Immigration, Wages, and Employment under Informal Labor Markets

Lukas Delgado-Prieto*

March, 2024

Abstract

This paper studies the labor market impacts of the Venezuelan immigration 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, I build a model in which a firm substitutes formal for informal labor in response to lower informal wages. Consistent with the model's predictions, I document that the decrease in formal employment is driven by small firms that use both labor types in production and by workers earning the minimum wage.

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

JEL Codes: F22, O15, O17, R23.

*Department of Economics, Universidad Carlos III de Madrid, Spain. Email: ludelgad@eco.uc3m.es. ORCID: 0000-0002-9210-854X.

1 Introduction

Despite many studies over the last three decades, the impact of immigration on native wages and native employment remains one of the most relevant, albeit disputed, issues in empirical labor economics (see Dustmann, Schönberg, and Stuhler (2016) for a summary of findings). In the latest years, Latin American countries have witnessed a substantial surge in the number of migrants due to the Venezuelan crisis. As of November 2023, the region has hosted over 6.5 million people that left Venezuela (UNHCR, 2023). These massive and sudden inflows can influence various socioeconomic outcomes in the host countries, both in the short and long run.

This paper studies the labor market impacts of the Venezuelan immigration in Colombia, where the supply of migrants surged from 0.2% in 2015 to 4.1% in 2019 (this means around 1 million more working-age Venezuelans).[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 in the short run. This effect, however, can be different in settings with a large informal sector and binding minimum wages in the formal sector, as informal labor income adjusts more flexibly while formal wages are more rigid. Thus, the impacts on formal and informal employment can be different (Kleemans and Magruder, 2018; Corbi, Ferraz, and Narita, 2021).[2]

This paper's main contribution is to empirically show that immigration affects the wages and employment of natives in Colombia and that these impacts are driven by the informal and formal sectors, respectively. Moreover, this paper develops a model that explains why these sectors experience an asymmetric response. For the empirical part, I take advantage of varying treatment intensity across areas. In Colombia, certain areas received vast inflows of migrants, while others barely received any. This is the primary source of variation exploited in the empirical strategy.

The main assumption needed for this research design is the parallel trends assumption (PTA). In practice, this assumption may not be satisfied as migrants can endogenously sort into the areas that offer the best economic conditions, which can lead to differential trends in the outcomes even without any impact of immigration. I use an event study research design with an instrumental variable (IV) to deal with the potential sorting and other endogeneity issues. 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 Grossman (1982).

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 migrants from Venezuela reduces native hourly wages. A one percentage point (pp) increase in the share of employed migrants from Venezuela over the employed population reduces local native wages by around 1.5% and 1.7%.[4] Importantly, these estimates are robust to the two chosen instruments, alternative definitions of local labor markets, and different wage datasets. Compared to other studies analyzing this immigration shock in Colombia, my wage estimate lies in between the sizeable negative coefficient of Caruso, Canon, and Mueller (2021) and the more positive or insignificant ones of Morales-Zurita et al. (2020) and Lebow (2024).[5] Compared to other migration episodes, I find a wage estimate that is more negative than several recent papers (Dustmann, Schönberg, and Stuhler, 2017; Aksu, Erzan, and Kırdar, 2022; Monras, 2020) except for the wage estimate of Edo (2020), which is similar in magnitude.[6] I then complement the effect on average wages with the distributional impacts of immigration, where I find that wage losses are concentrated at the bottom 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 migrants from Venezuela reduces local employment by between 0.9% and 1.3%. On the intensive margin, natives work more hours in the informal sector in response to the immigration shock.

I then argue that a critical 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 (MW), allows wages to adjust flexibly to a labor supply shock. This is relevant as almost all Venezuelans are employed by the informal sector. Specifically, for natives, I find a

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

[4]Throughout this paper, wages can refer to labor income as they contain both the labor income of self-employed and wages of employees.

[5]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 G, I show that the difference between the wage estimates of this paper and that one is driven by the sample period they analyze (until 2017) and the empirical specification they choose (panel data regression). This motivates, in part, the analysis performed in this paper.

[6]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%.

negative effect on informal salaried employment and a considerable decrease in informal wages; they decrease by 1.9% for a one pp increase in the immigration shock. This is explained by the industry of commerce, hotels, and restaurants, where most migrants end up working.

For formal employment, I find a reduction of 1.9% for a one pp increase in the immigration shock that is concentrated in native workers earning the MW, while for formal wages, I find an insignificant effect around zero. The insignificant impact is determined partly by the downward rigidity imposed by the MW, which is binding for many formal workers in Colombia (in 2015, 40% of formal workers earned around the MW). Then, I exploit that small firms are more constrained by the MW to document that the most negative drop in formal employment happens among workers in the smallest firms.

From a policy perspective, a relatively high MW can magnify the adverse employment effects in the formal sector. According to OECD (2023), the ratio of the minimum wage to the median in Colombia is the highest of all OECD countries, being 82.6% in 2015. Relative to other countries in the same year, the relevance of the minimum wage is lower (e.g., for Turkey, it is 69.9%; for Mexico, it is 38.7%; and for the US, it is 35.8%). In that sense, the MW plays a more critical role in the Colombian setting, but still, in the Turkish case it is high enough to explain the asymmetric results of immigration in the formal and informal sectors (Gulek, 2022). More generally, the combination of how large the informal sector is and how high the minimum wage is can determine the response to immigration shocks in other contexts.

To rationalize the findings for formal employment, I first show that firms usually combine formal and informal labor in their production function. Importantly, this combination mainly depends on firm size: when formal firms are smaller, the share of informal workers increases on average. Thus, I build a model with a representative firm that hires two types of inputs, formal and informal labor, with different costs as in Ulyssea (2018) but allowing them to be imperfect 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 increases with the size of informal labor within the firm (the cost of evasion). In this framework, I derive the elasticity of formal labor demand when there is a change in informal wages. This elasticity's sign depends on the elasticity of substitution between informal and formal workers. If they are strong substitutes, the model predicts that the firm will respond to lower informal wages by substituting formal for informal workers, decreasing formal employment. 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 the general functioning of informal labor markets. In that sense, the *labor-labor* substitution pattern to the immigration shock can be a margin of adjustment, for instance, to minimum wage increases too, as stated in Clemens (2021).

Regarding heterogeneous effects, formal employment decreases the most for low-skilled natives (who are more substitutable with informal workers), and informal employment declines the most for salaried workers. In contrast, informal labor income decreases in most subgroups (except high-skilled and older natives). Then, I find insignificant estimates by studying the price response on a bundle of goods and services to complement the impact on wages and employment. These findings indicate a more substantial supply effect from lower wages rather than a higher demand from migrant consumption of goods and services.

Literature. This paper contributes to different strands of the literature. First, to the literature on the labor market effects of immigration (Dustmann, Schönberg, and Stuhler, 2017; Edo, 2020; 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 helps to alleviate concerns from previous immigration adjustments conflating short-run impacts (Jaeger, Ruist, and Stuhler, 2018). Not many immigration events have these characteristics; 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 most evidence 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 influx of Syrian refugees in Turkey (Del Carpio and Wagner, 2015; Ceritoglu et al., 2017; Aksu, Erzan, and Kırdar, 2022), a case study with a migration shock similar to the Colombian one. For instance, Aksu, Erzan, and Kırdar (2022) finds 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 and are less educated than natives, and Turkey did not implement an open border policy with migrants like Colombia. My contribution is to build a theoretical model linking the labor market's informal and formal sectors. 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 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) in Indonesia and Corbi, Ferraz, and Narita (2021) in Brazil find that internal migration, where the binding MW is high, creates asymmetric responses in the informal and formal sectors. They find that the most negative wage effects are in the informal sector, while the most negative employment effects are in the formal sector. These results are similar to the ones in this paper, but I study international (not internal) migration. Altogether, it is suggestive of how binding minimum wages, interacted with the size of the informal sector, determines the wage and employment response for natives in the presence of an immigration shock. A consistent argument with Calderón-Mejía and Ibáñez (2016), that studies the impact of forced internal displacement in Colombia and emphasizes that the high MW in the formal sector impedes a smooth adjustment to labor supply shocks.

Last, this paper makes several contributions to previous studies that estimate the impact of the Venezuelan immigration on Colombia's labor market (Penaloza-Pacheco, 2022; Morales-Zurita et al., 2020; Santamaria, 2020; Caruso, Canon, and Mueller, 2021; Lebow, 2024; Rozo and Vargas, 2021). To start, I use an event study design with continuous treatment while using two different instruments that test for the presence of preexisting trends.[7] The use of this research design

[7] The event study design is motivated by the fact that the static coefficient in most previous studies comes from a panel IV regression that may be interpreted as a weighted average of treatment effects. The issue is that these weights may even be negative, for instance, because the timing of treatment varies 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 of the lack of correlation between migration rates in 1973 and 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 (2024) 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.

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.[8] In addition, the dynamic specification of the event study has the advantage of being precise about the timing of the 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.[9] Finally, I use a population census that measures Venezuelans and returning Colombians from Venezuela with the greatest detail, reducing the measurement error in standard surveys and consequently any attenuation bias on the estimates (Aydemir and Borjas, 2011).

The rest of the paper is structured as follows. Section 2 briefly overviews the Venezuelan crisis and its institutional background. Section 3 describes the data with 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 provides robustness checks, and Section 8 concludes.

2 Venezuelan Crisis and Institutional Context

2.1 Overview of the Venezuelan Crisis

Since Hugo Chávez assumed the presidency of Venezuela in 1998, the nation gradually moved towards a state-based economy, characterized by a diminished private sector and a dominant oil industry. In 2013, after more than 14 years of Chávez leadership, Nicolás Maduro succeeded him after his passing. With Maduro as president, the Venezuelan economy suffered a sizable negative shock in 2015, triggered by a nearly 50% drop in oil prices. This shock directly impacted the government's primary revenue source, reducing the funding for social programs and subsidies for essential products like medicines and food. The social discontent exploded in 2017 when Maduro's party won most state elections with apparent signs of fraud. Massive flows of Venezuelans started

[8] For instance, Morales-Zurita et al. (2020), Caruso, Canon, and Mueller (2021) and Lebow (2024) 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 Appendix Figure F.1), resulting in a source of bias in their estimates.

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

leaving the country in 2016 due to the worsening 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 found that 96.2% of all Venezuelans were poor and 79.3% were extremely poor (UCAB, 2020). In this context, the Venezuelan exodus began with voluntary and involuntary immigration. As of 2019, more than 1 million working-age Venezuelans were in Colombia, according to the labor force survey (GEIH, by its acronym in Spanish). Most of these Venezuelans plan to stay in Colombia for over 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). Therefore, the immigration shock I analyze combines the arrival of refugees with voluntary migrants.

2.2 Regulatory Framework for Venezuelans

Before 2018, undocumented Venezuelans needed a special visa to work in the formal sector. 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 undocumented Venezuelans. It provided a new framework for work called the Special Permit of Permanence (PEP, by its acronym in Spanish).[10] The PEP's goal was to foster legal and more accessible employment for Venezuelans without the need for sponsor companies or investments.[11] 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,

[10]In July 2018, the salient president of Colombia, Juan Manuel Santos, unexpectedly announced the creation of a special permit to work for all Venezuelans registered in the Administrative Record of Venezuelan Migrants (RAMV, by its acronym in Spanish).

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

2021).[12] Even if the short-run impacts of this policy seem insignificant, they are part of the overall effect of the Venezuelan immigration on natives' labor market outcomes. Still, with the dynamic specification, I show there were significant wage effects even before this regularization happened.

3 Data

I use two primary datasets 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 2018. GEIH is a monthly cross-sectional survey that characterizes the Colombian labor market. It covers approximately 240,000 households annually and is the survey with the most detailed sample coverage. Both datasets are administered by the National Statistics Office of Colombia (DANE, by its acronym in Spanish) and are available on their webpage.

To measure the Venezuelan immigration, DANE implemented a migration module in the GEIH survey with questions on where the person was born, where they lived 12 and 60 months ago, and the reasons for migrating since 2013. In this study, I use data from 2013–2019 to differentiate natives from migrants in urban areas for the main outcomes.[13] I exclude rural areas from the analysis as their labor markets behave differently, and they have very few immigrants.

I also exploit supplementary databases for the construction of the instruments and additional outcomes. The first is the RAMV survey, which characterized the population of undocumented Venezuelans in Colombia.[14] I take the information from which state in Venezuela immigrants are coming to build the distance instrument. 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 newly registered firms in Colombia with tax records, to construct the information of newly registered firms. Last, I use price data that is representative of 23 capital cities in Colombia.

[12]In Ecuador, Olivieri et al. (2021) find that providing work permits to Venezuelan workers would increase their average earnings.

[13]Moreover, I include 2011 and 2012 in the analysis to have additional pre-treatment periods, assuming all survey respondents were Colombian.

[14]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.

3.1 Descriptive Statistics for Natives and Migrants

For the descriptive statistics, I focus on three main groups of interest. The first is of native Colombians residing permanently in Colombia, the second is Venezuelans who emigrated to 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 main analysis, I focus only on Colombians who did not migrate from Venezuela in the previous year. This is because the sample size for the Venezuelan migrants and Colombian returnees, mainly when split by region-year cell, is very small. With this in mind, Appendix A presents a table with descriptive statistics regarding the age profile, level of education, and sex composition of the different groups according to the different years of arrival.

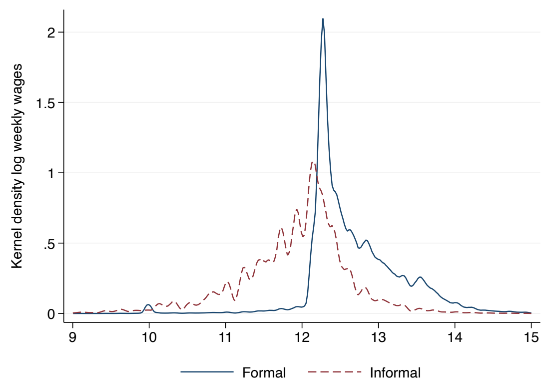
First, Venezuelan immigrants arriving in Colombia tend to be young, although their average age seems to grow. Before 2017, the highest share of arrivals was in the range of 0–14 years; after 2017, the range changed to 14–28 years. Second, returning Colombians are more concentrated in older ages; before 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, skills. Suggesting that migrants are substitutes, more than complements, of natives. Another relevant takeaway is that Venezuelan arrivals seem more educated in recent years. Fourth, regarding sex composition, the share of males and females is 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 Colombian Labor Market

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 where workers do not contribute to social security (i.e., the pension and health system). This employment category belongs to the *informal sector*, which aggregates all informal workers. The second category is where workers contribute to social security. This employment 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.

