Formal Effects of Informal Labor and Work Permits Evidence from Venezuelan Refugees in Colombia

Dany Bahar Isabel Di Tella Ahmet Gulek*

November 1, 2024

Most recent draft here.

Abstract

The Venezuelan refugee crisis in Colombia is a one-of-its-kind setting to study the trade-offs of providing work permits to refugees. In 2015 refugees began arriving without work permits. Then, the government granted work permits to existing refugees in several waves between 2017-2018. We analyze each of these two shocks in turn. First, using a synthetic shift-share design a la Gulek and Vives-i Bastida (2023), we find that the mostly informal labor supply of refugees displaced formal and informally employed natives in salaried jobs. This suggests high substitutability between informal and formal labor in production. Second, using a triple and a quadruple difference in differences design exploiting variation across regions, time, and skill group exposure, we find that work permits allow high-skill refugees to find formal salaried jobs and work closer to their skill level, reducing the skill mismatch in the economy. This comes at a cost to some natives, who lose their formal jobs, and at a benefit to others, whose wages rise.

JEL Classification: D22, J21, J46.

Keywords: Immigration, Refugee crises, Work permits, Informality

*Bahar: Brown University and Harvard Growth Lab (email: dany_bahar@brown.edu); Di Tella: PhD student, MIT Department of Economics (e-mail: iditella@mit.edu); Gulek: PhD student, MIT Department of Economics (e-mail: agulek@mit.edu)

1 Introduction

Over the last decade, the number of people forcibly displaced across international borders has nearly tripled, from approximately 16 million in 2012 to 46 million in 2022 (UNHCR, 2022).[1] Unlike many other migrants, refugees usually arrive without work permits and are mostly hosted by developing countries with sizeable informal sectors. Host countries can provide work permits, which allows refugees to formally participate in the labor force (Clemens et al., 2018). However, there is limited empirical evidence of the effects of these policies. This paper aims to close that gap.

The context of our study is Colombia, a country that by 2023 hosted 2.5 million forcibly displaced Venezuelans, one of the largest refugee-like populations in the world. Venezuelans began to arrive in 2015, mostly without permission to work. Over the course of several waves, starting in 2017, Colombia provided Venezuelans with a migratory status that included work permits. By doing this, Colombia became the only developing country to receive large numbers of refugees and later grant them work permits en masse.[2] The provision of these work permits provide quasi-experimental variation that allows us to identify the labor market effects of both (1) an informal labor supply shock and (2) the subsequent provision of work permits.

We start by adapting the canonical labor demand framework to lay out several testable predictions for these two shocks which we later validate empirically. In the model, a representative firm can use both informal and formal labor in production. We start by analyzing the effect of an increase in the informal labor supply. The first prediction is that it unambiguously reduces native informal employment due to greater competition in the informal sector. The second prediction is that the effect on native's formal employment is ambiguous. On the one hand, more informal workers could increase the productivity of formal workers due to Q-complementarity; on the other hand, they increase competition if returns to labor are diminishing and there is a high degree of subsitutability between them. Thus, the effect on overall formal native employment is an empirical question. Then, we turn to the work permit shock. The third and fourth predictions are that this policy leads some natives to gain informal jobs while others lose formal jobs because of refugee reallocation from the informal to the formal sector. However, this reallocation leads to a wage increase in the informal sector so that firms demand more formal workers. The fifth prediction is that this process results in an increase in the economy's total formal employment.

[1]This includes refugees, asylum seekers and other people in need of international protection. We will refer to them as refugees for the remainder of the paper.

[2]In contrast, Syrian refugees in Turkey do not have permits to this day (Gulek, 2023), and Ukranian refugees received permits almost immediately in all EU countries (European Commission, n.d.).

Next, we empirically test the model's implications. First we analyze the impact of refugee arrival on natives' employment. Although most studies in this space focus on overall employment, we distinguish between salaried and nonsalaried work where the latter is mostly self-employment. The reason is that salaried work is partially determined by firm demand while non-salaried work depends only on worker's preferences. For salaried workers, we find that Venezuelan arrival displaced both informal and formal native workers.[3] The former is consistent with our first prediction, while the latter suggests that informal and formal labor are highly subsitutable in production as it relates to our second prediction. At the same time, we find that the immigration shock *increases* non-salaried employment, especially for men without high-school degrees. This result suggests that non-salaried jobs can act as a buffer to unemployment. The distinction between salaried and non-salaried workers is crucial. If we pooled them into one category we would find null effects in the aggregate.

Identification in this setting comes from a shift-share design, where travel distance between Colombian and Venezuelan cities operates as a share and the total number of refugees in Colombia in a given year acts as a shift. To deal with the fact that regions close to the border could follow different economic trajectories from other regions –and indeed data on pre-trends suggest that is the case– we employ the Synthetic Instrumental Variables method, henceforth SIV, developed in Gulek and Vives-i Bastida (2023). This method uses synthetic controls to address unmeasured confounding in instrumental variable difference in differences (IV-DiD) settings similar to ours, and helps us relax the share exogeneity assumption embedded in our design (Goldsmith-Pinkham et al., 2020)).

To test the model's second set of predictions, we study the effect of awarding work permits to irregular migrants on native's labor market outcomes. Using a triple differences in differences design that exploits the region, timing, and bite of work permits, we validate our model's third and fourth predictions: permits cause natives to gain informal jobs due to lower competition in the informal sector and to lose formal jobs due to greater competition in the formal sector. Here, identification relies on the observation that permits are more effective in helping highly educated refugees find formal jobs compared to those with less education. Due to data limitations we are not able to empirically test the fifth prediction. However, having validated four predictions already, we use the model to quantify the fifth prediction. We estimate that all the work permits issued during our study's time period allowed 68,000 Venezuelans to transition from informal to formal jobs, creating a net increase

[3] In terms of magnitudes, a 1 percentage point (pp) increase in the refugee/native ratio decreases the native informal salaried employment rate by 0.26 pp, and decreases the formal salaried employment rate by 0.16 pp. These estimates represent, based on average values at baseline, about a 3% drop in informal salaried employment (baseline of 9%) and less than a 1% drop in formal salaried employment (baseline of 18%).

of 24,440 formal jobs in the economy. Under the conservative assumption that all these workers received the minimum wage, we estimate that the total tax revenue from granting work permits was around 22.5 million USD annually.

Lastly, we further investigate whether work permits can facilitate a better match between migrants and employers, therefore creating overall productivity gains in the economy. We present suggestive evidence that work permits lead to wage increases for more exposed natives in more formal/skill-intensive industries like Finance relative to less formal/skill-intensive industries like Construction and Transportation. Our quadruple DiD design shows that this effect is unlikely to be driven by industry, skill-sex, region or time specific trends, or any triple combination of such trends. These results are consistent with relatively higher skilled refugees leaving low-skill industries and moving to high-skill industries. It is likely that these reallocation dynamics generate productivity gains for the economy as a whole.

This paper makes a contribution to the economic literature by comprehensively reexamining labor market responses to immigration shocks. Examples of such episodes include the Mariel Boatlift (Card, 1990), the Algerian war of independence (Hunt, 1992), the Yugoslav wars (Angrist and Kugler, 2003), and the Syrian refugee crisis (Del Carpio and Wagner, 2015). These episodes often entail two separate shocks: the arrival of the immigrants and subsequent government policies that often result in either granting or withholding work permits. Ours is the first attempt at estimating separate effects for these two events, which is one of the main contributions of our paper. Researchers that combine immigration shocks and their subsequent policies, could be under- or overestimating policy effects since they are being confounded with the effects of the shock, and viceversa. For example, in our context we show that whereas the informal immigration shock displaces natives in both the informal and formal sectors, granting work permits to immigrants increases natives' prospects in the informal sector and further increases the overall amount of formal jobs in the economy.[4]

Second, the quasi-experimental methods we use complement the economic literature at large by studying the dynamics of informal and formal labor markets given an exogenous shock to labor supply. Initial contributions in this field were largely theoretical (Rauch, 1991; Amaral and Quintin, 2006), while more recent efforts have focused on calibrating/estimating structural models (Bosch and Esteban-Pretel, 2012; Meghir et al., 2015; Ulyssea, 2018). One exception is Gulek (2023), who studies the labor market consequences of the Syrian

[4] In the context of Venezuelan refugees in Colombia a number of studies show inconclusive results on employment due to the immigration shock (Santamaria, 2020; Caruso et al., 2021; Penaloza-Pacheco, 2022; Lebow, 2022; Delgado-Prieto, 2022; García-Suaza et al., 2024) and only one paper studies the effects of granting work permits Bahar et al. (2021) finding negative but negligible effects on the formal employment rate of Colombian workers. Our paper claims that the reason for the inconsistency of results across this literature is that these papers do not distinguish between immigration flow and the granting of work permits that happened almost immediately, as we do.

refugee shock in Turkey. He also finds that the arrival of informal refugees causes native disemployment in informal and formal salaried jobs, and low-skill native men transition to non-salaried jobs. However, this paper goes beyond that since the Colombian setting allows us to empirically test for the effect of work permits on natives. We go another step further and present strongly suggestive evidence that work permits reduce the mismatch in the economy, resulting in further productivity gains overall.

Third, our paper also presents a methodological contribution to the literature on immigration by implementing SIV. Shift-share designs exploiting travel distance or past settlement instruments have long been used in the literature on immigration since Card's seminal work (Card, 2009). Identification in these designs often come from the exogeneity of shares (Goldsmith-Pinkham et al., 2020). However, we show that the travel distance, which is heavily used by the literature on Colombia (Delgado-Prieto, 2022), is correlated with unobserved confounders and hence likely to fail the exogeneity assumption. SIV is an appropriate tool for researchers when pre-trends exist in shift-share designs (Gulek and Vives-i Bastida, 2023). Implementing it sets us apart from the literature.

The rest of this paper is structured as follows. First, Section 2 provides the background on Colombian labor markets and the Venezuelan refugee crisis. Then Section 3 introduces the economic framework and lays out its theoretical predictions. Sections 4 and 5 empirically evaluate these predictions as they relate to the immigration shock and the work permit shock respectively. Section 6 employs the model to answer the questions where we fall short with the data, and finally Section 7 presents the empirical results on mismatch reduction. Section 8 concludes.

2 Background and Data

2.1 The Venezuelan Refugee Crisis and the Colombian Policy Response

Venezuela began to experience out-migration in the early 2000's in response to President Chavez's political and economic reforms. Following his death in 2013, the economic and political crisis deepened. The first waves of refugees arrived in Colombia in late 2015, but their numbers remained small until mid-2016. As the economic crisis intensified in the following months, there was a substantial increase in Venezuelans seeking refuge in Colombia. Figure 1a shows how the number of Venezuelan refugees in Colombia has evolved over time. In 2022, Venezuelans made up approximately 5% of the Colombian population.

Figure 1b shows the distribution of the number of Venezuelans per 100 natives in Colombia at the department level, our unit of analysis, in 2019. As is standard in refugee crises, Venezuelans are more densely located in regions closer to the border (Caruso et al., 2021). Distance to the populous governorates in Venezuela strongly predicts the number of refugees per native in a given region, which constitutes one of the key building blocks of our identification strategy.

In light of the large inflows of immigrants from Venezuela, the Colombian government created a temporary visa named *Permiso Especial de Permanencia* (PEP) that allowed Venezuelans to obtain formal employment and benefit from social security. We provide a longer description in Appendix B but include a brief summary here for context. PEP was granted in waves. The first two waves—the first in July 2017 and the second in February 2018—were targeted to Venezuelans who had proper documentation and a migratory status recognized by the law. Consequently, these first two waves were largely selected, with access limited to the highly skilled migrants. Under the first two waves of the PEP program, nearly 182,000 permits were issued.

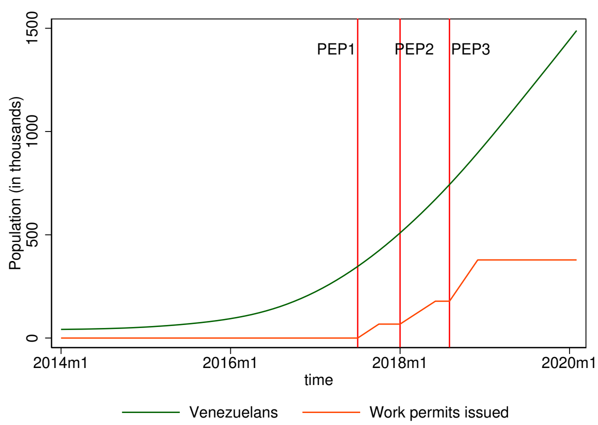
The third and the largest wave of PEP was a byproduct of a nationwide registration of irregular migrants. As the number of undocumented, hence PEP-1 & 2 ineligible, Venezuelan immigrants was increasing, the Colombian government implemented a nation-wide census known as RAMV between April and June of 2018 to count and characterize the undocumented Venezuelan population in Colombia. Importantly, registering was not explicitly or implicitly linked to the possibility of obtaining formal migratory status. RAMV identified 442,462 undocumented Venezuelan migrants in Colombia, belonging to 253,575 different households.

On July 25, 2018, just days before his term ended, President Juan Manuel Santos unexpectedly issued a decree allowing all undocumented Venezuelans listed in the RAMV to apply for a new round of PEP. Sixty-four percent, or around 280,000, of the total undocumented migrants registered in the RAMV received a PEP.

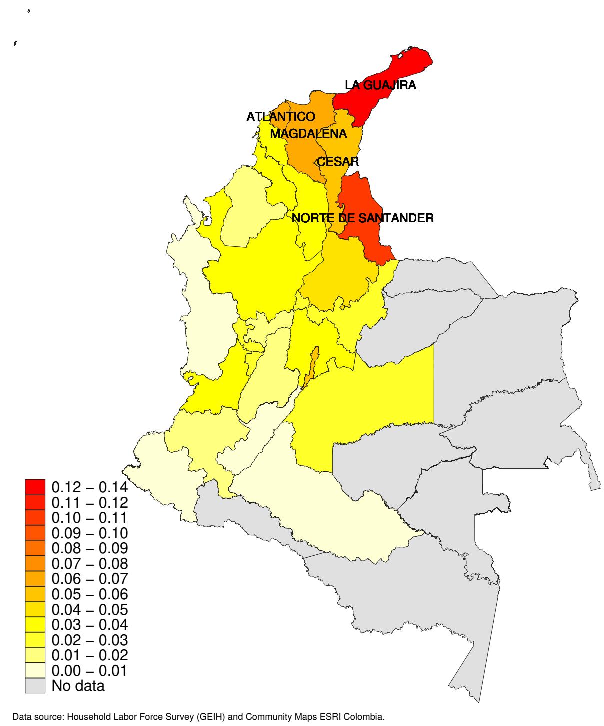
The fact that PEP-1 & 2 required formal documentation whereas PEP-3 did not resulted in differently skilled Venezuelans finding formal jobs on different timelines. Figure 1c shows the number of formally employed Venezuelans for six education-sex cells. There are three skill groups: less than high-school degree (low-skill), at least a high-school degree (middle-skill), and at least a college degree (high-skill); and each education cell is further divided into men and women. Figure 1c shows that high-skilled men began obtaining formal jobs in the third quarter of 2017, during PEP-1 and PEP-2, while the middle-skill men gained formal jobs only after PEP-3. Low-skill men, and women of all skills did not observe consistent and significant increases in formal job uptake.

To sum up, the Venezuelan refugee crisis in Colombia was not entirely an informal labor supply shock. It was an informal labor supply shock early on, but the granting of work permits enabled some of the highest skilled Venezuelans to eventually find formal jobs.

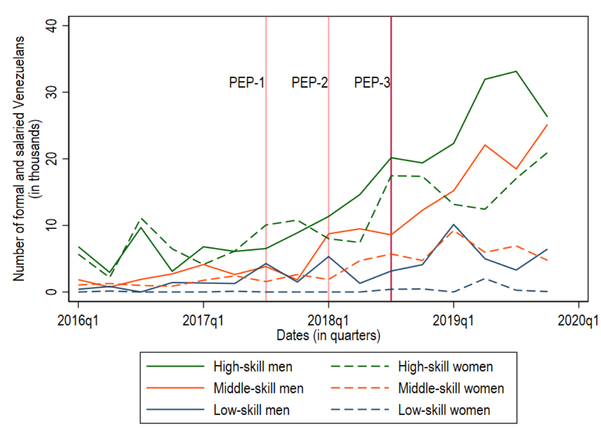
Figure 1: The Venezuelan Refugee Crisis in Colombia



(a) Immigration and Work permits



(b) Cross-sectional distribution of the Venezuelan to native ratio



(c) Formal salaried employment of Venezuelans

2.2 Labor market statistics

Information about the informal and formal labor market outcomes of Venezuelan and native workers comes from the 2007–2019 Colombian Household Labor Force Surveys (GEIH in Spanish) conducted by the Colombian Statistical Institute (DANE in spanish).[5] GEIH surveys around 20,000 households per month and is representative at the department level, of which there are 24. Since 2013, GEIH also contains questions about respondent's citizenship and place of residency one and five years prior. Throughout this paper we use the term natives to refer to all non-Venezuelans unless noted otherwise.

Key Outcomes.– We create four broad employment categories: salaried informal and formal, and non-salaried informal and formal workers. *Salaried workers* are those employed by private companies receiving a regular wage. *Non-salaried workers* are everyone else, mainly self employed people. We omit government jobs for simplicity. This is an economically meaningful separation of jobs into those that depend largely on the labor demand of firms and those that do not. It matters for our interpretation of the impact of irregular migration. For instance, a native who loses their formal salaried job and transitions to unpaid family work or self-employment will still appear as "employed" in the GEIH. Consequently, focusing on the overall native employment rate misses important and meaningful compositional changes in the labor force. To address this problem, we study salaried and non-salaried employment separately, and focus on salaried employment as the key outcome of interest. The employment statistics for different types of natives can be found in Appendix Table A.1.

