*Jeremy Lebow**

The labor market effects of Venezuelan migration to Colombia: reconciling conflicting results†

Abstract

The recent mass migration of Venezuelans to Colombia has become a focal point for economists interested in the labor market effects of migration in developing countries. Existing papers studying this migration wave have consistently found negative effects on the hourly wages of native Colombians, which are most concentrated among less-educated natives working in the informal sector. However, the magnitude and significance of this wage effect varies substantially across papers. I explore the potential specification choices that drive this variation. Differences in how migration is measured are particularly important: exclusion of a subset of migrants from the migration measure, according to characteristics such as time of arrival, amounts to an omitted-variable bias that will tend to inflate the estimated wage effect. In my own analysis based on the total migration rate across 79 metropolitan areas and by using an instrument based on historical migrant locations, I estimate a native hourly wage effect of −1.05% from a 1 percentage point increase in the migrant share or an effect of −0.59% after controlling for regional time trends, alongside little-to-no effect on native employment. Native movements across occupation skill groups and geography are small and do not play a meaningful role in mitigating local wage effects. Wage effects are also larger in cities that have a higher baseline informality rate and lower ease of starting a business.

Current version: February 17, 2022
Keywords: immigration, labor markets, informality
JEL codes: F22, J21, J46

Corresponding author: Jeremy Lebow
jal99@duke.edu

† This paper was previously circulated under the title "Refugees in the Colombian Labor Market: The Consequences of Occupational Downgrading".

Department of Economics, Duke University, 213 Social Sciences, Durham, N.C. 27708

© The Author(s). 2022. Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
Cite as: Lebow. *IZA Journal of Development and Migration* (2022) 13:05
https://doi.org/10.2478/izajodm-2022-0005

1 Introduction

Between 2015 and 2019, approximately 1.8 million Venezuelans fled into neighboring Colombia, largely fleeing poverty and violence induced by the economic and political crisis in Venezuela. This large and unprecedented migration wave increased Colombia's population by almost 4% and stimulated debate over the potential positive and negative economic consequences of the migration. The Colombian case is part of an alarming trend of increasing forced displacement around the world: the United Nations High Commissioner for Refugees (UNHCR) estimates that the number of forcibly displaced worldwide increased from 41 million to 79.5 million between 2010 and 2019.

In the canonical framework, an exogenous increase in labor supply induces a wage decrease when labor demand is sloping downward or a decrease in employment when wages fall below the reservation wage. Labor is distinguished by characteristics such as education, experience, or sector, and the effect of migration on any particular subgroup thus depends on the composition of migrants and their degree of substitutability with natives (Peri, 2016; Ottaviano and Peri, 2012; Borjas, 2003; Altonji and Card, 1991). Firms also react to increased labor supply by investing in additional capital, and we therefore expect persistence of economic effects to decrease with the speed of capital adjustment. Moreover, migrants boost consumer demand, transfer human capital and networks (Bahar *et al.,* 2019, 2020), and can stimulate firm technological upgrading and native occupational upgrading (Foged and Peri, 2016). Thus, the effect of migration on native labor outcomes is not always expected to be negative and will vary depending on the characteristics of both the migration wave and the labor market. In general, when there are detrimental effects, we expect to see them most concentrated on natives with similar skills and working similar jobs as migrants.

The context of Venezuelan migration to Colombia is unique and interesting for various reasons. First, this is a developing country setting, in which around 60% of natives and 90% of Venezuelan migrants are in the informal sector, defined according to enrollment in mandatory health and pension schemes. The informal sector has no minimum wage and tends to have high turnover rates, increasing wage flexibility (Agudelo and Sala, 2017; Guriev *et al.,* 2019).[1] Furthermore, around 25% of natives and 30% of migrants are self-employed own-account (with no employees) workers and, thus, compete directly over prices. Second, Venezuelan migrants and Colombian natives speak the same language and have a similar cultural background, which increases their substitutability in the workforce (Braun and Mahmoud, 2014). This also limits the scope for natives to respond to migration by upgrading to communication-intensive tasks (Peri and Sparber, 2009; Peri *et al.,* 2020). Third, Colombia has experienced extensive internal migration from decades of civil war and has an unemployment rate that has hovered between 8% and 11% since 2010, indicating limited capacity to mobilize capital to absorb an expanding workforce (Calderón-Mejía and Ibáñez, 2016; Morales, 2018). Finally, an important characteristic of this migration is the occupational downgrading of migrants: while Venezuelan migrants and Colombian natives have similar levels of education, Venezuelan migrants are heavily concentrated in occupations, such as restaurant work, construction, street vending, and domestic service, which typically require less education. Because migrants in Colombia

1 Many Venezuelans in Colombia have access to legal work status. However, in practice, the vast majority remain in the informal sector, where lack of work status is not a barrier to employment.

are mostly competing with less-educated workers in the informal sector, we expect economic effects to be most concentrated among these natives, with potential benefits for more-educated natives (Dustmann *et al.,* 2013).

In the first part of this paper, I study the effects of migration on native Colombians' economic outcomes. I do this using variation in the migration rate across 79 metropolitan areas constructed according to commuting patterns, data on the labor market outcomes of migrants and natives obtained from the official labor survey of Colombia (the Colombian National Integrated Household Survey (*Gran Encuesta Integrada de Hogares* [GEIH]), and an instrumental variable (IV) strategy based on historical migration rates. I find that a 1 percentage point (pp) increase in the migrant share of the population decreases native hourly wages by 1.05%, and this decreases to −0.59% after accounting for region-specific time trends. In 2019, the migrant share across metro areas varied from 1% at the 10th percentile to 8.6% at the 90th percentile, over which an effect size of −0.59 is associated with a 4.5% wage decrease. While these wage effects do not vary significantly by age and gender, they are larger for less-educated natives, especially those who are in the informal sector, are self-employed, or are working in low-skill occupations. These magnitudes are larger than those typically observed in the literature and are consistent with evidence that the economic effects of migration tend to be largest in middle-income developing countries, especially for less-educated workers in the informal sector (Verme and Schuettler, 2021).

Unlike with earnings, I find little evidence for effects on the employment margin, consistent with Colombian workers having low reservation wages. There is no effect of migration on unemployment. Among natives younger than 25 years of age, migration causes a reduction in labor force participation, which is partially explained by a reduction in school dropouts; however, this is not robust to the dropping of metro areas close to the Venezuelan border with high migration rates.

This analysis is "nonstructural" in that it makes no assumptions about mechanisms, which may include any of those discussed at the start of this Introduction. This approach is the most common in the literature studying the labor market effects of migration, and it is informative about the total effects of migration overall and for subgroups of natives, for example, by age, gender, education, or sector (Dustmann *et al.,* 2016).[2]

After estimating this baseline specification, I conduct a sensitivity analysis for various specification choices, including choice of instrument, unit of geographic variation, migration data source, and definition of the migration share. The motivation for this is that a variety of studies have used a similar approach to study Venezuelan migration in Colombia (Caruso *et al.,* 2021; Delgado-Prieto, 2021; Penaloza-Pacheco, 2021; Santamaria, 2020; Bonilla-Mejía *et al.,* 2020). While they are consistent in finding negative hourly wage effects concentrated on less-educated natives, they find wildly different magnitudes for those wage effects, ranging from −0.5% to −7.6% in response to a 1 pp increase in the migrant share.[3] I show that the

2 In another paper (Lebow, 2022), I study this same migration wave by estimating a production function with imperfect substitutability between migrants and natives (Ottaviano and Peri, 2012; Manacorda *et al.*, 2012). This entails making explicit assumptions about the structure of production, which allows me to estimate counterfactual wage effects under alternative scenarios, such as one in which migrants do not downgrade. The estimated wage effects loosely match the nonstructural estimates presented in this paper.

3 Another paper (Rozo and Vargas, 2021) studies the effects of Venezuelan migration on right-wing voting in Colombia. It also looks at economic outcomes and finds imprecise negative wage effects for natives. However, the magnitudes are not directly comparable with those in other papers because the authors estimate the effect of a change in the predicted migrant share, rather than the observed migrant share.

smaller estimates can be explained by failing to use an instrument to account for the endogenous sorting of migrants into locations, which biases the estimated wage effect toward zero. I show that among the various factors that explain the larger estimates, the most important is related to implicit assumptions made while calculating the migrant share of the population. Specifically, by restricting migration to include only those who arrived over the past 12 months or by excluding Colombian-born return migrants, the estimated wage effect become substantially inflated due to what I argue is best understood as omitted-variable bias.[4] This is because the location of previously arrived migrants (or return migrants) is strongly correlated with that of recently arrived migrants (or Venezuelan-born migrants), and it is reasonable to expect that both groups affect the local economy.

To illustrate this concept, let "M1" represent the share of the population of migrants who arrived in the past 12 months, and let "M5" represent the share who arrived in the past 13–60 months. Regress the average native log-wage in region k and year t on $M1_{kt}$ with region and year fixed effects[5]:

$$\ln W_{kt} = \beta M1_{kt} + \gamma_k + \delta_t + \epsilon_{kt} \tag{1}$$

Assume $M1_{kt}$ and ϵ_{kt} are uncorrelated conditional on the fixed effects. Let the true effect of $M1_{kt}$ and $M5_{kt}$ on $\ln W_{kt}$ be α_1 and α_5, respectively. Then, β, the marginal effect of $M1_{kt}$ on $\ln W_{kt}$ (suppressing conditionality on the fixed effects from the notation for brevity), is obtained as follows:

$$\hat{\beta} = \frac{\partial \ln W_{kt}}{\partial M1_{kt}} = \alpha_1 + \alpha_5 \frac{\partial M5_{kt}}{\partial M1_{kt}} \tag{2}$$

Thus, excluding $M5_{kt}$ from the regression generates an omitted-variable bias equal to $\alpha_5 \frac{\partial M1_{kt}}{\partial M5_{kt}}$. Only if the excluded group is uncorrelated with the included group, or if the excluded group has no effect on native labor market outcomes, will this omission not generate a bias. If an instrument is being used for $M1_{kt}$, in which case, $M1_{kt}$ can be replaced with $\widehat{M1}_{kt}$ in the above equations, then the problem persists so long as the instrument is also correlated with the excluded group. In this case, the instrument based on historical migration rates is correlated with both 1-year and 5-year migrant shares, as well as with both foreign-born and return migrant shares. In the analysis in this paper, a regression of $M5_{kt}$ on the predicted $\widehat{M1}_{kt}$ and year and metro area fixed effects generates a coefficient of 2.6, such that even a small value of α_5 will substantially bias $\hat{\beta}$. For example, the true values of α_1 and α_5 could be −1.5 and −0.8, respectively (consistent with a model in which the effects of migration dissipate over time), and this would generate a coefficient $\hat{\beta} = 3.5$ [since $-1.5 - (0.8 \times 2.6) = -3.5$], much larger than the true α_1 of −1.5. As I will show, this is the estimate of $\hat{\beta}$ when I only include past-year arrivals in the migrant share. As made clear in Eq. (2), many other plausible effect sizes α_1 and α_5 would also be consistent with this estimate.

Of course, there are many theoretical reasons to believe that the economic effects of migration may differ according to migrant characteristics, such as time of arrival, return-migrant

4 Return migrants are those who were born in Colombia, migrated to Venezuela in the decades preceding the Venezuelan crisis, and then returned to Colombia during the crisis. They make up around 20% of migrants who arrived from Venezuela between 2014 and 2019.

5 The fixed effects are not necessary for this exercise, but I have included them to match the typical specification in the literature.

status, or demographic characteristics. For example, the economic effects of migration tend to dissipate as capital mobilizes. In order to study this, one could include both $M1_{kt}$ and $M5_{kt}$ together on the right-hand side, and this is a useful approach if there is sufficient independent variation in these variables to precisely estimate these coefficients. However, when an instrument is being used to account for the endogeneity of the migrant share, this approach requires two instruments to generate independent variation in $M1_{kt}$ and $M5_{kt}$. In practice, such instruments are often difficult to find. If the groups are correlated, then in the absence of independent exogenous variation in each group, one must either estimate the average effect of both groups jointly, assume that the excluded group has no effect on labor market outcomes, or accept the bias that results from studying a group in isolation.[6]

It is worth noting that this discussion is closely related to one in the migration economics literature around the use of the "skill-cell" approach, in which data are divided into education-experience cells and labor outcomes are regressed on the cell-specific migrant share and cell fixed effects. This approach identifies the "partial effect" of migration on wages within an education–experience group given fixed supplies in other groups. It therefore does not account for potential effects on workers across cells, which is a necessary component of the total wage effect (Ottaviano and Peri, 2012; Dustmann *et al.,* 2016).

In the remainder of the paper, I conduct additional analysis that extends on the existing literature. I fail to find evidence for nonlinear effects of migration on the logarithm of wages, though there is not enough variation in the data to identify nonlinear effects at very high migration rates. I document a small internal migration response to the Venezuelan migration, and I confirm that native spatial arbitrage does not bias labor market estimates in this context (Borjas, 2003; Borjas and Katz, 2007; Monras, 2020). I study native employment across occupation skill groups ranked according the premigration mean years of schooling in each occupation. I find little average effect on occupational skill level and small movements among some demographic groups. Men with completed secondary schooling experienced minor upgrading from low- to middle-skill occupations, while men with postsecondary education experienced minor downgrading from high- to middle-skill occupations, alongside increases in self-reported underemployment. There is also a small movement of natives *out* of the formal sector in response to the migration. Thus, while some studies have found that migration stimulates native upgrading to higher-skill occupations (Peri and Sparber, 2009; Foged and Peri, 2016; Peri *et al.,* 2020), in this context, this did not occur on a large scale, though there were winners and losers among some subgroups. Finally, I document that wage effects are slightly larger in locations with higher baseline informality rates and lower ease of starting a business as measured in the World Bank Doing Business report, indicated that local economic characteristics are relevant for the economic consequences of migration.

Importantly, the results from this paper are short term, and both theory and existing empirical evidence predict that the effect of migration on native wages should recover and possibly become positive in the long term (Edo, 2020; Verme and Schuettler, 2021). However, this is not necessarily true for the distributional consequences of migration, which often persist.

6 Note that it is not satisfactory to simply use an instrument for one group, $M1_{kt}$, while controlling for the other, $M5_{kt}$. If $M5_{kt}$ is correlated with the instrument and the error term, this is an "endogenous controls" problem and β_b will be inconsistent. See Frölich (2008) for a discussion of endogenous controls in ordinary least squares (OLS) and two-stage least squares (2SLS) models. Nonparametric methods may allow for consistent estimates in the absence of a second instrument for $M5_{kt}$, but this is demanding and requires extensive independent variation in the endogenous variables.

The results motivate policies to support lower-income workers during large migration waves, as well as further research to better understand the mechanisms that drive the aggregate and distributional consequences of migration in the developing-country setting, to enable policymakers to minimize the costs and maximize the benefits of migration.

This paper proceeds as follows. Section 2 reviews the literature, and Section 3 gives the background on Venezuelan migration to Colombia. Sections 4 and 5 review the data and empirical specifications, and Section 6 presents the baseline results. Section 7 tests the sensitivity to various specification choices and uses these results to reconcile findings in the literature. Sections 8 and 9 conduct additional robustness tests and analysis, and Section 10 concludes.

2 Literature

The magnitude of the average wage effect found in this paper is large but not unheard of in the literature, in the context of a large, sudden migrant arrival and short-run outcomes. For example, Edo (2020) finds that the repatriation of Algerians to France led to native wage effects between −1.3% and −2%, though wages recovered after 10 years. Dustmann *et al.* (2017), studying the 1991 inflow of Czech workers into Germany, find that the corresponding elasticity is a smaller 0.13% fall in native wages alongside reductions in employment. Studies of migration from the former Soviet Union to Israel in the 1990s also find small negative wage effects concentrated among less-educated natives, which disappear after 4–7 years (Cohen-Goldner and Paserman, 2011; Friedberg, 2001). For various other episodes of forced displacement in high-income settings, the literature has found little-to-no effects on native employment or wages. This includes Cuban refugees in Miami in the mid-1980s, refugee dispersal in the United States between 1980 and 2000, migration from former Yugoslavia into the European Union (EU), and Puerto Ricans in Orlando after Hurricane Maria (Peri *et al.,* 2020; Peri and Yasenov, 2019; Clemens and Hunt, 2019; Mayda *et al.,* 2017; Card, 1990), though the lack of null results has been disputed in some cases (Borjas and Monras, 2017; Borjas, 2017). In other settings, effects are positive: Foged and Peri (2016) find that refugee dispersal in Denmark in the late 1980s increased native low-skill wages and increased the complexity of native jobs. In a recent meta-analysis of the forced-displacement literature, Verme and Schuettler (2021) find that native wage and employment effects are typically insignificant, and when significant, they tend to be negative. These negative effects tend to be the largest for less-educated and informal workers, in middle-income countries, and when the migrant supply increase is large relative to the native workforce. These effects tend to dissipate after 5 years.

An emerging literature studies the economic effects of forced displacement in low- and middle-income countries. An important example is the Syrian refugee migration to Turkey, where the increase in supply was also mostly in the informal sector. Studies of this migration wave find negative effects on native informal employment, alongside positive effects on formal employment (Del Carpio and Wagner, 2015; Ceritoglu *et al.,* 2017; Tumen, 2016; Altındağ *et al.,* 2020) and increasing native task complexity (Akgündüz and Torun, 2018). This could result from language and cultural barriers creating the potential for communication-intensive occupational upgrading by natives (Peri and Sparber, 2009). While many studies find negligible wage effects in Turkey, Aksu *et al.* (2018) find a negative wage effect in the informal sector, which decreases with education, and Cengiz and Tekgüç (2021) find imprecise negative

wage effects in the informal sector using a synthetic control method. As discussed, one reason that wage effects may be larger in Colombia is that Venezuelan migrants speak the same language, increasing substitutability with natives. Another is that employment effects are smaller in Colombia, leaving effects concentrated along the earnings margin.[7] In contrast, Fallah *et al.* (2019) find no economic effects of Syrian refugees in Jordan, potentially explained by low refugee labor force participation alongside increases in EU aid and trade concessions.

Studies of the economic effect of Venezuelan migration in other countries in Latin America have also found evidence for negative native wage effects mostly concentrated on less-educated and informal workers in Ecuador (Olivieri *et al.,* 2021), Brazil (Zago, 2020), and Peru (Morales and Pierola, 2020).[8] In Brazil, migration is found to affect participation rather than wages when using a synthetic control method (Ryu and Paudel, 2022). Papers have also highlighted migrant occupational downgrading as a mechanism behind the unequal wage effect in Colombia: Lebow (2022) estimates a production function that allows for imperfect substitutability between migrants and natives and finds that migrant occupational downgrading led to increases in the wage effect for less-educated natives, alongside decreases in total productivity. Lombardo *et al.* (2021) also show that the magnitude of the negative wage effect is increasing in the density of the migrant share along the native wage distribution. Finally, there are studies of the labor market consequences of internal displacement during Colombia's civil war between the 1980s and the early 2000s. These papers have found negative wage effects for urban workers ranging from −0.09% to −1.4%, which are consistently larger for low-skill and informal workers (Calderón-Mejía and Ibáñez, 2016; Morales, 2018). My results are consistent with these estimates.

