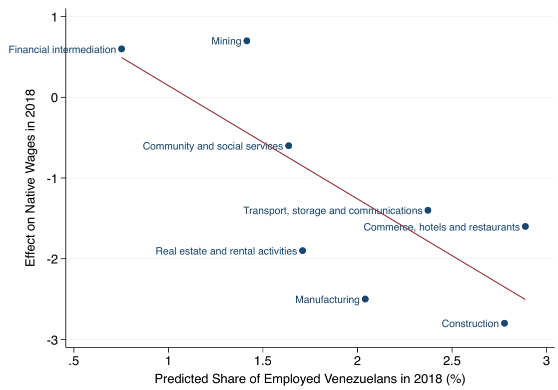
(b) IV (distance) informal sector



Note: These figures report the coefficients of the second-stage regression of the immigration rate $\Delta M_{d,2018}$ with the distance instrument. The outcome is log hourly wages or log employment. Standard errors are clustered at the department level. I use department survey weights from GEIH 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. The F-statistic for distance instrument is 157.9. Hourly wages are in real terms using the monthly CPI from DANE.

Finally, I exploit area-industry variation by eight industries in the Colombian economy. Figure 13 plots the estimated effects against the predicted share of employed Venezuelans in those industries. The figure shows that the industries with larger shares of employed Venezuelans experience the largest declines in native wages.

Figure 13: **Native wage effects by industry**



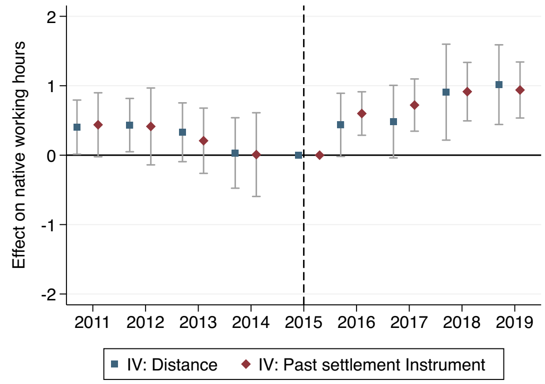
Note: The coefficient measures the difference between 2018 and 2015. I remove agricultural industries as the analysis is restricted to the urban population, and electricity, gas, and water supply due to the small sample available. I calculate the predicted share regressing the actual share of Venezuelan workers in each industry for 2018 against the share of informal workers in each industry in 2015. I restrict the sample to permanent Colombian residents between 18 and 64 years in urban areas. I use department survey weights from GEIH to construct aggregate outcomes. I use past settlements as an instrument; its first-stage F-statistic is 35.5. Hourly wages are in real terms using the monthly CPI from DANE.

7.2 Intensive Margin Effect

Next, I estimate the impact of the immigration shock on natives' labor supply decisions at the regional-level, measured as the usual amount of working hours per week in each region. Figure 14 shows that natives are working more hours in all the post-treatment years. However, it is not possible to rule out compositional effects of part-time workers moving relatively more than full-time workers out of employment.[48]

[48] I find that for employment, part-time workers have a more negative point estimate than full-time workers, but both estimates are insignificant.

Figure 14: **Event study estimates on log working hours of natives**



Note: I use a 95% confidence interval. I restrict the sample to permament Colombian residents between 18 and 64 years in urban areas. I use department survey weights from GEIH 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. The F-statistic for distance instrument is 157.9 and for past settlement instrument is 35.5.

7.3 Impact on Prices

The wage and employment responses from immigration are not the only channels of adjustment in the presence of a supply shock. To complement the findings, I also analyze the price response of a bundle of goods and services. Two opposing mechanisms can affect prices. The first is the demand effect that arises from the relatively higher consumption of goods and services from migrants. The second 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 some markets, driving prices down. Hence, the final impact on prices is ambiguous.[49]

In Appendix Figure A.3a, I find insignificant estimates close to zero on the overall consumer price index (CPI), suggesting a relatively stronger supply effect from the immigration shock. Complementary, I compare the impact on nominal and real wages to find strikingly similar effects. Still, these findings do not imply that the immigration shock does not increase demand at all, probably additional government transfers or increased consumption of goods and services increase the local demand, but this is partly offset in the price analysis by the supply and search channels mentioned

[49] Previous studies have found that low-skilled immigration in the US reduces prices of non-tradable goods and services where migrants are more likely to compete with natives (Cortes, 2008); the mass inflow of Soviet Union immigrants to Israel in the 1990s lowered the price of goods, from higher price elasticity and lower search costs of immigrants (Lach, 2007); and immigration in the UK decreases the prices of non-tradable goods and services while prices of tradable goods increase (Frattini, 2008).

above. This is consistent with the theoretical model of Borjas (2013), which states that in the short-run the supply effect dominates the demand effect if capital is fixed.

8 Robustness Checks

The first robustness check relates to excluding areas on the Venezuelan border from the analysis. Note that these departments could be more affected by the Venezuelan crisis through fewer trade links or lower business interactions. The main results of this exercise, displayed in Appendix Figures A.6a and A.6b, yield no significant variations in the estimate. In addition, I restrict the sample to natives residing in the current department for more than one year, removing possible compositional bias of natives reallocating between departments due to the Venezuelan immigration. This is an important check since, according to Figure 7a, there is an out-migration response of natives. Importantly, when imposing this sample restriction, the estimates tend to be more negative (see Appendix Table A.3). Finally, to account for unobservable shocks related to proximity to the Venezuelan border, I explicitly control with quintiles of distance to the nearest crossing bridge with Venezuela. The coefficients are all less negative, especially for wages using distance as the instrument (–1.3% versus –1.6%). This difference, however, is plausible as there can be a correlation between the instrument and the quintiles of distance used.

The second robustness test targets the identifying assumptions of the instruments. For past settlements, there might be a correlation between the distribution of Venezuelans in 2005 and current economic trends. To test this, I construct two further historically lagged census shares of Venezuelans in Colombia from IPUMS (2019). Appendix Figure A.8 displays the coefficients using that instrument with three distinct shares (2005, 1993, and 1973) and shows 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 this, I use the change in real log GDP 2015–2018 and share of total exports in USD to Venezuela with respect to the world as controls. The corresponding estimates of their effect on native wages, reported in Appendix Table A.3, confirm the statistical significance of these effects (with a less negative point estimate for GDP and a more negative one for trade).

Finally, I compute residual wages in a preliminary stage to check whether aggregating outcomes without netting out individual observables is important in the analysis. As shown in Appendix Table A.3, the residual wage has a slightly more negative estimate for the distance instrument and a slightly more positive one for the past settlement instrument. Thus, there is little gain in regressing previous individual characteristics in the event study analysis.

9 Conclusion

This paper analyzes the impact of the Venezuelan mass migration on the Colombian labor market. By comparing departments with different exposure to the Venezuelan immigration, I find a negative effect on hourly wages and employment for natives. The differences between my findings with previous studies in Colombia can be explained by the specification I choose, an event study design with IV that rules out potential confounding issues; the immigration rate used, which is built with census data, reducing the measurement error of migrants; and the number of periods analyzed, which captures the dynamic effects of the treatment more accurately.