We distinguish between *formal* and *informal* employment based on respondent's self report of social security coverage in GEIH. By law, Colombian employers must contribute to social security on their workers behalf.[6] This measure is a good predictor of formality for two reasons. First, there is no incentive for workers to misreport their insurance status. It is not illegal to work informally, only to employ informally. Second, the descriptive statistics on formal and informal employment using insurance status are consistent with the general knowledge on the informal sector (Ulyssea, 2020). Across regions and industries, the informality rate (defined as the ratio of employment that is informal) decreases with education. It is higher in less developed regions and in industries like agriculture, which are known to rely on informal labor.

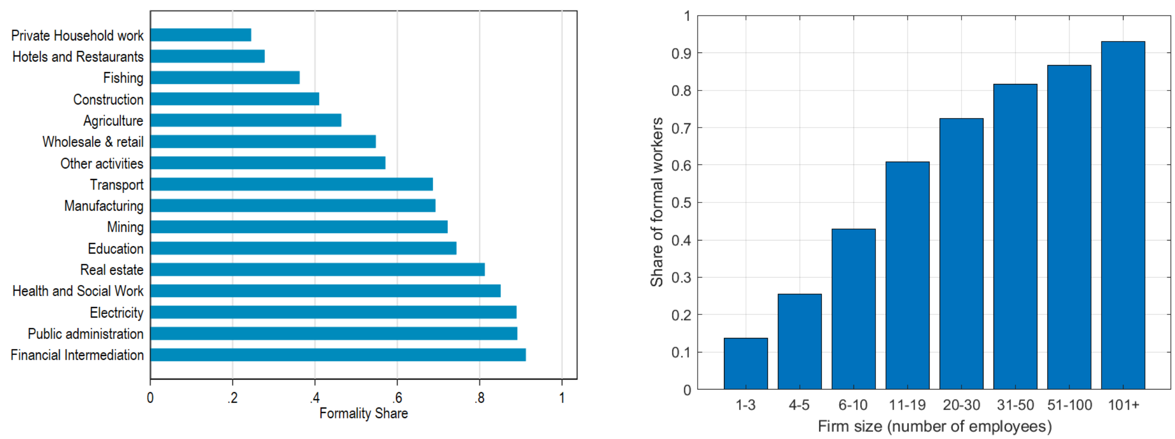
Summary Statistics.– Figure 2 shows the formality rate of salaried workers across select industries and firm sizes. Although there is important heterogeneity, formal and informal workers coexist everywhere. Panel A shows the formality rate across sectors, ranging from 24% in private household work to 91% in financial intermediation. Panel B shows that the

[5] This dataset is available on DANE's website.

[6] Law 100 of 1993

formality rate increases as firms get bigger: ranging from 6% in firms with 1 employee, to 91% in firms with more than 101 employees. This relationship can be rationalized by the fact that larger firms are more visible and therefore more likely to be punished if hiring informally (Ulyssea, 2020).

Figure 2: Ratio of informal workers across industries and firm size



(a) Industries

(b) Firm size

3 Model

The aim of this section is to formalize the economic forces by which (1) an informal labor supply increase, and (2) the granting of work permits can impact natives' labor market outcomes and make testable predictions. We focus on salaried jobs in this section, and present a stylized model of non-salaried jobs in the Appendix Section F. Here, we employ the canonical labor demand framework with a representative firm that can use both informal and formal labor in production. Given the widespread coexistence of informal and formal workers shown in Section 2.2, the representative firm assumption is a benign simplification.

Labor Demand: Following Ulyssea (2018), we assume that the firm pays a different cost to hire formal or informal salaried workers. To hire a formal worker, the firm must pay a payroll tax τ_w. To hire an informal worker, it pays an increasing and convex expected cost $\tau(\cdot)$. This cost structure predicts that the probability of being informally employed decreases with firm size, which is consistent with the Colombian data. It can be rationalized by the fact that larger firms are more likely to be discovered and punished for hiring informally (De Paula and Scheinkman, 2011). Additionally, there are sector-specific wages w_i and w_f for informal and formal workers respectively.

The firm chooses the number of informal workers, l_i, and the number of formal workers, l_f, to maximise operating profits:

$$\max_{l_i, l_f} F(l_i, l_f) - \tau(l_i)w_i - (1+\tau_w)w_f l_f \tag{1}$$

It takes wages as given and produces a homogeneous good whose price is normalized to one. The system of equations that determines labor demand for each type of worker can then be derived from the first order conditions of equation 1.

Labor Supply: For simplicity, native labor supply for salaried jobs is given exogenously by $L_i^S(w_i)$ for the informal sector and $L_f^S(w_f)$ for the formal sector. We assume that each sector's labor supply is independent of the other sector's wage. This could be justified, for example, if natives were endowed with either informal or formal labor, and therefore could only work in one sector.[7]

3.1 The effect of an informal labor supply shock

Refugees without work permits can only provide informal labor. Figure 3a illustrates how we model the refugee shock as an outward shift in the informal labor supply curve $L_i^S(w_i)$. The effect of this shock on natives' labor market outcomes depends on the labor demand elasticities in both sectors.

To make progress, we impose a CES production function $F(l_i, l_f) = (\eta l_i^\rho + (1-\eta)l_f^p)^{\alpha/\rho}$ as in Gulek (2023). Relative productivity is given by the parameter η, the elasticity of substitution is given by $\sigma = \frac{1}{1-\rho}$, and the returns to scale in labor is governed by α.

Given the CES form, the labor demand elasticities w.r.t. the informal wage are given by:

$$\begin{aligned}
\epsilon_{L_i,w_i} &= -\frac{1-\rho-(\alpha-\rho)s_f}{(1-\rho+\gamma)(1-\rho)-(\alpha-\rho)[(1-\rho+\gamma)s_f+(1-\rho)s_i]} \\
\epsilon_{L_f,w_i} &= -\frac{(\alpha-\rho)s_i}{(1-\rho+\gamma)(1-\rho)-(\alpha-\rho)[(1-\rho+\gamma)s_f+(1-\rho)s_i]]}
\end{aligned} \tag{2}$$

where $s_i = \frac{\eta L_i^\rho}{\eta L_i^\rho + (1-\eta)L_f^\rho}$ is the informal share in production, and vice versa for s_f.[8]

Equation 2 encapsulates two straightforward findings, summarized in equation 3. First, $\epsilon_{L_i,w_i} < 0$ for all possible parameter values. This means that lower informal wages lead to higher informal labor demand, i.e., the informal labor demand curve is downward sloping. Second, the sign of ϵ_{L_f,w_i} depends on the sign of $\alpha - \rho$. Specifically, when ρ is greater than

[7] For a model where natives can move between formal and informal jobs, see Meghir et al. (2015).

[8] We refer the interested reader to Gulek (2023) for more details and full derivations.

α, the elasticity of formal labor demand with respect to informal wages turns positive. This implies that lower informal wages lowers formal labor demand.

$$\begin{aligned}
&\epsilon_{L_i,w_i} < 0 \\
&\epsilon_{L_f,w_i} \begin{cases} > 0 & \text{if } \alpha < \rho \\ < 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{3}$$

To understand the intuition behind this result, consider the change in the marginal productivity of a formal worker when an informal worker is hired:

$$\frac{\partial(log\frac{\partial F}{\partial L_f})}{\partial L_i} = (\alpha - \rho)L_i s_i$$

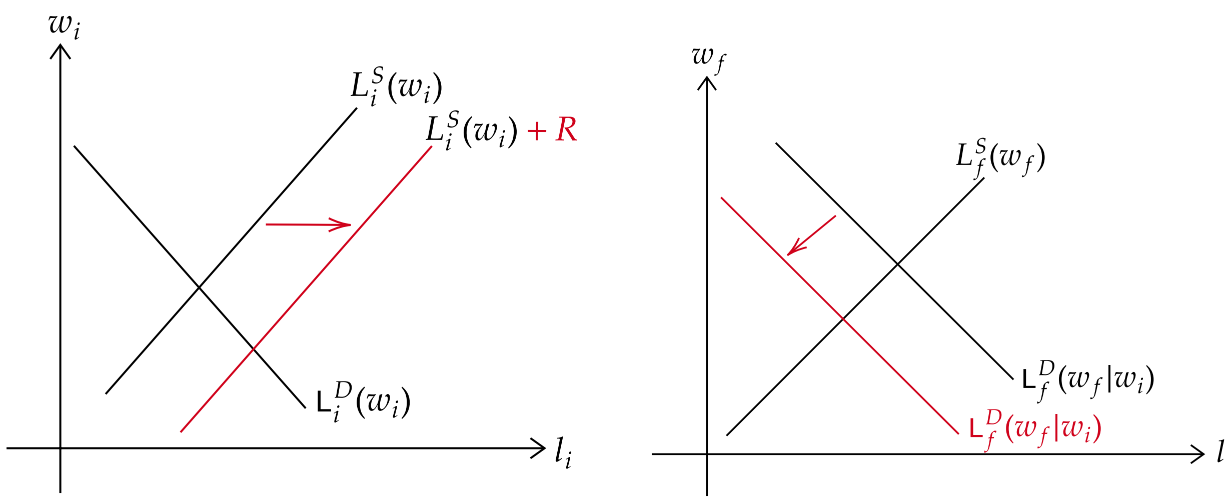
In a scenario where $\alpha > \rho$ such as when the production function exhibits constant returns to scale (CRTS, $\alpha = 1$) and formal and informal workers are not perfect substitutes ($\rho < 1$), hiring an informal worker raises the productivity of formal workers due to Q-complementarity between the two types of labor. This complementarity increases the demand for formal labor, resulting in a negative elasticity of formal labor demand with respect to informal wages $\epsilon_{L_f,w_i} < 0$. Nevertheless, as α decreases, the addition of any worker leads to productivity declines for existing workers due to diminishing returns. When α falls below ρ, the productivity loss caused by technological limits (such as fixed capital in the short term) exceeds the productivity gains from Q-complementarity. Thus, a decrease in informal wages, triggered by a surge in the supply of informal labor, could prompt firms to displace formal workers.

Figure 3 illustrates what happens when $\alpha < \rho$. The (assumed to be inelastic for display purposes) refugee labor supply shock shifts the total informal labor supply curve from $L_i^S(w_i)$ to $L_i^S(w_i)+R$. Panel A shows the response in the informal sector. Informal wages necessarily lower, except for corner cases such as a perfectly elastic native labor supply curve. Panel B shows the response in the formal sector. The informal wage decline has the second-order effect of shifting the formal labor demand curve leftwards from $L_f^D(w_f|w_i)$ to $L_f^D(w_f|w_i')$. Firms replace their formal workers with informal ones.

3.2 The effect of granting work permits to refugees

In this framework, granting work permits has two effects. They are illustrated in figure 4 for the case of $\alpha < \rho$. First, refugees endowed with permits leave the informal sector and work in the formal sector to benefit from higher formal sector wages. This shifts the total informal labor supply curve to the left from $L_i^S(w_i) + R$ to $L_i^S(w_i) + R^i$ in Panel A. It shifts

Figure 3: Informal LS shock when $\alpha < \rho$



(a) Informal

(b) Formal

the total formal labor supply curve to the right from $L_f^S(w_f)$ to $L_f^S(w_f) + R^f$ in Panel B. As refugees leave the informal sector, natives gain informal jobs due to lower competition. As refugees enter the formal sector, natives lose formal jobs due to greater competition. Second, higher informal wage have the second-order effect of causing firms to increase their formal labor demand from $L_f^D(w_f|w_i')$ to $L_f^D(w_f|w_i'')$. This is because formal and informal labor are highly substitutable, that is $\alpha < \rho$. Natives gain some formal jobs but not enough to overturn the first order effect. In total, natives gain informal jobs and lose formal jobs, and the total number of formal jobs in the economy increases.

Predictions of the Model

Using the canonical labor demand framework to model informal and formal salaried jobs, we make five predictions relating to the effects of an informal immigration shock and the granting of work permits.

In response to an informal labor supply shock:

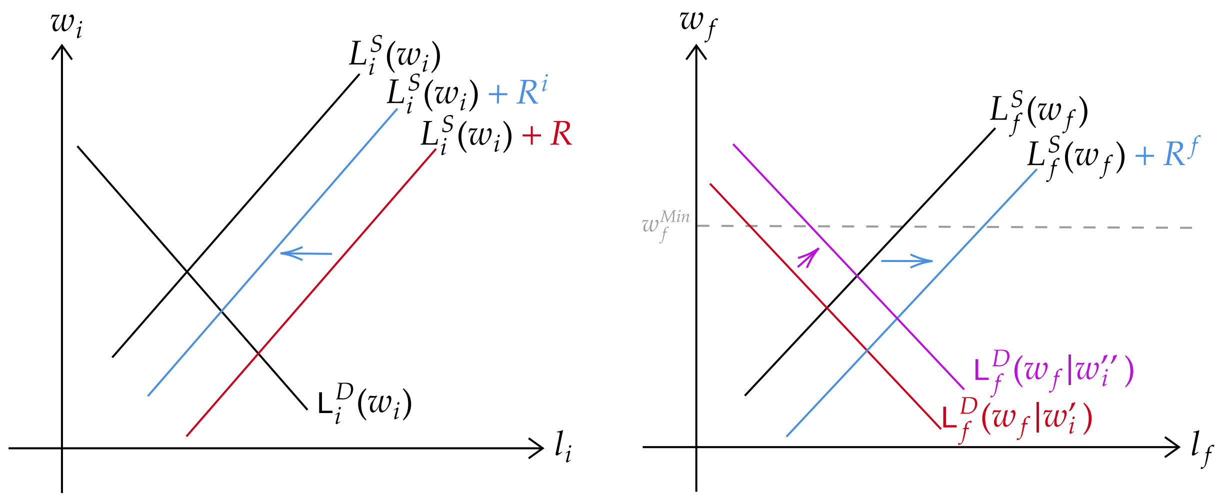
(1) natives lose informal salaried jobs

(2) natives lose (gain) formal salaried jobs if informal and formal labor are substitutes (complements)

In response to providing migrants with work permits:

(3) natives gain informal salaried jobs

Figure 4: Effects of Work Permits when $\alpha < \rho$



(a) Informal (b) Formal

(4) natives lose formal salaried jobs

(5) total formal jobs in the economy increase if informal and formal labor are substitutes

In the rest of the paper we refer back to these predictions while explaining our reduced-form results on salaried employment. Explaining our results on non-salaried employment requires microfounding the labor supply decision of natives, which we do in Appendix Section F. We omit this extension from the main text for brevity. In Section 4 we empirically validate the model's predictions relating to (1) and (2), and in Section 5 we validate the model's predictions relating to (3) and (4). Due to data limitations we cannot validate prediction (5) directly. Instead, we estimate the model to quantify the mechanism highlighted by prediction (5) in Section 6.

4 The effects of Immigration

In this section, we estimate the impact of Venezuelan migration on the labor market outcomes of Colombians. For identification we rely on a shift-share design where the share is a weighted "distance to border" exposure variable, and the shift is the number of Venezuelans in a given year. Event study estimates reveal pre-trends in our data. This means our distance exposure measure is correlated with local labor market trends. Therefore, the share exogeneity assumption embedded in our design is likely invalid (Goldsmith-Pinkham et al., 2020). To make progress, we employ the Synthetic IV (SIV) methodology (Gulek and

Vives-i Bastida, 2023). SIV relaxes the share-exogeneity assumption by partialing out the unmeasured confounders (i.e., the pre-trends) by using synthetic controls in IV settings. We provide details of our estimation technique below.

4.1 Setting up the distance instrument

Equation 4 describes how we construct the distance exposure Z_r.[9] For each Colombian department r we calculate the weighted average of the distance $d_{r,s}$ between it and each of the 23 Venezuelan governorates s.[10] The weight π_s is the population share of the Venezuelan governorate s.

$$Z_r = \sum_{s=1}^{23} \pi_s \frac{1}{d_{r,s}} \tag{4}$$

We standardize the distance exposure Z_r to have mean zero and standard deviation of one so that we obtain economically meaningful coefficients when estimating the reduced-form.

To measure whether the distance exposure is a good predictor of the treatment, the Venezuelan to native ratio $R_{r,t}$, we estimate an event study model of the form:

$$R_{r,t} = \sum_{j \neq 2015} \lambda_j (\text{year}_j \times Z_r) + f_r + f_t + \epsilon_{r,t} \tag{5}$$

where the distance exposure Z_r is interacted with year dummies year_j, and f_r and f_t are region and year fixed effects. Figure 5 plots the coefficients λ_j from equation 5. It shows that the instrument is a strong predictor of the treatment $R_{r,t}$ in all post-treatment periods. The joint F-statistic in the years 2016–2019 is 280. It also shows how the treatment intensity increases continuously after 2015.[11] By 2018, a 1 standard deviation increase in the distance-exposure is associated with a 2 pp increase in Venezuelan/native ratio.

To finalize the shift-share instrument, we interact the share Z_r with shift S_t: $Z_{r,t} = Z_r * S_t$, where S_t is the number of refugees in Colombia in a given year.