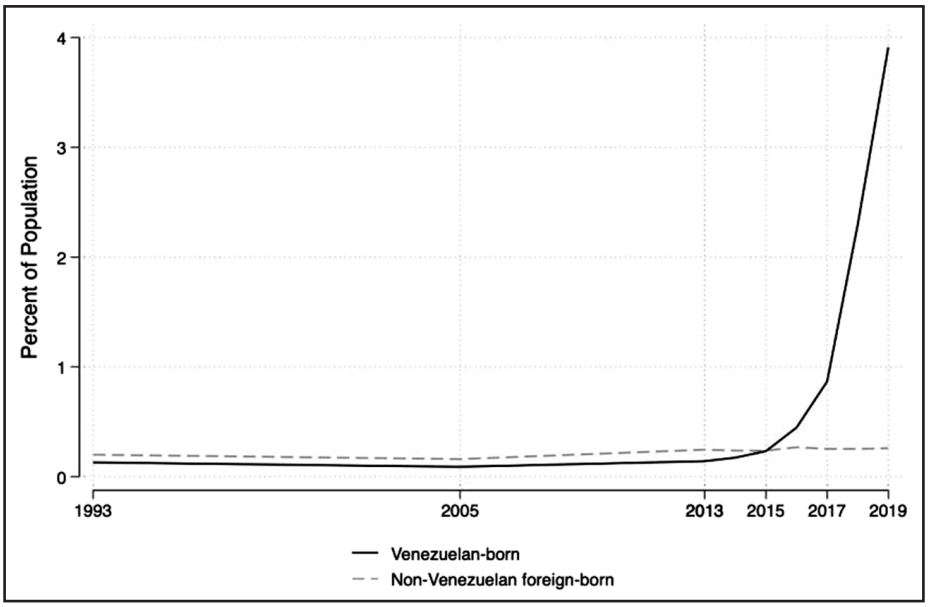
3 Background on Venezuelan Migration to Colombia

Between 2015 and 2019, around 4.5 million people fled Venezuela, making Venezuelans the second-largest internationally displaced population after Syrians (UNHCR, 2019). The primary reasons for this migration were to escape poverty and violence induced by the political and economic crisis. Following the sudden collapse of global oil prices in 2014, Venezuela entered an economic recession that led to hyperinflation by 2016. By 2018, the gross domestic product (GDP) had contracted by 45% since 2013, and around 90% of the population was estimated to be living in poverty. More than 20% of the population was undernourished, access to water and electricity became increasingly scarce, and an estimated 85% of essential medicines were scarce. The murder rate also rose to one of the highest in the world (Wilson Center, 2019; Reina *et al.,* 2018; World Bank, 2018). The primary reasons for migration that Venezuelans cite include shortages of food and medicine, violence and insecurity, lack of access to social services, and fear of political persecution (UNHCR, 2018).

7 Various papers in Africa have evaluated, in the low-income country setting, the impact of forced displacement on nearby communities. This setting can be characterized by refugees hosted in camps and large inflows of foreign aid, which differs heavily from the Colombian context as I discuss in Section 3. These studies have generally found positive effects on local employment and household consumption (Alix-Garcia and Saah, 2010; Maystadt and Verwimp, 2014; Ruiz and Vargas-Silva, 2015; Alix-Garcia *et al.*, 2018).

8 More recent estimates from Peru show less clear evidence of negative wage effects (Boruchowicz *et al.*, 2021). A possible explanation is migrant selection - Venezuelans who go to Peru are even more educated on average than those who go to Colombia.

Figure 1 Foreign-born population in Colombia.



Sources: Colombian National Integrated Household Survey (GEIH) (2013–2019); Population Census (1993, 2005).

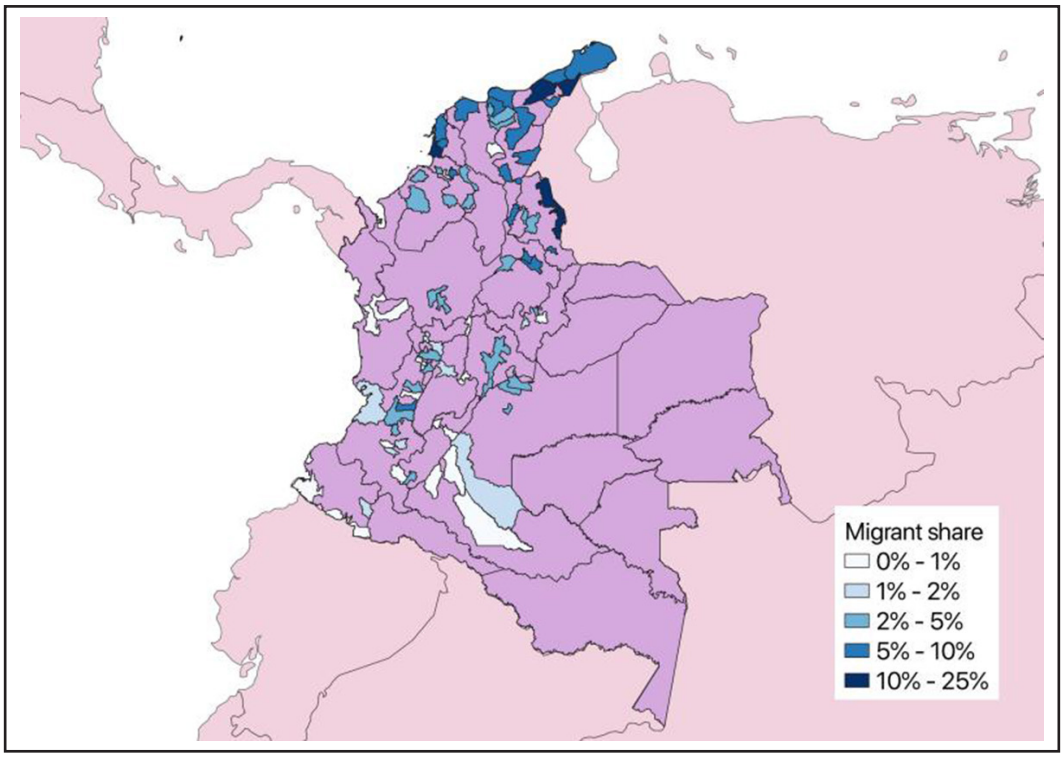
Colombia, the neighbor closest to the population centers of Venezuela, received an estimated 1.8 million of these migrants, more than any other country. Figure 1 shows that this is the first time that Colombia has received a large migration wave from another country: in the 1993 census, 0.13% of the population was Venezuelan born and 0.2% was born in a different foreign country, and these rates remained relatively constant until the onset of the Venezuelan migration in 2015.[9] The arrival rate increased in 2016 and again in 2017, with the majority of migrants arriving between 2018 and 2019. The results presented in this paper therefore reflect very short-run economic effects.

Figure 2 shows the migrant share of the population across 79 metro areas in 2019, where a migrant is defined as someone who was living in Venezuela 5 years ago. There is extensive variation in these migrant shares across Colombia. They tend to be largest closer to the Venezuelan border, in many cases exceeding 10% of the metro area population. In Cúcuta and Riohacha, two cities close to the primary entry points along the Venezuelan border, the migrant shares are around 16% and 11%, respectively. In Bogotá, Medellín, and Cali, the three largest cities in Colombia, the shares range between 4% and 5%. For other cities, they remain below 1%.

The majority of migrants crossed at a handful of official border crossings, which required a passport or a visa that allowed for short-term access to the border regions. However, those without legal documents could pass around border checkpoints on paths commonly known as "trochas". The Colombian government created a temporary resident visa beginning in January 2017 (the *Permiso Temporal de Permanencia* [PEP]), which allowed documented migrants to access the formal labor force and additional education and health services. This status was offered to a large number of undocumented migrants starting in April 2018 through a process called the RAMV. This was intended to regularize the growing number of migrants who

9 The lack of migration into Colombia pre-2015 reflects the fact that Venezuela was historically a recipient, rather than a source, of immigrants. Favorable economic conditions and generous social programs attracted migrants from across Latin America, including Colombians fleeing the decades-long civil war in Colombia (Freitez, 2011).

Figure 2 Venezuelan migrants in Colombia (2019).



Sources: Colombian National Integrated Household Survey (GEIH) (2019).

had either entered illegally or overstayed their temporary permit (the *Registro Administrativo de Migrantes Venezolanos* [RAMV]). By the end of 2019, according to official numbers from the Colombian migration authorities, an estimated 754,000 Venezuelan migrants were regularized, corresponding to approximately 42% of the Venezuelan migrants estimated to be in the country (*Migración Colombia*, 2019). Furthermore, Bahar *et al.* (2021) show that the 2018 RAMV regularization had little-to-no effect on native labor outcomes, likely because registered migrants mostly remained in the informal sector. In summary, there was little control over who entered the country and where migrants went within Colombia, and despite many having access to regularization, the majority of migrants remained undocumented and informal. Both documented and undocumented migrants are included in my data and I do not observe documentation status.

Another important characteristic of this migration wave is that, unlike with many other global episodes of forced displacement, relatively little international aid has been dedicated to the reception of Venezuelan migrants. By 2019, international funding for the Venezuelan migration crisis was at $580 million, compared to $7.4 billion that went to displaced Syrians in the first 4 years of the Syrian crisis (Bahar and Dooley, 2019). Furthermore, few migrants are living in camps, and the Colombian government has been unable to mobilize large-scale aid or investment into areas most affected by migration (Migration Policy Institute, 2020).

Migration can also affect local economies by increasing local consumer demand, especially if migrants carry savings (Verme and Schuettler, 2021; Cortes, 2008). However, it is estimated that Venezuelan migrants' household expenditures in Colombia are less than half that of natives, reflecting migrants' low income in Colombia and the fact that savings for many Venezuelans was wiped away by inflation (Tribín-Uribe *et al.,* 2020). Delgado-Prieto (2021)

studies the effect of migration on the consumer price index across 23 capital cities in Colombia and finds precise null effects, though there is evidence for small increases in the cost of education and small decreases in the cost of health care. Overall, the demand effects of migration in this context appear to be modest, and I do not study the effects on prices in this paper. The literature would benefit from a more detailed analysis of the effect of Venezuelan migration on the local prices of different goods and services. To the extent that increased demand directly affects wages, this will be captured in the estimated wage effects.

4 Data and Descriptive Statistics

Labor market outcomes come from the GEIH, which is a large-sample cross-sectional survey collected by the National Administrative Department of Statistics (*Departamento Administrativo Nacional de Estadística* [DANE]) and is the official source for labor market indicators in Colombia. Since 2013, this survey has included a migration module that can be used to identify migrants, and I therefore also use this survey to measure migration shares across regions and over time. This module also allows me to restrict labor market outcomes to nonmigrants, to avoid any compositional effects driven by arriving migrants.

It is important to note that the GEIH is not intended to be representative of Venezuelan migrants in Colombia. Another potential source of the migrant share across regions is the 2018 census, but this comes with two caveats.[10] First, security concerns and other unexpected logistical constraints led this census to undercount the Colombian population by 8.5%, generating skepticism over the ability of the census to accurately measure the migrant population (*El Espectador*, 2019). Second, this only measures a snapshot of the migrant population from January to September 2018, missing the dynamics of migrant arrival before 2018 and the large inflow of migrants in the last quarter of 2018 and 2019. For these reasons, the GEIH is increasingly being used to track the Venezuelan population across Colombia over time (Graham *et al.,* 2020; Tribín-Uribe *et al.,* 2020). In practice, migration rates under the census and the GEIH are closely correlated ($\rho = 0.77$ across metro areas in my sample in 2018). In Section 7, I test sensitivity to using the 2018 census to measure the migrant share, and I show that results are comparable but slightly larger in magnitude when using the census to estimate the effects of migration through 2018.[11]

I define migrants as anyone who was living in Venezuela 5 years ago. Importantly, this includes Colombian-born return migrants, who make up around 20% of all migrants from Venezuela during this period. Many of these returnees migrated to Venezuela during Colombia's decades-long civil war. They are on average older and less educated than Venezuelan-born migrants, and there are various reasons to believe that their effect on the labor market may differ.[12] However, many have been in Venezuela for decades (the first large migration to Venezuela was in the 1970s) and may be more comparable with Venezuelan-born migrants than with

10 Another possible source is the official estimate from the Colombia Migration Unit, which is imputed based on a combination of border flows, registration of undocumented Venezuelans, and the 2018 census. However, these numbers likely undercount undocumented migrants, especially before 2018 and in locations farther from the border (Tribín-Uribe *et al*., 2020; Graham *et al*., 2020).

11 The complete 2005 and 2018 censuses were provided by DANE. I will also use the 1993 census 10% subsample, which was accessed via Integrated Public Use Microdata Series (IPUMS) (2019).

12 In the working-age metro sample of the 2019 GEIH, the average age of return migrants and Venezuelan-born migrants is 38.7 years and 30.0 years, respectively. The completed years of schooling is 8.2 years and 10.4 years, respectively.

nonmigrants in terms of characteristics, such as networks and work experience. In practice, their location in Colombia is closely correlated with that of Venezuelan-born migrants ($\rho = 0.71$ across metro areas in 2019). Therefore, as I discussed in Section 1, excluding them from the migrant share could introduce bias. Indeed, I show in Section 7 that the magnitude of the wage effect increases substantially when return migrants are excluded. While, ideally, one could study the effect of each group in isolation, these effects are empirically difficult to untangle due to lack of independent variation and in the absence of a separate instrument for each group.

Colombia is divided into a capital district and 32 departments, which are further divided into 1,122 municipalities. In order to generate a geographic unit that represents a contiguous labor market in which workers compete, I group municipalities into metropolitan areas according to commuting patterns. Following Duranton (2015), I use a recursive algorithm based on a 10% commuting threshold using commuting data from the 2005 census, which is the latest year before the start of the migration period in which such data are available.[13] The algorithm results in 184 metro areas with at least 30,000 residents in 2005. These can be compared with the 23 metro areas officially defined by the GEIH: in some cases, they are identical (for example, in the case of Medellín), and in other cases, they are distinct (for example, the constructed areas of Bogotá, Calí, and Baranquilla all include substantially more municipalities than the administrative metro areas). Given that the administrative metro areas are politically determined considering factors such as allocation of city resources and jurisdiction of city government activity, it is preferable to use metro areas constructed according to economic criteria (Duranton, 2015).

A concern about measurement error arises when the GEIH is used at such a fine geographic level, since it is only designed to be representative at the level of the department or the 23 official metro areas. I address this in various ways. First, I restrict analysis to the 79 metro areas that contain at least 300 observations per year. These represent around 80% of the Colombian population and 90% of the Venezuelan migrant population. Second, I test the robustness of results to increasing the annual observation threshold to 1,000, essentially restricting analysis to the 23 metro areas for which the survey is officially representative.

Third, in my analysis, I use an instrument based on migration shares from the complete 2005 National Census, which will mitigate the effect of measurement error in the endogenous variable. Finally, I test robustness to using alternative geographic units (including the department level) in Section 7, and I find that choice of geographic unit only has a moderate effect on the results.

I now describe the characteristics of working-age (15–64) migrants and natives in the 79 metro areas of analysis in the 2019 GEIH using sampling weights, which has a national sample of 447,264 natives and 21,730 migrants. Table A1 shows that migrants are gender balanced and 5 years younger than natives on average. Figure A1 shows that migrants come most heavily from the middle of the education distribution: 44% have completed secondary education relative to 35% of natives. However, >20% of migrants have some postsecondary education, indicating that this is not a low-skill migration wave, and the educational profile of migrants and natives is broadly similar. However, by plotting the distribution of migrants and natives across occupations ranked according to mean years of completed schooling for natives, Figure A2 shows that

13 Specifically, a municipality is grouped with another if >10% of its residents commute to work in that municipality. They are then treated as a single unit in the next round of the algorithm, and this is repeated until no more municipalities meet this threshold. In practice, the number of metro areas generated by this method does not depend on the choice of commuting threshold. See Duranton (2015) for details.

the majority of migrants are concentrated in occupations that tend to employ less-educated natives.[14] In particular, they are severely overrepresented in street vending, restaurant work, construction, domestic service, and beautician work. Thus, despite migrants and natives being similarly educated, we expect economic consequences to be larger for less-educated natives working in heavily affected occupations. In terms of economic outcomes, migrants have higher labor force participation than natives (79.4% relative to 71.7%) and higher unemployment rates (14.8% relative to 11.4%). They work more hours and have a median hourly wage that is around 70% of that of natives. They are also more likely to be own-account workers and almost 90% are informal.

5 Empirical Specification

I estimate the following regression across metro areas (c) and years (t) from 2014 to 2019:

$$Y_{ct} = \beta M_{ct} + \gamma_c + \delta_t + \epsilon_{ct} \tag{3}$$

where M_{ct} is the migrant population as a share of the 2014 population. Metro fixed effects, γ_c, hold constant all time-invariant metro characteristics; year fixed effects, δ_t, adjust for national-level changes in labor outcomes. Standard errors are clustered at the metro level to allow for the error to be correlated within metro areas over time, and observations are weighted by the population within each metro–year cell.[15]

Labor market outcomes, Y_{ct}, are averaged at the metro–year level and include labor force participation, unemployment, hours worked in a typical week, and log hourly wage, including all overtime, benefits, and other transfers. This information is collected for all workers regardless of self-employment or formality status, and going forward, I use hourly wage to refer to hourly earnings for wage workers or hourly profits for self-employed. These outcomes are residualized from a regression on gender, age, and education fixed effects, though this adjustment has little effect on the results.[16] Residual log wages are windsorized at the top and bottom 0.5% of observations to reduce the impact of extreme outliers.

There has been recent concern regarding bias in two-way fixed-effect models when there are dynamic treatment effects (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Borusyak and Jaravel, 2021; De Chaisemartin and d'Haultfoeuille, 2020). Intuitively, in two-way fixed-effect models, groups that are "treated" later in the study period use previously treated groups as a control, even though they may still be experiencing a dynamic treatment effect. While these papers generally deal with models that have a binary treatment, the same concern in principle extends to a continuous treatment framework.[17] In the case of this paper, lagged economic effects of migration in preceding years can bias the estimate of the treatment

14 Occupations are recorded in the GEIH using the 82 classifications of the 1968 International Standard Classification of Occupations (ISCO-68). They are ranked by the average years of schooling for natives between 2010 and 2015.

15 I choose not to use the GEIH sampling weights since they are not designed to be representative at this level, but my results are not sensitive to this decision.

16 Age is grouped into 5-year intervals, and education is grouped into less than primary, less than secondary, completed secondary, and postsecondary.

17 One paper, by De Chaisemartin and d'Haultfoeuille (2020), develops an estimator robust to heterogeneous treatment effects, which works with a nonbinary treatment. However, it only extends to discrete treatments (as opposed to continuous) and is not identified when treatment takes too many values. When I apply the estimator developed in this paper with migrant shares grouped into intervals, the resulting confidence intervals are not informative.

effect in later years. However, I do not expect this to be a large problem because the majority of migrants arrive in 2018 or 2019, at the end of the study period. To confirm that this is not biasing $\hat{\beta}$, I show a robustness specification in which I regress the 2014–2019 change in outcome on the change in the migrant share and I find very similar results. This implies that most of the estimation is driven by differences in total arrivals across metro areas and not by different timings of the arrival across metro areas.

One cannot simply compare labor market outcomes in metro areas that experienced different migration levels because migrants select destinations according to endogenous characteristics. For example, if migrants are more likely to settle in cities with greater expected growth in available jobs or wages, this would induce a positive correlation between ϵ_{ct} and M_{ct} and an upward bias in the estimate of β. I deal with this endogeneity by constructing the following instrument:

$$Z_{ct} = M_{c,\,2005} \times M_{Nat,t}^{-c} \tag{4}$$

where $M_{c,2005}$ is the 2005 Venezuelan share of the population in metro area *c* according to the complete 2005 population census, and $M_{Nat,t}^{-c}$ is the "leave-one-out" national share of migrants from Venezuela to all metro areas in year *t*, excluding migration into area *c* (Card, 2001; Tabellini, 2020). This "leave-out" factor reduces the correlation between national-level inflows and large inflows into certain cities, which may be correlated with time-changing characteristics of those cities.