The massive influx of Venezuelan immigrants in Colombia causes a decrease in natives' informal wages, since immigrants are mostly employed in the informal sector, and a decrease of natives' formal employment, in part due to the binding minimum wage. Thus, this labor supply shock affects the informal and formal sectors differently, the former via wage reductions and the latter via employment reductions. To explain these findings, I develop a model of a firm that can substitute formal for informal labor if formal and informal workers are strong substitutes as a response to lower informal wages. Overall, in settings with large informal labor markets, migration can lead to asymmetric responses between the informal and formal sectors, especially when the minimum wage is binding in the formal sector. Future research may focus on firm-level outcomes to shed light on the consequences of substituting formal for informal labor. For instance, how this affects production or productivity within the firm.

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.

Altındağ, O., Bakış, O., and Rozo, S. V. (2020). Blessing or burden? impacts of refugees on businesses and the informal economy. *Journal of Development Economics*, 146:102490.

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.

Arellano-Bover, J. and San, S. (2020). The role of firms in the assimilation of immigrants. *Available at SSRN 3594778*.

Autor, D., Dorn, D., and Hanson, G. H. (2021). On the persistence of the china shock. Technical report, National Bureau of Economic Research.

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., Ibáñez, A. M., and Rozo, S. V. (2021). Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees. *Journal of Development Economics*, 151:102652.

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. (2006). Native internal migration and the labor market impact of immigration. *Journal of Human resources*, 41(2):221–258.

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

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

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

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. (2021). Spillover effects of the venezuelan crisis: migration impacts in colombia. *Oxford Economic Papers*, 73(2):771–795.

Ceritoglu, E., Yunculer, H. B. G., Torun, H., and Tumen, S. (2017). The impact of syrian refugees on natives? labor market outcomes in turkey: evidence from a quasi-experimental design. *IZA Journal of Labor Policy*, 6(1):1–28.

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

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Del Carpio, X. V. and Wagner, M. C. (2015). The impact of Syrian refugees on the Turkish labor market. *World Bank policy research working paper*, (7402).

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. (2020). The impact of immigration on wage dynamics: Evidence from the algerian independence war. *Journal of the European Economic Association*, 18(6):3210–3260.

Edo, A. and Rapoport, H. (2019). Minimum wages and the labor market effects of immigration. *Labour Economics*, 61:101753.

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

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

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.

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.

Kleemans, M. and Magruder, J. (2018). Labour market responses to immigration: Evidence from internal migration driven by weather shocks. *The Economic Journal*, 128(613):2032–2065.

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

Lebow, J. (2021a). Immigration and occupational downgrading in Colombia. *Unpublished Manuscript*.

Lebow, J. (2021b). The Labor Market Effects of Venezuelan Migration to Colombia: Reconciling Conflicting Results.

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: Venezuelan Migrants in the Ecuadorian Labor Market.

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.

RAMV (2018). Informe Final Registro Administrativo de Migrantes Venezolanos en Colombia. Technical report. Available on: https://www.refworld.org.es/pdfid/5b2957524.pdf (last accessed 28th of November 2021).

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.

Rozo, S. V. and Vargas, J. F. (2021). Brothers or invaders? how crisis-driven migrants shape voting behavior. *Journal of Development Economics*, 150:102636.

Samaniego de la Parra, B., Bujanda, L. F., et al. (2020). Increasing the Cost of Informal Workers: Evidence from Mexico. Technical report.

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

Santamaria, J. (2020). ?When a Stranger Shall Sojourn with Thee?: The Impact of the Venezuelan Exodus on Colombian Labor Markets.

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 Online Appendix

A. Descriptive Statistics

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
2015	0.267	0.239	0.170	0.246	0.078	0.493	0.583	0.191	0.176	783,888	46,627,550
2017	0.260	0.238	0.171	0.248	0.084	0.493	0.563	0.207	0.183	761,148	47,456,897
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
2015	0.558	0.269	0.111	0.062	0.000	0.486	0.508	0.121	0.146	475	28,667
2017	0.401	0.362	0.179	0.056	0.001	0.510	0.493	0.211	0.201	3,577	205,277
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
2015	0.165	0.312	0.285	0.225	0.012	0.540	0.629	0.292	0.071	1,062	51,436
2017	0.151	0.198	0.283	0.305	0.063	0.488	0.670	0.256	0.072	1,504	84,412
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. Shares do not sum up to 1 because missing values. In this table, the shares are calculated using national survey weights from GEIH. College refers to all the technical levels of education after high school. In panels (b) and (c) I restrict to the population that responded they were living in Venezuela in the last year. Source: GEIH, 2013-II to 2019.

Appendix Table A.2 shows in which industries Venezuelans and Colombians are working. Immigrants are overrepresented with respect to natives in two industries. The first one is commerce, hotels, and restaurants, where almost half of all Venezuelan workers have a job (46.9%), while the corresponding share for Colombians workers is $\approx$ 30%. The second one is in construction (11.1% versus 7%). Conversely, immigrants are underrepresented relative to natives in two main industries of employment: real estate, business, and rental activities (6% versus 9.5%) and community, social and personal services (15.6% versus 23.5%). I also compute the share of informal workers within industries to find that for agriculture, commerce, and construction, the proportion of informal workers is around 70% of all workers.

Table A.2: Distribution of workers by place of birth and informality across industries

Industry	Colombians	Venezuelans	Share of Informality
Agriculture, livestock, hunting, forestry and fishing	3.6	1.4	73.7
Mining and quarrying	0.7	0.4	28.5
Manufacturing industry	13.5	12.3	49.6
Electricity, gas and water supply	0.6	0.2	6.1
Construction	7.0	11.1	67.4
Commerce, hotels and restaurants	30.1	46.9	71.4
Transport, storage and communications	9.6	5.6	58.1
Financial intermediation	1.9	0.6	13.8
Real estate, business and rental activities	9.5	6.0	46.1
Community, social and personal services	23.5	15.6	37.5
N (Workers, 18-64)	1,979,144	24,706	2,003,850

Note: In this table, the shares are calculated using national survey weights from GEIH. The shares for columns 1 and 2 should sum up to 100% after adding the unknown occupation share. Column 3 refers to the share of informal workers within each industry. I restrict the sample to permanent Colombian residents and Venezuelans between 18 and 64 years in urban areas. Source: GEIH, 2013-II to 2019.

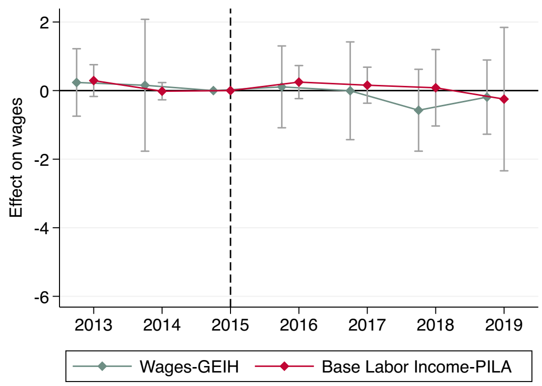
B. Comparison of Administrative and Survey data

The administrative data to affiliation to social security (PILA, by its acronym in Spanish) is used as an additional robustness check. Importantly, PILA covers only formal employment hence the comparison population of the GEIH survey is workers that contribute to the health system. Moreover, the municipal information in the administrative form of the worker is not verified by national authorities until 2019. Hence, the regional estimates using PILA might not be the first-best option. Still, to have consistent estimates, the measurement error of this variable must be uncorrelated with the selected instrument. The results for formal wages with the administrative data yields closely similar ones when using the GEIH survey.[A.1] For formal employment, I find negative coefficients on both sources, yet not significant for PILA, in part due to the measurement problem of the location variable.