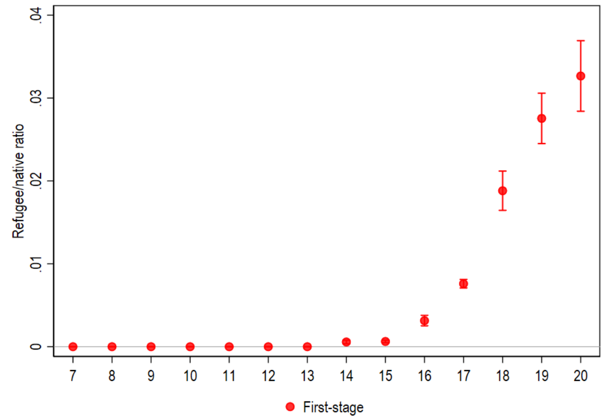
To visually and flexibly assess the pattern of outcomes $y_{r,t}$ captured by the distance exposure relative to the beginning of the refugee crisis we estimate an even study design. The basic nonparametric event study specification takes the form

[9] Travel distance is a strong predictor of migrant settlement and has been routinely used to construct instruments in settings similar to our since Angrist and Kugler (2003).

[10] City centers in each state or department are used to calculate the travel distance.

[11] In the GEIH data, we can separate between Venezuelans and Colombians only after 2013. In our analysis we assume that there were no Venezuelans in Colombia before this date, which is a safe assumption as in 2014 only 0.16% of the working-age population were Venezuelan.

Figure 5: Event Study of the First-stage



Notes: The regression equation is: $R_{i,t} = \sum_{j\neq 2010} \theta_j(\text{year}_j \times Zr) + f_r + f_t + \eta_{r,t}$, where the instrument Z_r is standardized to have mean zero and standard deviation of one to have economically meaningful coefficients, f_r and f_t are region and year fixed effects. Standard errors are clustered at the region level. The 95% confidence intervals are shown.

$$y_{r,t} = \sum_{j\neq 2015} \theta_j(\text{year}_j \times Z_r) + f_r + f_t + \epsilon_{r,t} \tag{6}$$

where the distance share Z_i is interacted with year dummies year_j, and f_i and f_t are region and year fixed effects. The coefficients θ_j compare the outcomes in regions close to and away from the border. For these to have a causal interpretation the identifying assumption is that, absent the treatment, their outcomes would evolve similarly. To establish this claim, it is customary to ensure they did so at least prior to the treatment. However, this is not the case for several of the outcomes we look at. In the pre-period (2007–2015), regions closer to the border observed decreases in salaried employment rates relative to regions further from the border. This leads to a negative trend that is correlated with the instrument.

Visual inspection of the nonparametric event study estimates for various outcomes does not suggest a particular parametric solution, such as adding linear trends, that fit all outcomes of interest. To make progress, we employ the synthetic IV (SIV) methodology developed by Gulek and Vives-i Bastida (2023).

4.2 The synthetic IV method

Synthetic IV is a non-parametric method that combines the instrumental variable strategy with synthetic controls. We describe the methodology in greater detail in Appendix D, and refer the reader to Gulek and Vives-i Bastida (2023) for a full treatment.

In a nutshell, the procedure is as follows. First, find synthetic control (SC) weights solving the standard synthetic control program for pre-treatment outcomes. Use these weights to generate synthetic data (outcome $\hat{y}_{it}^{SC}$, treatment $\hat{R}_{it}^{SC}$, and instrument $\hat{Z}_{it}^{SC}$). Then, subtract the synthetic data from the real data to obtain *debiased* data ($\tilde{y}_{it} = y_{it} - \hat{y}_{it}^{SC}$, $\tilde{R}_{it} = R_{it} - \hat{R}_{it}^{SC}$, $\tilde{Z}_{it} = Z_{it} - \hat{Z}_{it}^{SC}$). Finally, estimate the desired model using the *debiased* data.

Intuitively, *debiasing* the data in this way addresses the pre-trend problem. However, it does not address the fact that immigrants can choose their location based on contemporaneous economic shocks. This is still addressed by the instrument Z. Put differently, SIV addresses the unobserved confounding problem via synthetic control and the endogeneity problem via IV.

In section 4.4 we show the reduced-form event study results by estimating equation 6 separately using IV and SIV, with the only difference being that the instrument and the outcome are their debiased versions in SIV. To show economically meaningful results, we also present 2SLS estimates of the following system of equations using the debiased data:

$$\begin{aligned} \text{LATE:} \quad \tilde{y}_{i,t} &= \beta^{LATE}\tilde{R}_{i,t} + f_i + f_t + \epsilon_{i,t} \\ \tilde{R}_{i,t} &= \theta\tilde{Z}_{i,t} + g_i + g_t + \eta_{i,t} \end{aligned} \tag{7}$$

4.3 Threats to Identification

In our empirical strategy, we essentially compare regions close to the border with those further away. There are two main reasons why this comparison may not be valid.

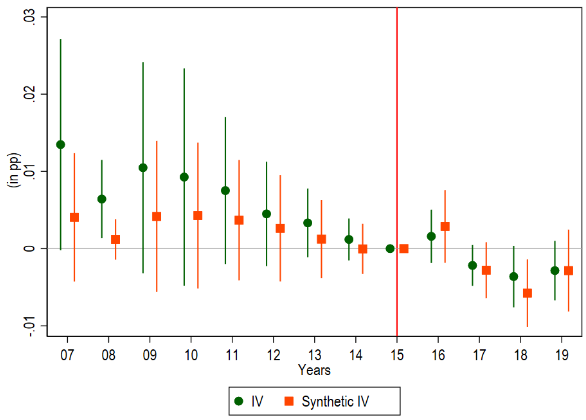
First, if border regions traded with Venezuela to a greater extent than other regions, then Venezuela's economic crisis would likely impact them more. Empirically, Venezuela was not a major trade partner of any region in Colombia. Moreover, we do not find significant changes in trade flows correlated with the instrument. Appendix Figure C.4 provides more details.

Second, the Stable Unit Treatment Value Assumption (SUTVA) is violated if markets reequilibrate across space through the movement of capital and people. We address this concern by focusing on the short run, since these adjustments arguably take several years. We also show that in Colombia, the refugee crisis impacted internal migration of natives only in small quantities. Appendix Figure C.5 shows that regions closer to the border faced slightly more out-migration and less in-migration. However, these effects are small in magnitude and likely do not bias our estimates in an economically meaningful way.[12]

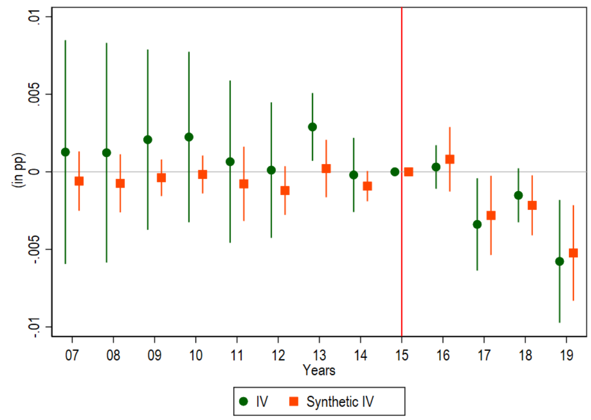
[12] It should be noted that any such bias from reequilibration of spatial markets would cause us to underestimate the effect of refugees on labor markets.

Figure 6: Effects on Native Salaried Employment

(a) Informal Employment



(b) Formal Employment



Note: IV estimates come from the reduced form regression: $y_{i,t} = \sum_{j \neq 2015} \theta_j (year_j * Z_i) + f_i + f_t + \epsilon_{i,t}$. Synthetic IV estimates come from the same reduced-form design where the dependent variable and the instrument are the debiased versions after applying the first step of the SIV procedure described in text. Standard errors are clustered at the department level.

4.4 Results

Next, we review the empirical results. First we look at the event study estimates. These provide supporting evidence for the use of the Synthetic IV methodology since they demonstrate the weakness of the parallel trends assumption in our setting. Second, we present the 2SLS estimates for the heterogeneity analysis across education-sex cells. Given the evidence of the event study analysis, we focus on the Synthetic IV methodology in this second part.

Event study estimates

Figure 6 plots the θ_j coefficients from equation 6 using the standard IV and the new SIV designs. The outcome variable is the native salaried employment rate in the informal sector (Panel A) and in the formal sector (Panel B).

In Panel A, the IV design estimates present a negative trend between 2007–2014. In other words, the difference between the native informal salaried employment rates close to and away from the border was steadily declining. The trend undermines the validity of the exclusion restriction assumption since it reveals that the instrument is still correlated with some unobserved confounder. This raises questions about the validity of the IV design that has been used extensively in the growing set of work studying the effect of Venezuelan migration on Colombian labor markets (Caruso et al., 2021; Delgado-Prieto, 2022; García-Suaza et al., 2024).

Using the Synthetic IV methodology, we no longer observe a pre-trend in Panel A. This implies a good pre-treatment fit of the matching step (Gulek and Vives-i Bastida, 2023). The SIV estimates document a larger decline in informal salaried employment rates between 2017–2019. This suggests that the migration shock decreased the native informal salaried employment rate.

Note that in 2017 and 2018, the SIV estimates are larger in absolute terms than the IV estimates. This means that, absent the migrant treatment, informal salaried employment in regions close to the border would have risen relative to other regions. Put differently, the negative trend between 2007–2014 not only slows down but also flips sign in the post period. This is not surprising as the GDP growth rate also flips sign during this period. The GDP per capita in Colombia increased from $4762 in 2007 to $8167 in 2014, but then decreased to $5936 in 2016. If the unobserved confounders leading to the negative trend in the pre-period were related to economic growth, for example if the more developed non-border regions spearheaded GDP growth, then the 2015 slowdown could have generated this change in trend.

In Panel B we observe formal native salaried employment losses in the post period. Since the instrument does not predict systematic differences between border and non-border regions in the pre-period, relaxing the share exogeneity assumption via SIV does not change our conclusions.

What do these results imply about our predictions? Work permits were very limited until the beginning of 2018. Hence, before then the migration shock constituted a solely informal labor supply shock. The loss of native informal salaried employment in 2017 validates the model's first prediction: an informal labor supply shock causes natives to lose informal jobs. Our second finding, that natives also lose formal salaried jobs, implies that informal and formal labor are highly substitutable in production.

2SLS Estimates

Next, we present the results from estimating equation 7 in figure 7. Panel A compares the outcomes for formal and informal salaried native employment rates. Panel B compares the outcomes for salaried, non-salaried and overall employment. In each subfigure, the first row shows the results on all natives, and rows 2–7 show results across the six education-sex groups defined earlier.

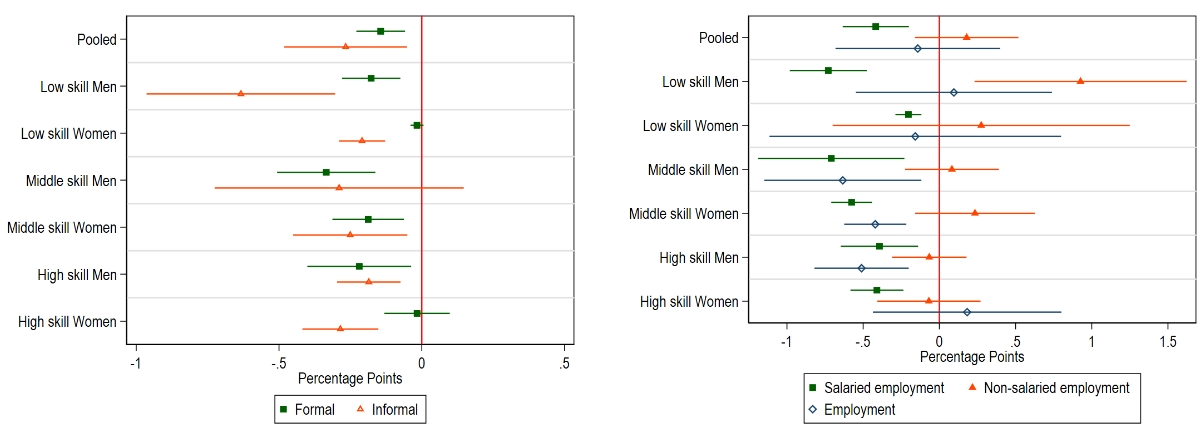
The first row of Figure 7a shows that that the mostly informal immigration shock caused significant declines in the informal and formal native salaried employment rates. A 1 pp increase in migrant/native ratio decreases the native salaried employment rate by 0.27 pp in the informal sector and 0.14 pp in the formal sector. These effects are consistent with model predictions 1&2 and imply that formal and informal labor are highly substitutable in production. To understand the economic significance, consider an economy with 1000 working age natives. Based on 2015 employment statistics of natives, there would be on average 270 salaried workers, of which 180 were formal and 90 were informal. Imagine that 10 refugees arrived to this city, hence a 1 pp increase in refugee/native ratio. According to our estimates, this leads to a 2.7 workers losing informal salaried jobs (out of 90 informal salaried workers), and 1.4 workers losing formal salaried jobs (out of 180 formal salaried workers). Hence, 10 immigrants, having a 68% employment rate, leads to 4.1 natives losing salaried jobs.

Rows 2–7 of 7a shows that that almost all skill groups lose both informal and formal salaried jobs. Recall that only some middle-skill and high-skill male refugees were able to find formal jobs after receiving work permits. The fact that formal job losses of natives are not only coming from high and middle-skill men is reassuring of our preferred mechanism.

Figure 7: LATE estimates

(a) Informal vs Formal

(b) Salaried vs Non-salaried



The estimates come from the Synthetic IV methodology described in text. Panel A shows the estimated effects on the informal and formal salaried employment rates across all skill levels, and Panel B shows the estimated effects on salaried and non-salaried employment rates. Standard errors are clustered at the region level.

We argued that focusing on salaried employment is crucial when studying labor demand. Figure 7b provides empirical evidence that salaried and non-salaried jobs are driven by different economic forces. It shows the 2SLS estimates of equation 7, where the left hand side variable is the salaried, non-salaried, and total employment rates of natives. We abstract away from formal and informal divide since most non-salaried jobs are informal in Colombia.

As in Figure 7a, row 1 shows the results on all natives, and rows 2–7 show heterogeneity by education-sex cells. We find three notable results. First, consistent with Figure 7a, we find decreases in salaried jobs throughout education-sex cells. Second, we find a sizeable, albeit statistically insignificant, increase in non-salaried jobs in the aggregate. Across education-sex cells, this effect comes mostly from low-skill workers. We also document positive but statistically insignificant effects on low-skill women, middle-skill men and women. We find null results on high-skill men and women.

What do these results imply about our understanding of immigrants' effects on host countries' labor markets? First, from a theoretical perspective, they imply that for natives without college degrees the outside option of salaried jobs is not unemployment but rather non-salaried employment. This is especially true for low-skill men. From an empirical perspective, the results show that studying overall employment rates can miss important economic adjustments due to two components of employment (salaried and non-salaried) being impacted in opposite signs. In this setting we do not find statistically significant changes in the overall native employment rate in the pooled regression. This could lead us to erroneously infer that the massive inflow of immigrants had no impact on the labor market outcomes of natives.[13] Across education-sex cells, we find statistically significant effects on overall employment rates only in the case of natives with at least high-school degrees for whom transition to non-salaried jobs is relatively minuscule in magnitude.

4.5 Robustness Checks

Industry heterogeneity

We first check whether the effects we find on salaried and non-salaried employment rates of natives are disproportionately driven by particular industries. We do so by estimating equation 7 where the outcome variable is the employment rate of major industry groups following ISIC standards. Figure A.1 shows the results. Salaried employment rates go down across various industries, except for Health and Social Services. The native escape to non-salaried employment occurs mostly in Manufacturing, Transportation, and to some degree in Agriculture, Health and Social Services, and Education. This transition to non-salaried employment masks employment losses in these industries.

[13] Alternatively, if we had studied the changes in informal and formal employment rates as opposed to salaried employment rates, we would have found only decreases in formal employment rates and no change in informal employment rates, similar to Delgado-Prieto (2022), as the transition to non-salaried jobs is almost entirely informal. This would be the wrong conclusion based on our results, and would require perfectly elastic native labor supply to justify theoretically.

Quality Checks of SIV

The performance of synthetic control based estimators is contingent upon certain conditions being satisfied in practice, such as sufficient signal to noise ratio in the training data to prevent overfitting (Abadie and Vives-i Bastida, 2022). SIV is not an exception. Gulek and Vives-i Bastida (2023) provide three robustness checks that practitioners should implement when using SIV:

1. Show the new first stage: the debiased instrument should have strong predictive power over debiased treatment

2. Show the pre-treatment fit: the debiased data should not show significant pre-trends.

3. Apply back test: back testing can reveal the degree by which a good pre-treatment fit is due to over-fitting.

We perform all these checks and show them in the Appendix Section G. Overall, we find that our main results remain robust to these checks.

First, notice that SIV creates different weights for each outcome of interest, hence it creates different debiased treatment and instrument for each outcome. Figure G.8 shows the scatterplot of the F-stats from different specifications we used in Figure 7. Of the 35 specifications, the minimum F-statistic of the instrument in the post-period (2016–2019) is 32, with more than 75% of the specifications having an F-stat of more than 100. Overall, we maintain a strong first-stage in all of our main estimates.

Second, we show that applying SIV reduces the pre-trends in the training period. This implies a good pre-treatment fit and is necessary for the SIV estimates to remain consistent. We further improve upon the pre-treatment fit by adding a constant shift (Doudchenko and Imbens, 2016). This is equivalent to computing a synthetic control where the outcome variable is measured in deviations from its pre-treatment means (Abadie and Vives-i Bastida, 2022). Our main results remain robust. Figure G.9 displays the event study estimates of the results shown in Figure 7a using SIV with and without a constant shift, and Figures G.10, G.11 and G.12 do the same for the results shown in Figure 7b.

