Worker Responses to Immigration Across Firms: Evidence from Colombia

Lukas Delgado-Prieto*

January, 2026

Abstract

The labor market effects of immigration depend on how firms adjust, yet this aspect remains unexplored in developing countries. This paper studies the mass influx of Venezuelan migrants into Colombia using employer-employee data. As immigrants land in informal employment, formal employment for minimum-wage natives falls, reflecting their substitutability with lower-cost informal workers. The negative effects are stronger in small formal firms, which rely more on informality. A machine learning analysis shows that firm-level factors explain more of the heterogeneity in worker-level impacts. These findings highlight that informality amplifies firms' role in shaping workers' immigration adjustments.

Keywords: Immigration, Minimum wages, 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, Jonas Hjort, Patrick Kline, Luigi Minale, Joan Monras, Jesse Rothstein, and Jon Piqueras for their comments and suggestions. I want to thank participants at the Workshop of Firm-Level Impacts of Labour Supply Shocks, the Oslo Trade Meeting Coffee, 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. This paper uses confidential administrative records on social security contributions obtained from Colombia's Ministry of Health. Due to confidentiality restrictions, the data cannot be publicly shared. This paper was awarded the Best Immigration Economics Paper at the OECD Immigration Conference 2023. This project received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101043127.
Department of Economics, University of Oslo, Norway. Email: laprieto@econ.uio.no.

1 Introduction

Over the past decade, several countries around the globe have experienced substantial population outflows (e.g., Afghanistan, Ukraine, Syria, and Venezuela). The majority of these migrants and refugees are moving to neighboring, developing countries. As of June 2023, seventy-five percent of the 110 million forcibly displaced individuals were hosted by low- and middle-income countries (UNHCR, 2023). The labor market in these contexts features the coexistence of both formal and informal labor in production (Ulyssea, 2018), as well as the prevalence of small businesses (McKenzie, 2017). Given that migrants disproportionately work informally and in small firms (Delgado-Prieto, 2024), it is likely that immigration effects on formal natives are shaped by the size of the firm they work for, beyond the traditional skill-substitution channel. In fact, there is a growing literature that documents the relevant role of firms when analyzing immigration adjustments in various developed countries (Amior and Stuhler, 2022; Beerli et al., 2021; Clemens and Lewis, 2022; Doran et al., 2022; Orefice and Peri, 2024). This paper shows that, in settings with high informality and binding minimum wages, firms play a more central role in determining how formal workers and the formal labor market respond to immigration shocks.

To do so, I examine the labor market impacts of one of the most significant episodes of immigration in recent history: the Venezuelan mass migration to Colombia, which brought more than 2 million migrants with skills similar to those of natives and who had access to work permits through massive regularization programs, although the vast majority end up working informally. By exploiting the uneven arrival of migrants across local labor markets and tracking workers over time, I quantify worker-level impacts and study heterogeneity over firm-level characteristics.[1] With the longitudinal administrative data, I cover the universe of formal workers and firms in the country, while I complement it with census information to measure the stock of migrants locally.[2] To address endogenous sorting of migrants into economically favorable areas, I construct two distinct

[1]I address more carefully with this data 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 leaves employment or relocates to other regions, the wage estimates are not accurately identified (Borjas and Edo, 2021; Dustmann et al., 2023).

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

instruments: past Venezuelan settlement locations and the local distance to the nearest crossing bridge between Colombia and Venezuela. Using these instruments in a difference-in-differences research design (DiD-IV), I find a negative impact on individual formal employment and native wages in the short- to medium-term (up to 4 years).

The negative employment effects are concentrated in small formal firms and are driven by workers earning the minimum wage, most likely the least educated within the formal sector. For these low-wage workers, a 1 percentage point (pp) increase in the immigration shock in a given labor market decreases the probability of formal sector employment by 1.5 pp. These workers are likely to transition towards informality, but may also move into unemployment or inactivity. In this context, the relatively high and binding minimum wage limits the space for downward wage adjustments and increases the risk of job displacement. Approximately 40% of all formal workers, including self-employed individuals, earned the minimum wage in 2015, and in small firms this share exceeds 50%. The negative impact on wages, in contrast, primarily affects native workers earning above the minimum wage, but again for those in the smallest formal firms. The concentration of immigrants in small firms, which face binding minimum-wage constraints and tend to hire more workers informally, suggests potential mechanisms for the observed findings. Still, it lacks a conceptual framework that can formally rationalize the underlying mechanisms.

Therefore, I construct a model of an imperfectly competitive labor market based on Card et al. (2018) that features heterogeneous firms, some of which pay a minimum wage. This model incorporates labor input costs similarly to Ulyssea (2018), while allowing for imperfect substitution of labor inputs, as in Delgado-Prieto (2024). The model indicates that aggregate substitutability between formal and informal workers in production must be high to observe negative responses in formal employment and wages.[3] It also reveals that firm-level responses to immigration depend on their reliance on informal labor for production, ultimately suggesting that smaller firms have more elastic formal labor demand than larger firms. Notably, the finding that workers in smaller firms experience more adverse impacts on formal employment than those in larger firms ultimately

[3]This follows 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).

suggests strong substitutability between formal and informal labor in small firms. The labor-labor substitution I point to can be a response to immigration shocks that may extend to other economic shocks and policy changes, such as increases in the minimum wage (Clemens, 2021).

Following the model's insights, I estimate the canonical "AKM" regression (Abowd et al., 1999) to construct firm-fixed effects (FEs) representing firm pay premiums or firm productivity, and worker FEs as a proxy for workers' constant skills portable across employers. In line with the model, I find that native workers in lower-paying firms are more adversely affected in terms of employment than those in high-paying firms. Regarding wages, a potential explanation for the observed negative effects is the reallocation of workers across firms, in which workers may move from high-paying to low-paying firms or vice versa (Gyetvay and Keita, 2023). Still, I find no differential sorting of native workers after the immigration shock, potentially due to institutional factors or economic conditions.[4] Last, I examine complementary firm-level outcomes that highlight other margins of adjustment to immigration. I find that firms poach fewer workers from other formal firms, especially those that, at baseline, recruit the least from them, consistent with the view that formal workers become relatively more expensive for some firms, leading them to hire more from outside the formal sector. Additionally, formal firm exit rates are higher in areas with higher migrant inflows, while formal firm creation rates remain unchanged up to the third post-treatment year.

I then leverage the large size of the matched employee-employer dataset to systematically estimate heterogeneous 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, thereby identifying the worker and firm variables that explain most of the heterogeneity in worker-level effects. Using this algorithm, I first identify the subgroups most affected by immigration, both in employment and wages. Then, based on the frequency of these variables across all decision trees in the causal forests, I construct a variable importance statistic. I show that, with this measure, firm pay premiums or firm size are consistently ranked higher, indicating they

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

explain more of the heterogeneity in employment and wage effects than worker characteristics, such as job tenure, age, sex, or wages in the pre-shock period. All in all, both regression analysis and causal forests complement the theoretical framework's prediction that workers' adjustments to immigration shocks depend on how their firms respond to these shocks, and that response might be influenced by the institutional factors in which they operate, for instance, through binding minimum wage constraints or a looser regulatory environment that fosters informality.

To my knowledge, this is one of the first papers to examine the impact of immigration in developing countries using matched employer-employee data.[5] With the universe of formal job matches over time, I can document more accurately firm-level sources of heterogeneity previously unexplored, as prior research in developing countries primarily focused on the effects across worker characteristics or industries.[6] Although worker and firm characteristics are interrelated through sorting (e.g., minimum wage workers often work in smaller firms), I document that the heterogeneity in this setting stems more from firm-specific factors, even after controlling for the firm's industry and worker baseline characteristics.

It is important to note that these findings reflect only short- and medium-term effects from 2015 up to 2019. As workers and firms adapt over time, a longer-term analysis may demonstrate quantitatively different results (Monras, 2020). However, such analysis is constrained in this context due to the onset of the COVID-19 pandemic. Nevertheless, the findings indicate the need for concurrent policies to support native workers. Given Colombia's relatively high minimum wage, reducing other labor costs for formal employers could help sustain formal employment (Morales and Medina, 2017; Kugler et al., 2017). Additionally, stricter enforcement of fines for hiring informal workers could deter the substitution of formal workers with informal ones (de la Parra and Bujanda, 2024).

[5]In developed countries is not relatively new, 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.

[6]An exception is Delgado-Prieto (2024), which studies heterogeneity at the regional level using two broad firm-size categories from the labor force survey (GEIH, by its acronym in Spanish). Because these results are less precise and may reflect other margins of adjustment, such as employment inflows or regional mobility, the administrative data are better suited for studying firm-level heterogeneity.

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) examines the abolition of immigration restrictions and shows 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. In contrast, Clemens and Lewis (2022) exploits the low-skill H-2B visa lottery to find that firms able to hire low-skill immigrants increase production with zero employment effects on natives. Additionally, there is evidence that firms attract natives from lower- to higher-paying firms post-immigration (Orefice and Peri, 2024; Gyetvay and Keita, 2023).[7] 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. In my paper, this implication stems from the high informality context and from how small firms substitute formal labor for low-priced informal labor in production. In contrast, Amior and Stuhler (2022) argues that small, low-paying firms in Germany tend to hire the most migrants because of lower reservation wages, thereby crowding out natives as they exert monopsony power and move along their labor demand curve. These two novel mechanisms jointly indicate which firms respond more to immigration shocks and how, ultimately shaping workers' outcomes.

Second, I contribute to the literature that estimates the individual impacts of immigration. Initial studies, such as Bratsberg and Raaum (2012) and Foged and Peri (2016), exploited licensing requirements in the Norwegian construction sector and refugee dispersal policies in Denmark, respectively, to estimate worker-level or individual effects of immigration.[8] 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

[7]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. Moreover, Marchal et al. (2023) finds that these gaps narrow depending on the exporting activity of the employing firm, as migrants have an informational rent when the exporting destination coincides with their region of origin.

[8]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.

of immigration (Dustmann et al., 2023). This allows me to understand the main drivers of labor market adjustments to immigration in two novel ways: by studying worker-level effects, including firm-level heterogeneity, and by exploiting a machine-learning method that estimates heterogeneous effects in a data-driven manner.

I also build on the literature on how workers respond to labor demand shocks, such as those coming from import competition (Autor et al., 2014), local unemployment shocks (Yagan, 2019; Redondo, 2022), and mass layoffs (Gulyas et al., 2019). My findings show that formal firms in contexts of high informality more strongly determine wage and employment impacts on formal natives, both theoretically and empirically. This important result complements the labor literature by shifting the focus from worker characteristics or industries 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.[9]

Lastly, this paper contributes to the growing literature on the impact of international migration in developing countries. I document negative wage and employment effects for formal natives at the worker-level, while evidence from related studies is mixed. In Colombia, Bonilla-Mejía et al. (2024) and Delgado-Prieto (2024) report negative regional employment effects in the formal sector, whereas Caruso et al. (2021) and Lebow (2022) find only imprecise negative effects on regional formal wages. By contrast, studies in Turkey document positive regional employment effects for men in the formal sector (Del Carpio and Wagner, 2015; Ceritoglu et al., 2017; Aksu et al., 2022), and evidence from Peru indicates positive effects for highly educated natives (Groeger et al., 2024). Beyond differences in institutional factors and labor markets, these studies use cross-sectional surveys for their regional-level analysis. Although labor force surveys provide crucial information about the informal sector, they remain limited in their ability to test firm-level mechanisms.[10] Surveys are typically restricted to broad geographic or firm-size categories, making it

[9] 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.

[10] They also enable a direct comparison between worker- and regional-level wage estimates, highlighting the importance of accounting for compositional changes among workers, which survey data can only partially address.

challenging to study firm heterogeneity. Specifically, the analysis in Delgado-Prieto (2024) was constrained to 24 departments and two firm-size categories, whereas, using administrative records, I study heterogeneous effects across the whole distribution of firm size and productivity. Moreover, my paper is the first to use panel administrative data in a developing-country setting to quantify worker-level effects of immigration and to examine heterogeneity across firm characteristics. This approach identifies precisely which workers are most affected and addresses the conceptual differences between regional and individual labor-market responses emphasized by Dustmann et al. (2023), thereby helping to explain why my findings may differ from those of studies on Colombia, Peru, and Turkey.

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

2 Institutional Context and Data

2.1 The Venezuelan Mass Migration

Colombia and Venezuela share an extensive land border, historically characterized by a dynamic relationship marked by 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 amid Venezuela's unprecedented socio-economic and political deterioration, which triggered massive outflows of people since 2013 and peaked around 2021. As a result, several countries in

Latin America are receiving vast numbers of migrants, especially Colombia, Perú, and Ecuador (UNHCR, 2019). Colombia has by far been the main destination, hosting more than 2 million Venezuelans, including around 1 million of working age, equivalent to 4.1% of the working-age population as of 2019 (DANE, 2019).

The Venezuelan exodus is unprecedented in Latin America's recent migration history, comparable globally to the Syrian and Ukrainian exoduses. Turkey has been the major host country for Syrian refugees, with various papers analyzing the labor market impacts of this immigration shock (e.g., Del Carpio and Wagner (2015); Ceritoglu et al. (2017); Aksu et al. (2022); Gulek (2024)). However, the Colombian context differs from the Turkish one. Venezuelans speak the same language as Colombians and can be categorized as voluntary migrants or forcibly displaced refugees. They also have access to work permits from the Colombian government's generous open-border policy. Since 2018, irregular Venezuelans in Colombia have been eligible for the Special Permit of Permanence (PEP, by its Spanish acronym), which enables them to work for at least 2 years, provides access to essential services, and facilitates their integration into Colombian society.[11] However, as of 2019, about 90% of Venezuelan immigrants remain employed in the informal sector (see Online Appendix Figure A.1a). Limited formalization can be driven by migrants' lack of skills certification and by employers' low incentives to formalize informal migrants, given unchanged enforcement costs of informality during regularization.[12] Consequently, Venezuelans are disproportionately concentrated at the lower end of the wage distribution (Delgado-Prieto, 2024), reflecting significant occupational downgrading despite their comparable or higher education levels relative to Colombians during the 2016–2019 period (see Online Appendix Figure A.1b).

The informal sector, therefore, absorbs most of the immigration shock in Colombia. In line with a positive labor supply shock, Delgado-Prieto (2024) finds a significant negative effect of 1.9% on informal wages for every unit increase in the migrant share. However, this shock also has direct

[11] In 2021, to overcome the limitations of the PEP renewals, the government enacted a Temporary Protection Statute for Migrants (ETPV, by its acronym in Spanish) offering up to ten years of regularization.

[12] In Colombia, informal employment can imply penalties up to 400 times the monthly minimum wage. However, labor inspections are infrequent, and when workers report these violations to the Ministry of Labor, disputes generally favor employees (de la Parra et al., 2024).

effects on the formal sector. Because firms, especially smaller ones, use both formal and informal labor in production, they can substitute formal for informal labor in response to lower informal wages, particularly when the two types of labor are substitutable. As a result, formal employment is directly affected by the immigration shock due to the firm's production function. Similar substitution responses between the informal and formal sectors have been observed in other contexts as well, including Brazil (Corbi et al., 2021), Indonesia (Kleemans and Magruder, 2018), and Turkey (Gulek, 2024). Moreover, certain formal firms or self-employed workers may face increased competition in the product or service markets from informal firms that face lower labor costs (Rozo and Winkler, 2021; Bandiera et al., 2023). Therefore, examining how formal workers adjust across the firm-size and productivity distributions is crucial for understanding the heterogeneous responses of immigration. I also investigate the effects of immigration using recent machine-learning techniques to identify the worker and firm characteristics that drive the worker-level heterogeneity. The data's granularity enables a comprehensive understanding of the interactions among immigration, minimum wages, and firms, and how these ultimately affect formal workers in developing countries.

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*). The 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 contributions of workers, calculated according to their reported base income, to the health system in Colombia. Each observation 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.[13]

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 migration questions in the most recent census, such as the year of arrival for 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.[14]

For the analysis, I constructed a dataset with 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 aged 25 to 55 as of 2015, assigning the immigration shock to each worker based on their location in 2015. This restriction narrowed the sample down to 7,123,223 workers.[15]

I then transformed the municipality variable into a more standard definition of local labor markets by adjusting the methodology of Sanchez-Serra (2016).[16] The adjustment resulted in 109 functional urban areas (FUAs) after excluding small or rural municipalities, producing a sample of 6,706,035 workers.[17] 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 to avoid mechanically lower wages and to those with positive wages, requiring that the worker be employed in the

[13] 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.

[14] The GEIH 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) documents that surveys can attenuate immigration effects due to measurement error of migrants.

[15] 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.

[16] The municipality variable in the PILA has the issue that some firms with establishments in smaller cities may report their workers under the nearest capital city where their main establishment is located. This could lead to a misrepresentation of local employment, but as long as it is unrelated to the local immigration shocks, it is less of a concern.

[17] 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. Online Appendix Table H.1 shows the sample distribution by FUAs, with only 5.9% of workers excluded from the analysis.

comparison post-treatment year. Thus, the sample varies slightly by year, making it an unbalanced panel. 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.

Descriptive statistics for formal workers. Online Appendix Table A.1 shows descriptive statistics for natives, foreigners, and Venezuelans with PEP permits broken down by age, sex, and wages over time.[18] Venezuelans with PEP in the formal sector tend to be younger, predominantly male, and earn lower wages compared to both natives and other foreigners. The share of Venezuelans with PEP working in the formal sector is still small, suggesting that the regularization's impact on the Colombian labor market is limited (Bahar et al., 2021) due to a weak first-stage. Other foreigners, in contrast, earn substantially higher salaries than natives, yet this gap has been shrinking in recent years. The number of other foreigners in the formal sector has also risen, suggesting it may include Venezuelans holding other types of documentation.

Online Appendix Figure A.2 illustrates how binding the minimum wage is for a large share of formal workers, with around 40% of all formal workers earning it in 2015. [19] 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 declines during positive labor supply shocks or negative demand shocks, and why job displacement becomes more likely instead. During the analysis period (2015–2019), the minimum wage increased in real terms by less than 3% each year, mitigating concerns about additional employment impacts.[20]

[18] 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 documentation are classified as Venezuelan regularized migrants, and those with a foreign ID or passport are classified as foreigners. Note that the PEP program began around 2018 to facilitate the regularization of Venezuelan migrants. Thus, these migrants were not identifiable before.

[19] To contribute to the social security system, workers must declare a labor income equal to or above the monthly minimum wage (up to 25 minimum wages), so many self-employed workers who can decide how much their basic income is report the minimum wage even if they earn more or less. Adjusting labor income is more rigid for employees, as it also affects the additional 20.5% of labor income that firms pay toward workers' health and pension contributions. If under-the-table payment agreements exist (Feinmann et al., 2022), any changes in these payments in response to the immigration shock might bias the observed wage effects.

[20] Online Appendix Table A.2, I aggregate worker-level data from the PILA into seven firm size categories, excluding self-employed workers, to describe other insights. Male workers are predominant across all formal firm sizes,

3 Stylized Model

In this section, I develop a partial equilibrium model with heterogeneous firms and worker types to motivate how firms can shape immigration effects in a context of informality, and to guide the discussion of the empirical findings.

The model's market structure consists of J formal firms that decide how many workers to hire from two types. 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).[21] Natives and migrants are perfect substitutes for informal labor $I = I^N + I^M$, and 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 ν_{i,L_j} depending on the labor group they belong $L \in \{I, F\}$. This yields a worker-specific job valuation at each firm, implying that firms face upward-sloping labor supply curves, which they internalize when choosing posted wages to maximize profits.[22] In this model, the indirect utility of worker i employed at firm j is:

$$\upsilon_{i,L_j} = \beta_L \ln w_{L_j} + a_{L_j} + \nu_{i,L_j}. \tag{1}$$

Under the assumption that ν_{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 labor supply functions can be expressed as:

$$\ln I_j = \ln(\mathcal{I}\lambda_I) + \beta_I \ln w_{I_j} + a_{I_j}, \tag{2}$$

especially in small to medium-sized firms (with 10 to 999 workers), where over 60% of workers are male. Smaller firms employ older workers on average (39.6 years), while larger firms employ younger workers (36 years). Average wages rise with firm size, ranging from approximately 272 USD in firms with 1-4 workers to 531 USD in firms with more than 1,000 workers. Online Appendix Figure B.1 presents a histogram of firms by size, overlaid with the total number of employees in each firm size category. While most firms employ 1 to 4 employees, the distribution of employees across firm sizes is more evenly spread.

[21] 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, I do not consider the general equilibrium effects of potential workers' transitions between the formal and informal sectors.

[22] For instance, preferences for working in a firm may refer to distance to the workplace or interactions with coworkers.

$$\ln F_j = \ln(\mathcal{F}\lambda_F) + \beta_F \ln w_{F_j} + a_{F_j}. \tag{3}$$

In these equations, 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, $\frac{d \ln L(w_{L_j})}{d \ln w_{L_j}} = \beta_L$ 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 product of labor.[23] Adding the monopsony angle to the model is important for explaining equilibrium wage differentials across firms, consistent with the findings of the AKM model. However, my focus differs from that of Amior and Stuhler (2022), which examines the mediating role of monopsony power under immigration shocks. Instead, I explore how having firms that can hire both formal and informal labor generates different predictions about the impact of immigration shocks on workers, assuming monopsony power remains constant across firms.

The profit function of firm j, depends on a productivity shifter T_j, a price of the good P_j, and a production function Q_j:

$$\max_{I_j, F_j} \pi_j = P_j T_j Q_j - \tau(I_j) w_{I_j}(I_j) I_j - (1+\tau_F) w_{F_j}(F_j) F_j. \tag{4}$$

Here, $\tau(I_j)$ represents a convex cost that is increasing in the firm's informal labor size. These convex costs are important for matching the stylized fact that informal labor decreases with firm size (Ulyssea, 2018) and for capturing the cost of evasion related to government law enforcement. Particularly, I assume that $\tau(I_j) = I_j^{\eta}$ with $\eta \geq 0$. The τ_F represents the constant payroll taxes firms must pay for formal workers, and the production function takes a constant elasticity of substitution (CES) form: $Q_j = (\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho})^{\frac{1}{\rho}}$. Thus, formal and informal workers can possess different skills and, therefore, are imperfect substitutes, with an aggregate elasticity of substitution common across all firm types given by $\sigma = \frac{1}{1-\rho}$. To close the model, P_j is the inverse demand function defined as $P_j = D_j (T_j Q_j)^{-(1-\varepsilon)}$, where $\varepsilon^D = -1/(1-\varepsilon)$ is the elasticity of product demand and

[23] I omit 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.

D_j is the firm-specific product demand as in Card et al. (2018).[24]

In this partial equilibrium framework, I study the effect of an immigration shock that shifts the aggregate informal labor supply outward ($d\mathfrak{I}$), consistent with the fact that around 90% of immigrants are informally employed. I examine the firms' response across the wage and employment margin, so the wage and employment elasticity for each type of worker in firm j is $\varepsilon_{w_{L_j},\mathfrak{I}}$ and $\varepsilon_{L_j,\mathfrak{I}}$, respectively. Having firm-level responses to an immigration shock is one novelty of this framework. Unsurprisingly, in Online Appendix F I show that the elasticity of informal labor is always positive ($\varepsilon_{I_j,\mathfrak{I}} > 0$) and the elasticity of informal wages is always negative ($\varepsilon_{w_{I_j},\mathfrak{I}} < 0$) in response to 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 each firm change in response to the shock:

$$\varepsilon_{w_{F_j},\mathfrak{I}} = \Omega_j s_{I_j}(\varepsilon - \rho). \tag{5}$$

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 $\Omega_j = \frac{1}{\xi_{I_j}\xi_{F_j} - (\varepsilon-\rho)^2 s_{I_j}\beta_I s_{F_j}\beta_F}$ is a positive parameter.[25] Firstly, if informal workers are sufficiently close substitutes to formal workers (such that $\rho > \varepsilon$), 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 ($\varepsilon_{w_{F_j},\mathfrak{I}} \downarrow$).

The corresponding expression regarding formal employment is equal to:

$$\varepsilon_{F_j,\mathfrak{I}} = \Omega_j s_{I_j}(\varepsilon - \rho)\beta_F. \tag{6}$$

The implications in terms of the substitution parameter (i.e., $\rho > \varepsilon$) hold similarly as for formal

[24] I do not distinguish whether the produced good is tradable or non-tradable (Burstein et al., 2020). Besides, I exclude any spillover labor demand shock arising from migrants' consumption of goods and services (Galaasen et al., 2025).

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

wages, though the response is now adjusted by β_F. This is consistent with Clemens and Lewis (2022) framework, where a high elasticity of substitution across immigrants and natives relative to the output elasticity yields negative employment effects. Notably, as the relative contribution of informal workers to production increases ($s_{I_j} \uparrow$), the adjustment on formal employment is again more negative ($\varepsilon_{F_j,\mathfrak{I}} \downarrow$) as long as informal and formal workers are close substitutes.[26]

In this model, for certain low-productivity firms, formal wages can be downwardly rigid due to the existence of a minimum wage (e.g., $\overline{w}_{F_{Min}} \geq w_{F_j}$), which results in a muted formal wage margin (i.e., $\varepsilon_{\overline{w}_{F_{Min}},\mathfrak{I}} = 0$). In Online Appendix F, I derive the remaining elasticities under this condition for those firms defined as demand constrained (j_m). Importantly, the formal employment elasticity would be higher in absolute terms than for the firms where the minimum wage does not bind.[27] That is, firms that are demand constrained by the minimum wage change their formal employment more in response to the informal supply shock than unconstrained firms.

The stylized model I propose points to two main conclusions. First, when substitutability between formal and informal workers is high, an informal labor supply shock generally reduces formal wages and employment. Second, the extent to which the minimum wage is binding and the production structure depends on informality 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. Additionally, if they pay their formal workers close to the minimum wage, the formal employment response becomes even more important.

In the model, a firm's informal production share is inversely related to its size. Large firms face higher marginal costs of hiring additional informal workers because the cost of informal labor is convex. Therefore, this motivates the empirical analysis of worker responses across the firm-size

[26] Due to the absence of educational information in the administrative data and to maintain broader predictions, the model abstracts from skill heterogeneity and focuses solely on the formal-informal distinction. While this approach prevents the analysis of differential impacts on high- and low-skilled natives within each sector, my data lack clear empirical counterparts to test such additional predictions.

[27] As a reference, the expression would be equal to $\varepsilon_{F_{jm},\mathfrak{I}} = \frac{s_{I_{jm}}(\varepsilon-\rho)}{\left((1-\varepsilon)+(\varepsilon-\rho)s_{I_{jm}}\right)\left(1+\beta_I\left((1+\eta-\rho)-(\varepsilon-\rho)s_{I_{jm}}\left(1+s_{F_{jm}}\frac{(\varepsilon-\rho)}{(1-\varepsilon)+(\varepsilon-\rho)s_{I_{jm}}}\right)\right)\right)}$.

distribution to validate the model's predictions. I also extend the worker-level analysis to worker responses across firms' productivity using the estimated AKM firm effects, as these are also linked to the firm's informal production share.

Alternative mechanisms. Beyond these predictions, the stylized model is agnostic about other potential mechanisms. These include product-market competition effects, formal firm entry and exit, and the reallocation of workers across firms. To start, consider a market in which formal and informal firms produce similar goods and compete on prices. An immigration shock that lowers labor costs only for informal firms grants them a cost advantage that can translate into lower prices, larger market shares, and weaker labor demand at formal firms (Rozo and Winkler, 2021; Bandiera et al., 2023). Evidence on price pass-through and market-share changes on informal firms is, however, limited to test this channel. A second extension would allow for formal firm entry and exit, which could help explain declines in formal labor demand (Gulek, 2024), though some of this margin is already captured by within-firm substitution between formal and informal inputs in the current framework. A third extension would introduce spillovers across firms, such as through worker reallocation or production networks, which can shape the adjustments to immigration (Akgündüz et al., 2024; Gyetvay and Keita, 2023). Yet the empirical evidence below suggests worker mobility across firms is limited, so a model centered on reallocation may add little in this context. My focus with this model is on a single mechanism: adjustments in labor inputs within formal firms, where the coexistence of informal and formal labor in production suggests that changes in relative prices can affect labor input choices (Corbi et al., 2021; Kleemans and Magruder, 2018).

4 Empirical Strategy

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

$$\Delta Y_{i,l,t} = \delta_t + \theta_t \Delta M_{l,2018} + X_i' \beta_t + \Delta u_{i,l,t}. \quad (7)$$

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 a baseline year. 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.[28] Broadly, this specification is akin to Foged and Peri (2016) regression with individual fixed effects, and 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 2015 is the baseline year and $\theta_{2015} = 0$ 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 across all specifications at the level of the treatment, which is the FUA (equal to $G = 109$).

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, and the θ coefficient measures changes in the probability in pp. As in Yagan (2019), I use average employment in the pre-shock period to transparently account for workers' varying labor trajectories in the formal sector. In the event study figures, however, I use the simple difference with the base period ($e_{ilt} - e_{il,2015}$) to avoid pre-treatment coefficients being mechanically close to zero. Second, the wage outcome is $\frac{w_{i,l,t} - w_{i,l,2015}}{w_{i,l,2015}}$, and the θ coefficient 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

[28] Education and occupation data are unavailable in the PILA. Due to measurement errors, industry information is used as a robustness control rather than as a baseline control, as these codes were unverified until 2019.

employed in the formal sector in that given year, while if employed, they are equal to their wages. 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:

$$\Delta M_{l,2018} = \frac{L_{Ven,l,post}}{L_{l,2018}}, \tag{8}$$

where the numerator is the stock of employed migrants from Venezuela (aged 18 to 64) in local labor market l who arrived in Colombia between 2016 and 2018. Employed migrants include both Venezuelans and returning Colombians from Venezuela. The denominator $L_{l,2018}$ is the total employed population in the local labor market.[29] I focus and interpret mainly the coefficient of 2018 in the regressions (i.e., θ_{2018}) to match the census year and avoid rescaling the shock as for the coefficients of other periods. Lastly, this constant immigration shock is useful because it leverages the full census count rather than a survey to construct migration shares, and it allows for transparent placebo tests of pre-trends within the same analysis, as illustrated by Dustmann et al. (2017).

Because migrants self-select into areas with better economic opportunities, the immigration shock $\Delta M_{l,2018}$ is likely endogenous, and its effects would be upward-biased. In Figure 2a and 2b) I present evidence for this showing ordinary least squares (OLS) estimates. Thus, to consistently estimate the effect of immigration on the outcome variables, I instrument the immigration shock $\Delta M_{l,2018}$ with the distance to the nearest crossing bridge to Venezuela and with past Venezuelan settlements. The motivation for the IV approach is the following.

Distance plays a crucial role in the Venezuelan immigration, given that Colombia and Venezuela share a 2,220-kilometer land border. As a result, the entry of individuals into the local labor market l is affected by the travel distance between the two countries, which imposes both time and eco-

[29] 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 Online Appendix Table B.1, row 4).

nomic constraints on immigrants. The instrument of distance d_l is defined as the sum of the linear and quadratic air kilometers distance to the nearest main crossing bridge with Venezuela:

$$f(d_l) = d_l + d_l^2. \tag{9}$$

A potential challenge to this instrument is that border areas may face additional economic shocks (such as reduced trade or business transactions) compared with areas farther from Venezuela, potentially violating the exclusion restriction. Online Appendix Figure B.2a offers suggestive evidence that the trade shock induced by the Venezuelan crisis began several years before the immigration shock. In particular, exports from border departments to Venezuela remained consistently close to zero post-treatment, so any potential threat must stem from long-lasting lagged effects. Still, I observe insignificant employment and wage effects in the largest firms, which are more likely to be exporters affected by trade shocks. I also show in Figure B.2b that trends in log GDP for border and non-border departments were similar before the immigration shock, indicating that any trade-related impact on broader economic activity was limited. A department-level regression of the immigration shock on log exports to Venezuela before 2015 yields insignificant coefficients, yet with expected large values in 2012 and 2013, and a regression on log GDP yields coefficients that are close to zero in the pre-treatment periods (see Online Appendix Figures B.3a and B.3b). More strictly, given that border areas also exhibit higher rates of immigrant regularization (Bahar et al., 2021), I exclude them from the main analysis to reduce potential confounding. The resulting point estimates remain similar, although wage effects become statistically insignificant. Taken together, this evidence supports the requirement that distance satisfies the exogeneity assumption $E[f(d_l)\Delta u_{lt}] = 0$.

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

$$z_l = \left(\frac{Ven_{l,2005}}{\sum_l Ven_{l,2005}} * L_{Ven,post} \right) / L_{l,2005}, \tag{10}$$

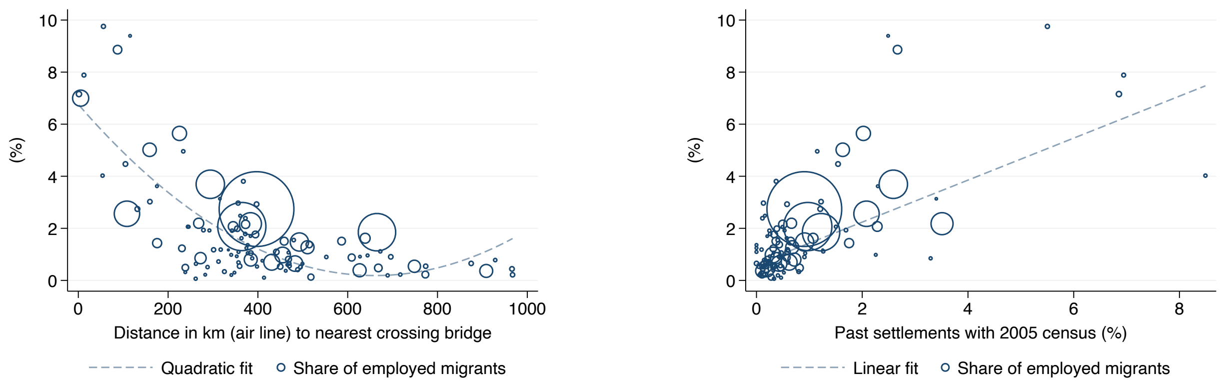
where the first term represents the share of Venezuelans in FUA l, according to the 2005 population census, normalized by the working-age population $L_{l,2005}$ in l at 2005, whereas $L_{Ven,post}$ is the number of employed migrants in Colombia who arrived between 2016 and 2018 according to the census. Since I have only one origin country, this corresponds to a just-identified shift-share instrument, where exogeneity relies on the share of Venezuelans (Goldsmith-Pinkham et al., 2020). The motivation for using past settlements as an instrument is that newly arriving immigrants tend to relocate to areas with established Venezuelan communities. To mitigate concerns about exclusion restrictions, I select a sufficiently lagged share in 2005 that captures persistent network effects without being systematically linked to current labor-demand shocks. A department-level regression testing pre-trends in log GDP and log exports to Venezuela with a similar instrument yields coefficients near zero for both outcomes (see Online Appendix Figures B.3a and B.3b). This indicates that past settlements predict new arrivals but remain uncorrelated with time-varying shocks, supporting the exogeneity assumption $E[z_l \Delta u_{lt}] = 0$.

Figures 1a and 1b depict graphically the first stage of the immigration shock $\Delta M_{l,2018}$ for the 109 FUAs included in this analysis (see Online Appendix Map H.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 immigration shock in the FUAs until a point at which longer distances no longer imply lower immigration shocks, causing the curve's slope to bend upward. Based on past settlements, the second instrument shows a positive and almost 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, whereas that study uses only 24 departments for analysis due to limitations in the labor force survey's sample.

Figure 1: **Immigration shocks and the instruments**

(a) Distance

(b) Past settlements



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 1a and 1b 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 (2SLS) will weight each FUA by the number of individual observations available. With this in mind, the first-stage model at the FUA level is:

$$\Delta M_{l,2018} = \delta + f(d_l) + z_l + \upsilon_l \tag{11}$$

where $f(d_l)$ 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 $\Delta M_{l,2018}$. In rows 1–3 of Online Appendix Table B.1, I estimate the main outcomes using both instruments jointly and separately. The results for the main outcomes using only the distance instrument yield slightly more negative coefficients than those obtained with both instruments, though the estimates are noisier. Using only past settlement instruments, the coefficients are also more negative than when combining both. The decision of which instruments to choose embeds a trade-off between different exclusion restrictions and the relevance of the combined instruments. Given that the estimates from each instrument are not substantially different, jointly they improve the precision of the estimates, and the R^2 in Online Appendix Table A.3 of the first stage regression is higher with both, although the F-statistic is lower, I combine them for the main analysis.

5 Worker-level 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 worker characteristics. One advantage of the dynamic DiD specification is the ability to test for differential trends in outcomes before the immigration shock occurs. Figure 2a shows no significant pre-trends for employment or wages, supporting the assumption that workers in less immigration exposed areas are a suitable counterfactual of the outcomes of workers in more exposed areas.

In the post-treatment periods, OLS coefficients are close to zero, presumably upward-biased, while the 2SLS regression helps mitigate this bias, resulting in more negative coefficients. Since pre-trends are relatively 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 this reason, I only analyze the 2SLS coefficients from here on.

In 2018, the census year, a one pp increase in the immigration shock in an area reduces the probability of formal sector employment by 1.1 pp (see Figure 2a).[30] To interpret this coefficient, the probability of formal sector employment based on the labor force survey 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.[31] In Figure 2b, I find a coefficient of –0.6% on formal wages in 2018 for a one pp increase in the immigration shock.[32] Consequently, a worker at the 75th percentile of exposure

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

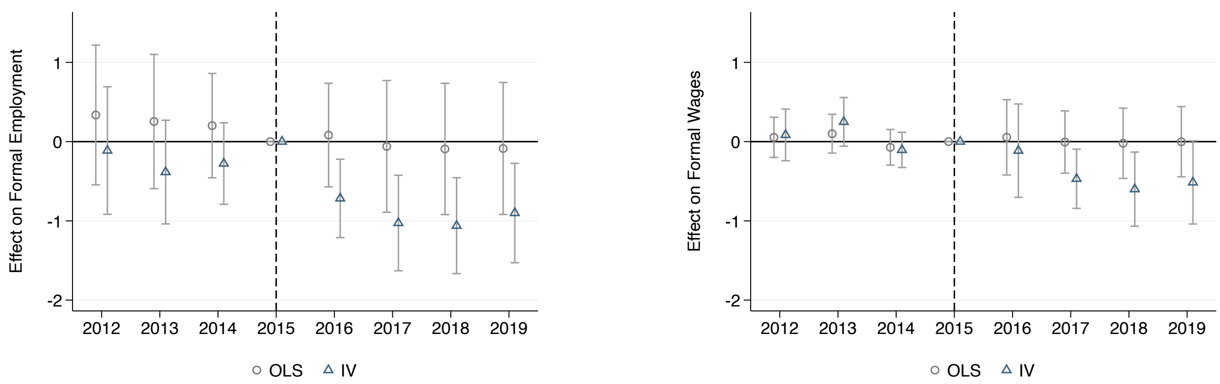
[31] The 25th and 75th percentile migration rate is 0.6% and 2.1%, respectively. So, (2.1-0.6)*2.4=3.6.

[32] In comparison to Osuna Gomez and Medina-Cortina (2023), which studies the deportation of migrants from the United States to Mexico, they find that a one pp increase in the deportation shock decreases the probability of formal

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 2: **Event study estimates on individual wages and employment**

(a) Employment (b) Wages



Note: I estimate Equation (7) 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.

I then examine the heterogeneity in wage and employment estimates based on workers' and firms' characteristics prior to the immigration shock, specifically using 2015 characteristics. The coefficients for each subgroup come from separate regressions of the main empirical specification (7).

The first worker characteristic I examine is job type, given its flexibility in adjusting to labor market shocks, and I categorize workers as employees or self-employed. In Colombia, self-employment accounts for about half of the employed population, predominantly in the informal sector, though a substantial share is in the formal sector (around 18% of all native formal workers were self-employed in 2015). Online Appendix Figure A.4 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 decide individually

employment among low-wage workers by about 0.11 pp and reduces formal wages by around 0.2%. These results are consistent with my findings, though the magnitudes are somewhat smaller.

whether to contribute to the social security system, making it less costly for them to exit the formal sector than for employees with more rigid labor contracts.[33]

To present heterogeneous effects across multiple dimensions, I now examine only the post-treatment estimate for 2018. Still, Online Appendix Table G.1 shows no systematic pre-trends across different worker or firm categories for employment or wages.[34] Figure 3 shows that older workers experience a larger decline in the probability of formal sector employment than younger workers, similar to Dustmann et al. (2017). In contrast, the pattern is less pronounced for wages, with similar negative estimates across all age groups.[35] In terms of sex, the impact on employment and wages is alike for both men and women.

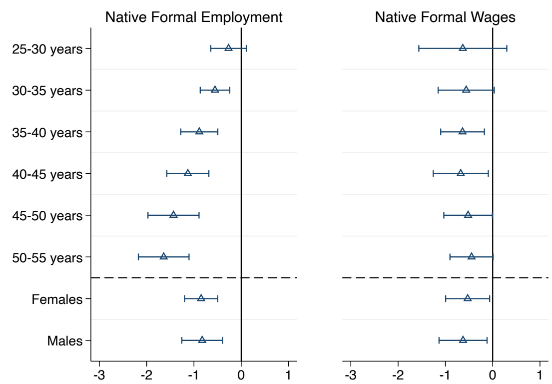
I then analyze the impact on earnings, which captures the joint effect of wage and employment changes on workers. Online Appendix Figure A.8a shows that workers over 30 experience a relatively similar reduction in earnings as their confidence intervals overlap. This suggests that older workers are more likely to be displaced from the formal sector, but younger workers face persistent wage losses.

[33]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 wage point estimates are more negative for the self-employed than for employees.

[34]I do not show pre-trends before 2012. Exposed areas may have exhibited different trends earlier, but three years of flat pre-trends before treatment suggest that any differential growth before 2012 is unlikely to generate lagged effects that could explain post-treatment trends.

[35]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 Online Appendix Figure A.3).

Figure 3: **Employment and wage estimates by age group, 2015–2018**



Note: I estimate Equation (7) 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 further identify heterogeneous effects, I now examine the workers' job tenure up to the base period of 2015. Online Appendix Figure A.5 categorizes natives by their job tenure, ranging from 0 to over 9 years. Notably, the immigration shock on employment is more severe for workers with lower tenure at the firm. This result is partly explained by severance payments, which increase with tenure, making it more costly for firms to dismiss longer-tenured workers. Additionally, longer-tenured employees can accumulate more firm-specific human capital, making them less substitutable by migrants with similar characteristics but lacking country- and sector-specific skills.[36]

Because immigration can disproportionately affect natives with specific skills, I estimate worker-level impacts across the baseline wage distribution, which serves as a proxy for education. For this analysis, I categorize native workers into seven bins based on their local wage distribution. Figure 4 illustrates the uneven effects of immigration: native workers earning the minimum wage, and

[36]The main analysis shows that older workers and those with shorter tenure face the greatest employment losses from the immigration shock. To explore this heterogeneity, I combine baseline age with tenure. Online Appendix Table A.4 indicates that age matters more: natives under 35 face no significant employment impact, while those over 35 face negative effects, especially with short tenure (1 pp vs. 0.3 pp). Wage effects show no clear patterns by age or tenure.

having the lowest skills within the formal sector, experience the most negative impact on formal employment, while workers above in the wage distribution are unaffected. Specifically, a 1 pp increase in the immigration shock leads to a 1.5 pp decrease in the probability of formal sector employment for minimum-wage workers. Because the immigration shock considers all employed immigrants locally, whether formal or informal, I need to multiply the coefficient by the size of the formal sector (around 0.5) and by the share of minimum-wage workers within the formal sector (approximately 0.4) to accurately benchmark the effects. This calculation indicates that one additional employed immigrant displaces roughly (0.5*0.4*1.5=) 0.3 native minimum-wage workers in the formal sector. 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.[37] Additionally, the presence of a large informal sector is relevant, as minimum-wage workers tend to be less educated and more easily substitutable with informal workers.

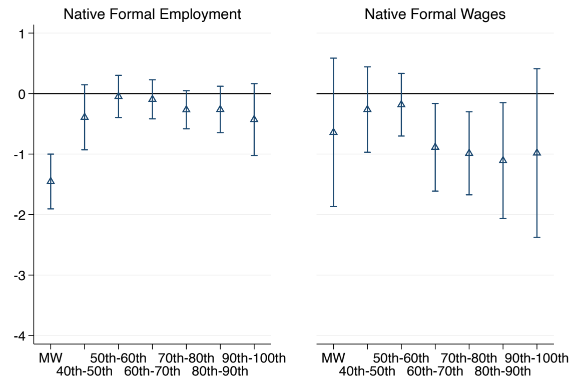
Conversely, formal workers who earn the minimum wage are the least affected by the immigration shock in terms of wages, partly due to their downward wage rigidity. In contrast, for workers between the 60th and 90th percentiles of the local wage distribution, who earn around two to three times the minimum wage on average, I observe a negative wage effect of between 1% and 1.2%. Increased competition from mid- and high-skilled migrants entering the formal sector is a potential explanation. Online Appendix Table A.1 shows a sharp increase in other foreigners between 2015 and 2019, who often earn high wages. On top of that, the contraction of the formal sector may strengthen the bargaining power of formal firms, which could affect the formal wages of incumbent workers earning above the minimum wage.[38] I turn to estimate the impact on earnings to find which workers are more affected overall. Online Appendix Figure A.8b shows that the only significant negative impact on earnings is concentrated among workers who earn the minimum wage before immigrants arrive, reflecting a more substantial adverse effect from the employment

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

[38] This does not necessarily imply an absolute wage decrease. 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.

margin.

Figure 4: **Employment and wage estimates by individual wage at baseline, 2015–2018**



Note: I estimate Equation (7) 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 .

Threats to the identification assumptions. Even if border areas have similar trends before immigrants arrive as non-border areas, they may be more exposed to post-treatment shocks derived from the Venezuelan crisis, potentially violating the exclusion restriction. To address this, I include a Bartik-type local labor demand control, based on the interaction between the pre-shock industry composition of local employment in each FUA and national post-shock employment trends by industry, to capture latent demand shocks across areas. Online Appendix Figure B.4 shows that the results across the wage distribution remain essentially unchanged, suggesting that differential demand shocks are less of a concern in this setup. I also exclude border areas from the sample to further mitigate the concern that they drive the main results, and I even find more negative point estimates for employment and wages, although the wage effects become insignificant (see Online Appendix Table B.1, row 2).[39] In addition, I include controls for the baseline wage of the worker,

[39]Since Bogotá accounts for 32.7% of formal employment in the sample, I also exclude it and find that although coefficients become less negative, especially for wages, they remain significant (see Online Appendix Table B.1, row 3). I also adjust nominal wages to real terms using the national CPI, which yields slightly more positive wage effects. Finally, top-coding wages at the 99th percentile show similar estimates.

as a proxy of workers' education, in the main regression to find similar results for wages and employment, with slightly more negative effects for wages. By further controlling for the baseline wage distribution, any remaining threat must arise from a specific shock that affects only workers with similar wages, ages, genders, and job types in the more exposed areas.

Comparison of worker-level and regional-level effects. Most of the migration literature focuses on regional responses when examining the effects of immigration shocks.[40] 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 the employment decomposition they introduce to shed light on these different responses.[41] 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. In contrast, the regional-level estimate from cross-sectional data combines all three adjustment margins. These two responses are complementary and address different policy questions. Online Appendix Figure A.6 shows 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%).[42]

[40]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.

[41]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.

[42]I estimate each component coefficient by running separate regional regressions of each share on the instrumented local immigration shock and use the local formal employment in the base period as regression weights.

More formally, the employment decomposition equals:

$$-1.3\% = \frac{E_{r1} - E_{r0}}{E_{r0}} = -\underbrace{\frac{E_{r,Out}}{E_{r0}}}_{\text{Outflows}} + \underbrace{\frac{E_{r,In}}{E_{r0}}}_{\text{Inflows}} - \underbrace{\frac{E_{r,Move}}{E_{r0}}}_{\text{Relocation}} . \tag{12}$$

In this analysis, the only significant margin is 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 Colombia's large informal sector, which enables firms to hire informally after displacing formal workers, and to its relatively less restrictive job-protection regulations compared to Germany.

Regarding wage estimates, the worker-level response shows a decrease of 0.6%, while the regional-level estimate in Delgado-Prieto (2024) is insignificant and nearly zero. As noted in Dustmann et al. (2023), worker-level wage regressions capture changes in the price of labor while holding constant the population composition. In contrast, regional-level wage regressions jointly capture changes in the selection and composition of workers due to inflows and outflows, as well as 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 mainly displacing minimum-wage workers, while positively selecting those who remain employed, thereby mechanically increasing regional formal wages (see Figure 4). However, immigration lowers labor costs in certain mid- and high-wage subgroups, thereby lowering regional formal wages on average. These counteracting effects explain the insignificant formal wage effect at the regional level, in contrast to the negative wage effect at the worker level, highlighting the need to analyze immigration effects on both aggregate local labor markets and individuals within these markets.

Individual panel data also offer insights into inter-regional movements in response to immigration. For example, Foged and Peri (2016) documents that younger workers in Denmark are more mobile following refugee arrivals. Online Appendix Table A.5 shows the effect on regional movements by age groups. Younger formal workers tend to move more, though the coefficients are insignificant. While point estimates decrease with age, all remain insignificant, indicating that mobility is a less relevant adjustment margin in this context.

6 Worker-level Responses Across Firms

The theoretical model suggests that formal workers in smaller and less productive firms are more responsive to 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 fewer than 11 workers. Second, more than 50% of workers in the smallest formal firms earn the minimum wage and are thus on the margin of informality. Third, smaller firms employ a higher share of informal workers, a share that decreases as firms grow. Therefore, the impact of immigration on workers in small firms is expected to be more pronounced.

I categorize workers by firm size in 2015, the year before the immigration shock, and examine worker-level employment and wage coefficients for 2018, the census year. 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.[43] Figure 5 shows that native workers in firms with fewer than 50 employees in the pre-shock period experience the most negative effect on the post-shock probability of remaining in formal employment, whereas workers in larger firms are less affected. After controlling for firms' industry, I find similar, and even more detailed, patterns, indicating that this is not driven by specific industries populated by small firms (see Online Appendix Figure B.5). In line with the model's predictions, small firms, which rely more on informal labor and face lower penalties for hiring informal workers, find it profitable to substitute formal workers with informal ones, whether they are natives or migrants, when the two are close substitutes.[44] This is useful for transparently showing that trade shocks from the Venezuelan crisis

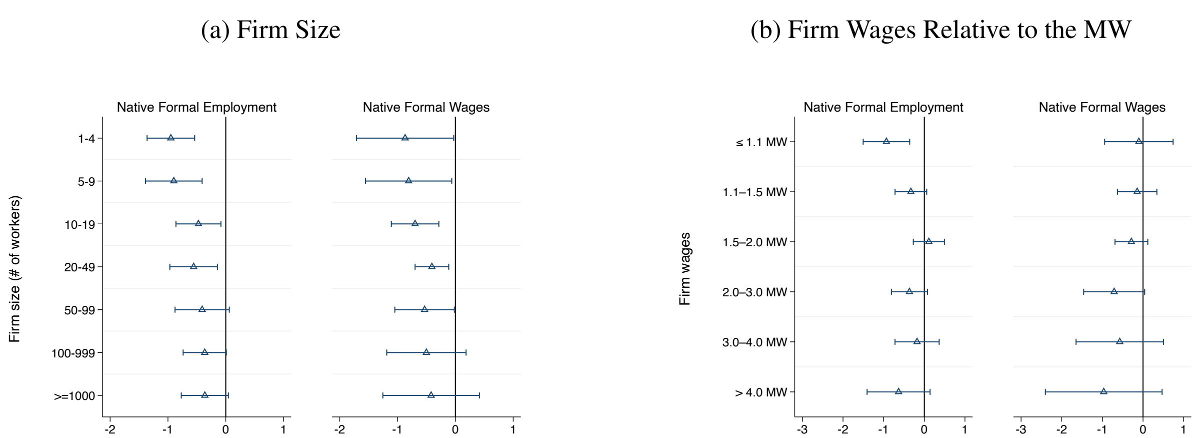
[43]Online Appendix Table G.1 shows that there are no systematic pre-trends across different firm categories, providing support for the parallel trends assumption.

[44]Delgado-Prieto (2024) indicates with survey data on the informal sector that after migrants arrive, the share of informal labor increases more in the smallest firms, indicating workforce composition changes due to this labor substitution.

are less of a concern in this context, as the main effects are observed in the small formal firms that are non-exporters and hence are less likely to be impacted by trade shocks.

Workers in the smallest firms (those with fewer than 10 employees) face slightly more negative wage point estimates than workers in larger firms, but the patterns are less clear than those for employment.[45] This aligns with model predictions, where firms adjust wages less heavily in response to an immigration shock due to the presence of minimum wages and labor market power. In particular, Figure 5 supports the theoretical prediction that firms paying close to the MW respond entirely through employment and not through wages due to wage-setting constraints (Bhuller et al., 2025).

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



Note: I estimate Equation (7) 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. 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.

Because administrative records do not contain worker-level information for the informal sector,

[45]Firm age is an alternative characteristic that can determine immigration impacts on workers. Online Appendix Figure A.7 presents results based on the worker's pre-shock firm age. Workers in younger firms suffer larger employment and wage losses, partly because younger firms tend to be smaller. Still, as Fort et al. (2013) suggests, firm responses over the business cycle may vary by firm age and size. Thus, in Appendix Table A.6, I separate these two categories and find that native workers in the youngest firms experience significant declines in both employment and wages, with the magnitude varying by firm size. In contrast, among older firms, only the smallest ones experienced employment losses.

I cannot directly observe movements between informal and formal employment. To address this limitation, I construct a proxy that captures how closely formal firms are connected to the informal sector, complementing standard measures of firm size. The key idea is to use firms' hiring patterns to infer their position in the formal labor market. Specifically, I build an insider index inspired by the poaching index in Bagger and Lentz (2019), which classifies firms by the source of their new hires. The intuition is simple: some formal firms mainly hire workers with formal-sector experience, while others predominantly recruit workers new to the formal sector. The latter group consists largely of labor market entrants or workers transitioning from informality. Firms in this second group effectively serve as entry points into formal employment.

Formally, the index measures the share of a firm's hires in year t that come from within the formal sector:

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

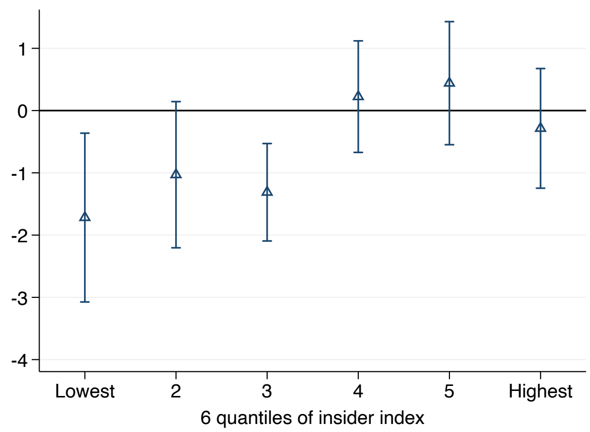
where $N_{j,t}^{In}$ denotes the number of workers hired by firm j in year t who were previously employed in the formal sector, and $N_{j,t}^{Out}$ denotes hires from outside the formal sector.[46] A higher value of $\kappa_{j,t}$ indicates that a firm primarily recruits experienced formal workers, while a lower value indicates greater reliance on labor market entrants or workers coming from informality. I then compute changes in the insider index between 2015 and 2018, $(\kappa_{j,2018} - \kappa_{j,2015})$, and assign these changes to workers based on their employer in each year. This allows me to study how firms' hiring composition responds to the immigration shock.

Figure 6 presents results across six quantiles of the pre-shock insider index. Firms that were initially more reliant on hiring from outside the formal sector experience a sizable decline in their insider index after immigration. A 1 pp increase in the migration rate reduces the insider index of firms in the lowest quantile by about 1.7 pp. In contrast, firms that primarily hire from within the formal sector exhibit little change. These patterns suggest that the formal firms more closely

[46] Hiring data are available from 2007 to 2018, observed in February and August. The index is undefined for firm–years with no hiring.

linked to the informal sector, hire relatively fewer formal-sector workers following immigration, consistent with a substitution mechanism in which they reduce formal-sector hiring as it becomes relatively more expensive.

Figure 6: **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.

Exit and entry of formal firms. A supplementary explanation for previous employment findings is the lack of formal firm dynamism, particularly in relation to the opening of new formal firms and the transition towards full informality within existing formal firms. Hence, I estimate the likelihood that firms hire formal workers entirely after the immigration shock. For this exercise, I count the number of firms in each area over time and perform regional-level regressions, excluding self-employed workers. Table 1 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 the firm exit rate by 1.2%. 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 up to three years.

Table 1: **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, excluding self-employed workers. Firms are observed in August of each year. Source: PILA, 2015–2018.

Heterogeneity by firm pay premiums. The model indicates that firm productivity is an important predictor of the effects of immigration. Thus, with access to the job matches of formal workers and firms in Colombia, I can construct a measure of firms' wage premiums, serving as a proxy of their productivity (Card et al., 2018). To achieve this, I 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 $\ln w_{it}$. The AKM model is:

$$\ln w_{it} = \alpha_i + \psi_{j(i,t)} + X_{it}'\beta + \varepsilon_{it}. \tag{14}$$

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 using data from 2010 to 2015 for August ($T = 6$), restricting the sample to the largest set of firms connected by workers' mobility.[47]

[47] In these models, ψ_j is identified through workers' movements across firms, assumed to be exogenous conditional on worker and firm fixed effects (i.e., $E[\varepsilon_{it}|\alpha_i, \psi_j(i,t), X_{it}] = 0$). 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. Since I rely solely on the estimated vector of firm FEs $(\hat{\psi}_1, ..., \hat{\psi}_J)$ and worker FEs $(\hat{\alpha}_1, ..., \hat{\alpha}_N)$, in which limited mobility bias only affects their precision not their consistency, this is not a concern. Nevertheless, I employ the leave-out method proposed by Kline et al. (2020) to address the limited mobility bias when decomposing the sources of wage inequality in Colombia.

Online Appendix Table D.2 presents the decomposition of the variance of wages $Var(\ln w_{it})$ in Colombia's formal sector. Worker effects explain 50.1% of the variance of wages, and firm effects explain 15.7%, in line with the literature cited in Card et al. (2018). Compared with middle-income countries, the share of variance explained by firm effects is lower in South Africa and Mexico and similar to that in Brazil (Kline, 2024). A potential explanation is Colombia's relatively high and binding minimum wage, which compresses the firm wage policies of lower-paying firms, decreasing the importance of firm effects. Lastly, the positive sorting of high-wage workers into high-wage firms accounts for another 21.6% of the variance. In four European countries and the US, this sorting explains between 10% and 20% of the wage variance (Bonhomme et al., 2023).[48]

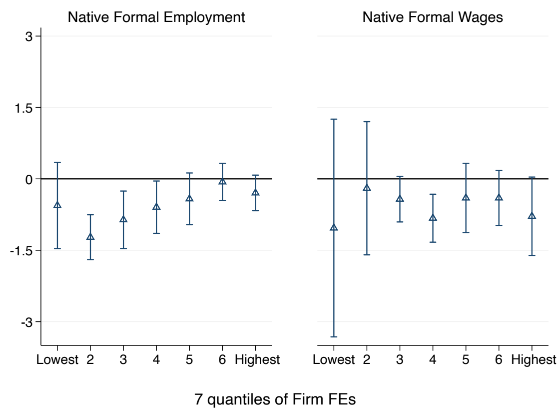
Using the estimated $\hat{\psi}_j$, which are relative to the largest employer in the country, I now categorize workers into seven quantiles of firm FEs, from lowest- to highest-paying firms, to analyze heterogeneity. Online Appendix Table D.1 reports employment and wage descriptive statistics by quantile, noting that the smallest firms are often absent from this analysis because of the largest connected set restriction. Figure 7 shows that workers at lower-paying firms experience negative employment effects while having insignificant wage changes. This contrasts with workers at higher-paying firms who do not experience negative employment effects.[49] To determine whether employment or wage changes prevail, I estimate the earnings outcome across quantiles of firm FEs. Online Appendix Figure A.8c shows that workers in the lowest-paying firms experience a more pronounced decline in earnings than workers from middle- to high-paying firms.[50]

[48]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).

[49]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).

[50]Online Appendix Figure A.9 presents a similar analysis but dividing workers into seven quantiles based on worker FEs $\hat{\alpha}_i$. High-wage workers exhibit the most negative wage point estimates. Conversely, low-wage workers show a null wage effect, while the employment effect is slightly more negative.

Figure 7: **Employment and wage estimates by quantiles of firm FEs, 2015–2018**



Note: I estimate Equation (7) 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.

Worker sorting across firms. Recent papers 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 of high- and low-paying workers into high- and low-paying firms, which could provide a complementary explanation for the observed negative wage effects.[51] I construct the outcome using $\hat{\psi}_j$ values from Equation (14) and exploiting the movements of workers between firms in the post-treatment period. More concretely, the outcome measures the change in the AKM firm FEs in 2018 relative to 2015: $\hat{\psi}_{i,\{j=2018\}} - \hat{\psi}_{i,\{j=2015\}}$. The difference is zero if the worker remains in the same firm during that period.[52]

To determine whether low- or high-wage workers are sorting into higher-paying firms after the

[51] For France, Orefice and Peri (2024) finds 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.

[52] 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.

immigration shock, I present results across seven quantiles of worker FEs. A positive coefficient suggests a positive sorting effect from immigration. Online Appendix Figure A.10a shows that none of the categories yield significant results, indicating no differential sorting due to immigration.[53] 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, Online Appendix Figure A.10b shows there is no evident reallocation of workers between larger or smaller firms post-immigration shock.

7 Heterogeneous Treatment Effects With Machine Learning

In Appendix E, I present a conventional heterogeneity analysis that shows the negative effects of immigration on employment and wages vary by the intersection of worker and firm characteristics. In this section, I move beyond ad hoc subsample splits and adopt a data-driven approach to systematically uncover treatment effect heterogeneity. I use a machine learning framework to identify the subgroups most affected by immigration and the characteristics driving these differences. Specifically, I use the 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. 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.[54] The benchmark specification that the algorithm uses is as follows:

$$\Delta Y_{i,l,2018} = TE(x_i)\Delta\hat{M}_{l,2018} + \Delta\varepsilon_{i,l,2018}, \tag{15}$$

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

[54] I use the `grf` package in `R` based on Athey et al. (2019) to estimate the causal forests.

x_i represents the value of the variables in X_i, and $TE(x_i)$ 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 shock after regressing the observed rate on the instruments. This is done because the algorithm does not allow for multiple instruments. 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. Notably, the wage variable indicates whether the worker earns the minimum wage, and self-employed workers are excluded from this section because 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 and are explained in more detail in Online Appendix C. 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.[55] Using many trees with a minimum node size helps mitigate overfitting concerns.

The first output of this procedure is shown in Online Appendix Figures C.2a and C.2b. These histograms display the predicted treatment effects for wage and employment outcomes, 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 the conditional average treatment effect (CATE), each OOB observation is assigned to a final node of each tree of the forest based on its characteristics.[56] In the histograms, the long dashed line represents the average CATEs, while the short dashed line is the average treatment effect from the standard regression of Equation (7). For both outcomes, the average coefficient from the causal forest aligns with the standard regression, reflecting the accuracy of the average prediction.

[55] I set the tunable parameters of the algorithm to their default values, including honest splitting, and chose a relatively small minimum node size of 300 for precision. In further cross-validation exercise, results hold when I substantially increase the minimum node size.

[56] For all trained trees, the algorithm counts how many times each observation falls into the same terminal node as the training sample to compute similarity weights. It calculates the weighted mean of TE across trees using these weights to determine the treatment effect $TE(x_i)$.

I then use the CATEs 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 (Q1 represents the most negative effect, while Q5 represents the most positive effect). The goal is to compare characteristics between these quintiles rather than making inference from the predicted CATEs.

Online Appendix Tables A.7a and A.7b provide an overview of worker and firm characteristics in the pre-shock period. Native workers experiencing the most negative employment effects are generally older, have the lowest job tenure, and earn the lowest baseline 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 baseline wages. The smallest firms also employ these workers, but in terms of pay premiums, they work in middle- to high-paying firms.[57] From a policy perspective, having the distribution of CATEs is useful for designing targeted measures to mitigate the adverse impacts of immigration in the most affected subgroups.

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 8a and 8b show an overview of the effects across these two dimensions. The greatest negative impact on employment is concentrated among minimum-wage earners in small and medium-sized firms. In contrast, the most negative wage impacts are concentrated among high-wage workers and gradually disappear as the worker's firm grows. Hence, both employment and wages have the greatest negative impacts in smaller firms, but differ across workers' wage distributions, complementing previous results.

Figures 8c and 8d then 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 among high-wage workers in low- to middle-paying firms.

[57] In Online Appendix Figures B.6a and B.6a, 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.

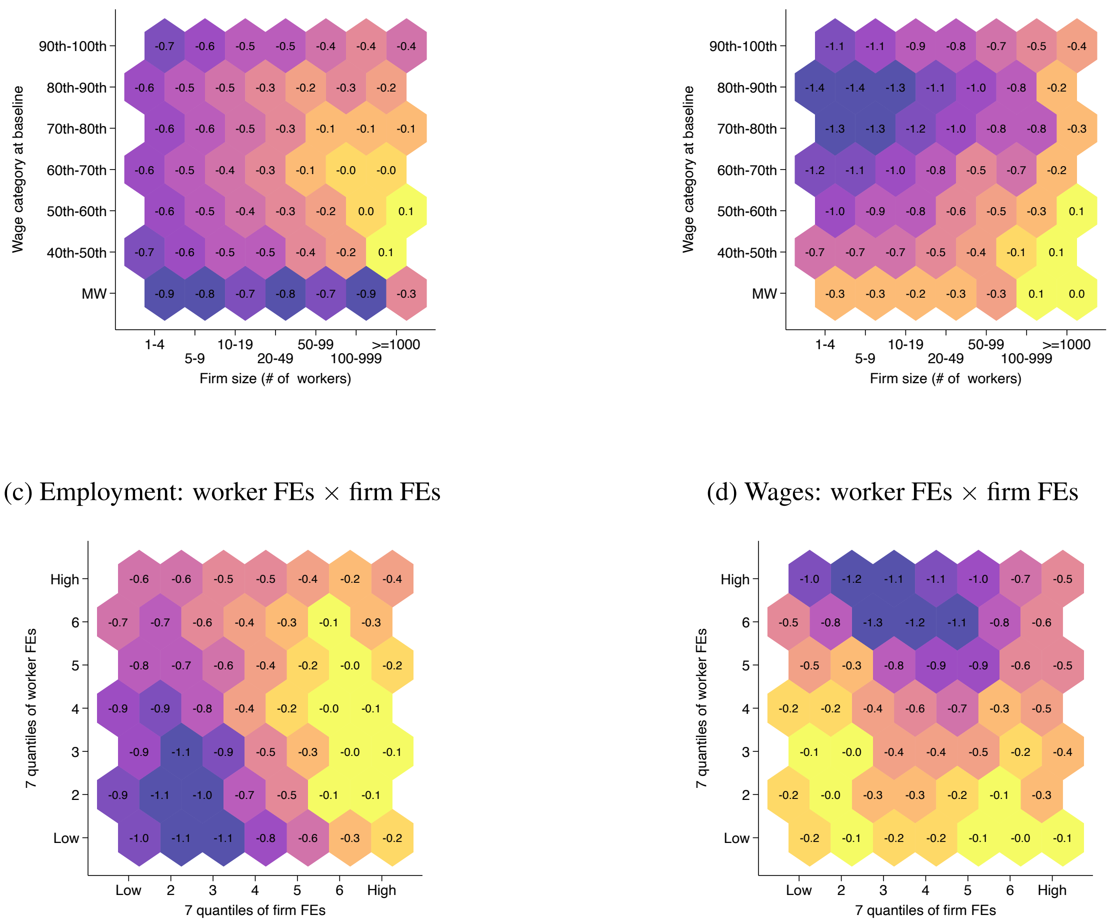
Figure 8: **Heat plots of treatment effects by worker and firm characteristics, 2015–2018**

(a) Employment: worker wage category × firm size

(b) Wages: worker wage category × firm size

(c) Employment: worker FEs × firm FEs

(d) Wages: worker FEs × firm FEs



Note: Each hexagon reports the average individual treatment effect within a subgroup defined by worker and firm characteristics, estimated using a causal forest trained on the OOB sample. Outcomes are changes in employment or wages in 2018 relative to the pre-shock period, with predicted immigration share as the treatment. The sample is restricted to natives aged 25 to 55. Standard errors are clustered at the FUA level. The causal forest is trained on 50% of the main sample.

An approach to summarize these findings involves using a variable importance measure commonly employed in random forests. This measure quantifies how frequently each variable is used as a split within the causal forest, up to the fourth depth of each three. Variables that appear more often are thus more important for explaining treatment effect heterogeneity. In practice, the algorithm assigns each variable a share of all considered splits, with the shares summing to one across all variables, and generates a ranking that serves as a proxy for classifying the key drivers of heterogeneity.

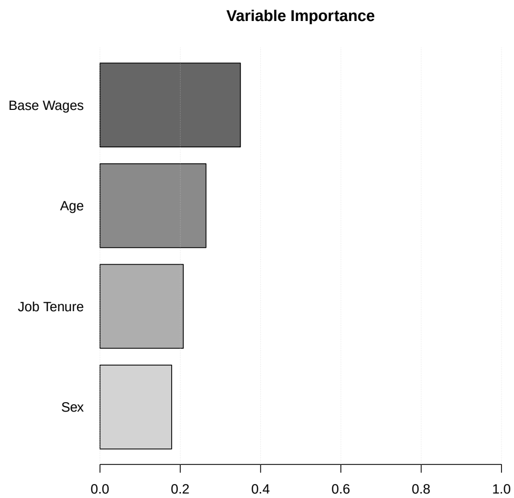
I run the algorithm both with and without firms' variables to illustrate differences in the importance measure. I use the change in employment for each individual between 2018 and the average pre-shock period employment as the outcome. When excluding firms' variables, I find in Figure 9a that base wages, age, and job tenure are more important for explaining the heterogeneity in employment impacts. Notably, firm FEs become the most important variable when including firms' variables, followed by firm size and age. It is important to discuss that 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.

I then use the individual wage growth between 2018 and 2015 as the outcome. Without incorporating firms' variables, Figure 9c shows that base wages are the most important variable, followed by age and job tenure. Notably, when including firms' variables in the causal forest, Figure 9d shows that firm-specific pay premiums, followed by base wages and firm size, become the most important. Note that firm pay premiums and firm size are positively correlated, though not so strongly (correlation coefficient of 0.19). The variable importance analysis reveals that firm-specific pay premiums and firm size are more important than other worker characteristics in explaining wage and employment heterogeneity. Firm FEs appear in 37% of all splits for wages and 30% for employment in the causal forest.[58] This emphasizes that firms in a context of informality can drive more of the heterogeneity in immigration effects on natives, aligning with the findings of Arellano-Bover and San (2024), which show the importance of firms and job mobility in assimilating immigrants into the Israeli labor market.

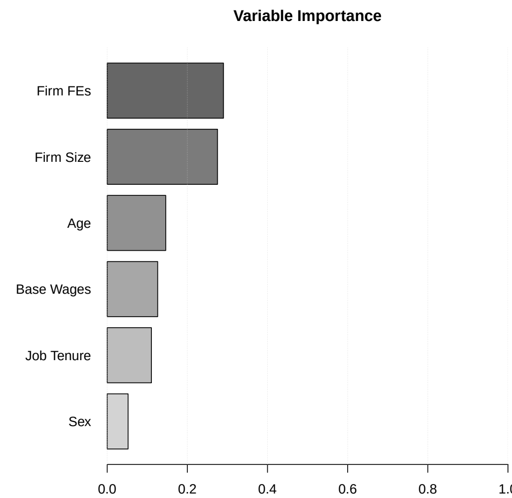
[58]Since base wages are a function of the unobserved firm and worker FEs, I include the constructed worker FEs $\hat{\alpha}_i$ into the algorithm instead of base wages. This adjustment reduces the sample size as every worker must be observed more than once. After including worker FEs, firm FEs and firm size remain the most important variables in explaining the heterogeneity of treatment effects for both employment and wages (see Online Appendix Figures A.11a and A.11b).

Figure 9: **Variable Importance for Formal Employment and Wages in the Causal Forest, 2015–2018**

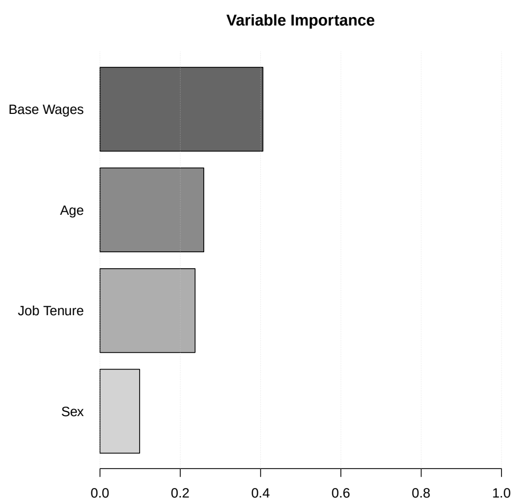
(a) Employment: Without Firm Variables



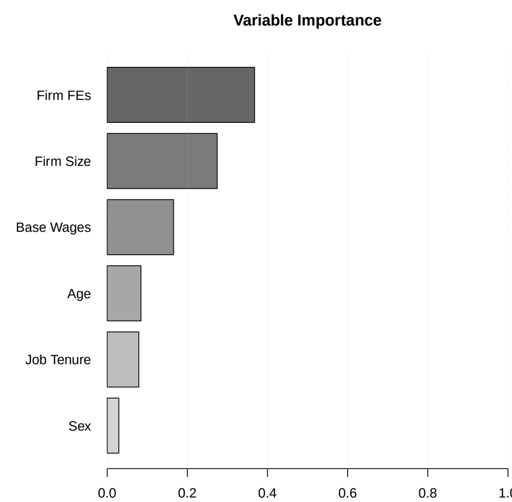
(b) Employment: With Firm Variables



(c) Wages: Without Firm Variables



(d) Wages: With Firm Variables



Note: Variable importance is a simple 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 constructed with a decay exponent of -2 and a maximum depth of 4. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300.

Robustness Checks. Since firm pay premiums correlate with industry type (Card et al., 2024), I include the weighted average of firm FEs at the industry level in the algorithm and find that firm FEs remain the most important variable for both wages and employment (see Online Appendix Figures C.4a and C.4b).[59] Besides, early tree splits may not receive sufficiently larger weights than later ones, so I use a stronger decay exponent of -4 instead of -2 (node k split is weighted $1/4$ of node $k-1$), prioritizing earlier splits. The variable importance ranking remains similar, with firm FEs and firm size leading (see Online Appendix Figures C.3a and C.3a).

Lastly, because variables with fewer possible values can appear mechanically less important (Strobl et al., 2007), I discretize continuous variables into six or seven categories. The importance ranking remains similar for wages but shifts slightly for employment, with firm FE ranked third (see Online Appendix Figures C.5a and C.5b). Nevertheless, exploiting the full range of variable values allows the algorithm to capture non-linearities and interactions more effectively than arbitrary categorization. Another concern is that the algorithm maximizes differences in treatment effects without testing pre-trends within subgroups, assuming strict exogeneity. Online Appendix G address this by checking pre-trends across worker and firm subgroups, finding mostly insignificant estimates. Finally, although recent work proposes hypothesis testing for variable importance in random forests (Hapfelmeier et al., 2023), these methods are not yet available for causal forests and are computationally infeasible in settings with millions of observations.

8 Conclusion

This is the first paper to examine an immigration-induced supply shock in a developing country, exploiting administrative data that covers the entire universe of formal workers and firms. This has several advantages over prior studies. First, administrative panel data 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

[59] I group industries into 19 broader categories based on ISIC revision 4 to address potential misclassification in PILA's 4-digit codes.

the data can identify firm-level heterogeneity, which is crucial for understanding the mechanisms at play following immigration shocks in a context with informality and binding minimum wages. Third, the universe of formal job matches, combined with machine learning, enables the construction of the drivers of heterogeneity that amplify the impact of immigration on formal workers.

My findings suggest that the arrival of Venezuelan immigrants has a negative impact on the employment and wages of individuals in the Colombian formal sector in the short term. However, this coefficient masks important heterogeneous responses. Minimum-wage workers are crowded out of the formal sector, whereas higher-wage workers are not displaced but instead experience negative wage growth. The negative effect on employment and wages is also concentrated in small formal firms. This result is consistent with the theoretical framework, which predicts that small firms that rely more heavily on informal labor for production will reduce formal employment and wages more significantly following an immigration shock if formal and informal labor are substitutes. Beyond firm size, firm pay premiums also explain heterogeneity in immigration effects, as workers in low-paying firms experience a more negative impact on their earnings.

Given substantial heterogeneity across workers and firms, I use a data-driven approach to identify the most important variables that explain the variation in employment and wage effects. Throughout this analysis, firm pay premiums consistently emerge as the most important variable, followed by firm size in most cases. In summary, focusing solely on workers' characteristics when analyzing the labor market impacts of immigration yields an incomplete picture of the main sources of adjustment. Indicating that after immigrants arrive, the focus should not be only on *who* the worker is, but also on the *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.

Akgündüz, Y. E., Aydemir, A., Cilasun, S. M., and Kirdar, M. G. (2024). Propagation of immigration shocks through firm-to-firm trade networks.

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.

Bandiera, A., Dinarte Diaz, L., Rozo, S. V., Schmidt-Padilla, C., Sviatschi, M. M., and Winkler, H. (2023). The unintended consequences of deportations: Evidence from firm behavior in el salvador. *Economic Development and Cultural Change*, 71(4):1331–1358.

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.

Bhuller, M., Delgado-Prieto, L., Hermo, S., and Lorentzen, L. (2025). Wage-setting constraints and firm responses to demand shocks. *arXiv preprint arXiv:2512.19622*.

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.

Burstein, A., Hanson, G., Tian, L., and Vogel, J. (2020). Tradability and the labor-market impact of immigration: Theory and evidence from the united states. *Econometrica*, 88(3):1071–1112.

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., 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.

Clemens, M. A. and Lewis, E. G. (2022). The effect of low-skill immigration restrictions on us firms and workers: Evidence from a randomized lottery. Technical report, National Bureau of Economic Research.

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.

de la Parra, B. S. and Bujanda, L. F. (2024). Increasing the cost of informal employment: Evidence from mexico. *American Economic Journal: Applied Economics*, 16(1):377–411.

de la Parra, B. S., Otero-Cortés, A., and Morales, L. F. (2024). The labor market effects of facilitating social security contributions under part-time employment contracts: Evidence from colombia. *Labour Economics*, 91:102630.

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.

Feinmann, J., Hsu Rocha, R., and Lauletta, M. (2022). Payments under the table. *Available at SSRN 4270665*.

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.

Galaasen, S., Kostøl, A. R., Monras, J., and Vogel, J. (2025). The labor supply curve is upward sloping: The effects of immigrant-induced demand shocks. Technical report, National Bureau of Economic Research.

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

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.

Gulek, A. (2024). Formal effects of informal labor: Evidence from the syrian refugees in turkey. *Available at SSRN 4264865*.

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.

Kline, P. M. (2024). Firm wage effects. Working Paper 33084, National Bureau of Economic Research.

Kugler, A., Kugler, M., and Prada, L. O. H. (2017). Do payroll tax breaks stimulate formality? evidence from colombia?s reform. *National Bureau of Economic Research*, (w23308).

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. (2022). The labor market effects of venezuelan migration to colombia: reconciling conflicting results. *IZA Journal of Development and Migration*, 13(1).

Marchal, L., Ourens, G., and Sabbadini, G. (2023). When immigrants meet exporters: A reassessment of the immigrant wage gap. *Documents de travail du Centre d'Économie de la Sorbonne*.

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.

Morales, L. F. and Medina, C. (2017). Assessing the effect of payroll taxes on formal employment: The case of the 2012 tax reform in colombia. *Economía*, 18(1):75–124.

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.

Osuna Gomez, D. and Medina-Cortina, E. (2023). The effect of immigration enforcement abroad on immigrants' home-country firms. *Available at SSRN 4620336*.

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

Rozo, S. V. and Winkler, H. (2021). Is informality good for business? the impacts of inflows of internally displaced persons on formal firms. *Journal of Human Resources*, 56(4):1141–1186.

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

Worker Responses to Immigration Across Firms: Evidence from Colombia

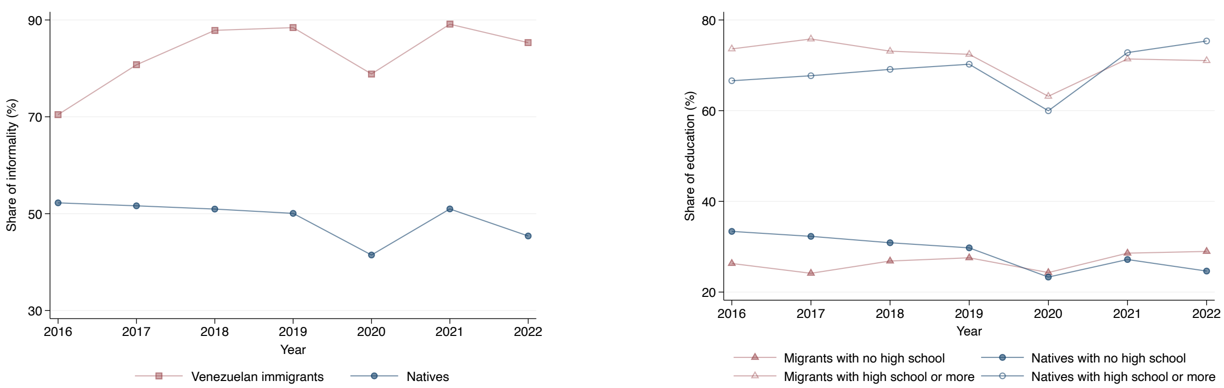
Lukas Delgado-Prieto

A Supplementary Results

Figure A.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.

Table A.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	423.3	506.2	7,335,933
2015	37.2	11.1	0.56	0.497	424.2	493.1	8,391,804
2017	37.8	11.4	0.55	0.497	415.7	482.8	8,064,238
2019	38.2	11.7	0.55	0.498	440.0	510.2	8,363,166

(b) Venezuelans with PEP

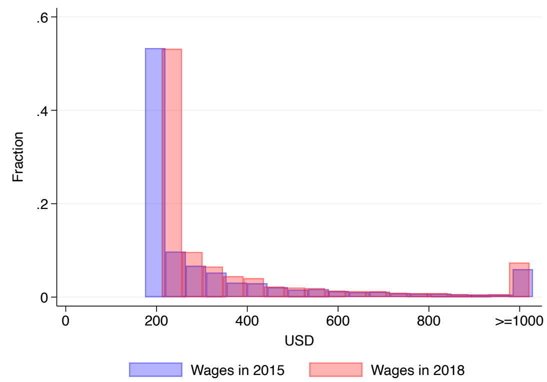
	Age		Male (%)		Real Wages (USD)		
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	N
2018	30.8	7.8	0.68	0.467	245.4	99.8	12,842
2019	31.8	8.1	0.67	0.472	250.9	99.4	42,752

(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	1,315.8	1,542.6	20,977
2015	39.5	10.2	0.64	0.481	1,275.7	1,472.3	27,729
2017	39.3	10.3	0.63	0.483	1,061.8	1,338.9	31,552
2019	39.6	10.4	0.58	0.494	999.3	1,304.5	39,703

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.

Figure A.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.

Table A.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.

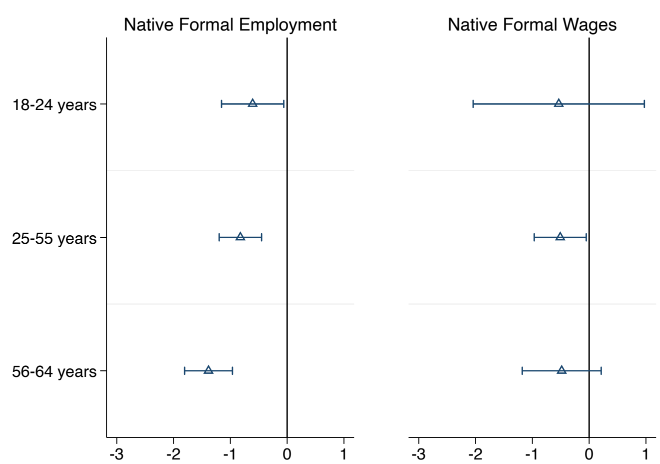
Table A.3: **First stage: The inflow of Venezuelan immigrants and the two instruments**

	(1) $\Delta M_{l,2018}$	(2) $\Delta M_{l,2018}$	(3) $\Delta M_{l,2018}$
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. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

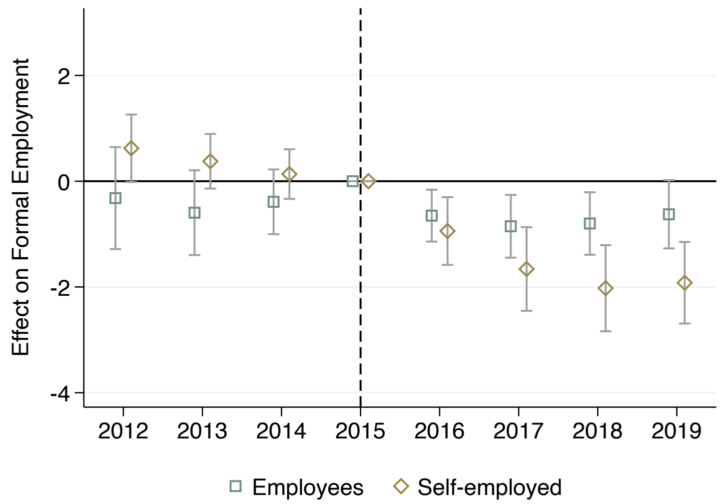
Note: This table reports the coefficient of the first-stage of the immigration shock $\Delta M_{l,2018} * 100$ with distance and distance squared to the nearest crossing bridge and past settlements as explanatory variables.

Figure A.3: **Estimates by extended age categories, 2015–2018**



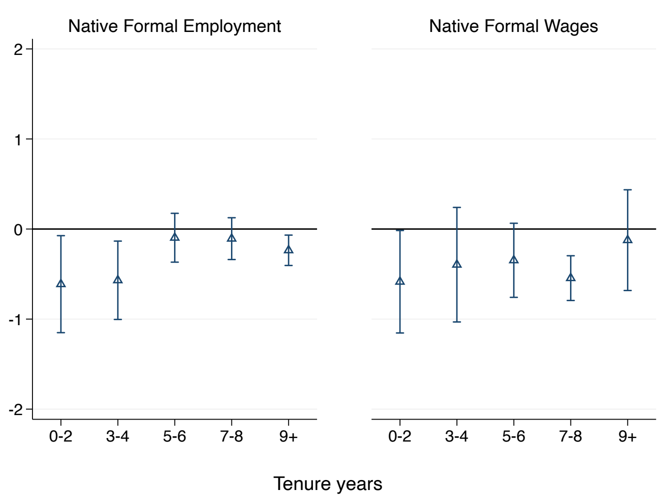
Note: I estimate Equation (7) separately by subgroups. The sample is restricted to natives between 18 and 64 years old. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. I use as controls sex with a dummy for self-employed in the base period. Standard 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.

Figure A.4: **Event study estimates on employment by job type**



Note: I estimate Equation (7) separately by year and type of job. 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.5: **Estimates by job tenure, 2015–2018**



Note: I estimate Equation (7) 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. I constructed pre-shock job tenure from 2007 to 2015. 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.

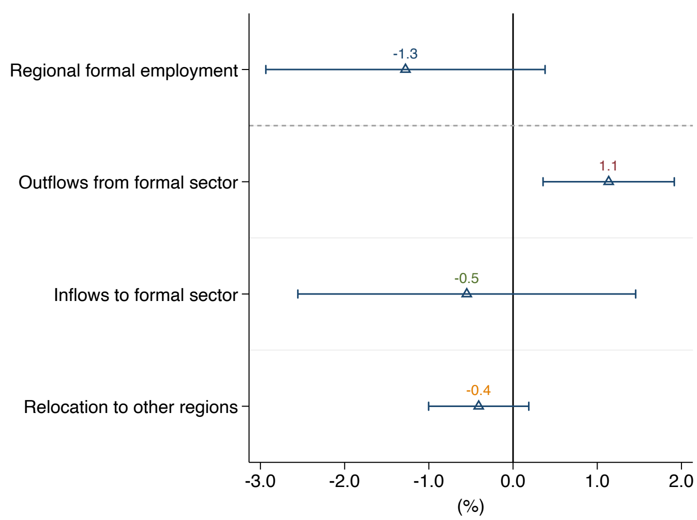
Table A.4: **Employment and wage estimates by age and job tenure, 2015–2018**

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 (7) 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. I constructed pre-shock job tenure from 2007 to 2015. 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: **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.

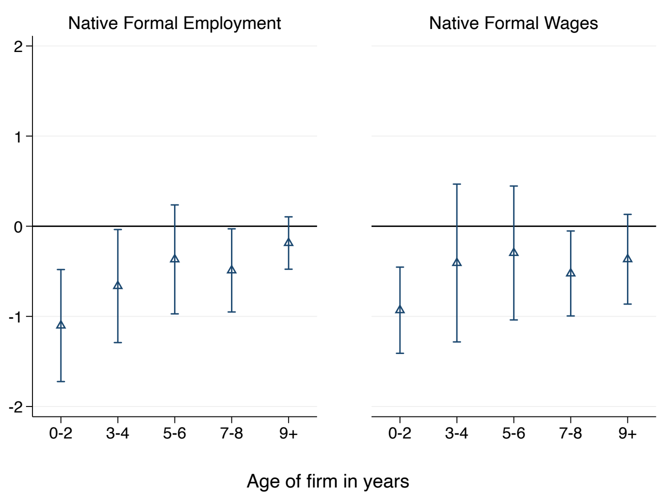
Table A.5: **IV estimates on regional changes of formal workers by age group, 2015–2018**

Age Group	25-30	30-35	35-40	40-45	45-50	50-55
Prob. of Changing FUA	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.7: **Estimates by age of firm, 2015–2018**



Note: I estimate Equation (7) 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 has appeared 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.6: **Employment and wage estimates by firm size and age of firm, 2015–2018**

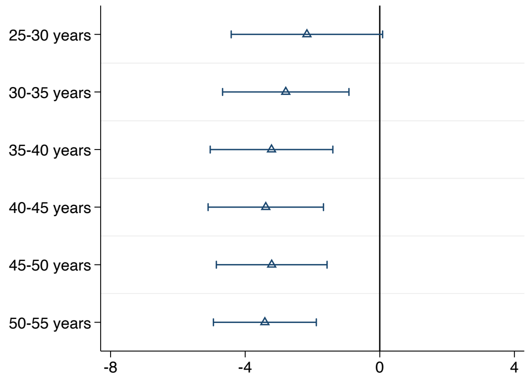
Firm's size	1 to 19 workers		Above 19 workers	
Age of firm	0 to 4 years	5 to 9+ years	0 to 4 years	5 to 9+ years
Prob. of Employment	-0.762**	-0.757***	-1.015**	-0.176
	(0.279)	(0.156)	(0.347)	(0.170)
N	479,715	498,842	923,272	3,700,822
Wages	-1.021*	-0.554	-0.603*	-0.395
	(0.432)	(0.305)	(0.304)	(0.286)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

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

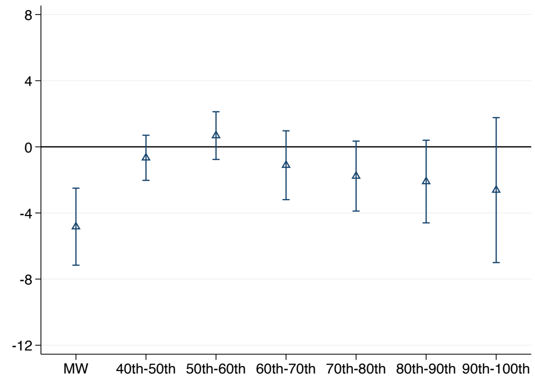
Note: I estimate Equation (7) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. The firm's age is the number of years the firm has appeared discontinuously 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.8: **Earnings estimates by different worker and firm characteristics, 2015–2018**

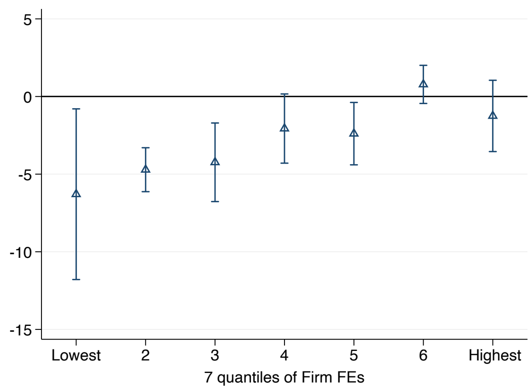
(a) By Age Groups



(b) By Wage Categories

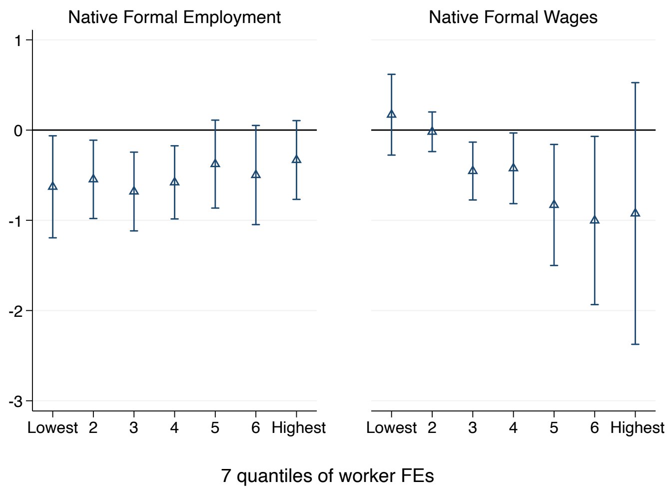


(c) By Pay Premiums



Note: I estimate Equation (7) 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) are already multiplied by 100. Workers are observed in August of each year. Source: PILA, 2013–2019.

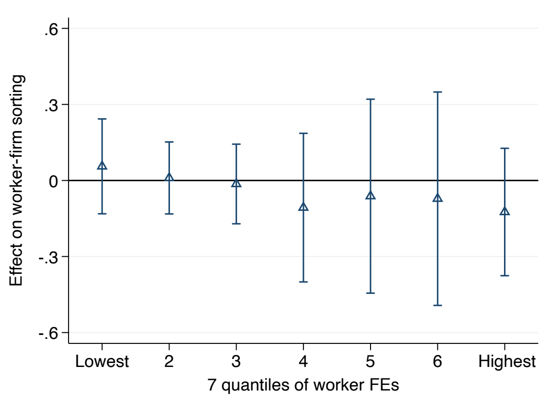
Figure A.9: **Estimates by quantiles of worker FEs, 2015–2018**



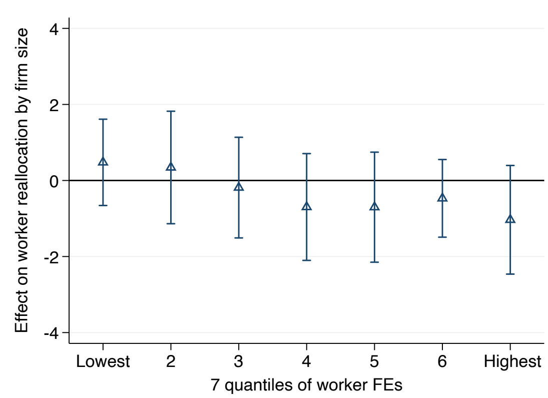
Note: I estimate Equation (7) separately by subgroups. The sample is restricted to native employees between 25 and 55 years old who appear more than once in PILA. Dependent variables are employment relative to the pre-shock period and wages relative to the base period. I compute Worker FEs in the first 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 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.10: **Reallocation estimates by quantiles of worker FEs, 2015–2018**

(a) By 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. I use as controls 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.

Table A.7: **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
Quantile of Firm FEs (1-7)	3.8	5.3	6.0	6.3	6.5

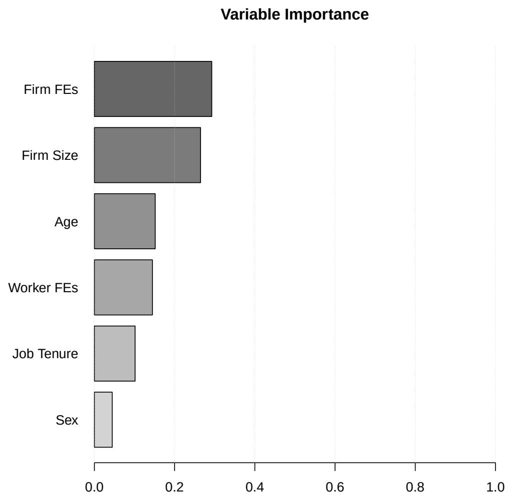
(b) Formal wages

	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
Quantile of Firm FEs (1-7)	5.7	5.8	5.6	5.5	5.5

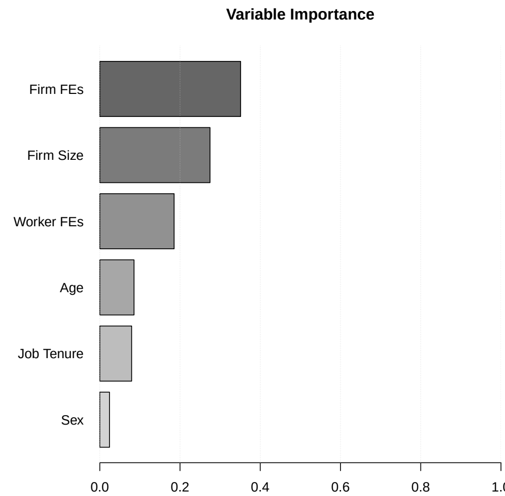
Note: These tables report the average or median statistics for quintiles of treatment effects (Q1 is the most affected and Q5 is the least affected) in terms of employment and wages, according to the predictions of the trained causal forest using the OOB sample. I constructed pre-shock job tenure from 2007 to 2015. The wages are transformed from Colombian pesos to USD using 2020 exchange rates from the World Bank. Source: PILA, August 2015.

Figure A.11: **Variable importance for formal employment and formal wages in causal forest with worker and firm FEs, 2015–2018**

(a) Employment



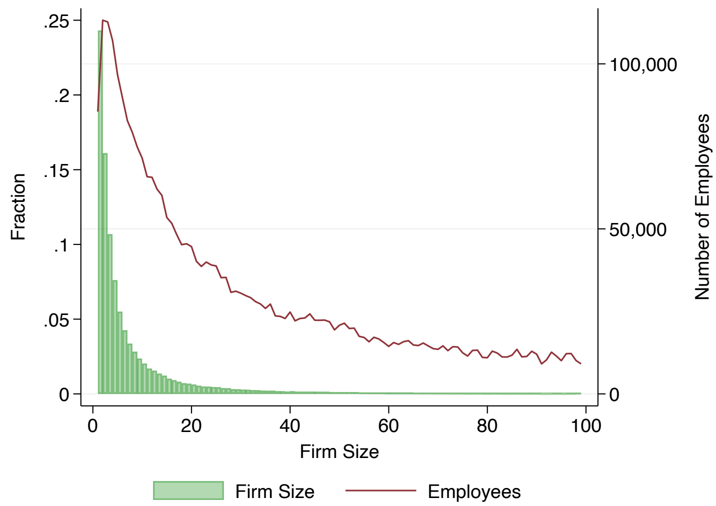
(b) Wages



Note: Variable importance is a simple 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 constructed with a decay exponent of -2 and a maximum depth of 4. 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

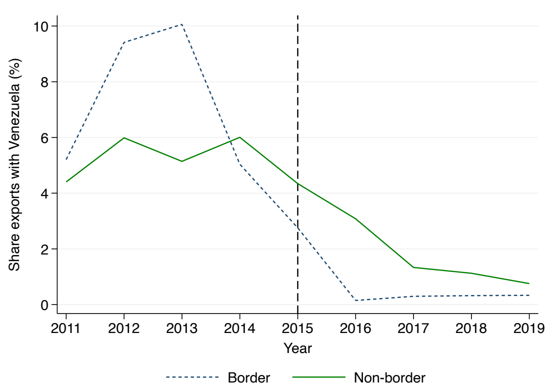
Figure B.1: **Firm size distribution and total employees**



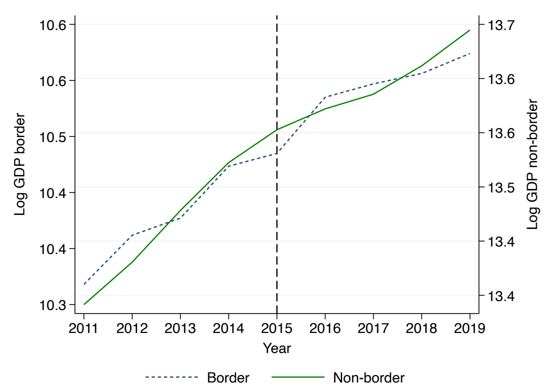
Note: The upper bound of firm size is restricted to 100 workers for the figure. The chosen bin width is 1. Only workers who contribute as employees are taken into account. Source: PILA, August 2015.

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

(a) Share Exports with Venezuela

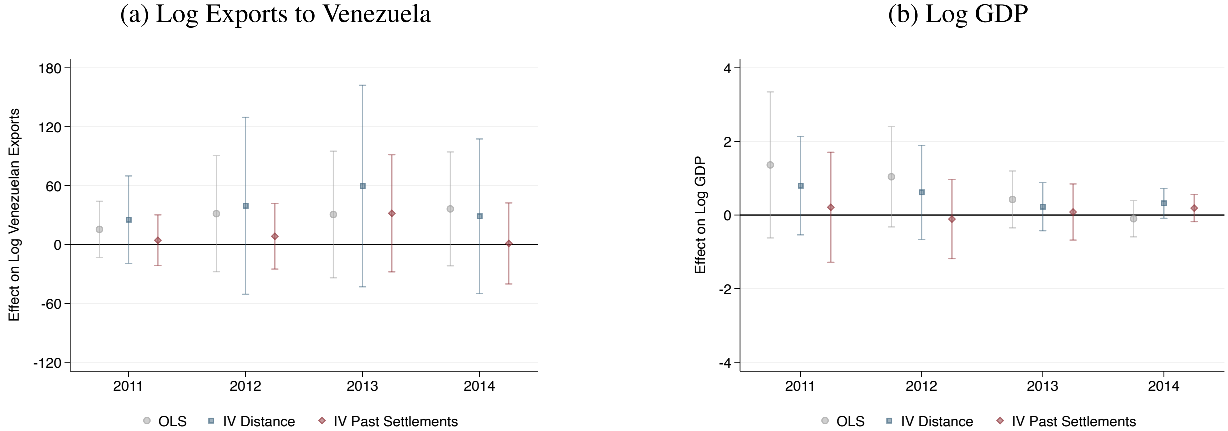


(b) Log GDP



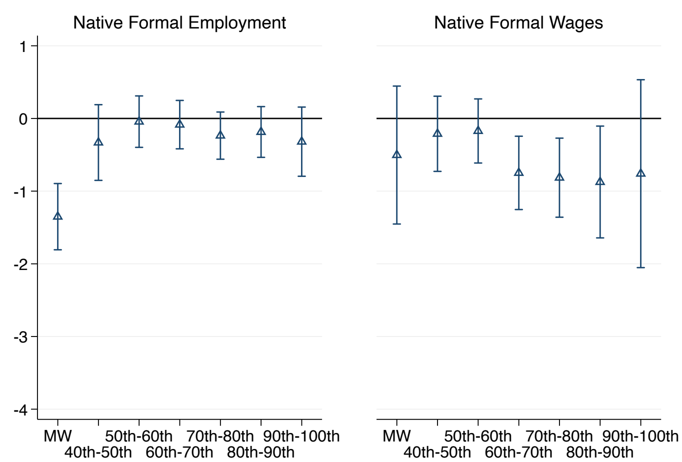
Note: Border departments are *Norte de Santander, La Guajira, and César*. Non-border departments are the rest. Source: Panel (a) *Exportaciones*-DANE, 2011–2019. Panel (b) DANE-*Cuentas Nacionales*, 2011–2019.

Figure B.3: **Pre-treatment estimates on Venezuelan exports and GDP at the department-level**



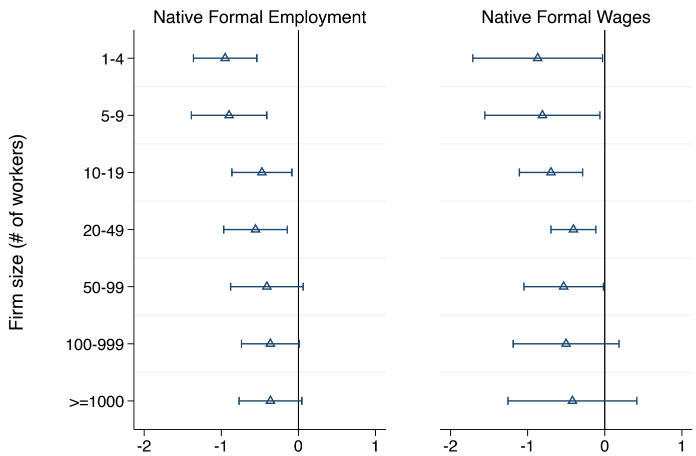
Note: The coefficients come from a department-level regression on 24 departments, based on the migration shock and instruments used in Delgado-Prieto (2024), and they are relative to the outcome in 2015. The coefficients (in percent) are already multiplied by 100. Source: Panel (a) *Exportaciones*-DANE, 2011–2019. Panel (b) DANE-*Cuentas Nacionales*, 2011–2019.

Figure B.4: **Employment and wage estimates by individual wage at baseline including a local labor demand control, 2015–2018**



Note: I estimate Equation (7) separately by subgroups including a Bartik-type control. The control is defined as: $\sum_i s_{i,l,2015} \cdot E_{i,2018}$ where $s_{i,l,2015}$ represents the employment share of industry i in each l in 2015, and $E_{i,2018}$ captures the national-level employment industry in 2018. I construct this measure using two-digit industries based on the ISIC revision 4 classification. 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 .

Figure B.5: **Employment and wage estimates by firm size including industry fixed effects, 2015–2018**



Note: I estimate Equation (7) separately by subgroups. I include 15 industry codes according to ISIC revision 4 classification. 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.

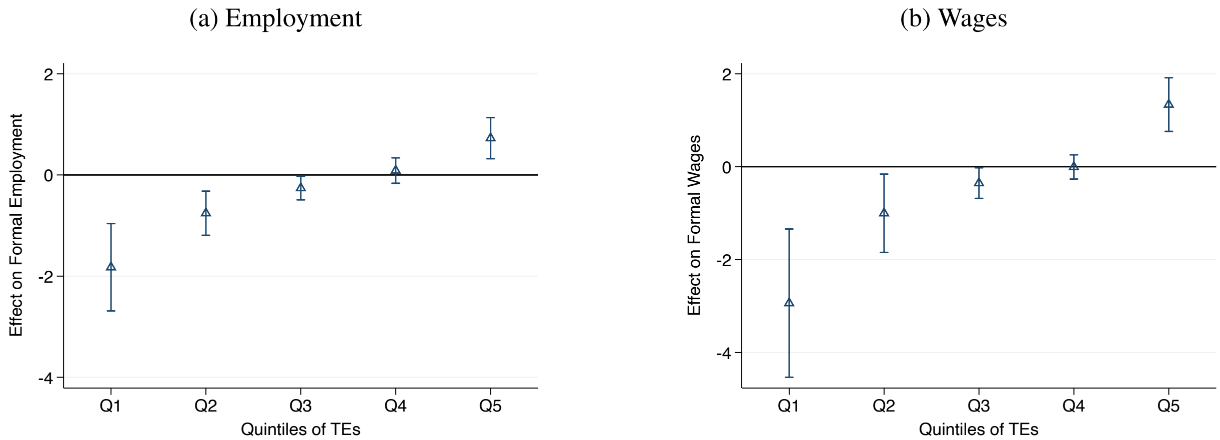
Table B.1: **Robustness checks for the main outcomes, 2015–2018**

	Employment	Wages
Baseline Specification (Both Instruments)	-0.841***	-0.600*
	(0.192)	(0.239)
N	6,706,035	4,090,973
Distance Instrument Only	-0.901**	-0.667*
	(0.290)	(0.339)
N	6,706,035	4,090,973
Past Settlements Instrument Only	-0.953***	-0.749*
	(0.267)	(0.330)
N	6,577,923	4,015,648
Excluding Border Areas with Venezuela*	-1.019*	-0.768
	(0.414)	(0.559)
N	6,577,923	4,015,648
Excluding Bogotá	-0.777***	-0.470**
	(0.180)	(0.173)
N	4,338,192	2,619,237
Adjusting ΔM_l Denominator Using 2005 Census	-0.639***	-0.440*
	(0.168)	(0.203)
N	6,706,035	4,090,973
Including Additional Controls$\star$	-0.828***	-0.689*
	(0.176)	(0.326)
N	6,064,430	4,090,973
Using Real Wages		-0.520*
		(0.207)
N		4,090,973
Top-Coding Local Wages Above 99th Percentile		-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 shock $\Delta M_{l,2018}$. 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.6: **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 (7). 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.

C Machine Learning

Specifically, the machine learning 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 (15) to estimate treatment effects for each partition. Choose the variable with its cutoff value that maximizes the difference in treatment effects using this formula:

$$(TE_l - TE_r)^2.^{b} \tag{C.1}$$

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 of 300 observations, the difference in sample size between the two

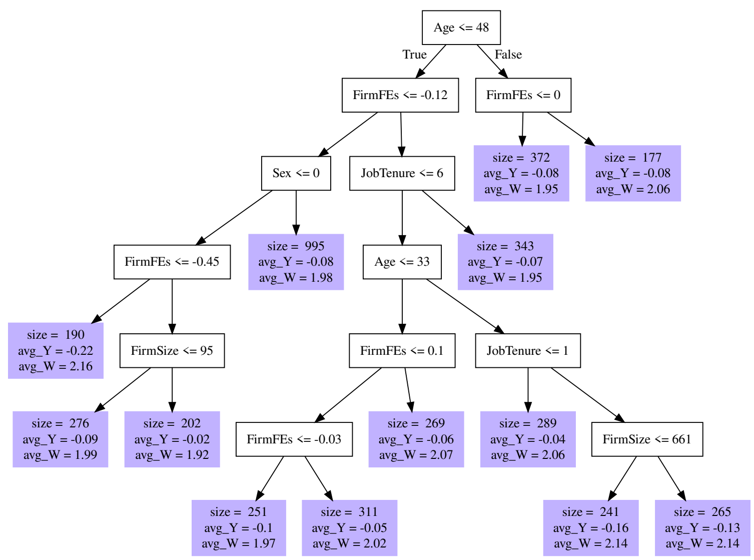
partitions is large (the maximum imbalance split is 5%), or when the split would only yield a difference in the point estimates of the treatment effects that are relatively small.[c]

[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).

[c]As an illustration of a decision tree within the causal forest algorithm, I use a 1% random sample of the main data. Online Appendix Figure C.1 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.

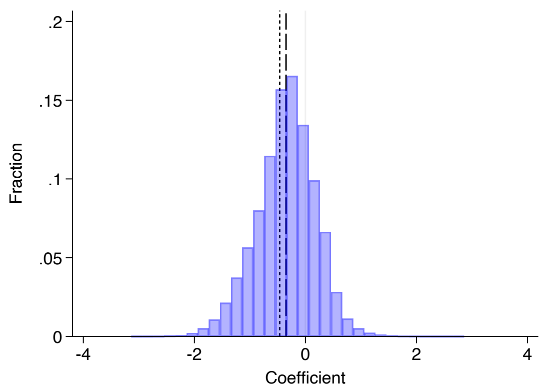
Figure C.1: **Example of decision tree**



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.

Figure C.2: **Histogram of treatment effects for formal employment and formal wages in the causal forest, 2015–2018**

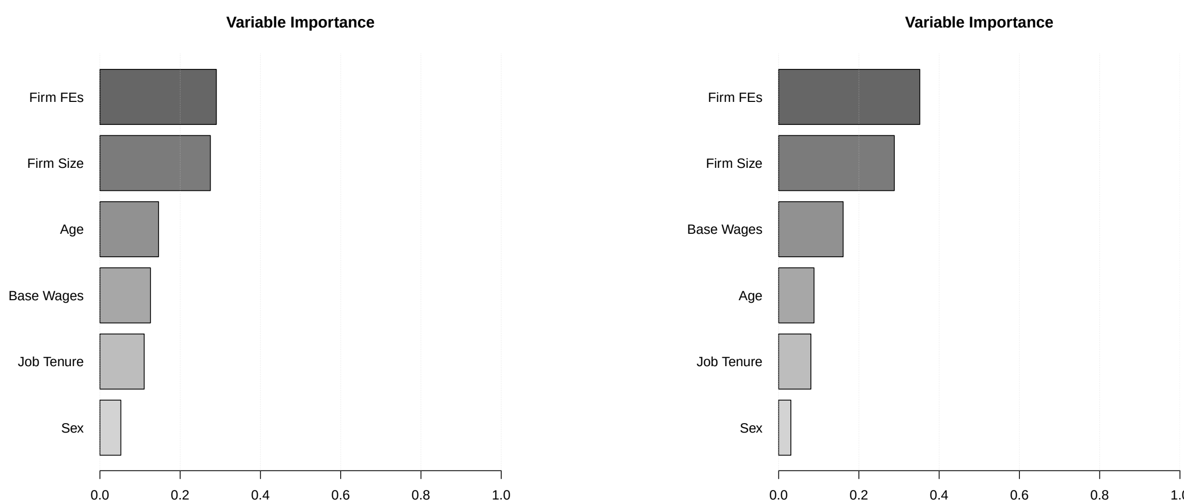
(a) Employment (b) Wages



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.

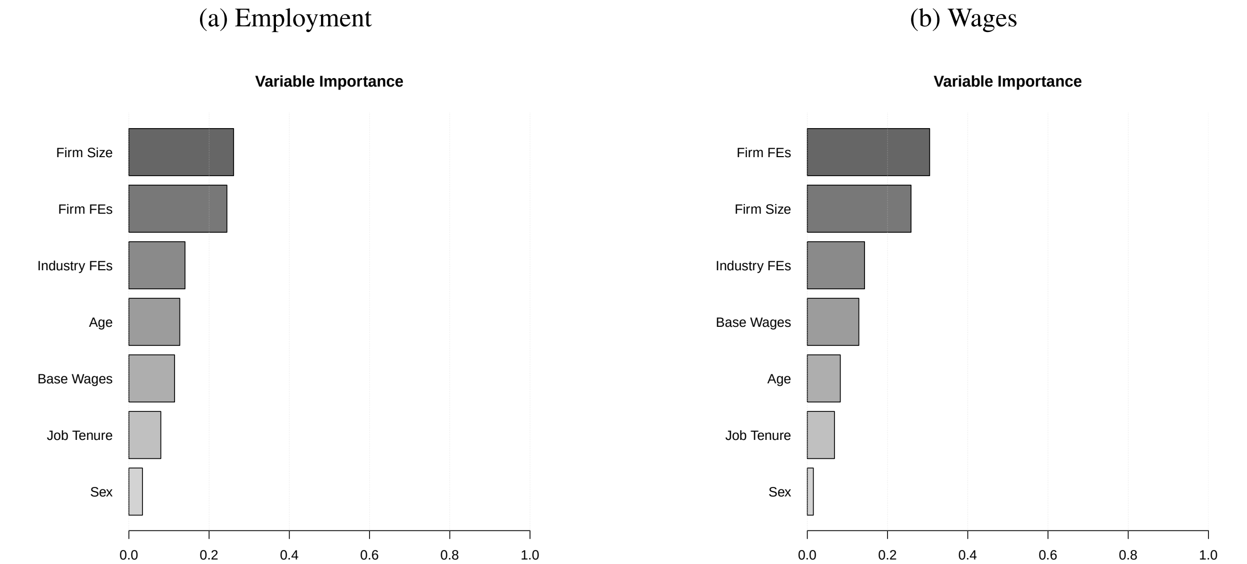
Figure C.3: **Variable importance for formal employment and formal wages in the causal forest with stronger decay exponent, 2015–2018**

(a) Employment (b) Wages



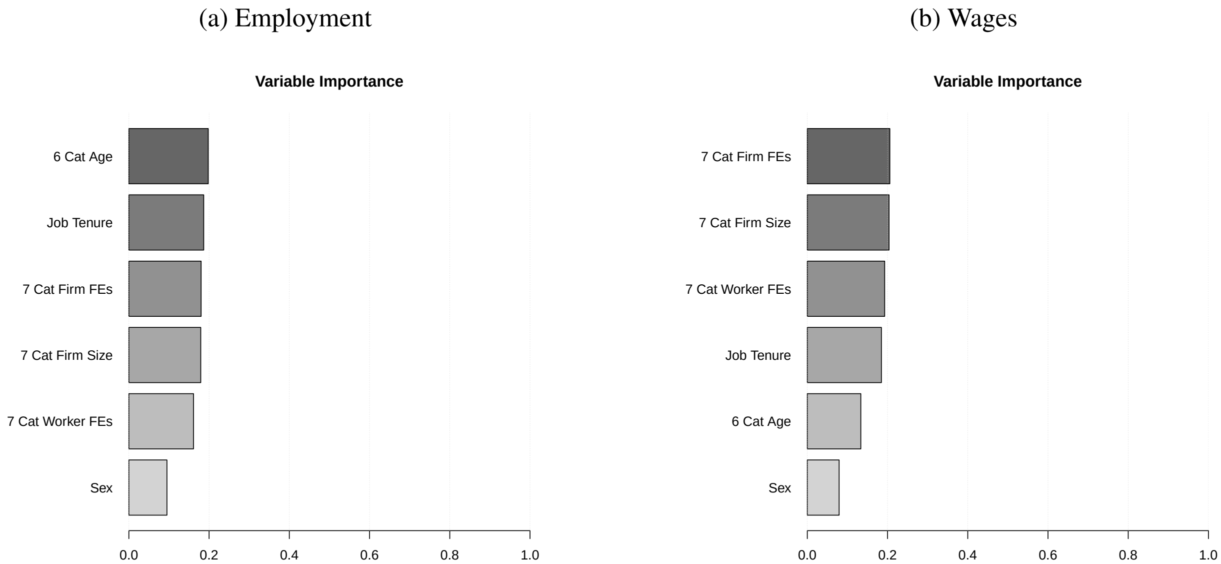
Note: Variable importance is a simple weighted sum of how often the feature f appears in the splits across the leaves of each tree in the forest. For this exercise, the forest comprises 2,000 trees. The analysis sample is restricted to native employees aged 25 to 55. The importance measure is constructed with a decay exponent of -2 and a maximum depth of 4. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300.

Figure C.4: **Variable importance for formal employment and formal wages in causal forest with industry, 2015–2018**



Note: Variable importance is a simple 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 constructed with a decay exponent of -2 and a maximum depth of 4. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300.

Figure C.5: **Variable importance for formal employment and formal wages in the causal forest for categories, 2015–2018**



Note: Variable importance is a simple 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 constructed with a decay exponent of -2 and a maximum depth of 4. The estimation is performed using clusters of FUAs in the causal forest. Additionally, the minimum node size for splitting is set to 300.

D Construction of the Main Sample

The administrative records of the 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.[D.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

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

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 ($\ln w_{it}$). Table D.1 shows descriptive statistics by the seven quantiles of firm FEs and Table D.2 shows the decomposition of the variance of wages $Var(\ln w_{it})$.

With this in mind, the definitions of the variables I use in the main analysis are as follows.

1. **Formal wages**. I use each worker's nominal contribution to the health system 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 consider all individuals who appear in the PILA with a national identity card to be natives. I take all the natives in the sample with a non-negative wage as employed.

3. **Firms**. I only include workers classified as employees in the firm-level data, then aggregate by the firm identifier.

Table D.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 the 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 D.2: Variance decomposition of $\ln w_{it}$

Component	Value	Share of $Var(\ln w_{it})$
$Var(\alpha_i)$	0.222	50.1%
$Var(\psi_{j(i)})$	0.069	15.7%
$2 * Cov(\alpha_i, \psi_{j(i)})$	0.095	21.6%
$Var(\ln w_{it})$	0.443	100%
$Corr(\alpha_i, \psi_{j(i)})$	0.384	

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

E Heterogeneity by Worker and Firm Characteristics

Workers experience varying employment and wage effects based on their characteristics and the types of firms they worked for before immigrants arrived. To determine the subgroups most affected in a standard way, I restrict the regressions to the intersection of subgroups where earlier findings indicate more negative coefficients.

First, Table E.1a 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, whereas for self-employed workers, it 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 E.1b 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 1.9% reduction for a 1 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 E.1: **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.

F Derivation of Model in Section 3

In this Online Appendix section, I explain the derivations of the Equations in subsection 3. First, to derive the firm-specific optimal wages, I maximize the profit Equation (4) for each type of worker:[F.1]

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

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

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

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

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

$$\ln w_{F_j} = \ln\left(\frac{\beta_F}{1+\beta_F}\right) + \ln(D_j T_j^{\varepsilon} \varepsilon \alpha_F) + (\rho - 1)\ln F_j - \ln(1+\tau_F) + \left(\frac{\varepsilon-\rho}{\rho}\right)\ln(\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho}). \tag{F.4}$$

Below, I analyze the firms where the minimum wage binds for formal workers ($w_{F_{Min}}$), such that $\overline{w}_{F_{Min}} \geq w_{F_j}$, as firms' optimal choices would be distorted. This is more likely to happen in low-productivity firms. Broadly, this model predicts that firms with higher productivity (T_j) or demand (D_j) will pay higher wages, holding amenities constant. I then study how firm-level wages respond to an immigration shock that shifts the aggregate informal labor supply outwards ($d\mathfrak{I}$)[F.3]:

$$\frac{d\ln w_{I_j}}{d\mathfrak{I}} \cdot \mathfrak{I} = (\rho - 1 - \eta)\frac{d\ln I_j}{d\ln \mathfrak{I}} + \left(\frac{\varepsilon-\rho}{\rho}\right)\frac{(\alpha_I \rho I_j^{\rho-1}\frac{dI_j}{dI} + \alpha_F \rho F_j^{\rho-1}\frac{dF_j}{dI})}{\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho}} * \mathfrak{I}, \tag{F.5}$$

$$\frac{d\ln w_{F_j}}{d\mathfrak{I}} \cdot \mathfrak{I} = (\rho - 1)\frac{d\ln F_j}{d\ln \mathfrak{I}} + \left(\frac{\varepsilon-\rho}{\rho}\right)\frac{(\alpha_I \rho I_j^{\rho-1}\frac{dI_j}{dI} + \alpha_F \rho F_j^{\rho-1}\frac{dF_j}{dI})}{\alpha_I I_j^{\rho} + \alpha_F F_j^{\rho}} * \mathfrak{I}. \tag{F.6}$$

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

$$\varepsilon_{w_{I_j},\mathfrak{I}} = -(1+\eta-\rho)\varepsilon_{I_j,\mathfrak{I}} + (\varepsilon-\rho)(s_{I_j}\varepsilon_{I_j,\mathfrak{I}} + s_{F_j}\varepsilon_{F_j,\mathfrak{I}}), \tag{F.7}$$

$$\varepsilon_{w_{F_j},\mathfrak{I}} = -(1-\rho)\varepsilon_{F_j,\mathfrak{I}} + (\varepsilon-\rho)(s_{I_j}\varepsilon_{I_j,\mathfrak{I}} + s_{F_j}\varepsilon_{F_j,\mathfrak{I}}). \tag{F.8}$$

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 (2) and (3) after an immigration shock:

[F.2] If $\beta_L = 9$ then workers are paid 90% of their marginal product of labor.

[F.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 that there are no strategic interactions among them.

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

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

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

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

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

Rearranging these expressions, I find that:

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

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

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

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

Here, I define $\Omega_j = \frac{1}{\xi_{I_j}\xi_{F_j} - (\varepsilon-\rho)^2 s_{I_j}\beta_I s_{F_j}\beta_F}$. Last, I replace ξ_{I_j} inside of (F.15) to find the Equation (5) in the main text. Next, I plug Equation (5) inside Equation (F.13) to find that:

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

In this case, the elasticity is going to be negative $\varepsilon_{w_{I_j},\mathcal{I}} < 0$.[F.5] Finally, after finding that 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 (F.16) into Equation (F.9):

[F.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.

[F.5]It suffices for $\varepsilon_{w_{I_j},\mathcal{I}} < 0$ that $1 \geq s_{I_j}\big(1 + s_{F_j}\Omega_j(\varepsilon-\rho)\beta_F\big)$, which always holds when $\rho > \varepsilon$. If $\rho < \varepsilon$, the inequality still holds since $1+\eta-\rho > \varepsilon-\rho$.

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

After simplifying the previous expression, I find that:

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

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 are close substitutes or not.

Case when the minimum wage binds for certain firms ($\overline{w}_{F_{Min}} \geq w_{F_j}$).

Among firms paying the minimum wage, it is useful to distinguish between two regimes. Some firms will be *demand-constrained* by the minimum wage, in the sense that formal employment is pinned down by equating $\text{MRPL}_{F_j} = (1+\tau_F)\overline{w}_{F_{Min}}$, while labor supply at the minimum wage is slack, meaning they do not hire all workers willing to work for that wage. Other firms will be *supply-constrained*, in which case formal employment is directly pinned down by the labor supply equation evaluated at the minimum wage and does not respond to an informal supply shock as the minimum wage is held fixed. In what follows, we focus on demand-constrained firms that pay the minimum wage and respond to the immigration shock.

Define the demand-constrained firms as j_m. In those firms, the formal wage response is muted, $\varepsilon_{\overline{w}_{F_{Min}},\mathcal{I}} = 0$, so that the formal employment elasticity is equal to:

$$\varepsilon_{F_{jm},\mathcal{I}} = \frac{(\varepsilon-\rho)\, s_{I_{jm}}}{(1-\varepsilon)+(\varepsilon-\rho)\, s_{I_{jm}}}\, \varepsilon_{I_{jm},\mathcal{I}}. \tag{F.19}$$

Then, the informal wage elasticity for these firms can be derived from Equations F.11 and F.10, and it is equal to:

$$\varepsilon_{w_{I_{jm}},\mathcal{I}} = -(1+\eta-\rho)\big(1+\beta_I\varepsilon_{w_{I_{jm}},\mathcal{I}}\big)+(\varepsilon-\rho)\Big(s_{I_{jm}}\big(1+\beta_I\varepsilon_{w_{I_{jm}},\mathcal{I}}\big)+s_{F_{jm}}\frac{(\varepsilon-\rho)\, s_{I_{jm}}}{(1-\varepsilon)+(\varepsilon-\rho)\, s_{I_{jm}}}\big(1+\beta_I\varepsilon_{w_{I_{jm}},\mathcal{I}}\big)\Big). \tag{F.20}$$

This expression simplifies to:

$$\varepsilon_{w_{I_{jm}},\mathcal{I}} = \frac{-(1+\eta-\rho)+(\varepsilon-\rho)s_{I_{jm}}\left(1+s_{F_{jm}}\frac{(\varepsilon-\rho)}{(1-\varepsilon)+(\varepsilon-\rho)s_{I_{jm}}}\right)}{1+\beta_I\left((1+\eta-\rho)-(\varepsilon-\rho)s_{I_{jm}}\left(1+s_{F_{jm}}\frac{(\varepsilon-\rho)}{(1-\varepsilon)+(\varepsilon-\rho)s_{I_{jm}}}\right)\right)}. \tag{F.21}$$

This negative elasticity is larger in absolute value than that of firms for which the minimum wage does not bind, reflecting the fact that adjustment occurs entirely through the informal margin. From this expression, the informal employment elasticity is:

$$\varepsilon_{I_{jm},\mathcal{I}} = 1+\beta_I \varepsilon_{w_{I_{jm}},\mathcal{I}} = \frac{1}{1+\beta_I\left((1+\eta-\rho)-(\varepsilon-\rho)s_{I_{jm}}\left(1+s_{F_{jm}}\frac{(\varepsilon-\rho)}{(1-\varepsilon)+(\varepsilon-\rho)\,s_{I_{jm}}}\right)\right)}. \quad \text{(F.22)}$$

This elasticity is again positive and larger than that of firms where the minimum wage does not bind. Finally, combining this result with equation (F.19), the formal employment elasticity for demand-constrained firms paying the minimum wage is:

$$\varepsilon_{F_{jm},\mathcal{I}} = \frac{(\varepsilon-\rho)\,s_{I_{jm}}}{\left((1-\varepsilon)+(\varepsilon-\rho)\,s_{I_{jm}}\right)\left(1+\beta_I\left((1+\eta-\rho)-(\varepsilon-\rho)s_{I_{jm}}\left(1+s_{F_{jm}}\frac{(\varepsilon-\rho)}{(1-\varepsilon)+(\varepsilon-\rho)\,s_{I_{jm}}}\right)\right)\right)}. \quad \text{(F.23)}$$

For firms that are supply constrained at the minimum wage, formal employment is pinned down by the labor supply function and does not respond to the immigration shock, so that $\varepsilon_{F_{jm},\mathcal{I}} = 0$.

G Additional Pre-Trends Checks

This subsection of the Online Appendix tests for differential trends in outcomes across worker and firm characteristics.

Table G.1: Event study estimates on pre-treatment periods of Figure 3

	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 G.2: Event study estimates on pre-treatment periods of Figure 4

	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 G.3: Event study estimates on pre-treatment periods of Figure 5

	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 G.4: Event study estimates on pre-treatment periods of Figure 7

	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.

H Size and Location of FUAs

Table H.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 H.2		
No FUA assigned	417,188	(5.9)
Total	7,123,223	(100)

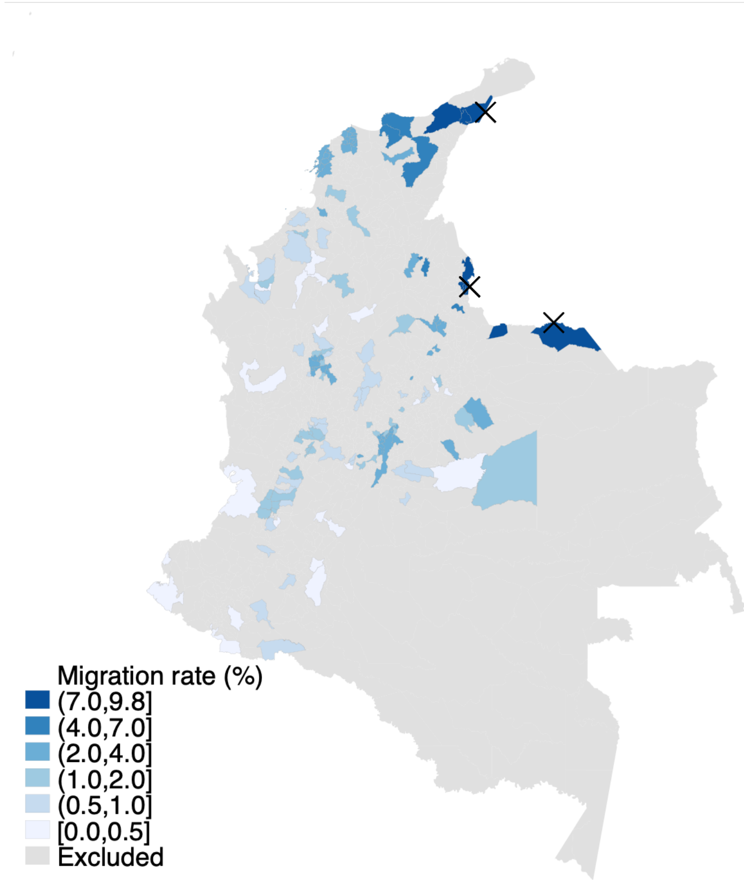
Note: This Table reports the number of workers from the PILA by FUAs 1 to 53. The name represents the main city of FUA, but often they aggregate multiple municipalities according to Sanchez-Serra (2016). The sample is restricted to natives aged 25 to 55. Workers are observed in August of each year. Source: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

Table H.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 the 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 H.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.