Worker Responses to Immigration Across Firms: Evidence from Colombia*

Lukas Delgado-Prieto†

October, 2024

[Click here for the latest version]

Abstract

Millions of migrants have recently relocated to developing countries where the informal sector is large and small firms are prevalent in the labor market. The interaction between workers, firms, and immigration shocks in these contexts remains largely unexplored. In this paper, I study the mass arrival of migrants from Venezuela in Colombia using comprehensive administrative records that cover the universe of formal workers and firms. As immigrants concentrate in the informal sector, formal employment for natives earning the minimum wage declines, reflecting their high substitutability with informal workers who become less costly. Across firms, the negative formal employment and wage effects are concentrated in small firms. To rationalize this, I construct a model of heterogeneous firms that employ both formal and informal labor to show that small firms' responses are more pronounced as they rely more on informal labor for production. Finally, I develop causal forests to find that the heterogeneity in employment and wage effects is explained more by firm characteristics rather than worker characteristics. These results indicate that the firm dimension is key for a complete understanding of the impact of immigration on native workers and the labor market overall.

Keywords: Immigration, Formal labor markets, Causal forest.

JEL Codes: F22, O15, O17, R23.

*I am extremely grateful to Jan Stuhler for his guidance and support in this project. I am also thankful to Michael Amior, Jaime Arellano-Bover, Esther Arenas-Arroyo, Agostina Brinatti, David Card, Juanjo Dolado, Juan Carlos Escanciano, Jesús Fernández-Huertas, Albrecht Glitz, Patrick Kline, Luigi Minale, Jesse Rothstein, and Jon Piqueras for their comments and suggestions. In general, I want to thank participants at the Lisbon Migration Workshop, the 13th Conference on Immigration in OECD, the 2023 EALE Conference, the 2023 ENTER Jamboree, the 1st HUMANS LACEA, the Labor Lunch at UC Berkeley, the IRLE Seminar, the UC3M Applied Reading Group, the 23rd IZA Summer School in Labor Economics, and the Junior Seminar of the Economics of Migration for the helpful discussions. I am indebted to the Ministry of Health and Social Protection of Colombia for providing access to their data. This paper was awarded the Best Immigration Economics Paper at the OECD Immigration Conference 2023. This paper was previously circulated as: "Immigration and Worker Responses Across Firms: Evidence from Administrative Records in Colombia" and "The Role of Workers and Firms in the Impact of Immigration".

†Department of Economics, University of Oslo, Norway. Email: laprieto@econ.uio.no.

1 Introduction

Over the past decade, several countries worldwide have experienced substantial population exodus.[1] The majority of these migrants and refugees are relocating to neighboring destinations, which are often developing countries.[2] Compared to developed economies, the labor market in these contexts is characterized by the interplay between the formal and informal sectors (Ulyssea, 2018), as well as the prevalence of small firms (McKenzie, 2017). Given that migrants are disproportionally concentrated in small firms (Delgado-Prieto, 2024), it becomes relevant to examine the labor market impacts of immigration on workers from the firm side. However, the existing migration literature often overlooks the role of firms when analyzing the effects of immigration, partly due to data constraints. This paper shows that firm heterogeneity is an important dimension to analyze in order to fully understand how native workers and the labor market overall adjust to immigration shocks.

To do so, I study the labor market impacts of one of the most significant episodes of immigration in recent history, the Venezuelan mass migration to Colombia, using comprehensive administrative data that covers the universe of formal workers and firms in the country.[3] By exploiting the uneven arrival of migrants across local labor markets with the panel structure of the data, I quantify impacts at the worker level across various worker and firm characteristics.[4] While an increasing number of studies analyze how firms shape, for instance, wage inequality (Card et al., 2013) or immigrant assimilation (Arellano-Bover and San, 2024), less is known about how firms shape natives' adjustments to immigration shocks.

To my knowledge, this is one of the first papers to study the impact of immigration in developing countries equipped with matched employee-employer data.[5] This reveals important sources of heterogeneity not previously explored, as prior research primarily focused on the effects across

[1] For example, Afghanistan, Ukraine, Syria, and Venezuela, among others.

[2] As of June 2023, 75% of the 110 million forcibly displaced individuals were hosted by low and middle-income countries (UNHCR, 2023).

[3] Throughout this paper, formal workers refer to workers who contribute monthly to the health system in Colombia.

[4] By tracking workers over time, I can address more carefully the compositional changes in the employed population following the arrival of immigrants, typically omitted using regional outcomes. For instance, recent papers emphasize that when a specific set of workers move out of employment or to other regions, the wage estimates are not correctly identified (Borjas and Edo, 2021; Dustmann et al., 2023).

[5] In developed countries, Bratsberg and Raaum (2012) for Norway, Foged and Peri (2016) for Denmark, Dustmann et al. (2017) for Germany and Orefice and Peri (2024) for France have recently analyzed immigration shocks exploiting administrative data.

worker or industry characteristics. Although worker and firm characteristics are interrelated (e.g., minimum wage workers often work in smaller firms), I find that most of the heterogeneity in migration impacts in this setting stems from the firm side, even after controlling for the firm industry. The strong interaction between the formal and informal sectors in small, less productive firms provides insights into this heterogeneity. The immigration shock, concentrated in the informal sector, drives down informal wages (Delgado-Prieto, 2024), which in turn reduces the labor demand in the formal sector, particularly in small firms that can substitute formal labor more easily for informal labor.[6] The labor-labor substitution is a crucial response to immigration that may extend to other economic shocks and policy changes, such as increases in the minimum wage (Clemens, 2021). Importantly, this substitution implies that the employment and wage effects vary massively across the firm size distribution, motivating the worker-level analysis across firms done in this paper.

My empirical strategy compares similar formal workers in regions with varying exposure to migration over time. To address the endogenous sorting of migrants into economically favorable areas, I construct two distinct instruments: past settlements of Venezuelans and distance to the nearest crossing bridge with Venezuela. Using these instruments in a differences-in-differences research design (DiD-IV), I find a persistent negative impact on individual formal employment and wages of natives.

The negative employment effects are concentrated in small formal firms and are driven by workers earning the minimum wage before the immigration shock. For these low-wage workers, a one percentage point (pp) increase in the share of employed migrants in a given labor market decreases the probability of formal sector employment by 1.5 pp. The relatively high and binding minimum wage limits the space for downward wage adjustments (approximately 40% of all formal workers in Colombia earned the minimum wage in 2015) and increases their chances of job displacement.[7] Regarding wages, the negative impact primarily affects native workers earning above the minimum wage and those employed in the smallest formal firms. The concentration of migrants in small firms, which face binding constraints from the minimum wage and employ a higher share of informal workers, helps to explain these findings.

[6] This application aligns with the first Hicks-Marshall rules of Derived Demand: “The demand for anything is likely to be more elastic, the more readily substitutes for the thing can be obtained” (Hicks, 1932).

[7] The 40% share is after applying the main sample restriction, which includes self-employed workers.

To rationalize how firms interact with immigration effects, I construct an imperfectly competitive labor market model based on Card et al. (2018) with heterogeneous firms, but incorporating the labor inputs cost similarly to Ulyssea (2018) and allowing for imperfect substitution of labor inputs as in Delgado-Prieto (2024). The model indicates that the aggregate substitutability between formal and informal workers must be high to observe a negative formal employment and wage response. It additionally reveals that firm-level responses to immigration depend on their initial share of informal labor in production, predicting that smaller firms have higher elasticities of formal employment and formal wages with respect to informal labor, in absolute terms, relative to larger firms. This prediction aligns with the empirical finding that workers in smaller firms experience more adverse impacts on formal employment and wages than those in larger firms, which ultimately suggests a higher substitutability of formal and informal labor in small firms.

Next, I exploit the canonical Abowd et al. (1999) (AKM hereafter) framework to estimate firm fixed effects (FEs), representing firm-specific pay premiums and worker FEs as a proxy of workers' skills. A key contribution of this paper is understanding the sources of wage and employment losses stemming from immigration using these constructed variables. I find that workers from middle-paying firms during the pre-shock period experience the largest wage losses, whereas workers from low-paying firms, bounded by the minimum wage, are less affected. A potential explanation for the observed wage effects is the reallocation effects of immigration, where workers may be moving from high- to low-paying firms or vice-versa (Gyetvay and Keita, 2023).[8] However, I find no differential sorting of native workers post-immigration shock.[9] Regarding employment, native workers in low-paying firms are more adversely affected than those in high-paying firms, similarly to when dividing between low- and high-wage workers.

Next, I focus on certain firms' outcomes that illustrate other margins of adjustment to immigration. First, I show that firms opt-out of the formal sector for new hires, especially those that poach fewer workers from other formal firms. Second, the rate of firms exiting from the formal sector is higher in places with more migrant inflows, whereas the formal firm creation rate does not change. Altogether, these findings suggest the need for policies to support affected native workers.

[8]Another explanation, according to Amior and Stuhler (2022), is that when the share of firms in the low-pay sector grows due to immigrants, firms in the high-pay sector exert greater monopsony power, thereby reducing workers' wages.

[9]Furthermore, there is no evidence that formal workers are relocating to other regions following the arrival of immigrants.

Given Colombia's relatively high minimum wage, reducing other labor costs for formal employers can help to maintain formal employment. Simultaneously, stricter enforcement of fines for informal worker hiring can deter the substitution between formal and informal workers.

In the second part of this paper, I leverage the longitudinal aspects of the matched employee-employer dataset to systematically estimate immigration effects using recent machine learning methods (Athey and Imbens, 2016; Athey et al., 2019). I implement causal forests to obtain reduced-form estimates from random subsamples, identifying the worker and firm variables that explain most heterogeneity in immigration effects. From this algorithm, first, I identify the subgroups most affected by immigration in terms of both employment and wages. Then, based on the frequency of these variables in the splits of all the decision trees in the causal forests, I construct a variable importance statistic. With this measure, I show that firm pay premiums or firm size are consistently ranked higher, indicating they explain more the heterogeneity in employment and wage effects than worker characteristics, such as job tenure, age, sex, or wages in the pre-shock period. In summary, both regression analysis and the causal forests highlight the relevant role of firms in the adjustment of workers to immigration shocks, which is the main contribution of this paper.

Related Literature. This paper contributes to different strands of the labor economics literature. First, it adds to the literature that analyzes how firms shape native and migrant outcomes. Several papers emphasize that firms influence workers' outcomes through different channels. For example, Beerli et al. (2021) exploits the abolition of immigration restrictions, showing that this led to the growth of Swiss firms, thereby improving the wages and job opportunities of highly educated natives. Doran et al. (2022), using the H-1B visa lottery, shows that winning firms crowd out their existing workers with H-1B visa holders. Additionally, there is evidence that firms attract natives from lower- to higher-paying firms post-immigration (Orefice and Peri, 2024; Gyetvay and Keita, 2023).[10] In this respect, my findings show that immigration effects are concentrated on natives working in small and low-paying firms, similar to Amior and Stuhler (2022), though the mechanisms differ substantially. In my paper, this implication stems from the developing context and how small firms can substitute formal for low-priced informal labor after immigrants arrive. In contrast, Amior and Stuhler (2022) argues small and low-paying firms in Germany tend to hire

[10] Regarding migrant outcomes, Arellano-Bover and San (2024) and Dostie et al. (2021) find that firm pay premiums explain around one-fifth of the immigrant-native wage gap in Israel and Canada, respectively.

most migrants due to their lower reservation wages, crowding out more natives in these firms as they exert their monopsony power.

Next, I contribute to the literature that estimates the individual impacts of immigration. Early studies like Bratsberg and Raaum (2012) and Foged and Peri (2016) used licensing requirements in the Norwegian construction sector and refugee dispersal policies in Denmark, respectively, to estimate worker-level effects of immigration.[11] By incorporating all movements of natives between areas and excluding employment inflows, my analysis reduces the attenuation of wage estimates highlighted by Borjas (2006) and identifies the individual effects of immigration (Dustmann et al., 2023). This allows me to understand the main drivers behind labor market adjustments to immigration in two novel ways: by studying immigration with firm heterogeneity and by using machine learning methods to estimate heterogeneous effects in a data-driven manner.

I also extend to the literature on how workers react to other labor market shocks, such as import competition (Autor et al., 2014), local employment shocks (Yagan, 2019; Redondo, 2022), and mass layoffs (Gulyas et al., 2019). My findings show that firms play a significant role in determining the wage and employment impacts of immigration shocks, both theoretically and empirically. This important result complements the labor literature by shifting the focus from worker or industry characteristics to firm heterogeneity, aligning with Gulyas et al. (2019) who find higher earning losses among workers in high-paying firms post-displacement using causal forests.[12]

Lastly, this paper contributes to the literature on the impact of immigration in developing countries (see related papers of Bonilla-Mejía et al. (2024), Caruso et al. (2021), Delgado-Prieto (2024), and Lebow (2024) for Colombia; Del Carpio and Wagner (2015), Ceritoglu et al. (2017), and Aksu et al. (2022) for Turkey; and Groeger et al. (2024) for Peru). Unlike most studies that use cross-sectional surveys, I use panel administrative data to quantify the individual employment and wage effects of immigration for the first time in a developing country, addressing the conceptual differences between regional and individual labor market responses highlighted by Dustmann et al. (2023). This helps to explain why my findings may differ from previous studies that analyze the impact of immigration in the Colombian or Turkish setting.

[11] More recent papers that quantify worker-level effects of immigration include Hoen (2020) for Norway, Ortega and Verdugo (2022) for France, and Kuosmanen and Meriläinen (2022) for Finland.

[12] Yakymovych et al. (2022) uses similar methods with Swedish administrative data to identify workers most vulnerable to job displacement, finding that older, less-educated, and manufacturing workers experience the most significant earning losses.

The remainder of the paper is structured as follows. Section 2 discusses the Venezuelan crisis-induced immigration shock and describes the datasets. Section 3 outlines the empirical strategy and identification assumptions. Section 4 presents results at the worker level by worker characteristics. Section 5 develops a model with heterogeneous firms and shows results by firm characteristics. Section 6 introduces the machine learning approach and discusses the main findings. Section 7 provides robustness tests. Finally, Section 8 concludes.

2 Institutional Context and Data

2.1 The Venezuelan Mass Migration

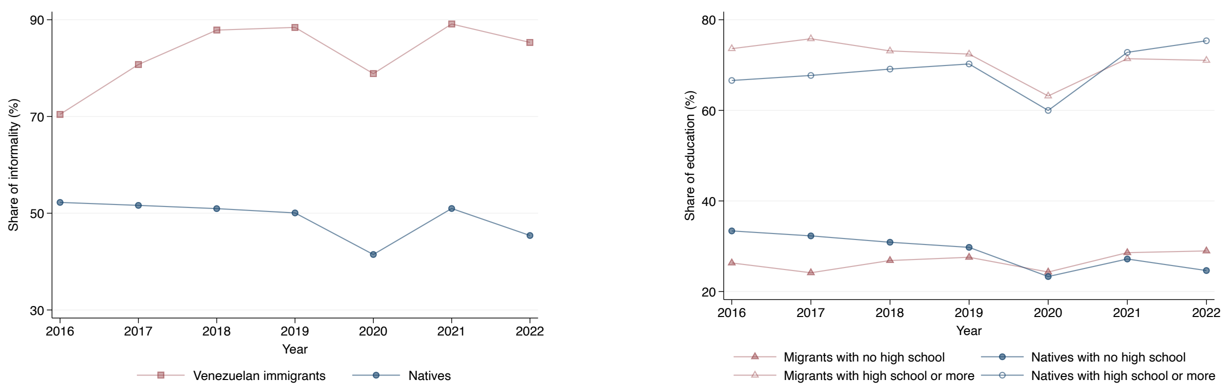
Colombia and Venezuela share an extensive territorial border historically characterized by a dynamic relationship with frequent economic and cultural interactions. People often moved back and forth between the two countries, but generally, Colombians settled in Venezuela. This trend intensified after 1950, fueled by the oil boom in Venezuela and the internal conflict in Colombia. The economic opportunities presented by Venezuela's oil industry attracted many Colombians to emigrate, seeking better livelihoods and prospects for their families. Recently, the trend reversed with Venezuela's unprecedented socio-economic and political deterioration that triggered massive outflows of people since 2013, both voluntarily and forcedly. As a result, several countries in Latin America are receiving vast numbers of migrants, especially Colombia, Perú, and Ecuador (UNHCR, 2019). By far, Colombia has been the major receiver country with more than 1 million working-age Venezuelans, or 4.1% of the working-age population living in Colombia, as of 2019 (DANE, 2019).

The Venezuelan exodus is unmatched in the recent migration history in Latin America. Worldwide, there are only two contexts with similar numbers, namely, the Syrian and the Ukrainian exodus. In the first case, Turkey has been the major receiver country of Syrians, with various papers analyzing this immigration shock (e.g., Del Carpio and Wagner (2015); Ceritoglu et al. (2017); Aksu et al. (2022)). However, the Colombian context is different from the Turkish one. First, Venezuelans speak the same language as Colombians and second, Colombia's government has implemented an open border policy in which all Venezuelan immigrants can have a work permit. In particular, after 2017, all irregular Venezuelans in Colombia have access to the Special Permit of Permanence (PEP, by its Spanish acronym). This allows them to work for a specific period, provides access to essential services, and facilitates their integration into Colombian society.[13] So, this is a setting with voluntary migrants and forcibly displaced refugees with access to the formal labor market. Still, Figure 1a shows that around 90% of Venezuelan immigrants were employed in the informal sector in 2019 and, according to Delgado-Prieto (2024), most of them were concentrated at the bottom of the native wage distribution. This fact relates to the occupational downgrading of Venezuelans since they have similar average levels of education compared to their Colombian counterparts and are even more educated in my period of analysis, from 2016 to 2019 (see Figure 1b).

Figure 1: **Descriptive statistics of immigrants and natives**

(a) Informality

(b) Education



Note: I define informality based on the contributions to the social security system and define the share over the employed population. For education, the share is over all the population. I aggregate information with national survey weights. Source: GEIH, 2016–2022.

As described above, the immigration shock in Colombia occurs mainly in the informal sector, so why focus on the formal sector in this paper? First, Delgado-Prieto (2024) shows that there is a strong negative effect on the wages of the informal sector following the arrival of migrants, and as firms can combine formal and informal labor in production (especially the smallest firms), they will substitute formal for informal employment in response to lower informal wages when both types of labor have a high substitutability. So, formal employment is the most affected in response to the immigration shock, even if immigrants primarily work in the informal sector.[14] For this

[13] To overcome the limitations of the PEP, the government enacted in 2021 a Temporary Protection Statute for Migrants (ETPV, by its acronym in Spanish) that grants up to ten years of regularization for Venezuelan immigrants.

[14] The asymmetric responses across the informal and formal sectors in the face of migration shocks have been

reason, focusing on formal workers' adjustments across the firm size or productivity distribution is crucial for capturing all the responses that could not be addressed with survey data in Delgado-Prieto (2024). I then explore these heterogeneous effects using recent machine-learning methods, providing a better understanding of the results.

2.2 Data

The main dataset I use in this paper is the *Planilla Integrada de Liquidación de Aportes* (PILA), which contains administrative records from the Colombian social security system managed by the Health Ministry (*Ministerio de Salud y Protección Social*). PILA contains information on the universe of formal workers in tax-registered firms. It excludes informal workers and informal firms but includes self-employed formal workers. The PILA is based on the monthly contribution of the worker, according to their reported base income, to the health system in Colombia. Each observation in PILA is a worker-employer match for a given year and month. The dataset contains worker-level information on labor income, sex, age, job type (employee or self-employed), foreign status, municipality, and the firm identifier for each job. In the main analysis, I focus on the years between 2012 and 2019 for August.[15]

The other main dataset I use in this paper is the 2018 Colombian census (CNPV, by its acronym in Spanish). As Colombian authorities recognize the significance of the Venezuelan exodus, they included migratory questions in the most recent census, like the year of the arrival of all migrants, which I exploit to construct the immigration shock. The census provides the most reliable source of information on the local stock of migrants in the country.[16]

For the analysis, I constructed a dataset encompassing all individuals recorded in PILA between 2012 and 2019, with rows representing individuals and columns their yearly variables. The dataset includes 18,430,987 workers who appeared in at least one of these eight years. I then restricted to full-time native workers between 25 and 55 years old as of 2015, assigning the immigration shock to all these workers based on their location in 2015. This restriction narrowed the sample down to

documented in other contexts too (see Corbi et al. (2021) for Brazil and Kleemans and Magruder (2018) for Indonesia).

[15] I choose August to exclude the seasonal characteristics of other months (e.g., December-January or March-April) and because the census recollection ended in October of 2018, omitting arrivals of migrants in November and December of that year.

[16] The labor force survey (GEIH, by its acronym in Spanish) also measures the number of Venezuelan immigrants in Colombia at a higher frequency but not at the local level I exploit. Furthermore, Aydemir and Borjas (2011) document that surveys can attenuate immigration effects due to measurement error of migrants.

7,123,223 workers.[17]

Subsequently, I transformed the municipality variable into a more standard definition of local labor markets or commuting zones by adjusting the methodology of Sanchez-Serra (2016).[18] This adjustment resulted in 109 functional urban areas (FUAs) after excluding small or rural municipalities, producing a sample of 6,706,035 workers.[19] This is the sample I use for the employment analysis over time (a balanced panel) as the worker can be employed or not in the comparison year. For the wage analysis, I further restricted the sample to workers with 30 days of employment per month with positive wages, requiring that the worker be employed in the post-treatment year of comparison. Thus, the sample varies slightly by year, making it an unbalanced panel. It is important to note that these restrictions are common in the literature (Gulyas et al., 2019; Yagan, 2019).

Descriptive Statistics for Formal Workers. Table 1 shows descriptive statistics for natives, foreigners, and Venezuelans with PEP broken down by age, sex, and wages over time.[20] In terms of observable characteristics, Venezuelans with PEP in the formal sector tend to be younger, predominantly male, and earn lower wages compared to both natives and other foreigners (see Table 1). Additionally, other foreigners also earn substantially higher salaries than natives. The share of Venezuelans with the PEP working in the formal sector is small, supporting the finding that the impact of the PEP regularization on the Colombian labor market is limited (Bahar et al., 2021). It is important to note that informal workers, who account for about half of all employed individuals in Colombia, are not observable in the administrative data.