Third, we apply back-testing by limiting the training period from 2007–2015 to 2007–2012. Figure G.14 shows that our results remain robust, which implies that we do not suffer from over-fitting. This is to be expected as our original training period (2007–2015) is long compared to our post-treatment period (2016–2019). Moreover, we show that our main results also remain robust to applying shorter training periods while applying demaned SIV in Figure G.14.

5 The effects of Work Permits

Next, we study the effect of providing work permits to migrants on the native population.

5.1 Identification

A key identification challenge is that the work permit treatment correlates with the migrant treatment both across time and space. Appendix Figure B.2 shows that regions receiving more migrants also received more work permits. When controlling for the refugee treatment, there is no spatial variation left to identify the effect of work permits.

To make progress, we exploit the fact that work permits had little to no bite for women and low-skill men. As shown in Section 2.1, PEP only enabled high-skill and middle-skill men to find formal jobs. Other Venezuelan refugees were not affected. Moreover, the high and middle skill men were impacted by the work permits at different times. This variation across skill cells provides a third dimension in which the work permit treatment intensity varies in addition to space and time. It allows for a triple DiD design that identifies the effect of work permits by partialing out the effects of immigration.

The triple differences in differences design

The difference in exposure to work permits amongst migrants forms the basis for a triple differences in differences design. Middle and high skilled migrant men are closest substitutes to similarly skilled native men. Therefore, we argue that middle and high skilled native men were more exposed to the provision of work permits than other native education-sex groups.

We first compare the employment outcomes of high-skill native men to the remaining education-sex cells (omitting middle-skill men), before and after the provision of work permits in the third quarter of 2017, in regions close to and far away from the border (Equation 8). We repeat this exercise for middle-skill native men (omitting high-skill men) before and after the fourth quarter of 2018 (Equation 9). This comparison can be written in an event-study model, separately for each treated cell, as follows:

$$Y_{brt} = \sum_{j \neq 2017q3} \beta_j \mathbb{I}_{b=HS-M} Z_r \mathbb{I}_{t=j} + f_{br} + f_{bt} + f_{rt} + \epsilon_{brt} \tag{8}$$

$$Y_{brt} = \sum_{j \neq 2018q4} \gamma_j \mathbb{I}_{b=MS-M} Z_r \mathbb{I}_{t=j} + f_{br} + f_{bt} + f_{rt} + \epsilon_{brt} \tag{9}$$

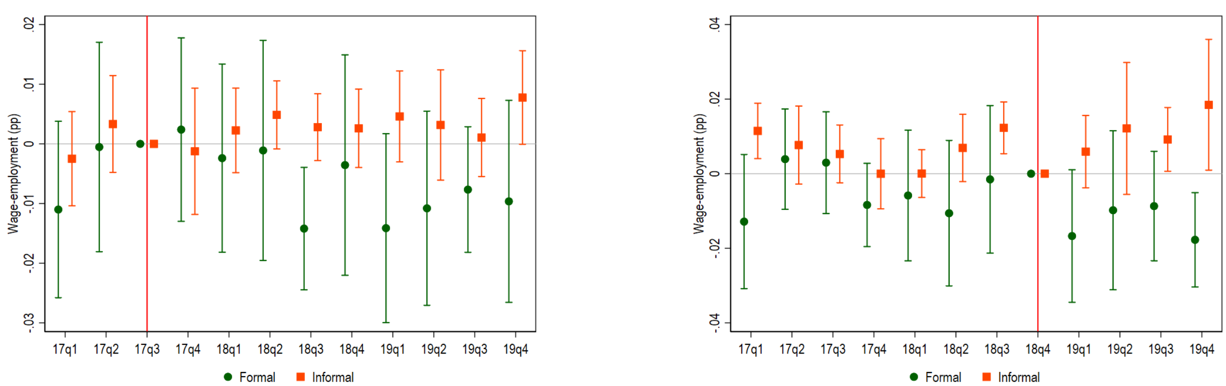
where f_{br} are bin*region fixed effects that capture anything about a region that might lead to differential outcomes by skill-cell, f_{bt} are bin*time fixed effects that capture any secular trends in labor market outcomes by skill-cell, and f_{rt} are region*time fixed effects that help us partial out regional trends. Crucially, under the assumption that the immigration shock itself does not impact the high-skill or middle-skill native men differently compared to other education-sex cells, these region*time fixed effects partial out the effect of immigration from the effect of work permits.

The identifying assumption is that, absent the provision of work permits, the gap between employment in regions close to and far away from the border, between treated men and others, would have evolved similarly over time. This is a statement, in differences, of the familiar parallel trends assumption. Visually, we would expect all the β_j and γ_j coefficients in equations 8 and 9 to be centered around zero absent treatment.

Additionally, the differential timing of the work permit treatment provides a sanity check for our identification strategy. We expect high-skill male natives to be impacted five quarters earlier than middle-skill male natives. Arguably, the likelihood that a confounder impacts high-skill and middle-skill men in the exposed regions five quarters apart is significantly low.

5.2 Results

Figure 8: Event-study estimates of education-sex exposure to work permits



(a) High-skill men vs Control

(b) Middle-skill men vs Control

The estimates come from triple DiD event-study designs: Panel A: $Y_{brt} = \sum_{j\neq 2017q3} \beta_j \mathbb{I}_{b=HS-M} Z_r \mathbb{I}_{t=j} + f_{br} + f_{bt} + f_{rt} + \epsilon_{brt}$ and Panel B: $Y_{brt} = \sum_{j\neq 2018q4} \gamma_j \mathbb{I}_{b=MS-M} Z_r \mathbb{I}_{t=j} + f_{br} + f_{bt} + f_{rt} + \epsilon_{brt}$. Control cells include low-skill men and all skill types of women. Standard errors are clustered at region-bin level.

Figure 8 plots the β_j and γ_j estimates in panels A and B respectively. Panel A shows that work permits reduce the formal salaried employment of high-skill men. They weakly increase, albeit insignificantly, their informal salaried employment. These effects are apparent by the third quarter of 2018, after the first two waves of PEPs. Panel B shows the same results for middle skill men. They too observe decreases in formal salaried employment and increases in informal salaried employment. Both results are statistically significant by the fourth quarter of 2019.

These intent-to-treat (ITT) effects provide evidence for the third and fourth predictions of the model. As work permits shift the labor supply shock from the informal sector to the formal sector, natives lose formal jobs due to increased competition and gain informal jobs due to decreased competition.

We conduct two more exercises to ensure that these effects are not driven by outlier cells within the region-time-skill dimension and report them in Appendix C. First, we estimate region-specific DiD estimates comparing the employment outcomes of high-skill and middle-skill men to the rest, before and after the issuance of work permits. We plot these estimates along the distance exposure dimension in Figure C.6. This figure shows that our results are not driven by an outlier region having skill-cell specific shocks. Second, we estimate education-sex cell specific DiDs, comparing employment outcomes in regions close to and away from the border, before and after the issuance of work permits. We plot these estimates for each education-sex cell in Figure C.7. This exercise shows that our results are not driven by an outlier skill-cell in one of the control groups. Together, these exercises reveal that our triple DiD estimates are not driven by outlier cells, which increase the reliability of our conclusions.

Data limitations prevent us from testing the fifth and final prediction of our model: that work permits increase the total formal employment in the economy. GEIH data is too noisy to observe with precision the number of Venezuelans across region*skill*time cells, and the administrative records of formal workers with PEPs do not allow us to separate workers across skill cells.

However, the fact that 4 out of the 5 predictions are validated in the data gives us confidence that our simple model is a good approximation of the world. Therefore, since we cannot empirically measure the effect of work permits on formal job creation, we use the model to quantify it in the next section.

6 Model Estimation and Counterfactuals

This section discusses the estimation of the full model with firm heterogeneity. It closely follows Gulek (2023). We use a minimum distance estimator. Firm heterogeneity is introduced to obtain additional moments for identification. Section 6.1 sets up the full model, while Section 6.2 describes the estimation method, identification, and the model's fit.

6.1 Introducing Firm heterogeneity in productivity

Building on the representative firm framework of Section 3 we allow for firms to have different productivities denoted by $\theta \in \{\theta_1, \ldots, \theta_K\}$, which enters firms' production function in a Hicks-neutral way:

$$F(\ell_i, \ell_f; \theta) = \theta(\eta \ell_i^\rho + (1-\eta)\ell_f^\rho)^{\frac{\alpha}{\rho}}$$

Firm of type θ's objective function is given by:

$$\max_{\ell_i, \ell_f} F(\ell_i, \ell_f; \theta) - \ell_i^{1+\gamma} w_i - (1+\tau_w) w_f \ell_f$$

The first-order conditions determine the labor demand functions of each firm of type θ:

$$\begin{aligned} \alpha\eta\ell_i^{\rho-1-\gamma} Y^{\frac{\alpha-\rho}{\alpha}} &= w_i(1+\gamma) \\ \alpha(1-\eta)\ell_f^{\rho-1} Y^{\frac{\alpha-\rho}{\alpha}} &= w_f(1+\tau_w) \end{aligned}$$

where $Y(\theta) = \theta(\eta\ell_i^\rho + (1-\eta)\ell_f^\rho)^{\frac{\alpha}{\rho}}$ is the output produced by the firm of type θ. Solving these two equations for $L_i(\theta)$ and $L_f(\theta)$ determines the informal and formal labor demanded by firms of type θ. The total labor demand curves are given by aggregating these group-specific labor demand curves.

Given K types of firms with productivities $\theta \in \{\theta_1, \ldots, \theta_K\}$, let n_j and m_j denote the ratio of informal and formal labor hired by firms of type θ_j. The aggregate informal labor demand elasticities w.r.t. informal wages are then given by weighted averages of group-specific elasticities:

$$\begin{aligned} \overline{\epsilon_{L_i,w_i}} &\coloneqq \sum_{j=1}^{K} \epsilon_{L_i,w_i}(\theta_j) n_j \\ \overline{\epsilon_{L_f,w_i}} &\coloneqq \sum_{j=1}^{K} \epsilon_{L_f,w_i}(\theta_j) m_j \end{aligned}$$

where the group-specific labor demand elasticities are given by:

$$\begin{aligned} \epsilon_{L_i,w_i}(\theta) &= -\frac{1-\rho-(\alpha-\rho)s_f(\theta)}{(1-\rho+\gamma)(1-\rho)-(\alpha-\rho)[(1-\rho+\gamma)s_f(\theta)+(1-\rho)s_i(\theta)]} \\ \epsilon_{L_f,w_i}(\theta) &= -\frac{(\alpha-\rho)s_i(\theta)}{(1-\rho+\gamma)(1-\rho)-(\alpha-\rho)[(1-\rho+\gamma)s_f(\theta)+(1-\rho)s_i(\theta)]]} \end{aligned}$$

where $s_i(\theta) = \frac{\eta\ell_i(\theta)^\rho}{(\eta\ell_i(\theta)^\rho+(1-\eta)\ell_f(\theta)^\rho)}$ is the share of informal labor in production for firms of type θ.

We partition the vector of parameters into two groups based on whether they are calibrated or estimated. $\alpha = 0.49$ is calibrated based on the share of labor in production in Colombia (acquired from Penn World Tables), informal wage w_i and formal wage w_f for the low-skilled are estimated using the labor force surveys, the labor tax rate is set to its statutory value $\tau_w = 0.39$. The value of τ_w corresponds to the effective tax rate for minimum wage earners.

6.2 Estimation Method

We take the parameters defined in the first step as given and use a Minimum Distance estimator to obtain the remaining model parameters. The model has three core parameters $\{\gamma, \eta, \rho\}$ and K productivity measures θ_K that need to be estimated. The estimator proceeds in two steps. First, it uses the model to generate the informal and formal labor demanded by each firm type. Second, it uses these inputs to compute the set of moments computed from actual data and the SIV estimates. The estimate is obtained as the parameter vector that best approximates these moments.

Let $\hat{m}_N = \frac{1}{N}\sum_{i=1}^{N} m_i$ denote the vector of moments computed from data, which can include, for example, the share of informal workers hired by firms of different sizes. Let the model-generated counterpart of these moments be denoted by $m(\Phi;\Psi)$. Define $g_N(\Phi;\Psi) = \hat{m}_N - m_s(\Phi;\Psi)$; the estimator is then given by

$$\hat{\Phi} = \arg\min_{\Phi} Q(\Phi;\Psi) = \{g_N(\Phi;\Psi)' W_N g_N(\Phi;\Psi)\} \tag{10}$$

where W_N is a positive, semi-definite weighting matrix. For simplicity, we use a diagonal matrix where each element is the inverse of the square of the empirical moment. This way, percentage deviations from the moments take equal weight.

Moments and Identification

We use nine moments from the data and our SIV estimates to form the vector $\hat{m}_N$. GEIH asks respondents how many people work in their establishment, and group results in 7 categories: between 2–3, 4–5, 6–10, 11–19, 20–30, 31–50, and 51–100 workers. We follow this structure of the GEIH. The moments we choose are (i) the size of firms in different groups (calculated using GEIH), (ii) the informality rate of firms in different groups (calculated using GEIH), (iii) the ratio of informal and formal labor demand elasticities (estimated in the empirical section).

This section's main goal is not to provide a rigorous proof of identification. Nonetheless, here we explain how the observed variations in data, combined with the outcomes of reduced-form analyses and the structure of the underlying model, help determine the model's parameters. In this model, the sole means by which firms can augment their output is by increasing their workforce, as labor constitutes the exclusive input in the production process. Consequently, the distinction between larger and smaller firms hinges entirely upon disparities in their productivities denoted as θ. More productive firms choose to expand their workforce. The parameter γ, which governs the marginal cost of employing informal workers, predominantly hinges on the extent to which larger firms opt for formalization at the intensive margin. For all types of firms, the share parameter η is linked to the relative productivity of formal and informal workers and, thus, is determined by the proportion of informal workers in the overall economy. The elasticity of substitution between informal and formal workers is primarily dictated by demand elasticities. For instance, the sign of the formal labor demand elasticity in isolation provides set identification for ρ as $\rho > \alpha \iff \epsilon_{L_f,w_i} > 0$. Similarly, the relative magnitudes of the elasticities of informal and formal labor demand, expressed as $\frac{\epsilon_{L_f,w_i}}{\epsilon_{L_i,w_i}} = \frac{(\alpha-\rho)s_i}{1-\rho-(\alpha-\rho)s_f}$, assist in pinpointing ρ. Holding the share of informal labor constant, this ratio exhibits a declining trend with respect to ρ.

Estimates and Model Fit

Table 1: Parameter Values

Parameter	Description	Source	Value
τ_w	Payroll tax	Statutory values	0.326
w_i	Informal wages	Calibrated	3.17
w_f	Formal wages for the low-skilled	Calibrated	5.949
α	Cobb-Douglass coefficient	Calibrated	0.49
γ	Intensive mg. cost of informal labor	Estimated	1.11
η	Informal share parameter	Estimated	0.86
ρ	CES elasticity parameter	Estimated	0.91
θ_1	Productivity of firms between 2–3 workers	Estimated	65
θ_2	Productivity of firms between 4–5 workers	Estimated	184
θ_3	Productivity of firms between 6–10 workers	Estimated	254
θ_4	Productivity of firms between 11–19 workers	Estimated	307
θ_5	Productivity of firms between 20–30 workers	Estimated	355
θ_6	Productivity of firms between 31–50 workers	Estimated	413
θ_7	Productivity of firms between 51–100 workers	Estimated	492
$\sigma_{i,f}$	Elasticity of substitution between informal and formal workers	Implied	11
ϵ_{L_i,w_i}	Elasticity of informal labor demand w.r.t. informal wages	Implied	-0.65
ϵ_{L_f,w_i}	Elasticity of formal labor demand w.r.t. informal wages	Implied	1.41

Note: Note: Formal and informal hourly wage (for salaried work) estimates are expressed in Pesos using the 2007–2015 sample of the GEIH.

Table 1 shows the values of all parameters. The most critical estimate is that the CES elasticity parameter ρ is 0.91, which implies an elasticity of substitution between informal and formal labor of 11. To the best of our knowledge, this is one of the first papers to estimate this elasticity.[14] This relatively high elasticity is consistent with the Colombian context, where informal employment is often in the same sectors and even in the same firms as formal employment. It also supports the assumption of perfect substitutability between informal and formal workers in the recent structural literature on the informal sector (Ulyssea, 2018, 2020), and is quantitatively similar to the other estimate in the literature coming from Turkey (Gulek, 2023).

Appendix Table A.3 shows how the model performs compared to all of the targeted moments in the data. The model matches most of the moments of the data quite well. In general, it overestimates the informality of smaller firms and underestimates it for the larger firms.

Counterfactual: How many formal jobs have been created due to work permits?

Given the model parameters, we can estimate the effect of providing work permits to refugees on formal job creation. Here is how we proceed. From Figure 1c, we estimate how many Venezuelans were able to find formal jobs thanks to work permits. This enables us to estimate how much the total informal labor supply in the economy changes. From model parameters we determine the informal labor demand elasticity ϵ_{L_i,w_i}. This allows us to determine how much informal wages would change given the change in informal labor supply due to work permits. Lastly, the formal labor demand elasticity times the estimated change in informal wages gives us the percentage change in the formal labor demand, which in turn gives us the number of formal jobs that are created in the economy. We estimate that from 240,000 work permits issued between PEP1–3, around 68,000 Venezuelans became able to transition from informal to formal sectors.[15] This in turn has created around 24,440 formal jobs in the economy. Even if these jobs were minimum wage jobs, the total tax revenue from granting

[14]There are two other papers that we are aware of that estimate this elasticity. Schramm (2014) studies the equilibrium effects of taxation on sectoral choice, work hours and wages in Mexico. She finds this elasticity to be around 1.8. Gulek (2023) studies the effect of informal Syrian refugees on Turkish labor markets and estimates this elasticity to be around 10. Informal and formal workers working in different sectors and firms in Mexico as opposed to working in the same firms as in Turkey and Colombia could explain this discrepancy.