This variable is strongly predictive of subsequent migrant shares, creating the strong first stage needed for consistent two-stage least squares (2SLS) estimates. Figure A3 displays the positive linear relationship between the 2005 Venezuelan share (which ranges from 0% to 0.9%) and the 2019 migrant share (which ranges from 0% to 25%). There is strong persistence over time in the distribution of Venezuelan migrants across metro areas, likely driven by the tendency of migrants to locate where they have migrant networks (Beaman and Magruder, 2012; McKenzie and Rapoport, 2007; Munshi, 2003). When the instrument in Eq. 4 is used with metro and year fixed effects, the first-stage Kleibergen-Paap (K-P) Wald statistic is 23.[18]

The variable $M_{Nat,t}^{-c}$ is assumed to be exogenous, driven primarily by push factors in Venezuela, and uncorrelated with any changes occurring in Colombia. Unemployment in Colombia steadily increased from 8.5% to 10% from 2014 to 2019, indicating that there were no favorable labor market conditions attracting migrants over this period. Furthermore, migration did not increase from any other country. Of greater concern is the potential endogeneity of $M_{c,2005}$, which may be correlated with *changes* in economic outcomes between 2014 and 2019 in metro area *c* through channels other than the mass migration. This concern is mitigated by the fact that historical migrant shares were determined 2 decades before the onset of the Venezuelan exodus, well before the election of Hugo Chávez. While I use the 2005 census to construct the instrument, I show that results using the 1993 census are similar. The correlation between the migrant shares in these years is 0.89.[19] The instrument would also be invalid if migration before 2005 stimulated dynamic economic responses or if subsequent migration

18 The Kleibergen-Paap Lagrange multiplier (LM) test tests the null hypothesis that the structural equation is underidentified. With a single endogenous regressor, this reduces to a standard first-stage *F*-statistic that is heteroskedasticity robust.

19 I prefer to use the 2005 census to construct the instrument because I have access to the complete census for this year, which helps to minimize measurement error.

correlated with current economic trends (Jaeger *et al.*, 2018). However, there was almost no migration into Colombia between 2005 and 2015, and the 2005 migrant share was miniscule relative to the current migration – in no metropolitan area in 2005 was the Venezuelan-born share >1%.

It remains possible that, between 2014 and 2019, migrants historically selected and moved into cities that had differential economic trends. While this cannot be formally tested, I conduct various checks and robustness tests recommended by the literature (Goldsmith-Pinkham *et al.*, 2020). First, I check for a correlation between the 2005 shares and the preperiod economic outcomes. Table A2 shows a regression of the 2005 Venezuelan share on a set of metro area characteristics measured in 2014, before the onset of the migration. The results show that a 1% increase in wages is associated with a 0.02 standard deviation (SD) decrease in the 2005 migrant share. Hours, labor force participation, unemployment, and total population are all insignificant and small in magnitude. Importantly, the value of R^2 is small, indicating that only 9% of the variation in the 2005 migrant share can be explained by all of these variables together. The largest concern is that places with more immigration in 2005 had slightly lower wages in 2014.

These fixed metro characteristics are controlled for in the analysis. It is arguably more relevant to test whether historical migrant shares are correlated with pre-trends in economic outcomes leading up to the migration. I run the following event-study model:

$$Y_{ct} = \sum_{y=2010}^{2014} [\sigma_y M_{c,2005} * (t = y)] + \sum_{y=2016}^{2019} [\sigma_y M_{c,2005} * (t = y)] + \gamma_c + \delta_t + \epsilon_{ct} \tag{5}$$

where σ_y measures the differential change in the outcome for cities that had a relatively higher 2005 migration share (in SDs) relative to the excluded year of 2015, the year preceding the onset of the migration. The 95% confidence interval (CI) around the coefficient for each primary outcome is presented in Figure A4. Hourly wages show a decreasing trend from 2011 to 2013, but this levels off and is flat in the 2 years before the migration. After 2015, it drops precipitously, in line with the 2SLS results. Hours per week is stable in the preperiod and increases after 2015, also consistent with the 2SLS results. However, pre-trends are observed for unemployment rates (which increase steadily during the preperiod and then flatten after the migration) and participation rates (which decrease between 2013 and 2014 and again after the migration). This motivates an adjustment for pre-trends when using this instrument to study unemployment and participation, which I include as a robustness check.

6 Results

6.1 Primary results

The ordinary least squares (OLS) and 2SLS results for the population average are presented in Table 1. According to Column 1, a 1 pp increase in the migrant share is associated with a 0.73% decrease in residual hourly wages. After endogenous selection is accounted for in the 2SLS model, this magnitude increases to a 1.05% decrease, with a 95% CI that includes [−1.48, −0.62].[20] At the −1.05 point estimate, a movement from the 10th percentile to the 90th percentile

20 To place this in the literature, this elasticity is slightly smaller that the values of −1.3 to −2 found by Edo (2020) when looking at the repatriation of Algerians to France; but, this is substantially larger than the value of −0.13 found by Dustmann *et al.* (2017) regarding the cross-border flow of Czech workers into the German workforce.

Table 1 Labor market effects of immigration

	(1)	(2)	(3)
	OLS	**2SLS**	**Test (1) = (2) (*p*-value)**
ln(hourly wage)	−0.73*	−1.05***	0.209
	(0.42)	(0.22)	
ln(hours/week)	0.08	0.27	0.399
	(0.26)	(0.27)	
Unemployment	0.01	−0.08	0.255
	(0.13)	(0.07)	
Labor force participation	−0.10	−0.21*	0.355
	(0.12)	(0.11)	
K-P Wald stat		23.35	
N	474	474	
Year FE, City FE	X	X	

Notes: Outcomes are residualized and multiplied by 100. Observations are weighted by city–year population. Column 3 presents a Hausman test for endogenous regressors with robust standard errors. Cluster-robust standard errors are in parentheses.2SLS, two-stage least squares; FE, fixed effect; K-P Wald stat., Kleibergen-Paap Wald statistic; OLS, ordinary least squares.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

of the migrant share across metro areas in 2019 (from 1% to 8.6%) is associated with a 7.98% decrease in wages. This effect is economically meaningful, considering that average wages for natives in the sample increase by around 4% over the period 2014–2019. That the 2SLS is more negative than the OLS indicates that there was positive selection of migrants into locations that had bigger increases in wages, though this difference is not significant according to a Hausman test for endogenous regressors.

There is little effect on hours or unemployment, and a small negative effect is found on participation, which is significant at the 10% level. Specifically, a 1 pp increase in the migrant share causes a 0.21 pp decrease in labor force participation, with a 95% CI that includes [−0.43, 0.01]. This is associated with a 1.6 pp decrease in participation from the 10th percentile to the 90th percentile of the migrant share, relative to a 2014 sample mean of 72.6%. Thus, the majority of the effect on natives was seen on the earnings margin rather than on the employment or participation margin.

In Table 2, I split the outcomes by gender, age, and education.[21] The results show that the effect on wages is slightly stronger for men, though the difference is not significantly significant. Though there are various examples of female participation being more responsive to changes in labor supply (Verme and Schuettler, 2021; Dustmann *et al.*, 2017), effects on participation are only slightly more negative for women and not significantly different. A priori, one could also expect labor market effects to be more severe for younger natives, since migrants are younger on average. However, as shown by Lebow (2022), migrants work in occupations that tend to employ older natives. Thus, it is not surprising that negative wage effects are slightly stronger for older natives, although, again, the differences are insignificant.

21 This is done by averaging the outcome within metro–year–group cells. The first stage *F*-statistic changes as the population-based cell weights change. Observations remain clustered at the metro level because treatment is assigned at this unit.

Table 2 2SLS estimates by demographic groups

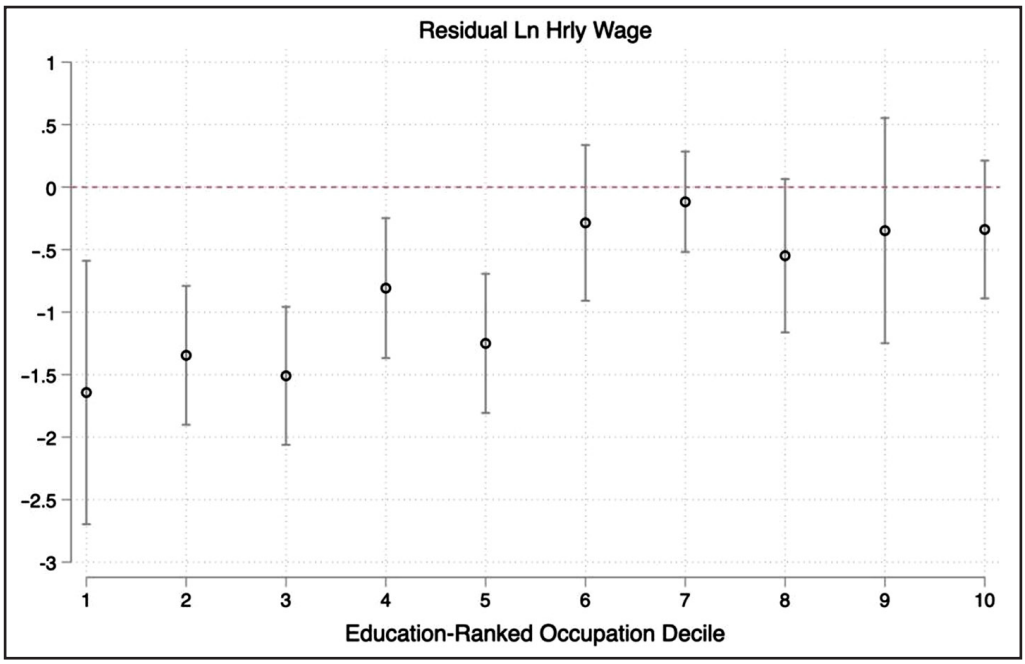
	(1)	(2)	(3)	(4)	(5)
	ln(hourly wage)	**ln(hours/week)**	**Unemployment**	**LFP**	**K-P Wald stat.**
All	−1.05***	0.27	−0.08	−0.21*	23.35
	(0.22)	(0.27)	(0.07)	(0.11)	
Male	−1.16***	0.37**	−0.05	−0.18**	24.34
	(0.22)	(0.17)	(0.08)	(0.08)	
Female	−0.91***	0.19	−0.12	−0.22*	22.22
	(0.25)	(0.39)	(0.08)	(0.13)	
Age 15–24 years	−0.94***	0.14	−0.01	−0.54***	21.34
	(0.20)	(0.46)	(0.12)	(0.18)	
Age 25–34 years	−0.97***	0.31	−0.13	−0.08	22.11
	(0.22)	(0.21)	(0.10)	(0.07)	
Age 35–44 years	−1.10***	0.29	−0.06	−0.07	23.39
	(0.23)	(0.20)	(0.07)	(0.06)	
Age 45–54 years	−1.01***	0.34	−0.10**	−0.11	26.11
	(0.26)	(0.25)	(0.05)	(0.10)	
Age 55–64 years	−1.21***	0.24	−0.02	−0.10	23.71
	(0.30)	(0.38)	(0.05)	(0.11)	
Less than secondary	−1.42***	0.50	−0.05	−0.23*	28.72
	(0.23)	(0.34)	(0.06)	(0.12)	
Secondary	−0.86***	0.32	−0.02	−0.12**	25.64
	(0.22)	(0.21)	(0.09)	(0.06)	
Postsecondary	−0.75*	−0.02	−0.18*	−0.24	17.22
	(0.38)	(0.23)	(0.10)	(0.16)	

Notes: Outcomes are residualized and multiplied by 100. All models include Year FE and City FE. Observations are weighted by city–year–group population. Cluster-robust standard errors are in parentheses.FE, fixed effect; K-P Wald stat., Kleibergen-Paap Wald statistic; LFP, labor force participation.
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

The results also show that the decrease in labor force participation is concentrated among workers younger than 25 years of age, with a magnitude of −0.54 pp from a 1 pp increase in the migrant share, likely reflecting lower labor force attachment for these workers. In Table A11, I show that this is partially driven by reduced school dropouts. Among workers younger than 25 years of age, there is a 0.39 pp increase in the share of migrants who are not working and are attending school, and a net 0.2 pp increase in school attendance relative to a mean of 54.6. However, these are only significant at the 10% level, and the results are not robust to dropping the metro areas located closest to the Venezuelan border, which are in the right tail of the migrant share distribution. I further discuss the importance of this robustness check in Section 8.

I expect the wage effect to be larger for less-educated workers given the concentration of migrants in low-skill occupations, and the results confirm this. The hourly wage effect, respectively, for those with less-than-secondary, secondary, and postsecondary education is −1.42%, −0.86%, and −0.75%. This pattern of negative wage effects decreasing in the education of natives is also observed when the sample is split by occupation groups, defined by ranking occupations

Figure 3 The effect of migration on residual ln(hourly wage) is separately estimated via 2SLS within occupation skill groups.



Note: 95% confidence intervals are presented around each point estimate.

according to mean preperiod native years of schooling and split into deciles. This is shown in Figure 3, where the coefficient hovers around −1.5 in the lowest three deciles, around −1 in Deciles 4 and 5, and between −0.5 and 0.0 for the top five deciles.[22]

I also split outcomes for workers by the "type" of work, grouped into four mutually exclusive categories: formal salaried, informal salaried, own account, and employer (representing 29%, 20%, 47%, and 4% of native workers, respectively). Self-employed workers may experience different effects of competition with migrants considering that they compete directly over prices rather than wages. Self-employed workers who hire employees, on the other hand, may experience increases in profits if the price of labor decreases. I split salaried workers according to formality status because we expect informal workers to face more direct competition with migrants and to have more flexible wages.[23] Figure A5 shows that the wage effects are driven by informal salaried and own-account workers. Formal salaried workers also experience a wage decrease, but the coefficient is 50% smaller and not statistically different from zero. As I show in Section 9.2, this may be partly explained by the small exit out of formal work, which could mitigate the estimated wage effects for formal workers. The sample of employers is too small to generate precise estimates about the effects of migration on their profits.

7 Reconciling Results in the Literature

Various papers (Caruso *et al.,* 2021; Delgado-Prieto, 2021; Bonilla-Mejía *et al.,* 2020; Santamaria, 2020; Penaloza-Pacheco, 2021) have also studied the effects of Venezuelan migration to Colombia on native employment and earnings. While they are generally consistent in finding

22 Estimates within occupation groups may be biased if workers switch occupations in response to the migration. In Section 9.2, I show that there are very small effects on movements across these groups.

23 More than 90% of own-account workers are informal, so I do not split that group into formal and informal.

null or small employment effects and negative wage effects for natives, the magnitude of the estimated wage effect varies drastically, with elasticities ranging from −0.5% to −7.5% from a 1 pp increase in the migrant share. This variation could be driven by various specification choices, including the geographic unit of analysis, the instrument, and the formula and data used to measure the migrant share of the population. In this section, I explore the sensitivity of the magnitude of the wage effect to these factors, to explain the dispersion in the literature and to shed light on the specification choices that may be most consequential when estimating the economic effects of migration.

In Table A3, I start by cross-interacting three dimensions of sensitivity: the chosen instrument, the geographic unit of analysis, and whether or not return migrants (born in Colombia and living in Venezuela before the migration) are included in the migrant share. The first dimension that I vary is the choice of instrument. Table A3 shows that the OLS model consistently generates smaller wage effects, consistent with migrants positively sorting into high-wage locations. This already helps to explain some variation in the literature: the papers that report smaller coefficients of −0.5, viz., Penaloza-Pacheco (2021) and Santamaria (2020), use OLS and thus do not account for migrant positive sorting.[24,25]

The papers that use 2SLS differ in their choice of the "share" component of the shift-share IV, using either historical migrant shares based on the 1993 or 2005 census or the inverse driving distance to the Venezuelan border.[26] The wage effects based on the 1993 and 2005 census IVs are similar in magnitude, driven by the high correlation in the migrant shares across these years. The IV based on distance to the border also produces similar results, though the magnitudes tend to be slightly larger. That these IVs generate similar results is not surprising given that border proximity is correlated with historical migrant shares. However, one may be concerned that border proximity is less likely to satisfy the exclusion restriction than historical migrant shares. These historical shares were small, determined decades earlier, and induced little immigration before 2015, while border proximity has the potential to be directly affected by economic changes related to the crisis. In Section 8, I treat border proximity as an omitted variable, correlated with historical migrant shares and potentially with trends in the outcome, rather than as a source of exogenous variation. I will show that controlling for a time trend interacted with border proximity somewhat reduces the magnitude of the wage effect.

Second, these papers differ in their unit of analysis, either at the level of the 24 departments or 23 administrative metro areas. The primary advantage of conducting analysis at these

24 The goal of this analysis is not to fully replicate the coefficients from each paper. There remain differences in how variables are calculated and analysis is conducted across papers. For example, Santamaria (2020) uses Google search keywords to measure migrant locations across departments. The goal is instead to identify the factors that explain the variation in results.

25 In an alternate approach, Penaloza-Pacheco (2021) estimates the average effect for the border departments La Guajira and Norte de Santander using a group of hand-selected departments as a control group or a synthetic control method. These methods estimate total wage effects of –13% and –9.4%, respectively, or –1.03% and –0.75% from a 1 pp increase in the migrant share. These are more consistent with my own estimates, though they should be interpreted as the average effect specifically for these two departments.

26 Distance to the border can be replaced with a summed distance to each Venezuelan department weighted by the share of Colombian expatriates living in each department (Caruso *et al.*, 2021; Delgado-Prieto, 2021). Here, I simply use driving distance to the border, which is closely correlated with this measure. Driving distance is calculated using Open Street Maps, from the central municipality of the metro area to the closest Venezuelan border crossing. I continue to use the "leave-out" instrument, which does not affect results.

levels, relative to the 79 commuting-zone (CZ)-based metro areas that I construct, is the mitigation of measurement error concerns since the GEIH is representative at these levels.

The disadvantage is that departments are large while most migrants are located in cities, and the official metropolitan areas are somewhat arbitrarily defined according to political considerations (Duranton, 2015). Table A3 shows that the wage effect tends to be smaller at the administrative metro area level and biggest at the department level, but results are generally similar across these specifications.[27] The geographic level of analysis is, therefore, not a factor that drives large differences across papers, though when the department level is combined with the distance IV, the coefficient becomes notably larger (−1.27 relative to −1.05 using the 2005 census IV at the CZ metro level, which I refer to, going forward, as the "baseline specification").[28]

Third, not all of these papers choose to include Colombian return migrants when calculating migration shares. Table A3 thus shows that the coefficients are inflated when return migrants are excluded. The coefficient increases from −1.05 to −1.24 in the baseline specification and further to −1.48 using the distance IV at the department level. As discussed, Venezuelan-born and return migrant shares are correlated. Therefore, to exclude return migrants from the migrant share is to assume that they have no effect on local labor markets and to attribute any changes in the labor market outcomes to the Venezuelan-born migrant population. While it is possible that the increased effect size of −1.24 is driven by a larger effect of Venezuelan-born migration on wages, it is equally likely that it is driven by omitted-variable bias, which will have a drastic effect on the estimated magnitude. To see this, consider the framework outlined in Section 1. A regression of the foreign-born migrant share (predicted by the instrument) on the return migrant share with metro and year fixed effects generates a coefficient of 0.2. Thus, if the effect of Venezuelan-born migration and return migration are both −1.05, one would expect a coefficient of −1.26 [since $-1.05 - (1.05 \times 0.2) = -1.26$], very close to the observed result. It would also be possible for the effect of Venezuelan-born migration to be −1.1 and of return migration to be −0.7 [since $-1.1-(0.7 \times 0.2) = -1.24$], or for the effect of Venezuelan-born migration to be −0.9 and of return migration to be −1.7 [since $-0.9-(1.7 \times 0.2) = -1.24$] Thus, a range of possible effects are consistent with the observed change in magnitude, including scenarios in which Venezuelan-born migration has a larger or smaller effect than return migration.