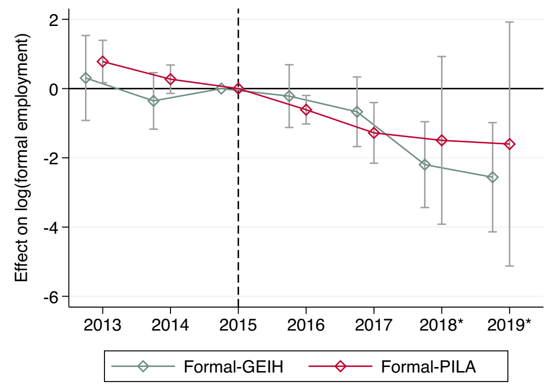
[A.1] In PILA is basic income and refers to the amount of labor income it is declared in the contributions of the social security. The information was retrieved on the seventh of May of 2021.

Figure A.1: 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 the administrative data. Note: I use a 95% confidence interval. The sample is not restricted and covers the universe of formal workers in the department. The department survey 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

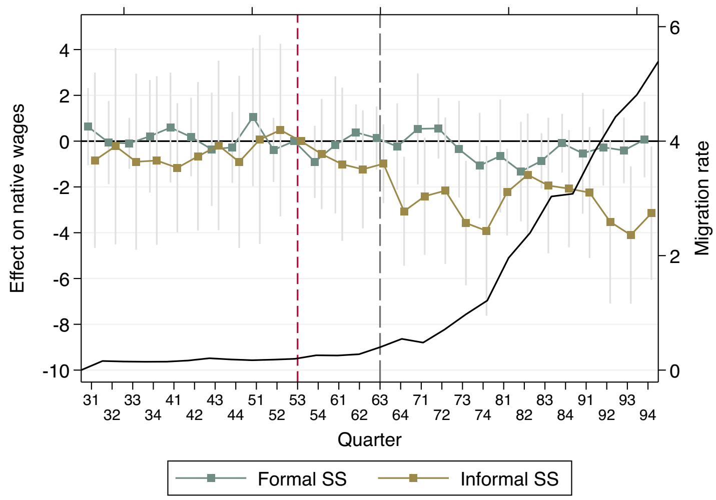
I perform the previous analysis at a higher time-frequency, using quarters instead of years. The base period of comparison is 2015-3, when the Venezuelan government closed the border. The empirical specification is the same as before, interchanging the yearly subscript y with the quarterly one q. I estimate the following model in an IV setup using the distance instrument:

$$Y_{dq} = \gamma_q + \gamma_d + \sum_{q=2013-1, q\neq 2015-3}^{2019-4} \beta_q \Delta M_{d,2018} + u_{dt} \tag{A.1}$$

Native Wages

First, note that the red dashed line in Figure A.2 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 immigration rate build from the GEIH survey. With this in mind, β_q measures the native wage effect across the formal and informal sectors. Interestingly, after the grey line (re-opening of the border), the impact is more pronounced for informal workers, while for formal workers, the effect is insignificant. The relation between the effect on informal wages and the immigration rate seems to be non-monotonic, as increases in the thick black line (immigration rate) do not necessarily translate into more negative impacts on informal wages.

Figure A.2: Event study estimates on log hourly wages by quarters and affiliation to social security



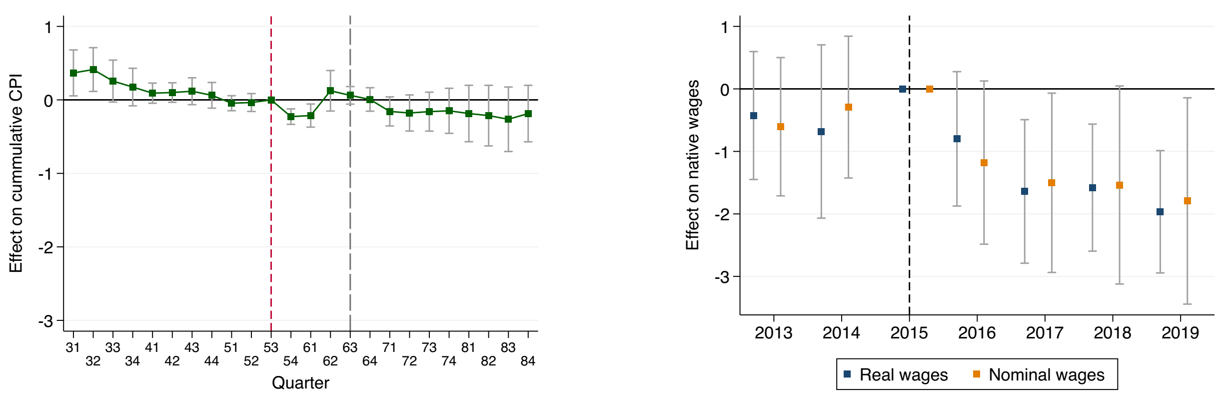
Note: The black line is the quarterly immigration rate from the GEIH survey defined as before. I use a 95% confidence interval. The sample is restricted to permanent Colombian residents between 18 and 64 years in urban areas. Department survey weights from GEIH are used to construct aggregate outcomes. 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

Figure A.3: Event study estimates on CPI and wages

(a) CPI

(b) Real versus nominal



Note: Capital cities in the regression are $N = 23$ per year. I use a 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. The base year is 2008 ($index = 100$)

D. Event Study with a Time-Varying Immigration Rate

As a robustness check, I also build an immigration rate that varies in post-treatment years from the GEIH survey. A potential caveat of this approach is that relative to 2015, migration arrivals before 2017 are small (as shown in Figure 3b) and therefore might not explain the observed changes in wages or employment, yielding noisy estimates for those years. For this and other reasons mentioned above, $\Delta M_{d,2018}$ is the primary explanatory variable in the analysis.

Given this discussion, equation (1) can be rewritten into a first-difference model, where I interchange $\Delta M_{d,2018}$ from the census with ΔM_{dt} from the GEIH survey. Hence, for each year $t = \{2013, .., 2019\}$ I estimate separate regressions of the following form:

$$Y_{dt} - Y_{d,2015} = \delta_t + \theta_t \Delta M_{dt} + u_{dt} - u_{d,2015}. \tag{A.2}$$

Note that when ΔM_{dt} is replaced by $\Delta M_{d,2018}$, the coefficient of interest in equations (1) and (A.2) is identical ($\beta_t = \theta_t$).[A.2] The definition of the time-varying immigration rate is as follows:

$$\Delta M_{dt} = \frac{L_{Ven,d,t} - L_{Ven,d,2015}}{L_{Total,d,2015}}, \tag{A.3}$$

where the numerator measures the employed Venezuelans (between 18 and 64 years) in d for all the post-years t from the GEIH survey, relative to the base period (2015). The denominator in turn is fixed on the base year, following Card and Peri (2016).