[15]We also validate this number using admin records. By the end of 2019, 73360 Venezuelans with PEPs were paying social security through their jobs. Using this number as opposed to 68,000 does not change our counterfactual estimates in a meaningful way.

work permits would be around 22.5 million USD annually.[16]

7 Do work permits reduce mismatch in the economy?

Section 5 shows that providing migrants with work permits leads them to transition from informal to formal work, and cause natives to transition from formal to informal work. At face value, the simplest interpretation is that work permits reshuffle immigrants and natives between informal and formal sectors. However, there could also be productivity gains if immigrants change the type of jobs or tasks they do. Consider an engineer who works informally at a restaurant because he does not have a work permit. Permits may enable him to work at a large factory that only hires formal labor and perform tasks closer to his skills. Such a transition could increase competition among natives in these high-skill industries and reduce their wages. There could also be positive knowledge spillovers from the decline in skill-mismatch in the economy, which could increase natives' wages in these high-skill industries. In this section, we look for evidence of such spillovers.

In the Appendix Section E, we use the canonical labor demand framework to model how work permits enabling high-skill refugees to transition from low-skill to high-skill jobs can impact natives' wages. We show that absent changes in skill-enhancing productivity shocks, a transition of refugees from low-skill to high-skill industries necessarily lowers natives' wages in skill-intensive industries. However, if the knowledge spillovers from refugees is skill-enhancing, e.g., the engineer refugees' human capital makes other engineers more productive compared to janitors, then natives' wages in skill-intensive jobs need not decrease and actually might increase. This provides a testable prediction for the existence of knowledge spillovers: if the wage gap between high-skill and low-skill natives does not decrease or even increases in more skill intensive industries, then we can conclude that knowledge spillovers exist.

To test this hypothesis, we employ a quadruple DiD design where we compare the wages of more exposed natives (high and middle-skill) to less exposed natives, in regions close to and away from the border, before and after the work permits are issued, and in more/less exposed industries based on their underlying formality or skill intensity. Specifically, we

[16] A similar counterfactual exercise can be found in Gulek (2023), who estimates what would have happened if Syrian refugees were all given work permits in Turkey. Since he does not know what ratio of refugees would be able to find formal jobs given work permits, he can only give a range of estimates on the effect of work permits. In contrast, because we know the exact number, we can point estimate the effect of granting work permits on formal job creation.

augment the triple DiD design of section 5 and add a fourth difference across industries:

$$ln(w_{ibrt}) = \beta D_{brt} * Exposure_i + f_{ibr} + f_{ibt} + f_{irt} + f_{brt} + \epsilon_{ibrt} \tag{11}$$

where w_{ibrt} is a statistic of natives' wages at industry-skill-region-time level, $Exposure_i$ measures the share of formal workers or the share of high skill workers at the industry level, D_{brt} is the treatment intensity at skill-bin*region*time level (i.e., D_{brt} is equal to the distance instrument Z_r for middle skill men after the fourth quarter of 2018, and for high skill men after the third quarter of 2017.), f_{ibr} are three-way fixed effects that account for level differences at the industry-skill-region level, f_{ibt} accounts for trends at the industry*skill level (e.g., allows for high-skill/low-skill wage-gaps in Oil to follow a common trend across regions), f_{irt} accounts for trends at the industry*region level (e.g., allows for average wages in Agriculture to follow different trends in each region), and f_{brt} accounts for trends at the skill-region level. The rich set of three-way fixed effects ensures that we can account for several factors that can impact natives' wages and allow us to exploit the variation in treatment that is purged from bias from unobserved confounders.

To facilitate the interpretation of this quadruple DiD design motivated by the model, we first perform a preliminary triple DiD analysis, similar to the design of section 5, for each of the major industries i in Colombia separately:

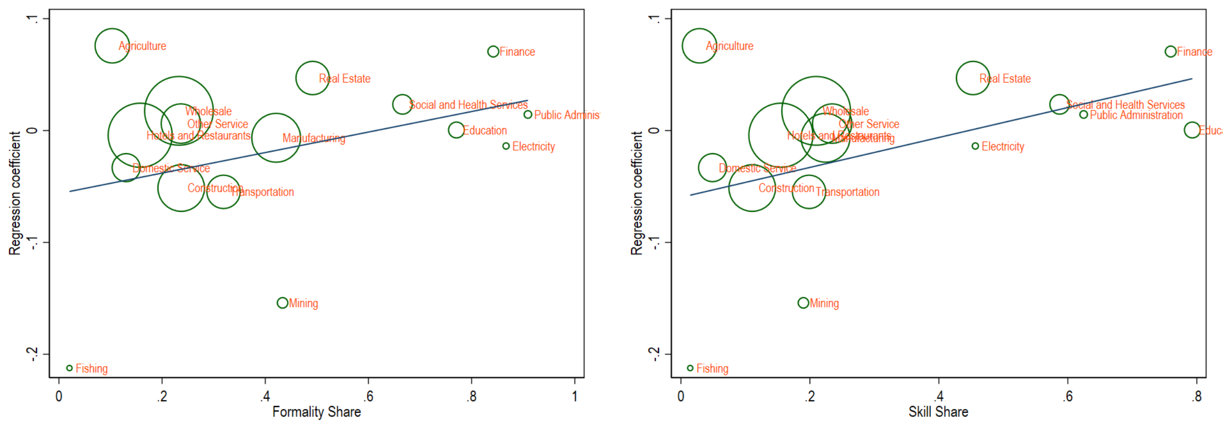
$$ln(w^i_{brt}) = \beta^i D_{brt} + f^i_{br} + f^i_{bt} + f^i_{rt} + \epsilon^i_{brt} \tag{12}$$

Figure 9 plots the industry specific β^i estimates against formality rate in Panel A and skill-intensity in Panel B. In industries with low shares of high-skill or formal workers, the more exposed natives (high and middle-skill natives in regions closer to the border) experience decreases in wages after the issuance of work permits. However, the industry-specific estimates increase as we move from less skill and formal-intensive industries like Fishing to more skill and formal-intensive industries like Finance. The slope of the linear line, which is a visual representation of the quadruple DiD design, is positive. It provides evidence in favor of knowledge spillovers. Put differently, whereas on average high-skill natives lose wages compared to low-skill natives in the economy as a result of the work permit shock (which is consistent with natives losing some formal jobs in the economy as shown by section 5), the decline is non-existent in the more exposed industries based on their skill and formal intensities, which implies that there is an offsetting productivity shock in these industries. We argue that this is evidence for the existence of knowledge spillovers.

To test for the existence of knowledge spillovers more formally, we estimate equation 11 while clustering the standard errors at skill-region level. Table 2 reports the estimates of β

Figure 9: Visual Quadruple DiD design

(a) Formality (b) Skill



Note: Each circle represents a triple DiD estimate from equation: $ln(w^i_{brt}) = \beta^i D_{brt} + f^i_{br} + f^i_{bt} + f^i_{rt} + \epsilon^i_{brt}$, where superscript i denotes an industry, and the dependent variable is the mean wage in a given bin-region-time cell. Circles sizes denote the industry-size. The x axis denotes the ratio of formal workers in Panel A, and the ratio of workers with college degrees in Panel B.

separately for skill and formality exposures. The first column reports the effect on average wages. Across both the formality and skill-intensity dimensions, average wages increase in more exposed industries relative to less exposed ones. These effects are statistically significant at 10% for the formality measure and at 5% for the skill-intensity measure.

Table 2: Quadruple DiD results

Hourly Earnings	(1) Mean	(2) p10	(3) p25	(4) p50	(5) p75	(6) p90
Formality	0.0457*	0.0851***	0.0629***	0.0405	0.0655**	0.0288
	(0.0272)	(0.0252)	(0.0238)	(0.0310)	(0.0297)	(0.0404)
Skill intensity	0.0854**	0.0947***	0.0907***	0.0818**	0.117***	0.0741
	(0.0329)	(0.0332)	(0.0290)	(0.0402)	(0.0341)	(0.0486)

Note: Regression estimates come from the quadruple DiD design: $ln(w_{ibrt}) = \beta D_{brt} * Exposure_i + f_{ibr} + f_{ibt} + f_{irt} + f_{brt} + \epsilon_{ibrt}$, where w_{ibrt} is a statistic of natives' wages at industry-skill-region-time level, $Exposure_i$ is the share of formal workers in industry i in row 1 and the share of college graduates in row 2, D_{brt} is the treatment intensity at skill-bin*region*time level (i.e., D_{brt} equals to distance instrument Z_r for middle skill men after the fourth quarter of 2018, and for high skill men after the third quarter of 2017.), fs denote the three-way fixed effects. column 1 shows the effects on mean wages, and columns 2–6 show the effects on the 10th, 25th, 50th, 75th, and 90th percentiles of the wage distribution, respectively. Standard errors are clustered at the skill-bin*region level.

Taken at face value, this result provides evidence for skill-enhancing knowledge spillovers. However, it could also be driven by compositional effects in either the bottom or the top of the income distribution. For example, work permits could have enabled refugees to displace the lowest earners in the skill-intensive industries, which could explain the increase in average wages. To show robustness, we estimate the effect across the wage distribution in columns 2–6. Estimates remain positive and are significant at both the 25th and 75th percentiles, which rules out compositional changes driving our results.

We acknowledge that skill mismatch, and its decline thereof, can have various effects on the economy. We do not argue that we can estimate the total effects of reducing skill mismatch, nor do we argue that we have a gold-standard identification strategy that isolates the effects of skill mismatch in the economy. Our argument is that reduction of skill mismatch is an important mechanism by which work permits can improve the economy of host countries, and it is therefore worth testing empirically for the existence of such forces. Our setting is not ideal given that the immigration shock and the work permit shock are occurring jointly, but our quadruple DiD design has the benefit of absorbing a large number of alternative economic forces that can impact natives' wages. The evidence is consistent with skill-biased productivity gains that correlate with the work permit treatment. Given that we absorb a large number of alternative economic forces in our three-way fixed effects, we conclude that the evidence is highly suggestive that work permits have caused this change. Future work can investigate occupational changes caused by work permits on both refugees and natives in more detail and try to isolate different mechanisms by which work permits can impact local economies.

8 Conclusion

This paper provides a theoretical and empirical analysis of how informal immigration and the granting of work permits impact labor markets, using the Venezuelan refugee crisis in Colombia as a quasi-experiment. The findings illuminate our understanding of the informal economy and have important policy implications.

We estimate that an increase in the informal labor supply due to the influx of mostly informal Venezuelan refugees significantly lowers natives' salaried employment rates in both the informal and formal sectors, indicating high substitutability between informal and formal workers. In fact, we estimate the elasticity of substitutition between informal and formal workers to be around 11, which supports the assumption of perfect substitutability in the recent structural literature on the informal sector (Ulyssea, 2018, 2020), and is consistent with the other estimates coming from immigration episodes (Gulek, 2023). It also shows that natives, especially low-skill men, who lose their salaried jobs transition to non-salaried jobs. These opposing forces hide immigrants' effect on natives' employment rates. Only by focusing on salaried jobs do we find the disemployment effects of immigration.

Moreover, we show that work permits enabled middle and high-sill male natives to find formal jobs. Immigrants' transition from the informal to the formal sector decreases natives' formal salaried employment rates and increases natives' informal salaried employment rates. These effects are all predicted by our simple model. Furthermore, we show suggestive evidence that the work permits increase the relative wages of high-skill natives in formal or skill intensive industries. This is consistent with skill-biased technology shocks, which we interpret as a knowledge spillover from refugees' working closer to their skill level.

Lastly, we estimate a model of the informal sector and use it to quantify how many formal jobs must have been created by providing work permits. We estimate that work permits enabled 68,000 Venezuelans to transition from the informal to the formal sector. The decrease in the informal labor supply caused firms to demand more formal workers, leading to 24,440 more formal jobs in the economy. Assuming these jobs were minimum wage jobs, the total tax revenue from granting work permits would be around 22.5 million USD annually.

In conclusion, this research provides valuable insights into the complex effects of refugee crises on host country economies and the role of work permits, or the lack thereof, in explaining these effects. Future research can dive into isolating different mechanisms by which work permits can impact the host economies.

References

Abadie, Alberto and Jaume Vives i Bastida, “Synthetic controls in action,” *arXiv preprint arXiv:2203.06279*, 2022.

Amaral, Pedro S and Erwan Quintin, “A competitive model of the informal sector,” *Journal of monetary Economics*, 2006, *53* (7), 1541–1553.

Angrist, Joshua D and Adriana D Kugler, “Protective or counter-productive? labour market institutions and the effect of immigration one unatives,” *The Economic Journal*, 2003, *113* (488), F302–F331.

Bahar, Dany, Ana María Ibáñez, and Sandra V Rozo, “Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees,” *Journal of Development Economics*, 2021, *151*, 102652.

Bosch, Mariano and Julen Esteban-Pretel, “Job creation and job destruction in the presence of informal markets,” *Journal of Development Economics*, 2012, *98* (2), 270–286.

Card, David, “The impact of the Mariel boatlift on the Miami labor market,” *ILR Review*, 1990, *43* (2), 245–257.

— , “Immigration and inequality,” *American Economic Review*, 2009, *99* (2), 1–21.

Carpio, Ximena V Del and Mathis Wagner, *The impact of Syrians refugees on the Turkish labor market*, The World Bank, 2015.

Caruso, German, Christian Gomez Canon, and Valerie Mueller, “Spillover effects of the Venezuelan crisis: migration impacts in Colombia,” *Oxford Economic Papers*, 2021, *73* (2), 771–795.

Clemens, Michael, Cindy Huang, Jimmy Graham et al., “The economic and fiscal effects of granting refugees formal labor market access,” *Center for Global Development Working Paper*, 2018, *496.*

Delgado-Prieto, Lukas, “Dynamics of Local Wages and Employment: Evidence from the Venezuelan Immigration in Colombia,” 2022.

Doudchenko, Nikolay and Guido W Imbens, “Balancing, regression, difference-in-differences and synthetic control methods: A synthesis,” Technical Report, National Bureau of Economic Research 2016.

European Commission, "EU Solidarity with Ukraine," https://eu-solidarity-ukraine.ec.europa.eu/information-people-fleeing-war-ukraine/fleeing-ukraine-your-rights-eu_en. Accessed: 2023-03-10.

García-Suaza, Andrés, Juan Miguel Gallego Acevedo, Juan Diego Mayorga, Angie Mondragón-Mayo, Carlos Sepúlveda, and Alexander Sarango-Iturralde, "Venezuelan Immigration in Colombia: Labor Market Dynamics," *The Center for Growth and Opportunity*, 2024.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik instruments: What, when, why, and how," *American Economic Review*, 2020, *110* (8), 2586–2624.

Gronau, Reuben, "Leisure, home production, and work–the theory of the allocation of time revisited," *Journal of political economy*, 1977, *85* (6), 1099–1123.

Gulek, Ahmet, "Formal Effects of Informal Labor Supply: Evidence from the Syrian Refugees in Turkey," 2023. Available at SSRN: https://ssrn.com/abstract=4264865.

_ **and Jaume Vives i Bastida**, "Synthetic Instrumental Variables in Diff-in-Diff Designs with Unmeasured Confounding," 2023. Available at https://economics.mit.edu/sites/default/files/inline-files/Synthetic_Wald%20%283%29.pdf.

Hunt, Jennifer, "The impact of the 1962 repatriates from Algeria on the French labor market," *ILR Review*, 1992, *45* (3), 556–572.

Lebow, Jeremy, "The labor market effects of Venezuelan migration to Colombia: reconciling conflicting results," *IZA Journal of Development and Migration*, 2022, *13* (1).

Meghir, Costas, Renata Narita, and Jean-Marc Robin, "Wages and informality in developing countries," *American Economic Review*, 2015, *105* (4), 1509–46.

Paula, Aureo De and José A Scheinkman, "The informal sector: An equilibrium model and some empirical evidence from Brazil," *Review of Income and Wealth*, 2011, *57*, S8–S26.

Penaloza-Pacheco, Leonardo, "Living with the neighbors: the effect of Venezuelan forced migration on the labor market in Colombia," *Journal for Labour Market Research*, 2022, *56* (1), 1–32.

Rauch, James E, "Modelling the informal sector formally," *Journal of development Economics*, 1991, *35* (1), 33–47.

Santamaria, Julieth, “when a stranger shall sojourn with thee’: The impact of the venezuelan exodus on colombian labor markets,” *University of Minnesota, Department of Applied Economics, working paper*, 2020, *51422.*

Schramm, Heidi R, “The equilibrium effects of income taxation on formal and informal labor markets,” in “NEUDC 2014 Working Paper. Presented at the Northeast Universities Development Consortium Conference” Boston University Boston 2014.

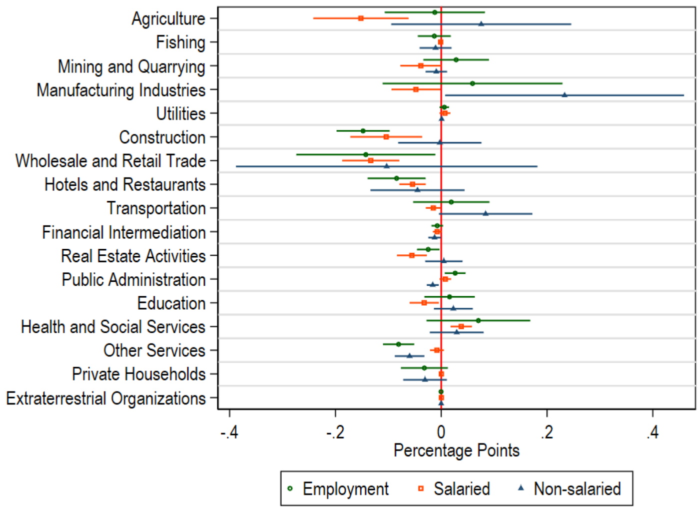
Ulyssea, Gabriel, “Firms, informality, and development: Theory and evidence from Brazil,” *American Economic Review*, 2018, *108* (8), 2015–47.