Delgado-Prieto (2021), using a specification in which the change in outcomes relative to the baseline period is regressed on the 2018 migrant share, finds a significant wage elasticity of −1.7. An important difference in this paper is the use of the 2018 census to calculate migration rates. As discussed in Section 4, there are two caveats with the census: it undercounts the Colombian population by 8.5%, and it only measures migration in the first three quarters of 2018. Given that the GEIH is also not designed to be representative of the migrant population, it is not obvious a priori whether one is more desirable than the other. However, the fact that the GEIH can measure the migration rates in other years, rather than only in 2018, is a large

27 The instrument is also adjusted to calculate both the historical shares and the national shift at each geographic level. Distance to the border for departments is calculated from the department capital.

28 The remaining paper that estimates a smaller wage elasticity of around −0.5 is by Bonilla-Mejía *et al.* (2020), using the 2005 census IV at the administrative metro area level. This result can be explained by the fact that they use total wages, not hourly wages, as the outcome. While this generates an effect size of −0.78 in that baseline specification (as can be seen in Table 1), it falls to −0.5 using the administrative metro areas. As we have seen, part of the effect on native total income is mitigated by an increase in hours worked.

advantage. To test sensitivity to the choice of data set, I run this specification using both the GEIH and the census to measure the migrant share in 2018.[29]Particularly, the specification is

$$\Delta Y_{c,2014-2018} = \beta M_{c,2018} + \epsilon_c \tag{6}$$

First, the analysis shows that, using a 2014–2018 difference model and using the GEIH to measure the migrant share, the results are very similar to the baseline specification. Second, once the 2018 census is used to measure the migrant share, the wage effect magnitude increases by around 20%, and this is true using each instrument and geographic unit of analysis. Thus, if we believe that the census is a more accurate measure of migrant locations, then the true wage effect becomes slightly larger, to −1.26 in the baseline specification. However, this is still not large enough to reconcile the primary results from Delgado-Prieto (2021): using the department unit (as in the paper), the wage coefficient ranges from −1.36 to −1.53 depending on which instrument is used.

The remaining difference in this paper's specification that explains the larger coefficient is the exclusion of return migrants in the migrant share, which – as we have seen – further inflates the wage coefficient. This is presented in Panel B of Table A4, where using the 2018 census, the 2SLS wage coefficient now ranges from −1.46 to −1.77 using CZ metro areas. At the department level, it ranges between −1.72 and −2.02. To conclude, the large coefficient found in Delgado-Prieto (2021) is not driven by stopping the analysis in 2018. It can be explained by a combination of using the 2018 census rather than the GEIH and, more importantly, excluding return migrants from the migrant share. If the reader's preferred specification is to measure total migration with the 2018 census using the migrant enclave IV, then the proper coefficient is between −1.09 and −1.13 using the administrative metro areas, between −1.21 and −1.26 using the CZ metro areas, or between −1.32 and −1.36 at the department level.[30]

A final paper is by Caruso *et al.* (2021), which finds a negative wage effect of −7.6% from a 1 pp increase in the migrant share, almost five times larger than the estimate by Delgado-Prieto (2021). One difference in this paper is that their analysis ends in 2017, before the majority of migrants arrived. They also conduct analysis at the department level combined with an inverse border distance instrument, which – we have seen – tends to generate a larger wage effect. However, the most consequential difference is that they define the migrant share as the share of the population that arrived from Venezuela over the past 1 year, rather than 5 years. Similar to the case of excluding return migrants, this will generate omitted-variable bias considering that these shares are highly positively correlated, and this will inflate the wage effect if previously arrived migrants also have a depressing effect on wages. In this example, this is also akin to

29 There are a few additional changes in how I use the GEIH in this specification so as to allow the results to be comparable with those estimated using the 2018 census. First, the migrant share includes migrants from all countries as opposed to only those from Venezuela. This is because the census does not ask for country of origin. However, this does not add very much noise because, according to the GEIH, 95% of foreigners who arrived over this period came from Venezuela. Second, the migrant share is taken as a fraction of the current population rather than the 2014 population, and again, this does not have a large effect on results.

30 Another result unique to Delgado-Prieto (2021) is the large estimated negative employment effect of –1.7% from a 1 pp increase in the migrant share. Using log total employment as the outcome in my primary specification, I find an insignificant negative effect of –0.2. However, this increases to –0.7 using the distance IV and increases further to –1.5 when data are restricted to the year 2018. This is consistent with Delgado-Prieto (2021)'s result that the employment effect of –1.7 coinciding with the 2018 migrant surge remains unchanged and, in fact, diminishes in 2019, despite large numbers of migrants continuing to arrive.

using a migration flow rather than a migration stock on the right-hand side.[31] In the two-way fixed effect framework, the resulting coefficient thus measures the effect of a change in the *flow* of migrants who arrived over the previous year. For example, in Bogotá, the 5-year migrant share was 0.48% in 2016 and 1.05% in 2017. Likewise, the 1-year migrant share was 0.24% in 2016 and 0.56% in 2017, the latter approximately reflecting the change in the 5-year migrant share over this period. When the model compares the change in the 1-year share, it attributes changes in the average wage in Bogotá between 2016 and 2017 to a change in migration of 0.32 pp, rather than the 0.57 pp increase in the migrant stock that actually occurred. In doing so, the researcher implicitly assumes that the economic effects of migration do not persist for >1 year.

In Table A5, I run the baseline specification using the migrant share defined in terms of 1-year flows, and I do this both through the full period and using data only through 2017 to test sensitivity to the period of analysis. The 1-year flow estimates are substantially larger across all specifications. In the baseline specification, the wage coefficient increases to −3.53.[32] Using the distance IV at the department level, the coefficient increases further to −4.17. When data are restricted to 2017, the coefficient inflates even further, to −4.80 in the baseline specification and −6.33 using the distance IV at the department level. Thus, the magnitude found by Caruso *et al.* (2021) can be explained primarily by the use of flows rather than stocks, second by the termination of analysis in 2017, and third by conducting analysis at the department level combined with an instrument based on distance to the border.[33]

To conclude, differences in these specification choices are able to explain the variation in the average wage effect reported in the literature. Choice among the candidate instruments and geographic unit of analysis leads to small differences in magnitudes, while choices relating to the measure of the migrant share are most consequential: when the migrant share is restricted to only a portion of the migrant population, or to those who arrived within the past year, the wage effect is substantially inflated. When the 2018 census is used to measure migration through 2018, the predicted wage effects also increase slightly. It is worth noting that, while the magnitude of the wage effect varies substantially, most of these papers find that the wage effect is larger for less-educated natives.

8 Additional Robustness

Armed with a preferred specification that uses an instrument based on historical migrant shares, variation across CZ metropolitan areas, and measuring the *total* migrant share using the GEIH, I now conduct various additional robustness tests for all four labor market outcomes, for the population average and by education group. The most important result from this section is that, after controlling for region-specific time trends, the wage effect falls to −0.59.

31 According to the GEIH, only 0.23% of the population had come from Venezuela in the past 5 years in 2014. Thus, using the 5-year migration measure until 2019 essentially measures the stock of migrants in the country who arrived since 2014.

32 Returning again to the framework in Section 1, this would be consistent with, for example, 1-year and 5-year migrations having an effect of −1.5 and −0.8, respectively, considering that the regression of the predicted 5-year share on the 1-year share generates a coefficient of 2.5, and $-1.5 - (0.8 \times 2.5) = -3.5$.

33 Indeed, in Table A4 in the Online Appendix, Caruso *et al.* (2021) run a specification in which they replace the 1-year migration measure with a 5-year migration measure, and the results are comparable with what I find here. However, these are not the primary results reported in the Abstract and Introduction sections.

I start by testing whether proximity to the Venezuelan border is a relevant omitted variable, since it is correlated with both the 2019 and historical migrant shares. In particular, cities very close to the border have experienced changes over this period in economic activity, commuting flows from Venezuela, and violent crime (Knight and Tribin, 2020). The cities with the highest migrant shares, including all of those with a migrant share >15%, are located within 100 km driving distance from the Venezuelan border. To ensure that these cities are not driving the results, I drop these six cities in Column 2 of Table 3. The instrument loses power (the first-stage Wald statistic falls to 13), and standard errors increase substantially. However, the coefficient values remain stable overall and within education group, implying that wage effects are not driven by these six metro areas. Effects on labor force participation, on the other hand, are eliminated when these cities are excluded.

Proximity to the Venezuelan border may remain an omitted variable if, among cities with distances of >100 km from the border, those closer to the border experienced decreases in trade with Venezuela over this period. While Venezuela used to be a top trading partner of Colombia, its trade shares steadily declined during the 2000s such that it represented a small share of imports and exports by 2010. However, trade persisted longer for departments closer to the border. In Column 3, I therefore control for total imports and exports with Venezuela at the department level, and the results are highly robust.[34]

To more flexibly account for distance to the border, I control for a linear trend interacted with inverse driving distance to the border. When I do this in Column 4, the estimated average wage elasticity falls to −0.53, but the first-stage F-statistic becomes weak and the standard error increases drastically, such that the coefficient is not statistically different from the original specification. I conclude that it is not feasible to isolate variation in the predicted migrant share from this distance measure. Thus, it remains possible that 2SLS results are partially explained by trends associated with border proximity.[35]

Considering this, another method to control for unobserved heterogeneity across broad geographic areas is to control for linear trends interacted with fixed effects for the four regions of Colombia defined by DANE: Pacific, Caribbean, Central, and Eastern, the last of which incorporates Bogotá. Accounting for regional time trends was also found to be important in the case of Syrian migration to Turkey (Aksu *et al.,* 2018). The region dummies are interacted with a time trend rather than year fixed effects to preserve power in the first stage. The results in Column 5 show that the wage effect is mitigated to −0.59 and remains statistically different from zero, though it is again not statistically different from the baseline estimate. This indicates that part of the 2SLS wage effect is driven by regional time trends, and after controlling for them, the true wage effect decreases. None of the papers discussed in Section 7 complete this robustness check. If I estimate the model using the 2018 census to measure migration as in Column 2, Row 2 of Table A4, addition of region fixed effects brings the coefficient down from −1.26 to −0.70.[36]

34 Trade data were downloaded from DANE and measure the net weight of all Venezuelan imports and exports with origin or destination in each department.

35 In all robustness checks with a linear trend, I can instead use interactions with year fixed effects, and results are qualitatively similar, though the standard errors are less precise. I can also use linear distance instead of driving distance, and results are similar. These are available upon request.

36 It is also notable that the wage effect for less-educated natives decreases the most when region trends are included, but the standard errors also increase the most, such that the effect for the less-than-secondary group is not significantly different from the baseline estimate and includes a wide range of plausible estimates. Thus, we are unable to isolate the effect from region trends for this subgroup.

Table 3 2SLS estimates' robustness

	(1)	(2)	(3)	(4)	(5)	(6)
	Original 2SLS	**Drop <100 km from border**	**Control trade with Venezuela**	**Year trend × inverse distance to border**	**Year trend × region**	**Year trend × pre-trend**
			ln(hourly wage)			
All	−1.05***	−1.02	−1.07***	−0.53	−0.59**	−1.06***
	(0.22)	(0.68)	(0.21)	(0.67)	(0.28)	(0.22)
Less than secondary	−1.42***	−1.43**	−1.45***	−0.45	−0.57	−1.42***
	(0.23)	(0.64)	(0.22)	(0.57)	(0.40)	(0.23)
Secondary	−0.86***	−0.93	−0.86***	−0.54	−0.32	−0.87***
	(0.22)	(0.68)	(0.22)	(0.58)	(0.23)	(0.22)
Postsecondary	−0.75*	−0.65	−0.76**	−0.33	−0.71**	−0.82**
	(0.38)	(0.99)	(0.35)	(1.18)	(0.30)	(0.36)
			ln(hours/week)			
All	0.27	−0.50*	0.27	−0.68***	0.47*	0.28
	(0.27)	(0.28)	(0.28)	(0.24)	(0.24)	(0.26)
Less than secondary	0.50	−0.58	0.51	−0.74**	0.67**	0.52
	(0.34)	(0.37)	(0.35)	(0.31)	(0.30)	(0.34)
Secondary	0.32	−0.29	0.32	−0.37	0.50**	0.32*
	(0.21)	(0.28)	(0.21)	(0.25)	(0.23)	(0.17)
Postsecondary	−0.02	−0.52*	−0.04	−0.80***	0.16	−0.02
	(0.23)	(0.28)	(0.25)	(0.22)	(0.22)	(0.22)
			Unemployment			
All	−0.08	−0.15	−0.09	−0.38**	−0.25*	−0.12
	(0.07)	(0.20)	(0.07)	(0.16)	(0.13)	(0.08)
Less than secondary	−0.05	−0.13	−0.06	−0.34**	−0.17*	−0.07
	(0.06)	(0.17)	(0.06)	(0.13)	(0.10)	(0.07)
Secondary	−0.02	−0.21	−0.02	−0.41***	−0.21	−0.04
	(0.09)	(0.19)	(0.08)	(0.16)	(0.14)	(0.08)
Postsecondary	−0.18*	−0.21	−0.18**	−0.47**	−0.37**	−0.19*
	(0.10)	(0.24)	(0.09)	(0.24)	(0.18)	(0.10)
			LFP			
All	−0.21*	0.07	−0.20*	0.05	−0.20**	-0.13
	(0.11)	(0.18)	(0.11)	(0.23)	(0.09)	(0.09)
Less than secondary	−0.23*	0.09	−0.23*	0.04	−0.20*	−0.17
	(0.12)	(0.23)	(0.12)	(0.31)	(0.11)	(0.11)
Secondary	−0.12**	−0.02	−0.09	0.00	−0.14***	−0.11
	(0.06)	(0.15)	(0.07)	(0.17)	(0.05)	(0.07)
Postsecondary	−0.24	0.13	−0.25	0.17	−0.26**	−0.21
	(0.16)	(0.16)	(0.16)	(0.19)	(0.13)	(0.15)
K-P Wald stat.	23.35	13.07	28.14	13.95	88.40	23.61
Number of metro areas	79	73	79	79	79	79

Notes: Outcomes are residualized and multiplied by 100. All models include Year FE and City FE. See text for description of robustness checks. Observations are weighted by city–year–group population. Cluster-robust standard errors are in parentheses.FE, fixed effect; K-P Wald stat., Kleibergen-Paap Wald statistic; LFP, labor force participation.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Next, motivated by the presence of pre-trends in unemployment and participation, in Column 6, I interact a linear year trend with the change in the outcome between 2013 and 2015, allowing linear changes over the migration period to vary flexibly with preperiod trends. As expected, this has little effect on the wage and hour estimates. Despite potential pre-trends for unemployment, the unemployment coefficients remain stable. The negative coefficient on participation decreases in magnitude and becomes insignificant for each education group, implying that the small decreases in participation are partially driven by trends that began before the Venezuelan exodus. This is consistent with evidence from Delgado-Prieto (2021) that native employment effects are mitigated after accounting for pre-trends.

Finally, in Table A6, I conduct various checks that have no effect on the estimated coefficients. First, I drop Bogotá, which is the largest city in the sample, to ensure that it is not disproportionately driving results.[37] Next, I drop all metro areas with an annual sample size of <1,000 observations, resulting in 27 major cities closely overlapping with the 23 official areas for which the GEIH is representative. This is to ensure that measurement error within small areas is not driving results. Third, I show that a first-difference model, in which the change in the outcome between 2014 and 2019 is regressed on the change in the migrant share, instrumented with the change in the instrument, produces results very similar to the two-way fixed-effects framework. This ensures that the bias of using two-way fixed effects in a context with potentially dynamic treatment effects is small, which was expected in this case, considering that most migrants did not arrive until close to the end of the study period. Fourth, there is a concern that the metro population, which is correlated over time, is in the denominator of both the endogenous variable and the instrument, and this may induce a spurious correlation that drives the first stage (Clemens and Hunt, 2019; Kronmal, 1993).[38] To test this, I replace the Venezuelan *share* of the population, M_{ct}, with a variable that represents the total number of Venezuelans. I then flexibly control for the baseline population by interacting it with year fixed effects. The results, in Column 10, show that the first stage remains strong, and that an increase of 100 migrants significantly reduces wages by 0.66%. At the mean population of 18,756, a 1% increase in the migrant share is therefore associated with a −1.23% wage effect, closely matching the initial estimates. Similar results are seen for other outcomes and subgroups. The final robustness check, regarding native internal migration, is discussed in Section 9.3.

9 Additional Results

9.1 Nonlinear effects

The analysis thus far assumes a log-linear relationship between wages and migration. To allow for nonlinearity in the migrant share, in Table A7, I run a quadratic model in which I include the squared migrant share as an endogenous variable and the squared instrument as an additional instrument. Overall and within education groups, the curvature is positive, but it is insignificant and extremely small in magnitude. With every 1 pp increase in the migrant share, the wage effect diminishes by 0.02 off a base effect of −1.5.

37 I can also drop all metro areas one by one, and this is available upon request.

38 For example, as discussed by Clemens and Hunt (2019), citing Kronmal (1993), "One would find storks-per-woman to be a strong instrument for babies-per-woman even if storks are irrelevant to babies, and that framework could show spuriously that babies cause any regional outcome that is correlated with the number of women in the region."

To visualize the wage effect across the distribution of the migrant share, I residualize the log wage and migrant share predicted by the instrument from the metro and year fixed effects, and I use them to estimate a locally weighted scatterplot smoothing (LOWESS) model in Figure A6. Focusing on the middle of the migrant share distribution, which contains the majority of observations, one sees that the slope is relatively constant and, if anything, becomes more negative at higher migrant shares. In other words, the effect of migration on wages becomes slightly larger as the migrant share increases. However, at very high and low migrant shares, the wage effect becomes flat and potentially even positive, though these tails are being determined by a small number of observations. It is clear that outliers are not responsible for driving the negative wage effect. To summarize, a linear model seems broadly appropriate, though it may be masking some negative curvature for small increases in migration and without sufficient variation in the data to identify nonlinear effects among very large or small migrant shares.

9.2 Occupation skill group and informality

It is also possible to have natives respond to migration by changing their occupation or type of work. In particular, given that migrants are concentrated in low-skill occupations, natives may have benefited by upgrading to higher-skill occupations where their labor is more complementary (Foged and Peri, 2016). While work transitions cannot be observed directly in the GEIH because it is cross sectional, in Table 4, I look at changes in total employment across occupation skill groups, again defined by ranking the mean native education in occupations before 2015, this time split into quintiles rather than deciles. These results are not conditional on working, so they reflect a combination of occupational movements and changes in employment (which is why rows do not sum to 1.0). This is done to avoid coefficient changes in one group being driven by workers exiting from a different group.

The results show small movements into middle-skill occupations on average, coming from both the lowest- and the highest-skill groups. The analysis by demographic group reveals that the upgrading from low- to middle-skill occupations is concentrated among men with completed secondary education (but not postsecondary education). Likewise, downgrading from high- to middle-skill occupations occurred among men with postsecondary education, especially in the age range of 35–44 years. However, these effects are small in magnitude: a 1 pp increase in the migrant share causes around a 0.15 pp shift out of low-skill occupations for men and people with completed secondary education. Women do not experience any occupational shift but are more likely to exit employment out of high-skill occupations. The analysis also demonstrates that the workforce exit among workers younger than 25 years of age is mostly out of low-skill occupations.