I use this setting to show the possible endogeneity of the immigration rate. Suppose the outcome of interest is $ln(wages) = ln(w_{dt})$ such that the estimated coefficient of interest becomes

$$\hat{\theta}_t = \frac{\hat{Cov}(\Delta M_{dt}, \Delta ln(w_{dt}))}{\hat{Var}(\Delta M_{dt})}, \tag{A.4}$$

where $ln(w_{dt}) - ln(w_{d,2015}) = \Delta ln(w_{dt})$.[A.3] Next, plugging model (A.2) in the last expression yields

$$plim(\hat{\theta}_t) = \theta_t + \frac{Cov(\Delta M_{dt}, \Delta u_{dt})}{Var(\Delta M_{dt})}. \tag{A.5}$$

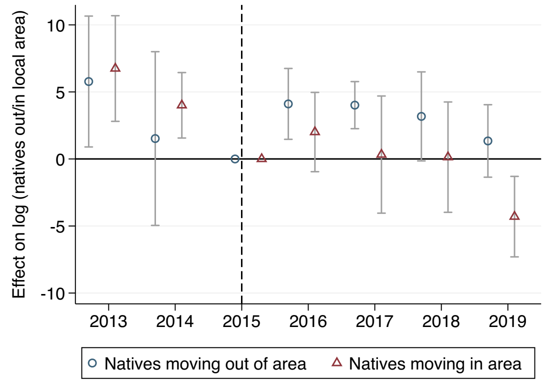
Thus, even if the time-invariant heterogeneity γ_d is removed, a bias can still emerge if migration is driven by unobservables in the departments (i.e., $E[\Delta M_{dt} \Delta u_{dt}] \neq 0$) that change over time. For instance, it could be the case that economic conditions relative to the base year are correlated with immigration inflows.

[A.2] Although the coefficient in equation (1) and (A.2) is the same when using a time-invariant treatment variable, the standard errors in (1) are going to be smaller or equal. This is because regression (1) groups observations of units over time, allowing to control for arbitrary serial correlation of errors, while regression (A.2) gives first difference errors with respect to a specific time interval for every unit. With this in mind, (1) uses clustered standard errors, while for (A.2), is only possible to use robust standard errors. In any case, the difference in the size of standard errors is negligible.

[A.3] 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.

E. Robustness Checks

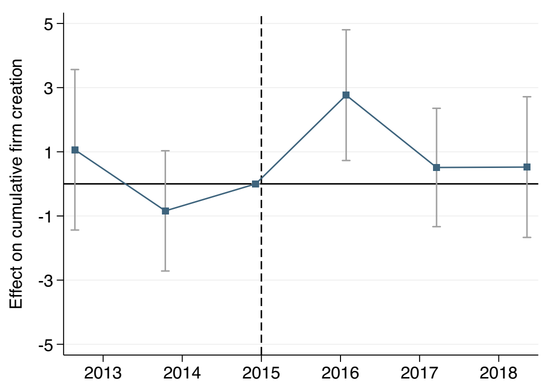
Figure A.4: Event study estimates on movements across geographical areas



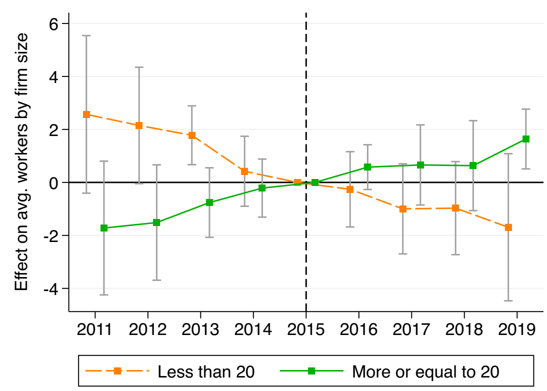
Note: I use a 95% confidence interval. Measures of geographical movements come from GEIH migration module. I use department survey weights from GEIH to construct aggregate outcomes. The base period is 2015. β_t from equation (1) are the plotted coefficients with explanatory variable $\Delta M_{d,2018}$.

Figure A.5: **Event study estimates on cumulative firm creation and average category of workers by firm size**

(a) IV (distance) on firm creation



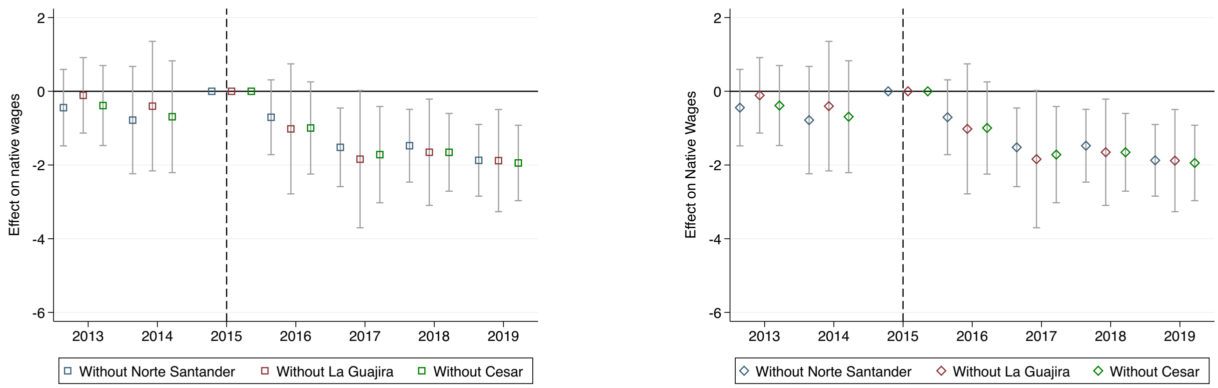
(b) IV (distance) on avg. category of workers



Note: The sample is restricted to registered, or formal, new firms that are in the databases of the corresponding state agencies. The outcome for panel (a) is the logarithm of new formal firms, and the outcome for panel (b) is the proxy average change in category of firm size. The explanatory variable is $\Delta M_{d,2018}$. I use a 95% confidence interval. The base period is 2015. β_t from equation (1) are the plotted coefficients. The F-statistic for distance instrument is 157.9. Source: (a) Confecamaras, 2013–2018. (b) GEIH, 2013–2019.