_ , “Informality: Causes and consequences for development,” *Annual Review of Economics*, 2020, *12*, 525–546.

Appendix

A Data Appendix

Figure A.1: Heterogeneity by Industry



The estimates come from the Synthetic IV methodology described in text. Standard errors are clustered at the region level. Industry definitions follow ISIC standards.

Table A.1: Summary Statistics

Sex	Pooled			Men			Women		
Skill	Pooled	Low	High	Pooled	Low	High	Pooled	Low	High
Panel A: Employment									
All	0.643	0.614	0.673	0.777	0.786	0.768	0.516	0.439	0.589
Formal	0.213	0.089	0.340	0.257	0.131	0.397	0.171	0.047	0.290
Informal	0.430	0.524	0.333	0.520	0.655	0.371	0.344	0.392	0.299
Panel B: Salaried Employment									
All	0.242	0.157	0.328	0.301	0.227	0.382	0.186	0.087	0.281
Formal	0.151	0.066	0.238	0.186	0.100	0.281	0.117	0.031	0.200
Informal	0.091	0.092	0.090	0.115	0.127	0.101	0.069	0.056	0.081
Panel C: Non-salaried Employment									
All	0.372	0.453	0.289	0.446	0.555	0.326	0.302	0.349	0.256
Formal	0.035	0.021	0.049	0.042	0.028	0.058	0.028	0.014	0.041
Informal	0.337	0.432	0.239	0.404	0.527	0.267	0.274	0.335	0.215
Panel D: Income (in thousands)									
All	807	496	1089	872	565	1212	712	363	946
Formal	1298	766	1444	1346	807	1546	1231	649	1322
Informal	538	444	690	618	512	826	420	324	538
Panel E: Wages (Hourly)									
All	4661	2771	6375	4700	2905	6695	4605	2514	6006
Formal	7101	3731	8021	7055	3841	8248	7165	3421	7749
Informal	3330	2589	4528	3440	2701	4895	3165	2389	4118
Panel F: Wages (salaried empoyment)									
All	4926	3103	5804	4965	3208	6091	4865	2831	5463
Formal	5949	3562	6628	5980	3668	6888	5904	3219	6307
Informal	3169	2756	3588	3242	2823	3807	3058	2607	3352

Note: Sample: Those between ages of 15–65 in years 2007–2015. Source: Author's calculations using GEIH.

Table A.2: 2SLS estimates reported in main text

	Salaried Employment			Non-Salaried Employment			Employment		
	All	Informal	Formal	All	Informal	Formal	All	Informal	Formal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Pooled									
	-0.417***	-0.267***	-0.144***	0.179	0.023	0.006	-0.142	-0.311*	-0.071
	(0.111)	(0.110)	(0.044)	(0.173)	(0.169)	(0.073)	(0.275)	(0.183)	(0.112)
Panel B: Low-skill Men									
	-0.729***	-0.633***	-0.177***	0.927***	0.962***	0.015	0.095	0.131	-0.297***
	(0.129)	(0.169)	(0.052)	(0.355)	(0.343)	(0.028)	(0.328)	(0.309)	(0.076)
Panel C: Low-skill Women									
	-0.203***	-0.209***	-0.017	0.275	0.113	0.061*	-0.158	-0.144	0.030
	(0.044)	(0.041)	(0.012)	(0.497)	(0.454)	(0.037)	(0.488)	(0.492)	(0.042)
Panel D: Middle-skill Men									
	-0.709***	-0.289	-0.335***	0.083	0.037	0.023	-0.634***	-0.437**	-0.212*
	(0.245)	(0.223)	(0.088)	(0.157)	(0.156)	(0.095)	(0.263)	(0.198)	(0.110)
Panel E: Middle-skill Women									
	-0.576***	-0.251***	-0.188***	0.234	0.220	0.051	-0.421***	-0.229**	-0.077
	(0.068)	(0.102)	(0.064)	(0.200)	(0.160)	(0.043)	(0.104)	(0.109)	(0.087)
Panel F: High-skill Men									
	-0.393***	-0.186***	-0.219***	-0.066	-0.377***	-0.074	-0.510***	-0.616***	-0.047
	(0.129)	(0.057)	(0.093)	(0.125)	(0.093)	(0.107)	(0.158)	(0.106)	(0.107)
Panel G: High-skill Women									
	-0.410***	-0.285***	-0.017	-0.069	-0.055	-0.048	0.182	-0.027	0.169
	(0.089)	(0.068)	(0.058)	(0.173)	(0.160)	(0.065)	(0.316)	(0.142)	(0.261)

Table A.3: Model Fit

Size of firm	Data	Model	
2–3 workers	GEIH	2.5	2.5
4–5 workers	GEIH	4.5	5.0
6–10 workers	GEIH	8.0	8.6
11–19 workers	GEIH	15.0	14.9
20–30 workers	GEIH	25.0	23.6
31–50 workers	GEIH	40.5	37.7
51-100 workers	GEIH	75.5	62.5
Share of informality			
2–3 workers	GEIH	0.87	1.00
4–5 workers	GEIH	0.74	0.94
6–10 workers	GEIH	0.57	0.65
11–19 workers	GEIH	0.39	0.41
20–30 workers	GEIH	0.28	0.27
31–50 workers	GEIH	0.18	0.18
51-100 workers	GEIH	0.13	0.11
Ratio of demand elasticities	SIV estimates	-0.44	-0.46

B Work Permits

The *Permiso Especial de Permanencia* (PEP) was a work permit created by Colombia in 2017 specifically to address the large influx of Venezuelan migrants. It was unprecedented in its scope and ease of access (/generosity?). As will be detailed below, there were few if any requirements, and the application was completed online. Once approved (usually in under two weeks), recipients then simply had to print out their ID number. The permits were granted for 90 days but automatically renewed for a period of 2 years unless an infraction took place. Holders of a PEP could work legally in any capacity, but also access the public healthcare and educational system. For this reason, they were attractive for all ages. By the end of the program, over xx,xxx Venezuelans held a PEP.

On July 25th of 2017, the ministry of foreign affairs created the first PEPs under resolution 5797[17]. They were implemented by the office of migration *(Unidad Administrativa Especial Migración Colombia)* under resolution 1272 three days later. During this first wave of PEP, applications were received between August 3rd and October 31st 2017. Eligibility was restricted to Venezuelans who had entered the country legally before July 28th and did not have a criminal record. This is an important point. To enter Colombia legally, migrants needed to have passed through an approved port of entry and received a tourist stamp in their passport. Venezuelan passports are notoriously expensive, with some estimates putting the cost upwards of 200$.[18] This implies that most Venezuelans who hold a passport are middle/high skill I'm sure there's a better way to phrase this.

Resolutions 0740 and 0361, published on February 5th and 6th of 2018, implemented a second round of PEPs. These were available to Venezuelan migrants who, as before, had entered the country legally before February 2nd. The application window was open between February 7th and June 7th 2018.

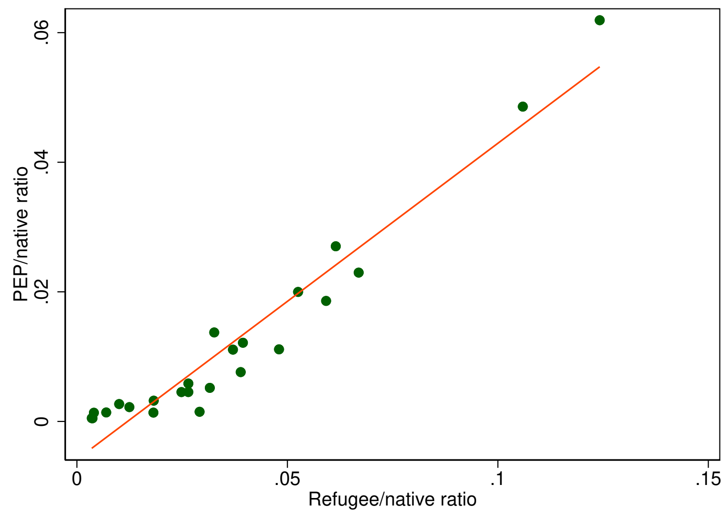
In March of 2018, the Colombian government set out to register undocumented Venezuelans for the stated purpose of designing better public policies (Decree 542). They called it RAMV (*Registro Administrativo de Migrantes Venezolanos*). This registry was open between April 6th and June 8th of 2018. During this period, surveyors travelled to areas with high concentrations of undocumented Venezuelans and carried out the registration in person. Most of these areas were close to the border. In the end, xx,xxx Venezuelans were included in the RAMV.

Decree 1288 of July 25th 2018, together with resolutions 6370 and 2033 of August 1st and 2nd, established that Venezuelans registered in the RAMV were also eligible for PEPs.

[17] All the laws, decrees, and resolutions can be found here

[18] Los pasaportes más caros y baratos de América Latina (y cuánto duran), BBC Mundo

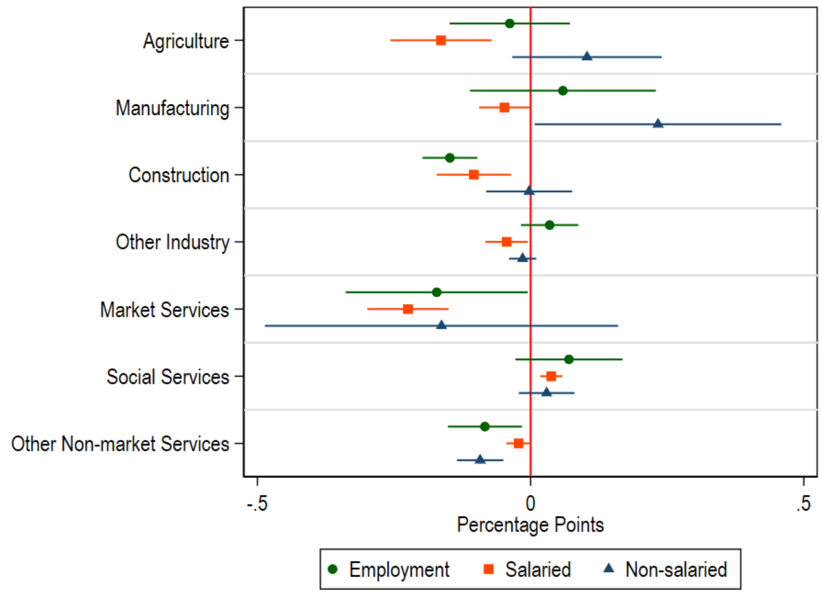
Figure B.2: Spatial correlation between work permit treatment and refugee treatment intensity



This brought about the third wave of PEPs. Those registered in the RAMV were invited to apply between August 2nd and December 21st of 2018.

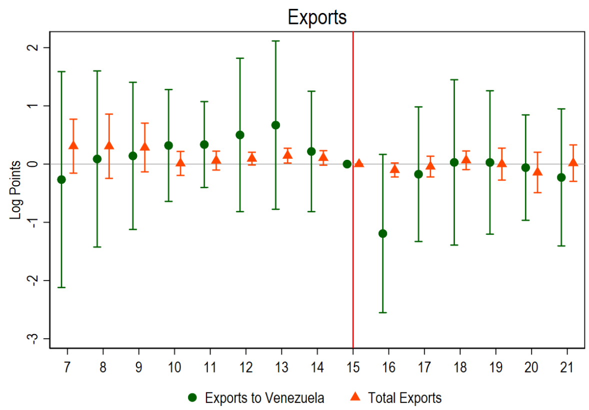
PEPs were first of their kind in providing formal labor market access to migrants easily and at low cost (add a citation here for how great and unique PEPs are - Danny maybe you have an article on this?). However, anecdotally, they encountered a series of difficulties. The first, was that many employers initially did not understand that PEPs operated as work permits. There was a big push by the government to educate the public about their functionality following the first wave of PEPs (citation needed). Secondly, there were some technological challenges in their implementation. For example, the ID number provided by PEPs was x digits long, while Colombian national IDs were x+y digits long. This meant that many systems could not process the migrant's new IDs. Finally, and most seriously, there began to be a lot of fraud. Since PEPs were simply printed from the migration office's website, some people began to replace the name of the recipient Basically what I'm trying to say here is that "gangs" would take print outs, replace the names and scan them again to create what looks like a PEP. These considerations led the government to replace the PEP program with the PPT (correct?) program. As part of this program, IDs were issued using a more sophisticated set of biometrics.

Figure C.3: Industry Heterogeneity of Refugee Effects

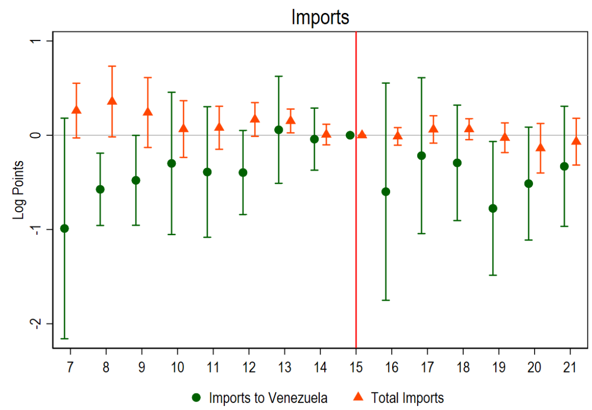


C Additional Results

Figure C.4: The evolution of trade flows in the region. Correlation between distance and trade



(a) Exports



(b) Imports

Figure C.5: Effect of distance on Internal migration

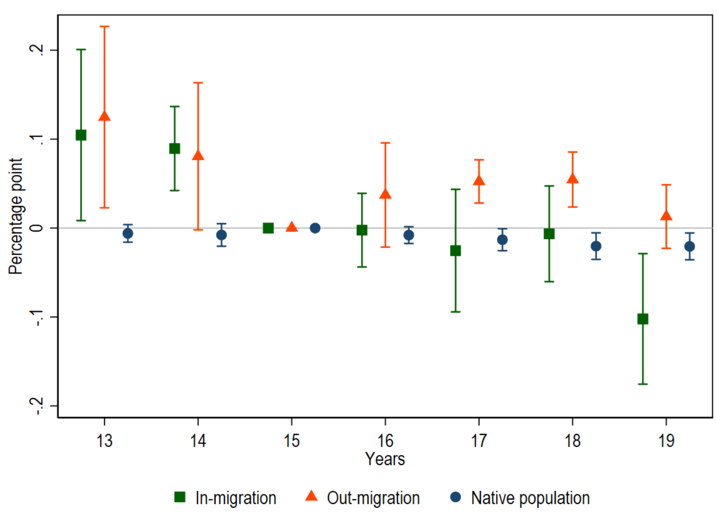
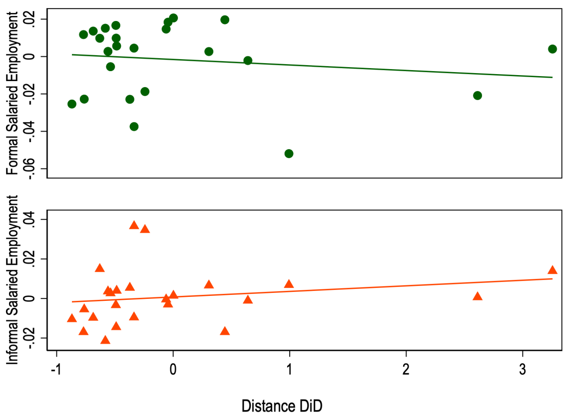
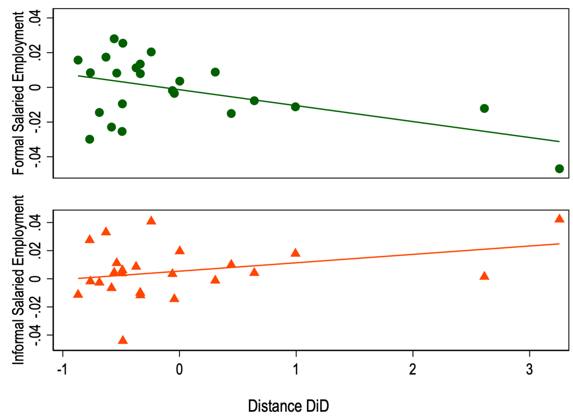


Figure C.6: Skill-sex DiD within regions



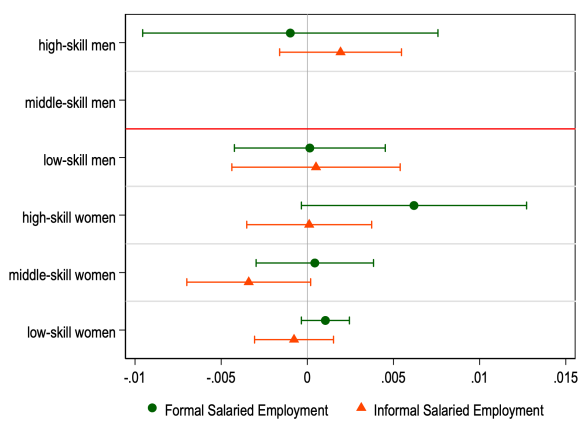
(a) High Skill Men vs rest



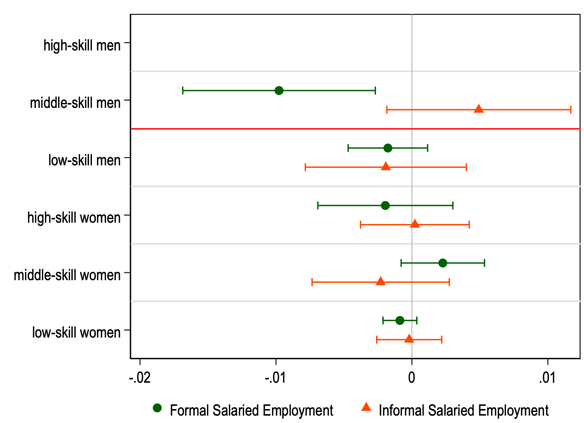
(b) Middle Skill Men vs rest

Notes: Each dot is a coefficient from a department-specific DiD estimation comparing employment outcomes across skill cells and time. To arrive at the triple DiD estimate we compare these estimates from regions closer to the border (left hand side of the x axis) to the estimates from regions further from the border (right hand side of the x axis).