Another measure indicative of occupational downgrading or upgrading is self-reported underemployment. The GEIH includes a question that asks workers whether they would like to change their job in order to improve their use of skills or training. In Column 6, I show that a 1 pp shift in the migrant share causes an increase in underemployment of 0.29 pp on average. This effect is the strongest for the demographic groups which, according to the previous analysis, experienced occupational downgrading: men and workers with postsecondary education. Interestingly, these magnitudes are larger than those for the occupation skill groups. This indicates that Venezuelan migration induced a change in self-reported underemployment among workers who did not change occupation skill group, which could reflect changing

Table 4 2SLS effects on employment by occupation skill group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Occupation					Underemployed	K-P Wald stat.
	Group 1	Group 2	Group 3	Group 4	Group 5		
All	−0.06***	−0.03	0.07	−0.01	−0.09***	0.29**	25.69
	(0.02)	(0.08)	(0.04)	(0.02)	(0.02)	(0.13)	
Female	0.02	−0.05	0.02	−0.03	−0.07***	0.24*	25.35
	(0.03)	(0.05)	(0.06)	(0.02)	(0.02)	(0.12)	
Male	−0.15***	−0.00	0.12***	0.02	−0.11***	0.35**	26.10
	(0.03)	(0.10)	(0.02)	(0.03)	(0.02)	(0.14)	
Age 15–24 years	−0.16***	−0.19**	−0.01	−0.04	−0.01	0.15	24.77
	(0.03)	(0.07)	(0.07)	(0.05)	(0.02)	(0.11)	
Age 25–34 years	−0.11***	0.01	0.11**	0.15***	−0.11**	0.49***	23.15
	(0.03)	(0.13)	(0.05)	(0.04)	(0.05)	(0.16)	
Age 35–44 years	0.00	0.08	0.08**	−0.03	−0.18***	0.36**	25.02
	(0.03)	(0.09)	(0.04)	(0.04)	(0.04)	(0.14)	
Age 45–54 years	−0.04	0.07	0.12***	−0.08*	−0.04	0.35**	29.18
	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.15)	
Age 55–64 years	0.09**	−0.12	0.02	−0.05	−0.06***	0.09	29.02
	(0.04)	(0.09)	(0.08)	(0.03)	(0.02)	(0.08)	
Less than secondary	−0.09***	−0.04	−0.01	−0.02*	−0.01**	0.18	31.28
	(0.03)	(0.07)	(0.06)	(0.01)	(0.00)	(0.11)	
Secondary	−0.14***	−0.04	0.15***	−0.00	0.00	0.28*	26.91
	(0.03)	(0.06)	(0.05)	(0.05)	(0.01)	(0.15)	
Postsecondary	−0.02	−0.02	0.12**	0.04	−0.20***	0.49***	19.22
	(0.02)	(0.09)	(0.05)	(0.04)	(0.07)	(0.17)	

Notes: All models are 2SLS linear probability models for probability of employment in a group (multiplied by 100), not conditional on working, with Year FE and City FE. Occupations are ranked according to mean education of natives pre-2015 and grouped into quintiles. Workers are underemployed if they say they would like to change jobs to improve use of skills or training. Observations are weighted by city–year–group population. Cluster-robust standard errors are in parentheses.2SLS, two-stage least squares; FE, fixed effect; K-P Wald stat., Kleibergen-Paap Wald statistic.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

task requirements within an occupation group or decreased satisfaction with work caused by migration.

Finally, I study whether natives changed their formality status or shifted into own-account employment. In Table 5, I divide jobs into the same four mutually exclusive categories already discussed (formal salaried, informal salaried, own account, and employer). There is evidence that natives left the formal market-wage sector and went into nonemployment. As discussed by Delgado-Prieto (2021), this may have resulted from firms substituting informal migrant labor for formal native labor. This effect is persistent across all demographic groups. Thus, there is no evidence for natives upgrading to the formal sector. There are also no significant shifts of natives into informal wage work, own-account work, or being an employer. The workforce exit unique to people younger than 25 years of age, which was primarily out of lower-skill occupations, is driven by a reduction in own-account work.

Table 5 2SLS effects on employment by type of work

	(1)	(2)	(3)	(4)	(5)
	Formal salaried	**Informal salaried**	**Own account**	**Employer**	**K-P Wald stat.**
All	−0.09***	−0.02	−0.04	0.01	25.69
	(0.02)	(0.07)	(0.09)	(0.02)	
Female	−0.09***	−0.00	−0.03	0.01	25.35
	(0.03)	(0.06)	(0.10)	(0.01)	
Male	−0.08**	−0.03	−0.05	0.02	26.10
	(0.03)	(0.10)	(0.09)	(0.03)	
Age 15–24 years	−0.10***	−0.02	−0.31***	−0.01	24.77
	(0.03)	(0.08)	(0.11)	(0.00)	
Age 25–34 years	−0.02	−0.01	0.07	−0.00	23.15
	(0.05)	(0.08)	(0.11)	(0.02)	
Age 35–44 years	−0.09**	−0.04	0.10	0.04	25.02
	(0.04)	(0.08)	(0.09)	(0.03)	
Age 45–54 years	−0.10**	−0.00	0.02	0.03	29.18
	(0.04)	(0.06)	(0.08)	(0.03)	
Age 55–64 years	−0.10**	−0.03	−0.01	0.05**	29.02
	(0.05)	(0.05)	(0.11)	(0.02)	
Less than secondary	−0.10***	−0.05	−0.06	0.02	31.28
	(0.02)	(0.08)	(0.10)	(0.03)	
Secondary	−0.05	−0.07	0.06	−0.00	26.91
	(0.04)	(0.07)	(0.11)	(0.02)	
Postsecondary	−0.08*	0.01	−0.04	0.03	19.22
	(0.05)	(0.06)	(0.10)	(0.02)	

Notes: Models are 2SLS linear probability models for probability of employment in a group (multiplied by 100), not conditional on working, with Year FE and City FE. Wage-sector workers are split by formality, defined according to compliance with mandatory health and pension schemes. Self-employed workers are split into own account (with no employees, predominantly informal) and employer. Observations are weighted by city–year–group population. Cluster-robust standard errors are in parentheses. 2SLS, two-stage least squares; FE, fixed effect; K-P Wald stat., Kleibergen-Paap Wald statistic.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

In Table A8, I conduct each of the robustness checks from the previous section for this analysis. This shows that, while the downgrading is robust to each check, the upgrading goes away once I include the linear trend interacted with inverse distance to the border or region fixed effects. The increase in self-reported underemployment and decrease in formal salaried work are robust to each check.

9.3 Internal migration

An extensive literature documents that natives often respond to immigration by migrating away from the affected areas, perhaps because of changing labor market conditions or direct disutility from migrant exposure (Borjas, 2003; Borjas and Katz, 2007; Monras, 2020). This can occur through both increases in outmigration or reductions in inmigration, and it might

bias geography-based estimates of labor market impacts in two ways. First, it creates a compositional change in the inhabitants of a metro area. Second, it could remove from each city individuals who experienced differential labor market effects of the migration; for example, if those who outmigrated would have experienced large counterfactual wage decreases had they stayed, this would bias upward the estimated wage effect.

A major strength of the GEIH is the ability to see where each individual was living 5 years ago, including their previous municipality of residence in Colombia. Using this information, I can calculate the number of migrants who left or entered a metro area over a 5-year period. I then take this as a share of the 5-year lagged population to get the out- and inmigration rates for each metro area. Using the same 2SLS framework, I can study the causal effect of Venezuelan arrivals on this in- and outmigration. I restrict analysis to 5-year migration flows because they are less likely to be driven by measurement error or short-term temporary movements than 1-year flows. I only include data from 2014 and 2019 (thus identical to the difference model in Table A6) and thus measure the changes in the out- and inmigration rates from 2014 to 2019 induced by the arrival of migrants over this period.

Table 6 shows that, on average, Venezuelan arrivals cause a small increase in outmigration and a decrease in inmigration, but both effects are insignificant. Outmigration becomes significant and larger in magnitude among people older than 45 years of age and with secondary education or less. Among those with completed secondary education, a 1 pp increase in Venezuelan arrivals causes a 0.1 pp increase in outmigration over the period 2015–2019 relative to 2010–2014, or 1.5% of the base migration rate for this group. Moving from the 10th percentile to the 90th percentile of metro areas, this is associated with a 0.76 pp increase in outmigration. The magnitudes of changes in inmigration are slightly larger and are also concentrated among people with completed secondary education, but they are not significantly different from zero using a 10% test size. These OLS and 2SLS results are also similar, implying little selection of Venezuelans into areas that had differential trends in internal migration. Finally, Table A9 shows that these results also become very imprecise and insignificant after dropping metro areas close to the border. The effect on outmigration is actually reversed and becomes significantly negative, after including region-specific time trends.

To ensure that this does not bias my primary results, in Column 6 of Table A6, I assign all individuals to their metro area of residence 5 years before the survey year, and the results are stable. This robustness check holds constant the composition of the sample, thus eliminating any potential for compositional change to drive results. Furthermore, if one makes the assumption that outmigrants earn in their new location as much or more than what they would have had they stayed (or in the case of diverted inmigrants, had they come to the city), then this also generates a lower bound on the magnitude of the true wage effect. Thus, under the assumption of weakly improving wages for internal migrants, these estimates provide an upper bound on the true negative wage effect. Given the small magnitude of the internal migration response, this bias is likely to be small.

9.4 Regional heterogeneity

Thus far, results have been reported for the national average. Yet, the economic effects of migration may vary according to regional economic characteristics. Cities with high

Table 6 Effects on native internal migration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Outmigration			Inmigration		
	OLS	2SLS	Sample mean	OLS	2SLS	Sample mean	K-P Wald stat.
All	0.06	0.03	6.58	−0.04	−0.08	10.17	17.78
	(0.06)	(0.03)		(0.25)	(0.11)		
Male	0.08	0.06	6.76	0.01	−0.04	10.16	18.13
	(0.06)	(0.04)		(0.26)	(0.12)		
Female	0.05	0.01	6.42	−0.07	−0.11	10.17	17.48
	(0.06)	(0.03)		(0.23)	(0.11)		
Age 15–24 years	0.07	0.05	7.67	−0.02	−0.06	14.88	16.69
	(0.08)	(0.06)		(0.32)	(0.16)		
Age 25–34 years	0.05	−0.03	9.49	−0.09	−0.13	13.02	16.74
	(0.08)	(0.05)		(0.32)	(0.15)		
Age 35–44 years	0.05	0.03	6.64	−0.06	−0.10	9.03	17.02
	(0.09)	(0.05)		(0.25)	(0.11)		
Age 45–54 years	0.07	0.07**	3.97	−0.04	−0.09	5.62	20.18
	(0.05)	(0.03)		(0.14)	(0.06)		
Age 55–64 years	0.04	0.05*	3.09	−0.04	−0.09	4.24	20.09
	(0.03)	(0.03)		(0.13)	(0.07)		
Less than secondary	0.10*	0.07**	5.54	−0.01	−0.04	8.87	21.90
	(0.06)	(0.04)		(0.25)	(0.12)		
Secondary	0.08	0.10***	6.62	−0.08	−0.17	10.04	17.97
	(0.08)	(0.03)		(0.23)	(0.11)		
Postsecondary	0.00	−0.07	7.70	−0.02	−0.01	11.71	13.30
	(0.08)	(0.05)		(0.28)	(0.14)		

Notes: Sample includes the years 2014 and 2019. In- and outmigration rates are taken as shares of the 5-year lag population and multiplied by 100. All models include Year FE and City FE. Observations are weighted by city–year–group population. Cluster-robust standard errors are in parentheses.2SLS, two-stage least squares; FE, fixed effect; K-P Wald stat., Kleibergen-Paap Wald statistic; OLS, ordinary least squares.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

unemployment may have less capacity to absorb incoming migrants. The 2014 unemployment rate varies considerably across metro areas in Colombia, from 4% at the 5th percentile to 15% at the 95th percentile. The rates of informal and own-account work may also be relevant considering that migrants compete more directly with natives in these sectors. Business climate considerations are also important, both for documented migrants who would like to register a business and for natives who may respond to the increase in labor supply by forming or expanding a business. Consistent with these hypotheses, Aracı *et al*. (2021) find that the labor market consequences of Syrian refugees in Turkey are smaller in more-developed regions of Turkey.

In Table 7, I interact both the endogenous migrant share and the instrument with various preperiod metro-level economic characteristics measured in the GEIH, GDP per capita at the department level measured by DANE, as well as business climate indicators at the department

Table 7 2SLS wage effects interacted with regional characteristics

	ln(hourly wage)								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Migrant share	−0.97***	−1.14**	−1.02***	−0.55	−1.28***	−0.94***	−1.23***	−1.01***	−1.01***
	(0.26)	(0.55)	(0.23)	(0.39)	(0.23)	(0.30)	(0.20)	(0.37)	(0.30)
Migrant share interacted with: 2014 mean ln(hourly wage)	−0.36								
	(0.46)								
2014 unemployment rate		0.07							
		(0.29)							
2014 informal rate			−0.62*						
			(0.34)						
2014 own-account rate				−0.75**					
				(0.35)					
2017 WB DB (starting a business)					0.29*				
					(0.16)				
2017 WB DB (construction permits)						−0.31			
						(0.50)			
2017 WB DB (registering property)							−0.36		
							(0.43)		
2017 WB DB (paying taxes)								−0.06	
								(0.23)	
2014 per capita GDP									−0.13
									(0.32)
K-P Wald stat.	12.38	9.51	14.11	13.27	36.46	11.55	8.73	9.03	13.31
N	474	474	474	474	474	474	474	474	474

Notes: Both the endogenous variable and the instrument are interacted with the interaction variable indicated in each row. All interaction variables are centered around zero and divided by the standard error. 2014 metro-level economic indicators are measured using the GEIH. Doing Business measures at the department level were taken from the World Bank 2017 Doing Business report. The values of 2014 per capita GDP at the department level downloaded from DANE. All models include Year FE and City FE. Observations are weighted by city–year population. Cluster-robust standard errors are in parentheses. DANE, National Administrative Department of Statistics; FE, fixed effect; GDP, gross domestic product; GEIH, Colombian National Integrated Household Survey; K-P Wald stat., Kleibergen-Paap Wald statistic; WB DB, World Bank Doing Business.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

level taken from the 2017 World Bank Doing Business report.[39] Importantly, these results may not reflect the causal effect of regional heterogeneity since metro areas may differ according to unobservable characteristics. Furthermore, in some cases, the first-stage *F*-statistic falls below 16. Nonetheless, the results are indicative of the importance of regional heterogeneity for the economic effects of migration.

There is no significant heterogeneity in the wage effect according to baseline hourly wages, unemployment, or per capita GDP. However, the effect does become significantly more negative as the baseline informality and own-account rates increase, which may be driven by the

39 This report evaluates the regulatory environment across Colombia's department capitals in four fields: ease of starting a business, obtaining construction permits, registering property, and paying taxes. For details, see World Bank (2017). I assume that noncapital metro areas face the same regulatory environment as the capital. Results are similar when the analysis is restricted to administrative metro capitals (available upon request).

fact that migrants compete more directly with natives in these sectors. Moreover, as hypothesized, the wage effect becomes less negative as the ease of starting a business increases, though this is only significant at the 10% level. For departments 1 SD above and below the mean score in ease of starting a business, the wage effect is −0.99 and −1.57, respectively. None of the other Doing Business indicators are significant.

Table A10 shows that the result of wage effects increasing in informality and decreasing in ease of doing business are robust to the same list of robustness checks previously considered, though, in some cases, the standard error increases and the first-stage *F*-statistic becomes weak. The informality interaction coefficient becomes even larger and more significant with the inclusion of region-specific time trends, while the ease of doing business interaction coefficient remains stable. Overall, the results are suggestive that the ease of doing business and the size of the informal sector are relevant contributors to the wage effects of migration. From a policy perspective, this may indicate that it is desirable to not only reduce formality and facilitate business formation but also encourage or subsidize migrant relocation to areas that are better prepared according to these characteristics [as discussed regarding the Colombian setting, for example, in Bahar *et al*. (2018)].

10 Conclusion

The migration from Venezuela to Colombia presents a unique opportunity to better understand the short-term effects of mass migration on native labor market outcomes in a developing country and in a context where natives and migrants share a similar culture and language. The results show little effects on the employment margin: there was a small decrease in participation and increase in school attendance among people younger than 25 years of age, but this was driven entirely by cities close to the Venezuelan border and was mitigated after adjusting for pre-trends. However, there were negative and robust effects on native hourly wages most pronounced for informal and less-educated workers, consistent with low reservation wages and high wage flexibility. These results are not biased by the small increase in native internal migration that occurred in response to the Venezuelan arrival, or by the small shifts in occupational skill group that benefited some workers and harmed others. They are consistent with a pattern in the literature in which the economic consequences of migration are most pronounced in developing countries, especially for less-educated and informal workers, and clearly motivate policy responses to mitigate the economic consequences of migration. They also motivate additional research to better understand the drivers of these economic consequences in developing countries. That these wage effects are moderately stronger in metro areas with higher baseline informality rates and lower ease of starting a business indicates that local economic conditions are a determinant of the labor market effects of migration and motivates the formulation of policies to facilitate business formation or encourage migrant relocation according to local economic conditions. The role of migrants' occupational downgrading is explored extensively by Lebow (2022), and the results indicate that migrant downgrading plays an important role in concentrating wage effects among lower-income natives.

The robustness and sensitivity analysis identified two important caveats for the estimated average native wage effect of –1.05% from a 1 pp increase in the migrant share. First, when the

2018 census rather than the GEIH is used to measure the migrant share, the magnitude of the average wage effect increases to −1.26%. Second, after controlling for regional time trends, the wage effect decreases to −0.59% (or −0.70% using the 2018 census), with a 95% CI ranging from [−1.14, −0.04].

Finally, I demonstrated the specification choices that explain the variation in estimated wage effects in the literature studying Venezuelan migration to Colombia. The papers that find small or insignificant wage effects do not use an instrument and thus fail to account for the positive sorting of migrants into favorable locations. Larger wage effect magnitudes of −1.7% and −7% found in the literature are only partially explained by differences in the instrument, geographic unit, time period of analysis, or data used to measure migration. The most important determinant is that, when a subset of the migrant population is excluded from the migrant share or when the migrant share is defined to only include recently arrived migrants, the wage effect is inflated substantially. This is likely driven by omitted-variable bias and presents an important lesson for migration researchers.

Declarations

Availability of data and material

The data for this paper was provided by the Departamento Administrativo Nacional de Estadística (DANE) in Colombia and it is not publicly available at the level of geography that I use. While I therefore cannot provide the micro-data used in analysis, I can provide the Stata code and guidance for how to request this data directly from DANE.

Competing interests

The author, Jeremy Lebow, declares that he has no relevant or material financial interests that relate to the research described in this paper.

Funding

No additional funding, beyond the author's PhD stipend, was received for this research.

Author's contributions

This is a solo-authored paper, and it was completed entirely by the author without any other contributors.

Acknowledgment

I thank Duncan Thomas, Daniel Xu, Arnaud Maurel, and Erica Field, along with two anonymous referees and participants at the Duke Labor and Development workshops, for all of their helpful comments. I am also grateful to the *Departamento Administrativo Nacional de Estadística (DANE)* in Colombia for generously providing the data used in this paper.

References

Agudelo, Sonia A.; Hector Sala (2017): Wage Rigidities in Colombia: Measurement, Causes, and Policy Implications. Journal of Policy Modeling 39(3), 547-567.

Akgündüz, Yusuf Emre; Huzeyfe Torun (2018): Two and A Half Million Syrian Refugees, Skill Mix and Capital Intensity. Technical Report. GLO Discussion Paper No. 186.