Figure A.6: Event study estimates excluding border departments for native wages

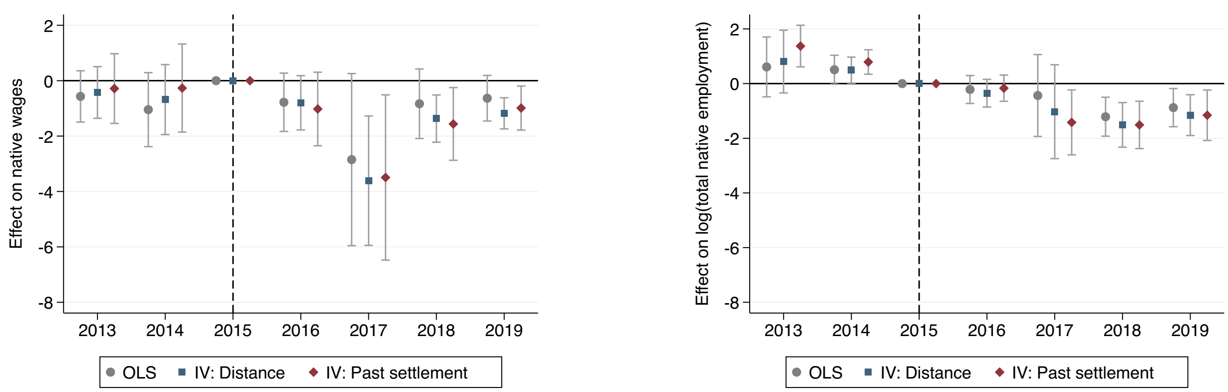
(a) IV (distance) estimates (b) IV (past settlement) estimates



Note: Departments in the regression are $N = 23$. I use a 95% confidence interval. I restrict the sample to permament Colombian 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.

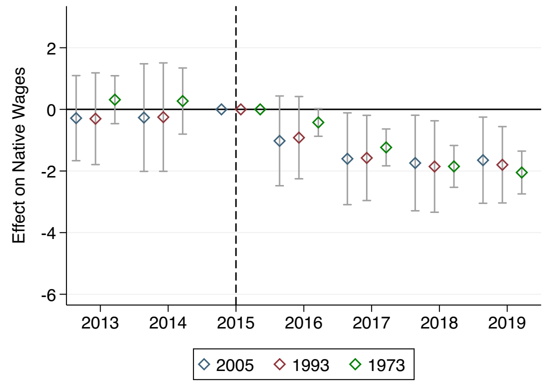
Figure A.7: Event study estimates using ΔM_{dt} from GEIH survey as explanatory variable

(a) IV native wages (b) IV native employment



Note: The explanatory variable is $\Delta M_{d,2018}$ before 2017, and after is ΔM_{dt}. I use a 95% confidence interval. I restrict the sample to permament Colombian residents between 18 and 64 years in urban areas. I use department survey weights from GEIH to construct aggregate outcomes. The base period is 2015. θ_t from equation (A.2) are the plotted coefficients. Hourly wages are in real terms using the monthly CPI from DANE.

Figure A.8: Event study estimates using different historical shares for the past settlement instrument



Note: The departments in the regression are $N = 24$ for 2005 and 1993, and $N = 22$ for 1973. I restrict the sample to permament Colombian residents between 18 and 64 years in urban areas. I use a 95% confidence interval. I use department survey weights from GEIH 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 35.5, 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.

Table A.3: **Robustness checks: wages and employment estimates for natives, 2015–2018**

	(1) Wages	(2) Employment	(3) Wages	(4) Employment
Instrument	**Distance**		**Past settlement**	
Main results⋆	-1.579**	-1.466***	-1.741*	-1.125*
	(0.446)	(0.358)	(0.682)	(0.459)
Control trade with Venezuela 2015	-1.917*	-1.533**		
	(0.684)	(0.413)		
Control change GDP 2015–2018	-1.521**	-1.460***		
	(0.492)	(0.367)		
Control quintiles of distance to Venezuela	-1.407	-1.120*	-1.825	-0.643
	(0.715)	(0.480)	(1.037)	(0.707)
Residualized wages	-1.925**		-1.548*	
	(0.516)		(0.556))	
Excluding natives that moved in $t-1$	-1.694**	-1.630***	-1.838*	-1.573***
	(0.465)	(0.368)	(0.697)	(0.374)
N	24	24	24	24
Trend-adjusted	No	Yes	No	Yes

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

Note: This table reports the coefficients of the second-stage regression of the instruments with the immigration rate $\Delta M_{d,2018}$. The coefficient measures the effect in 2018 relative to 2015. I restrict the sample to permanent Colombian residents between 18 and 64 years in urban areas. The standard errors are clustered at the department level. The variables are in logarithms thus the coefficients are interpreted as percentages. I use department survey weights from GEIH to construct aggregate outcomes. Trend-adjusted 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. I construct control of distance 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 the 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 the pre-trend.

F. Testing Caruso, Canon, and Mueller (2021) Wage Estimates

In here, the wage estimates of Caruso, Canon, and Mueller (2021) are replicated to test how sensitive they are to changes in periods or in the empirical specification.[A.4] First, in Table A.4 I replicate the central point estimate they found on hourly wages for a one pp increase in their immigration rate, I find a –7.4% compared to the original –7.6%.[A.5] Next, I estimate the same regression but add one more year of data. After adding 2018, the point estimate is less negative (–4.6%), a three pp difference in the coefficient relative to the one until 2017. The first conclusion is that as migration increases in Colombia, one additional year of data drastically reduces the impact of immigration on wages. Next, I use the DiD specification with the immigration rate from GEIH until 2017 to show the difference in the coefficient using the same time periods as Caruso, Canon, and Mueller (2021). In this case, the coefficient is drastically less negative (–3.6%), close to a four pp difference. Last, when using my specification until 2018, I find a coefficient of –1.4%. All in all,

[A.4] For details on their specification, I encourage the reader to check their paper and Lebow (2021b) paper that summarizes different wage effects found for Colombia.

[A.5] I do not match their coefficient completely since I have a slightly different distance instrument and the sample used is not the same.

the main difference in wage estimates from Caruso, Canon, and Mueller (2021) and my paper is due to the amount of periods analyzed and the specification used.

Table A.4: Replication of Caruso, Canon, and Mueller (2021) wage results

	(1)	(2)	(3)	(4)
	Until 2017	Until 2018	Until 2017	Until 2018
Panel regression with IV	-7.402***	-4.620***		
	(1.828)	(1.136)		
DiD with IV			-3.610**	-1.367**
			(1.126)	(0.411)
N	1,443,621	1,720,798	24	24

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

Note: This table reports the coefficients of the second-stage regression of the distance instrument with the immigration rate from the GEIH survey. The sample is restricted to workers between 15 and 64 years. In the panel regression, I use the same set of controls and construct an immigration rate as Caruso, Canon, and Mueller (2021) explain on their paper. Source: GEIH 2013–2018.

G. General Equilibrium Model

To complement the partial equilibrium model in the theoretical framework, I introduce a general equilibrium model inspired by Dustmann, Schönberg, and Stuhler (2017). The main difference is that I use two different labor inputs instead of two different skill groups. As in the theoretical framework, I model a firm that can hire workers L_f paying the official payroll taxes and also can hire workers L_i off the books to avoid complying with the contributions to the social security, but now firms also use capital.

The output for the firm is given by a cobb-douglas production function under constant returns to scale,

$$y(A, K, L) = AK^{\alpha}L^{(1-\alpha)} \tag{A.6}$$

Specifically, the labor function has a CES form

$$L = (\alpha_i L_i^{\rho} + \alpha_f L_f^{\rho})^{\frac{1}{\rho}} \tag{A.7}$$

where $\sigma = \frac{1}{1-p}$ is the elasticity of substitution between formal and informal workers, so $\rho \leq 1$.