Figure C.7: Spatial DiD within skill-sex groups



(a) High Skill Men vs rest



(b) Middle Skill Men vs Rest

Notes: Each row plots two coefficients from a skill-specific spatial DiD estimation comparing employment outcomes across regions close to and away from the border. Under the null that work permits had no impact, the estimates across rows would be the similar, i.e., lie on a vertical line. The triple DiD can be achieved by comparing the estimate of high-skill and middle-skill men against the average estimates of other skill-sex groups. Here, we also see that our triple DiD estimates are not driven by an outlier DiD estimate within the building blocks, for example, in one of the control groups.

D Synthetic Instrumental Variables

In this section, we explain briefly the idea behind the Synthetic IV methodology and how it works. we refer the reader to Gulek and Vives-i Bastida (2023) for a full treatment.

The synthetic IV (SIV) is a non-parametric method that combines instrumental variable strategy with synthetic controls. SIV partials out the unmeasured confounding in IV-DiD settings such as exposure or shift-share designs, and under standard conditions, is asymptotically normal when the standard two-stage least squares is not. We employ SIV in two steps.

Step 1: For each region i, we find synthetic control weights w_{ij} by solving the standard synthetic control program for past outcomes Y_{jt}. We use these weights to generate the synthetic outcome Y^{SC}, treatment R^{SC}, and instrument Z^{SC}

$$
\begin{aligned}
\hat{Y}_{it}^{SC} &= \sum_{j \neq i} \hat{w}_{ij}^{SC} Y_{jt}, \\
\hat{R}_{it}^{SC} &= \sum_{j \neq i} \hat{w}_{ij}^{SC} R_{jt}, \\
\hat{Z}_{it}^{SC} &= \sum_{j \neq i} \hat{w}_{ij}^{SC} Z_{jt},
\end{aligned}
$$

Then, we compute the *debiased* values.

$$
\begin{aligned}
\tilde{Y}_{it} &= Y_{it} - \hat{Y}_{it}^{SC}, \\
\tilde{R}_{it} &= R_{it} - \hat{R}_{it}^{SC}, \\
\tilde{Z}_{it} &= Z_{it} - \hat{Z}_{it}^{SC}.
\end{aligned}
$$

Step 2: We estimate the first-stage and the reduced-form using the debiased data.

The intuition behind SIV is simple. Step 1 creates a synthetic "control" for all the regions in the data that follow similar trajectories before the refugee shock begins. This addresses the "pre-trend" problem. However, it does not solve the problem that immigrants can choose their location based on contemporaneous economic shocks. This is still addressed by the instrument Z. Put differently, SIV addresses both the unobserved confounding problem via SC and the endogeneity problem via IV by combining both.

E Model explaining knowledge spillovers

The goal of this section is to rationalize the results from the quadruple difference in differences design in Section 5. There are R regions and J industries, each region-industry pair is populated with a representative firm, which uses low-skill and high-skill labor in production. For simplicity, we abstract away from informal and formal labor differences. Introducing a nested CES of informal and formal labor within each group would not change our predictions. Each firm's production function is given by

$$\begin{aligned} y_{rjt} &= f(H_{rjt}t, L_{rjt}, A_{H,rjt}, A_{L,rjt}) \\ &= A_{rjt}(A_{H,rjt}H^{\rho}_{rjt} + A_{L,rjt}L^{\rho}_{rjt})^{(1/\rho)} \end{aligned}$$

where A_{rjt} is a Hicks-neutral productivity measure for each firm, $A_{H,rjt}$ and $A_{L,rjt}$ are skill-enhancing productivity parameters for high and low-skill labor, respectively. Labor markets are competitive, workers are paid their marginal productivities. After some algebra, one can show:

$$ln(\frac{w_{H,rjt}}{w_{L,rjt}}) = ln(\frac{A_{h,rjt}}{A_{l,rjt}}) + (\rho - 1)ln(\frac{H_{rjt}}{L_{rjt}}) \tag{13}$$

This equation shows that the relatve wages of high-skill and low-skill workers in an industry-region cell depends on (1) the relative magnitudes of their productivitities and (2) their relative supplies.

In practice, both the relative technology measures and the relative skill intensities can evolve over time. For example, certain industries may become more productive over time, immigration can alter the skill ratios at the region level, etc.

Work permits can impact relative wages in two ways. First, absent work permits, high-skill refugees cannot work in certain industries due because of their high formality measures. Work permits enable them to alter their industry composition and work in more skill and formal intensive industries. This will change the relative supply of high and low-skill workers in a region-industry cell: the ratio of high to low skill labor in more formal intensive industries and in regions closer to the border will increase. Second, we acknowledge the possibility that the high-skill workers may not be purely a labor supply shock. Some refugees bring valuable human capital: specific know-how, ideas, etc. These can also change the relative productivities of high and low skill labor. In particular, we let the skill ratios and technologies in region-industry cells to evolve:

$$
\begin{aligned}
ln(\frac{H_{rjt}}{L_{rjt}}) &= \delta_{rt} + \delta_{jt} + \delta_{rj} + D_{rjt} + \epsilon_{rjt} \\
ln(\frac{A_{H,rjt}}{A_{L,rjt}}) &= g_{rt} + g_{jt} + g_{rj} + \theta_{rjt} + \eta_{rjt}
\end{aligned} \tag{14}
$$

where the fixed effects δ and g allow for the skill ratios and the productivities to differ at region-year, industry-year, region-industry level; D_{rjt} is the change in effective skill ratios due to work permits, θ_{rjt} allows for the possibility that high skill refugees impact the technology parameters in the exposed industry-region cells. To get at our main quadruple DiD design, plug in these terms into equation to get:

$$
ln(\frac{w_{H,rjt}}{w_{L,rjt}}) = \underbrace{f_{rt} + f_{jt} + f_{rj}}_{\text{F: Fixed effects}} + \underbrace{(\rho - 1)D_{rjt} + \theta_{rjt}}_{\text{E: exposure terms}} + \epsilon_{rjt} \tag{15}
$$

In practice, if we had only two skill groups, we could run this regression at region-industry-year level. However, we show that work permits impact high-school graduate and college graduate refugees at different times, which we exploit for additional statistical power. To obtain a regression equation at skill-region-industry-time level, we write:

$$
\begin{aligned}
& ln(\frac{w_{H,rjt}}{w_{L,rjt}} = \underbrace{f_{rt} + f_{jt} + f_{rj}}_{\text{F}} + \underbrace{(\rho - 1)D_{rjt} + \theta_{rjt}}_{\text{E}} + \epsilon_{rjt} \\
& \Rightarrow ln(w_{H,rjt}) = ln(w_{L,rjt}) + F + E + \epsilon_{rjt} \\
& \Rightarrow ln(w_{H,rjt}) = f_{rjt} + F + E + \epsilon_{rjt} \\
& \Rightarrow ln(w_{rjt}) = f_{rjt} + \mathbb{1}(skill = h)(F + E) + \epsilon_{rjt} \\
& \Rightarrow ln(w_{srjt}) = f_{rjt} + f_{srt} + f_{sjt} + f_{srj} + (\rho - 1)D_{srjt} + \theta_{srjt} + \epsilon_{srjt}
\end{aligned} \tag{16}
$$

where we let f_{rjt} to capture the wages of the baseline skill-level workers (i.e., low-skill workers), and define $D_{srjt} = D_{rjt}$ if $s = h$, 0 otherwise.

If work permits do not create any change in the technology through knowledge spillovers, then $\theta_{srjt} = 0$. In this case, we should see a decrease in relative wages of middle to high skill male natives, in more skill/formal intensive industries, in regions closer to the border, after the granting of work permits. The fact that we observe a positive slope in figure **??** suggests that $\theta_{srjt} > 0$, meaning that the high-skill refugees make the high-skill workers in the exposed region-industries more productive compared to low-skill workers. This is one way to interpret what we refer to as the knowledge spillover from reduction of skill mismatch in the main text.

F Model Extension: Nonsalaried employment

Non-salaried jobs constitute a significant part of the labor market in developing countries. In practice, these jobs can create an alternative to unemployment for salaried workers, which can impact our inference on the effects of immigration on labor markets. For example, if the natives displaced in salaried jobs transition to self-employment, then the overall employment rate of natives could be unimpacted by immigrants despite natives' being displaced in salaried jobs.[19]

To capture this intuition, we model the non-salaried jobs using the standard home production framework, following Gronau (1977) for simplicity. Individuals are endowed with time T, which they can use to allocate between leisure l, salaried employment h_s which pays constant wages w_s, and nonsalaried-employment (i.e., home production) h_n. Production from non-salaried work is given by the concave function f. Home production and market goods are perfect substitutes. Consumers get utility from leisure and what they consume: $U(c, l)$, they consume what they produce at home or buy at the market $c = f(h_n) + w_s h_s$, and are subject to time constraint: $T = l + h_n + h_s$.

Assuming an interior solution, which can be guaranteed with functional form assumptions, we get $f'(h_n) = w_s$: people work in non-salaried jobs until the marginal return from non-salaried work equals salaried work. As wages fall, for example due to an immigration shock, people transition to non-salaried jobs as f is strictly concave. How much people transition to non-salaried jobs depends on the (inverse of) curvature of the home production function f. Appendix Table F.4 shows that the concavity of f is higher for women and increases with education, which implies that for comparable changes in wages, the transition to non-salaried jobs would be higher for low-skilled men, which is what we find in the data.

We note that this simple model is just to highlight the overarching intuition: some people can transition from salaried to non-salaried jobs upon facing negative shocks in the labor market. If we had panel data on workers, we could track transitions between salaried employment, non-salaried employment, unemployment and out of labor force status. The transition matrix between these states would enable us to create a more realistic model that captures the incentives to remain in

the differences between home production and market production, whether non-salaried jobs reduce the probability of finding salaried jobs,

[19] For example, Gulek (2023) shows that low-skill Turkish natives who lose salaried jobs due to the arrival of Syrian refugees transition to non-salaried employment, leading to a null effect of Syrian immigrants on natives' overall employment.

Table F.4: The concavity of home production for different groups of natives

	Men	Women	Low-skill	Middle-skill	High-skill
$hours$	0.0437***	0.0671***	0.0530***	0.0525***	0.0659***
	-0.000242	-0.000213	-0.000195	-0.000267	-0.000395
$hours^2$	-0.000282***	-0.000507***	-0.000344***	-0.000346***	-0.000543***
	-0.00000203	-0.0000024	-0.00000184	-0.00000248	-0.00000425
N	761990	636745	710154	351996	336585

Note: The outcome is the natural logarithm of monthly income in all columns. Standard errors are in parenthesis. * p<0.10, ** p<0.05, *** p<0.01

G SIV Quality Checks

In this section we describe the robustness checks we perform for the SIV methodology. Gulek and Vives-i Bastida (2023) suggest three sets of checks to validate the robustness of the SIV methodology. We perform the robustness checks of the synthetic IV methodology suggested by Gulek and Vives-i Bastida (2023):

1. **Check your first stage**: "given that the debiasing procedure can lead to a weaker first stage, in cases with strong omitted variable bias if the synthetic IV estimator exhibits a weak first stage researchers may be worried using the synthetic estimator."

2. **Check your pre-treatment fit**: "if the debiased outcomes exhibits large deviations in the pre-treatment period or an event study design reveals pre-trends, it is likely that the synthetic estimator will be biased."

3. **Back test**: "given that the finite sample bias depends on the expected pre-treatment fit, back testing can reveal whether good pre-treatment fit was due to over-fitting (biasing the estimator) or not."

We perform all these robustness checks and report the results here.

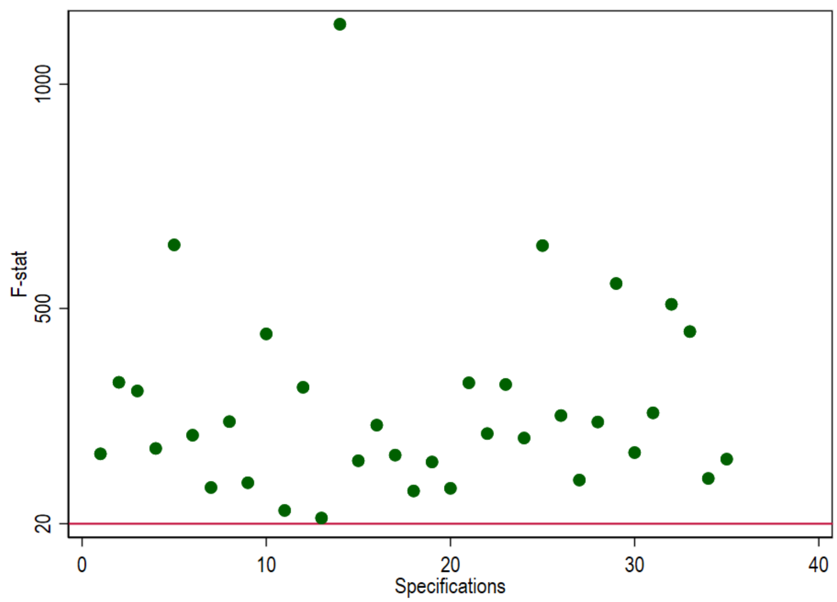
First, we need to ensure that a strong first stage remains after debiasing. For each of the 81 outcomes reported in Figure 7, we estimate equation 5 using debiased data after applying both baseline SIV and SIV with demeaned outcomes as explained in the text. Figure **??** shows the t-stats of the individual coefficients of the instrument interacted with the year dummies in 2017, 2018, and 2019, and Figure **??** shows the joing significance of these three coefficients for each specification. Our debiasing procedure not only maintains the positive first-stage coefficients, i.e., (debiased) inverse distance predicts more (debiased) refugee treatment, but the coefficients remain significant, with the smallest t-stat being 5 out of 486 estimates, and the smallest joint F-statistic being 59 with the 5th percentile is 98.

Second, we present the event-study figures using IV and SIV for all the outcomes reported in Figure 7. Figures G.10, G.11, and G.12 present the event study estimates on salaried, non-salaried, and overall employment rate of natives using IV and the two versions of SIV.

Third, we match on deviations from pre-treatment mean instead of levels in the synthetic control problem and repeat the main analyses. For almost all outcomes, this provides qualitatively similar results to baseline SIV.

Fourth, we perform backtesting. Instead of using the 2007–2015 values of the outcome in the matching process, we use values from 2010–2015. Figure **??** compares results reported in Figure 7 reported in the main text with the version using a different training set. Overall, results are highly comparable across the two specifications.

Figure G.8: F-statistic of the debiased instrument



F-statistic of the instrument is calculated by estimating the equation $\widetilde{R_{it}} = \beta\widetilde{Z_{it}} + f_i + f_t + \eta_{it}$, where $\widetilde{R_{it}}$ and $\widetilde{Z_{it}}$ are the debiased versions of the treatment and instrument, respectively. f_i and f_t are region and time fixed effects. Standard errors are clustered at the region level. F-statistic of the instrument $\widetilde{Z_{it}}$ from the 35 main specifications shown in Figure 7 are plotted.