More precisely, I define informal employment in this paper based on whether the worker contributes to the health or pension system. Figure 1 plots the wage density for all the workers in the two sectors. The figure shows strong bunching around the MW for formal workers, which is much less pronounced for informal workers. Thus, many workers in the formal sector have a binding restriction; this "stickiness" helps explain why there are no wage losses for formal workers in response to an immigration shock.

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



Note: Informality is defined according to contribution to the social security system. All the wage data are stacked across periods and departments in this figure. In addition, the sample is restricted to full-time workers between 18 and 64 years old. Log wages are in weekly real terms using monthly CPI from DANE. The kernel function is Epanechnikov, and the bandwidth is optimal. 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 57.3% in 2013 to 51.4% in 2019. The opposite occurs for Venezuelan workers, where the same rate increases from 61.8% to 88.8% in the same period, indicating that almost all new arrivals of Venezuelans got employment in the informal sector.

Analyzing the workers' perspectives of the two sectors helps understand the setting better. Being part of the formal sector offers several benefits for workers, including a guaranteed minimum wage, access to paid vacations, severance assistance, and annual service bonuses that enhance overall working conditions. In the formal sector, workers contribute 8% of their monthly salary to the health and pension system. In contrast, those in the informal sector do not pay taxes but still can access the health system as subsidiaries. So, the main distinction is the provision of a pension upon retirement. If migrants discount the future heavily, they can see the contributions as costly, but overall, the pecuniary factors would be more beneficial for workers in the formal sector. In terms of non-wage attributes, informal jobs may offer more flexibility or autonomy that migrants can prefer. López García (2015) provides a more extended discussion on why workers end up in informal jobs.

Next, comparing both Venezuelans and Colombians, in 2019, Venezuelans have a higher labor force participation rate (85.4% versus 78.8%), a higher employment rate (72.6% versus 69.8%), and a higher unemployment rate (14.9% versus 11.5%), see Tables 1a and 1b. The higher employability of migrants can 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 (44.1% of all native workers in 2019 were self-employed), and relative to Venezuelans, the share is slightly higher (48% in 2019). Note that 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 rates)

	LFP	Employment	Unemployment	Informality	Self-employment	N (15-64)	Population
2013	80.2	72.2	9.9	57.3	47.2	326,205	21,382,610
2015	80.4	72.6	9.7	54.7	45.2	436,178	22,105,009
2017	79.8	71.4	10.6	52.9	45.1	426,432	22,675,397
2019	78.8	69.8	11.5	51.4	44.1	411,864	22,493,469

(b) Venezuelans (in rates)

	LFP	Employment	Unemployment	Informality	Self-employment	N (15-64)	Population
2013	84.0	72.6	13.7	61.8	45.9	544	34,206
2015	77.5	69.4	10.4	64.9	48.7	900	47,312
2017	84.4	71.0	15.9	81.6	50.9	3,834	195,451
2019	85.4	72.6	14.9	88.8	48.0	17,030	1,039,797

Note: LFP stands for labor force participation. This table calculates the rates using national survey weights from GEIH. I restrict the sample to the population aged 18- 64 in urban areas. In addition, panel (a) restricts to natives living in Colombia for more than one year. The denominator in each rate is the following: for labor force participation and employment rates, it is the working-age population; for the unemployment rate, it is the labor force population; and for the informality and self-employment rates, it is the employed population. Source: GEIH, 2013 to 2019.

Last, Appendix Figure A.1 computes the extent of migrants' occupational downgrading, showing the likelihood that actual and predicted wages of migrants are in a certain percentile of the native wage distribution. The results show that migrants are overrepresented at the bottom of the distribution, meaning they are nearly three times more likely to be in the lowest wage percentiles than natives, while they are underrepresented in the upper part. Appendix Figure A.1 also shows Venezuelan predicted wages, constructed using the returns to education and experience of natives, but with their observed levels of education and experience. This measure of migrant downgrading indicates that migrants consistently downgrade, as the solid line is over the dashed line between the 20th and 80th percentiles of the distribution.

4 Empirical Specification

4.1 Event Study

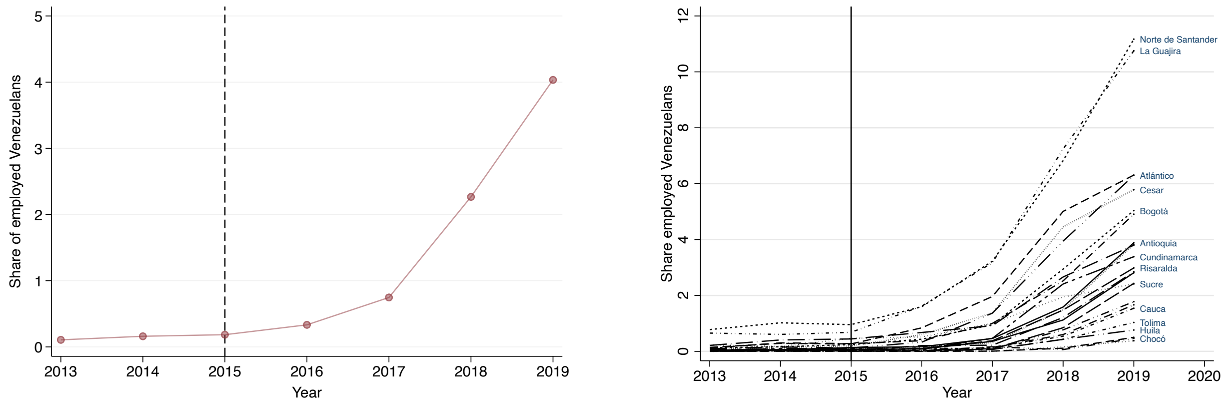
I use an event study that compares pre- and post-treatment trends as the main empirical strategy. The baseline year of comparison is 2015, the last year before the massive increase of Venezuelans in Colombia (see Figure 2a). Having data until 2019 is an improvement relative to Caruso, Canon, and Mueller (2021). These authors analyzed this migration event up to 2017, but migrants' arrivals doubled in 2018 and again in 2019. Appendix G shows that this limitation is critical to explain their sizable negative wage effect.[15] As a motivation for the spatial approach undertaken here, Figure 2b shows the differential intensity of migrant arrivals across departments in Colombia. Note that departments have a local labor market in their capital city connected with the surrounding smaller cities, with some degree of independence from labor markets in other departments.

[15] A recent paper by Lebow (2022) summarizes the differences in the wage effects of immigration in Colombian studies.

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

(a) National migration

(b) Migration by departments



Note: Panel (a) is at the national level. 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 DANE, 2013-II to 2019.

More precisely, I estimate an event study regression that interacts year dummies (T_t) with the treatment variable to obtain pre- and post-treatment coefficients that measure the dynamic impact of the Venezuelan immigration relative to the base year 2015. Specifically, I estimate the following equation:

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

Here, the standard errors are clustered at the department level, γ_t and γ_d represent fixed effects of year and department, and $\Delta M_{d,2018}$ is a time-invariant treatment variable defined as:[16]

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

where the numerator is the stock of employed migrants from Venezuela (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 migrants from Venezuela in d whose year of arrival was 2015 or earlier. Employed migrants are Venezuelans or returning Colombians from Venezuela, and the denominator $L_{Total,d,2018}$ is the total employed population in the department.[17]

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

[17]In the Appendix F, I show that results are almost identical if I use a denominator from the GEIH survey in 2015

Considering a time-invariant (i.e., the migration share in 2018) instead of a time-varying (i.e., immigrant share in the current year t) treatment variable is helpful for three reasons. First, this allows me to use the full count of the 2018 census instead of a survey to construct migration shares, thereby reducing the influence of sampling error in migration shares that arises in traditional surveys (Aydemir and Borjas, 2011).[18] Second, I remain agnostic regarding the timing of the effects of increased migration. In contrast, a more traditional panel regression with time-varying shares relates migration shares in year t to native employment and wages that same year, thereby implicitly assuming that immigration immediately affects native labor market outcomes. Third, using a time-constant migration shock allows me to integrate the placebo tests on pre-trends for the years $t < 2015$ transparently within the same analysis.

Given these advantages, the use of a time-invariant treatment variable has become common even in settings in which the treatment builds up over time (see, for instance, Dustmann, Schönberg, and Stuhler (2017) or Autor, Dorn, and Hanson (2021)).[19] Of course, the coefficient estimates from equation (1) need to be interpreted in conjunction with evidence on the timing of the migration shock, as shown in Figure 2a. For example, the coefficient β_{2016} would be rescaled small simply because most of the shock has not happened yet. For ease of interpretation, I focus on the coefficient β_{2018}, which is already scaled as it matches the census year. Moreover, as a robustness test, I also show wage and employment results in Appendix E.1b using a time-varying migration measure from the GEIH survey.

Next, in the spatial setup, the estimates can be attenuated if there exist mobility responses of inputs, say of native workers or capital, from areas more affected by the immigration shock to areas less affected (Borjas, 2006). In my setting, I test this hypothesis with retrospective information from the GEIH survey to show that there are no changes in natives' inflows to the most affected areas but slightly more natives' outflows from these areas (see Appendix Figure F.2). As a robustness check, I exclude the internal movers from the estimation sample to find similar results.

The last limitation is the small number of treated units ($N = 24$), which increases type I error

(see Figures F.3a and F.3b).

[18]Moreover, the survey weights were built using the 2005 census, and its reference frame was not updated until the 2018 census, creating additional uncertainty when constructing migration shares using the survey.

[19]The use of a time-invariant treatment variable is particularly attractive in settings in which the intensity of the treatment increases proportionally across regions over time, i.e., in which the measure of exposure for different years differs only in scale. As shown in Figure 2b, this is the case in my setting.

rates (Pustejovsky and Tipton, 2018). For this reason, to control for the over-rejection of the null hypothesis, I implement the wild cluster bootstrap method of Roodman et al. (2019), reporting the p-value for the main estimates. Moreover, I built a more detailed definition of local labor markets that yields more areas ($N = 51$) to find that the estimates are more negative than the department sample. Showing the number of treated areas does not drive the negative wage and employment coefficients (see Appendix C). Still, as the standard errors are quite similar in both cases and the sampling error is much higher for the larger sample, especially by subgroups, I focus on the sample of the 24 representative departments in the analysis. Last, I do not use sample weights of departments 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.[20]

4.2 Instrumental Variables

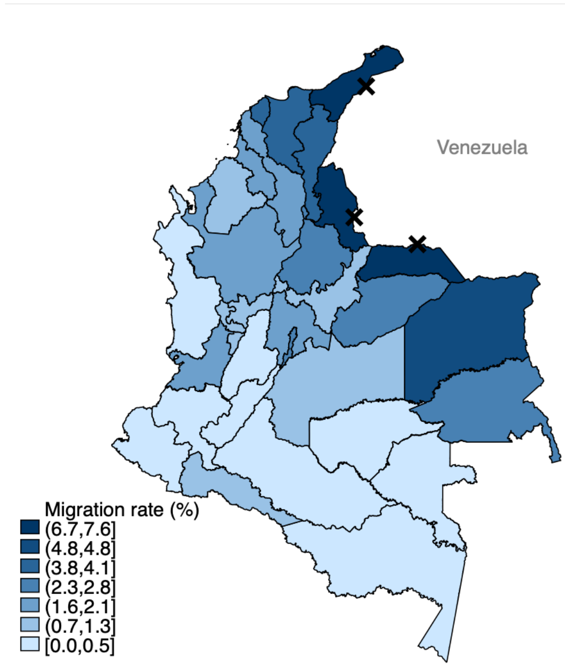
If migrants self-select into the departments where the economic conditions are better, $\Delta M_{d,2018}$ is going to be endogenous, and the OLS estimates downward biased.[21] To show it geographically, Figure 3 plots the Colombian map with the immigration rate $\Delta M_{d,2018}$ by departments. The highest immigration rates are observed in the departments closer to the Venezuelan border, especially near the main crossing bridges.[22] 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 3).

[20]Regression weights are typically 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 more negative estimates for native wages and more positive ones for employment (see Appendix Figure F.6a), and also results by the formal and informal sectors hold (see Appendix Figures F.7a and F.7a).

[21]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.

[22]The data for the outcomes is available to 24 departments, not to all 33 in the country. The missing nine departments, mostly located in the Amazonia and Orinoquia regions, only account for 2.8% of Colombia's total population, according to the 2018 census.

Figure 3: **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.

To address this, I estimate equation (1) after instrumenting it with two variables: (i) 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 I construct it 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.[23]

[23] 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. I decompose the instrument 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 Cristóbal concentrate 0.7 of the weights, which sum up to 1. In fact, those cities are closer to the border with Colombia; hence, effectively, the instrument compares cities closer to the border with those further away. To name some characteristics of these migrants, Maracaibo is an industrial city focusing on oil extraction, while San Cristóbal relies more on the service

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, 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 immigrants. A threat to this identification strategy arises if the border states suffer other economic shocks, such as less trade than the counterpart far-located states (violation of the exclusion restriction). To deal with these concerns, Figure F.1 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. Still, I run placebo tests for differences in pre-treatment outcomes to find insignificant pre-trends for wages and employment.[24] Furthermore, I drop each of the border states and all of them simultaneously from the main regression to find closely similar results.

Next, I include export patterns with Venezuela as a control (measured as the share of regional exports in USD to Venezuela over total exports in 2015 for every department) to find that the wage and employment estimates are slightly more negative. I also include a control of department GDP (a proxy of business cycles) to show that estimates for wages and employment are similar, yet GDP may be a "bad control" (Angrist and Pischke, 2008, p. 49). Last, I show that the decrease in formal employment is concentrated in the smaller firms, presumably less affected by trade and more affected by migration (immigrants work mainly in small firms). Hence, the main concerns of the exclusion restriction of this instrument may be alleviated.

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

$$z_{2,d} = \left(\frac{Ven_{d,2005}}{Ven_{2005}} * M_{2018}\right)/L_{d,2012}, \tag{4}$$

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 $L_{d,2012}$ in d before immigrants arrive as in Card (2001), whereas M_{2018} are Venezuelan arrivals to Colombia in 2018.

The validity of the past settlements instrument $z_{2,d}$ relies on the fact that the network effects in that location attract new arrivals to department d, while current economic trends in d are unlikely to be systematically related to lagged immigration shares (when those shares are lagged sufficiently).

and commerce sector.

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

If this holds, the instrument is valid because the lagged immigrant location is related to new arrivals (relevance) but not to current economic conditions (exogeneity). However, this assumption can 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 F, I show the results for wages and employment of natives using shares from the census of 1973, and they are not significantly altered. Last, as immigrants tend to locate in similar places over time, the past settlements instrument is serially correlated, at least for the US (Jaeger, Ruist, and Stuhler, 2018). This is problematic because the short-run response of immigration can be conflated with the general equilibrium adjustment of previous waves of immigrants. In this paper, it is possible to break the serial correlation since migrant arrivals surged rapidly in a country with very few immigrants in the past.

Formally, the exclusion restriction of the instruments is expressed as $E[z_{i,d}\Delta u_{dt}] = 0$ for $i = 1, 2$.[25] 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.[26] The distance instrument explains 89.7% of the immigration rate's variation, while the past settlements instrument explains 55.6%. The positive coefficient of the distance instrument indicates that areas closer to the border receive more migrants than areas located farther away. On the other hand, the positive coefficient of the past settlements instrument indicates areas where migrants were established before are receiving more migrants now.[27]

[25] 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, I use both instruments in the first-stage regression 5 to find that the null hypothesis of the instruments being exogenous is not rejected.

[26] In practice, for instrumenting the interaction of year dummies with the immigration shock of equation (1), I interact the instruments with year dummies.

[27] Recently, Lee et al. (2022) 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 287.1, and the past settlements instrument's F-statistic is 33.7.

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

	(1) $\Delta M_{d,2018}$	(2) $\Delta M_{d,2018}$
Distance ($z_{1,d}/100$)	0.437***	
	(0.026)	
Past settlements ($z_{2,d}$)		40.208***
		(6.927)
Constant	-1.294***	0.995***
	(0.236)	(0.250)
N	24	24
R^2	0.897	0.556
F st	287.1	33.7

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.

5 Baseline Estimates

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

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 (2024), 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 4a).[28] 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.[29]

Figure 4a shows that the OLS estimates are negative and, even if they could be upward biased

[28] Instruments predict a higher 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 between 0.24 and 0.27, depending on the instrument.

[29] The construction of the wage (or labor income) variable is shown in Appendix I. Importantly, it covers all types of labor income, including self-employed earnings.

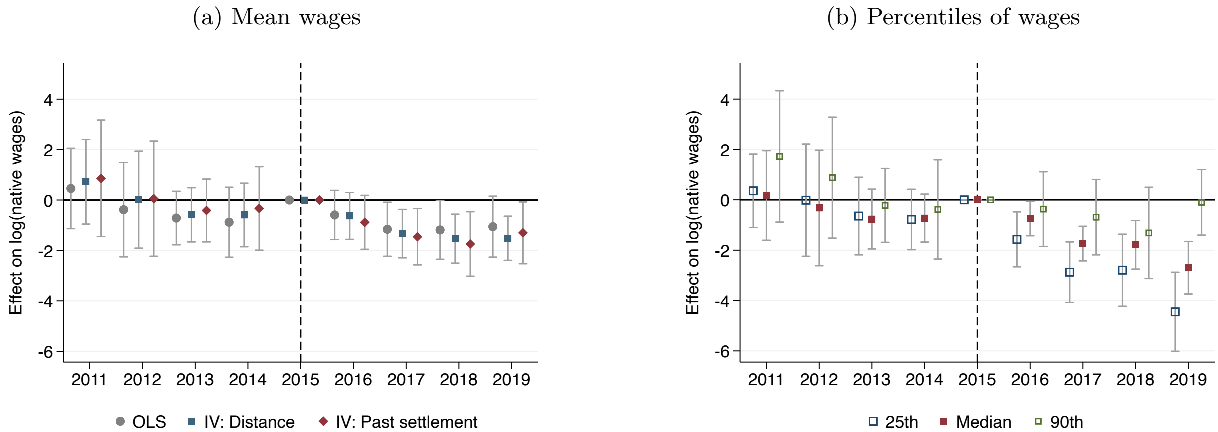
due to omitted variables, they do not differ much from the IV ones due to the high R^2 of the first-stage regression (see Table 2). The wage estimates are also negative and significant when using the two instruments separately.[30] A one pp increase in the share of employed Venezuelans decreases the wages of natives by around 1.5% and 1.7%, depending on the instrument selected. The point estimate of the distance instrument is significant when using the wild cluster bootstrap method (see Table 3). To further deal with compositional effects, I only focus on full-time workers in the wage analysis, and I show in the Appendix F that using residual wages (having controlled for individual characteristics such as age, years of schooling, and sex), instead of observed wages, yields similar results.

Scaling up these estimates, the total shock, according to the census, is a 2.1 pp increase in the employed population in Colombia (in absolute numbers, this means around 316,000 more employed immigrants relative to 2015). Hence, the total impact on wages for 2018 is around –3.2% and –3.7%. Note that the shock may 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.

Next, I analyze the distributional effects of immigration by taking specific percentiles of the distribution of wages as an outcome. Figure 4b shows the results of this exercise, where the outcome variable is the log wage at different percentiles. In effect, the native wages at the bottom part of the distribution are most affected by immigration. At the median, a more robust estimate to outliers and censored data, the coefficient is –1.8% 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.

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

Figure 4: **Event study estimates on hourly wages of natives**



Note: In the 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 the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. In 2011 and 2012, I assume all survey respondents were 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 the outcome. The F-statistic for the distance instrument is 287.1, and for the past settlements instrument is 33.7. In panel (b), I use the distance instrument. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

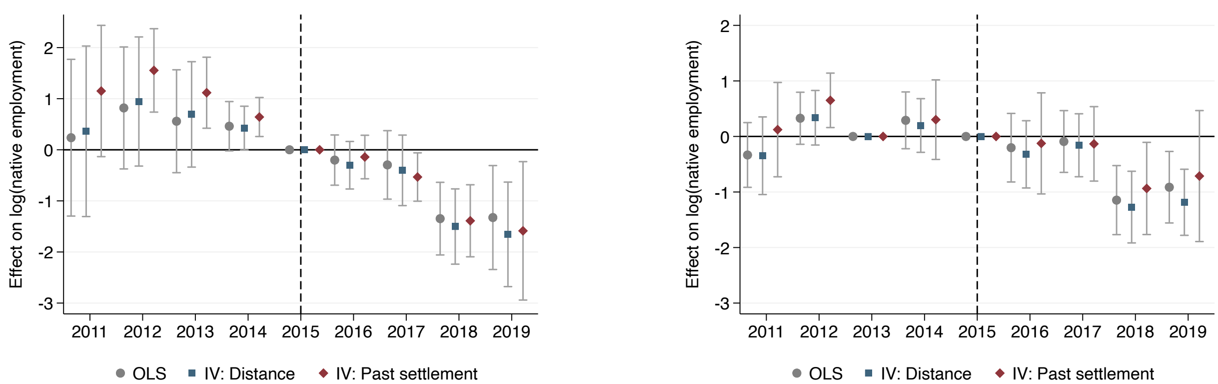
The explanation for such a negative finding hinges on several factors pointing to the high substitutability of natives and migrants. In effect, migrants speak the same language as natives, overcoming communication skills problems. In addition, they share cultural traits, which reduces wage discrimination, 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 leads to flexible wage cuts when migrants arrive.

Last, I construct a time-varying immigration rate ΔM_{dt} from the GEIH survey as the explanatory variable to test how the dynamic effects vary (see Appendix Figure E.1a). Interestingly, the estimates follow a similar (negative and significant) pattern but differ in magnitude from those obtained with the fixed $\Delta M_{d,2018}$ from the census. In 2018, the census year, the estimates 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, there are estimates similar to those of 2018. Thus, the effect in 2019 is similar, even if more migrants arrive. Last, the correlation of the yearly immigration rate ΔM_{dt} from GEIH with other post-treatment years is between 0.93 and 0.98, showing that the constant $\Delta M_{d,2018}$ from the census can capture accurately the dynamics of the treatment.

Employment Responses. I regress equation (1) for log native employment using OLS and the two instruments. Figure 5b shows that, in contrast to wages, with the instruments there are significant differences in the trends before the arrival of immigrants (see Figure 5a). This indicates that regions with higher predicted immigration rates were growing less than those with lower predicted immigration rates, suggesting a violation of the PTA with IV required for identifying the causal parameter. To address this problem, I control for the pre-trends in the employment regression of 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 ($\beta_{2013} = 0$).

Figure 5: **Event study estimates on native employment**

(a) Employment (b) Employment with trend adjustment



Note: In the 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 the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. Trend-adjusted estimates have, as a control, the growth in employment from 2013 to 2015. The F-statistic for the distance instrument is 287.1, and for the past settlements instrument is 33.7. The coefficients (in percent) are already multiplied by 100.

In Table 3, I show results for native employment, with and without the adjustment for pre-trends, while making inference with standard errors and the wild cluster bootstrap method. After adjusting for pre-trends, a one pp increase in the immigration rate reduces, on average, 1.3% local native employment using distance as the instrument and by 0.9% using past settlements as the instrument (coefficients of Table 3, column 4). The coefficient of the distance instrument is the only significant according to the wild cluster bootstrap p-values.[31]

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

	(1)	(2)	(3)	(4)
	Wages		Employment	
Panel A: OLS				
	-1.185*	-0.778	-1.347***	-1.146***
	(0.563)	(0.694)	(0.344)	(0.300)
Wild cluster bootstrap p-value	0.138	0.434	0.012	0.036
Panel B: IV				
Distance instrument				
	-1.533**	-1.240	-1.502***	-1.272***
	(0.469)	(0.607)	(0.356)	(0.312)
Wild cluster bootstrap p-value	0.038	0.040	0.014	0.222
Past settlement instrument				
	-1.743**	-1.556*	-1.389***	-0.936*
	(0.619)	(0.666)	(0.341)	(0.401)
Wild cluster bootstrap p-value	0.080	0.348	0.172	0.587
Trend-adjusted	No	Yes	No	Yes
N	216	216	216	216

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 with the immigration rate $\Delta M_{d,2018}$. The coefficient measures the effect in 2018 relative to 2015. In the 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 the second step, I use the department-level data to estimate β_{2018} 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).

6 Labor Market Linkages between the Formal and Informal Sector

In this section, I first introduce a theoretical model that helps to rationalize the empirical findings. Then, I document the differential response of wages and employment in the informal and formal sectors to the immigration shock.

6.1 Model

First, I build a partial equilibrium model inspired by Ulyssea (2018) with two labor inputs, formal and informal labor. Different from Ulyssea (2018) and Altındağ, Bakış, and Rozo (2020) model, I

[31] Appendix Figure E.1b 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.

flexibly allow for aggregate imperfect substitution of labor inputs as they could have, for instance, different sets of skills or work for different types of firms. This is key to rationalizing the empirical findings.

To motivate the firm's profit maximization problem, I show that firms combine formal and informal labor in production. To do this, I use a cross-sectional survey that distinguishes between informality at the firm and worker level (EMICRON, by its acronym in Spanish).[32] This exercise aims to describe how different the workforce composition of firms is, in terms of formal and informal labor, depending on their size. On average, all informal firms (that do not pay taxes) hire mostly informal workers (who do not contribute to the social security system). This contrasts with formal firms, where formal and informal labor are combined in their production function. Interestingly, the proportion of formal workers increases as the firm grows (see Table 5). The 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 4: **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 have the tax registry (RUT), while informal firms do not. Informal workers do not contribute to the health or pension system. I calculate the shares using survey weights. Only owners of firms with less than 10 workers are surveyed. Due to the small sample size, I aggregate firms with 7 to 9 workers. Source: EMICRON-DANE, 2019.

For the model, my objective is to represent in a tractable manner the effect of immigration on a local labor market, as my empirical analysis identifies the regional, not the firm-level, effects of immigration. With this in mind, there is a representative formal firm that hires workers L_f paying the official payroll taxes but also hires workers L_i off the books to avoid complying with the

[32] For firms, formality is defined based on the payment of direct taxes. For workers, formality is defined based on the contribution to the health or pension system. The EMICRON survey only covers owners of firms with less than 10 workers.

contributions to the social security system.[33] Moreover, the firm may be a self-employed worker who chooses to be formal or informal. The profit function of the firm is 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, I assume that $\tau(L_i) = L_i^\eta$ with $\eta \geq 0$, which captures the cost of evasion related to law enforcement exerted by the government.[34] The τ_f represents the payroll taxes the firm must enroll in 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 and productivity parameters are standardized such that $\alpha_i + \alpha_f = 1$. Why would informal and informal workers be imperfect substitutes? Ulyssea (2018) instead assumes that in the same skill group, they are perfect substitutes within the firm, and this assumption is reasonable given that many workers move between the formal and informal sectors. However, the object of interest when considering the labor market impact of immigration is the elasticity of substitution on the aggregate level, not the firm level. Even if workers are perfect substitutes within certain firms, informal and formal workers may be imperfect substitutes in aggregate production (e.g., because small firms employing many informal workers produce different goods than large firms hiring only formal workers). Last, the aggregate output price p is defined according to an inverse demand function: $p = C^{1-\epsilon} Q^{-(1-\epsilon)}$ as in Borjas (2013), where C is the number of consumers (natives plus migrants) and $\epsilon^D = 1/(1-\epsilon)$ is the price elasticity of demand in absolute value.[35]

Assuming the firm is competitive in the labor market (takes prices w_i and w_f as given), informal

[33] 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 of a representative firm.