Aksu, Ege; Refik Erzan; Murat Güray Kırdar (2018): The Impact of Mass Migration of Syrians on the Turkish Labor Market. Technical Report. Working Paper No. 1815.

Alix-Garcia, Jennifer; David Saah (2010): The Effect of Refugee Inflows on Host Communities: Evidence from Tanzania. The World Bank Economic Review 24(1), 148-170.

Alix-Garcia, Jennifer; Sarah Walker; Anne Bartlett; Harun Onder; Apurva Sanghi (2018): Do Refugee Camps Help or Hurt Hosts? The Case of Kakuma, Kenya. Journal of Development Economics 130, 66-83.

Altındağ, Onur; Ozan Bakiş; Sandra. Rozo (2020): Blessing or Burden? Impacts of Refugees on Businesses and the Informal Economy. Journal of Development Economics 146, 102490.

Altonji, Joseph G.; David Card (1991): The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives, John M. Abowd and Richard B. Freeman, editors, in: Immigration, Trade, and the Labor Market. Chicago: University of Chicago Press, 201-234.

Aracı, Doğu Tan; Murat Demirci; Murat G., Kirdar (2021): Development Level of Hosting Areas and the Impact of Refugees on Natives' Labor Market Outcomes. Technical Report. Institute of Labor Economics (IZA).

Bahar, Danny; Meagan Dooley (2019): Venezuela Refugee Crisis to Become the Largest and Most Underfunded in Modern History. Technical Report. Brookings.

Bahar, Dany; Ana María Ibáñez; Sandra V. Rozo (2021): Give Me Your Tired and Your Poor: Impact of A Large-Scale Amnesty Program for Undocumented Refugees. Journal of Development Economics 151, 102652.

Bahar, Dany; Cem Ozgüzel; Andreas Hauptmann; Hillel Rapoport (2019): Migration and Post-Conflict¨ Reconstruction: The Effect of Returning Refugees on Export Performance in the Former Yugoslavia.

Bahar, Dany; Meagan Dooley; Cindy Huang (2018 Dec.): Integrating Venezuelans into the Colombian Labor Market: Mitigating Costs and Maximizing Benefits. Policy Brief. Brookings.

Bahar, Dany; Prithwiraj Choudhury; Hillel Rapoport (2020): Migrant Inventors and the Technological Advantage of Nations. Research Policy 49, 103947.

Beaman, Lori; Jeremy Magruder (2012): Who Gets the Job Referral? Evidence from a Social Networks Experiment. American Economic Review 102(7), 3574-3593.

Bonilla-Mejía, Leonardo; Leonardo Fabio Morales-Zurita; Didier Hermida; Luz Adriana Flórez; Leonardo BonillaMejía; Leonardo Fabio Morales; Didier Hermida-Giraldo; Luz A. Flórez (2020): The Labor Market of Immigrants and Non-Immigrants Evidence from the Venezuelan Refugee Crisis. Borradores de Economía, No. 1119.

Borjas, George J. (2003): The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. The Quarterly Journal of Economics 118(4), 1335-1374.

Borjas, George J. (2017): The Wage Impact of the Marielitos: A Reappraisal. ILR Review 70(5), 1077-1110.

Borjas, George J.; Joan Monras (2017): The Labour Market Consequences of Refugee Supply Shocks. Economic Policy 32(91), 361-413.

Borjas, George J.; Lawrence F. Katz (2007): The Evolution of the Mexican-born Workforce in the United States. George J. Borjas, editor. Mexican Immigration to the United States. Chicago: University of Chicago Press, 13-56.

Boruchowicz, Cynthia; Martinelli, César; W. Parker, Susan (2021): Economic Consequences of Mass Migration: The Venezuelan Exodus in Peru (May 17, 2021). Available at SSRN: https://ssrn.com/abstract=3847897 or http://dx.doi.org/10.2139/ssrn.3847897.

Borusyak, Kirill; Xavier Jaravel (2021): Revisiting Event Study Designs. Unpublished Working Paper.

Braun, Sebastian; Toman Omar Mahmoud (2014): The Employment Effects of Immigration: Evidence from the Mass Arrival of German Expellees in Postwar Germany. The Journal of Economic History 74(1), 69-108.

Calderón-Mejía, Valentina; Ibáñez Ana María (2016): Labour Market Effects of Migration-Related Supply Shocks: Evidence from Internal Refugees in Colombia. Journal of Economic Geography 16(3), 695-713.

Callaway, Brantly; Pedro H.C. Sant'Anna (2021): Difference-in-Differences with Multiple Time Periods. Journal of Econometrics 225.2, 200-230.

Card, David (1990): The Impact of the Mariel Boatlift on the Miami Labor Market. ILR Review 43(2), 245-257.

Card, David (2001): Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. Journal of Labor Economics 19(1), 22-64.

Caruso, German; Christian Gomez Canon; Valerie Mueller (2021): Spillover Effects of the Venezuelan Crisis: Migration Impacts in Colombia. Oxford Economic Papers 73.2, 771-795.

Cengiz, Doruk; Hasan Tekgüç (2021): Is it Merely a Labor Supply Shock? Impacts of Syrian Migrants on Local Economies in Turkey. ILR Review, 0019793920978365.

Ceritoglu, Evren; H. Burcu Gurcihan Yunculer; Huzeyfe Torun; Semih Tumen (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), 5.

Clemens, Michael A.; Jennifer Hunt (2019): The Labor Market Effects of Refugee Waves: Reconciling Conflicting Results. ILR Review 72(4), 818-857.

Cohen-Goldner, Sarit; M. Daniele Paserman (2011): The Dynamic Impact of Immigration on Natives' Labor Market Outcomes: Evidence from Israel. European Economic Review 55(8), 1027-1045.

Cortes, Patricia (2008): The Effect of Low-Skilled Immigration on US Prices: Evidence from CPI Data. Journal of Political Economy 116(3), 381-422.

De Chaisemartin, Clément; Xavier d'Haultfoeuille (2020): Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects. American Economic Review 110(9), 2964-2996.

Del Carpio, Ximena V.; Mathis C. Wagner (2015): The Impact of Syrian Refugees on the Turkish Labor Market. World Bank Policy Research Working Paper No. 7402.

Delgado-Prieto, Lukas (2021): Immigration, Wages and Employment Under Informal Labor Markets. SocArXiv. October 30. doi:10.31235/osf.io/acr4v

Duranton, Gilles (2015): A Proposal to Delineate Metropolitan Areas in Colombia. Revista Desarrollo y Sociedad No. 75, 223-264.

Dustmann, Christian; Tommaso Frattini; Ian P. Preston (2013): The Effect of Immigration Along the Distribution of Wages. Review of Economic Studies 80(1), 145-173.

Dustmann, Christian; Uta Schönberg; Jan Stuhler (2016): The Impact of Immigration: Why Do Studies Reach Such Different Results? Journal of Economic Perspectives 30(4), 31-56.

Dustmann, Christian; Uta Schönberg; Jan Stuhler (2017): Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment. The Quarterly Journal of Economics 132(1), 435-483.

Edo, Anthony (2020): The Impact of Immigration on Wage Dynamics: Evidence from the Algerian Independence War. Journal of the European Economic Association 18(6), 3210-3260.

El Espectador (2019): Más de cuatro millones de personas quedaron por fuera del censo. Technical Report.

Fallah, Belal; Caroline Krafft; Jackline Wahba (2019): The Impact of Refugees on Employment and Wages in Jordan. Journal of Development Economics 139, 203-216.

Foged, Mette; Giovanni Peri (2016): Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data. American Economic Journal: Applied Economics 8(2), 1-34.

Freitez, Anitza (2011): La emigración desde Venezuela durante la última década. Revista temas de coyuntura 63.

Friedberg, Rachel M. (2001): The Impact of Mass Migration on the Israeli Labor Market. The Quarterly Journal of Economics 116(4), 1373-1408.

Frölich, Markus (2008): Parametric and Nonparametric Regression in the Presence of Endogenous Control Variables. International Statistical Review 76(2), 214-227.

Goldsmith-Pinkham, Paul; Isaac Sorkin; Henry Swift (2020): Bartik Instruments: What, When, Why, and How. American Economic Review 110(8), 2586-2624.

Goodman-Bacon, Andrew (2021): Difference-in-differences with Variation in Treatment Timing. Journal of Econometrics 225.2, 254-277.

Graham, Jimmy; Martha Guerrero Ble; Daphne Panayotatos; Izza Leghtas (2020): From Displacement to Development: How Colombia Can Transform Venezuelan Displacement into Shared Growth. Policy paper from the Center for Global Development. Available at: https://www.cgdev.org/publication/displacement-development-how-colombia-can-transform-venezuelan-displacement-shared

Guriev, Sergei; Biagio Speciale; Michele Tuccio (2019): How do Regulated and Unregulated Labor Markets Respond to Shocks? Evidence from Immigrants During the Great Recession. The Journal of Law, Economics, and Organization 35(1), 37-76.

IPUMS (2019): Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN.

Jaeger, David A.; Joakim Ruist; Jan Stuhler (2018): Shift-share Instruments and the Impact of Immigration. Technical Report. National Bureau of Economic Research.

Knight, Brian G.; Ana Tribin (2020): Immigration and Violent Crime: Evidence from the Colombia Venezuela Border. Technical Report. National Bureau of Economic Research.

Kronmal, Richard A. (1993): Spurious Correlation and the Fallacy of the Ratio Standard Revisited. Journal of the Royal Statistical Society: Series A (Statistics in Society) 156(3): 379-392.

Lebow, Jeremy (2022): Immigration and Occupational Downgrading in Colombia. Unpublished Manuscript. Available at: www.jeremylebow.com/s/Lebow2021.pdf

Lombardo, Carlo; Julián Martínez Correa; Leonardo José Peñaloza Pacheco; Leonardo Carlos Gasparini (2021): The Distributional Effect of a Massive Exodus in Latin America and the Role of Downgrading and Regularization. Documentos de Trabajo del CEDLAS.

Manacorda, Marco; Alan Manning; Jonathan Wadsworth (2012): The Impact of Immigration on the Structure of Wages: Theory and Evidence from Britain. Journal of the European Economic Association 10(1), 120-151.

Mayda, Anna Maria; Chris Parsons; Giovanni Peri; Mathis Wagner (2017): The Labor Market Impact of Refugees: Evidence from the US Resettlement Program. US Department of State Office of the Chief Economist Working Paper, 4, 2017.

Maystadt, Jean-François; Philip Verwimp (2014): Winners and Losers Among a Refugee-Hosting Population. Economic Development and Cultural Change 62(4), 769-809.

McKenzie, David; Hillel Rapoport (2007): Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico. Journal of Development Economics 84(1), 1-24.

Migración Colombia (2019): Total de Venezolanos en Colombia Corte a 31 de Marzo de 2019. Technical Report.

Migration Policy Institute (2020): An Uneven Welcome: Latin American and Caribbean Responses to Venezuelan and Nicaraguan Migration. Technical Report.

Monras, Joan (2020): Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis. Journal of Political Economy 128(8), 3017-3089.

Morales, Fernando; Martha Denisse Pierola (2020): Venezuelan Migration in Peru: Short-term Adjustments in the Labor Market. Inter-American Development Bank.

Morales, Juan S. (2018): The Impact of Internal Displacement on Destination Communities: Evidence from the Colombian conflict. Journal of Development Economics 131, 132-150.

Munshi, Kaivan (2003): Networks in the Modern Economy: Mexican Migrants in the US Labor Market. The Quarterly Journal of Economics 118(2), 549-599.

Olivieri, Sergio; Francesc Ortega; Ana Rivadeneira; Eliana Carranza (2021): The Labour Market Effects of Venezuelan Migration in Ecuador. The Journal of Development Studies, 1–17. doi: 10.1080/00220388.2021.1988077

Ottaviano, Gianmarco I.P.; Giovanni Peri (2012): Rethinking the Effect of Immigration on Wages. Journal of the European Economic Association 10(1), 152-197.

Penaloza-Pacheco, Leonardo (2021): Living with the Neighbors: The Effect of Venezuelan Forced Migration on the Labor Market in Colombia. Unpublished manuscript. Available at: https://static1.squarespace.com/static/609c48182747022f80ae9ed1/t/609db0d0ece4d67fcb09179f/1620947220550/Migration.pdf

Peri, Giovanni (2016): Immigrants, Productivity, and Labor Markets. Journal of Economic Perspectives 30(4), 3-30.

Peri, Giovanni; Chad Sparber (2009): Task Specialization, Immigration, and Wages. American Economic Journal: Applied Economics 1(3), 135-169.

Peri, Giovanni; Derek Rury; Justin Wiltshire (2020): The Economic Impact of Migrants from Hurricane Maria. No. w27718. National Bureau of Economic Research.

Peri, Giovanni; Vasil Yasenov (2019): The Labor Market Effects of a Refugee Wave Synthetic Control Method Meets the Mariel Boatlift. Journal of Human Resources 54(2), 267-309.

Reina, Mauricio; Carlos Antonio Mesa; Tomás Ramírez Tobón (2018): Elementos para una pol´ıtica pu´blica frente a la crisis venezuela. Technical Report.

Rozo, Sandra V.; Juan F. Vargas (2021): Brothers or Invaders? How Crisis-driven Migrants Shape Voting Behavior. Journal of Development Economics 150, 102636.

Ruiz, Isabel; Carlos Vargas-Silva (2015): The Labor Market Impacts of Forced Migration. American Economic Review 105(5), 581-586.

Ryu, Hanbyul; Jayash Paudel (2022): Refugee Inflow and Labor Market Outcomes in Brazil: Evidence from the Venezuelan Exodus. Population and Development Review 48(1), 75-96.

Santamaria, Julieth (2020): 'When a Stranger Shall Sojourn with Thee': The Impact of the Venezuelan Exodus on Colombian Labor Markets. Mimeo: University of Minnesota.

Tabellini, Marco (2020): Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration. The Review of Economic Studies 87(1), 454-486.

Tribín-Uribe, Ana María; Adhvaryu Achyuta; Cesar Anzola; Oscar Iván Ávila-Montealegre; Leonardo Bonilla-Mejía; Juan Carlos Castro-Fernández; Luz Adriana Flórez; Ánderson Grajales-Olarte; Alexander Guarín-López; Franz Hamann-Salcedo; Didier Hermida-Giraldo; Gaurav Khanna; Francisco Javier Lasso-Valderrama; Carlos Medina-Durango; Ligia Alba Melo-Becerra; Juan Camilo Méndez-Vizcaino; Leonardo Fabio Morales-Zurita; Anant Nyshadam; Juan Jose Ospina-Tejeiro; Andrea Otero-Cortés; José David Pulido; Jorge Ramos-Forero; Mario Ramos-Veloza; Jorge Andrés Tamayo-Castaño; Santiago Velásquez; (2020): Migración desde Venezuela en Colombia: caracterización del fenómeno y análisis de los efectos macroeconómicos. Revista Ensayos Sobre Política Económica, 97, 1-74.

Tumen, Semih (2016): The Economic Impact of Syrian Refugees on Host Countries: Quasi-Experimental Evidence from Turkey. American Economic Review 106(5), 456-460.

UNHCR (2018): Global Trends: Forced Displacement in 2018.

UNHCR (2019): Global Trends: Forced Displacement in 2019.

Verme, Paolo; Kirsten Schuettler (2021): The Impact of Forced Displacement on Host Communities a Review of the Empirical Literature in Economics. Journal of Development Economics, 150, 102606.

Wilson Center (2019): Understanding the Venezuelan Refugee Crisis. Technical Report.

World Bank (2017): Doing Business in Colombia 2017. Technical Report.

World Bank (2018): Migración desde Venezuela a Colombia. Technical Report.

Zago, Rafael Alfena (2020): Local Labor Market Impacts of Immigration: The Case of Venezuela and Brazil. Mimeo: University of Oklahoma.

Appendix

Table A1 Characteristics of migrants and nonmigrants

	Nonmigrants	Migrants
Male (%)	48.5	49.6
	(50)	(50)
Age (years)	36.5	31.5
	(13.9)	(11.3)
Labor force participation (%)	71.7	79.4
	(45)	(40.5)
Unemployment (%)	11.4	14.8
	(31.7)	(35.5)
Median hourly wage (2010 USD)	2.3	1.6
	(6.1)	(4.8)
Hours/week	45.2	49.6
	(15.9)	(17.4)
Own account (%)	25.3	32.1
	(43.5)	(46.7)
Informal (%)	56	88.3
	(49.6)	(32.1)
N	447,264	21,730

Notes: Means are presented, SDs are in parentheses; data are restricted to urban residents of age 15–64 years. Migrant is defined as anyone in 2019 who was living in Venezuela 5 years ago. Population weights are applied. GEIH, Colombian National Integrated Household Survey; SD, standard deviation.

Source: GEIH (2019).

Table A2 Preperiod correlates of 2005 Venezuelan share of population

	(1)	(2)	(3)	(4)	(5)
Population (100,000)	−0.001	0.005	0.004	0.004	0.002
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
100*ln(hourly wage)		−0.017**	−0.018*	−0.018*	−0.024*
		(0.008)	(0.009)	(0.009)	(0.026)
100*ln(hours/week)			0.021	0.021	0.026
			(0.017)	(0.017)	(0.019)
100* unemployment rate				0.047	0.049
				(0.125)	(0.126)
100* LPP rate					0.041
					(0.047)
N	79	79	79	79	79
R^2	1.6e−04	0.051	0.073	0.073	0.09
2014 metro characteristic	$M_{c,2005}$	$M_{c,2005}$	$M_{c,2005}$	$M_{c,2005}$	$M_{c,2005}$

Notes: $M_{c,2005}$ is the 2005 Venezuelan share of the metro area, which ranges from 0% to 0.9%, and has been normalized to a mean of zero and SD of one. Observations are weighted by 2014 population. Robust standard errors are in parentheses. LPP, Labor force participation; SD, standard deviation.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A3 Wage estimate sensitivity

	Panel A: including return migrants			
	(1)	**(2)**	**(3)**	**(4)**
	OLS	**IV: 2005 census**	**IV: 1993 census**	**IV: inverse distance**
	Geographic unit: CZ metro areas			
Migrant share	−0.73*	−1.05***	−1.13***	−1.16***
	(0.42)	(0.22)	(0.21)	(0.17)
K-P Wald stat.		23.35	21.78	28.54
Number of units	79	79	79	79
	Geographic unit: administrative metro areas			
Migrant share	−0.67	−0.94***	−1.03***	−1.08***
	(0.49)	(0.26)	(0.20)	(0.16)
K-P Wald stat.		29.58	19.88	46.11
Number of units	23	23	23	23
	Geographic unit: department			
Migrant share	−0.83	−1.19***	−1.21***	−1.27***
	(0.51)	(0.28)	(0.30)	(0.18)
K-P Wald stat.		19.07	20.90	35.20
Number of units	24	24	24	24
	Panel B: excluding return migrants			
	(5)	**(6)**	**(7)**	**(8)**
	OLS	**IV: 2005 census**	**IV: 1993 census**	**IV: inverse distance**
	Geographic unit: CZ metro areas			
Migrant share	−0.87**	−1.24***	−1.32***	−1.34***
	(0.42)	(0.26)	(0.25)	(0.21)
K-P Wald stat.		26.42	21.52	29.50
Number of units	79	79	79	79
	Geographic unit: administrative metro areas			
Migrant share	−0.79	−1.09***	−1.18***	−1.24***
	(0.48)	(0.30)	(0.23)	(0.18)
K-P Wald stat.		30.17	17.76	39.35
Number of units	23	23	23	23
	Geographic unit: department			
Migrant share	−0.97*	−1.41***	−1.44***	−1.48***
	(0.52)	(0.34)	(0.35)	(0.21)
K-P Wald stat.		19.55	19.38	31.37
Number of units	24	24	24	24

Notes: The outcome, namely, log hourly wage, is residualized and multiplied by 100. All models include Year FE and Unit FE. Panel A includes Colombian-born migrants in the migrant share, while Panel B excludes them. Columns 2–4 use different share components of the shift-share IV, based respectively on the complete 2005 census (as in the main analysis), the 10% subsample of the 1993 census, and the inverse minimum driving distance from the metro area or department capital to the closest Venezuelan border crossing. Geographic units include the 79 CZ-defined metro areas (as in the main analysis), the 23 official metro areas of which the GEIH is representative, and the 24 departments of Colombia. Observations are weighted by unit–year population. Cluster-robust standard errors are in parentheses. CZ, commuting zone; FE, fixed effect; GEIH, Colombian National Integrated Household Survey; IV, instrumental variable; K-P Wald stat., Kleibergen-Paap Wald statistic; OLS, ordinary least squares.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A4 Wage estimate sensitivity: 2014–2018 difference model

	Panel A: including return migrants			
	(1)	(2)	(3)	(4)
	OLS	IV: 2005 census	IV: 1993 census	IV: inverse distance
	Geographic unit: CZ metro areas			
Foreigner share (2018 GEIH)	−0.26	−1.07***	−1.05***	−1.24***
	(0.60)	(0.39)	(0.35)	(0.24)
Foreigner share (2018 census)	−0.54	−1.26***	−1.21***	−1.45***
	(0.49)	(0.46)	(0.42)	(0.28)
Number of units	79	79	79	79
	Geographic unit: administrative metro areas			
Foreigner share (2018 GEIH)	−0.09	−0.87**	−0.89***	−1.07***
	(0.63)	(0.43)	(0.33)	(0.23)
Foreigner share (2018 census)	−0.39	−1.13**	−1.09**	−1.35***
	(0.54)	(0.57)	(0.43)	(0.29)
Number of units	23	23	23	23
	Geographic unit: department			
Foreigner share (2018 GEIH)	−0.25	−1.02***	−0.97**	−1.19***
	(0.59)	(0.39)	(0.41)	(0.23)
Foreigner share (2018 census)	−0.46	−1.36***	−1.32**	−1.53***
	(0.79)	(0.50)	(0.53)	(0.29)
Number of units	24	24	24	24
	Panel B: excluding return migrants			
	(5)	(6)	(7)	(8)
	OLS	IV: 2005 census	IV: 1993 census	IV: inverse distance
	Geographic unit: CZ metro areas			
Foreigner share (2018 GEIH)	−0.62	−1.58***	−1.53***	−1.80***
	(0.75)	(0.58)	(0.52)	(0.34)
Foreigner share (2018 census)	−0.65	−1.55***	−1.46***	−1.77***
	(0.55)	(0.56)	(0.51)	(0.34)
Number of units	79	79	79	79
	Geographic unit: administrative metro areas			
Foreigner share (2018 GEIH)	−0.36	−1.27**	−1.26***	−1.54***
	(0.76)	(0.63)	(0.48)	(0.32)
Foreigner share (2018 census)	−0.48	−1.38**	−1.31**	−1.65***
	(0.61)	(0.70)	(0.53)	(0.36)
Number of units	23	23	23	23
	Geographic unit: department			
Foreigner share (2018 GEIH)	−0.55	−1.54***	−1.43**	−1.77***
	(0.75)	(0.59)	(0.61)	(0.33)
Foreigner share (2018 census)	−0.55	−1.81***	−1.72**	−2.02***
	(0.99)	(0.67)	(0.70)	(0.38)
Number of units	24	24	24	24

Notes: See notes to Table A3. This model is a regression of the change in residual wages from 2014 to 2018 on the 2018 foreigner share, calculated using the GEIH or Census as a share of the 2018 population. The instrument is created analogously (interacted with the national migration rate from the GEIH or census). In no case does the K-P Wald statistic fall below 10. CZ, commuting zone; FE, fixed effect; GEIH, Colombian National Integrated Household Survey; IV, instrumental variable; K-P Wald stat., Kleibergen-Paap Wald statistic; OLS, ordinary least squares.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A5 Wage estimate sensitivity: 1-year migration measure

	Panel A: including return migrants			
	(1)	(2)	(3)	(4)
	OLS	IV: 2005 census	IV: 1993 census	IV: inverse distance
	Geographic unit: CZ metro areas			
1-year migrant share (data until 2019)	−1.45	−3.53***	−3.96***	−3.82***
	(1.29)	(0.80)	(0.90)	(0.67)
1-year migrant share (data until 2017)	−2.18	−4.80***	−5.15***	−5.74***
	(2.35)	(1.36)	(1.25)	(0.91)
Number of units	79	79	79	79
	Geographic unit: administrative metro areas			
1-year migrant share (data until 2019)	−1.38	−3.05***	−3.60***	−3.58***
	(1.67)	(0.89)	(0.82)	(0.59)
1-year migrant share (data until 2017)	−2.76	−4.40**	−5.03***	−5.65***
	(3.03)	(1.74)	(1.35)	(0.96)
Number of units	23	23	23	23
	Geographic unit: department			
1-year migrant share (data until 2019)	−1.80	−4.01***	−4.31***	−4.17***
	(1.65)	(1.01)	(1.19)	(0.63)
1-year migrant share (data until 2017)	−3.58	−5.38***	−5.29***	−6.33***
	(3.15)	(1.59)	(1.63)	(0.92)
Number of units	24	24	24	24
	Panel B: excluding return migrants			
	(5)	(6)	(7)	(8)
	OLS	IV: 2005 census	IV: 1993 census	IV: inverse distance
	Geographic unit: CZ metro areas			
1-year migrant share (data until 2019)	−1.74	−3.75***	−4.28***	−3.94***
	(1.29)	(0.85)	(1.06)	(0.75)
1-year migrant share (data until 2017)	−3.64	−5.51***	−5.92***	−6.64***
	(2.85)	(1.48)	(1.30)	(0.96)
Number of units	79	79	79	79
	Geographic unit: administrative metro areas			
1-year migrant share (data until 2019)	−1.58	−3.03***	−3.55***	−3.55***
	(1.57)	(0.87)	(0.79)	(0.59)
1-year migrant share (data until 2017)	−3.78	−4.85**	−5.49***	−6.42***
	(3.23)	(1.88)	(1.40)	(1.02)
Number of units	23	23	23	23
	Geographic unit: department			
1-year migrant share (data until 2019)	−1.97	−4.14***	−4.45***	−4.30***
	(1.63)	(1.03)	(1.21)	(0.67)
1-year migrant share (data until 2017)	−4.82	−6.68***	−6.50***	−8.07***
	(3.41)	(1.93)	(1.93)	(1.08)
Number of units	24	24	24	24

Notes: See notes to Table A3. This is the original two-way fixed-effects model, replacing the 5-year migrant share with the 1-year migrant share (the share of the population that arrived from Venezuela in the past 12 months) and varying the end of the period of analysis. In no case does the K-P Wald stat. fall below 10.

CZ, commuting zone; FE, fixed effect; GEIH, Colombian National Integrated Household Survey; IV, instrumental variable; K-P Wald stat., Kleibergen-Paap Wald statistic; OLS, ordinary least squares.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A6 2SLS estimates additional robustness

	(1)	(2)	(3)	(4)	(5)	(6)
	Original 2SLS	**Drop Bogotá**	**Drop metro yearly sample <1,000**	**2014–2019 Difference model**	**Kronmal specification**	**Assign natives to past metro**
			ln(hourly wage)			
All	−1.05***	−1.03***	−1.04***	−1.09***	−0.66***	−1.04***
	(0.22)	(0.22)	(0.23)	(0.27)	(0.15)	(0.20)
Less than secondary	−1.42***	−1.41***	−1.41***	−1.33***	−0.95***	−1.38***
	(0.23)	(0.24)	(0.25)	(0.27)	(0.17)	(0.22)
Secondary	−0.86***	−0.87***	−0.85***	−0.79***	−0.54***	−0.80***
	(0.22)	(0.22)	(0.23)	(0.28)	(0.15)	(0.21)
Postsecondary	−0.75*	−0.69*	−0.75*	−0.96**	−0.46**	−0.80**
	(0.38)	(0.37)	(0.39)	(0.48)	(0.23)	(0.36)
			ln(hours/week)			
All	0.27	0.28	0.27	0.27	0.17	0.29
	(0.27)	(0.27)	(0.27)	(0.32)	(0.17)	(0.27)
Less than secondary	0.50	0.52	0.51	0.51	0.34	0.51
	(0.34)	(0.34)	(0.34)	(0.39)	(0.22)	(0.34)
Secondary	0.32	0.32	0.30	0.26	0.20	0.32
	(0.21)	(0.21)	(0.21)	(0.26)	(0.12)	(0.21)
Postsecondary	−0.02	−0.02	−0.00	0.01	−0.01	0.02
	(0.23)	(0.22)	(0.22)	(0.25)	(0.12)	(0.22)
			Unemployment			
All	−0.08	−0.08	−0.09	−0.07	−0.05	−0.07
	(0.07)	(0.07)	(0.08)	(0.07)	(0.04)	(0.07)
Less than secondary	−0.05	−0.05	−0.06	−0.03	−0.04	−0.06
	(0.06)	(0.07)	(0.07)	(0.06)	(0.04)	(0.06)
Secondary	−0.02	−0.01	−0.03	0.02	−0.01	−0.01
	(0.09)	(0.08)	(0.09)	(0.09)	(0.05)	(0.08)
Postsecondary	−0.18*	−0.18*	−0.17*	−0.20*	−0.11*	−0.15
	(0.10)	(0.10)	(0.10)	(0.10)	(0.06)	(0.10)
			LFP			
All	−0.21*	−0.21**	−0.21**	−0.17*	−0.13**	−0.20**
	(0.11)	(0.11)	(0.10)	(0.10)	(0.07)	(0.10)
Less than secondary	−0.23*	−0.24*	−0.24**	−0.16	−0.16*	−0.24**
	(0.12)	(0.12)	(0.11)	(0.10)	(0.08)	(0.12)
Secondary	−0.12**	−0.12**	−0.12**	−0.10	−0.07**	−0.12**
	(0.06)	(0.06)	(0.06)	(0.07)	(0.04)	(0.05)
Postsecondary	−0.24	−0.25	−0.23	−0.27	−0.15*	−0.21
	(0.16)	(0.16)	(0.17)	(0.18)	(0.08)	(0.15)
K-P Wald stat.	23.35	23.41	19.28	30.89	24.17	24.47
Number of metro areas	79	78	27	79	79	79

Notes: Outcomes are residualized and multiplied by 100. All models include Year FE and City FE. Column 4 regresses the change in outcomes from 2014 to 2019 on the change in the migrant share. Column 5 replaces the migrant share with the number of migrants (divided by 100) and controls for baseline population interacted with year FEs. Column 6 assigns natives to their 5-year lagged metro area of residence. Observations are weighted by city–year–group population. Cluster-robust standard errors are in parentheses. 2SLS, two-stage least squares; FE, fixed effect; K-P Wald stat., Kleibergen-Paap Wald statistic; LFP, labor force participation.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A7 2SLS estimates with migrant share quadratic

	(1)	(2)	(3)	(4)
	ln(hourly wage)			
	All	Less than secondary	Secondary	Postsecondary
Migrant share	−1.40	−2.37**	−1.14	−0.79
	(1.12)	(1.11)	(1.02)	(1.67)
Migrant share^2	0.02	0.05	0.02	0.00
	(0.05)	(0.05)	(0.05)	(0.08)
Joint *F*-statistic	62.36	55.46	39.73	25.46
K-P Wald stat.	58.02	35.66	33.97	89.30

Notes: Linear and quadratic migrant shares are instrumented with the linear and quadratic forms of the IV. Wages are residualized and multiplied by 100. All models include Year FE and City FE. Observations are weighted by city–year–group population. Cluster-robust standard errors are in parentheses. 2SLS, two-stage least squares; FE, fixed effect; K-P Wald stat., Kleibergen-Paap Wald statistic.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A8 Robustness check of 2SLS effects on occupation skill group and type of work

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Original 2SLS	**Drop <100 km from border**	**Control trade with Venezuela**	**Year trend × inverse distance to border**	**Year trend × region**	**Year trend × pre-trend**	**Drop Bogotá**	**Drop metro yearly sample <1,000**	**2014–2019 difference model**	**Kronmal specification**	**Assign natives to past metro**
Occup skill Group 1	−0.06***	−0.05	−0.05**	0.04	−0.00	−0.06**	−0.07***	−0.07***	−0.06***	−0.04***	−0.06**
	(0.02)	(0.08)	(0.03)	(0.07)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Occup skill Group 2	−0.03	0.15	−0.05	0.21**	0.03	−0.01	−0.03	−0.03	−0.04	−0.02	−0.03
	(0.08)	(0.10)	(0.07)	(0.10)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.04)	(0.08)
Occup skill Group 3	0.07	0.16**	0.08	0.21***	0.05	0.09**	0.06	0.06	0.08*	0.04	0.06
	(0.04)	(0.08)	(0.05)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)
Occup skill Group 4	−0.01	0.03	0.00	0.04	0.02	−0.01	−0.01	−0.00	−0.00	−0.01	−0.01
	(0.02)	(0.05)	(0.03)	(0.06)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Occup skill Group 5	−0.09***	−0.11*	−0.09***	−0.14**	−0.11***	−0.09***	−0.08***	−0.09***	−0.09***	−0.06***	−0.09***
	(0.02)	(0.06)	(0.02)	(0.07)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Underemployed	0.29**	0.52	0.26**	0.57*	0.31**	0.29**	0.28**	0.30**	0.33**	0.18**	0.28**
	(0.13)	(0.32)	(0.12)	(0.32)	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)	(0.08)	(0.12)
Formal salaried	−0.09***	−0.10**	−0.08***	−0.07	−0.09*	−0.09***	−0.08***	−0.09***	−0.07***	−0.06***	−0.06***
	(0.02)	(0.05)	(0.03)	(0.04)	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)
Informal salaried	−0.02	0.07	0.01	0.18	0.09	−0.04	−0.03	−0.01	−0.07	−0.01	0.00
	(0.07)	(0.16)	(0.07)	(0.15)	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)	(0.05)	(0.06)
Own account	−0.04	0.09	−0.06	0.13	−0.08	−0.04	−0.04	−0.05	0.01	−0.03	−0.07
	(0.09)	(0.21)	(0.08)	(0.22)	(0.13)	(0.09)	(0.09)	(0.09)	(0.07	(0.05)	(0.09)
Employer	0.01	0.04	0.02	0.07	0.06***	0.02	0.01	0.02	0.02	0.01	0.01
	(0.02)	(0.05)	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02	(0.01)	(0.02)
Kleibergen-Paap Wald stat.	25.69	12.80	30.99	14.50	102.73	24.59	25.77	20.69	33.17	26.38	26.37
Number of metro areas	79	73	79	79	79	79	78	27	79	79	79

Notes: Outcomes are multiplied by 100. All models are 2SLS linear probability models for probability of employment in a group, not conditional on working, with Year FE and City FE. Observations are weighted by city–year–group population. See paper for description of robustness checks. Cluster-robust standard errors are in parentheses. 2SLS, two-stage least squares; FE, fixed effect; Occup, occupational.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A9 Robustness of 2SLS effects on native internal migration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Original 2SLS	**Drop <100 km from border**	**Control trade with Venezuela**	**Year trend × inverse distance to border**	**Year trend × region**	**Drop Bogotá**	**Drop metro yearly sample <1,000**	**Kronmal specification**
				Outmigration				
All	0.03	0.01	0.03	−0.09	−0.11**	0.05*	0.03	0.02
	(0.03)	(0.08)	(0.05)	(0.12)	(0.05)	(0.03)	(0.03)	(0.02)
Less than secondary	0.07**	0.07	0.07	−0.04	−0.09	0.10***	0.04	0.05**
	(0.04)	(0.10)	(0.06)	(0.14)	(0.06)	(0.03)	(0.03)	(0.03)
Secondary	0.10***	0.05	0.13***	0.04	0.04	0.10***	0.07**	0.06***
	(0.03)	(0.10)	(0.04)	(0.13)	(0.05)	(0.03)	(0.03)	(0.02)
Postsecondary	−0.07	−0.06	−0.13	−0.16	−0.24***	−0.04	−0.04	−0.04
	(0.05)	(0.08)	(0.08)	(0.15)	(0.08)	(0.04)	(0.05)	(0.03)
				Inmigration				
All	−0.08	0.03	−0.08	0.26	−0.03	−0.09	−0.07	−0.05
	(0.11)	(0.28)	(0.12)	(0.32)	(0.09)	(0.12)	(0.12)	(0.07)
Less than secondary	−0.04	0.20	−0.05	0.44	−0.01	−0.05	−0.02	−0.02
	(0.12)	(0.29)	(0.12)	(0.30)	(0.10)	(0.12)	(0.13)	(0.08)
Secondary	−0.17	−0.04	−0.19*	0.11	−0.08	−0.18*	−0.16	−0.11*
	(0.11)	(0.30)	(0.10)	(0.36)	(0.08)	(0.11)	(0.11)	(0.06)
Postsecondary	−0.01	−0.06	0.03	0.26	0.07	−0.02	−0.04	−0.00
	(0.14)	(0.30)	(0.15)	(0.41)	(0.14)	(0.14)	(0.13)	(0.09)
Kleibergen-Paap Wald stat.	17.78	8.88	38.19	6.30	61.57	17.79	14.00	19.49
Number of metro areas	79	73	79	79	79	78	27	79

Notes: Sample includes the years 2014 and 2019. Migration rates are taken as a share of the 5-year lag population. All models include Year FE and City FE. Observations are weighted by city–year–group population. See paper for description of robustness checks. Cluster-robust standard errors are in parentheses. 2SLS, two-stage least squares; FE, fixed effect.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A10 Robustness of 2SLS wage effects interacted with regional characteristics

Interaction variable	(1) Original 2SLS	(2) Drop <100 km from border	(3) Control trade with Venezuela	(4) Year trend × inverse distance to border	(5) Year trend × region	(6) Year trend × pre-trend	(7) Drop Bogotá	(8) Drop metro yearly sample <1,000	(9) 2014–2019 difference model	(10) Kronmal specification	(11) Assign natives to past metro
2014 mean ln(hourly wage)	−0.36	−0.69	−0.36	−0.31	−0.20†	−0.36	−0.47	−0.68	−0.48†	−0.63**	−0.35
	(0.46)	(0.60)	(0.46)	(0.45)	(0.42)	(0.46)	(0.49)	(0.62)	(0.50)	(0.29)	(0.40)
2014 unemployment rate	0.07	0.35	0.11	0.58**	−0.29	0.08	0.08	0.04	0.08	0.11	0.07
	(0.29)	(0.82)	(0.31)	(0.25)	(0.33)	(0.29)	(0.30)	(0.33)	(0.33)	(0.18)	(0.26)
2014 informal rate	−0.62*	−0.80**	−0.63*†	−0.56*	−0.76***	−0.60*	−0.62*	−0.64	−0.62†	−0.34	−0.53*
	(0.34)	(0.38)	(0.34)	(0.34)	(0.27)	(0.34)	(0.37)	(0.40)	(0.42)	(0.48)	(0.31)
2014 own-account rate	−0.75**	−1.15**	−0.81**†	−0.73**	−0.86***†	−0.73**	−0.75**	−0.80*	−0.71*†	−0.74*	−0.61*
	(0.35)	(0.45)	(0.38)	(0.33)	(0.33)	(0.35)	(0.37)	(0.42)	(0.43)	(0.40)	(0.33)
WB DB (starting a business)	0.29*†	0.43	0.34*†	0.41*	0.28†	0.29*†	0.28*†	0.27*†	0.26†	0.16†	0.26*†
	(0.16)	(0.35)	(0.18)	(0.24)	(0.25)	(0.16)	(0.16)	(0.15)	(0.16)	(0.11)	(0.14)

(*Continued*)

Table A10 Continued

Interaction variable	(1) Original 2SLS	(2) Drop <100 km from border	(3) Control trade with Venezuela	(4) Year trend × inverse distance to border	(5) Year trend × region	(6) Year trend × pre-trend	(7) Drop Bogotá	(8) Drop metro yearly sample <1,000	(9) 2014–2019 difference model	(10) Kronmal specification	(11) Assign natives to past metro
WB DB (construction permits)	−0.31	−0.31	−0.31	−0.29	−0.08†	−0.28	−0.36	−0.30	−0.35	−0.18	−0.26
	(0.50)	(0.51)	(0.51)	(0.52)	(0.38)	(0.49)	(0.51)	(0.53)	(0.61)	(0.23)	(0.47)
WB DB (registering property)	−0.36	−0.37	−0.38	−0.34	−0.38†	−0.37	−0.50	−0.33	−0.34	−0.20	−0.30
	(0.43)	(0.43)	(0.50)	(0.39)	(0.37)	(0.43)	(0.45)	(0.45)	(0.54)	(0.16)	(0.39)
WB DB (paying taxes)	−0.06	−0.28	−0.05	0.30	−0.49*	−0.05	−0.06	−0.07	−0.11	0.01	−0.05
	(0.23)	(0.53)	(0.25)	(0.25)	(0.27)	(0.24)	(0.23)	(0.24)	(0.27)	(0.12)	(0.22)
2014 per capita GDP	−0.13	−0.15	−0.14†	−0.20	0.09†	−0.11	−0.18	−0.12	−0.18†	−0.14	−0.14
	(0.32)	(0.33)	(0.31)	(0.33)	(0.37)	(0.31)	(0.30)	(0.33)	(0.35	(0.15)	(0.30)
N	474	438	474	474	474	474	468	162	79	474	474

Notes: Each row presents the interaction coefficient from a separate regression using the indicated interaction variable. †Indicates that the K-P Wald stat. is >16. See notes to Table 7. Cluster-robust standard errors are in parentheses. 2SLS, two-stage least squares; GDP, gross domestic product; K-P Wald stat., Kleibergen-Paap Wald statistic; WB DB, World Bank Doing Business.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

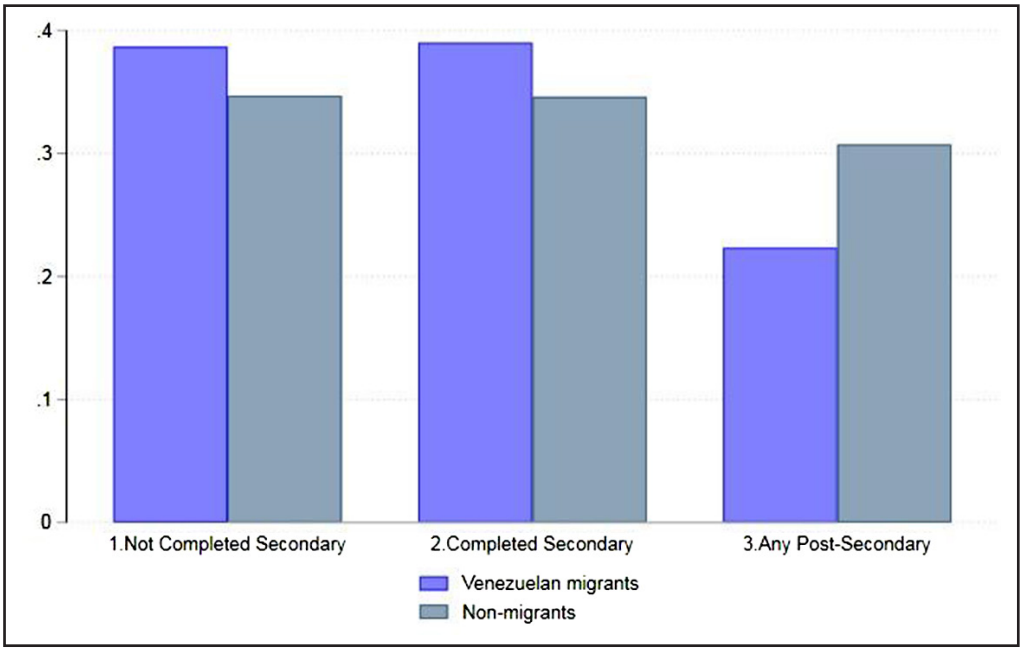
Table A11 School attendance 2SLS linear probability models

	In school	In school	In school
		In labor force	Out of labor force
	All ages (including border locations)		
Migrant share	0.07**	−0.03	0.11*
	(0.03)	(0.08)	(0.06)
Sample mean	18.76	7.14	11.62
	Age ≤24 years (including border locations)		
Migrant share	0.20*	−0.19	0.39*
	(0.10)	(0.24)	(0.23)
Sample mean	54.60	14.43	40.17
	All ages (excluding border locations)		
Migrant share	0.13	0.20*	−0.07
	(0.11)	(0.10)	(0.09)
Sample mean	18.82	7.17	11.65
	Age ≤24 years (excluding border locations)		
Migrant share	0.28	0.49*	−0.21
	(0.34)	(0.27)	(0.33)
Sample mean	54.69	14.45	40.24

Notes: All outcomes are multiplied by 100. All models include Year FE and City FE. Observations are weighted by city–year–group population. Excluding border locations drops cities within 100 km driving distance of the Venezuelan border. Cluster-robust standard errors are in parentheses. 2SLS, two-stage least squares; FE, fixed effect.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

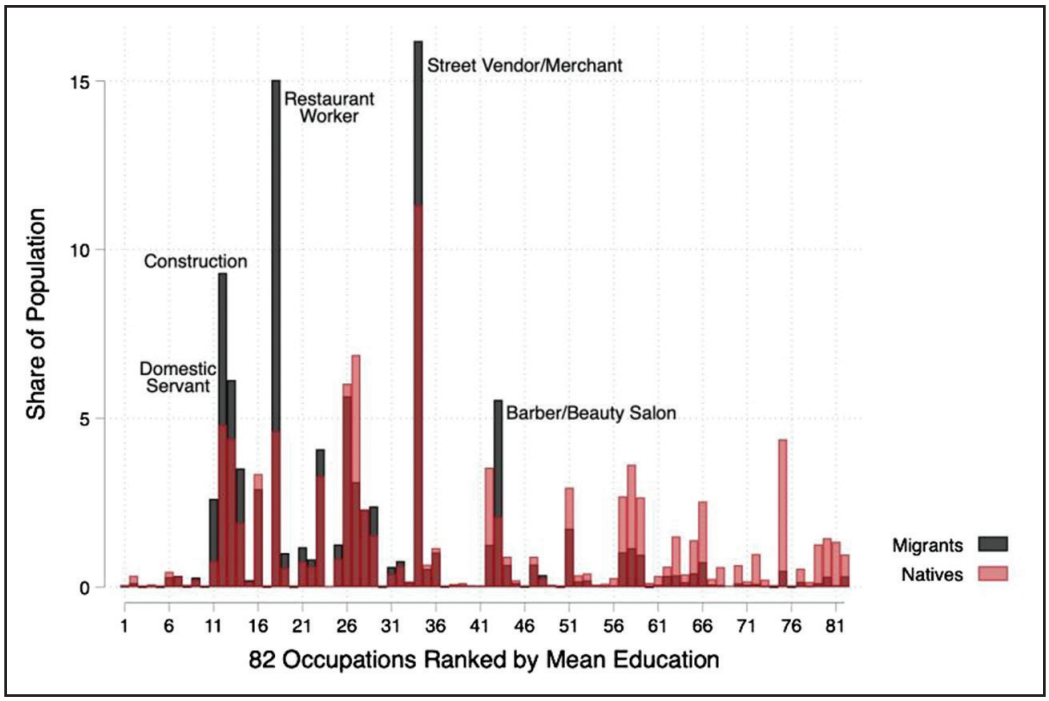
Figure A1 Education by migrant status.



Notes: Data restricted to urban residents of age 15–64 years. GEIH, Colombian National Integrated Household Survey.

Source: GEIH (2019).

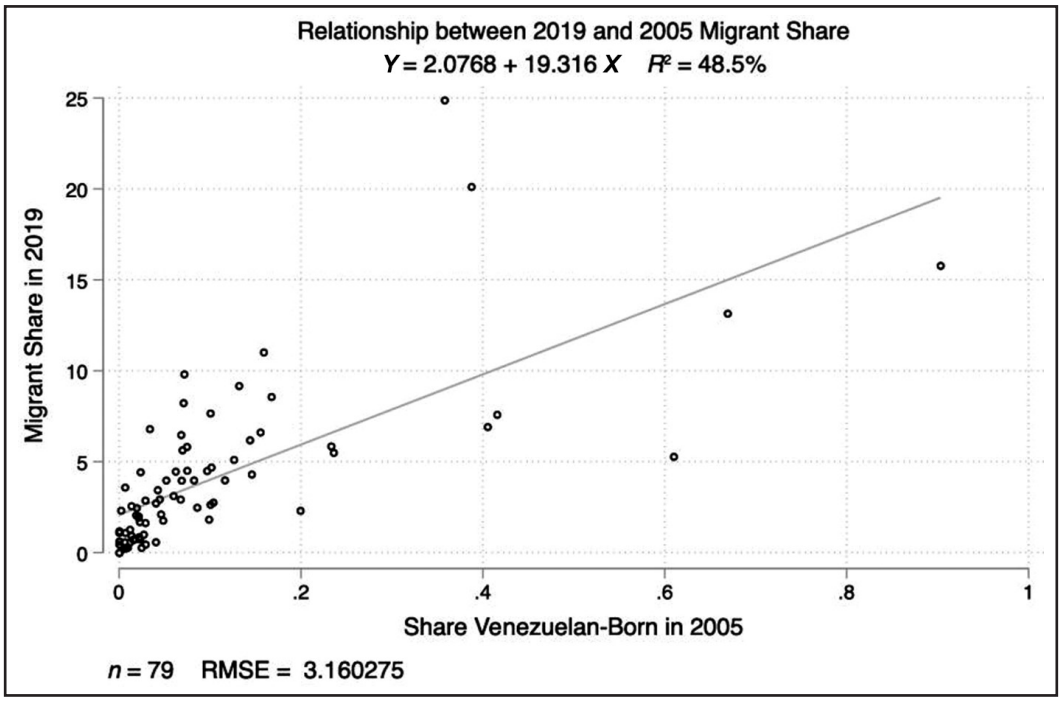
Figure A2 Concentration of migrants and natives across occupations.



Notes: Data restricted to urban residents of age 15–64 years. Occupations ranked according to mean years of completed schooling for natives in the GEIH between 2010 and 2015. GEIH, Colombian National Integrated Household Survey.

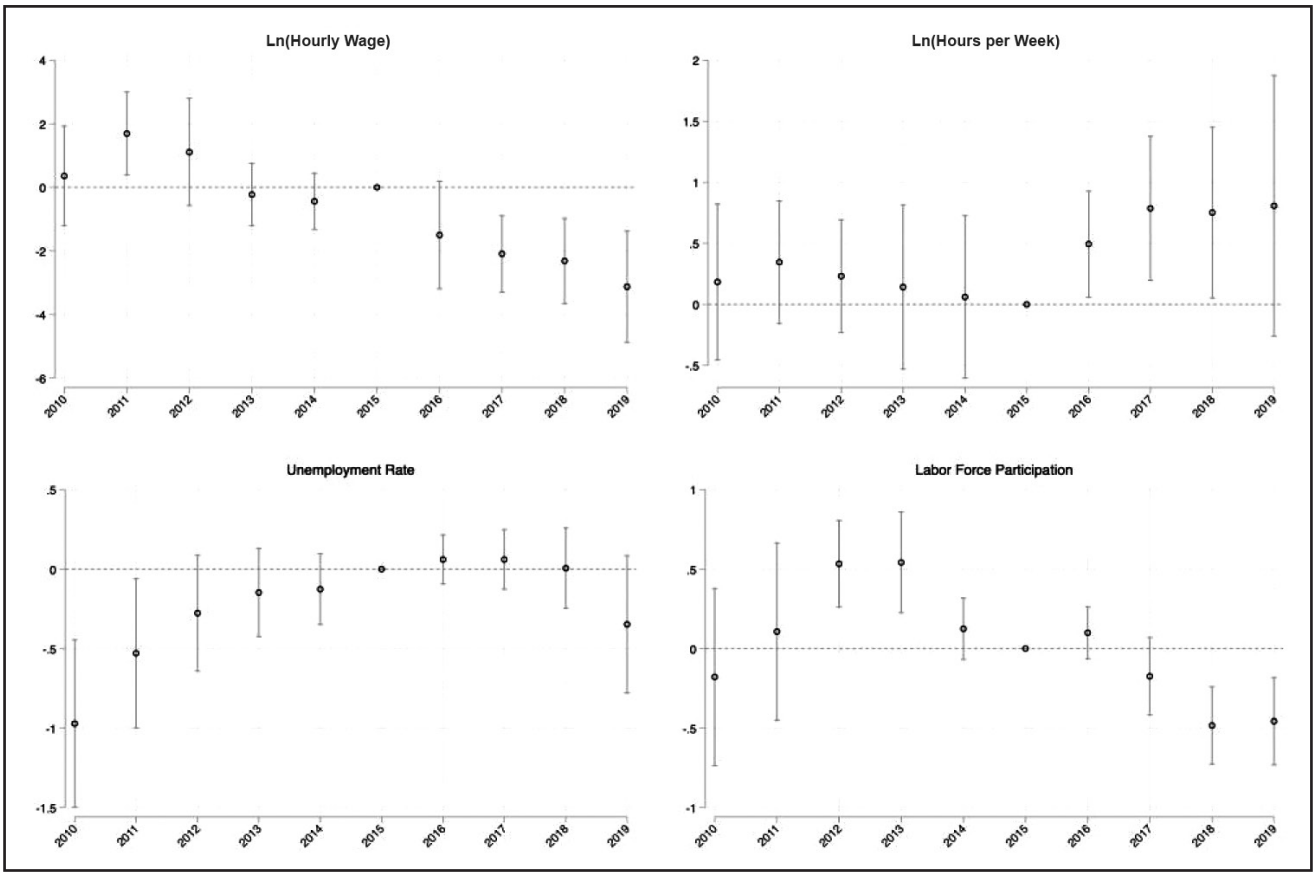
Source: GEIH (2019).

Figure A3 Relationship between 2019 and 2005 migrant shares.



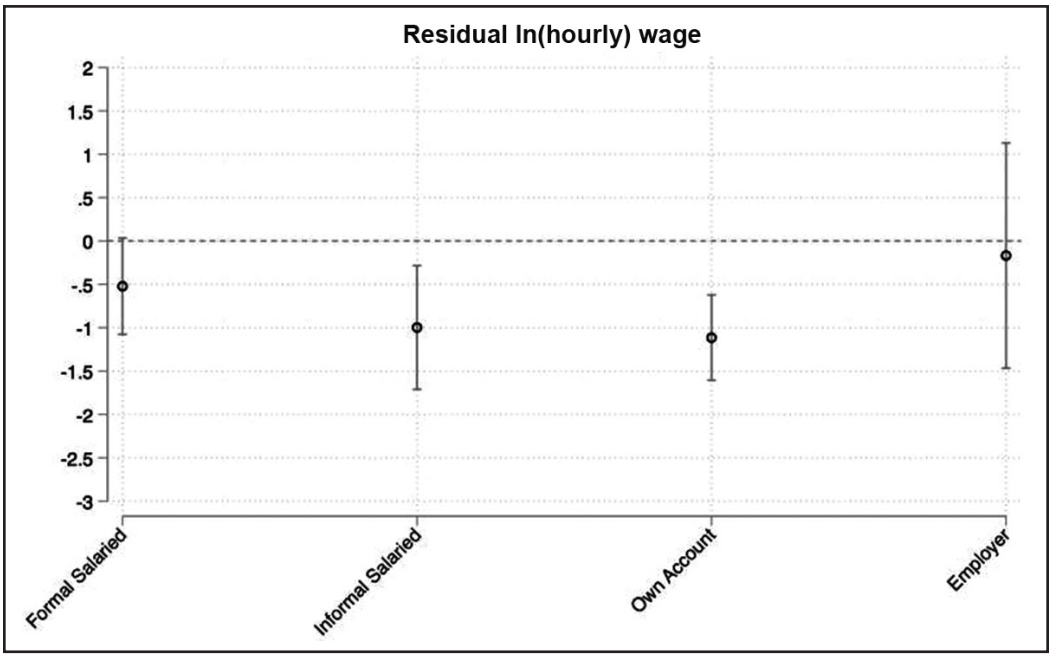
Note: RMSE, root mean square error.

Figure A4 Pre-trends in labor market outcomes.



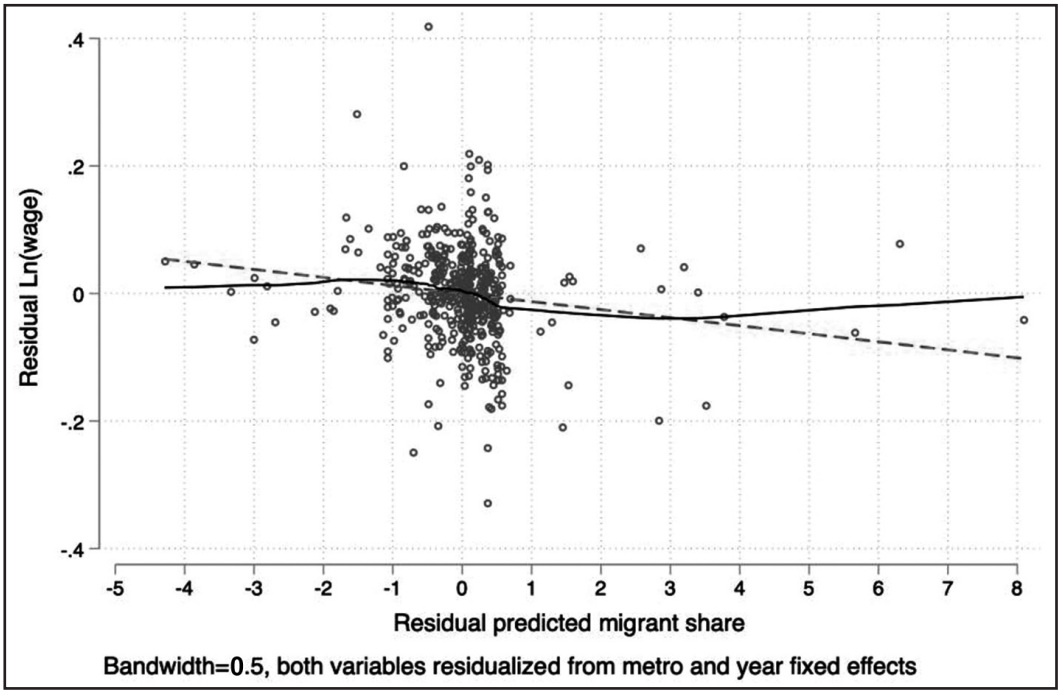
Note: Pre-trends estimated according to Eq. (5) in Section 5. They are 95% confidence intervals.

Figure A5 2SLS coefficient plot – by work sector.



Note: 2SLS, two-stage least squares. They are 95% confidence intervals.

Figure A6 LOWESS plot.



Note: LOWESS, locally weighted scatterplot smoothing. The red dotted line displays the least square line of best fit.