Figure G.9: Event-study estimates of Immigrants' impact on salaried employment of natives (IV vs SIV vs SIVdm)

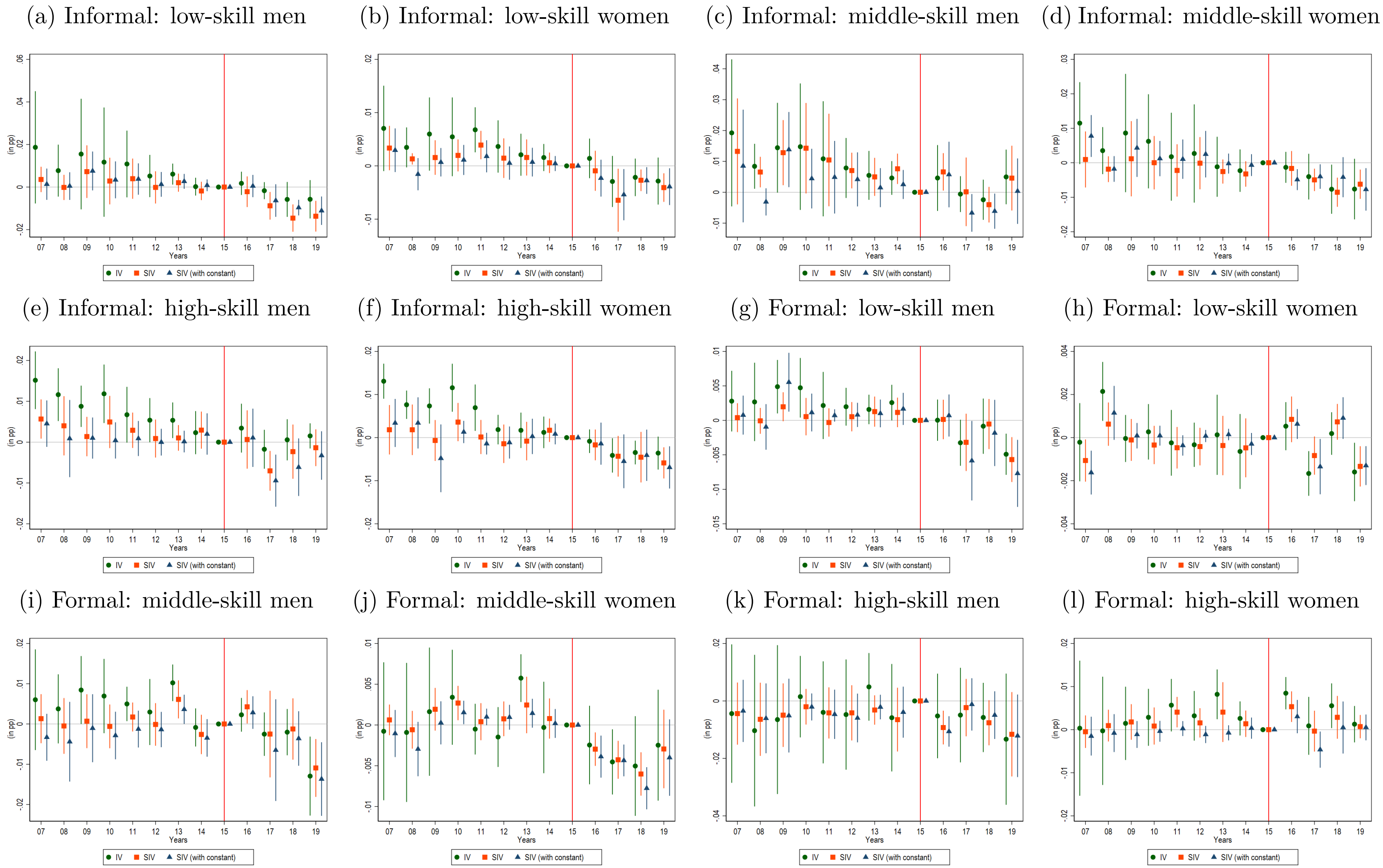
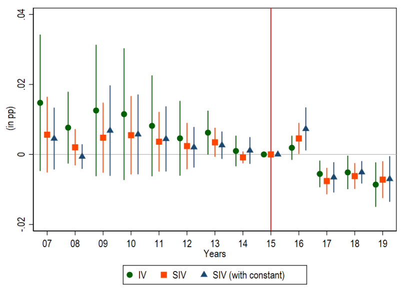
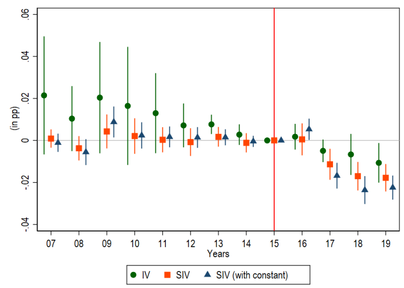


Figure G.10: Event-study estimates of Immigrants' impact on salaried employment of natives (IV vs SIV vs SIVdm)

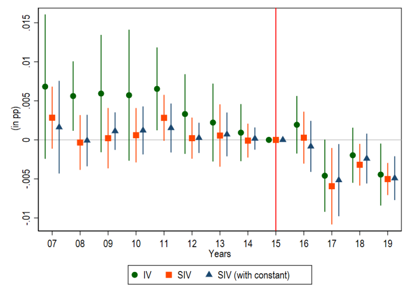
(a) Pooled



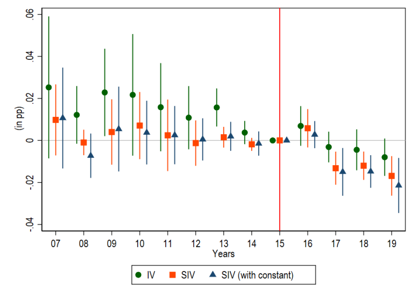
(b) Low-skill men



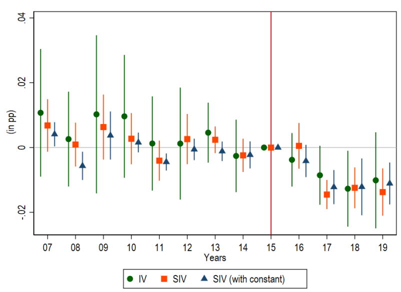
(c) Low-skill women



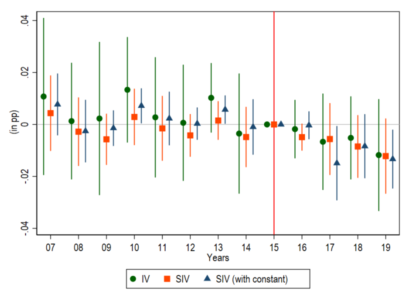
(d) Middle-skill men



(e) Middle-skill women



(f) High-skill men



(g) High-skill women

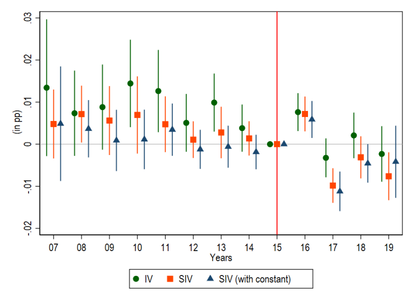
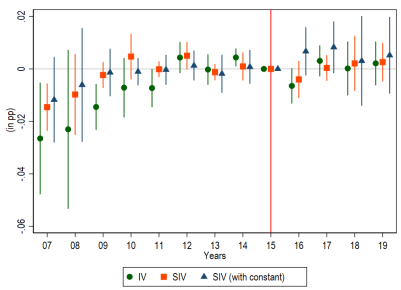
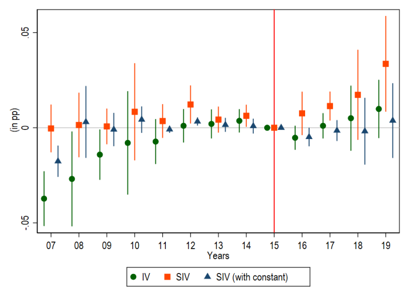


Figure G.11: Event-study estimates of Immigrants' impact on Non-Salaried employment of natives (IV vs SIV vs SIVdm)

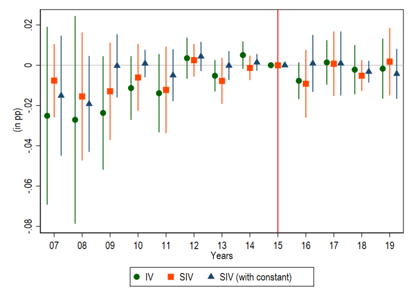
(a) Pooled



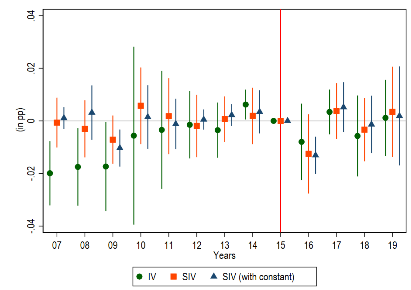
(b) Low-skill men



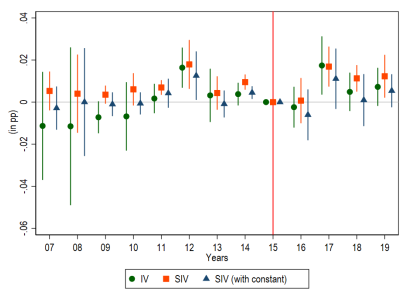
(c) Low-skill women



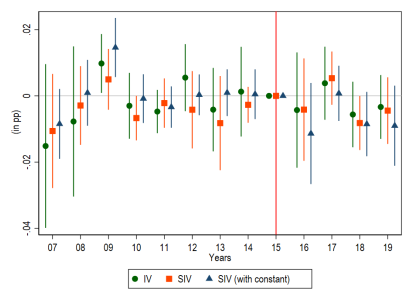
(d) Middle-skill men



(e) Middle-skill women



(f) High-skill men



(g) High-skill women

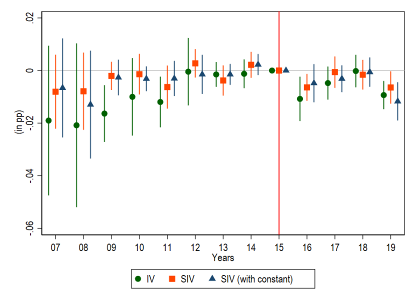
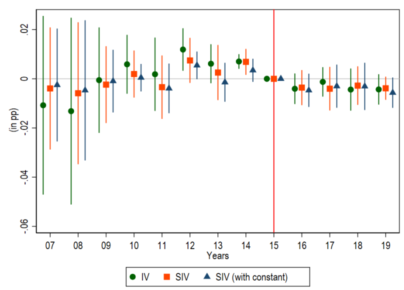
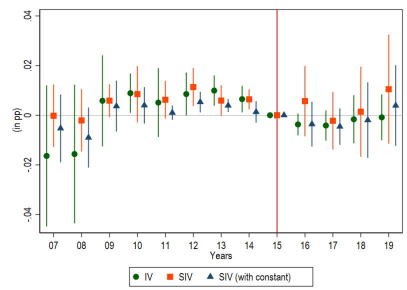


Figure G.12: Event-study estimates of Immigrants' impact on Employment of natives (IV vs SIV vs SIVdm)

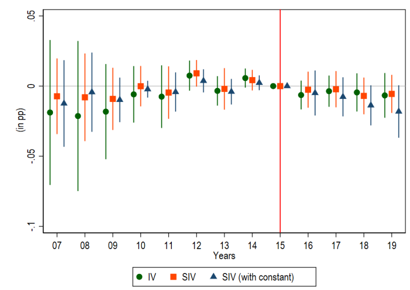
(a) Pooled



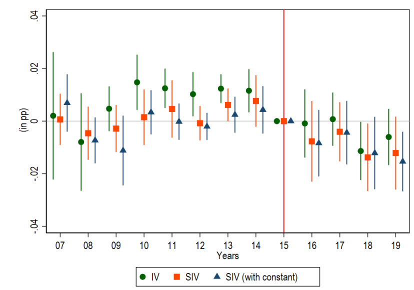
(b) Low-skill men



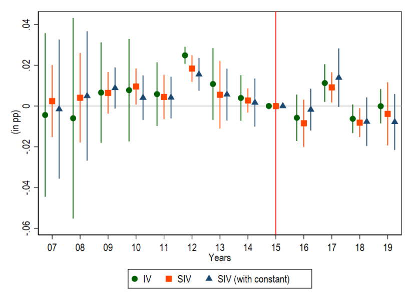
(c) Low-skill women



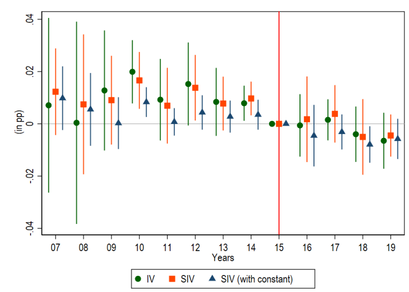
(d) Middle-skill men



(e) Middle-skill women



(f) High-skill men



(g) High-skill women

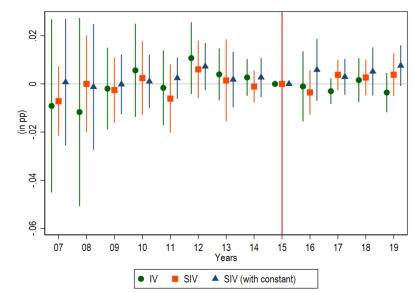
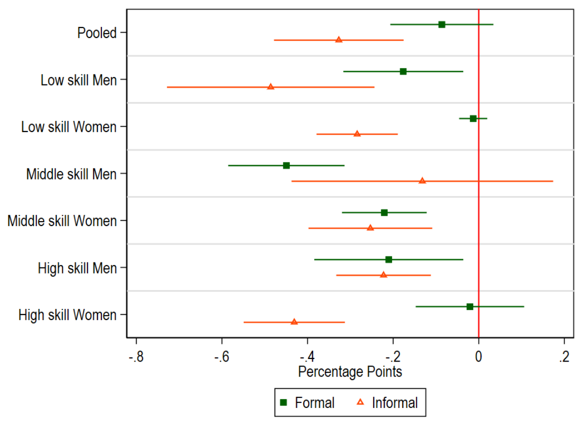
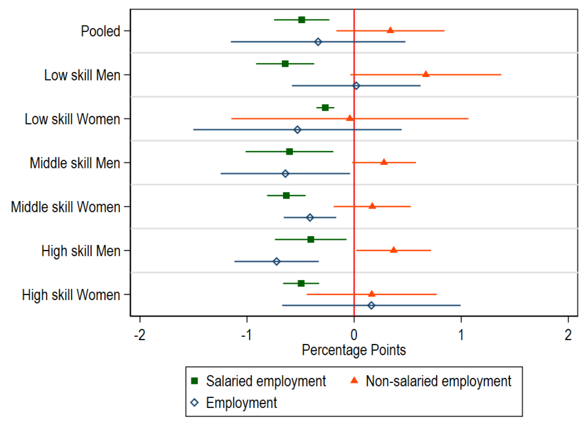


Figure G.13: LATE estimates: Training period 2007–2012

(a) Informal vs Formal



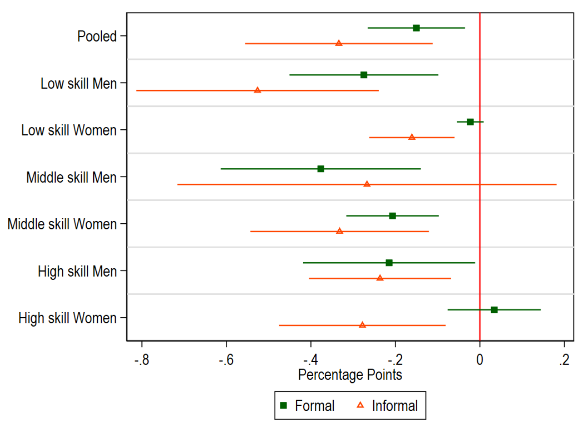
(b) Salaried vs Non-salaried



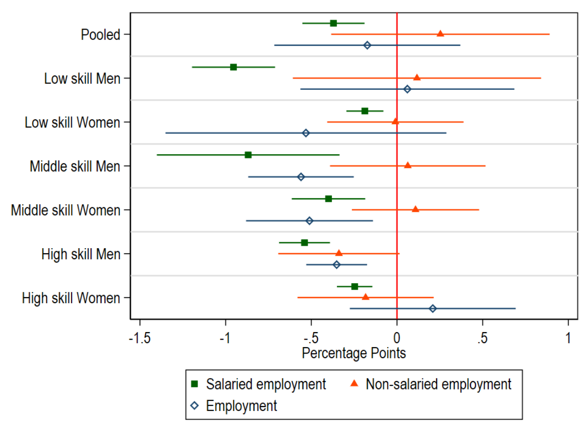
The estimates come from the Synthetic IV methodology described in text. Panel A shows the estimated effects on the informal and formal salaried employment rates across all skill levels, and Panel B shows the estimated effects on salaried and non-salaried employment rates. Standard errors are clustered at the region level.

Figure G.14: LATE estimates: demeaned SIV (SIVdm)

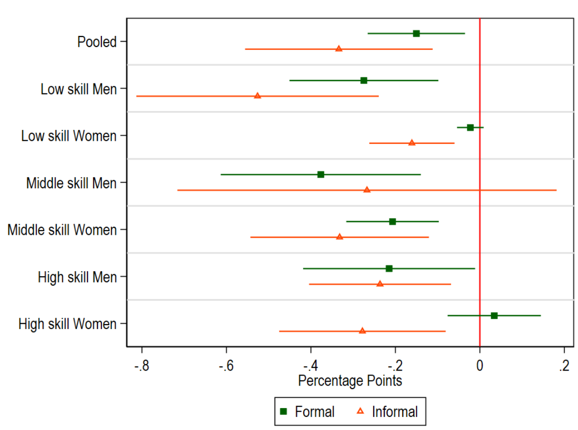
(a) Informal vs Formal (SIVdm 07-15)



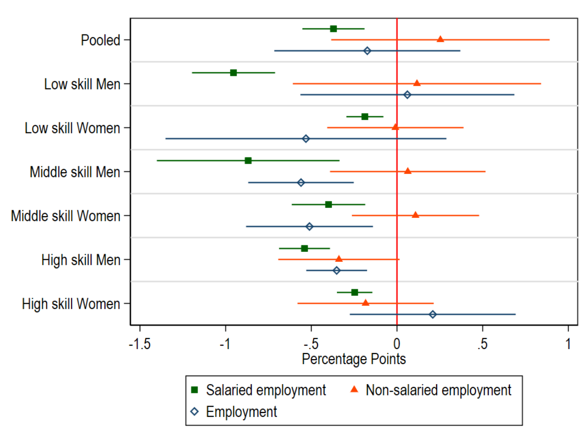
(b) Salaried vs Non-salaried (SIVdm 07-15)



(c) Informal vs Formal (SIVdm 07-12)



(d) Salaried vs Non-salaried (SIVdm 07-12)



The estimates come from the Synthetic IV methodology described in text. Panel A shows the estimated effects on the informal and formal salaried employment rates across all skill levels, and Panel B shows the estimated effects on salaried and non-salaried employment rates. Standard errors are clustered at the region level.