[17]Selecting workers observed in the base period excludes inflows of workers in the post-treatment period from the analysis. Also, part-time workers constituted less than 0.3% of the PILA workforce in 2015.

[18]A drawback of the municipality variable in the PILA is that certain firms with several establishments across the country report data for all employed workers in the municipality where their largest establishment is located, understating employment in smaller cities.

[19]The FUAs definition consists of the 53 most extensive urban areas in the country defined from population grid data, municipal boundaries, inter-municipal commuting flows, as well as 56 municipalities with more than 2,500 formal workers according to the restricted 2015 sample. I exclude San Andrés, Cumaribo, Leticia, and San José del Guaviare from this definition as they belong to islands or the Amazon region. Appendix Table F.1 shows the sample distribution by FUAs, with only 5.9% of workers excluded from the analysis.

[20]To identify foreign status in PILA, I use the type of document listed in the workers' health contribution. Workers with a national ID are classified as natives, those with the PEP document are classified as Venezuelan regularized migrants, and those with a foreigner ID or passport are classified as foreigners. Note that the PEP program began around 2018 to facilitate the regularization of Venezuelan migrants. Thus, it was not possible to identify these migrants before that period.

Table 1: **Descriptive statistics for natives and immigrants in the formal sector**

(a) Colombians

	Age		Male (%)		Real wages (USD)		
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	N
2013	37.0	10.8	0.56	0.496	421.4	504.0	7,335,989
2015	37.2	11.1	0.56	0.497	416.8	484.5	8,391,843
2017	37.8	11.4	0.55	0.497	411.4	477.8	8,064,282
2019	38.2	11.7	0.55	0.498	436.1	505.7	8,363,249

(b) Venezuelans with PEP

	Age		Male (%)		Real wages (USD)		
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	N
2018	30.8	7.8	0.68	0.467	243.7	99.0	12,842
2019	31.8	8.1	0.67	0.472	248.7	98.5	42,752

(c) Other foreigners

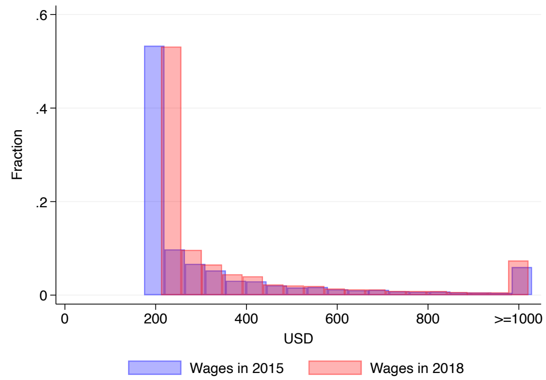
	Age		Male (%)		Real wages (USD)		
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	N
2013	40.2	10	0.61	0.487	1,310.0	1,535.9	20,978
2015	39.5	10.2	0.64	0.481	1,253.4	1,446.6	27,730
2017	39.3	10.3	0.63	0.483	1,050.8	1,325.0	31,553
2019	39.6	10.4	0.58	0.494	990.5	1,293.0	39,704

Note: This table reports the descriptive statistics for Colombians, foreigners, and Venezuelans with PEP between 18 and 64 years of age. Only workers with full days of employment recorded in PILA and a positive health contribution are considered for wages and the number of observations. I only observe Venezuelans with PEP since 2018. The real wages are deflated using the Consumer Price Index (CPI) from DANE for prices in 2018. Colombian pesos to USD using 2020 exchange rates from the World Bank. For self-employed workers, observed wages in PILA correspond to 40% or more of their actual wages by law, with the minimum wage as a lower bound. Source: PILA, 2013–2019, August.

Next, Figure 2 illustrates how binding the Colombian minimum wage is for a large share of formal workers. In 2015, around 40% of all formal workers earned the minimum wage.[21] Importantly, the national minimum wage must increase, by law, more than the inflation in the preceding year. This downward rigidity suggests why minimum wage workers generally do not experience real wage drops during positive labor supply shocks or negative demand shocks, and instead, job displacement becomes more likely. Last, in the period of analysis (2015-2019), the minimum wage increased in real terms by less than 3% each year, mitigating the concerns of additional impacts of minimum wage hikes on employment.

[21] To contribute to the health and pension system in Colombia, workers must declare a labor income equal or greater than the minimum wage, so many self-employed workers (who can decide how much is their basic income in PILA) report the minimum wage even if they earn more or less.

Figure 2: **Histogram of wages by years**



Note: The sample is restricted to native workers between 18 and 64 years with full employment days in the month and positive wages. Wages are in nominal terms. Colombian pesos to USD using 2020 exchange rates from the World Bank. The chosen bin width is 45. Source: PILA, 2015–2018.

Descriptive Statistics for Formal Firms. To describe patterns in the workforce composition of formal firms, I aggregate the worker information of the PILA at the firm level after excluding self-employed workers. I divide Table 2 into seven firm size categories to highlight relevant facts. First, male workers are predominant across all formal firm sizes, especially in small to medium-sized firms (between 10 and 999 workers), where over 60% of workers are male. Second, smaller firms employ older workers on average (39.6 years), while larger firms employ younger workers (36 years). Third, average wages increase with firm size, ranging from approximately 272 USD in firms with 1 to 4 workers to 531 USD in firms with more than 1,000 workers. Finally, Appendix Figure B.1 illustrates a histogram of firms by size, overlaid with the total number of employees in each firm size. Although the majority of firms fall into the 1 to 4 employee range, the distribution of employees is more evenly spread across different sizes of firms.

Table 2: **Descriptive statistics by firm size**

	Average				
Firm size (# of workers)	Employment	Male (%)	Age	Real wages	Firms
1-4	2	0.56	39.6	271.9	206,456
5-9	7	0.60	37.5	304.2	64,347
10-19	13	0.61	37.2	329.2	42,207
20-49	30	0.63	36.9	360.9	28,625
50-99	69	0.65	36.7	394.8	10,032
100-999	259	0.63	36.9	443.1	10,107
1000 and more	2677	0.58	36.0	530.8	859

Note: This Table reports the descriptive statistics for seven groups of firm size. I deflate real wages using the CPI from DANE for prices in 2018. Then, I transform Colombian pesos to USD using 2020 exchange rates from the World Bank. I only consider employees when constructing firm sizes. Source: PILA, 2015-August.

3 Empirical Strategy

To quantify the evolving impact of immigration on worker outcomes, I estimate the following differences-in-differences specification from $t = \{2012, .., 2019\}$ estimating separate yearly regressions of the following form:

$$\Delta Y_{i,l,t} = \delta_t + \theta_t \Delta M_{l,2018} + X_i' \beta_t + \Delta u_{i,l,t}. \tag{1}$$

Here, the outcome $\Delta Y_{i,l,t}$ is the difference in wages, employment, or earnings within workers' pre- and post-treatment years relative to 2015. The immigration shock $\Delta M_{l,2018}$ varies for each local labor market l, and the vector X_i contains individual characteristics in 2015; namely, dummies for six age groups interacted with dummies for sex and self-employment.[22] Broadly, this specification compares individuals with similar observables in the base period but who were working in local labor markets with different exposure to the immigration shock, which I will describe below in detail. Hence, as I'm following worker outcomes over time, θ_t measures the worker-level response to migration, where $\theta_{2015} = 0$ by construction. By taking differences, I net out any individual constant unobservables that may confound the impact of migration. Lastly, the intercept for each year is δ_t, and I cluster the standard errors in all the specifications at the level of the treatment, which are the FUAs (defined as G and equal to $G = 109$).

[22]Education information is not available, and I decide not to use industry information as a baseline control due to significant measurement errors, as the data processor has not verified the industry codes in PILA up to 2019.

The individual outcomes are more precisely defined as follows. First, the employment outcome is $e_{i,l,t} - \sum_{k=2013}^{2015} e_{i,l,k}/3$, where $e_{i,l,t}$ is the indicator of formal sector employment for worker i in local labor market l in period t. As in Yagan (2019), I consider the average employment in the pre-shock period to transparently allow for varying labor trajectories of workers in the formal sector. In the event study figures, however, I take the simple difference with the base period ($e_{ilt} - e_{il,2015}$) to avoid pre-treatment coefficients being mechanically around zero. Second, the wage outcome is $\frac{w_{i,l,t} - w_{i,l,2015}}{w_{i,l,2015}}$, so it measures the percentage change in wages $w_{i,l,t}$ for each worker i with respect to 2015, so the worker must be observed both in 2015 and t. Third, the earnings outcome is $\sum_{t=2016}^{t=2018} \frac{Earnings_{it}}{Earnings_{i,2015}}$ and it measures changes in the evolution of earnings normalized by the earnings in the pre-shock period, as in Autor et al. (2014). More precisely, the yearly earnings are zero if the worker is not employed in the formal sector in that given year, while if employed, they are equal to their wages, so this outcome yields a combined effect of the observed changes in employment and wages in the post-treatment periods.

The immigration shock $\Delta M_{l,2018}$ is defined as follows:

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

where the numerator is the stock of employed migrants from Venezuela (aged 18 to 64) in local labor market l who arrived in Colombia within the last five years before 2018, minus the stock of employed Venezuelan migrants in l who arrived in 2015 or earlier according to the census. Employed migrants include both Venezuelans and returning Colombians from Venezuela. The denominator $L_{Total,d,2018}$ is the total employed population in the local labor market.[23] I focus and interpret mainly the coefficient of 2018 in the regressions (i.e., θ_{2018}) to match the year of the census and avoid rescaling the shock as for the coefficients of other periods. Lastly, this constant immigration shock is useful because it exploits the full count of a census instead of a survey to construct migration shares. Moreover, it allows for transparent placebo tests on pre-trends within the same analysis, as illustrated by Dustmann et al. (2017).

Because migrants self-select into areas where the economic opportunities are better, the immi-

[23] Using a post-treatment period denominator can be a potential concern. Unfortunately, municipal-level employment and population data in Colombia before 2018 is scarce, with the most recent data available from the 2005 census. Nonetheless, I construct an immigration shock using the 2005 employed population in the denominator to find that the main results remain robust (see Appendix Table B.1, row 4).

gration rate $\Delta M_{l,2018}$ is likely to be endogenous, and its coefficient is upward biased (see ordinary least squares (OLS) estimates of Figure 4a and 4b). Thus, to consistently estimate the effect of immigration on the outcome variables, I instrument the immigration rate $\Delta M_{l,2018}$ with the distance to the nearest crossing bridge with Venezuela and with past settlements of Venezuelans. The motivation for the IV approach follows.

First, I use distance as an instrument for Venezuelan immigration since Colombia and Venezuela share 2,220 kilometers of terrestrial border. Therefore, arrivals to the local labor market l are a function of travel distance between the two countries, as distance imposes both time and economic constraints on Venezuelan immigrants. A potential threat to this identification strategy is that border areas may be more affected by economic shocks, such as reduced trade, compared to areas farther away from the Venezuelan crisis, thus violating the exclusion restriction.

Appendix Figure B.2a provides suggestive evidence that the trade shock resulting from the Venezuelan crisis began years before the immigration shock. Notably, border department exports to Venezuela consistently hovered around zero during the post-treatment period. Another critical piece of evidence is that I find insignificant employment and wage effects in the largest firms, which are likely more affected by trade shocks and less by immigration shocks since migrants disproportionally concentrate in small firms. In addition, I plot log GDP for border and non-border departments over time to show similar trends before the immigration shock, suggesting any trade impact on economic activity is limited (see Figure B.2b). Lastly, I exclude border areas from the main analysis and find similar point estimates, although they are insignificant for wages. Given this suggestive evidence in mind, it is formally required that distance satisfies the exogeneity assumption $E[f(dist_l)\Delta u_{lt}] = 0$.

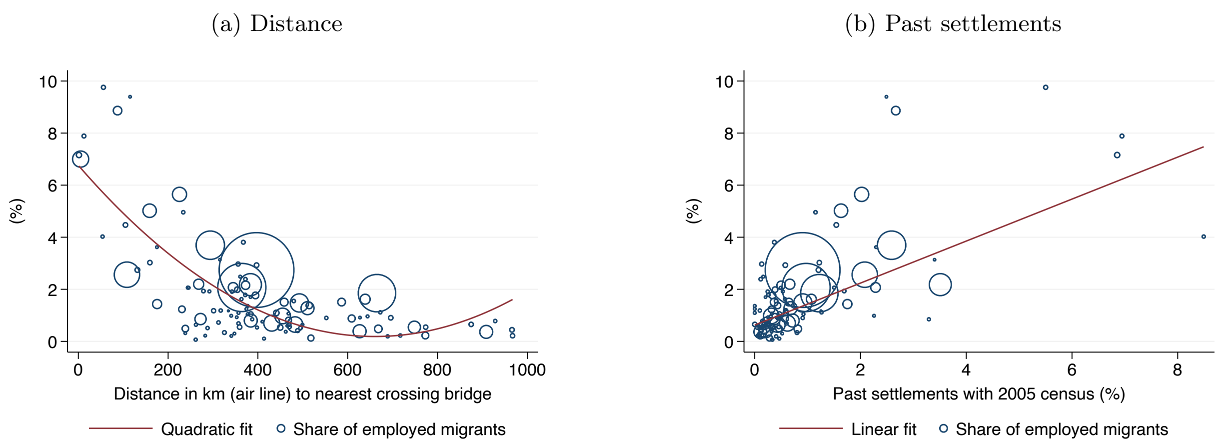
The other instrument follows the methodology of Altonji and Card (1991) and Card (2001), using past settlements of Venezuelans to predict new arrivals. It is defined as:

$$z_l = \left(\frac{Ven_{l,2005}}{Ven_{2005}} * M_{2018}\right) / L_{l,2005}, \tag{3}$$

where the first term represents the share of Venezuelans in FUA l, according to the 2005 population census, normalized by the working-age population $L_{l,2005}$ in l at 2005, whereas M_{2018} is the number of migrants in Colombia who arrived between 2016 and 2018 according to the census. I use past settlements as an instrument because newly arriving immigrants tend to move to areas where there are already established Venezuelan communities. For this instrument to be valid, past settlements must be related to new arrivals but not related to time-varying shocks, formally requiring $E[z_l \Delta u_{lt}] = 0$.

Figures 3a and 3b depict graphically the first stage of the immigration shock $\Delta M_{l,2018}$ for the 109 FUAs included in this analysis (see Appendix Map F.1 for the geographic distribution of the shock). These figures highlight the relevance and functional form of the instruments. For the first instrument, a larger distance from a crossing bridge decreases the share of employed migrants in the FUAs until a point where longer distances do not imply lower immigration rates, causing the slope of the curve to bend downward. The second instrument, based on past settlements, shows a positive and linear relationship with the immigration shock. The immigration shock at the FUA level is substantial, with some areas experiencing an increase in the share of employed migrants from 7% to 10% of their overall employed population. An improvement relative to Delgado-Prieto (2024) is that I construct more granular local labor markets using administrative data, while that study uses only 24 departments for the analysis due to the sample limitations of the labor force survey.

Figure 3: **Immigration rates and the two instruments**



Note: I weigh dots by formal employment according to the PILA in 2015. In (b), I exclude one area to narrow the x-axis values. Functional Urban Areas in Colombia (G=109). Source: CNPV, 2018.

I construct Figures 3a and 3b at the FUA level. However, since this paper aims to estimate the impact of immigration at the individual level, the first stage of the two-stage least squares regression (2SLS) will weigh each FUA differently based on the number of individual observations available. With this in mind, the first-stage model is:

$$\Delta M_{l,2018} = \delta + f(dist_l) + z_l + \upsilon_l \tag{4}$$

where $f(dist_l)$ includes linear and quadratic terms of the distance to the nearest crossing bridge, and z_l represents past settlements of Venezuelans. The error term υ_l captures the endogenous component of $\Delta M_{l,2018}$. I combine the two instruments in the analysis because past settlements and distance capture different exogenous components of migration while increasing the R^2 of the first-stage regression (see Table A.1).[24] As a result, I estimate Equation (1) using 2SLS with these two excluded instruments.

4 Worker Responses

This section examines the impact of immigration on formal wages and employment at the worker level and then explores the heterogeneity of these effects across different worker characteristics. First, I present wage and employment event study estimates using both OLS and 2SLS. One advantage of this empirical specification is the ability to test for differential trends in outcomes before the immigration shock occurs. Notably, I find no significant pre-trends for employment or wages that might confound the impact of immigration (see Figure 4a).

In the post-treatment periods, I find that OLS coefficients are close to zero, presumably downward biased, while the 2SLS regression helps mitigate this bias, resulting in more negative coefficients. Since pre-trends are rather stable before the arrival of immigrants with both methods, this suggests that immigrants are arriving in areas more likely to have positive demand shocks in the post-treatment periods, probably due to improving economic opportunities, which biases the OLS estimates towards zero. For that reason, I only analyze 2SLS coefficients hereafter. In 2018, the census year, a one pp increase in the share of employed migrants in an area reduces the probability of formal sector employment by 1.1 pp (see Figure 4a).[25]

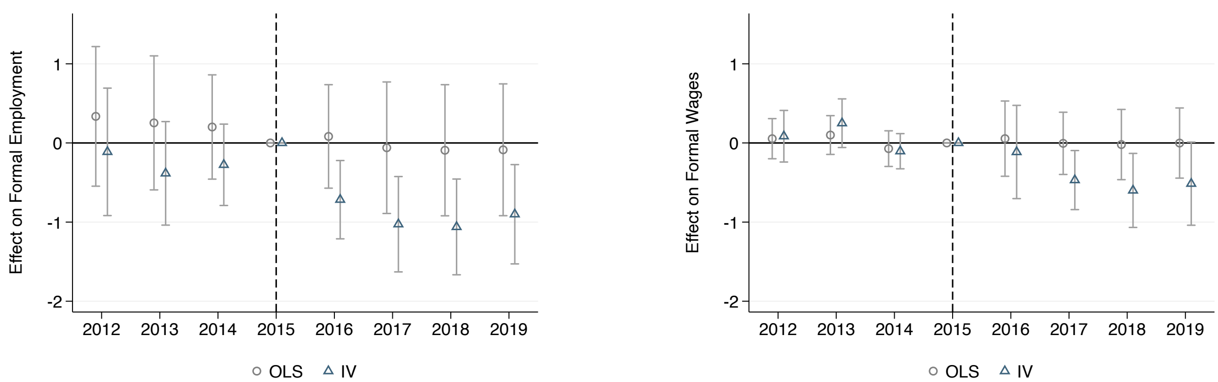
[24] Notably, the main coefficients do not change if I use one instrument instead of both.

[25] This regression uses $e_{i,l,2018} - e_{i,l,2015}$ as the dependent variable, capturing the change in the employment status

To interpret this coefficient, I quantified the probability of formal sector employment from the labor force survey, which is 0.42 for my main sample of workers aged 25 to 55 in 2015. Thus, a 1.1 pp drop corresponds to a 2.4% decrease relative to the mean. More broadly, a worker in an area at the 75th percentile of exposure relative to one at the 25th percentile of exposure experiences a 3.6% decrease in the probability of formal employment.[26] Regarding formal wages, I find a coefficient of –0.6% in 2018 for a one pp increase in the immigration shock (see Figure 4b).[27] Consequently, a worker at the 75th percentile of exposure relative to one at the 25th percentile of exposure experiences a drop of 0.9% in formal wages. Thus, the impact on wages is minor compared to the impact on employment.

Figure 4: **Event study estimates on individual wages and employment**

(a) Employment (b) Wages



Note: I estimate Equation (1) separately by year. The sample is restricted to natives aged 25 to 55. In panel (a), there are 6,706,035 workers, while in panel (b), this varies slightly by year as the worker must be employed in the post-treatment and base year. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for employment (in percentage points) and wages (in percent) are already multiplied by 100. Workers are observed in August of each year. Source: PILA 2012–2019.

For the remainder of the paper, I focus on the heterogeneity of wage and employment estimates based on workers' and firms' characteristics prior to the immigration shock, specifically using baseline characteristics from 2015. The coefficients for each subgroup come from separate regressions of the main empirical specification (Equation 1). Later, I develop a more systematic heterogeneity

from 2015 to 2018. For the heterogeneity analysis, the dependent variable is $e_{i,l,t} - \sum_{k=2013}^{2015} e_{i,l,k}/3$, which compares the post-shock employment status to the average pre-shock employment, yielding slightly less negative coefficients.

[26] The 25th and 75th percentile migration rate is 0.6% and 2.1%, respectively. So, (2.1-0.6)*2.4=3.6.

[27] I do not compare formal employment and wage estimates with other countries, as there are no papers, to the best of my knowledge, that estimate worker-level effects in developing countries. However, I do compare these estimates with regional-level estimates from the Colombian context later on.

analysis using a machine learning method.

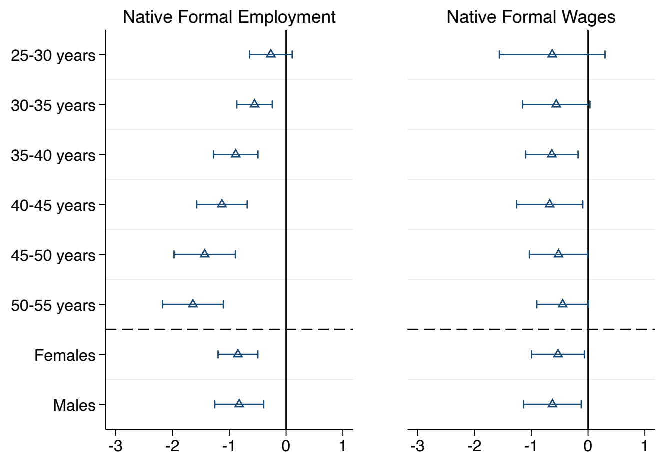
The first worker characteristic I examine is job type, categorizing workers as either employees or self-employed. In Colombia, self-employment represents about half of the employed population, predominantly in the informal sector, though there is a substantial portion in the formal sector (around 18% of all native formal workers were self-employed in 2015). Appendix Figure A.2 illustrates a greater decline in the probability of being a formal worker for self-employed natives than for employees, with pre-trends hovering around zero. Most self-employed individuals in the private sector can choose whether to contribute to the social security system, making it less costly for them to exit the formal sector compared to employees with labor contracts with firms.[28]

For clarity in presenting heterogeneous effects across multiple dimensions, I now show only the event study estimate for the year of the census, 2018, as the instruments do not predict differential wage or employment trends in the pre-treatment period. Moreover, Appendix Table E.1 shows no systematic pre-trends across different worker or firm categories for employment or wages. So, the next results I show are based on age groups and sex, which are also the controls I use in the main specification. Figure 5 shows that older workers experience a larger decline in the probability of formal sector employment compared to younger workers. In contrast, the pattern for wages is less clear, with similar negative estimates across all age groups. I extend the sample to include labor market entrants (18 to 24 years) and workers nearing retirement (56 to 64 years) in the base period. The highest negative effect on employment is observed among the oldest workers, suggesting they may be retiring earlier or shifting to the informal sector (see Appendix Figure A.1). For wages, no stark differences are observed across age groups. Finally, in terms of sex, the impact on employment and wages is similar for both men and women.

Next, I analyze the impact on earnings, which captures the combined effect of changes in employment and wages. Appendix Figure A.6a shows that workers over 30 experience a relatively similar reduction in earnings as their confidence intervals overlap. This suggests that while older workers are more frequently displaced from the formal sector, younger workers face larger wage losses.

[28]Labor income data for self-employed in PILA is noisy as it includes public and private contractors that typically report only 40% of their labor income by law. Nevertheless, the point estimates for wages are more negative for the self-employed than for employees.

Figure 5: **Employment and wage estimates by age group, 2015–2018**



Note: I estimate Equation (1) separately by subgroups. The sample is restricted to natives aged 25 to 55. The dependent variables are employment and wages relative to the base period. I use as controls the interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for employment (in percentage points) and wages (in percent) are already multiplied by 100. Workers are observed in August of each year. Source: PILA, 2013–2019.

To complement the pattern of employment effects by age group, I also calculate the labor supply elasticities at the extensive margin for each of these age groups (i.e., $\eta_w^s = \frac{\Delta L}{\Delta w}$). Appendix Table A.2 shows that as native workers age, their labor supply becomes more elastic. This means older workers are more responsive to employment changes due to wage fluctuations than younger workers. Consistent with this finding, Dustmann et al. (2017) also documents that the local labor supply elasticity in Germany increases with age.

To further unravel which types of workers are most affected, I now examine the number of years the employee has been employed in the same firm (i.e., job tenure) up to the base period of 2015. Appendix Figure A.3 splits the sample by job tenure of native workers, ranging from zero to more than nine years. Notably, the immigration shock on employment is more severe for workers with fewer years in the same firm. Still, the coefficient is less negative compared to older workers.[29] This result is partly due to severance payments, which increase with tenure, making it more costly for firms to dismiss longer-tenured workers. Additionally, longer-tenured employees often possess more firm-specific human capital, so they are less substitutable by migrants who have similar characteristics but lack country-specific skills.

[29] I construct job tenure from the first year PILA is available, 2007.

The evidence from these two analyses suggests that older workers and those with lower tenure experience the most significant drops in formal employment due to the immigration shock. To further clarify which workers are more affected, I combine their age and job tenure. Appendix Table A.3 shows that age is more influential for employment impacts than job tenure, as native workers below 35 have an insignificant effect on employment, regardless of their tenure. In contrast, native workers above 35 exhibit a significant negative effect on employment, with those with lower tenure experiencing a more pronounced effect than those with higher tenure (–1 pp versus –0.3 pp). Regarding wages, there are no clear differential effects across tenure and age.

4.1 Distributional Impacts of Immigration

Next, I estimate the impact of immigration on workers across the wage distribution. For this analysis, I categorize native workers into seven bins based on their local wage distribution in 2015. Figure 6 illustrates the uneven effects of immigration: native workers earning the minimum wage experience the most negative impact on formal employment, while workers at the rest of the wage distribution show insignificant changes in employment. Specifically, for low-wage workers, a one pp increase in the share of employed migrants in a given labor market leads to a 1.5 pp decrease in the probability of formal sector employment. To interpret this magnitude, I multiply this coefficient by the proportion of minimum wage workers within the formal sector (approximately 0.4) and the size of the formal sector (around 0.5). This calculation suggests that one additional employed immigrant displaces roughly (0.5*0.4*1.5=) 0.3 native minimum-wage workers in the formal sector. Conversely, formal workers who earn the minimum wage are the least affected by the immigration shock in terms of wages, partly due to their downward wage rigidity.

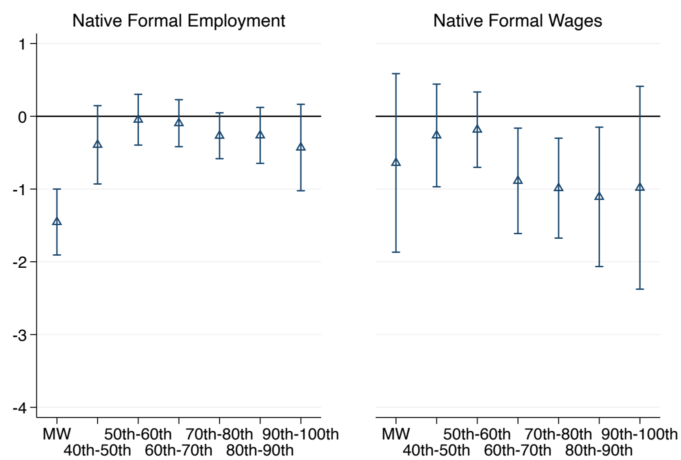
Due to the relatively high and binding nature of the minimum wage for around 40% of formal workers in the pre-shock period, these workers face a greater risk of job displacement.[30] Additionally, the substantial presence of a large informal sector explains part of the large coefficient, as minimum-wage workers tend to be less educated and more easily substitutable with informal workers, who become less costly following the arrival of migrants (Delgado-Prieto, 2024).

For workers between the 60th and 90th percentiles of the local wage distribution, earning around

[30] Conditional on being employed in the two periods, around 75% of minimum wage earners still earn the minimum wage after three years.

two to three times the minimum wage on average, I observe a negative wage effect between 1% and 1.2%. The increased competition from mid- and high-skilled migrants entering the formal sector can rationalize these results. Additionally, the contraction of the formal sector may strengthen the bargaining power of formal firms, which could affect formal wages. Note this does not necessarily imply an absolute decrease in wages. The coefficient reflects the average wage growth of native workers in areas with higher migration exposure compared to those with lower exposure, indicating relatively slower wage growth in more affected areas. Lastly, I estimate the impact on earnings to find which workers are more affected overall. Appendix Figure A.6b shows that the only significant negative impact on earnings is observed among workers who earn the minimum wage before immigrants arrive, reflecting a stronger effect from the disemployment margin.

Figure 6: **Employment and wage estimates by individual wage at baseline, 2015–2018**



Note: I estimate Equation (1) separately by subgroups. The sample is restricted to natives aged 25 to 55. The dependent variables are employment relative to the pre-shock period and wages relative to the base period. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for employment (in percentage points) and wages (in percent) are already multiplied by 100. Workers are observed in August of each year. Source: PILA, 2013–2019 .

4.2 Worker-level and Regional-level Effects

Most of the migration literature focuses on regional responses when examining immigration shocks.[31] Since regional responses aggregate multiple margins of adjustment to immigration, they can yield findings that differ from worker-level responses, as documented by Dustmann et al. (2023). To address this, I adapt to this setup the employment decomposition they introduce to shed light on these different responses.[32] Specifically, I decompose the changes in regional formal employment into three different components: (1) a displacement of incumbent workers –outflows from formal employment–, (2) hiring of new formal workers or inflows from other regions –inflows to formal employment–, and (3) relocation of existing employed formal workers to other regions.

In this analysis, worker-level employment estimates capture the outflows or displacement of incumbent native workers from the formal sector, while the regional-level estimate from cross-sectional data combines all three margins of adjustment. Appendix Figure A.4 illustrates the decomposition of the regional formal employment response at the FUA-level (–1.3%), breaking it down into the three components: outflows to non-employment, or the informal sector (1.1%), inflows from other regions, non-employment or the informal sector (–0.5%) and relocation to other regions (–0.4%). More formally, the employment decomposition equals:

$$-1.3\% = \frac{E_{r1} - E_{r0}}{E_{r0}} = -\underbrace{\frac{E_{r,Out}}{E_{r0}}}_{\text{Outflows}} + \underbrace{\frac{E_{r,In}}{E_{r0}}}_{\text{Inflows}} - \underbrace{\frac{E_{r,Move}}{E_{r0}}}_{\text{Relocation}}. \tag{5}$$

In this analysis, the most important and only significant margin is the outflows from the formal sector, a finding that contrasts with Dustmann et al. (2023), where inflows are the most prominent margin. The differences in results can be attributed to a large informal sector in Colombia, enabling firms to hire informally after displacing formal workers, and the relatively less restrictive job protection regulation in Colombia compared to Germany.

Regarding wage estimates, the worker-level response shows a decrease of 0.6%, while the

[31] Recent regional-level studies include Monras (2020) in the US and Muñoz (2024) in the EU. The first study documents that low-skilled Mexicans who left their country due to the peso crisis had a high transitory impact on local labor markets in the US. The second study exploits a trade liberalization in services across Europe to find negative regional effects on the employment of domestic workers.

[32] The main distinction, relative to Dustmann et al. (2023), is that the outflows and inflows margins in this study can be decomposed further into non-employment and the informal sector. Unfortunately, there is no panel data for the informal sector to measure these decompositions.

regional-level estimate in Delgado-Prieto (2024) is insignificant and nearly zero. These two responses are complementary and address different policy questions. As noted in Dustmann et al. (2023), worker-level wage regressions capture the change in the price of labor, holding population composition constant. In contrast, regional-level wage regressions jointly capture the change in the selection and composition of workers due to inflows and outflows alongside changes in labor prices. The differential estimate between the two can be rationalized as follows. The immigration shock alters the composition of employed natives by displacing mostly minimum wage workers, positively selecting those who remain employed, thereby mechanically increasing regional formal wages (see Figure 6). However, immigration reduces labor prices in certain mid- and high-wage subgroups, reducing regional formal wages on average. This explains the insignificant formal wage effect at the regional level versus the negative wage effect at the worker level, highlighting the need to analyze immigration effects on both aggregate local labor markets and individuals within local labor markets.

Individual panel data also offers insights into inter-regional movements as a response to immigration. For example, Foged and Peri (2016) document that younger workers in Denmark are more mobile following refugee arrivals. Appendix Table A.4 shows the effect on regional movements by age groups. Younger formal workers tend to move more, though coefficients are insignificant. While point estimates decrease with age, all remain insignificant, indicating that mobility is a less relevant adjustment margin in this context.

5 Immigration, Workers, and Firms

In this section, I first develop a partial equilibrium model with heterogeneous firms and types of workers to motivate and interpret the empirical findings across firms. Then, consistent with the model, I show that the effects for natives vary substantially depending on the type of firm they were employed in before the immigration shock. Last, I perform a standard heterogeneous analysis that motivates the next section's machine learning heterogeneity analysis.

5.1 Model

The market structure of the model consists of J formal firms that can hire two types of labor inputs. Specifically, firms hire formal workers F paying payroll taxes and informal workers I off the books to avoid paying the payroll taxes, as in Ulyssea (2018).[33] So, each firm $j = \{1, ..., J\}$ posts a pair of wages (w_{I_j}, w_{F_j}) that all workers i observe and decide to accept. Importantly, each firm has different work environments, measured by amenities a_{L_j}, and workers have idiosyncratic preferences ϵ_{i,L_j} depending on the fixed labor group they belong $L \in \{I, F\}$. This gives a worker-specific job valuation at each firm, implying firms face upward-sloping labor supply curves.[34] In this case, the indirect utility of worker i employed at firm j is:

$$\upsilon_{i,L_j} = \beta_L lnw_{L_j} + a_{L_j} + \epsilon_{i,L_j}. \tag{6}$$

Under the assumption that ϵ_{i,L_j} follows a type I extreme value distribution for each of the workers' types $L \in \{I, F\}$ and that the number of firms J is sufficiently large, Card et al. (2018) shows that the firm-specific supply functions can be expressed as:

$$lnI_j(w_{I_j}) = ln(\mathcal{I}\lambda_I) + \beta_I lnw_{I_j} + a_{I_j}, \tag{7}$$

$$lnF_j(w_{F_j}) = ln(\mathcal{F}\lambda_F) + \beta_F lnw_{F_j} + a_{F_j}. \tag{8}$$

In this case, the total number of informal workers in the market is $\mathcal{I}$ and of formal workers is $\mathcal{F}$, where λ_I and λ_F are constant parameters across firms. Moreover, $\frac{dlnL(w_{L_j})}{dlnw_{L_j}} = \beta_L$ is the elasticity of labor supply to the firm with respect to its wage. Hence, as $\beta_L \to \infty$, the supply functions become perfectly elastic, and firms have no monopsony power to set wages below the marginal productivity of labor.[35] Monopsony power is important for explaining equilibrium wage differentials across firms,

[33] In this model, I abstract from the decision of the firm to become formal or informal, as I focus only on the labor choices of formal firms. Moreover, workers' transitions between the formal and informal sectors are out of the model's scope.

[34] For instance, preferences for working in a firm may refer to distance to the workplace or interactions with coworkers (Card et al., 2018).

[35] Here, I exclude any market wage offered in an outside competitive sector as the comparative statics focus is on firm-level responses to immigration and not on market-level responses that have been thoroughly analyzed in the migration literature.

consistent with the findings of the AKM model. However, my focus differs from Amior and Stuhler (2022), who examine the role of monopsony power under immigration shocks. Instead, I explore how firm heterogeneity generates different predictions about the impact of immigration on workers, assuming monopsony power remains constant across firms.

Regarding firms, there is a productivity shifter T_j, a price of the good P_j, and a production function Q_j for each firm, such that the profit function of firm j is:

$$\max_{I_j, F_j} \pi_j = P_j T_j Q_j - \tau(I_j) w_{I_j}(I_j) I_j - (1 + \tau_F) w_{F_j}(F_j) F_j. \tag{9}$$

Here, $\tau(I_j)$ represents a convex cost that is increasing on the firm's informal labor size. These convex costs are important to match the stylized fact that informal labor decreases with firm size and captures the cost of evasion related to law enforcement exerted by the government. Particularly, I assume that $\tau(I_j) = I_j^{\eta}$ with $\eta \geq 0$. The τ_F represents the payroll taxes firms must pay for formal workers, and the production function takes the following form: $Q_j = (\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho})^{\frac{1}{\rho}}$. Thus, formal and informal workers are imperfect substitutes, and the aggregate elasticity of substitution common across all firm types is given by $\sigma = \frac{1}{1-p}$. To finish the setup, P_j is the inverse demand function defined as $P_j = D_j(T_j Q_j)^{-(1-\epsilon)}$, where $\epsilon^D = -1/(1-\epsilon)$ is the elasticity of product demand and D_j is the firm-specific product demand.[36]

In this partial equilibrium framework, I then analyze the impact of an immigration shock that shifts the aggregate informal labor supply outwards ($d\mathcal{I}$).[37] I study the firms' response across the wage and employment margin, so the wage elasticity for each type of worker in firm j is $\varepsilon_{w_{L_j},\mathcal{I}}$ and the employment elasticity for each type of worker in firm j is $\varepsilon_{L_j,\mathcal{I}}$. Allowing for firm-level responses to an immigration shock is the main contribution of this framework. Unsurprisingly, in Appendix D I show after some derivations that the elasticity of informal labor is always positive ($\varepsilon_{I_j,\mathcal{I}} > 0$) and the elasticity of informal wages is always negative ($\varepsilon_{w_{I_j},\mathcal{I}} < 0$) after an aggregate informal labor shock, regardless from whether informal and formal workers are close substitutes or not.

More interestingly, I show how formal wages and employment of firm j change in response to

[36] For simplicity, in this model, I do not distinguish if the produced good is tradable or non-tradable. Besides, I exclude any spillover labor demand shock arising from the consumption of goods and services from migrants.

[37] Figure 1a shows that around 90% of Venezuelan immigrants are employed in the informal sector.

the immigration shock:

$$\varepsilon_{w_{F_j},\mathcal{I}} = \Omega_j s_{I_j}(\epsilon - \rho). \tag{10}$$

Here, $s_{I_j} = \frac{\alpha_I I_j^\rho}{\alpha_I I_j^\rho + \alpha_F F_j^\rho}$ is the relative contribution of informal work to production before immigrants arrive and $\Omega_j = \frac{1}{\xi_{I_j}\xi_{F_j} - (\epsilon-\rho)^2 s_{I_j}\beta_I s_{F_j}\beta_F}$ is a positive parameter.[38] Firstly, if informal workers are close substitutes to formal workers (such that $\rho > \epsilon$), then the elasticity of formal wages with respect to aggregate informal labor is negative. Importantly, as the contribution of informal labor to production in firm j increases ($s_{I_j} \uparrow$), the elasticity of formal wages is more negative ($\varepsilon_{w_{F_j},\mathcal{I}} \downarrow$). Note that, for certain low productivity firms, formal wages can be downwardly rigid due to the existence of a minimum wage, so the formal wage margin is muted (i.e., $\varepsilon_{w_{F_j},\mathcal{I}} = 0$).

In terms of formal employment, the corresponding expression is equal to:

$$\varepsilon_{F_j,\mathcal{I}} = \Omega_j s_{I_j}(\epsilon - \rho)\beta_F. \tag{11}$$

The implications for formal employment in terms of the substitution parameter (i.e., $\rho > \epsilon$) hold similarly as for formal wages, though the response is now adjusted by β_F. Hence, as the relative contribution of informal workers to production increases ($s_{I_j} \uparrow$), the adjustment on formal employment is more negative ($\varepsilon_{F_j,\mathcal{I}} \downarrow$) as long as informal and formal workers are sufficiently close substitutes.

To summarize, the model I propose points to two main conclusions. First, when the substitutability between formal and informal workers is high, an informal labor supply shock negatively impacts formal wages and employment. Second, the production structure determines how responsive the firm is to such shocks. Specifically, firms that rely more heavily on informal labor for production will adjust formal wages and employment more intensely in response to an immigration shock.

In the model, a firm's informal production share is inversely related to its size. Large firms face higher marginal costs for hiring additional informal workers due to the convex cost of informal labor $\tau(I_J)$. Therefore, my empirical analysis first examines worker responses across the firm size

[38] To show that $\Omega_j > 0$, note that this can be simplified as $\Omega_j = (1 + (1+\eta-\rho)\beta_I)(1+(1-\rho)\beta_F) - (\epsilon-\rho)(s_{I_j}\beta_I + s_{F_j}\beta_F + (1+\eta-\rho)\beta_I s_{F_j}\beta_F + (1-\rho)\beta_F s_{I_j}\beta_I)$ which is always positive.

distribution to validate the model's predictions. Then, I extend the analysis to worker responses across firms' productivity, as it is also a measure linked to the firm's informal production share.

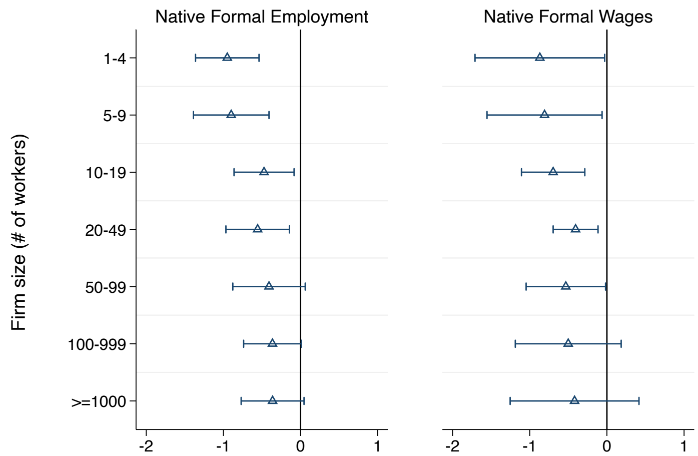
5.2 Worker Responses Across Firms

The previous model suggests that formal workers in smaller and less productive firms are more affected by an informal labor supply shock. Therefore, I turn to the data to empirically test these implications. The firm dimension is also key for this heterogeneity analysis for three reasons described in Delgado-Prieto (2024). First, over 80% of Venezuelan immigrants work in firms with less than 11 workers. Second, more than 50% of workers in the smallest formal firms earn the minimum wage and hence are in the margin of informality. Third, smaller firms employ a higher share of informal workers, a share that decreases as firms grow. Thus, the impact of immigration on workers in small firms is more salient, as they can substitute more easily formal for informal labor.

For this analysis, I categorize workers by firm size categories in 2015, the year before the immigration shock, and examine worker-level employment and wage coefficients for 2018, the year of the census. By comparing worker outcomes in small or large firms across local labor markets over time, I rule out time-varying effects of firm size from the impact of the immigration shock. Figure 7 shows that native workers in firms with fewer than 50 employees in the pre-shock period experience the most negative effect on the probability of formal employment, whereas workers in larger firms are less affected. In line with the model's predictions, small firms, which rely more on informal labor and face lower costs of being caught, find it profitable to substitute formal workers with informal ones when the two are close substitutes. Delgado-Prieto (2024) confirms that after migrants arrive, the share of informal labor increases more in smaller firms, indicating workforce composition changes due to this labor substitution.

Regarding wages, workers in the smallest firms (fewer than ten employees) face the most negative effects. Yet, workers in firms with less than 100 workers also experience a significant negative effect. This pattern aligns with model predictions, where smaller firms adjust wages more heavily in response to an immigration shock. These results are useful to transparently show that trade shocks from the Venezuelan crisis are less of a concern in this context, as the main effects are observed in the small formal firms directly impacted by migration and less likely by trade shocks.

Figure 7: **Employment and wage estimates by firm size, 2015–2018**



Note: I estimate Equation (1) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. I use as controls the interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. Workers are observed in August of each year. The coefficients for employment (in percentage points) and for wages (in percent) are already multiplied by 100. Source: PILA, 2013–2019.

Next, I quantify worker-level effects based on other relevant firm characteristics, specifically the firm's age up to the pre-shock year, using the information from administrative records as a proxy. Appendix Figure A.5 presents the results for native workers by the age of their firm in 2015. Workers in younger firms experience more negative impacts on employment than workers in older firms, while the wage pattern is less clear, with workers in younger firms being more affected. The positive correlation between firm size and age could explain previous findings, as smaller firms tend to be younger. However, Fort et al. (2013) document that firm responses during the business cycle vary by age and size, so I combine these characteristics to assess how worker-level effects vary. Appendix Table A.5 shows that native workers in the youngest firms face significant negative effects on employment and wages, and they vary depending on whether the firm is small or large. Conversely, native workers in older firms show a significant negative effect on employment only in the smallest firms.

5.3 Worker Responses by Firm Pay Premiums

With access to the full count of formal workers and firms in Colombia, I can construct a measure of firms' wage premiums, serving as a proxy of their productivity. To achieve this, I first estimate the standard AKM model proposed by Abowd et al. (1999), which decomposes the contribution of firm-specific and worker-specific constant characteristics to log formal wages lnw_{it}. The AKM model is expressed as follows:

$$lnw_{it} = \alpha_i + \psi_{j(i,t)} + X'_{it}\beta + \epsilon_{it}. \tag{12}$$

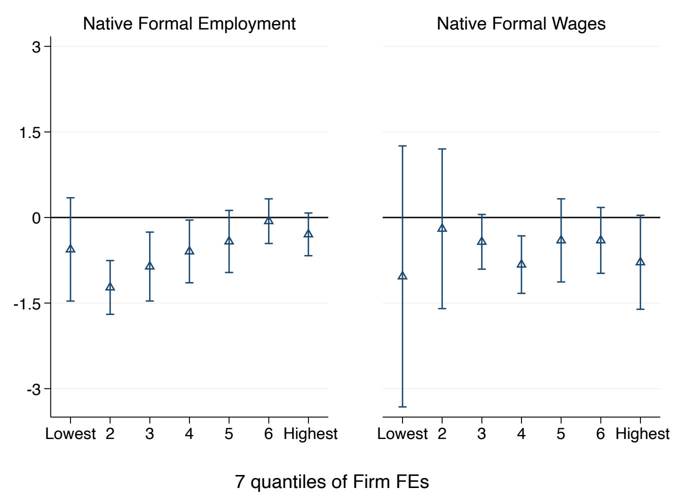
Here, α_i captures the unobserved worker effect, ψ_j captures the unobserved firm effect, and $j(i,t)$ refers to the firm j where worker i is working in t. X_{it} is a vector of controls that are age squared and cubic after being normalized and year FEs. Lastly, ϵ_{it} is the error term. To rule out possible endogenous workers' movement due to the immigration shock, I estimate the model with data from 2010 to 2015 for August ($T = 6$), restricting the sample to the largest set of firms connected by the mobility of workers.

In these models, ψ_j are identified through workers' movements across firms, assumed to be exogenous conditional on worker and firm fixed effects (i.e., $E[\epsilon_{it}|\alpha_i, \psi_j(i,t), X_{it}] = 0$). Several studies have indicated that AKM models present biases in estimating $Var(\hat{\psi}_j)$ or the $Cov(\hat{\alpha}_i, \hat{\psi}_j)$ when workers mobility across firms is limited, particularly within smaller firms or with few estimating periods (Bonhomme et al., 2023). Several strategies have been proposed to address this limitation. Given that I use only the estimated vector of firm FEs $(\hat{\psi}_1, ..., \hat{\psi}_J)$ and worker FEs $(\hat{\alpha}_1, ..., \hat{\alpha}_N)$, which are unaffected by limited mobility bias, this concern is mitigated (Bonhomme et al., 2023). Still, I also use the leave-out method proposed by Kline et al. (2020) to address the limited mobility bias when decomposing the sources of wage inequality.

Appendix Table C.2 presents the decomposition of the variance of wages $Var(lnw_{it})$ in Colombia's formal sector. Worker effects explain 50.2% of the variance of wages, and firm effects explain 15.7%, in line with the literature cited in Card et al. (2018). Additionally, the positive sorting of high-wage workers into high-wage firms explains another 21.6% of the variance. In four European countries and the US, this sorting explains between 10% to 20% of the wage variance (Bonhomme et al., 2023).[39]

Using the estimated $\hat{\psi}_j$, which are relative to the largest firm in the country, I now categorize workers into seven quantiles of firm FEs or firm-specific pay premiums, referred to as lowest- to highest-paying firms, to analyze the impact of immigration. Figure 8 shows that workers at low-paying firms experience negative employment effects while having insignificant wage changes. This contrasts with workers in middle-paying firms who experience negative wage and employment effects. A possible explanation is that the share of firms in the low-pay sector grows as immigrants predominantly work in these firms. Consequently, high-pay sector firms may extract higher rents from workers, thus reducing their wages, as shown in a model with on-the-job search in Amior and Stuhler (2022). Lastly, to determine whether employment or wage changes prevail, I estimate the earnings outcome across quantiles of firm FEs. Appendix Figure A.6c shows that workers in the lowest-paying firms experience a more pronounced decline in earnings than workers from middle- to high-paying firms.

Figure 8: **Employment and wage estimates by quantiles of firm FEs, 2015–2018**



Note: I estimate Equation (1) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old. The dependent variables are employment relative to the pre-shock period and wages relative to the base period. I compute Firm FEs in the first stage using the standard AKM framework, with age squared and its cubic as time-varying controls, for the period 2010-2015. I use as controls in the second stage interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for employment (in percentage points) and wages (in percent) are already multiplied by 100. Workers are observed in August of each year. Source: PILA, 2013–2019.

[39] The four European countries are Austria, Italy, Norway, and Sweden. The method they use for estimating the sorting in 6-year panels is the correlated random effects based on the grouping proposed by Bonhomme et al. (2019).

Appendix Figure A.7 presents a similar analysis by dividing workers into seven quantiles based on worker FEs $\hat{\alpha}_i$. High-wage workers exhibit more negative point estimates for wages and the least negative ones for employment. Conversely, low-wage workers show close to zero wage effects, while the employment effects are more negative.

5.4 Heterogeneity by Worker and Firm Characteristics

As shown previously, workers experience heterogeneous employment and wage effects based on their own characteristics as well as the type of firms they were employed at before immigrants arrived. To illustrate the groups most affected in a standard way, I restrict the sample to the intersection of subgroups where previous findings indicate more negative coefficients. In the next section, I introduce a systematic analysis of heterogeneity using a machine learning method.

First, Table 3a shows that for minimum wage earners in 2015, immigration reduces the probability of formal sector employment by 1.5 pp. For the medium age group, the impact is less negative at 1.2 pp, while for self-employed workers, the impact is more negative at 2.2 pp. When combining these three characteristics, there are 565,594 workers in the sample, for whom the negative effect of the immigration shock on the probability of being a formal worker is 2.6 pp.

Table 3b divides the sample into subgroups with the highest negative coefficients for native wages. It shows that for workers earning more than the minimum wage in 2015, migration reduces average wages by 0.7%. For workers in the smallest firms in 2015, the impact is more negative at 0.8%, while for workers in middle-paying firms in 2015, the estimate is also 0.8%. When combining these characteristics, there are 30,772 workers in the sample, and the effect on wages in 2018 is a reduction of 1.9% for a one pp increase in the immigration shock. However, this analysis is subject to arbitrary sample restrictions and smaller sample sizes, which can lead to differential effects partly due to random variation or statistical noise. Therefore, in the next section, I propose a method to estimate heterogeneous immigration effects in a data-driven way.

Table 3: **Most Affected Native Workers: Employment and Wages, 2015–2018**

(a) Employment

	(1)	(2)	(3)	(4)	(5)
Prob. of Employment	-0.841***	-1.453***	-1.188***	-2.194***	-2.647***
	(0.192)	(0.231)	(0.227)	(0.327)	(0.388)
Sample restriction					
Minimum wage earners	✗	✓	✗	✗	✓
Median age (35 years or more)	✗	✗	✓	✗	✓
Self-employed	✗	✗	✗	✓	✓
N	6,706,035	2,205,814	3,915,188	1,103,384	565,594
Clusters	109	109	109	109	109

(b) Wages

	(1)	(2)	(3)	(4)	(5)
Wages	-0.600*	-0.711*	-0.827**	-0.804**	-1.908**
	(0.239)	(0.315)	(0.320)	(0.260)	(0.477)
Sample restriction					
Above minimum wage	✗	✓	✗	✗	✓
Small firm (1 to 19 workers)	✗	✗	✓	✗	✓
Middle-paying firm (quantile 4)	✗	✗	✗	✓	✓
N	4,090,973	2,639,040	643,346	195,647	30,772
Clusters	109	109	109	109	109

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

Note: I estimate the equation separately by subgroups. For Panel A, the outcome variable is $e_{i,2018} - \sum_{t=2013}^{2015} e_{it}/3$ where e_{it} indicates formal sector employment. For Panel B, the outcome variable is $\frac{w_{i,2018}-w_{i,2015}}{w_{i,2015}}$ where w_{it} indicates wages in the formal sector. The sample is restricted to natives aged 25 to 55. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. Workers are observed in August of each year. Source: PILA, 2015?2018.

5.5 Sorting

Recent findings indicate that workers' reallocation across firms is a relevant response to labor market distortions. For example, Germany's introduction of a national minimum wage led low-wage workers to reallocate to higher-paying firms (Dustmann et al., 2022). I examine the reallocation effects of an immigration shock by analyzing changes in the sorting patterns of high- and low-paying workers into high- and low-paying firms.[40] I construct the outcome using $\hat{\psi}_j$ values from Equation (12) and exploiting the movements of workers between firms in the post-treatment period.

[40]For France, Orefice and Peri (2024) find that high-paying workers are moving more into high-paying firms following the arrival of migrants. Similarly, Gyetvay and Keita (2023) finds that German natives shift from low- to high-paying firms.

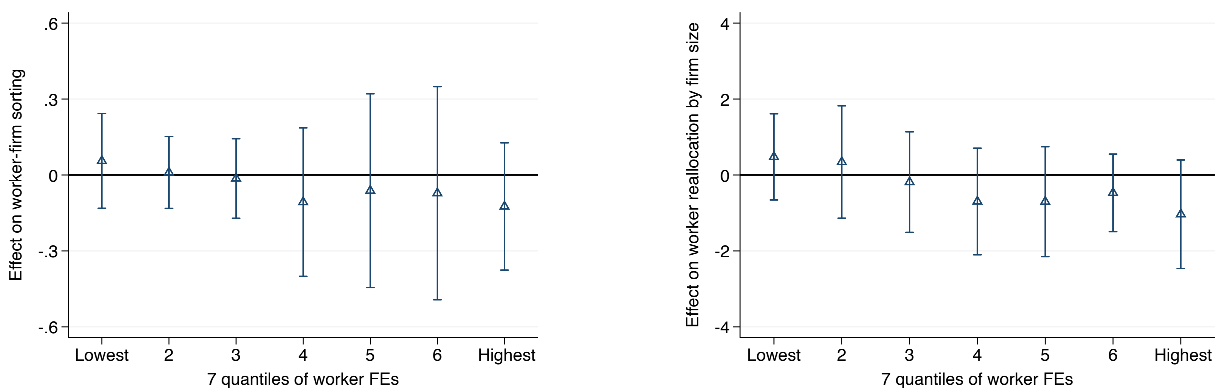
More concretely, the outcome measures the change in the AKM firm FEs in 2018 relative to 2015: $\hat{\psi}_{i,\{j=2018\}} - \hat{\psi}_{i,\{j=2015\}}$. If the worker remains in the same firm during that period, the difference is zero.[41]

To understand whether low- or high-wage workers are sorting into different-paying firms post-immigration shock, I present results across seven quantiles of worker FEs. A positive coefficient suggests a positive sorting effect from immigration. Figure 9a shows that none of the categories yield significant results, indicating no differential sorting due to immigration.[42] Therefore, the negative wage coefficient observed in workers from high-paying firms likely reflects lower wage growth within these firms due to increased competition from migrants. Furthermore, there is no evident reallocation of workers between larger or smaller firms post-immigration shock (see Figure 9b).

Figure 9: **Reallocation estimates by quantiles of worker FEs, 2015–2018**

(a) By firm pay-premiums

(b) By firm size



Note: The sample is restricted to natives aged 25 to 55. The dependent variable in (a) is the change in $\hat{\psi}_{i,\{j=2018\}} - \hat{\psi}_{i,\{j=2015\}}$ and in (b) is the change in the categories of firm size in 2018 relative to 2015, both measured in the base period. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. Workers are observed in August of each year. Source: PILA, 2013–2019.

[41]Firms FEs are based on the pre-policy period, excluding firms created post-2015. The estimated firm FEs are adjusted to positive values for the outcome.

[42]This lack of significant sorting may be due to the macroeconomic conditions in Colombia, where unemployment slightly increased during the study period.

5.6 Changes in Hiring Patterns of Formal Firms

A key mechanism in this paper is the substitution of formal for informal workers, which comes from the strong linkages between the formal and informal sectors. Due to the absence of informal worker-level data in administrative records, I construct an alternative measure of firms' connectedness with the informal sector for formal firms, aside from the standard firm size variable. To build this proxy, I develop an insider index akin to the poaching index in Bagger and Lentz (2019). The intuition of the index is that firms are divided by the share of hires that come from *outside* the formal sector, that is, workers who have not been employed in the formal sector before (most likely labor market entrants or workers from the informal sector) and *inside* the formal sector (workers previously employed in other formal firms or unemployed but formerly in the formal sector). Essentially, it reflects the revealed preferences of workers, as firms hiring more from the formal sector are more desirable, while those hiring more labor market entrants or from the informal sector act as "gatekeepers"' for formal sector entry. The insider index is constructed for each firm j before and after immigrants arrive:

$$\pi_{j,t} = \frac{N_{j,t}^{In}}{N_{j,t}^{In} + N_{j,t}^{Out}}, \tag{13}$$

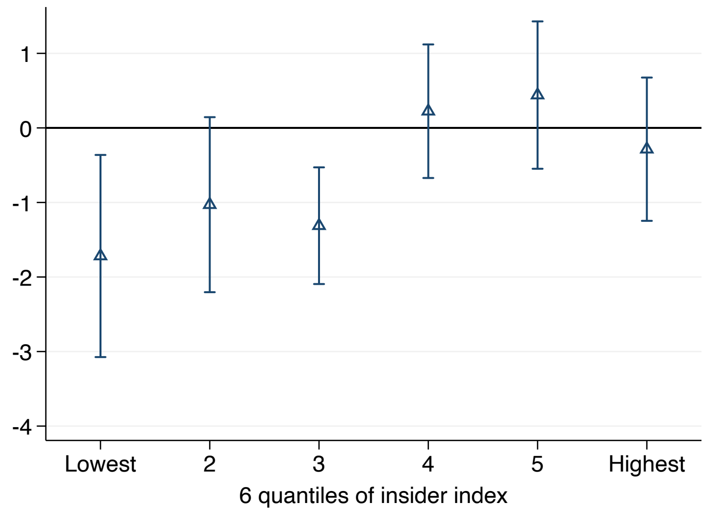
where $N_{j,t}^{In}$ is the number of firm j hires in year t hires from the formal sector, and $N_{j,t}^{Out}$ is the number of firm hires from outside the formal sector.[43] I then compute the difference in the insider index between 2018 and 2015 ($\pi_{j,2018} - \pi_{j,2015}$) at the worker level, based on the firm of employment in the two years.

Figure 10 shows results across six quantiles of the insider index based on the pre-shock period. Interestingly, formal firms that typically hire workers from the informal sector show a negative effect on their insider index post-immigration, indicating fewer hires that belong in some year to the formal sector (a 1 pp increase in the migration rate reduces the insider index of the lowest type of firms by around 1.7 pp). On the contrary, the insider index does not change much for firms with a higher share of hires within the formal sector. This measure is an important way of showing that some firms, particularly those more connected to the informal sector, are opting out or poaching

[43] Hiring data is available from 2007 to 2018, recorded for February and August each year. The index is missing if a firm did not hire in a year.

less from the formal sector for new hires.

Figure 10: **Index estimates by quantiles of the baseline insider index, 2015–2018**



Note: Dependent variable is the change in the insider index for workers employed in firm j in the base period between 2015-2018. Controls include interactions of sex with six age categories in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. Workers are observed in August of each year. Source: PILA, 2013–2019.

5.7 Exit and Entry of Formal Firms

Another significant margin of adjustment to the arrival of immigrants is formal firm dynamism, specifically related to the closures and openings of formal firms. I estimate the likelihood of firms exiting the formal sector entirely post-immigration. For this exercise, I count the firms in each FUA over time and perform regional-level regressions. Table 4 shows that formal firms experience negative growth in areas receiving more immigrants compared to those receiving fewer immigrants, though the coefficient is insignificant at the 5% level. When breaking this growth down into exits and entries, there is a marked increase in firm exits. A 1 pp increase in immigration increases by 1.2% the firm exit rate. This does not necessarily mean firms cease operations entirely, as they may continue hiring all workers informally. In contrast, the firm entry rate remains almost unchanged, indicating that immigrants are not spurring the creation of formal firms in the short term.

Table 4: **Decomposition of Firm Growth, 2015–2018**

	(1) Total Firms	(2) Firm Exit	(3) Firm Entry
$\Delta M_{l,2018}$	-1.127	1.190*	0.063
	(0.750)	(0.582)	(0.935)
N	109	109	109

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

Note: Regressions are estimated at the regional level for 109 FUAs weighted by their formal employment in 2015. The outcome variable in (1) is the percent growth in the number of firms, while in (2) and (3), I decompose the percent growth in terms of the exit and entry of firms, respectively. The sample is restricted to firms with at least one native employee. Firms are observed in August of each year. Source: PILA, 2015–2018.

6 Machine Learning

In this section, I employ machine learning to identify the subgroups most affected by immigration and to determine a proxy for the importance of firms in the labor market effects of immigration. Previously, I presented wage and employment effects for arbitrarily chosen population subgroups based on certain characteristics. To more accurately determine which variable explains most of the heterogeneity of immigration effects, I turn to a data-driven approach proposed by Athey and Imbens (2016) and generalized by Athey et al. (2019), recently implemented by Gulyas et al. (2019) and Yakymovych et al. (2022). This framework identifies the subgroups experiencing the greatest wage and employment losses through a recursive partitioning method that allows for non-linear effects and high-order interactions between firm and worker variables. Specifically, I use the generalized random forest (GRF) method from Athey et al. (2019) to build causal forests in the spirit of random forests (Breiman, 2001) but splitting the data according to a criterion on treatment effect heterogeneity.[44] The benchmark specification that the algorithm uses is as follows:

$$\Delta Y_{i,l,2018} = \tau(x_i)\Delta \hat{M}_{l,2018} + \Delta\epsilon_{i,l,2018}. \tag{14}$$

Here, x_i represents the value of the variables in X_i, and $\tau(x_i)$ denotes the treatment effect. The outcome of interest, $\Delta Y_{i,l,2018}$, is the change in individual employment or wages in 2018 relative to the pre-shock period. $\hat{M}_{l,2018}$ is the predicted immigration rate after regressing the observed rate on the instruments. This is done because the algorithm does not allow for multiple instruments.

[44] I use the `grf` package in `R` to estimate the causal forests.

Vectors of worker and firm variables, including the ones constructed from the AKM model, are the partitioning variables f included in the vector X_f. All these features or variables correspond to baseline characteristics in 2015 and include age, sex, job tenure, wages, firm FEs, worker FEs, and firm size. Self-employed workers are excluded from this section as their firm-related characteristics are not comparable to those of employees.

The procedure outlined in Athey and Imbens (2016) and Athey et al. (2019) for building causal trees involves multiple steps, which have been adapted to this setup. Broadly, the algorithm proceeds as follows:

1. Start with 50% of the full sample P.[a] I use the remaining out-of-bag (OOB) sample for estimation after the algorithm is trained.
2. Take a random subsample, without replacement, of P and choose a variable randomly from X_f and a value from all possible values for this selected variable.
3. For every possible value of one variable in X_f, the data is split into two partitions (say P_l and P_r) to run separate regressions of form (14) to estimate treatment effects for each partition. Choose the variable with its cutoff value that maximizes the difference in treatment effects using this formula:

$$(\tau_l - \tau_r)^2.^{b} \tag{15}$$

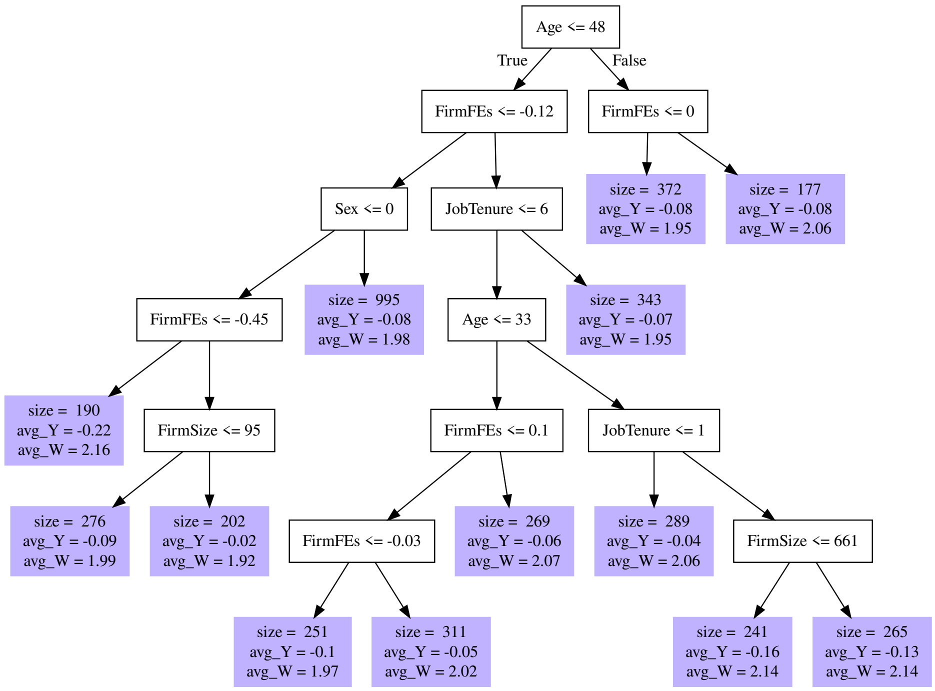
4. Observations with a value below or equal to the cutoff value are placed into a new left node, and observations with a value above are placed into a new right node of the decision tree.
5. Recursively forms the resulting nodes with this algorithm until the nodes reach a minimum node size, the difference in sample size between the two partitions is large, or when the split would only yield a difference in treatment effects relatively small.

[a] The 50% threshold is selected due to computational burden. This subsample is further cut by 50% to do sub-sample splitting to create similarity matrices.

[b] There are penalties in the algorithm for the imbalance of the splits. For instance, the squared difference criterion can include an additional term $\frac{n_l n_r}{N^2}$ to adjust for more balanced splits (n_l and n_r refer to the sample size of each partition, and total subsample refers to N).

As an illustration of a decision tree within the causal forest algorithm, I use a 1% random sample of the main data. Figure 11 shows how observations with specific characteristics are split to the right or left of the tree based on a cutoff value (48 years) after testing cut-off values in all variables to maximize the squared difference in treatment effects in this subsample. For the main algorithm, I estimate the causal forest using 2,000 decision trees with a minimum node size of 300 while clustering observations by FUAs.[45] Using a large number of trees with a minimum node size helps mitigate overfitting concerns.

Figure 11: **Illustration of decision tree**



Note: Dependent variable Y is employment changes in 2018 relative to the pre-shock period, and the predicted immigration shock in 2018 is W. This decision tree uses a 1% random sample of the data.

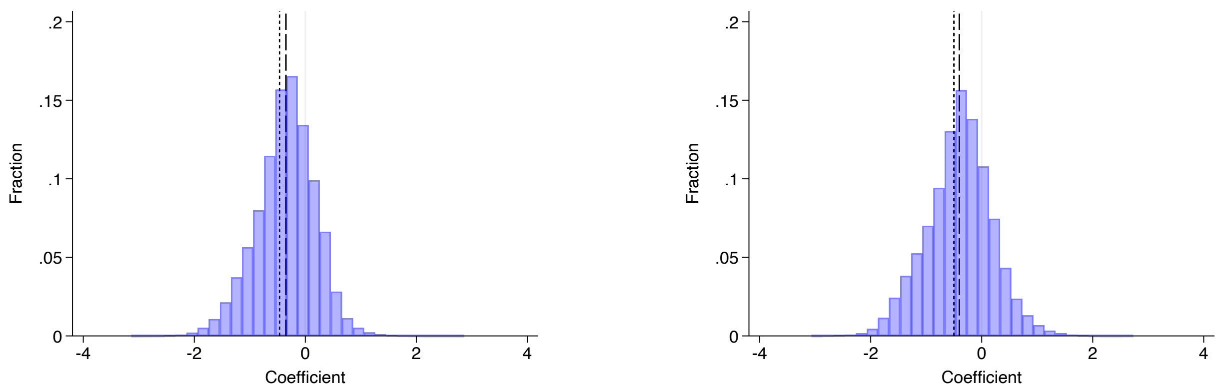
The first output of this procedure is shown in Figures 12a and 12b. These histograms display the predicted individual treatment effects for wages and employment, as determined by the trained causal forest. The treatment effects are derived from the OOB sample that is not used in the main algorithm. To estimate individual treatment effects, each OOB observation is assigned to a final node of each tree of the forest based on their characteristics. For all trained trees, the algorithm counts the times these observations fall into the same terminal node as the training sample to compute similarity weights. Using these weights, it calculates the weighted mean of τ across trees to determine the individual treatment effect $\tau(x_i)$. In the histograms, the long dashed line represents the average individual treatment effect, while the short dashed line is the average treatment effect from the standard regression of Equation (1). For both outcomes, the average coefficient from the causal forest aligns with the standard regression, reflecting the accuracy of the

[45] I set the tunable parameters of the algorithm to their default values, including honest splitting, and chose a relatively small minimum node size for precision. In further cross-validation, results hold when I substantially increase the minimum node size.

average prediction.

Figure 12: **Histogram of treatment effects for formal employment and formal wages in the causal forest, 2015–2018**

(a) Employment (b) Wages



Note: The short dashed line refers to the coefficient from the benchmark specification, and the long dashed line refers to the average predicted treatment effects that are estimated with the trained causal forest using the OOB sample. The number of trees is 2,000. The sample is restricted to natives aged 25 to 55. I use clusters at the FUA level for the causal forest. The causal forest uses 50% of the main sample. The minimum node size is 300.

Next, I use the individual treatment effects estimated by the algorithm to identify which subgroups are most affected by immigration. For this exercise, I divide native workers into quintiles based on their treatment effects for employment and wages (quintile 1 represents the most negative effect, while quintile 5 represents the most positive effect). The goal is to compare characteristics between these quintiles rather than making inference from the predicted individual treatment effects.

Tables 5a and 5b provide an overview of worker and firm characteristics in the pre-shock period. Firstly, native workers experiencing the most negative employment effects are generally older, have the lowest job tenure, and earn the lowest initial wages. These workers are also employed in the smallest firms and the lowest-paying firms. In contrast, those facing the most negative wage effects are relatively younger and earn the highest initial wages. The smallest firms also employ these workers, but in terms of pay premiums, they work in middle- to high-paying firms.[46] From a policy perspective, having the distribution of individual treatment effects is useful for designing targeted measures to mitigate the adverse impacts of immigration in the most affected subgroups.

[46]In Appendix Figures B.3a and B.3a, I verify that the quintiles of treatment effects from the causal forest align with the main regression. Notably, the estimates follow the same order across both wages and employment.

Table 5: **Descriptive statistics for native workers by quintiles of treatment effects**

(a) Formal employment

	Q1	Q2	Q3	Q4	Q5
Male (%)	0.7	0.6	0.5	0.5	0.5
Age of worker	42.8	40.3	38.5	35.1	31.1
Job tenure (1-9 years)	2.3	3.6	4.4	4.1	2.8
Monthly wages (USD)	324.8	462.6	521.8	478.4	336.2
Median firm size	79	105	276	510	1109
Quantiles of firm FEs (1-7)	3.8	5.3	6.0	6.3	6.5

(b) Formal wages

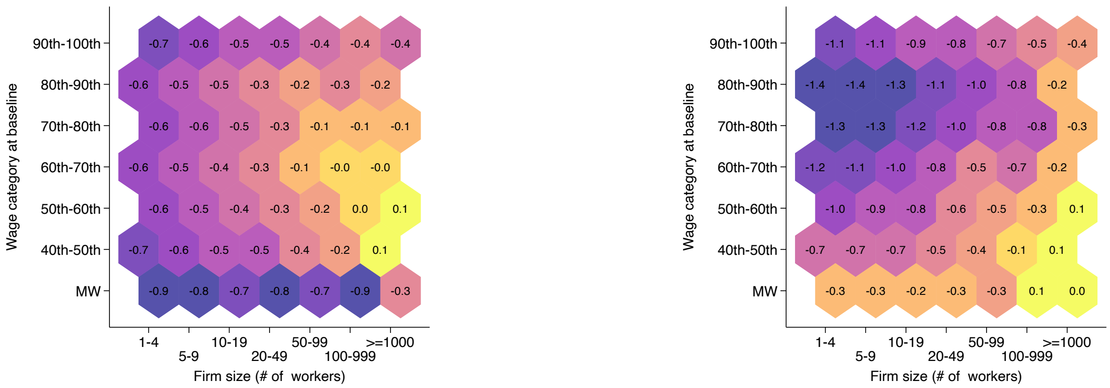
	Q1	Q2	Q3	Q4	Q5
Male (%)	0.6	0.6	0.6	0.6	0.5
Age of worker	36.6	38.5	38.8	38.1	37.5
Job tenure (1-9 years)	3.2	3.9	4.0	3.8	3.5
Monthly wages (USD)	559.5	466.2	419.3	379.0	393.7
Median firm size	86	189	242	309	892
Quantiles of firm FEs (1-7)	5.7	5.8	5.6	5.5	5.5

Note: These tables report the average or median statistics for quintiles of treatment effects (Q1 is the most affected and Q5 is the least affected) in terms of employment and wages, according to the predictions of the trained causal forest using the OOB sample. The wages are transformed from Colombian pesos to USD using 2020 exchange rates from the World Bank. Source: PILA, August 2015.

To better illustrate which subgroups are most affected by immigration, I construct heat plots that visualize the average individual treatment effects stratified by baseline wages and firm size. Figures 13a and 13b show an overview of the effects across these two dimensions. The most negative impacts on employment are concentrated among minimum-wage earners in small and medium-sized firms. In contrast, the most negative wage impacts are concentrated in high-wage workers and smoothly disappear as the worker's firm gets bigger. Hence, both employment and wages have the most negative impacts in smaller firms but differ along the wage distribution.

Figure 13: **Heat plot of treatment effects by wage category at baseline and firm size, 2015–2018**

(a) Employment (b) Wages

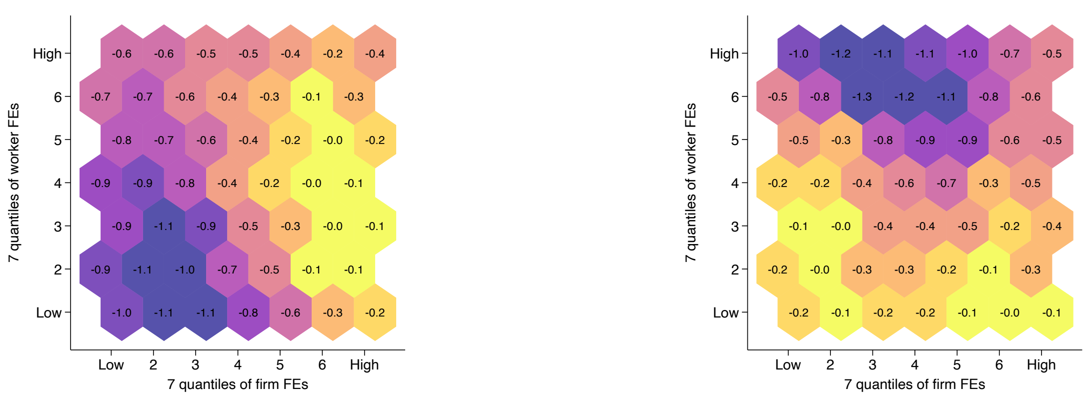


Note: Each hexagon is the average of individual treatment effects in the subgroup according to the trained causal forest using the OOB sample. The outcomes are the difference in individual employment or wages in 2018 relative to the pre-shock period, with the predicted immigration share as the treatment. The sample is restricted to natives aged 25 to 55. I use clusters at the FUA level for the causal forest. The causal forest uses 50% of the main sample.

Next, Figures 14a and 14b display average treatment effects by quantiles of firm FEs intersected with quantiles of worker FEs. Interestingly, the most negative employment effects are concentrated among low-wage workers in the lowest-paying firms. Conversely, the most negative wage effects tend to be concentrated in high-wage workers in middle-paying firms.

Figure 14: **Heat plot of treatment effects by quantiles of workers and firm FEs, 2015–2018**

(a) Employment (b) Wages



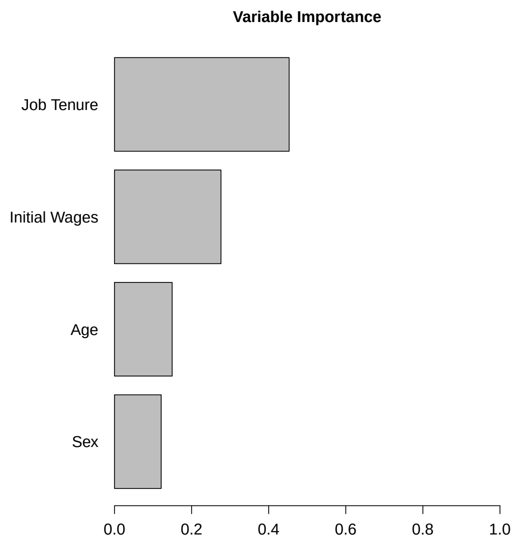
Note: Each hexagon is the average of individual treatment effects in the subgroup according to the trained causal forest using the OOB sample. The outcomes are the difference in individual employment or wages in 2018 relative to the pre-shock period, with the predicted immigration share as the treatment. The sample is restricted to natives aged 25 to 55. I use clusters at the FUA level for the causal forest. The causal forest uses 50% of the main sample.

A complementary way to summarize these findings is by using the variable importance measure. In this context, variables that frequently appear more as splits in the causal forest are categorized as more important for explaining treatment effect heterogeneity. This simple measure provides a ranking that can serve as a proxy to classify the sources of such heterogeneity. To start, I perform the algorithm both excluding and including firms' variables to illustrate differences in the importance measure. I use as an outcome the individual change in employment between 2018 and the average pre-shock period employment. When excluding firms' variables, I find that job tenure, followed by initial wages and age, are more important for explaining the heterogeneity in employment impacts (see Figure 15a). However, when including firms' variables, firm-specific pay premiums or firm FEs become the most important variable, followed by age and firm size. It is important to note this measure does not indicate the sign or magnitude of each variable's effect on employment but rather highlights which variable explains most of the heterogeneity in treatment effects. This emphasizes the relevant role of firms in the heterogeneity of immigration effects on natives, aligning with the findings by Arellano-Bover and San (2024), which shows the importance of firms and job mobility in the assimilation of immigrants in the labor market.

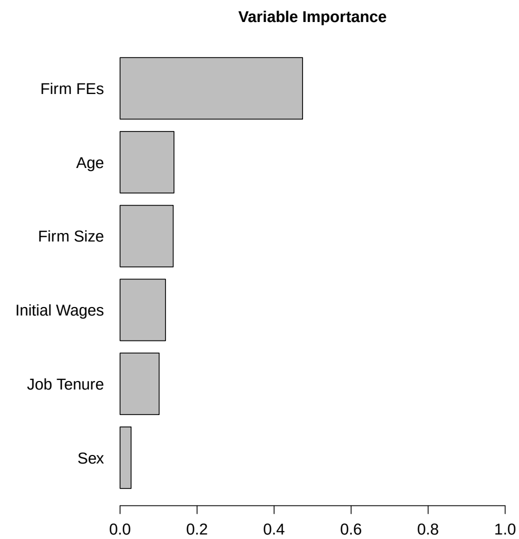
In a similar exercise for wage impacts, I use the individual wage growth between 2018 and 2015 as the outcome. Without incorporating firms' variables, initial wages are the most important variable, followed by age and job tenure (see Figure 15c). However, when including firms' variables in the causal forest, firm-specific pay premiums followed by initial wages and firm size become the most important (see Figure 15d). Note that firm pay premiums and firm size are positively correlated, though not so strongly (correlation coefficient of 0.19). Overall, the variable importance analysis reveals that firm-specific pay premiums and firm size are often more important than other worker characteristics in explaining the heterogeneity of wage and employment impacts. Firm FEs appear in 37% of all splits for wages and 30% for employment in the causal forest.

Figure 15: **Variable Importance for Formal Employment and Wages in the Causal Forest, 2015–2018**

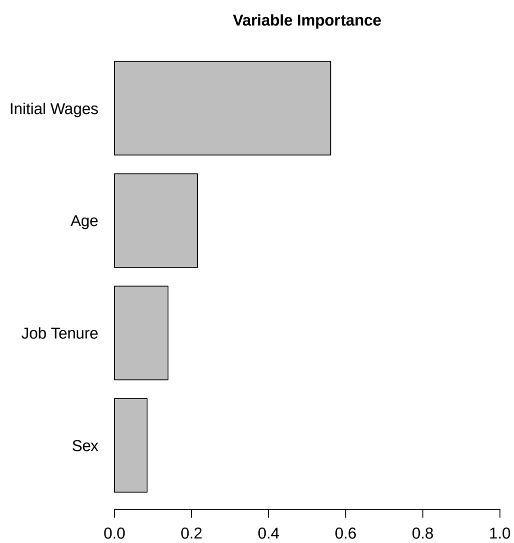
(a) Employment: Without Firm Variables



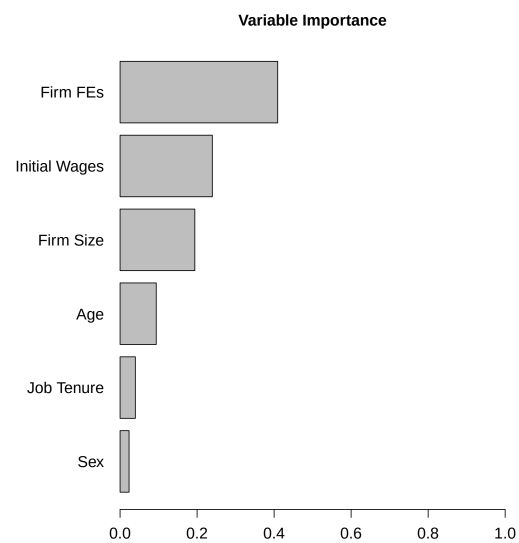
(b) Employment: With Firm Variables



(c) Wages: Without Firm Variables



(d) Wages: With Firm Variables



Note: Variable importance is a weighted sum of how often the feature f appears in the splits across the leaves of each tree in the forest. For this exercise, the forest comprises 2,000 trees. The analysis sample is restricted to native employees aged 25 to 55. The importance measure is normalized to sum to 1. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300.

Finally, given that initial wages are a function of the unobserved firm and worker FEs, I include the constructed worker FEs $\hat{\alpha}_i$ instead of initial wages into the algorithm. This adjustment reduces the sample size as every worker must be observed more than once. After including worker FEs, firm FEs remain the most important variable in explaining the heterogeneity of treatment effects for both employment and wages (see Appendix Figures A.8a and A.8b). Thus, firms' importance in the impact of immigration persists even after controlling for the quality of workers.

7 Robustness Checks

First, the exclusion restriction of the distance instrument might fail as border areas are more prone to time-varying shocks due to the Venezuelan crisis, such as trade shocks. Hence, I remove these border areas from the estimation sample, and I find more negative point estimates for employment and wages, but they become insignificant for wages (see Appendix Table B.1, row 2). Next, another concern is the concentration of employment in Bogotá (the capital of Colombia), as it accounts for 32.7% of the main sample. Hence, I exclude it from the analysis and find that while coefficients become less negative, especially for wages, they remain significant (see Appendix Table B.1, row 3).

Second, to improve the comparison of workers across local labor markets, I include wage categories based on the local wage distribution in the main regression as controls. Reassuringly, results remain similar for wages and employment, being more negative for wages. For the next robustness check, I adjust the nominal wages to real terms using the national CPI and find that wage effects are slightly less negative. Lastly, to exclude wage outliers, I top-code wages above the 99th percentile, finding that estimates remain consistent.

Third, I perform robustness checks for the machine learning method. Firstly, firm pay-premiums are correlated with the type of industry the firm belongs to, reflecting that some industries generally have higher or lower wage premia (Card et al., 2024). To address this, I include in the algorithm the industry of the firm, along with the firm FEs, to find that for wages and employment, the most important variable is still the firm FEs (see Appendix Figures B.5a and B.5a).[47] Secondly, to address the critique that the frequency of splits in the early nodes of trees is weighted as heavily as those in later nodes (where the sample size is smaller), I use a decay exponent in the variable importance measure. This approach prioritizes splits selected earlier in the tree-building process.[48]

[47]Due to possible misspecification in the 4-digit industry codes in PILA, I use a coarser definition of 19 industries based on ISIC revision 4.

[48]The decay exponent is -2, meaning that split frequencies in node k are weighted 1/2 compared to those in node $k-1$.

Here, the order of the variable importance is fairly similar for wages and employment, with firm size becoming the second most important variable after firm FEs. Interestingly, both variables capture the role of firms (see Appendix Figures B.4a and B.4a).

Last, the number of possible variable values can alter the variable importance ranking (Strobl et al., 2007). For instance, if a variable has a smaller set of values, it can mechanically appear in fewer nodes further in the tree. Therefore, I transform all continuous variables into categorical variables with six or seven categories. Appendix Figures B.6a and B.6b show that the order is similar for employment but for wages changes slightly, with firm FEs ranked second. Despite this, the benefit of the algorithm lies in exploiting the full range of a variable's values to allow for non-linear and interaction effects rather than arbitrarily aggregating values into categories. Another critique is that the algorithm maximizes the squared difference in treatment effects without evaluating pre-trends for all subgroups. It assumes the instrument's strict exogeneity. To address this, I check pre-trends for several subgroups in Appendix E and find mostly insignificant estimates across worker and firm variables. Finally, recent statistical literature proposes hypothesis testing of variable importance measures in random forests (see, for instance, Hapfelmeier et al. (2023)). The main idea is to perform sequential permutation tests to obtain a $p-value$ of each variable in the algorithm. While relevant for this study, these tests have not been developed for causal forests and are infeasible in high-dimensional setups.

8 Conclusion

This paper represents the first study to examine an immigration-induced supply shock in a developing country equipped with administrative data that covers the universe of formal workers and firms. This is an advantage in several dimensions. First, administrative panel data allows to track workers over time, addressing compositional changes that typically arise in the standard regional-level analysis of immigration based on cross-sectional surveys. Second, the matched employee-employer structure of the data enables to uncover heterogeneity across both firm and worker characteristics, which is key to understand the main mechanisms at play after immigration shocks. Third, the full count of formal firms, combined with a machine learning method, facilitates the construction of a proxy measure for the role of firms in the impact of immigration on workers.

Altogether, the findings indicate that the arrival of immigrants negatively impacts individual employment and wages in the formal sector. However, this coefficient masks heterogeneous responses. Specifically, minimum-wage workers are crowded out from the formal sector, whereas workers higher up in the wage distribution are not displaced but instead experience negative wage growth. Focusing on firm characteristics, the negative effect on employment and wages is concentrated in small formal firms. This result is consistent with the theoretical model, which predicts that small firms relying more on informal labor for production will reduce formal employment and wages more heavily following an immigration shock if formal and informal labor are substitutes. Beyond firm size, firm pay premiums also play a role in explaining immigration effects, as workers in low-paying firms experience a more negative effect on their earnings.

There are substantial heterogeneous effects along other worker and firm dimensions, so I use causal forests to identify which variable is most important to explain the heterogeneity in employment and wage effects. Throughout this analysis, firm-specific pay premiums appear prominently as the most important variable for heterogeneity, followed by firm size in most cases. In summary, focusing solely on workers' characteristics when analyzing the labor market impacts of immigration can lead to an incomplete perspective of the sources of adjustments. Suggesting that after immigrants arrive, the focus should not be only on *who* the worker is, but also on *type of firm* where the worker is employed.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

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.

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

Amior, M. and Stuhler, J. (2022). Immigration and monopsony: Evidence across the distribution of firms. Technical report, Working paper.

Arellano-Bover, J. and San, S. (2024). The role of firms and job mobility in the assimilation of immigrants: Former soviet union jews in israel 1990-2019.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.

Autor, D. H., Dorn, D., Hanson, G. H., and Song, J. (2014). Trade adjustment: Worker-level evidence. *The Quarterly Journal of Economics*, 129(4):1799–1860.

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

Bagger, J. and Lentz, R. (2019). An empirical model of wage dispersion with sorting. *The Review of Economic Studies*, 86(1):153–190.

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.

Beerli, A., Ruffner, J., Siegenthaler, M., and Peri, G. (2021). The abolition of immigration restrictions and the performance of firms and workers: evidence from switzerland. *American Economic Review*, 111(3):976–1012.

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., and Setzler, B. (2023). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*, 41(2):291–322.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739.

Bonilla-Mejía, L., Morales, L. F., Hermida, D., and Flórez, L. A. (2024). The labor market effect of south-to-south migration: Evidence from the venezuelan crisis. *International Migration Review*, 58(2):764–799.

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. and Edo, A. (2021). Gender, selection into employment, and the wage impact of immigration. Technical report, National Bureau of Economic Research.

Bratsberg, B. and Raaum, O. (2012). Immigration and wages: Evidence from construction. *The economic journal*, 122(565):1177–1205.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

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., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128(3):967–1015.

Card, D., Rothstein, J., and Yi, M. (2024). Industry wage differentials: A firm-based approach. *Journal of Labor Economics*, 42(S1):S11–S59.

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 nonwage compensation.

DANE (2019). Gran Encuesta Integrada de Hogares (GEIH)-Módulo de Migración. https://microdatos.dane.gov.co/index.php/catalog/641/get_microdata.

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. (2024). Immigration, wages, and employment under informal labor markets. *Journal of Population Economics*, 37(2):55.

Doran, K., Gelber, A., and Isen, A. (2022). The effects of high-skilled immigration policy on firms: Evidence from visa lotteries. *Journal of Political Economy*, 130(10):2501–2533.

Dostie, B., Li, J., Card, D., and Parent, D. (2021). Employer policies and the immigrant–native earnings gap. *Journal of Econometrics*.

Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., and Vom Berge, P. (2022). Reallocation effects of the minimum wage. *The Quarterly Journal of Economics*, 137(1):267–328.

Dustmann, C., Otten, S., Schönberg, U., and Stuhler, J. (2023). The Effects of Immigration on Places and Individuals - Identification and Interpretation. Technical report.

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.

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

Fort, T. C., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2013). How firms respond to business cycles: The role of firm age and firm size. *IMF Economic Review*, 61(3):520–559.

Groeger, A., León-Ciliotta, G., and Stillman, S. (2024). Immigration, labor markets and discrimination: Evidence from the venezuelan exodus in perú. *World Development*, 174:106437.

Gulyas, A., Pytka, K., et al. (2019). Understanding the sources of earnings losses after job displacement: A machine-learning approach. Technical report, University of Bonn and University of Mannheim, Germany.

Gyetvay, S. and Keita, S. (2023). Competition, mobility and immigration. Working paper.

Hapfelmeier, A., Hornung, R., and Haller, B. (2023). Efficient permutation testing of variable importance measures by the example of random forests. *Computational Statistics & Data Analysis*, page 107689.

Hicks, J. (1932). The theory of wages, london: Macmillan.

Hoen, M. F. (2020). Immigration and the tower of babel: Using language barriers to identify individual labor market effects of immigration. *Labour Economics*, 65:101834.

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.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components.

Econometrica, 88(5):1859–1898.

Kuosmanen, I. and Meriläinen, J. (2022). Labor market effects of open borders: Evidence from the finnish construction sector after eu enlargement. *Journal of Human Resources*.

Lebow, J. (2024). Immigration and occupational downgrading in colombia. *Journal of Development Economics*, 166:103164.

McKenzie, D. (2017). Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition. *American Economic Review*, 107(8):2278–2307.

Monras, J. (2020). Immigration and wage dynamics: Evidence from the mexican peso crisis. *Journal of Political Economy*, 128(8):3017–3089.

Muñoz, M. (2024). Trading nontradables: The implications of europe's job-posting policy. *The Quarterly Journal of Economics*, 139(1):235–304.

Orefice, G. and Peri, G. (2024). Immigration and worker-firm matching. *Review of Economics and Statistics*, pages 1–45.

Ortega, J. and Verdugo, G. (2022). Who stays and who leaves? immigration and the selection of natives across locations. *Journal of Economic Geography*, 22(2):221–260.

Redondo, H. (2022). From bricklayers to waiters: Reallocation in a deep recession.

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

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1):1–21.

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

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

UNHCR (2023). Mid-year trends 2023. https://www.unhcr.org/sites/default/files/2023-10/Mid-year-trends-2023.pdf.

Yagan, D. (2019). Employment hysteresis from the great recession. *Journal of Political Economy*, 127(5):2505–2558.

Yakymovych, Y., Nordström Skans, O., Vikström, J., Simon, L., and Athey, S. (2022). Worker attributes, aggregate conditions and theimpact of adverse labor market shocks.

Online Appendix

A Additional Results

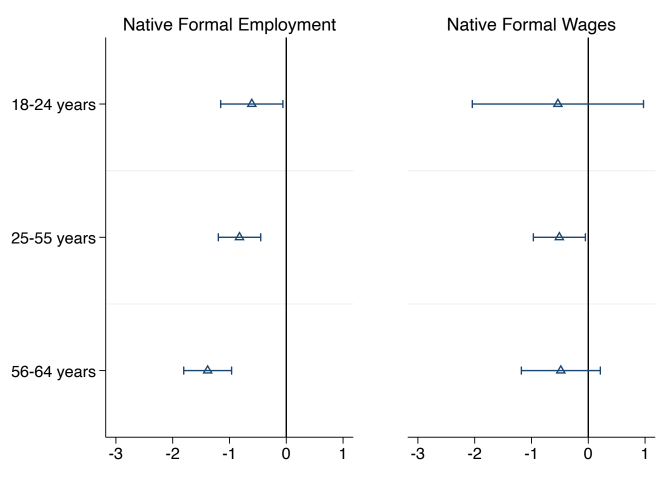
Table A.1: **First stage: The inflow of Venezuelan immigrants and the two instruments**

	(1)	(2)	(3)
	$\Delta M_{l,2018}$	$\Delta M_{l,2018}$	$\Delta M_{l,2018}$
Distance (/100)	-1.992***		-1.455***
	(0.272)		(0.350)
Distance (/100) squared	0.151***		0.107***
	(0.024)		(0.029)
Past settlements		0.703***	0.280*
		(0.160)	(0.130)
Constant	6.762***	1.040***	5.184***
	(0.715)	(0.149)	(1.000)
R^2	0.583	0.450	0.618
F	34.53	19.37	23.68
N	109	109	109

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

Note: This table reports the coefficient of the first-stage of the share of employed migrants $\Delta M_{l,2018} * 100$ with distance and distance squared to the nearest crossing bridge and past settlements as explanatory variables.

Figure A.1: **Estimates by extended age categories, 2015–2018**



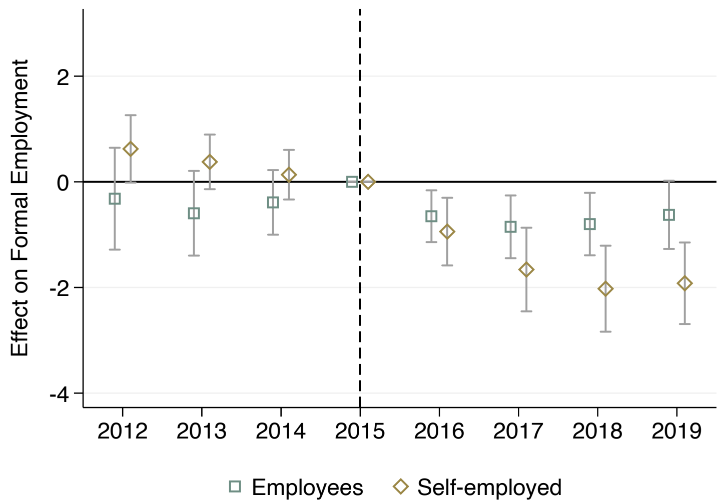
Note: I estimate Equation (1) separately by subgroups. The sample is restricted to natives between 18 and 64 years old. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. I use as controls sex with a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for employment (in percentage points) and wages (in percent) are already multiplied by 100. Workers are observed in August of each year. Source: PILA, 2013–2019.

Table A.2: **Labor supply elasticities by age group**

Age group	25-30	30-35	35-40	40-45	45-50	50-55
η_w^s	0.42	0.99	1.39	1.67	2.76	3.67

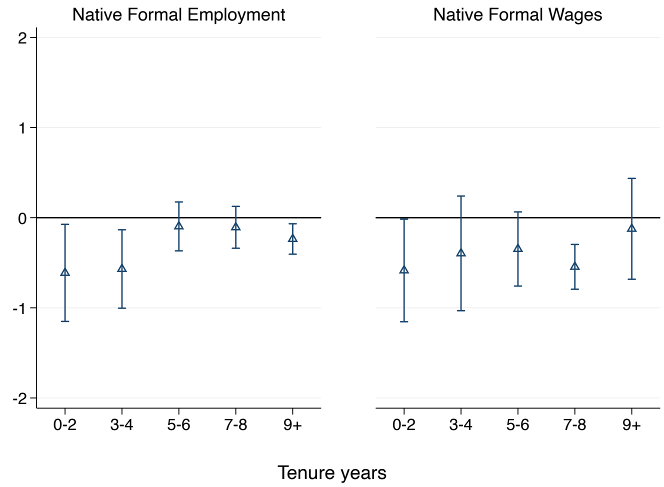
Note: The elasticity of labor supply is given by the reduced-form results from changes in native employment over changes in native wages in 2018.

Figure A.2: **Event study estimates on employment by job type**



Note: I estimate Equation (1) separately by year and characteristic. The sample is restricted to natives aged 25 to 55. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for employment (in percentage points) are already multiplied by 100. Workers are observed in August of each year. Source: PILA 2012–2019.

Figure A.3: **Estimates by job tenure, 2015–2018**



Note: I estimate Equation (1) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old. Dependent variables are employment and wages relative to the base period. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for employment (in percentage points) and wages (in percent) are already multiplied by 100. Workers are observed in August of each year. Source: PILA, 2013–2019.

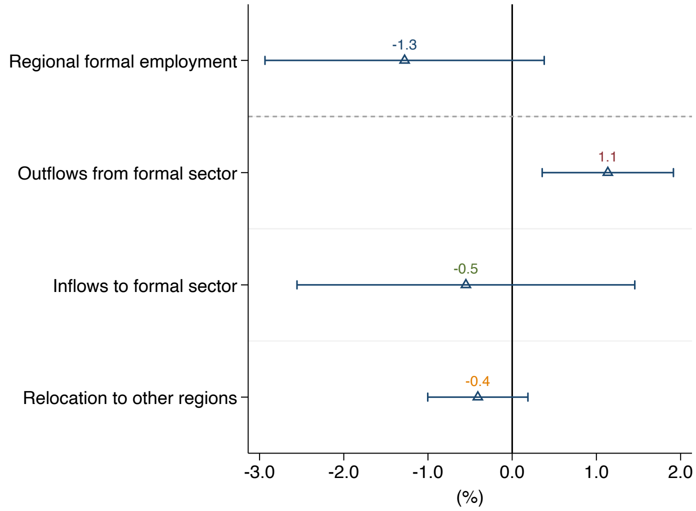
Table A.3: **Employment and wage estimates by age and job tenure, 2015–2018**

Worker's age	Below 35 years		Above 35 years	
Job tenure	0 to 4 years	5 to 9+ years	0 to 4 years	5 to 9+ years
Prob. of Employment	-0.138	0.209	-1.009**	-0.302***
	(0.195)	(0.226)	(0.315)	(0.086)
N	2,099,147	344,156	2,075,913	1,083,435
Wages	-0.479	-0.664*	-0.556	-0.194
	(0.344)	(0.279)	(0.354)	(0.182)
N	1,094,691	240,058	1,170,322	785,839
Clusters	109	109	109	109

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

Note: I estimate Equation (1) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. Workers are observed in August of each year. Source: PILA, 2013–2018.

Figure A.4: **Decomposition of formal employment, 2015–2018**



Note: Regressions are estimated at the regional level for 109 FUAs weighted by their formal employment in 2015. 95% confidence interval. The sample is not restricted by age groups. Regional formal employment is decomposed into outflows from formal employment in that region, inflows from non-employment or the informal sector, employed people in other regions, and relocation of formal workers to other regions. Source: PILA, 2015–2018.

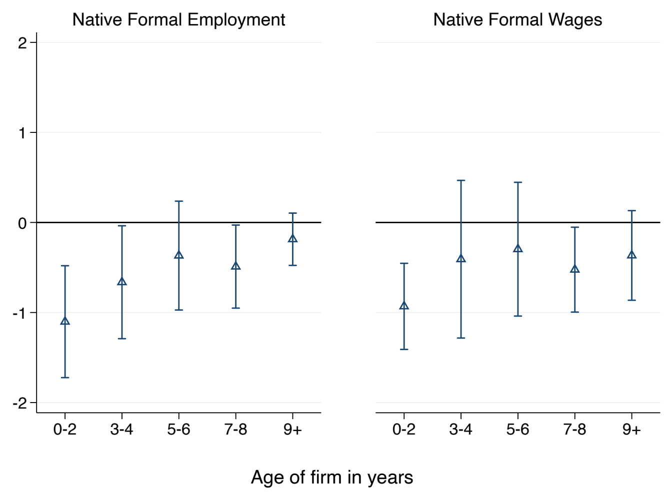
Table A.4: **IV estimates on regional changes of formal workers by age group, 2015–2018**

Age group	25-30	30-35	35-40	40-45	45-50	50-55
Prob. of changing region	0.200	0.088	-0.035	-0.156	-0.211	-0.254
	(0.400)	(0.404)	(0.354)	(0.307)	(0.266)	(0.209)
N	1,255,301	1,041,726	873,437	732,208	674,945	561,949
Clusters	109	109	109	109	109	109

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

Note: The outcome variable is an indicator that takes value one for workers that changed region in 2018 relative to 2015, and zero otherwise. The sample is restricted to natives aged 25 to 55. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The PILA had a measurement error with the regional variable in 2018, so the worker's location in February 2020 (when the health ministry started to verify this information) is used for the workers who present this error. Workers are observed in August of each year. Source: PILA, 2015–2018.

Figure A.5: **Estimates by age of firm, 2015–2018**



Note: I estimate Equation (1) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. The firm's age is the number of years the firm appears discontinuously in PILA. I use as controls the interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for employment (in percentage points) and wages (in percent) are already multiplied by 100. Workers are observed in August of each year. Source: PILA, 2013–2019.

Table A.5: **Employment and wage estimates by firm size and age of firm, 2015–2018**

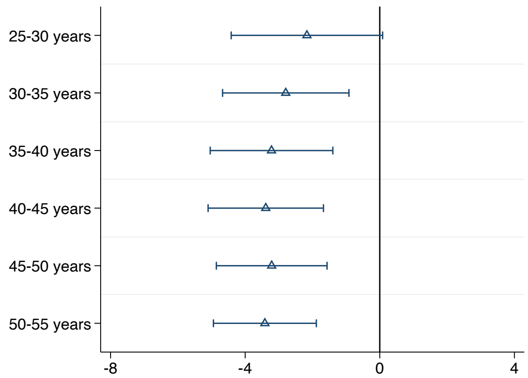
Firm's size	1 to 19 workers		Above 19 workers	
Age of firm	0 to 4 years	5 to 9+ years	0 to 4 years	5 to 9+ years
Prob. of Employment	-0.762**	-0.757***	-1.015**	-0.176
	(0.279)	(0.156)	(0.347)	(0.170)
N	479,715	498,842	923,272	3,700,822
Wages	-1.021*	-0.554	-0.603*	-0.395
	(0.432)	(0.305)	(0.304)	(0.286)
N	274,728	352,015	444,586	2,219,581
Clusters	109	109	109	109

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

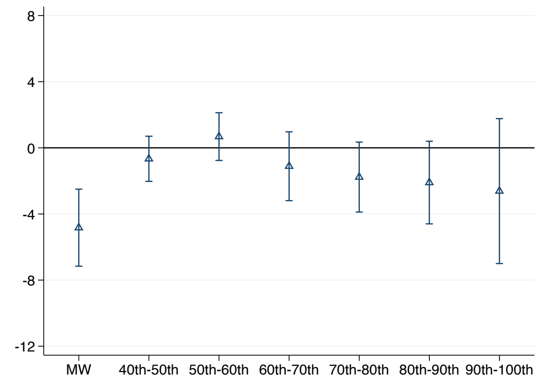
Note: I estimate Equation (1) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. The firm's age is the number of years the firm appears discontinuously in PILA. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. Workers are observed in August of each year. Source: PILA, 2013–2018.

Figure A.6: **Earnings estimates by different worker and firm characteristics, 2015–2018**

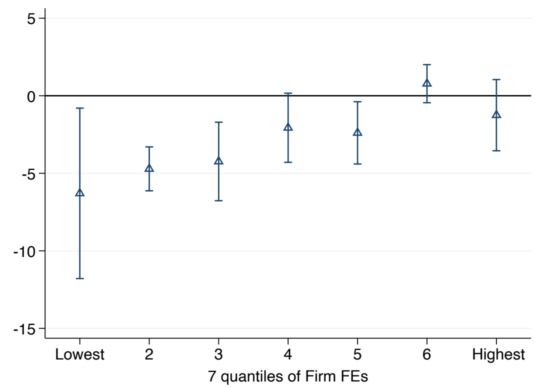
(a) By age groups



(b) By wage categories

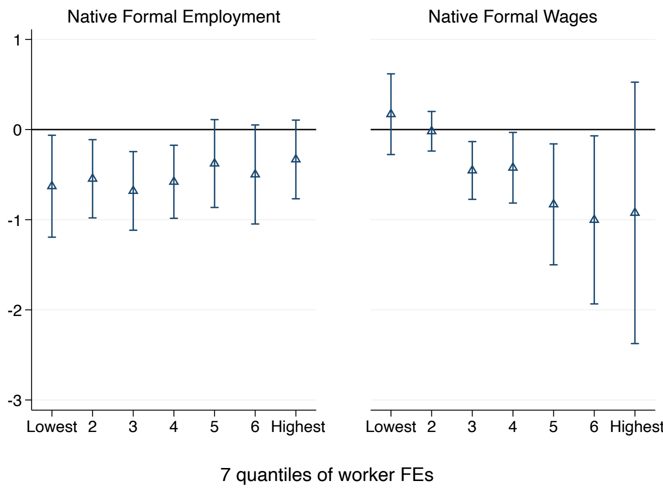


(c) By pay premiums



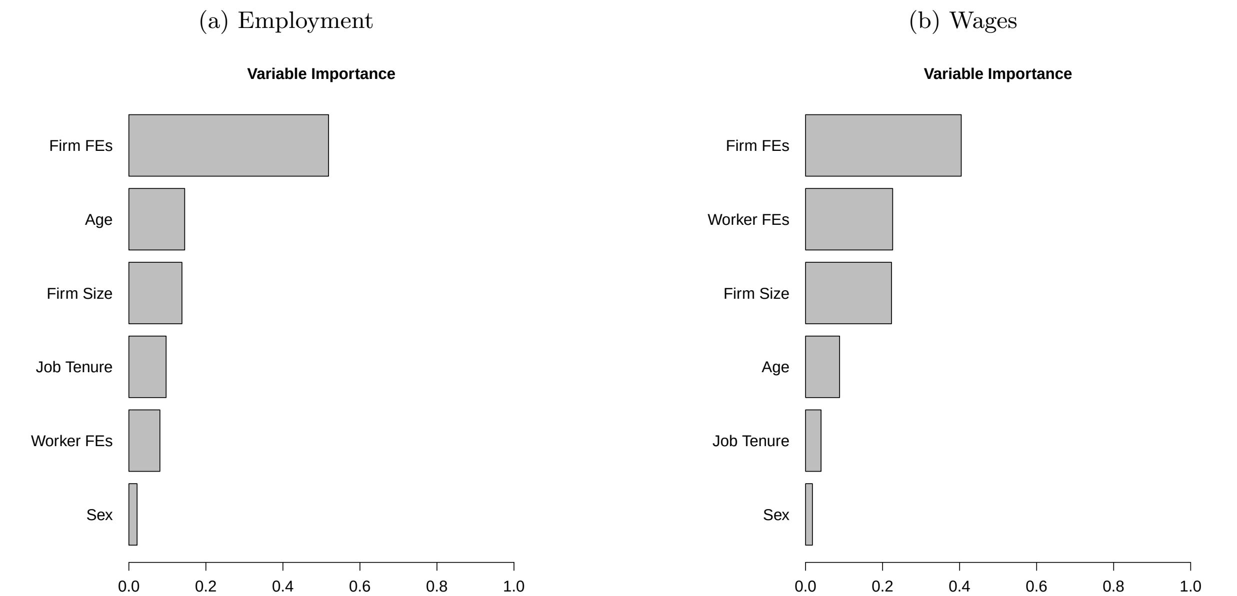
Note: I estimate Equation (1) separately by subgroups. The sample is restricted to natives aged 25 to 55. The dependent variable is cumulative earnings in the post-treatment period. I use as controls the interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The coefficients for earnings (in percent) is already multiplied by 100. Workers are observed in August of each year. Source: PILA, 2013–2019.

Figure A.7: **Estimates by quantiles of worker FEs, 2015–2018**



Note: I estimate Equation (1) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old who appear more than once in PILA. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. I compute Worker FEs in the first stage using the standard AKM framework, with age squared and its cubic as time-varying controls, for the period 2010-2015. I use as controls in the second stage are interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. The coefficients for employment (in percentage points) and for wages (in percent) are already multiplied by 100. Workers are observed in August of each year. 95% confidence interval. Source: PILA, 2013–2019.

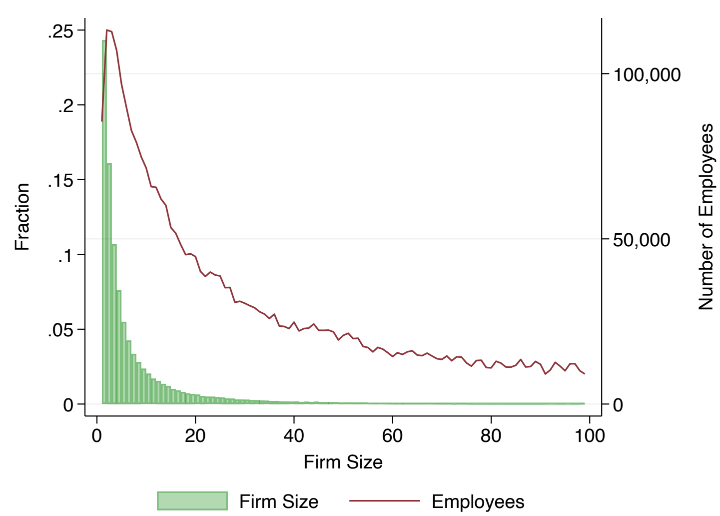
Figure A.8: **Variable importance for formal employment and formal wages in causal forest with worker and firm FEs, 2015–2018**



Note: Variable importance is a weighted sum of how often the feature f appears in the splits across the leaves of each tree in the forest. For this exercise, the forest comprises 2,000 trees. The analysis sample is restricted to native employees aged 25 to 55. The importance measure is normalized to sum to 1. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300.

B Robustness Checks

Figure B.1: **Firm size distribution and total employees**

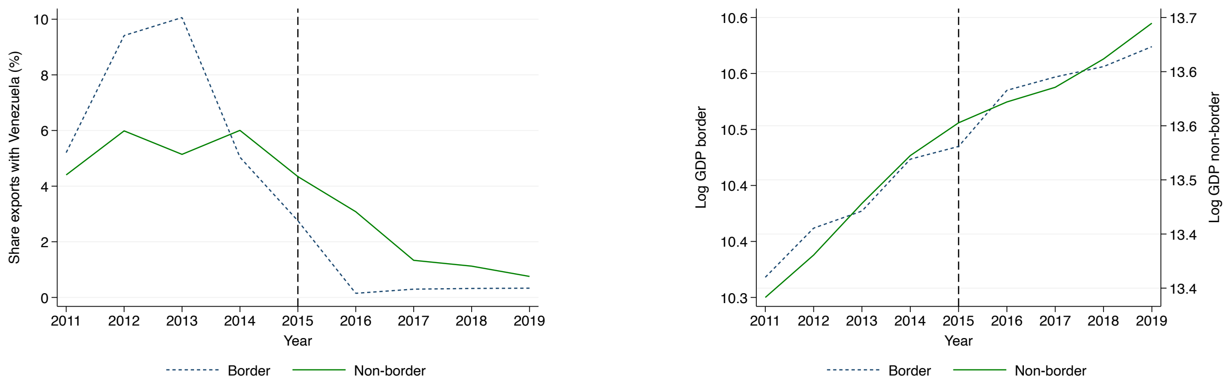


Note: The upper bound of firm size is restricted to 100 workers for the figure. The chosen bin width is 1. Only workers who contribute as employees are taken into account. Source: PILA, August 2015.

Figure B.2: **Evolution of trade and GDP for border and non-border departments**

(a) Share Exports with Venezuela

(b) Log GDP



Note: Border departments are *Norte de Santander, La Guajira, and César*. Non-border departments are the rest.
Source: Panel (a) Exportaciones-DANE, 2013–2019. Panel (b) DANE-Cuentas Nacionales, 2011–2019.

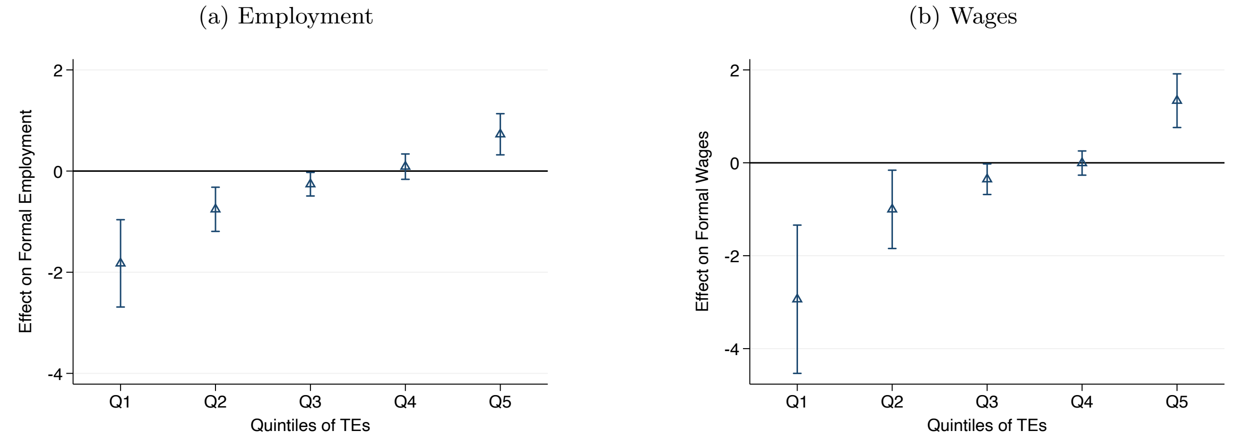
Table B.1: **Robustness checks for formal wages and formal employment, 2015–2018**

	Employment	Wages
Baseline	-0.841***	-0.600*
	(0.192)	(0.239)
N	6,706,035	4,090,973
Removing border areas*	-1.019*	-0.768
	(0.414)	(0.559)
N	6,577,923	4,015,648
Removing Bogotá	-0.777***	-0.470**
	(0.180)	(0.173)
N	4,338,192	2,619,237
Changing the denominator of ΔM_l with the 2005 census	-0.639***	-0.440*
	(0.168)	(0.203)
N	6,706,035	4,090,973
Further controls$\star$	-0.828***	-0.689*
	(0.176)	(0.326)
N	6,064,430	4,090,973
Real wages		-0.520*
		(0.207)
N		4,090,973
Top code local wages above 99%		-0.605*
		(0.241)
N		4,090,973

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

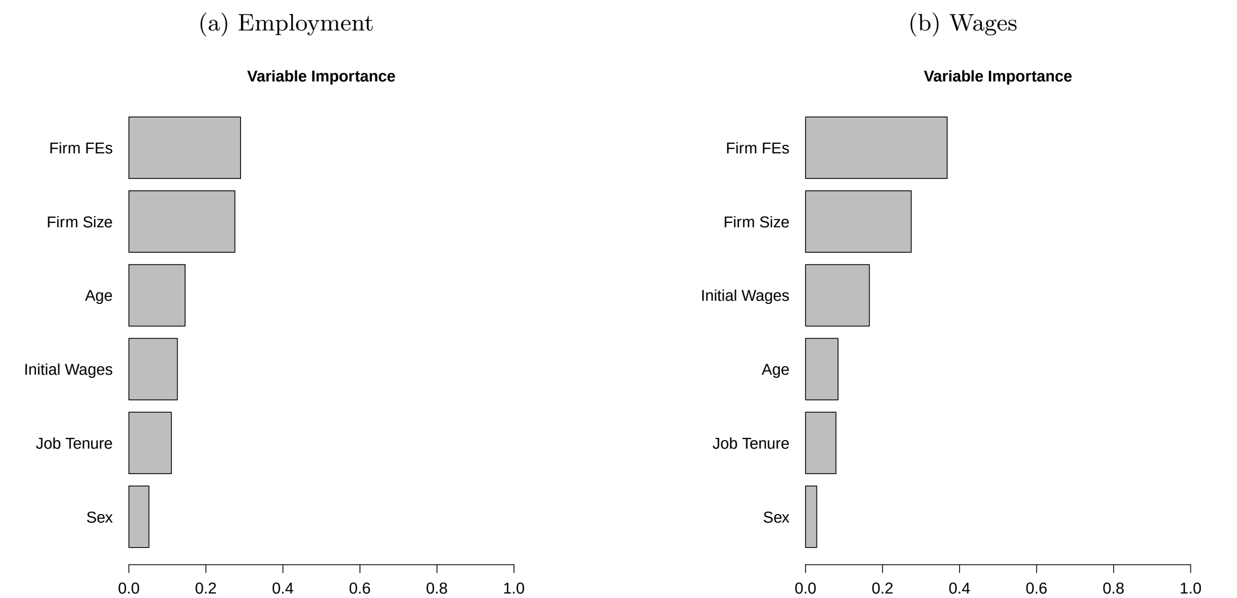
Note: This Table reports the coefficients of the second-stage regression of the instruments with the immigration rate $\Delta M_{l,2018}$. The outcome is the difference with the base period. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. *The border areas are Cucutá, Maicao and Arauca. $\star$ Further controls refer to dummies of seven wage quantiles. The sample is restricted to natives aged 25 to 55. Standard errors are clustered at the level of 109 geographic units. Workers are observed in August of each year. Source: PILA, 2015–2018.

Figure B.3: **Quintiles of treatment effects for formal employment and formal wages in the causal forest, 2015–2018**



Note: The quintiles of treatment effects are constructed using the individual treatment effects from the trained causal forest. The coefficients come from separate regressions of Equation (1). The sample is restricted to natives aged 25 to 55. I use clusters at the FUA level for the causal forest. Standard errors are clustered at the level of 109 geographic units. 95% confidence interval. The causal forest uses 50% of the main sample. The coefficients for employment (in percentage points) and wages (in percent) are already multiplied by 100.

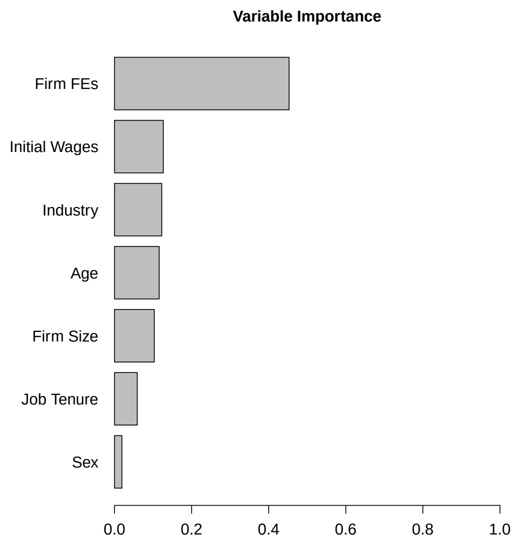
Figure B.4: **Variable importance for formal employment and formal wages in the causal forest with decay exponent, 2015–2018**



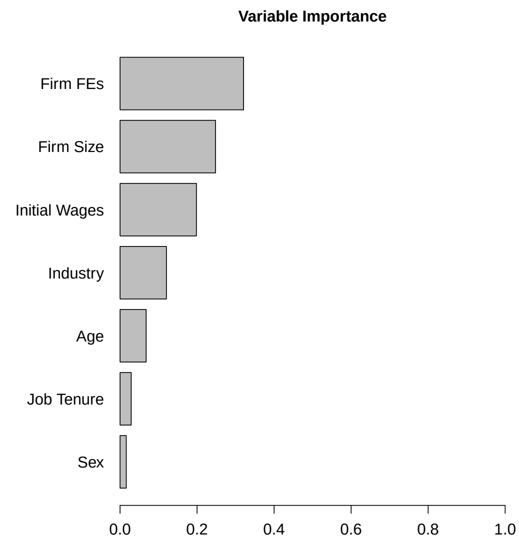
Note: Variable importance is a weighted sum of how often the feature f appears in the splits across the leaves of each tree in the forest. For this exercise, the forest comprises 2,000 trees. The analysis sample is restricted to native employees aged 25 to 55. The importance measure is normalized to sum to 1. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300.

Figure B.5: **Variable importance for formal employment and formal wages in causal forest with industry, 2015–2018**

(a) Employment

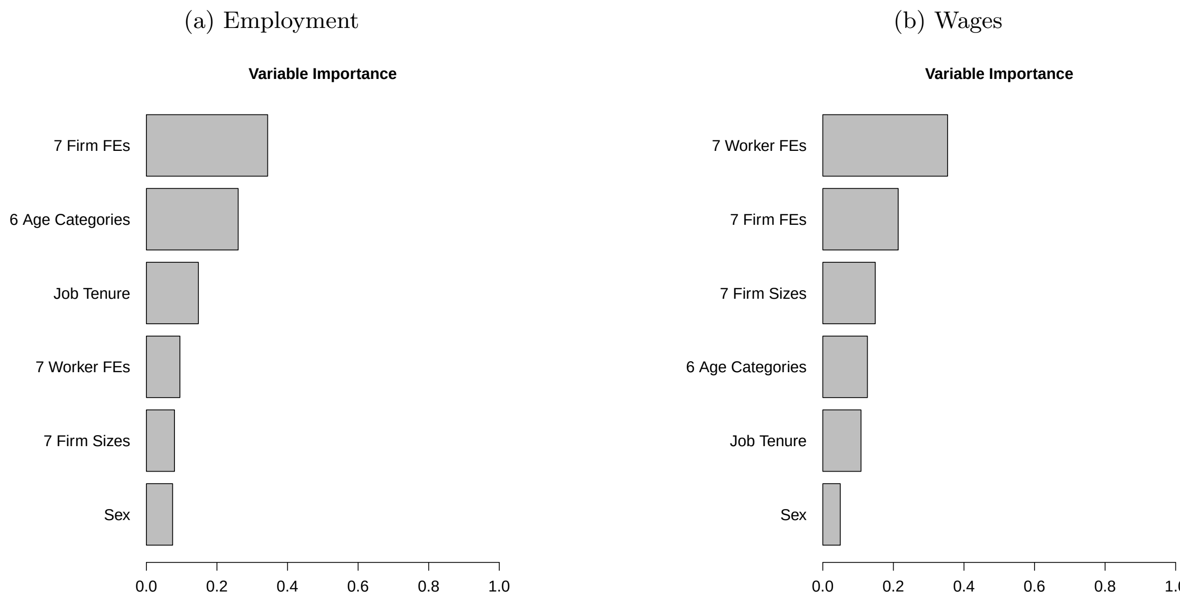


(b) Wages



Note: Variable importance is a weighted sum of how often the feature f appears in the splits across the leaves of each tree in the forest. For this exercise, the forest comprises 2,000 trees. The analysis sample is restricted to native employees aged 25 to 55. The importance measure is normalized to sum to 1. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300. The decay exponent is -2.

Figure B.6: **Variable importance for formal employment and formal wages in the causal forest for categories, 2015–2018**



Note: Variable importance is a weighted sum of how often the feature f appears in the splits across the leaves of each tree in the forest. For this exercise, the forest comprises 2,000 trees. The analysis sample is restricted to native employees aged 25 to 55. The importance measure is normalized to sum to 1. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300.

C Construction of the Sample

The administrative records of PILA are constructed at the contract level, as workers with more than one labor contract must pay contributions for each one. To transform to a worker-level dataset, first, I drop the contributions to the health system with type N, which are the ones that present corrections to their base income or changes to their labor status. Second, I sum the labor income for workers with multiple labor contracts and leave the job characteristics with the highest reported income for the worker.

To construct the sample for the AKM estimation, I restrict it to six years before the immigration shock to capture more movements of workers between firms. This sample uses the years 2010 to 2015 for August. The total sample consists of 32,195,048 worker-year observations after eliminating workers with non-positive wages, with less than 30 employment days per month, restricting to employees between 20 and 60 years, and leaving the highest wage job for workers with more than one contribution to the social security system.[C.1] Also, I exclude 3,931,843 additional workers because they do not belong to the largest connected set of firms and workers or appear only once in the estimation sample. Then, the nominal wages are transformed to real terms using the monthly CPI from DANE (with the base year 2018) and the logarithms of the final expression (lnw_{it}). Table C.1 shows descriptive statistics by the seven quantiles of firm FEs and Table C.2 shows the

[C.1] Around 5% of workers in PILA have more than one contribution.

decomposition of the variance of wages $Var(lnw_{it})$.

With this in mind, the definition of variables I use in the main analysis is as follows.

1. **Formal wages**. I use the nominal contribution to the health system of each worker in August. I only consider positive contributions, as zero indicates workers on leave for several reasons unrelated to wages or jobs. I focus on workers who reported 30 days of employment.

2. **Natives with formal employment**. I count all individuals who appear in PILA with a national identity card as natives. I take all the natives in the sample with a non-negative wage as employed.

3. **Firms**. I only leave workers classified as employees for the firm-level data and then aggregate by the firm identifier.

Table C.1: **Descriptive statistics by firm FEs**

	Average				
7 quantiles of $\hat{\psi}_j$	Employment	Male (%)	Age	Real wages (USD)	N
1	8	0.6	37.7	239.2	40,201
2	18	0.7	37.1	224.0	41,628
3	14	0.6	37.2	232.5	37,703
4	13	0.6	37.5	248.3	36,223
5	18	0.5	38.0	276.4	36,599
6	40	0.5	38.3	342.0	38,524
7	81	0.5	38.4	616.1	42,455

Note: This Table reports the descriptive statistics for different firm sizes recorded in PILA. Real wages are deflated using the CPI from DANE for prices in 2018. Colombian pesos to USD using 2020 exchange rates from the World Bank. Only workers who contribute as employees are taken into account. Source: PILA, August 2015.

Table C.2: Variance decomposition of lnw_{it}

Share of variance explained by:	
$Var(\alpha_i)$	50.2%
$Var(\psi_{j(i)})$	15.7%
$2Cov(\alpha_i, \psi_{j(i)})$	21.6%
$Corr(\alpha_i, \psi_{j(i)})$	.38

Note: This Table reports the variance decomposition of wages in the formal sector in Colombia using the largest connected set of workers and firms with the leave-out method proposed in Kline et al. (2020) with year FEs as the control variable. Source: PILA, August 2010–August 2015.

D Derivations of Model in 5.1

In this Appendix section, I explain the derivations of the Equations in subsection 5.1. First, to derive the firm-specific optimal wages, I maximize the profit Equation (9) for each type of worker:[D.1]

$$\frac{d\pi_j}{dI_j} = 0 \Leftrightarrow w_{I_j} = \left(\frac{\beta_I(1+\eta)}{1+\beta_I(1+\eta)}\right) D_j T_j^{\epsilon} \epsilon \alpha_I I_j^{\rho-1-\eta}(1+\eta)^{-1}(\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho})^{\frac{\epsilon-\rho}{\rho}}, \tag{D.1}$$

$$\frac{d\pi_j}{dF_j} = 0 \Leftrightarrow w_{F_j} = \left(\frac{\beta_F}{1+\beta_F}\right) D_j T_j^{\epsilon} \epsilon \alpha_F F_j^{\rho-1}(1+\tau_F)^{-1}(\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho})^{\frac{\epsilon-\rho}{\rho}}. \tag{D.2}$$

Here, workers' wages not only depend on their marginal productivity but also on the labor supply elasticities to the firm.[D.2] For clarity, l take logarithms of the wages in Equation (D.1) and (D.2):

$$lnw_{I_j} = ln\left(\frac{\beta_I(1+\eta)}{1+\beta_I(1+\eta)}\right) + ln(D_j T_j^{\epsilon}\epsilon\alpha_I) + (\rho-1-\eta)lnI_j - ln(1+\eta) + \left(\frac{\epsilon-\rho}{\rho}\right) ln(\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho}), \tag{D.3}$$

$$lnw_{F_j} = ln\left(\frac{\beta_F}{1+\beta_F}\right) + ln(D_j T_j^{\epsilon}\epsilon\alpha_F) + (\rho-1)lnF_j - ln(1+\tau_F) + \left(\frac{\epsilon-\rho}{\rho}\right) ln(\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho}). \tag{D.4}$$

In general, if I introduce a minimum wage for formal workers ($w_{F_{Min}}$) in this model such that $w_{F_{Min}} \leq w_{F_j}$, then formal workers must be paid the minimum wage and firms' optimal choices would be distorted. This is more likely to happen in low-productivity firms. Broadly, this model predicts firms with higher productivity (T_j) or demand (D_j) will pay higher wages, holding constant amenities. I then study how firm-level wages respond to an immigration shock that shifts the aggregate informal labor supply outwards ($d\mathcal{I}$)[D.3]:

$$\frac{dlnw_{I_j}}{d\mathcal{I}} \cdot \mathcal{I} = (\rho-1-\eta)\frac{dlnI_j}{dln\mathcal{I}} + \left(\frac{\epsilon-\rho}{\rho}\right)\frac{(\alpha_I \rho I_j^{\rho-1}\frac{dI_j}{d\mathcal{I}} + \alpha_F \rho F_j^{\rho-1}\frac{dF_j}{d\mathcal{I}})}{\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho}} * \mathcal{I}, \tag{D.5}$$

$$\frac{dlnw_{F_j}}{d\mathcal{I}} \cdot \mathcal{I} = (\rho-1)\frac{dlnF_j}{dln\mathcal{I}} + \left(\frac{\epsilon-\rho}{\rho}\right)\frac{(\alpha_I \rho I_j^{\rho-1}\frac{dI_j}{d\mathcal{I}} + \alpha_F \rho F_j^{\rho-1}\frac{dF_j}{d\mathcal{I}})}{\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho}} * \mathcal{I}. \tag{D.6}$$

Simplifying the last expressions and defining the derivates as the elasticities, I find that:

$$\varepsilon_{w_{I_j},\mathcal{I}} = -(1+\eta-\rho)\varepsilon_{I_j,\mathcal{I}} + (\epsilon-\rho)(s_{I_j}\varepsilon_{I_j,\mathcal{I}} + s_{F_j}\varepsilon_{F_j,\mathcal{I}}), \tag{D.7}$$

[D.1] In the derivations, I multiply by $\frac{w(L_j)}{w(L_j)}$ in the last term of FOCs to find the Equations on the text.

[D.2] If $\beta_L = 9$ then workers are paid 90% of their marginal productivity to the firm.

[D.3] Assuming that the supply shock does not affect the firm-specific demand and the firm-specific amenities for each group of workers. Besides, the number of firms is sufficiently large such that there are no strategic interactions between firms.

$$\varepsilon_{w_{F_j},\mathcal{I}} = -(1-\rho)\varepsilon_{F_j,\mathcal{I}} + (\epsilon-\rho)(s_{I_j}\varepsilon_{I_j,\mathcal{I}} + s_{F_j}\varepsilon_{F_j,\mathcal{I}}). \tag{D.8}$$

In these expressions, $s_{L_j} = \frac{\alpha_L L_j^\rho}{\alpha_I I_j^\rho + \alpha_F F_j^\rho}$ is the relative contribution of type of worker $L \in \{I, F\}$ to production. To further derive these elasticities, I use the changes in the firm-specific supply functions (7) and (8) after an immigration shock:

$$\varepsilon_{I_j,\mathcal{I}} = 1 + \beta_I \varepsilon_{w_{I_j},\mathcal{I}}, \tag{D.9}$$

$$\varepsilon_{F_j,\mathcal{I}} = \beta_F \varepsilon_{w_{F_j},\mathcal{I}}. \tag{D.10}$$

This yields a direct relationship between wages and employment as a function of the elasticities of supply to the firm.[D.4] Then, l replace Equations (D.9) and (D.10) into (D.7) and into (D.8):

$$\varepsilon_{w_{I_j},\mathcal{I}} = -(1+\eta-\rho)(1+\beta_I\varepsilon_{w_{I_j},\mathcal{I}}) + (\epsilon-\rho)(s_{I_j}(1+\beta_I\varepsilon_{w_{I_j},\mathcal{I}}) + s_{F_j}\beta_F\varepsilon_{w_{F_j},\mathcal{I}}), \tag{D.11}$$

$$\varepsilon_{w_{F_j},\mathcal{I}} = -(1-\rho)\beta_F\varepsilon_{w_{F_j},\mathcal{I}} + (\epsilon-\rho)(s_{I_j}(1+\beta_I\varepsilon_{w_{I_j},\mathcal{I}}) + s_{F_j}\beta_F\varepsilon_{w_{F_j},\mathcal{I}}). \tag{D.12}$$

Rearranging these expressions, I find that:

$$\varepsilon_{w_{I_j},\mathcal{I}} = \left(\frac{1}{\xi_{I_j}}\right)(-(1+\eta-\rho) + (\epsilon-\rho)(s_{I_j} + s_{F_j}\beta_F\varepsilon_{w_{F_j},\mathcal{I}})), \tag{D.13}$$

$$\varepsilon_{w_{F_j},\mathcal{I}} = \left(\frac{1}{\xi_{F_j}}\right)(\epsilon-\rho)s_{I_j}(1+\beta_I\varepsilon_{w_{I_j},\mathcal{I}}). \tag{D.14}$$

Here, I define $\xi_{I_j} = 1 + (1+\eta-\rho)\beta_I - (\epsilon-\rho)s_{I_j}\beta_I$ and $\xi_{F_j} = 1 + (1-\rho)\beta_F - (\epsilon-\rho)s_{F_j}\beta_F$. Then, replacing Equation (D.13) into (D.14) yields:

$$\varepsilon_{w_{F_j},\mathcal{I}} = \Omega_j s_{I_j}\beta_I(\epsilon-\rho)(\frac{\xi_{I_j}}{\beta_I} - (1+\eta-\rho) + (\epsilon-\rho)s_{I_j}). \tag{D.15}$$

Here, I define $\Omega_j = \frac{1}{\xi_{I_j}\xi_{F_j} - (\epsilon-\rho)^2 s_{I_j}\beta_I s_{F_j}\beta_F}$. Last, I replace ξ_{I_j} inside of (D.15) to find the Equation (10) in the main text. Next, I plug Equation (10) inside Equation (D.13) to find that:

$$\varepsilon_{w_{I_j},\mathcal{I}} = \left(\frac{1}{\xi_{I_j}}\right)(-(1+\eta-\rho) + (\epsilon-\rho)s_{I_j}(1+s_{F_j}\Omega_j(\epsilon-\rho)\beta_F)). \tag{D.16}$$

In this case, the elasticity is going to be negative $\varepsilon_{w_{I_j},\mathcal{I}} < 0$.[D.5] Finally, after finding that

[D.4] Here, the total number of formal workers $\mathcal{F}$ in the market is held constant. Besides, in this partial equilibrium framework, the response of one firm does not have spillover effects on other firms.

[D.5] To find that $\varepsilon_{w_{I_j},\mathcal{I}} < 0$ it is sufficient that $1 \geq s_{I_j}(1+s_{F_j}\Omega_j(\epsilon-\rho)\beta_F)$, which always happens when $\rho > \epsilon$. On the other hand, if $\rho < \epsilon$ then $\varepsilon_{w_{I_j},\mathcal{I}} < 0$ is also negative as $1+\eta-\rho > \epsilon-\rho$.

informal wages always decrease with the informal labor supply shock, the last adjustment to analyze is what happens to informal employment within the firm. For that, I plug Equation (D.16) into Equation (D.9):

$$\varepsilon_{I_j,\mathcal{I}} = 1 + \left(\frac{\beta_I}{\xi_{I_j}}\right)(-(1+\eta-\rho) + (\epsilon-\rho)s_{I_j}(1 + s_{F_j}\Omega_j(\epsilon-\rho)\beta_F)). \tag{D.17}$$

After simplifying the previous expression, I find that:

$$\varepsilon_{I_j,\mathcal{I}} = \frac{1}{\xi_{I_j}}(1 + (\epsilon-\rho)^2 s_{I_j}\beta_I s_{F_j}\beta_F\Omega_j). \tag{D.18}$$

Thus, in this case, a positive aggregate informal shock always increases informal labor within the firm ($\varepsilon_{I_j,\mathcal{I}} > 0$), independent of whether formal and informal workers' being close substitutes or not.

E Additional tests for Pre-Trends

This subsection of the Appendix tests for differential trends in outcomes according to the characteristics of workers and firms.

Table E.1: Event study estimates on pre-treatment periods of Figure 5

	Employment			Wages		
	2012	2013	2014	2012	2013	2014
25 to 30 years	-0.312	-0.458	-0.412	0.284	0.370	-0.019
	(0.702)	(0.655)	(0.428)	(0.461)	(0.234)	(0.124)
30 to 35 years	-0.249	-0.512	-0.305	0.104	0.226	-0.011
	(0.499)	(0.415)	(0.322)	(0.166)	(0.235)	(0.272)
35 to 40 years	-0.212	-0.381	-0.060	0.032	0.166	-0.011
	(0.364)	(0.327)	(0.196)	(0.296)	(0.253)	(0.273)
40 to 45 years	0.043	-0.155	-0.328	-0.012	-0.068	-0.501**
	(0.367)	(0.318)	(0.240)	(0.297)	(0.275)	(0.160)
45 to 50 years	0.191	-0.092	-0.101	0.110	0.566	-0.032
	(0.303)	(0.266)	(0.215)	(0.290)	(0.356)	(0.187)
50 to 55 years	0.121	0.005	0.103	0.413	0.369	-0.209
	(0.335)	(0.262)	(0.191)	(0.293)	(0.313)	(0.201)
Males	-0.449	-0.715	-0.473	0.272	0.317	-0.018
	(0.512)	(0.481)	(0.322)	(0.235)	(0.285)	(0.170)
Females	0.297	0.193	0.094	-0.011	0.199	-0.225*
	(0.376)	(0.310)	(0.223)	(0.209)	(0.178)	(0.095)

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

Note: I reduce the sample to a 10% random subsample of the entire dataset due to computational burden. The sample is restricted to natives aged 25 to 55. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. I observe workers in August of each year. Source: PILA, 2012–2015.

Table E.2: Event study estimates on pre-treatment periods of Figure 6

	Employment			Wages		
	2012	2013	2014	2012	2013	2014
Minimum wage	0.313	0.296	0.217	-0.538	-0.409	-0.526
	(0.269)	(0.213)	(0.181)	(0.825)	(0.700)	(0.435)
40th–50th	-0.295	-0.886	-0.190	0.262	0.190	-0.107
	(0.502)	(0.471)	(0.403)	(0.388)	(0.449)	(0.235)
50th–60th	-0.235	-0.766*	-0.284	0.058	0.234	-0.141
	(0.469)	(0.382)	(0.240)	(0.335)	(0.298)	(0.170)
60th–70th	-0.244	-0.100	-0.136	-0.360	0.401*	0.150
	(0.319)	(0.321)	(0.226)	(0.264)	(0.186)	(0.120)
70th–80th	-0.243	-0.475	-0.553**	0.918	0.730*	0.033
	(0.301)	(0.281)	(0.211)	(0.481)	(0.341)	(0.259)
80th–90th	-0.130	-0.432	-0.385*	0.435	0.367	-0.026
	(0.330)	(0.241)	(0.167)	(0.596)	(0.485)	(0.268)
90th–100th	0.330	-0.220	-0.146	-0.039	0.132	-0.477
	(0.483)	(0.173)	(0.136)	(0.288)	(0.269)	(0.297)

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

Note: I reduce the sample to a 10% random subsample of the entire dataset due to computational burden. The sample is restricted to natives aged 25 to 55. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. I observe workers in August of each year. Source: PILA, 2012–2015.

Table E.3: Event study estimates on pre-treatment periods of Figure 7

	Employment			Wages		
	2012	2013	2014	2012	2013	2014
1-4 workers	-0.070	-0.022	-0.193	0.263	0.476	0.264
	(0.353)	(0.456)	(0.368)	(0.534)	(0.389)	(0.262)
5-9 workers	-0.044	0.131	-0.484	-0.041	0.176	-0.088
	(0.481)	(0.433)	(0.360)	(0.300)	(0.570)	(0.222)
10-19 workers	-0.314	-0.352	-0.446	0.646	1.156**	0.240
	(0.736)	(0.493)	(0.322)	(0.500)	(0.356)	(0.188)
20-49 workers	-0.525	-0.573	-0.397	0.511*	0.638**	0.398*
	(0.607)	(0.622)	(0.384)	(0.220)	(0.213)	(0.177)
50-99 workers	-0.178	-0.565	-0.497	0.708**	0.877***	0.199
	(0.656)	(0.543)	(0.435)	(0.240)	(0.193)	(0.186)
100 to 999 workers	-0.168	-0.499	-0.211	0.648	0.583	-0.137
	(0.695)	(0.608)	(0.413)	(0.620)	(0.443)	(0.223)
More than 1000 workers	-0.239	-0.462	-0.168	0.202	0.465	0.134
	(0.478)	(0.474)	(0.362)	(0.390)	(0.344)	(0.212)

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

Note: I reduce the sample to a 10% random subsample of the entire dataset due to computational burden. The sample is restricted to native employees between 25 and 55 years old. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. I observe workers in August of each year. Source: PILA, 2012–2015.

Table E.4: Event study estimates on pre-treatment periods of Figure 8

	Employment			Wages		
	2012	2013	2014	2012	2013	2014
Lowest quantile	-0.170	-0.169	-0.217	0.121	0.352	-0.867
	(0.469)	(0.483)	(0.426)	(1.819)	(1.369)	(0.750)
2nd quantile	0.040	0.273	0.435	0.390	0.370	0.268
	(0.433)	(0.401)	(0.340)	(0.355)	(0.315)	(0.150)
3rd quantile	-0.695	-0.900	-1.053*	0.574**	0.369	0.049
	(0.548)	(0.537)	(0.457)	(0.201)	(0.230)	(0.120)
4th quantile	0.196	0.455	-0.084	-0.117	0.135	-0.248
	(0.482)	(0.438)	(0.272)	(0.235)	(0.277)	(0.140)
5th quantile	-0.470	-0.772	-0.466	0.480	0.645	0.094
	(0.462)	(0.451)	(0.287)	(0.447)	(0.407)	(0.188)
6th quantile	0.102	-0.385	0.119	0.143	0.383	0.244
	(0.416)	(0.430)	(0.292)	(0.221)	(0.249)	(0.204)
Highest quantile	-0.342	-0.650	-0.456	0.394	0.539	-0.063
	(0.378)	(0.362)	(0.263)	(0.446)	(0.341)	(0.157)

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

Note: I reduce the sample to a 10% random subsample of the entire dataset due to computational burden. The sample is restricted to native employees between 25 and 55 years old. Controls include interactions of sex with six age categories and a dummy for self-employed in the base period. Standard errors are clustered at the level of 109 geographic units. I observe workers in August of each year. Source: PILA, 2012–2015.

F Size and Location of FUAs

Table F.1: Number of observations by FUA I

	Observations	Percent
1. Bogotá	2,327,306	(32.7)
2. Medellín	983,096	(13.8)
3. Cali	593,447	(8.3)
4. Barranquilla	341,211	(4.8)
5. Cartagena	205,150	(2.9)
6. Bucaramanga	273,090	(3.8)
7. Cúcuta	110,123	(1.5)
8. Pereira	140,791	(2.0)
9. Ibagué	100,823	(1.4)
10. Manizales	103,401	(1.5)
11. Santa Marta	84,705	(1.2)
12. Pasto	70,170	(1.0)
13. Armenia	71,314	(1.0)
14. Villavicencio	106,493	(1.5)
15. Montería	71,007	(1.0)
16. Valledupar	76,072	(1.0)
17. Buenaventura	24,514	(0.3)
18. Neiva	71,376	(1.0)
19. Palmira	41,687	(0.6)
20. Popayán	62,422	(0.9)
21. Sincelejo	39,859	(0.6)
22. Barrancabermeja	35,095	(0.5)
23. Tuluá	25,123	(0.3)
24. Tunja	52,987	(0.7)
25. Riohacha	31,134	(0.4)
26. San Andres de Tumaco	7,960	(0.1)
27. Florencia	19,704	(0.3)
28. Apartadó	26,268	(0.4)
29. Giradot	14,920	(0.2)
30. Cartago	17,006	(0.2)
31. Maicao	6,263	(0.1)
32. Magangué	5,327	(0.1)
33. Sogamoso	18,220	(0.3)
34. Buga	21,072	(0.3)
35. Ipiales	8,754	(0.1)
36. Quibdó	15,687	(0.2)
37. Fusagasugá	12,899	(0.2)
38. Facatativá	18,796	(0.3)
39. Duitama	18,427	(0.3)
40. Yopal	43,279	(0.6)
41. Ciénaga	4,701	(0.1)
42. Zipaquirá	12,908	(0.2)
43. Rionegro	29,601	(0.4)
44. Ocaña	8,966	(0.1)
45. La Dorada	8,563	(0.1)
46. Caucasia	7,372	(0.1)
47. Sabanalarga	2,434	(0.03)
48. Aguachica	9,748	(0.1)
49. Espinal	6,439	(0.1)
50. Arauca	11,726	(0.2)
51. Santa Rosa de Cabal	4,887	(0.1)
52. El Carmen de Bolívar	1,411	(0.02)
53. Fundación	3,881	(0.1)
Continues in Table F.2		
No FUA assigned	417,188	(5.9)
Total	7,123,223	(100)

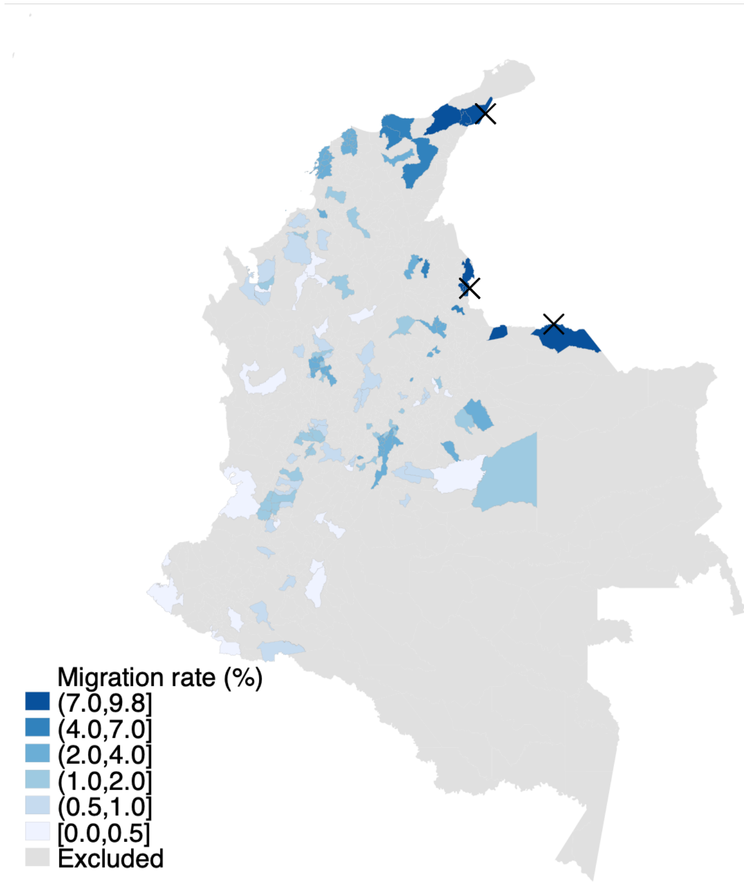
Note: This Table reports the number of workers from PILA by FUAs 1 to 53. The name represents the main city of FUA, but often they aggregate multiple municipalities according to Sanchez-Serra (2016). The sample is restricted to natives aged 25 to 55. Workers are observed in August of each year. Source: PILA, 2015.

Table F.2: Number of observations by FUA II

	Observations	Percent
54. Acacías	12,472	(0.2)
55. Madrid	8,922	(0.1)
56. La Ceja	8,662	(0.1)
57. Santander de Quilichao	8,505	(0.1)
58. San Gil	8,268	(0.1)
59. Mocoa	7,974	(0.1)
60. Pitalito	7,852	(0.1)
61. Albania	7,020	(0.1)
62. Tocancipá	7,007	(0.1)
63. Los Patios	6,137	(0.1)
64. Montelíbano	6,083	(0.1)
65. Turbo	5,830	(0.1)
66. Granada	5,298	(0.1)
67. El Carmen de Viboral	5,047	(0.1)
68. Chinchiná	4,903	(0.1)
69. Puerto Boyacá	4,761	(0.1)
70. Guarne	4,697	(0.1)
71. Zarzal	4,584	(0.1)
72. Puerto Asís	4,568	(0.1)
73. Chiquinquirá	4,526	(0.1)
74. Villa de San Diego de Ubaté	4,522	(0.1)
75. Garzón	4,454	(0.1)
76. Santa Rosa de Osos	4,406	(0.1)
77. Puerto Gaitán	4,380	(0.1)
78. Pamplona	4,348	(0.1)
79. Puerto Tejada	4,279	(0.1)
80. Caloto	4,136	(0.1)
81. Segovia	4,016	(0.1)
82. Puerto Berrío	3,989	(0.1)
83. Lorica	3,875	(0.1)
84. Sopó	3,832	(0.1)
85. Aguazul	3,627	(0.1)
86. Santa Fé de Antioquia	3,589	(0.1)
87. Cereté	3,526	(0.0)
88. Puerto López	3,412	(0.0)
89. Pradera	3,388	(0.0)
90. La Cruz	3,387	(0.0)
91. La Virginia	3,375	(0.0)
92. San Pedro de los Milagros	3,170	(0.0)
93. Tenjo	3,166	(0.0)
94. Villanueva	3,136	(0.0)
95. Sahagún	3,126	(0.0)
96. Melgar	3,099	(0.0)
97. Barbosa, Santander	3,042	(0.0)
98. Socorro	3,026	(0.0)
99. Carepa	2,999	(0.0)
100. Planeta Rica	2,893	(0.0)
101. Chigorodó	2,880	(0.0)
102. Yarumal	2,874	(0.0)
103. Paipa	2,873	(0.0)
104. Samacá	2,782	(0.0)
105. Barbosa, Antioquia	2,781	(0.0)
106. Saravena	2,730	(0.0)
107. El Cerrito	2,597	(0.0)
108. Amagá	2,534	(0.0)
109. Villeta	2,518	(0.0)

Note: This Table reports the number of workers from PILA by FUAs 54 to 109. The name represents the main municipality. The sample is restricted to natives aged 25 to 55. Workers are observed in August of each year. Source: PILA, 2015.

Figure F.1: **Map of FUAs with the immigration shock** $\Delta M_{l,2018}$



Note: The X represents the main three crossing bridges with Venezuela. The distance instrument is according to the nearest crossing bridge. Source: CNPV, 2018.