Hence, the profit function of the firm is equal to:

$$\max_{K, L_i, L_f} \pi = pAK^{\alpha}L^{(1-\alpha)} - L_i^{\eta} w_i L_i - (1+\tau_f) w_f L_f - rK \tag{A.8}$$

where L_i^{η} can be stated as a cost that is increasing on informal labor size, with $\eta \geq 1$, whereas τ_f represents the payroll taxes that the firm has to enroll when paying for formal workers. The price p is normalized to be equal to 1 and therefore it is fixed.

If wages are fully flexible (i.e., there are no minimum wages), the price of labor is given by its marginal product. After taking logs, wages in the formal sector are equal to:

$$
\begin{aligned}
logw_f = log[(1-\alpha)A] + \alpha[logK - logL] + \\
log\alpha_f + (\rho - 1)[logL_f - logL] - log(1+\tau_f)
\end{aligned} \tag{A.9}
$$

Then the corresponding (logged) wages in the informal sector are given by:

$$
\begin{aligned}
logw_i = log[(1-\alpha)A] + \alpha[logK - logL] + \\
log\alpha_i + (\rho - 1)[logL_i - logL] - \eta logL_i - log(1+\eta)
\end{aligned} \tag{A.10}
$$

Notice that the unit cost of capital is given by the marginal product for both firms:

$$
logr = log\alpha A + (\alpha - 1)(logK - logL) \tag{A.11}
$$

where it is assumed that the aggregate capital stock K depends on prices (r) and λ is the inverse elasticity of capital supply with respect to its price.

Competitive Equilibrium

The two labor demand equations (depending on labor input) should equate the labor supply in equilibrium. To derive equilibrium wage and employment responses, there are three possible cases that can occur, which have different implications and help to explain the empirical findings.

1. The first case is assuming natives labor supply (ϵ^u) is perfectly inelastic and no labor cost (i.e., $\epsilon^u = 0$ and $\eta = 0$). After some algebraic derivations in the subsection A I find that the relative wage effect is given by

$$
\partial logw_f - \partial logw_i = -\frac{1}{\sigma}\left[\frac{E_f^I}{E_f^N} - \frac{E_i^I}{E_i^N}\right]\partial I \tag{A.12}
$$

   The immigrant-induced supply shock relative to native equilibrium employment is given by $\partial I = \frac{\partial L^I}{L^N}$. This result yields the same conclusion as in Borjas (2013) heterogeneous labor model, where the impact of immigration on relative wages only depends on the elasticity of substitution σ. If formal and informal workers are perfect substitutes, immigration has no relative wage effect. If they are imperfect substitutes, the group that experiences the larger supply shock (measured as $\frac{E_g^I}{E_g^N}$ for $g = i, f$) will experience the largest relative wage loss. However, the assumptions in this case imply that the employment of natives will be the same after the immigration event, which in this setting is not occurring.

2. The second case is assuming natives labor supply is elastic but is the same for formal and informal workers, and there are no informal labor costs (i.e., $\epsilon^u = \epsilon_f^u = \epsilon_i^u$ and $\eta = 0$). Thus, the immigration event will also trigger a change in native employment, not only on relative wages as in the first case. After some algebraic derivations in the subsection A, I find that the change of native employment in each sector to an immigrant induced supply shift is given by:

$$\partial logL_f^N = \frac{\epsilon^u(\rho-1)\left[\frac{E_f^I}{E_f^N}(1-\upsilon\epsilon^u)-\phi(1-\frac{\upsilon}{\rho-1})\right]}{1-\epsilon^u(\rho-1)\left[1-\upsilon(\sigma+\epsilon^u)\right]}\partial I \tag{A.13}$$

$$\partial logL_i^N = \frac{\epsilon^u(\rho-1)\left[\frac{E_i^I}{E_i^N}(1-\upsilon\epsilon^u)-\phi(1-\frac{\upsilon}{\rho-1})\right]}{1-\epsilon^u(\rho-1)\left[1-\upsilon(\sigma+\epsilon^u)\right]}\partial I \tag{A.14}$$

First, the denominator is always positive as the aggregate slope of labor demand υ is negative, $\rho \leq 1$ and $\epsilon^u > 0$. In this setup, the native employment response will depend again on the relative density of immigrants to natives in each sector $g = i, f$, but also on the weighted average of immigration across the two sectors $\phi = s_i \frac{E_i^I}{E_i^N} + s_f \frac{E_f^I}{E_f^N}$, where s_g is the labor share of each sector in production. In this case, because migrants are mostly employed on the informal sector ($\frac{E_i^I}{E_i^N} > \phi$) wages and employment in the informal sector will decline more than wages and employment in the formal sector (if $\rho < 1$), which is not observed in the empirical findings, where native formal employment drops more than native informal employment. Finally, the relative effect on wages is given by,

$$\partial logw_f - \partial logw_i = -\frac{1}{\sigma}\left[\frac{E_f^I}{E_f^N} - \frac{E_i^I}{E_i^N}\right]\Psi\partial I \tag{A.15}$$

when the elasticity of labor supply is the same for the two workers, I find a similar relative wage effect as in the first case. Just that the effect is now mediated by the elasticity of labor supply and the elasticity of capital supply given by expression $\Psi = \frac{1-\upsilon\epsilon^u}{1-\epsilon^u(\rho-1)\left[1-\upsilon(\sigma+\epsilon^u)\right]}$.

3. The third case is assuming natives labor supply is different for formal and informal workers and informal labor cost is positive (i.e., $\epsilon_f^u \neq \epsilon_i^u$ and $\eta \neq 0$). First, the labor supply equations for each group are the following:

$$\partial logw_f = (1/\epsilon_f^u)\partial logL_f^N \tag{A.16}$$

$$\partial logw_i = (1/\epsilon_i^u)\partial logL_i^N \tag{A.17}$$

It is critical to establish that wages in the formal sector $\partial log\bar{w}_f$ are downwardly rigid by the minimum wage, thus in this case, the formal wage effect can be muted while having changes in native formal employment.

After some derivations in the subsection A, I find that native employment change in each sector to an immigrant induced supply shift is given by:

$$\partial logL_f^N = \frac{\epsilon_f^u(\rho-1)\left[\frac{E_f^I}{E_f^N}(1-\upsilon\epsilon_i^u)-\phi(1-\frac{\upsilon}{\rho-1})+\mu_\eta\right]}{1-(\rho-1)\left[\epsilon_f^u(1+s_f\delta)+\epsilon_i^u(1+s_i\delta)-\epsilon_f^u\epsilon_i^u\upsilon\right]+\omega_\eta}\partial I \tag{A.18}$$

$$\partial log L_i^N = \frac{\epsilon_i^u(\rho-1)\Big[\frac{E_i^I}{E_i^N}(1-\upsilon\epsilon_f^u)-\phi(1-\frac{\upsilon}{\rho-1})\Big]}{1-(\rho-1)\big[\epsilon_i^u(1+s_i\delta)+\epsilon_f^u(1+s_f\delta)-\epsilon_i^u\epsilon_f^u\upsilon\big]+\omega_\eta}\partial I \tag{A.19}$$

Note that $\delta = \frac{\upsilon}{\rho-1} - 1$ and $\omega_\eta = \eta\epsilon_i^u(1-\upsilon s_f\epsilon_f^u-(\rho-1)s_i\epsilon_f^u)$ are always positive, and thus the denominator is always positive. In this case, the native employment response is mediated by its own labor supply elasticity and the informal labor cost in expression $\mu_\eta = \eta\epsilon_i^u\big(\frac{E_f^I}{E_f^N}-\phi(1-\frac{\upsilon}{\rho-1})\big)$, which gives a different employment response compared to previous cases.