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

[35] 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.*

and formal workers supply labor inelastically, and workers do not move between labor types. I then analyze changes in the optimal labor choices of the firm when wages for informal workers change as a result of the immigration shock (I take formal wages as fixed in the short run due to the MW).[36] After some algebraic derivations in Appendix H, 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:

$$\varepsilon_{L_i,w_i} = -\frac{(1-\epsilon s_f - \rho s_i)}{(1-\epsilon)(1-\rho)+\eta(1-\epsilon s_f-\rho s_i)}, \tag{8}$$

$$\varepsilon_{L_f,w_i} = -\frac{s_i(\epsilon-\rho)}{(1-\epsilon)(1-\rho)+\eta(1-\epsilon s_f-\rho s_i)}. \tag{9}$$

In these equations, the denominator in both expressions is always non-negative as $\epsilon \leq 1$, $\rho \leq 1$, and $\eta \geq 0$, where η is the integer power of cost function ($\tau(L_i) = L_i^{\eta}$). In turn, $s_g = \frac{\alpha_g L_g^{\rho}}{\alpha_i L_i^{\rho}+\alpha_f L_f^{\rho}}$ is the relative contribution of labor ($g = i, f$) to production before immigrants arrive. Thus, in the short run, the sign of ε_{L_i,w_i} will always be negative if $\epsilon < 1$ or $\rho < 1$. 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 when informal wages decrease:

1. If $\epsilon^D > \sigma$, there are *scale effects*, and the firm will hire more formal and informal labor. For instance, when the demand for the good is inelastic ($\epsilon = 1$), and formal-informal workers are imperfect substitutes ($\rho < 1$), the firm increases production using both labor inputs.

2. If $\epsilon^D < \sigma$, there are *substitution effects*, and the firm will hire less formal and more informal labor (either from migrants or natives). Importantly, the size of the response will hinge on s_i, which implicitly depends on the existence of a MW. With a relatively large MW, the demand for formal workers is lower, and s_i is higher. That leads to a bigger substitution of labor.[37]

[36] 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 is 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.

[37] In the working paper version of this paper, I extend this theoretical framework with a general equilibrium model that includes capital responses and differential labor supply functions of migrants and natives. A strong substitutability between labor types also predicts a decrease in formal employment in this model.

6.2 Overall Impacts on the Informal and Formal Sector

After describing the main implications of the model, I turn to the data to test which case occurs in this setting. I report results for wages and employment at the regional level using the distance instrument for expositional clarity.[38] Figure 6a shows the effect of immigration on wages of formally and informally employed natives.[39] Regional informal wages experience a negative wage effect (2018 coefficient is –1.9% with wild bootstrap p-value of 0.010), while the coefficient for regional formal wages is insignificant.[40] In line with the *substitution effects* case of the model where formal wages are downward rigid, and formal-informal workers are strong substitutes.

Compared to the Turkish case, Aksu, Erzan, and Kırdar (2022) finds a positive effect on wages of formal workers, whereas Akgündüz and Torun (2020) documents an increase in natives' tasks complexity.[41] 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 in regional formal employment (2018 coefficient is –1.9% with wild bootstrap p-value of 0.018), while regional informal employment presents an insignificant but negative point-estimate. The role of the MW helps to explain these findings since formal workers have a high probability mass to the right of the MW (see Figure 1), and thus there are no further wage drops for certain workers. This can translate into higher employment losses as informal wages are freely adjusted, and formal-informal workers are substitutable for the firm (as shown in the model). To test this hypothesis, I divide formal employment by MW earners and others to show more directly that the employment losses are concentrated in the MW earners (see Appendix Table F.2). In that sense, native formal workers earning the MW either move to the informal sector or become unemployed. Last, as a robustness check, I remove all border departments from the estimation sample to find slightly more positive estimates (see Appendix Figures F.8a and F.8b) and test the reliability of the survey data by comparing it with administrative records. In Appendix C, I show that using both sources yields fairly similar wage and employment estimates

[38] Results are unaltered if I use the past settlement instrument or control for trade with Venezuela.

[39] 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 the other way.

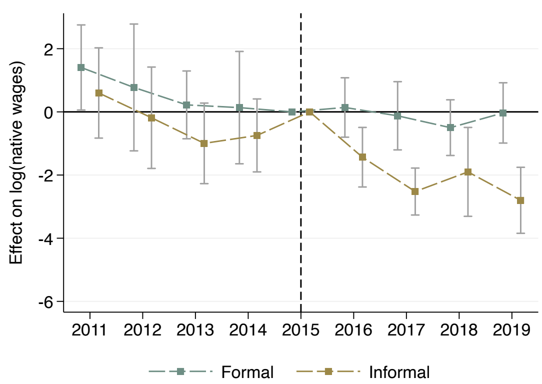
[40] Appendix Figure D.1 shows similar wage estimates using more frequent time windows (quarters instead of years). Also, pre-trends are not significant in this specification.

[41] Though, Dustmann et al. (2022) discusses that regional wages arising from cross-sectional data jointly measure the selection of workers and changes in the price of labor. Thus, there can be a compositional bias in the wage estimates.

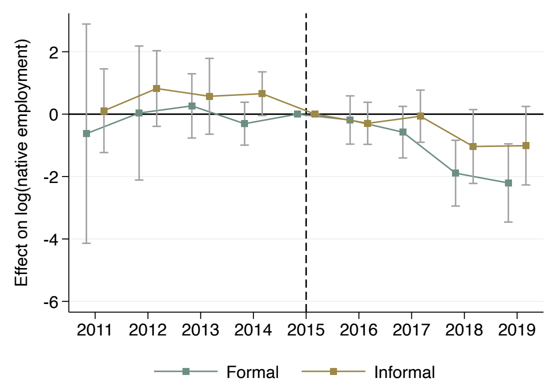
(see Appendix Figures C.1a and C.1b).

Figure 6: **Event study estimates by sector for natives**

(a) IV (distance) estimates on native wages



(b) IV (distance) estimates on native employment



Note: In the 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 the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. In panel (b), I do not use controls for pre-trends. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

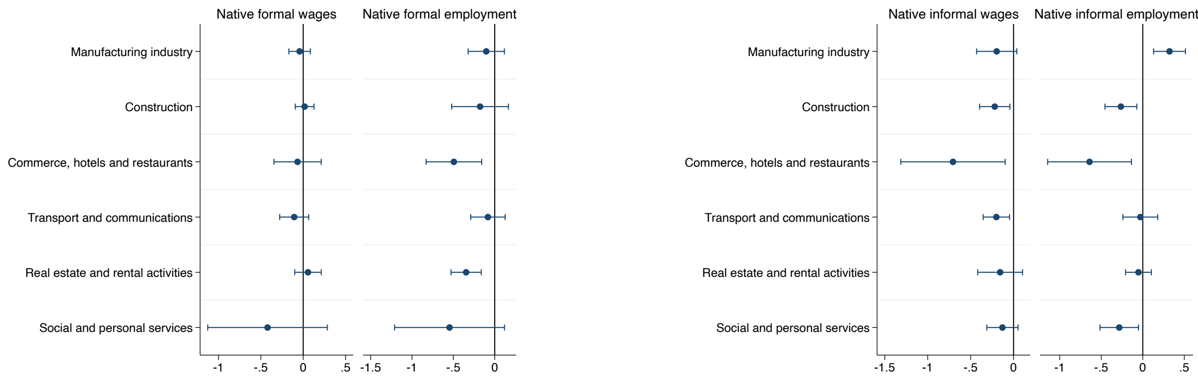
Because the effects across sectors can be driven by specific industries where immigrants concentrate, I then show which industry explains most of the results.[42] In particular, I decompose separately for the formal and informal sectors the regional employment response as the change in employment between period 1 and 0 for each industry i in the region r over the initial total employment in that region as $\frac{E_{r,1}-E_{r,0}}{E_{r,0}} = \sum_i \frac{E_{r,i,1}-E_{r,i,0}}{E_{r,0}}$. For wages, I decompose the wage change weighting by the industry initial size in each region as $\frac{w_{r,1}-w_{r,0}}{w_{r,0}} = \sum_i \frac{w_{r,i,1}-w_{r,i,0}}{w_{r,i,0}} * \frac{E_{r,i,o}}{E_{r,o}}$. I then use the fraction of each industry i as the outcome to yield its exact contribution to the overall employment and wage change in the informal and formal sectors. To start, Figure 7a shows the effects for the formal sector with no significant changes in formal wages, while for formal employment, the negative growth is explained by commerce, hotels and restaurants, and social and personal services, precisely the industries migrants are overrepresented (see Appendix Table A.2). Regarding the informal sector, Figure 7b shows that informal wages and employment are mostly affected again by commerce, hotels, and restaurants, while there is a positive employment impact on the manufacturing industry.

[42]Dividing the sample by industry and sector increases the sampling error considerably, so I focus only on the largest industries.

Figure 7: **Decomposition of estimates by industry for natives**

(a) IV (distance) formal sector

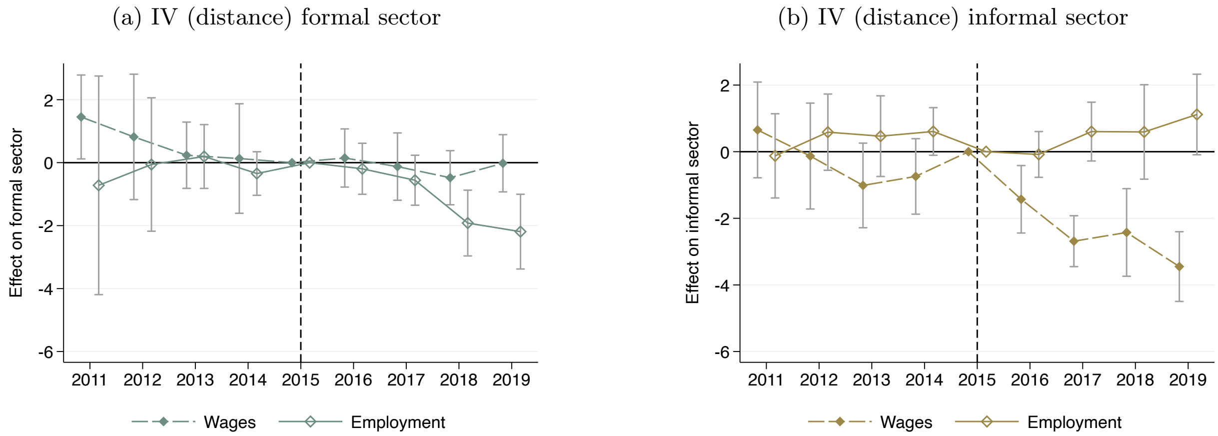
(b) IV (distance) estimates on informal sector



Note: In the 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 the second step, I use the department-level data to estimate the equation separately for each industry. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. The coefficient measures the difference between 2018 and 2015. I removed the industries with the smallest sample sizes. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

Next, I focus on overall employment (natives plus migrants) and total wages (counting natives and migrants) across the informal and formal sectors. The main takeaway is that there are insignificant effects for formal wages, while for informal wages, I find more negative wage estimates (see Figure 8a and 8b). In addition, formal workers get similarly crowded out after 2017, while informal employment increases when adding migrants, yet not significantly. Hence, the relative size of the informal sector in the local economies is growing due to the shrinkage of the formal sector.

Figure 8: **Employment and wage estimates by sector for natives plus migrants**



Note: In the 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 the 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. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. In 2011 and 2012, I assume all survey respondents were Colombian.The explanatory variable is $\Delta M_{d,2018}$. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

6.3 Results by Firm Size

I now investigate the typical firm size of employed immigrants. In 2018, nearly 90% Venezuelan immigrants were employed in firms with less than twenty workers. Hence, immigrants are over-represented 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 too.

Table 5: **Share of Venezuelan immigrants by firm size**

Firm Size	Distribution (%)
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 between 18 and 64 years old in urban areas. Source: GEIH, 2018.

I then focus only on formal firms to show how binding the minimum wage is 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 6). Thus, the immigration shock is more salient in smaller firms, as immigrants disproportionally work there, and they employ a larger share of minimum-wage workers that are more substitutable with informal workers.

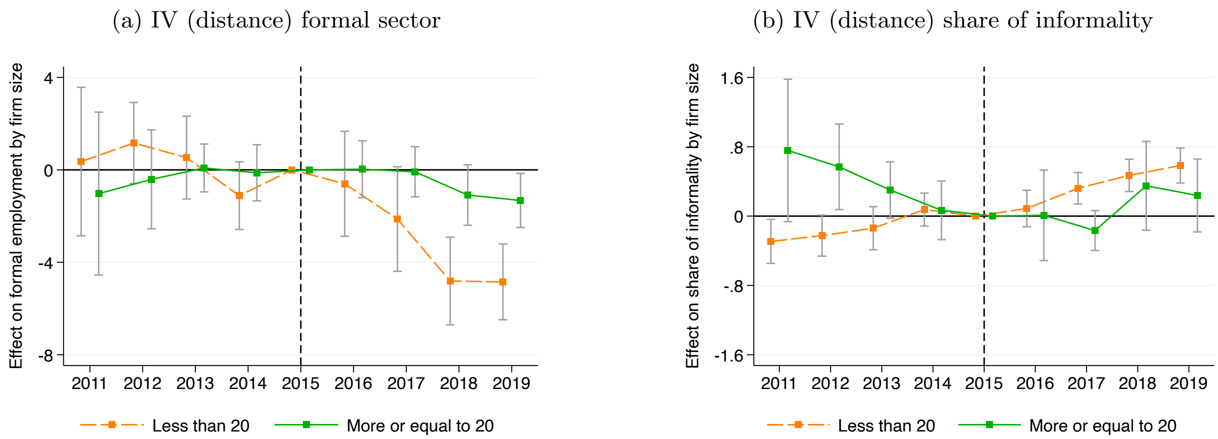
Table 6: **Composition of workers by firm size**

	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. The administrative data does not capture informal workers; only workers who contribute as employees are considered. Source: PILA, 2015-August.

For these reasons, I aggregate workers in two firm size categories: less than 20 workers (smaller firms) and 20 workers or more (bigger firms). As shown in Figure 9a, where the dependent variable is the logarithm of formal employment, smaller firms decrease more prominently formal employment in response to the immigration shock, in line with findings from Delgado-Prieto (2023) using Colombian administrative data.[43] Moreover, Figure 9b shows that the share of informal workers in smaller firms increases more compared to bigger firms.[44] This suggestive evidence indicates a possible change in the composition of the firm's workforce, with relatively more informal workers than formal workers.

Figure 9: **Event study estimates for formal employment and share of informal workers by firm size**



Note: In the 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 the 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. Departments in the regression are $N = 24$ per year. I do not explicitly control for pre-trends. I use a 95% confidence interval constructed from standard errors. In 2011 and 2012, I assume all survey respondents were Colombian. The F-statistic for the distance instrument is 287.1.

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 F.4a). Thus, employment changes by firm size are less driven by the entry of new formal firms.[45] 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 F.4b).[46] Overall, these findings support that employment changes are mainly due to changes in the workforce composition of smaller and

[43]Regarding formal wages, coefficients are insignificant for both firm sizes.

[44]Informal employment significantly increases for smaller firms but not as much to counteract the formal employment decrease.

[45]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.

[46]I measure the average number of workers who report working in a given firm size category with a specific question from the GEIH survey.

larger firms.[47]

6.4 Estimating the Elasticity of Informal Labor Demand and the Cross-elasticity

Figure 10 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 that market. Thus, if formal wages are fixed, the change in equilibrium employment in each sector can be expressed as:

$$\frac{\Delta L_i}{L_i} = \frac{\Delta w_i}{w_i} * \varepsilon_{L_i, w_i}, \tag{10}$$

$$\frac{\Delta L_f}{L_f} = \frac{\Delta w_i}{w_i} * \varepsilon_{L_f, w_i}. \tag{11}$$

Then, with the reduced-form estimates, I recover the values of the own-elasticity of informal labor demand and the cross-elasticity of formal labor demand, assuming there are no aggregate labor demand shifts.[48] First, the ratio between the change of total informal employment and the native informal labor income change yields the elasticity of labor demand $\varepsilon_{L_i, w_i} = \frac{0.59\%}{-1.89\%} = -0.31$.[49] This estimate measures the slope of the demand curve (see Figure 10), 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).[50] Last, the ratio between the total change of formal employment and native informal labor income change yields $\varepsilon_{L_f, w_i} = \frac{-1.92\%}{-1.89\%} = 1.02$, suggesting formal and informal workers are close to perfect substitutes in aggregate production.

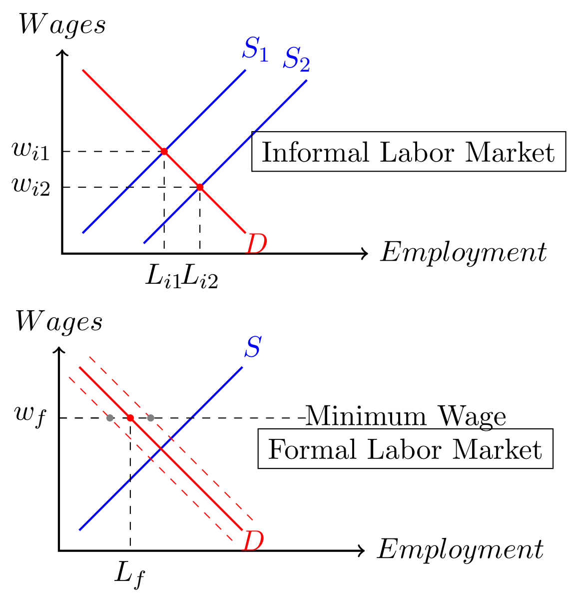
[47] Another mechanism to be analyzed is the effect on the creation of informal firms not registered in tax records.

[48] This assumption aligns with most of the literature, for a discussion of the wage effects of immigration allowing for demand shifts see Borjas (2013).

[49] 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.

[50] 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.

Figure 10: **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.

6.5 Heterogeneous Effects by Education, Job Type, Sex, and Age

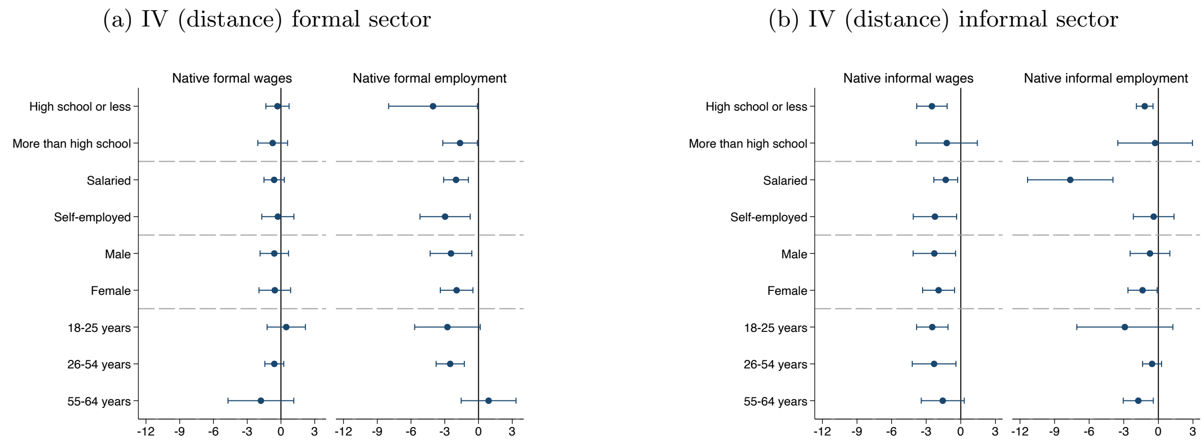
This subsection describes the heterogeneous effects of immigration across different native subpopulations. Because the immigration shock causes asymmetric employment and wage effects across the informal and formal sectors, I focus on the heterogeneous analysis of these two sectors. Figure 11a shows that wage effects in the formal sector are consistently around zero for the different subgroups. Note that even if some workers (e.g., natives with more than high school) earn more than the minimum wage, those workers are more likely to be covered by long-term contracts that prevent wage changes. On the other hand, formal employment effects are negative in most subgroups, especially for low-skilled natives (with high school or less) who are more substitutable with informal workers. In contrast, Figure 11b shows that wage effects in the informal sector are mostly negative and significant (only natives that are high-skilled and are old present an insignificant estimate).

It is worth noting that salaried or wage employment for natives in the informal sector decreases significantly, in line with the Turkish case (Aksu, Erzan, and Kırdar, 2022); for a 1 pp increase in the immigration shock, employment decreases by 7.6%.[51] Suggesting high substitutability between

[51] The reason the effect is so large is that salaried informal employment is a small subset of employment, whereas

informal natives and informal migrants in salaried jobs. In contrast, self-employment in the informal sector has an insignificant effect close to zero, and these contrasting effects help to explain why the impact on total native employment in the informal sector has a negative but insignificant coefficient.

Figure 11: **Native wages and employment estimates by subpopulations and by sector, 2015–2018**



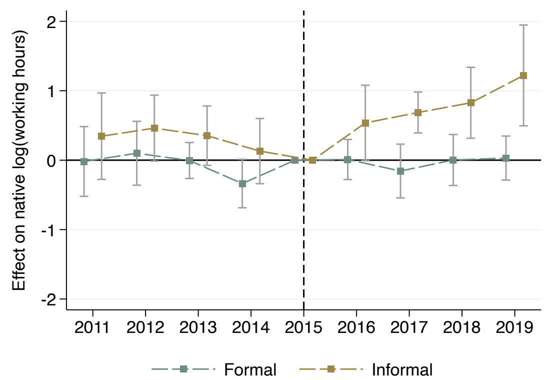
Note: In the first step, I use department survey weights from GEIH to construct department-level outcomes from individual data of natives between 18 and 64 years in urban areas. In the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 24$ per year. 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 the logarithm of hourly wages or employment; it is earnings instead of wages for self-employed workers. For employment, I use trend-adjusted estimates, controlled by the growth in native employment from 2015 to 2013 for each subpopulation. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

6.6 Intensive Margin

Next, I estimate the impact of the immigration shock on the intensive margin. In this case, the outcome is the log of average weekly working hours in each region. To rule out the compositional effects of part-time workers moving relatively more than full-time workers out of employment, I focus only on full-time workers for this analysis. Figure 12 shows that natives work more hours in all the post-treatment years in the informal sector in response to the immigration shock, while in the formal sector, working hours are unaffected.

the immigration shock is the share of overall employment.

Figure 12: **Event study estimates on log working hours of natives**



Note: In the 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 the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. The F-statistic for the distance instrument is 287.1, and for the past settlements instrument is 33.7. The coefficients (in percent) are already multiplied by 100.

7 Robustness Checks

The first robustness check relates to excluding areas on the Venezuelan border from the analysis. The Venezuelan crisis can affect these departments more through fewer trade links or lower business interactions. The main results of this exclusion, shown in Appendix Figures F.5a and F.5b, yield no significant variations in the estimates, even if I exclude all border departments at the same time (Figure F.6b) or focus on the effects across the informal and formal sectors (Figures F.8a and F.8b). Next, I restrict the sample to natives residing in the current department for more than one year, removing possible compositional changes of natives reallocating between departments due to the Venezuelan immigration. This is a relevant check as the internal migration responses of natives can attenuate the wage estimates. Importantly, when imposing this sample restriction, the estimates are even more negative (see Appendix Table F.1). 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. These coefficients are more positive for wages and employment when I use distance as the instrument.

The following robustness tests target the identifying assumptions of the instruments. For past settlements, there can be a correlation between the distribution of Venezuelans in 2005 and current economic trends. To test this, I construct two historically lagged Venezuelan census shares in Colombia from IPUMS (2019). Appendix Figure F.10 shows the coefficients using that instrument with three distinct shares (2005, 1993, and 1973), and there are no significant differences. Another threat to the instruments' identifying assumption is that trade or business patterns derived from the Venezuelan crisis could affect more severely geographically closer departments in which the instrument predicts more migrants. Hence, I use the change in real log GDP 2015–2018 and the share of regional exports in USD to Venezuela with respect to the world in 2015 as controls. The corresponding estimates of their effect on native wages, reported in Appendix Table F.1, confirm the statistical significance of these effects. Still, the estimates vary more with the past settlement instrument, while the coefficients of controls are insignificant in all specifications. Moreover, I replicate the event study figures for the informal and formal sectors, including as a control the share of regional exports and results again hold (see Appendix Figures F.9a and F.9a).

Next, I compute residual wages to check whether aggregating outcomes without netting out individual observables changes the results. As shown in Appendix Table F.1, the residual wage has a slightly more negative estimate for the distance instrument and a slightly more positive one for the past settlements instrument. Thus, there is a minor gain in regressing previous individual characteristics in the analysis, so compositional changes are less of a concern. Last, Borjas, Grogger, and Hanson (2012) discusses the differences between the mean of log wages and the log of mean wages. I use the log of mean wages for the main analysis, so I now use the mean of log wages for robustness. In this case, I find slightly more negative results than the baseline ones (see Appendix Table F.1).

8 Conclusion

This paper analyzes the impact of the Venezuelan mass migration on the Colombian labor market. By comparing regions with varying levels of exposure to the Venezuelan immigration, I find a negative effect on hourly wages and employment for natives. The differences between my findings with previous migration studies in Colombia are explained by the specification I chose, an event study design with IV that rules out potential confounding issues; the immigration rate I use (which is built with census data) reducing the measurement error of migrants; and the period of analysis, which captures accurately 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. It also causes a decrease in natives' formal employment, reflecting that firms substitute formal labor with high minimum wages for low-priced informal labor. Thus, the immigration shock affects the informal and formal sectors differently, the former via wage reductions and the latter via employment reductions. To explain these findings, I construct a model in which the firm substitutes formal for informal labor but only when formal and informal workers are strong substitutes. In settings with large informal labor markets, migration may lead to asymmetric responses between the informal and formal sectors, especially when the minimum wage is large and binding in the formal sector. Future research may focus on firm-level outcomes to show the consequences of substituting formal for informal labor on profits and production.

Acknowledgements. I am grateful to the editor, Shuaizhang Feng, and three anonymous referees for their insightful comments and suggestions. I am also 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 thank Jesús Fernández-Huertas, Jorge Pérez Pérez, Lorenzo Aldeco, and Sarah Schneider-Strawcynski for the helpful feedback. In general, I am indebted to all the comments from participants in the UC3M Applied Reading Group, the Junior Economics of Migration Seminar, 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 the DANE staff, especially from Geovanni Portilla. This paper was 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.

Funding. The author gratefully acknowledges the financial support from the Ministry of Science and Innovation in Spain through research grant PDI2019-108144GB-I00.

Declarations

Competing interests: The author declares no competing interests.

References

Akgündüz, Y. E. and Torun, H. (2020). Two and a half million syrian refugees, tasks and capital intensity. *Journal of Development Economics*, 145:102470.

Aksu, E., Erzan, R., and Kırdar, M. G. (2022). The impact of mass migration of syrians on the turkish labor market. *Labour Economics*, 76:102183.

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.

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. (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., Grogger, J., and Hanson, G. H. (2012). Comment: On estimating elasticities of substitution. *Journal of the European Economic Association*, 10(1):198–210.

Calderón-Mejía, V. and Ibáñez, A. M. (2016). Labour market effects of migration-related supply shocks: evidence from internal refugees in colombia. *Journal of Economic Geography*, 16(3):695–713.

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.

Clemens, J. (2021). How do firms respond to minimum wage increases? understanding the relevance of non-employment margins. *Journal of Economic Perspectives*, 35(1):51–72.

Corbi, R., Ferraz, T., and Narita, R. (2021). Internal migration and labor market adjustments in the presence of non-wage compensation. Technical report, Working paper.

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).

Delgado-Prieto, L. (2023). Immigration and worker responses across firms: Evidence from administrative records in colombia. Technical report, Working Paper, Universidad Carlos III de Madrid.

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., Otten, S., Schönberg, U., and Stuhler, J. (2022). Labor Market Effects of Immigration - Identification and Interpretation. Technical report.

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*.

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

Gulek, A. (2022). Formal effects of informal labor supply: Evidence from the syrian refugees in turkey. *Available at SSRN 4264865*.

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. (2022). The labor market effects of venezuelan migration to colombia: reconciling

conflicting results. *IZA Journal of Development and Migration*, 13(1).
Lebow, J. (2024). Immigration and occupational downgrading in colombia. *Journal of Development Economics*, 166:103164.
Lee, D. S., McCrary, J., Moreira, M. J., and Porter, J. (2022). Valid t-ratio inference for iv. *American Economic Review*, 112(10):3260–3290.
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.
López García, I. (2015). Human capital and labor informality in chile: a life-cycle approach.
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*.
OECD (2023). Earnings: Minimum wages relative to median wages.
Olivieri, S., Ortega, F., Rivadeneira, A., and Carranza, E. (2021). Shoring up economic refugees: Venezuelan migrants in the ecuadoran labor market. *Migration Studies*, 9(4):1590–1625.
Penaloza-Pacheco, L. (2022). Living with the neighbors: the effect of venezuelan forced migration on the labor market in colombia. *Journal for Labour Market Research*, 56(1):1–32.
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).
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 (2023). Venezuela situation. Technical report, Office of the United Nations High Commissioner for Refugees. Available on: https://www.unhcr.org/venezuela-emergency.html (last accessed 9th of January 2023).
Weber, S. and Péclat, M. (2017). A simple command to calculate travel distance and travel time. *The Stata Journal*, 17(4):962–971.

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					Sex	Schooling			Sample	
	0-14 (%)	15-28 (%)	29-40 (%)	41-64 (%)	65+ (%)	Male (%)	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+ (%)	Male (%)	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+ (%)	Male (%)	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 of 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.

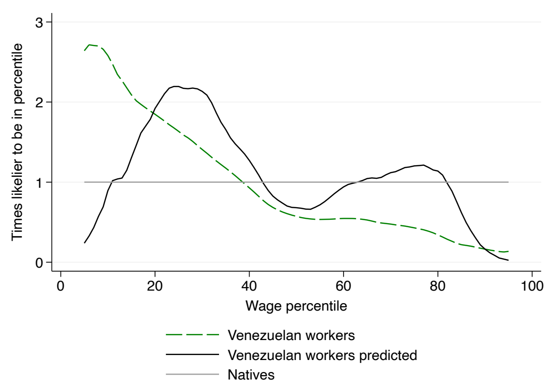
Appendix Table A.2 shows in which industries Venezuelans and Colombians are working. Immigrants are overrepresented with respect to natives in two industries. The first one is commerce, hotels, and restaurants, where almost half of all Venezuelan workers have a job (46.9%), while the corresponding share 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 and rental activities (6% versus 9.5%) and 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, hotels and restaurants; 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 and communications	9.6	5.6	58.1
Financial intermediation	1.9	0.6	13.8
Real estate and rental activities	9.5	6.0	46.1
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 sum up to 100% after adding the unknown industry 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.

Figure A.1: 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 the month as controls. I restrict the sample to full-time workers between 18 and 64 years old in urban areas. I top code wages above the 99% percentile of the wage distribution. Source: GEIH-2018.

B Results with Alternative Definition of Local Labor Markets

One potential shortcoming of the analysis is the number of treated areas per year ($N = 24$). To address this, I build a larger sample of functional urban areas (FUAs) in Colombia using the methodology of Sanchez-Serra (2016). This sample consists of the 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 increase considerably as the GEIH does not statistically represent hese areas.[B.1]

[B.1] The municipality information of the survey is not publicly available. I had to access secured computers from DANE research data centers to retrieve results.

This exercise compares the estimates for 24 departments and 51 FUAs constructed with the GEIH. I use a slightly different distance instrument as 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 in Dustmann, Schönberg, and Stuhler (2017). Regarding natives, when focusing on the FUAs, the coefficients tend to be more negative. For the departments, the employment estimate is –1.4%, while for the FUAs, the corresponding estimate is –1.8%.[B.2] Overall, the estimates I find are more conservative, but the main results hold when increasing the sample of analysis.[B.3] Thus, as standard errors are quite similar in both cases and the sampling error is higher for FUAs, I focus on the sample of the 24 representative departments in the main analysis.

Table B.1: **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	-1.66**	-1.41***
	(0.622)	(0.390)
F-st	16.27	14.93
N	24	24
Panel B: FUAs		
IV: Distance to nearest crossing bridge	-2.16***	-1.85**
	(0.564)	(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 migration 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 migration rate from the GEIH survey is constructed with only Venezuelan-born migrants in the numerator. 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.

C Comparison of Administrative and Survey data

I use the Colombian administrative data on the affiliation to social security (PILA, by its acronym in Spanish) as an additional check for the formal sector results. Note that the regional estimates using PILA are not the first-best option as the municipality information in the administrative form of the worker was not verified by national authorities until 2019.[C.1] Figure C.1a shows the results for formal wages with the administrative

[B.2] The employment increase in the larger sample is 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).

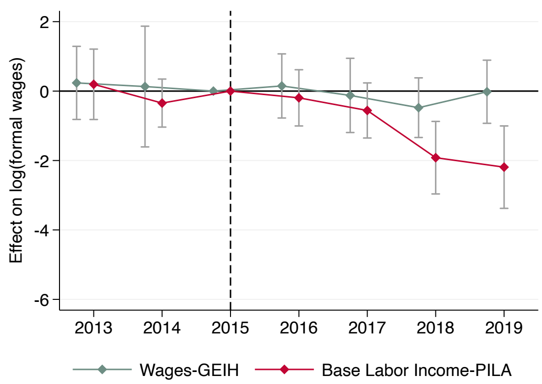
[B.3] Results for the larger sample also hold when dividing by the formal and informal sectors.

[C.1] In any case, to have consistent estimates, the measurement error of this variable must be uncorrelated with the selected instrument.

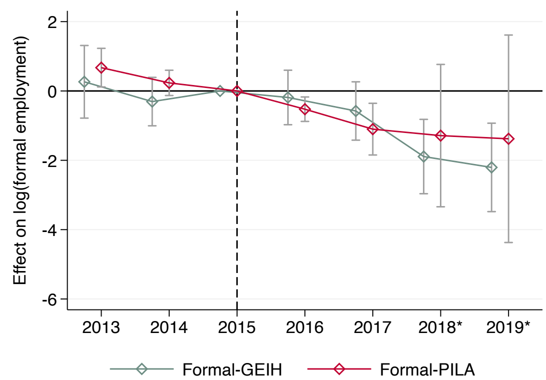
data, and they yield similar estimates when using the GEIH survey.[C.2] For formal employment, I find negative coefficients on both sources, yet not significant for PILA, in part due to the measurement problem of the geographical variable.

Figure C.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: In the 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 the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. The sample is not restricted and covers the universe of formal workers in the department. F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE. Source: GEIH 2013–2019 and PILA 2013–2019.

D Event study with quarterly information

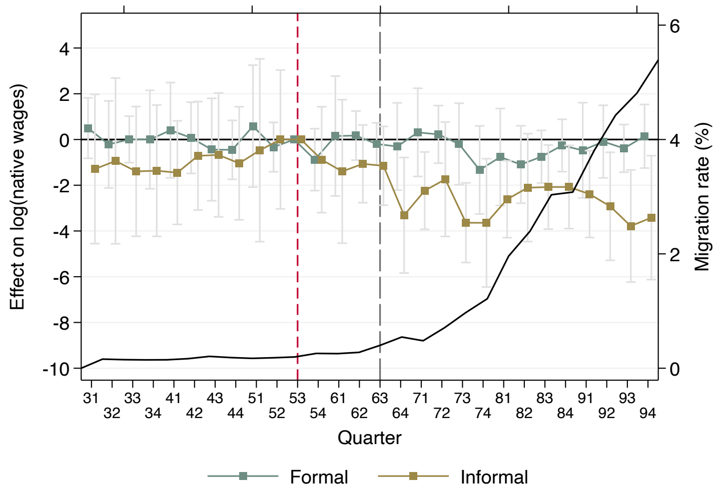
I perform the event study analysis at a higher time frequency, using quarters instead of years. The base period is 2015Q3 when the Venezuelan government closed the border. The empirical specification is the same as before, interchanging the yearly subscript y with the quarterly one q. I estimate the following model with the distance instrument:

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

First, note that the red dashed line in Figure D.1 represents the base period of analysis of the event study regression. The grey long-dashed line represents the re-opening of the border between Colombia and Venezuela. Finally, the thick black line is the quarterly immigration rate built 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 non-monotonic, as increases in the thick black line (immigration rate) do not necessarily translate into more negative impacts on informal wages.

[C.2] In PILA, it is basic income and refers to the amount of labor income declared in the social security contributions.

Figure D.1: Event study estimates on hourly wages by quarters and affiliation to social security



Note: The black line is the quarterly immigration rate from the GEIH survey. In the 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 the second step, I use the department-level data to estimate β_q from equation (D.1) and use clustered standard errors at the department level. I use a 95% confidence interval constructed from standard errors. The sample is restricted to permanent Colombian residents between 18 and 64 years in urban areas. The base period is 2015Q3. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

D.1 Impact on Prices

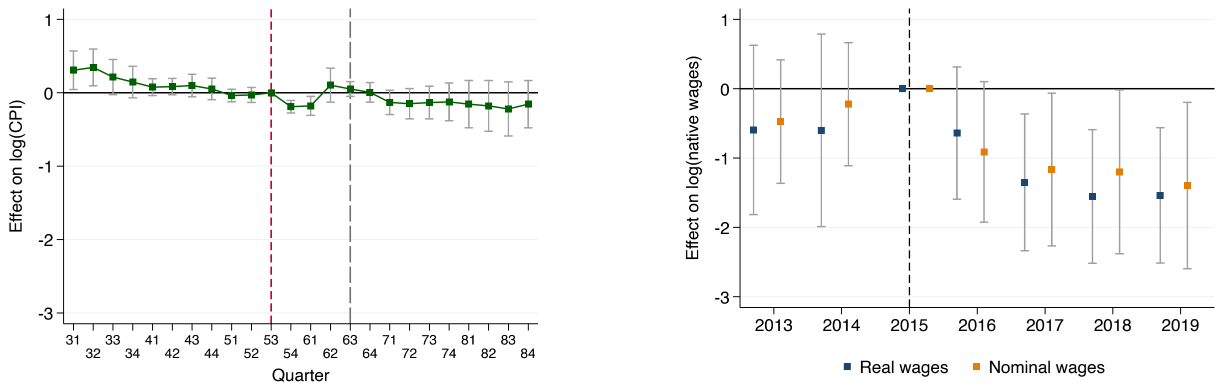
The wage and employment responses from immigration are not the only adjustment channels in a labor supply shock. To complement the findings, I also analyze the price response of a bundle of goods and services. Two opposing mechanisms affect prices. The first is the demand effect that arises from the relatively higher consumption of goods and services from migrants. The second is a negative supply effect from relatively lower wages that alter production costs. Moreover, there is also a search channel, as migrants can be more sensitive to prices, which spurs more competition in some markets, driving prices down. Hence, the final impact on prices is ambiguous.[D.1]

In Appendix Figure D.2a, I find insignificant estimates close to zero, using the distance instrument, on the overall consumer price index (CPI), suggesting a relatively more substantial 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 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 when capital is fixed.

[D.1] 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).

Figure D.2: 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 constructed from standard errors. The base period is 2015Q3. β_q from equation (12) are the plotted coefficients in (a), β_t from equation (1) are the plotted coefficients in (b), and standard errors are clustered at the department level. The F-statistic for the distance instrument is 287.1. The base year for the index is 2008. The coefficients (in percent) are already multiplied by 100.

E Event Study with a Time-varying Immigration Rate

As a robustness check, I also perform an event study analysis with an immigration rate that varies in post-treatment years using the GEIH survey. To do that, first, I rewrite equation (1) into a differences model, and then I interchange $\Delta M_{d,2018}$ from the census with ΔM_{dt} from the GEIH survey.[E.1] In this model, I estimate separate regressions for each year $t = \{2013, .., 2019\}$ of the following form:

$$Y_{dt} - Y_{d,2015} = \delta_t + \theta_t \Delta M_{dt} + u_{dt} - u_{d,2015}. \tag{E.1}$$

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

$$\Delta M_{dt} = \frac{L_{Ven,d,t} - L_{Ven,d,2015}}{L_{Total,d,2015}}, \tag{E.2}$$

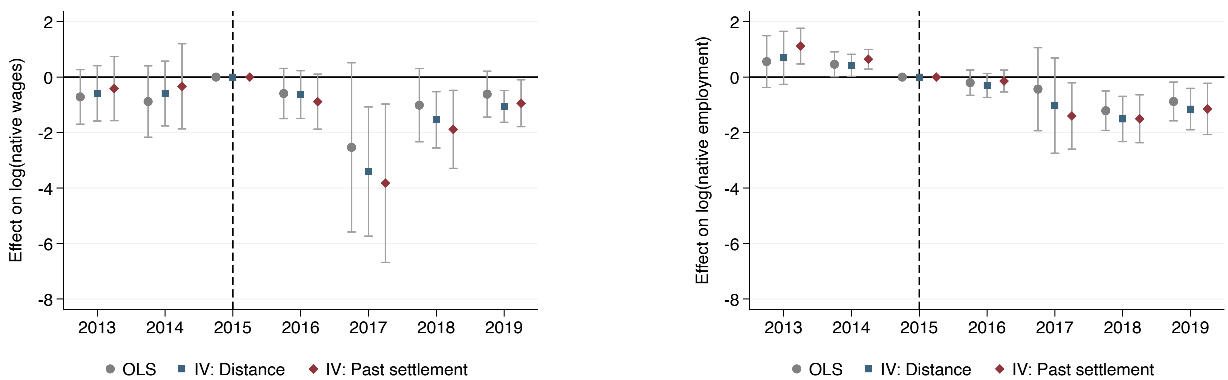
where the numerator measures the employed Venezuelans (between 18 and 64 years) in d for all the post-years t from the GEIH survey, relative to the base period (2015). The denominator is fixed on the base year, following Card and Peri (2016). Figures E.1a and E.1b show the results on native wages and employment using the time-variant migration variable.

[E.1] The migration rate from the survey is constructed with only Venezuelan-born migrants in the numerator.

[E.2] Although the coefficient in equation (1) and (E.1) is the same when using a time-invariant treatment variable, the standard errors are going to be different. This is because regression (1) groups observations of units over time, allowing to control for arbitrary serial correlation of errors, while regression (E.1) gives differenced errors relative to a specific time interval for every unit. With this in mind, (1) uses clustered standard errors, while for (E.1), it is only possible to use robust standard errors.

Figure E.1: Event study estimates using ΔM_{dt} from GEIH survey as explanatory variable

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



Note: In the 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 the second step, I use the department-level data to estimate θ_t from equation (E.1). Departments in the regression are $N = 24$ per year. The explanatory variable is $\Delta M_{d,2018}$ before 2017, and after is ΔM_{dt}. I use a 95% confidence interval constructed from standard errors. I restrict the sample to permanent Colombian residents between 18 and 64 years in urban areas. Hourly wages are in real terms using the monthly CPI from DANE.

With this model, I also show precisely how the endogeneity of the immigration rate biases the OLS estimates. Suppose the outcome of interest is $ln(wages) = ln(w_{dt})$ such that the estimated coefficient of interest becomes

$$\hat{\theta}_t = \frac{\hat{Cov}(\Delta M_{dt}, \Delta ln(w_{dt}))}{\hat{Var}(\Delta M_{dt})}, \tag{E.3}$$

where $ln(w_{dt}) - ln(w_{d,2015}) = \Delta ln(w_{dt})$.[E.3] Next, plugging model (E.1) in the last expression yields

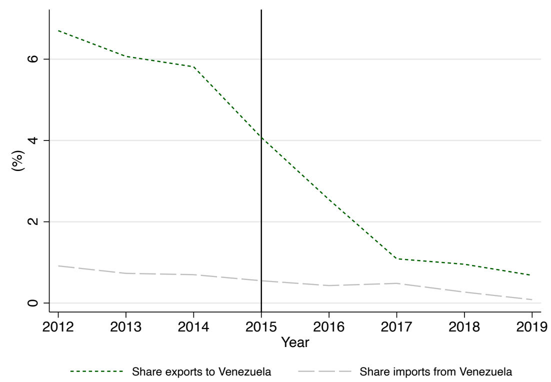
$$plim(\hat{\theta}_t) = \theta_t + \frac{Cov(\Delta M_{dt}, \Delta u_{dt})}{Var(\Delta M_{dt})}. \tag{E.4}$$

Thus, even if the time-invariant heterogeneity is removed, a downward bias still emerges if migration is concentrated in places with better economic trends. In other words, the unobservables that change over time will be correlated with the immigration inflows (i.e., $E[\Delta M_{dt} \Delta u_{dt}] \neq 0$).

[E.3] The resulting expression of $\hat{\theta}_t$ is explained as follows: the numerator measures the covariance between the inflow of migrants and the change in wages relative to the base period. At the same time, the denominator weights this covariance with the observed dispersion of migration.

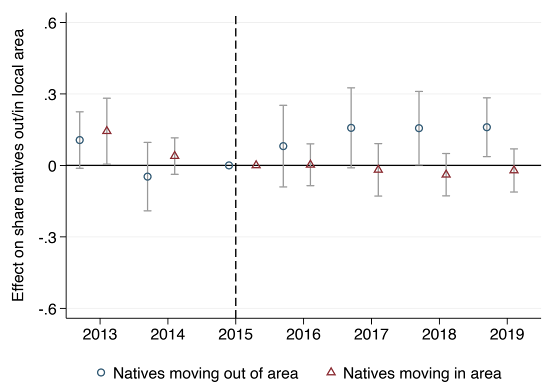
F Robustness Checks

Figure F.1: Trade with Venezuela



Note: The share of exports is constructed as total exports to Venezuela over total exports to the world, the same for imports. Source: DANE.

Figure F.2: Event study estimates on movements across geographical areas

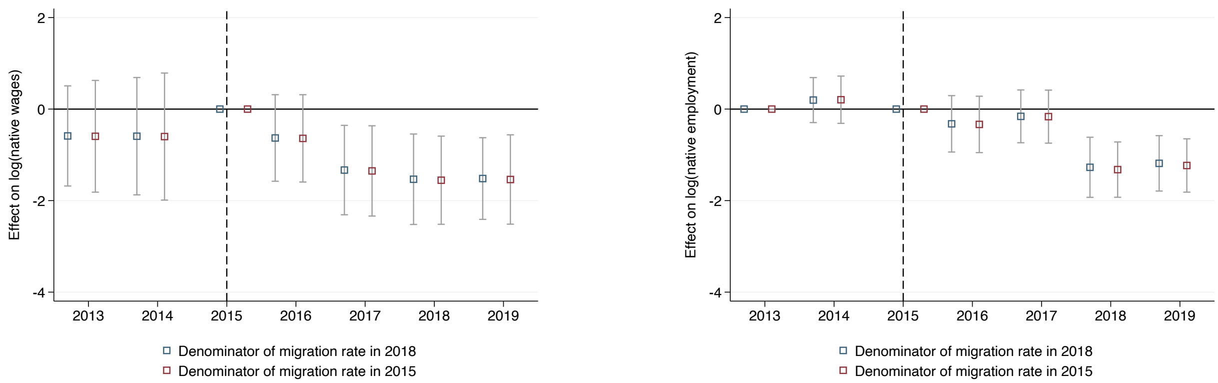


Note: In the first step, I use department survey weights from GEIH to construct department-level outcomes from individual data of natives between 18 and 64 years in urban areas. In the second step, I use the department-level data to estimate β_t from equation (1). The outcome is the inflows or outflows from that department in the last 12 months over the employed population in the base period. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors.

Figure F.3: Event study estimates using different denominators in ΔM_d from Equation (2)

(a) IV (distance) estimates on native wages

(b) IV (distance) estimates on native employment

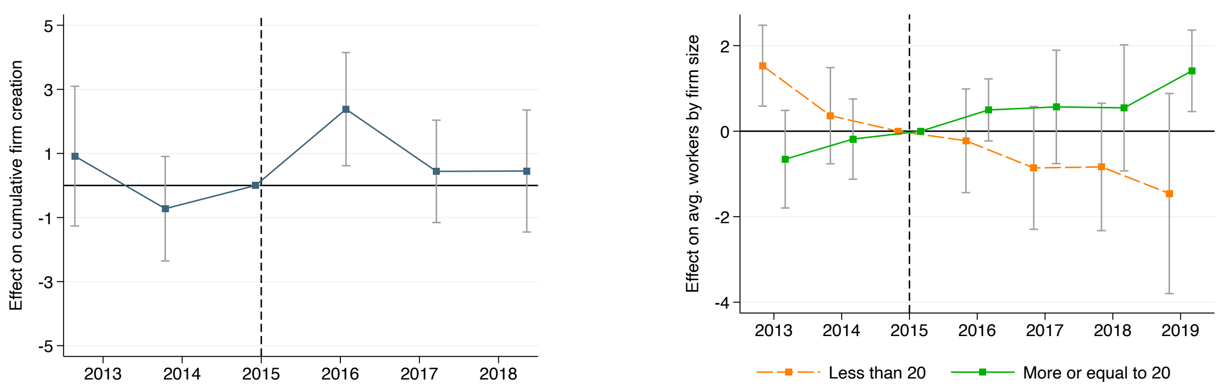


Note: In the first step, I use department survey weights from GEIH to construct department-level outcomes from individual data of natives between 18 and 64 years in urban areas. In the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. I construct the denominator of 2015 with the GEIH survey and the one of 2018 with the census. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

Figure F.4: Event study estimates on cumulative firm creation and average category of workers by firm size

(a) IV (distance) on firm creation

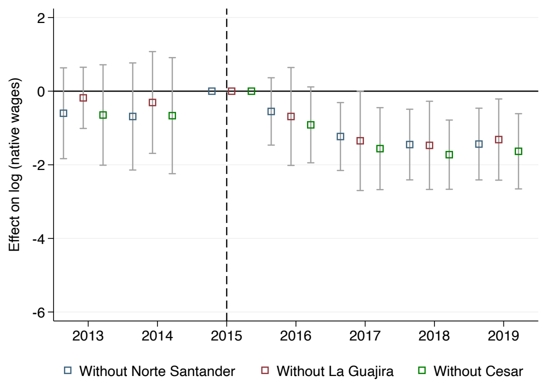
(b) IV (distance) on avg. category of workers



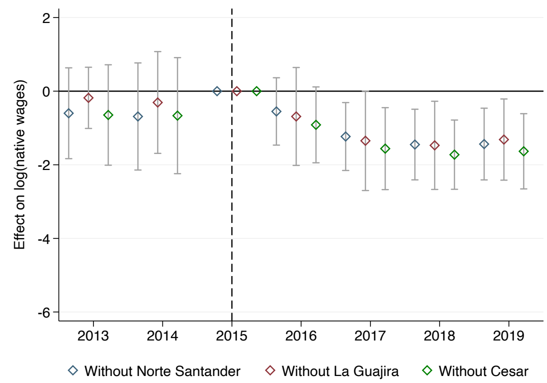
Note: The sample is restricted to registered or formal new firms that are in the databases of the corresponding state agencies. The outcome for panel (a) is the logarithm of new formal firms, and the outcome for panel (b) is the average proxy change in the firm size category. The explanatory variable is $\Delta M_{d,2018}$. I use a 95% confidence interval constructed from standard errors. The base period is 2015. The plotted coefficients are β_t from equation (1). The F-statistic for the distance instrument is 287.1. Source: (a) Confecamaras, 2013–2018. (b) GEIH, 2013–2019.

Figure F.5: Event study estimates excluding border departments for native wages

(a) IV (distance) estimates



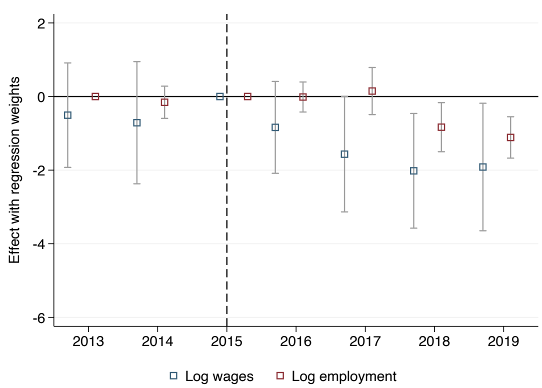
(b) IV (past settlements) estimates



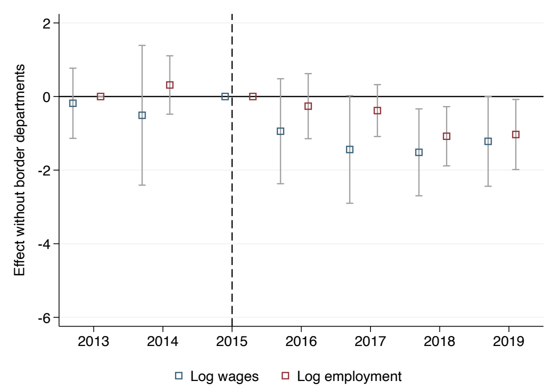
Note: In the first step, I use department survey weights from GEIH to construct department-level outcomes from individual data of natives between 18 and 64 years in urban areas. In the second step, I use the department-level data to estimate β_t from equation (1). Departments in the regression are $N = 23$ per year. I use a 95% confidence interval constructed from standard errors. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

Figure F.6: Event study estimates with additional checks

(a) Including weights in regression



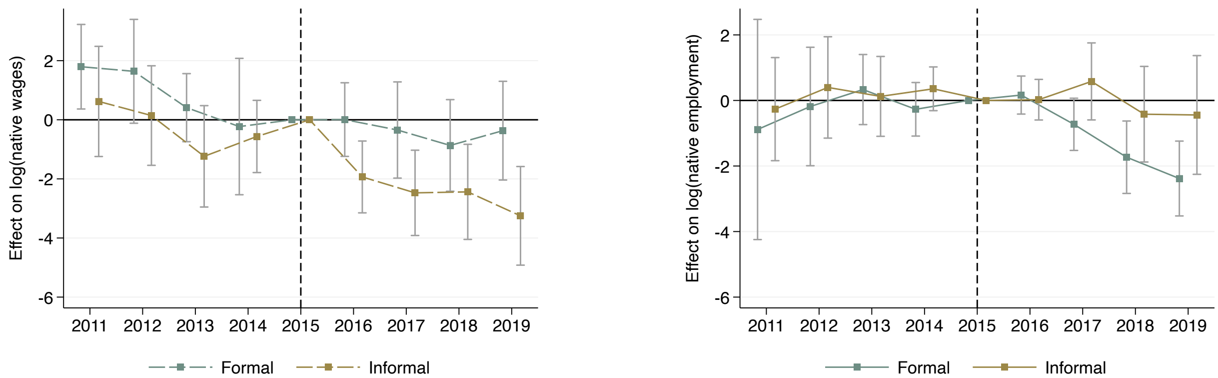
(b) Excluding all border departments



Note: In the 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 the second step, I use the department-level data to estimate β_t from equation (1). In panel (a), Departments in the regression are $N = 24$ per year, and in panel (b) are $N = 21$. In panel (a), I use employment in 2015 as a weight. For employment regressions, I use as a control the growth in employment from 2013 to 2015. I use a 95% confidence interval constructed from standard errors. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

Figure F.7: Event study estimates by sector for natives with survey weights

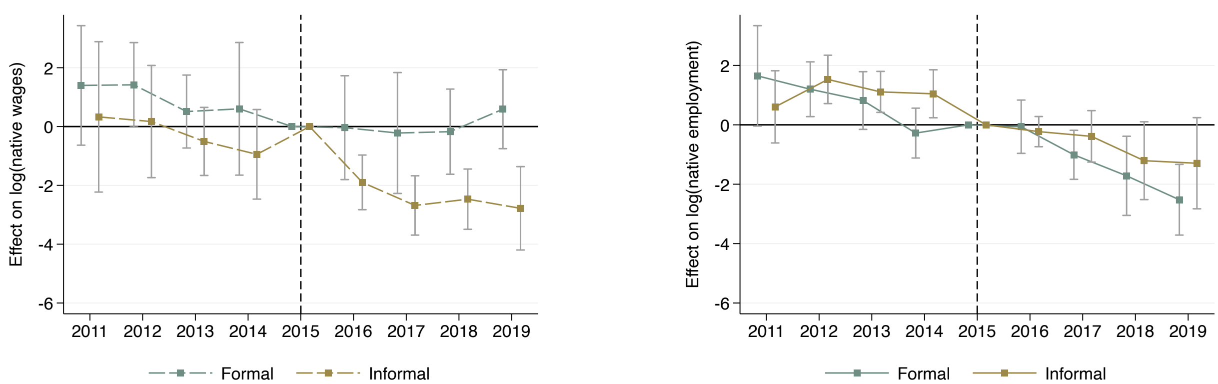
(a) IV (distance) estimates on native wages (b) IV (distance) estimates on native employment



Note: In the 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 the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. In panel (b), I do not use trend-adjustment controls. I use employment in 2015 in each sector as a weight. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

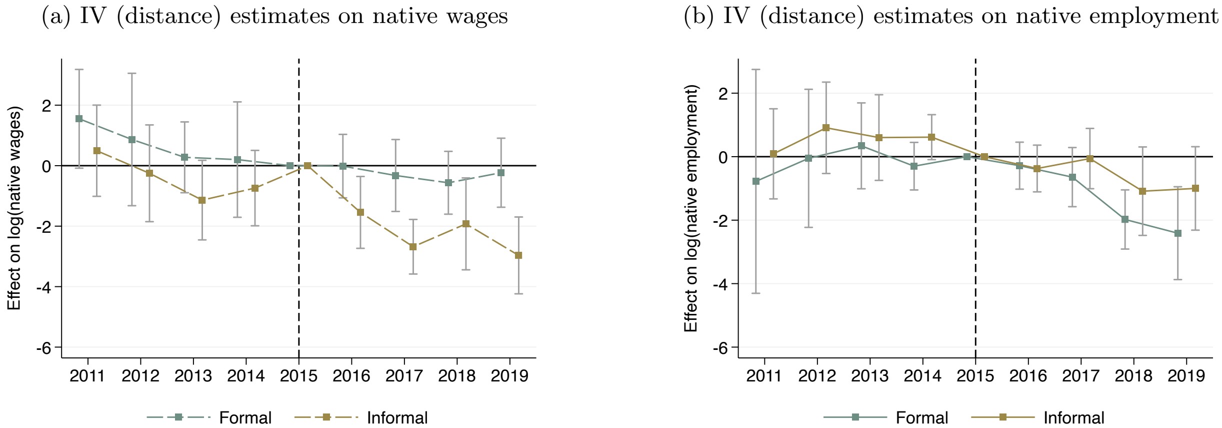
Figure F.8: Event study estimates by sector for natives without border departments

(a) IV (distance) estimates on native wages (b) IV (distance) estimates on native employment



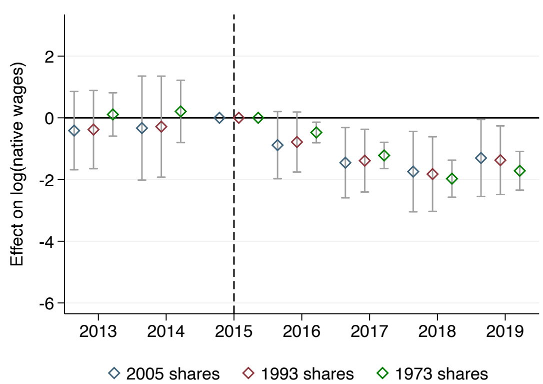
Note: In the first step, I use department survey weights from GEIH to construct department-level outcomes from individual data of natives between 18 and 64 years in urban areas. In the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 21$ per year. I use a 95% confidence interval constructed from standard errors. In panel (b), I do not use trend-adjustment controls. I use employment in 2015 in each sector as a weight. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

Figure F.9: Event study estimates by sector for natives with controls for exports to Venezuela



Note: In the 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 the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. Departments in the regression are $N = 24$ per year. I use a 95% confidence interval constructed from standard errors. In panel (b), I do not use trend-adjustment controls. I use employment in 2015 in each sector as a weight. The F-statistic for the distance instrument is 287.1. The coefficients (in percent) are already multiplied by 100. Hourly wages are in real terms using the monthly CPI from DANE.

Figure F.10: Event study estimates using different historical shares for the past settlements instrument



Note: In the 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 the second step, I use the department-level data to estimate β_t from equation (1) and use clustered standard errors at the department level. The Departments in the regression are $N = 24$ per year for 2005 and 1993 and $N = 22$ for 1973. I use a 95% confidence interval constructed from standard errors. The F-statistic for the past settlements instrument with shares of 2005 is 33.7, with shares of 1993 is 34.39, and with shares of 1973 is 43.02. The coefficients (in percent) are already multiplied by 100. Source: IPUMS for 1993 and 1973 and DANE for 2005.

Table F.1: Robustness checks: wages and employment estimates for natives, 2015–2018

	(1) Wages	(2) Employment	(3) Wages	(4) Employment
Instrument	**Distance**		**Past settlements**	
Control trade with Venezuela 2015	-1.555**	-1.365***	-1.865*	-1.128
	(0.514)	(0.338)	(0.800)	(0.554)
Coefficient of trade	0.041	0.128	0.084	0.092
	(0.325)	(0.128)	(0.365)	(0.167)
Control change GDP 2015–2018	-1.392**	-1.282***	-1.555*	-0.890*
	(0.469)	(0.311)	(0.600)	(0.350)
Coefficient of GDP	-0.251	0.015	-0.228	-0.041
	(0.211)	(0.172)	(0.210)	(0.183)
Control quintiles of distance to Venezuela	-1.450*	-0.987*	-1.960*	-0.525
	(0.645)	(0.418)	(0.859)	(0.563)
Residualized wages	-1.833***		-1.486**	
	(0.434)		(0.453)	
Excluding natives that moved in $t-1$	-1.604**	-1.411***	-1.816**	-1.310***
	(0.444)	(0.293)	(0.584)	(0.301)
Mean of log wages	-1.947***		-1.810**	
	(0.452)		(0.508)	
Trend-adjusted	No	Yes	No	Yes

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

Note: This table reports the coefficients of the second-stage regression 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 variables are in logarithms, so the coefficients are interpreted as percentages and are already multiplied by 100. I use department survey weights from GEIH to construct regional outcomes and then use the department-level data to estimate θ_{2018} from equation (E.1). 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. DANE constructs constant prices of GDP at the department level. 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, sex, and fixed effects of the department, year, and month. When excluding the natives that changed the department of residence, I do not control for pre-trends in employment as the sample restriction already controls the pre-trend.

Table F.2: Native formal employment by MW earners, 2015–2018

	(1) MW earners	(2) No MW earners
$\Delta M_{d,2018}$	-2.736*	-1.337
	(1.240)	(0.697)
N	216	216

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 with the immigration rate $\Delta M_{d,2018}$ using the distance as instrument. 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 variables are in logarithms, so the coefficients are interpreted as percentages and are already multiplied by 100. I use department survey weights from GEIH to construct regional outcomes and then use the department-level data to estimate β_{2018} from equation (1). MW earners are defined as formal workers with wages below 15% of the yearly minimum wage to include the transportation subsidy.

G Testing Caruso, Canon, and Mueller (2021) Wage Estimates

Here, I replicate the wage estimates of Caruso, Canon, and Mueller (2021) to test their sensitivity to changes in periods analyzed or in the empirical specification.[G.1] First, in Table G.1 I replicate similarly the point estimate of Caruso, Canon, and Mueller (2021). For a one pp increase in the immigration rate, I find a –7.4% wage decline.[G.2] Next, I estimate the same regression but add one year of data. After including 2018, the point estimate is less negative (–4.6%), which means a three pp difference in the coefficient relative to the one until 2017. The first conclusion is that as migration drastically increases in Colombia, one additional year of data substantially reduces the coefficient of immigration on wages, partly due to less statistical noise in the measure of the migration shock. Next, I show the differences in the coefficients using the yearly DiD specification with the immigration rate relative to 2015. In this case, the coefficient is substantially less negative (–3.4%), which means a difference of 4 pp. So, when using the DiD specification until 2018, I find a coefficient of –1.5%. Overall, the main difference in wage estimates from Caruso, Canon, and Mueller (2021), and this paper is due to the empirical specification that uses a different measure of the migration shock and the number of periods analyzed.

[G.1] For details, the reader may check their paper and the summary paper of Lebow (2022) that summarizes different wage effects of migration in the Colombian setting.

[G.2] I do not match the –7.6% coefficient as I use a different distance instrument, and the sample restriction is slightly different.

Table G.1: Replication of Caruso, Canon, and Mueller (2021) wage results

	(1) Until 2017	(2) Until 2018	(3) Until 2017	(4) Until 2018
Static TWFE	-7.402***	-4.620***		
	(1.828)	(1.136)		
Yearly DiD			-3.405**	-1.541**
			(1.123)	(0.489)
N	1,443,621	1,720,798	24	24
Clusters	24	24	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 static TWFE, I use the same set of controls and construct the immigration rate as Caruso, Canon, and Mueller (2021). Source: GEIH 2013–2018.

H Solution to the Model in Section 6.1

To derive equations (8) and (9) from the main text, I first combine the profit function with the price function. So, the maximization problem turns out to be:

$$\max_{L_i, L_f} \pi = C^{1-\epsilon} Q^{\epsilon} - \tau(L_i) w_i L_i - (1+\tau_f) w_f L_f \tag{H.1}$$

Then, I solve the maximization problem that implies market wages satisfy

$$(\tau'(L_i)L_i + \tau(L_i))w_i = C^{1-\epsilon}\epsilon\alpha_i L_i^{\rho-1}(Q)^{\epsilon-\rho}, \tag{H.2}$$

$$(1+\tau_f)w_f = C^{1-\epsilon}\epsilon\alpha_f L_f^{\rho-1}(Q)^{\epsilon-\rho}. \tag{H.3}$$

Rearranging market wages (from equation (H.2) and (H.3)) to leave in terms of labor, I get that:

$$L_i^{1-\rho}(\tau'(L_i)L_i + \tau(L_i)) = \left(\frac{C^{1-\epsilon}\epsilon\alpha_i}{w_i}\right)(\alpha_i L_i^{\rho} + \alpha_f L_f^{\rho})^{\frac{\epsilon-\rho}{\rho}} \tag{H.4}$$

$$L_f^{1-\rho} = \left(\frac{C^{1-\epsilon}\epsilon\alpha_f}{w_f(1+\tau_f)}\right)(\alpha_i L_i^{\rho} + \alpha_f L_f^{\rho})^{\frac{\epsilon-\rho}{\rho}} \tag{H.5}$$

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

$$logL_i = \frac{1}{1+\eta-\rho}((1-\epsilon)logC + log\epsilon\alpha_i - logw_i - log(1+\eta)) + \frac{\epsilon-\rho}{\rho(1+\eta-\rho)}log(\alpha_i L_i^{\rho} + \alpha_f L_f^{\rho}) \tag{H.6}$$

$$logL_f = \frac{1}{1-\rho}((1-\epsilon)logC + log\epsilon\alpha_f - logw_f - log(1+\tau_f)) + \frac{\epsilon-\rho}{\rho(1-\rho)}log(\alpha_i L_i^{\rho} + \alpha_f L_f^{\rho}) \quad (H.7)$$

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:

$$\frac{dlogL_i}{dw_i}|_{dw_f=0} * w_i = -\frac{1}{w_i(1+\eta-\rho)}w_i + \frac{\epsilon-\rho}{\rho(1+\eta-\rho)}\left(\frac{\rho(\alpha_f L_f^{\rho-1}\frac{dL_f}{dw_i} + \alpha_i L_i^{\rho-1}\frac{dL_i}{dw_i})}{\alpha_f L_f^{\rho} + \alpha_i L_i^{\rho}}\right)w_i \quad (H.8)$$

Simplifying, I get that:

$$\frac{dlogL_i}{dw_i}|_{dw_f=0} * w_i = -\frac{1}{1+\eta-\rho} + \frac{\epsilon-\rho}{1+\eta-\rho}(\alpha_f(L_f/Q)^{\rho}\varepsilon_{L_f,w_i} + \alpha_i(L_i/Q)^{\rho}\varepsilon_{L_i,w_i}) \quad (H.9)$$

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, I rewrite the last expression as:

$$\varepsilon_{L_i,w_i} = -\frac{1}{1+\eta-\rho} + \frac{\epsilon-\rho}{1+\eta-\rho}(s_f\epsilon_{L_f,w_i} + s_i\epsilon_{L_i,w_i}) \quad (H.10)$$

where $s_f = \alpha_f(L_f/Q)^{\rho}$ and $s_i = \alpha_i(L_i/Q)^{\rho}$ are the formal and informal labor shares in production and $s_f + s_i = 1$.

Then, when I differentiate formal labor with respect to informal wages, I get the following:

$$\varepsilon_{L_f,w_i} = \frac{\epsilon-\rho}{1-\rho}(s_f\epsilon_{L_f,w_i} + s_i\epsilon_{L_i,w_i}) \quad (H.11)$$

Combining the two last expressions, I find that:

$$\varepsilon_{L_i,w_i} = -\frac{1}{1+\eta-\rho} + \frac{1-\rho}{1+\eta-\rho}\varepsilon_{L_f,w_i} \quad (H.12)$$

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 (8) and (9) in the main text.

I Definition of Variables

Hourly real wages. The *inglabo* variable from the GEIH survey captures basic pay, pay in-kind, and income of the second activity. First, I subtract the income of the second activity and add allowances for food and transportation (according to ILO definition of wages).[I.1] Is worth noting that this variable captures any 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.[I.2] The base of the index (=100) is in December of 2018,

[I.1] https://www.ilo.org/wcmsp5/groups/public/---africa/---ro-abidjan/---ilo-pretoria/documents/publication/wcms_413782.pdf

[I.2] Information was taken from here https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc

$$RealWage_{imy} = \frac{NominalWage_{imy}}{CPI_{my}} * 100 \tag{I.1}$$

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 divided by the number of working hours that the respondent reported usually at this job in the week. In the next step, I only consider positive wages of full-time workers (with more than 30 working hours per week) 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 old in urban areas who reported working 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 the logarithms of that expression.

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