Finally, the relative effect on wages is given by:

$$\partial log w_f - \partial log w_i = \frac{-\frac{1}{\sigma}\Big[\frac{E_f^I}{E_f^N}(1-\upsilon\epsilon_i^u)-\frac{E_i^I}{E_i^N}(1-\upsilon\epsilon_f^u)+\mu_\eta\Big]}{1-(\rho-1)\big[\epsilon_f^u(1+s_f\delta)+\epsilon_i^u(1+s_i\delta)-\epsilon_f^u\epsilon_i^u\upsilon\big]+\omega_\eta}\partial I \tag{A.20}$$

This is the same expression as in Dustmann, Schönberg, and Stuhler (2017) with two skills groups, instead of sectors of employment, and when $\eta = 0$. Therefore, the relative wage effect will depend not only on the group that experiences the largest supply shock but also on the own labor supply elasticity and the size of the informal labor cost. So, if $\epsilon_f^u > \epsilon_i^u$ and $\frac{E_i^I}{E_i^N} > \frac{E_f^I}{E_f^N}$ formal wages relative to informal wages are going to increase. Regarding employment responses, the sign will depend on the aggregate slope of labor demand υ (which is a function of the elasticity of capital supply λ) relative to the substitution parameter. If $\upsilon < \rho - 1$, both native employment responses are negative. This decrease will depend on labor supply elasticities and informal labor costs.[A.6] In conclusion, this model states that to find a negative formal employment response, formal and informal workers need to have a high degree of substitutability accompanied by a proper capital response given in the expression υ. Besides, heterogeneous labor supply elasticities need to exist for formal and informal workers.

H. Algebraic Derivations

Derivations of the Theoretical Framework 6.3

To derive equations (10) and (11) from the main text, I combine first the profit function with the price function. So, the maximization problem turns to be:

$$\max_{L_i,L_f} \pi = C^{1-\epsilon}Q^\epsilon - \tau(L_i)w_iL_i - (1+\tau_f)w_fL_f \tag{A.21}$$

Then, rearranging market wages (from equation (8) and (9)) to leave in terms of labor we get that:

$$L_i^{1-\rho}(\tau'(L_i)L_i+\tau(L_i)) = \left(\frac{C^{1-\epsilon}\epsilon\alpha_i}{w_i}\right)(\alpha_iL_i^\rho+\alpha_fL_f^\rho)^{\frac{\epsilon-\rho}{\rho}} \tag{A.22}$$

[A.6] To find that, multiply the $(\rho-1)$ in the numerator of (A.18) inside the expression and rearrange to have as $\epsilon_f^u((\rho-1)\frac{E_f^I}{E_f^N}(1-\upsilon\epsilon_i^u)+\phi(\upsilon-(\rho-1))+\eta\epsilon_i^u((\rho-1)\frac{E_f^I}{E_f^N}+\phi(\upsilon-(\rho-1))))$, and this expression is negative when $\upsilon < (\rho-1)$. It is the same procedure for equation (A.19).

$$L_f^{1-\rho} = \left(\frac{C^{1-\epsilon}\epsilon\alpha_f}{w_f(1+\tau_f)}\right)(\alpha_i L_i^\rho + \alpha_f L_f^\rho)^{\frac{\epsilon-\rho}{\rho}} \tag{A.23}$$

To have a tractable solution, I specify the informal labor cost as $\tau(L_i) = L_i^\eta$ where $\eta = 0, 1, ..., N$. Then, taking logs of last expressions:

$$logL_i = \frac{1}{1+\eta-\rho}((1-\epsilon)logC + log\epsilon\alpha_i - logw_i - log(1+\eta)) + \frac{\epsilon-\rho}{\rho(1+\eta-\rho)}log(\alpha_i L_i^\rho + \alpha_f L_f^\rho) \tag{A.24}$$

$$logL_f = \frac{1}{1-\rho}((1-\epsilon)logC + log\epsilon\alpha_f - logw_f - log(1+\tau_f)) + \frac{\epsilon-\rho}{\rho(1-\rho)}log(\alpha_i L_i^\rho + \alpha_f L_f^\rho) \tag{A.25}$$

I differentiate previous equations with respect to informal wages. 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:

$$\frac{dlogL_i}{dw_i}|_{dw_f=0} * w_i = -\frac{1}{w_i(1+\eta-\rho)}w_i + \frac{\epsilon-\rho}{\rho(1+\eta-\rho)}\left(\frac{\rho(\alpha_f L_f^{\rho-1}\frac{dL_f}{dw_i} + \alpha_i L_i^{\rho-1}\frac{dL_i}{dw_i})}{\alpha_f L_f^\rho + \alpha_i L_i^\rho}\right)w_i \tag{A.26}$$

Simplifying, I get that:

$$\frac{dlogL_i}{dw_i}|_{dw_f=0} * w_i = -\frac{1}{1+\eta-\rho} + \frac{\epsilon-\rho}{1+\eta-\rho}(\alpha_f(L_f/Q)^\rho\varepsilon_{L_f,w_i} + \alpha_i(L_i/Q)^\rho\varepsilon_{L_i,w_i}) \tag{A.27}$$

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:

$$\varepsilon_{L_i,w_i} = -\frac{1}{1+\eta-\rho} + \frac{\epsilon-\rho}{1+\eta-\rho}(s_f\epsilon_{L_f,w_i} + s_i\epsilon_{L_i,w_i}) \tag{A.28}$$

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:

$$\varepsilon_{L_f,w_i} = \frac{\epsilon-\rho}{1-\rho}(s_f\epsilon_{L_f,w_i} + s_i\epsilon_{L_i,w_i}) \tag{A.29}$$

Combining the two last expressions I find that:

$$\varepsilon_{L_i,w_i} = -\frac{1}{1+\eta-\rho} + \frac{1-\rho}{1+\eta-\rho}\varepsilon_{L_f,w_i} \tag{A.30}$$

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 (10) and (11) in the main text.

Derivations of General Equilibrium Model

Recall that the elasticity of capital supply with respect to r is given by:

$$\frac{1}{\lambda} = \frac{\partial logK}{\partial logr} \tag{A.31}$$

Then, taking total derivatives of A.11 I get that:

$$\partial logr = (\alpha - 1)(\partial logK - \partial logL) \tag{A.32}$$

After plugging A.31 in A.32 I get:

$$\partial logK = -\frac{\alpha - 1}{\lambda - \alpha + 1}\partial logL \tag{A.33}$$

Then, I totally differentiate A.9 and A.10 to get that:

$$\partial logw_f = \alpha(\partial logK - \partial logL) + (\rho - 1)(\partial logL_f - \partial logL) \tag{A.34}$$

$$\partial logw_i = \alpha(\partial logK - \partial logL) + (\rho - 1)(\partial logL_i - \partial logL) - \eta\partial logL_i \tag{A.35}$$

I replace A.33 in previous expressions to find that:

$$\partial logw_f = \upsilon\partial logL + (\rho - 1)(\partial logL_f - \partial logL) \tag{A.36}$$

$$\partial logw_i = \upsilon\partial logL + (\rho - 1)(\partial logL_i - \partial logL) - n\partial logL_i \tag{A.37}$$

In this case, $\upsilon = -\frac{\lambda\alpha}{\lambda-\alpha+1}$ is the aggregate slope of labor demand. Assuming that immigrants and natives are perfect substitutes $L_g = L_g^I + L_g^N$, where $g = i, f$, and using the fact that there are no immigrants at baseline, total employment can be expressed as:

$$\partial logL_g = \frac{\partial L_g^I}{L} + \partial logL_g^N \tag{A.38}$$

Denoting employment shares of natives as $E_g^N = \frac{L_g^N}{L_f^N + L_i^N}$ and of immigrants as $E_g^I = \frac{L_g^I}{L_f^I + L_i^I}$ where $g = i, f$.

Then, I can express A.38 as:

$$\partial logL_g = \frac{E_g^I}{E_g^N}\frac{\partial L^I}{L^N} + \partial logL_g^N \tag{A.39}$$

Totally differentiating the aggregate CES function L I find that:

$$\partial logL = s_i\partial logL_i + s_f\partial logL_f \tag{A.40}$$

The shares of labor g to total labor aggregate are given by $s_g = \frac{\alpha_g L_g^\rho}{\alpha_i L_i^\rho + \alpha_f L_f^\rho}$.

Substituting expression A.39 into previous result I get that,

$$\partial logL = \phi\partial I + s_i\partial logL_i^N + s_f\partial logL_f^N \tag{A.41}$$

where $\phi = s_i\frac{E_i^I}{E_i^N} + s_f\frac{E_f^I}{E_f^N}$ is the weighted average of the relative density of immigrants across labor groups and $\partial I = \frac{\partial L^I}{L^N}$.

Derivations for each of the three cases. First, in the case the labor supply of natives is perfectly inelastic and informal labor cost is zero (i.e., $\epsilon^u = 0$ and $\eta = 0$). The native employment channel is muted and the relative wage effect is given by subtracting A.36 with A.37 after replacing inside with expression A.39.

For the second and third cases in the text, I will start with the two labor supply equations for every group A.16 and A.17.

Therefore, I plug these two labor supply equations in the two labor demand equations A.36 and A.37, I also use the labor expression A.41 and group-level expression A.39 to get that:

$$\begin{aligned}\partial logw_f = \upsilon(\phi\partial I + s_i\epsilon_i^u\partial logw_i + s_f\epsilon_f^u\partial logw_f) + (\rho-1)(\epsilon_f^u\partial logw_f \\ -(s_i\epsilon_i^u\partial logw_i + s_f\epsilon_f^u\partial logw_f) + (\frac{E_f^I}{E_f^N} - \phi)\partial I)\end{aligned} \tag{A.42}$$

$$\begin{aligned}\partial logw_i = \upsilon(\phi\partial I + s_i\epsilon_i^u\partial logw_i + s_f\epsilon_f^u\partial logw_f) + (\rho-1)(\epsilon_i^u\partial logw_i \\ -(s_i\epsilon_i^u\partial logw_i + s_f\epsilon_f^u\partial logw_f) + (\phi - \frac{E_i^I}{E_i^N})\partial I) - \epsilon_i^u\eta\partial logw_i\end{aligned} \tag{A.43}$$

Then, when solving for $\partial logw_f$ and $\partial logw_i$ I find that:

$$\partial logw_f = \frac{(\upsilon-(\rho-1))s_i\epsilon_i^u\partial logw_i + \upsilon\phi\partial I + (\rho-1)(\frac{E_f^I}{E_f^N} - \phi)\partial I}{1-\upsilon s_f\epsilon_f^u - (\rho-1)s_i\epsilon_f^u} \tag{A.44}$$

$$\partial logw_i = \frac{(\upsilon-(\rho-1))s_f\epsilon_f^u\partial logw_f + \upsilon\phi\partial I + (\rho-1)(\frac{E_i^I}{E_i^N} - \phi)\partial I}{1-\upsilon s_i\epsilon_i^u - (\rho-1)s_f\epsilon_i^u + \epsilon_i^u\eta} \tag{A.45}$$

The informal labor cost only affects the denominator of the second expression. Then, plugging the two last expressions into each other I get that,

$$\partial logw_f = \frac{\chi s_f\epsilon_f^u(\upsilon\phi + (\rho-1)(\frac{E_i^I}{E_i^N} - \phi)) + \kappa(\upsilon\phi + (\rho-1)(\frac{E_f^I}{E_f^N} - \phi))}{\psi\kappa - \chi^2 s_f\epsilon_f^u s_i\epsilon_i^u}\partial I \tag{A.46}$$

$$\partial logw_i = \frac{\chi s_i\epsilon_i^u(\upsilon\phi + (\rho-1)(\frac{E_f^I}{E_f^N} - \phi)) + \psi(\upsilon\phi + (\rho-1)(\frac{E_i^I}{E_i^N} - \phi))}{\psi\kappa - \chi^2 s_f\epsilon_f^u s_i\epsilon_i^u}\partial I \tag{A.47}$$

where $\kappa = 1 - \upsilon s_i\epsilon_i^u - (\rho-1)s_f\epsilon_i^u + \epsilon_i^u\eta$, $\psi = 1 - \upsilon s_f\epsilon_f^u - (\rho-1)s_i\epsilon_f^u$ and $\chi = \upsilon - (\rho-1)$. To find equations A.13 and A.14 in the second case, I simplify previous expressions assuming that $\epsilon^u = \epsilon_f^u = \epsilon_i^u$ and $\eta = 0$. Finally, for the third case, I simplify previous expressions and plug inside the two labor supply equations to find equations A.18 and A.19.

J. Definition of Variables

Hourly real wages. The *inglabo* variable from the GEIH survey captures basic pay, pay in-kind, and income of second activity. First, I subtract income of second activity and add allowances for food and transportation (according to ILO definition of wages).[A.7] Is worth noting that this variable captures any type of labor income, as non-salaried workers are part of my sample. Next, the nominal labor income variable is transformed into a real one using monthly CPI at the national level.[A.8] The base of the index (=100) is in December of 2018,

$$RealWage_{imy} = \frac{NominalWage_{imy}}{CPI_{my}} * 100 \tag{A.48}$$

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

Employed Colombians. 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 the GEIH survey are counted as employed. Then, I count (with department weights) all individuals in each department and year, and then I take logarithms of that expression.

Employed definition according to census. The census does not have all the standard labor force survey questions regarding occupation. It only has one question that asks about work in the last week. If the respondent state to work for a compensated income for at least 1 hour, I count them as occupied.

[A.7] https://www.ilo.org/wcmsp5/groups/public/---africa/---ro-abidjan/---ilo-pretoria/documents/publication/wcms_413782.pdf

[A.8] Information was taken from here https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc