Misallocation of the Immigrant Workforce: Aggregate Productivity Effects for the Host Country*

José Pulido† and Alejandra Varón‡

First version: September 2020
This version: March 2023
[Latest version here]

Abstract

Immigrants often face more obstacles in finding employment in their desired occupations compared to native workers. These barriers can result in misallocation of the immigrant workforce, which, in the case of large-scale migrations, can decrease aggregate labor productivity in the host country. We examine this phenomenon in the context of Colombia between 2015 and 2019, a period of mass migration from Venezuela. Using a model of occupational choice that accounts for discrimination and additional barriers preventing workers from selecting their preferred occupations, we quantify the extent of immigrants' misallocation and its impact on Colombian aggregate labor productivity. Our findings indicate that both frictions significantly misallocate Venezuelan immigrants. Removing them could increase Colombian productivity by 0.9%, making the contribution of immigration to Colombian economic growth up to 29% larger. Our study thus highlights the importance of reducing barriers for immigrants in the labor market, to promote productivity and growth in the host country.

Keywords: Immigration, misallocation, Roy model, discrimination, productivity.
J.E.L. Classification: F22, O15, J61, O24.

---

*The views expressed in this paper are exclusively ours and do not necessarily represent those of Banco de la República's Board of Governors. We would like to thank Hernando Vargas, Juan Ospina, Juan Castro, Emilio Fernández and Daniel Wills for helpful comments and suggestions; Sushanta Kumar Mallick for handling our manuscript, and to our associate editor and two anonymous referees for their valuable comments that significantly enhanced the quality of our paper. We also thank seminar participants at Banco de la República, Colombian Ministry of Finance and Public Credit, IADB Immigration Workshop and Andes-Rosario-Javeriana Universities Macro-Workshop. All errors are ours.

†Corresponding author, Banco de la República, Colombia. Permanent address: Cra 7 No. 14-78, 7th floor, Bogotá D.C. E-mail: jpulidpe@banrep.gov.co.

‡Universidad de los Andes, Colombia. E-mail: ma.varon10@uniandes.edu.co.

1 Introduction

The impacts of immigration on host countries' labor markets have typically been analyzed in terms of immigrants' assimilation process or the potential displacement effects on native workers.[1] However, an overlooked issue with possible macroeconomic implications is the effect of immigrants on the aggregate productivity of the host country through their impact on total labor misallocation. If immigrants face more barriers in finding jobs in their desired occupations compared to natives, then the overall labor misallocation across ocupations could rise due to immigration. As a result, there could be a decrease in aggregate labor productivity due to the loss of efficiency in the allocation of labor.[2] In cases where there is a large influx of immigrants, this fact may have sizable implications for the macroeconomic performance of the host country.

The objective of this paper is to evaluate this phenomenon in the case of Colombia between 2015 and 2019, a period when the country experienced a massive influx of migrants from Venezuela, its primary neighboring nation. During this time, the Latin American region saw a mass exodus of over 4.5 million Venezuelans as a result of their country's economic and governance crisis. Colombia became the principal destination for these migrants, receiving approximately two million immigrants by 2019, which increased the country's workforce by a significant 4.6%. We assess whether the occupational misallocation of immigrants from Venezuela in the Colombian labor market is effectively larger compared to that of non-migrants, and derive the implications of such labor misallocation for the Colombian aggregate labor productivity.

Using data from the Colombian household survey, we first document several facts that suggest a higher degree of occupational misallocation for migrants from Venezuela compared to non-migrants. Despite having, on average, more years of education than non-migrants, immigrants tend to work in occupations that require lower levels of education.[3] Moreover, even after controlling for differences in educational attainment and other observable characteristics, we find significant residual income gaps for immigrants. These gaps arise due both to the composition effect, where observationally equivalent immigrants work in lower-paying occupations relative to non-migrants, and to the presence of within-occupations gaps. The gaps persist even for new hires and after controlling for local working experience, and are

[1]For an extensive literature review of the empirical studies until the the early 1990s see Borjas (1994); and for the empirical studies for more recent years, see Kerr and Kerr (2011).

[2]The literature on resource misallocation emphasizes that, given some amount of factors endowments, the micro-level misallocation of these factors across heterogeneous uses generates sizable losses in aggregate TFP. For an extensive review of this literature from the perspective of heterogeneous firms, see Restuccia and Rogerson (2013) or Hopenhayn (2014).

[3]This educational mismatch, where migrant workers are "overeducated" (i.e., their schooling level is greater than is typical for workers in each occupation) is commonly observed in the cases of high-skilled migration. For a summary of evidence, see McDonald and Worswick (2015) and Borjas et al. (2019).

more pronounced in the formal sector, where barriers to finding desired jobs, such as a lack of documentation or recognition of educational degrees, are more likely to bind. Furthermore, these gaps are time-variant and positively correlated with the proportion of immigrants in the workforce.

Recent literature has suggested that residual-income gaps alone may not be sufficient to determine the presence of frictions in a context where workers self-select into different sectors or occupations (Hsieh et al., 2019; Pulido and Święcki, 2020). However, when taken together with the observed occupational allocations, the residual income gaps can discipline a structural model to quantify the additional frictions immigrants face that prevent them from working in their preferred occupations. Using this model, we can also estimate the aggregate implications of these frictions. For this reason, we continue by guiding our analysis through the lens of a standard Roy's (1951) model of occupational choice. Within this framework, each worker is endowed with a certain amount of human capital, obtained before the period of migration, and draws different levels of non-observable skills across occupations. The worker then chooses the occupation that maximizes her indirect utility, which depends on her real consumption and thus on the value of her efficiency units of labor, given her skills and human capital endowment.

We consider two types of frictions that may prevent migrant workers from choosing their preferred occupations. First, we introduce the possibility of pure discrimination in the labor market. Following Hsieh et al. (2019), this discrimination takes the form of an occupation-specific wedge between marginal products and wages, which is intended to capture the standard formulations of employer taste for discrimination in the literature (Becker, 1971; Altonji and Blank, 1999). Second, we allow for the possibility that even after accounting for implicit discrimination, immigrants may be forced to make involuntary occupational choices in a different proportion than natives.[4] This may be due to additional potential obstacles that immigrants face in finding jobs in their desired occupations, including a lack of professional connections or networks, issues with the recognition of educational degrees, or difficulties in obtaining permits to work legally. Similar to Pulido and Święcki (2020), we represent the extent of these barriers indirectly by assuming that a fraction of workers in each period is randomly assigned to work in occupations other than their desired ones, and by allowing these fractions to be different between native and immigrant workers. Our specification of frictions aims to capture the two primary obstacles that immigrants report in opinion surveys when asked about the main difficulties in finding a job: the lack of required documentation to perform the job and the perceived discrimination due to their nationality (DANE, 2021).

We identify the extent of both types of frictions using the implications of our self-selection model for the residual income gaps of immigrants. In the model, the gaps depend on both the relative occupational allocations of immigrants, reflecting how workers sort across occupa-

[4] In our model, we use the word "natives" to refer to the group of non-migrants. But it is worth to mention that in earlier years, a significant proportion of immigrants were return migrants (born in Colombia).

tions, and the frictions each group of workers faces in the labor market. Hence, by combining information from the residual income gaps and the allocation of labor for both natives and immigrants across occupations, we can quantify the extent of those frictions. Our findings suggest that discriminatory wedges have a considerable dispersion among occupations, indicating potential gains from reallocating workers across occupations when those wedges are removed. Furthermore, we find that the proportions of immigrants in each period who are forced to make involuntary choices tend to grow over time, coinciding with the rise in the immigration rate, increasing from 5% in 2015 to 9% in 2019. These values are larger than those obtained for native workers in all years (4.7%).

Once we have inferred our frictions, we conduct two counterfactual exercises to evaluate their quantitative importance. In the first exercise, we entirely remove each set of frictions for immigrants. This is a drastic reform that enables all immigrants to choose their occupations according to their efficient allocation, but it allows us to understand the relative importance of each set of frictions. Our findings suggest that removing both types of frictions entirely would lead at least one-third of immigrants to reallocate. Moreover, as it triggers general equilibrium effects, it also causes a small reallocation of natives, which amounts to around 0.4% of their workforce. The reallocation of the entire workforce would raise total output by as much as 0.9% in 2019, due to the increase in aggregate labor productivity as a result of the gains in allocative efficiency. Compared to the total impact of immigration to the Colombian output in our model that year (3.1%), our results suggest that this type of reform would boost the contribution of immigration to Colombian economic growth by 29% (i.e., 0.9% out of 3.1%). By decomposing the contribution of each type of frictions, we find that discrimination accounts for around two-thirds of the total gains from the extreme reform of eliminating all frictions. Furthermore, discrimination involves stronger general equilibrium effects than the frictions that force workers to choose random allocations.

Our second counterfactual aims to equalize immigrants' frictions with those inferred for natives. By this way, this reform provides a rough estimate of the aggregate productivity gains that could result from fully assimilating immigrants into the labor market, i.e., enabling them to face the same extent of barriers to finding a job as native workers. Thus, this exercise offers valuable input for policy analysis, allowing us to assess the potential implications of different programs aimed at helping immigrants compete in the labor market under similar conditions as natives. Our results suggest that reducing immigrants' frictions to a level similar to that inferred for natives would reallocate at least 9% of immigrants, resulting in an increase in aggregate productivity of up to 0.4% (2019). Compared to the total impact of immigration to the Colombian output, this type of reform would have led to a 13% upsurge in the growth of aggregate output attributable to immigration (i.e., 0.4% out of 3.1%).

Finally, we show that the macroeconomic gains resulting from our counterfactual exercises are robust even when considering non-trivial variations in the calibrated parameters and alternative model specifications. These specifications include allowing for differences in innate talent distribution across groups, accounting for time-variant discriminatory wedges, and controlling for local working experience in the vector of observables. Our analysis also reveals that the gains resulting from our proposed reforms, particularly the one that ensures immigrants face the same extent of barriers as native workers, are larger when improving the misallocation of immigrant labor in the formal sector.

Related literature

Our paper relates to a recent literature that quantitatively evaluates the role of misallocation of heterogeneous workers across sectors, locations or occupations in a context of self-selection (Lagakos and Waugh, 2013; Hsieh et al., 2019; Bryan and Morten, 2019; Pulido and Święcki, 2020; Adamopoulos et al., 2022). This quantification is usually tackled with a framework that encompasses a Roy's (1951) type model augmented by micro-level frictions. We apply this setting to study the extent of occupational misallocation of immigrants and its consequences on the host country's aggregate productivity, but our approach can be extended to study discrimination between more groups of workers, for example. Mainly, our specification borrows elements from the models used by Hsieh et al. (2019) and Pulido and Święcki (2020) for studying the allocation of all workers across occupations and sectors, respectively. As in these papers, we emphasize that reduced-form findings in income-gaps alone are not enough statistics to distinguish between sorting and misallocation. Hence, to make inferences about the extent of misallocation is necessary to nourish the empirical findings with the guidance of a structural model.

In the migration literature, although the educational mismatch of immigrants to jobs within local labor markets in cases where migrants are more educated on average than natives is well documented (Chiswick and Miller, 2011; Nielsen, 2011; Joona et al., 2014; McDonald and Worswick, 2015; Borjas et al., 2019), the exploration of the macroeconomic implications of this type of mismatch (or similar misallocations) has been less studied. Some recent studies address the consequences for allocative efficiency of the degree of mismatch of immigrants and local firms. For example, using French data, Orefice and Peri (2020) document that since immigrants have a larger dispersion on productivity, they increase the positive assortative matching between firms and workers. In the same vein, Burzynski and Gola (2019) show that immigration may trigger a similar sorting mechanism in the host country using a model with two types of workers who draw country-specific skills from different sets. Our paper relates to those studies but uses a different modeling approach. Instead of assuming immigrants are intrinsically different from natives because their skills are drawn from distributions with different attributes, we assume immigrants are misallocated due to the inherent frictions they face in the destination labor market. In the light of our studied episode of migration, where immigrants are very similar to natives, our modeling approach seems to be more appropriate. And, in the other way around, our paper also relates to the concurrent study of Birinci et al.

(2021), that uses an approach similar to ours, but to quantify the aggregate implications of immigrants misallocation in several developed countries.[5] The fact that most of the migration to developed countries shares the features of 'South-North' migrations, where immigrants and natives importantly differ in their backgrounds (and thus likely have different dispersions in their unobserved skills) makes the omission of assortative matching mechanisms in both studies less problematic in our case.

In broader terms, our paper enriches the migration literature that explores macroeconomic effects in the host country beyond the usual short-term impacts on the labor market outcomes for natives. Particularly, we contribute to a body of work looking at the effects of immigration on aggregate productivity (Peri, 2012; Lewis, 2013; Hornung, 2014; Ortega and Peri, 2014; Aleksynska and Tritah, 2015). This literature usually highlights channels such as changes in the demographic composition of the workforce, shifts in firms' production functions, and the boost to innovation and total factor productivity growth; see Nathan (2014) for a review. Compared with these studies our research focuses instead on a pure allocative mechanism, i.e. the way immigration shapes how total labor is allocated across heterogeneous occupations, a channel that has sizable impacts on aggregate productivity, as the recent misallocation literature suggests.

Our study also relates to two strands in the migration literature studying wages setting in the host country. First, to the studies that explore the roots of the observed residual-income disparities between immigrants and natives.[6] Those gaps are often associated to causes such as the lack of host country-specific human capital (e.g. language proficiency), barriers to access to the labor market or ethnic discrimination. In cases where migrants are culturally close to natives, such as the one studied here, host country-specific human capital is relatively easy to acquire (Isphording and Otten, 2014; Ingwersen and Thomsen, 2021), so we omit this channel in our model. Instead, we focus on frictions that can be rationalized as pure discriminatory preferences and barriers that impede individuals to work in their preferred occupations. Studies with evidence supporting the existence of discrimination include Rydgren (2004) for Sweden, Oreopoulos (2011) for Canada, and Weichselbaumer (2017) for Austria.[7] Second, although in a more subtle way, our paper also relates to the literature on the impact of immigration on natives' wages (see Dustmann et al., 2016 for a comprehensive review). The magnitudes of the effects found in this literature are usually dependent on the degree of

[5]Birinci et al. (2021) use a similar model of sorting than ours, although with only discriminatory wedges. For the US, they find that by decreasing immigrants' wedges to the level of their native counterparts, GDP would increase by around 4.4 percent.

[6]Generally part of the total income disparities between the two groups are due to observable differences (e.g. education or experience). However, usually a fraction of the gaps remain once observables are controlled for. As we stated earlier, these residual-income gaps are the main object of interest for us.

[7]A common finding in this literature is that although statistical discrimination (due to stereotypical thinking) seems to contribute to the phenomenon, there is also a component due to employers' preferences against working with minority group members, the way we choose to model discrimination.

labor substitutability (or complementary) between immigrants and natives. Our model, that is expressed in terms of efficiency units of labor, and which thereby factors out differences in human capital and other observables, considers within-occupation labor demands in which both efficiency units of labor are perfect substitutes. This approach is in line with: i) literature that does not find evidence to reject perfect substitution once all observables are controlled for (Borjas et al., 2011, 2012); ii) the fact that an important part of the explanation of why natives and immigrants can be imperfect substitutes in production is due to their different allocation across occupations (Peri and Sparber, 2009); and iii) arguments in favor of imperfect substitutability between natives and immigrants (e.g. natives have comparative advantage in communication skills) are invalid in cases where immigrants are culturally close to natives, such as the one studied here.

Finally, this study belongs to a recent collection of recent papers which have used the Venezuelan exodus to assess its consequences for the Colombian economy: Peñaloza (2022), Caruso et al. (2019), Santamaria (2022), Bonilla-Mejía et al. (2020), Bahar et al. (2021), Lebow (2022a,b), Rozo and Vargas (2021), Knight and Tribín-Uribe (2023) and Lebow et al. (2023). With exception of the latter three works, that explore how Venezuelan migration alters electoral outcomes, has incidence in crime events and changes trust towards foreigners, respectively, the remaining works mainly investigate the consequences for the Colombian labor market, focusing on natives and immigrants wages, unemployment and participation rates. Their results, in line with the findings in the migration literature, show relatively minor displacement effects from immigration in terms of employment for native workers; being the moderate effects mainly due to adjustments in the informal segment.[8] However, there are negative hourly wage effects concentrated on less educated natives, with magnitudes that vary across studies, see Lebow (2022b) for a comprehensive discussion.

The organization of this paper is as follows. Section 2 presents our empirical motivation. Section 3 introduces our Roy model of occupational choice with two groups of workers and frictions for immigrants. Section 4 discusses the procedure that allows us to infer the magnitude of the frictions and presents our baseline results. Section 5 performs our counterfactual exercises of suppressing frictions for migrants, and equating them to those of native workers. We also evaluate some departures from the baseline model. Finally, section 6 concludes.

2 Empirical motivation and data

The roots of the "Venezuelan exodus" lay in the country's economic and political turmoil that began at the end of the presidency of Hugo Chávez and was exacerbated during the presidency of Nicolás Maduro. Both governments were characterized by the implementation of

[8]Further, Bahar et al. (2021) show that granting work permits to Venezuelan immigrants does not seem to have generated short run impacts on labor outcomes.

a series of socialist reforms that included economic policies such as land expropriations, nationalizations, price and currency controls and systematic restrictions on private businesses (Vera, 2015; Gutiérrez S., 2017). These policies, coupled with political mismanagement and an international outlook in 2014 with a downfall in oil prices (Venezuela's main export commodity), led the country to suffer by 2015 the worst economic crisis in its history. The crisis was marked by hyperinflation and shortages of food and medicine and looting; that led to an escalation of starvation, disease, crime and mortality rates; a combination of factors that triggered Venezuelan migration (Mauricia, 2019; O'Neil, 2019).

According to the UN Refugee Agency (UNHCR) from 2015 to 2019 an estimated 4.5 million people fled Venezuela.[9] Colombia, Venezuela's main neighbor, was by far the main receptor of Venezuelan migrants. According to the Colombian household survey, where our data comes from (for a detailed description of our database see Appendix A.1), in 2019 the country hosted around 2 millions of Venezuelan migrants.[10] For the Colombian labor market, with a size of around 23.6 million economically active population before the migration started, this massive inflow of migrants implied a significant expansion of the workforce. Figure E.1 in Appendix E shows the inflow of Venezuelan migrants and their participation in the workforce. Since 2015 immigrants' share in the workforce have steadily risen, to reach 4.6% in 2019.

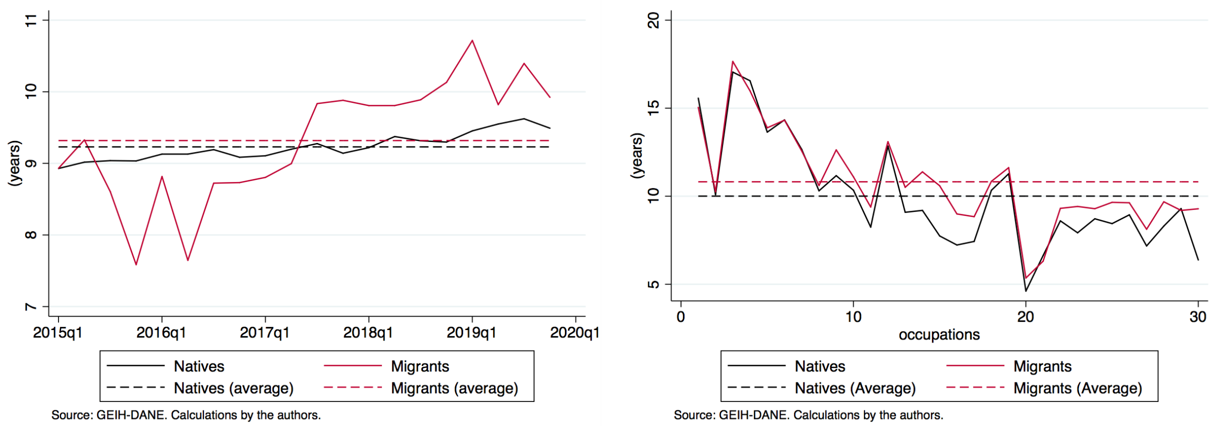
Unlike "South-North" migrations, immigrants in this case have a cultural and socio-economic background that is closely similar to that of the natives: Both countries share the same language, have comparable demographics, and, at the outset of the migration exodus, had comparable average incomes, among other attributes. On average, Venezuelan migrants in Colombia report even more years of education than natives, a pattern that was accentuated since 2017. Table D.1 in Appendix D displays some demographic characteristics (shares of males and average age and years of schooling) for immigrants and non-migrants, both in the whole household survey and in our restricted sample, which includes only employees. Differences in average schooling years are evident in those two those samples but are even present only for adults 25 years and older (see Panel A of Figure 1), a comparison that aims to control for the age composition of the migrant population, which is more biased towards people in productive ages.

[9] Relative to other major migration waves seen in recent history, the magnitude of Venezuela's migration is only smaller than the originated by the Syrian war (5.6 million).

[10] It is worth to say that given the large amount of irregular migrant inflows, (i.e. immigrants without legal documentation), household surveys would offer a more accurate picture about the dimension of the phenomenon than records of the migration authorities. Nevertheless, a comparison of the estimates between both sources, in which migration authorities compute the amount of irregulars by imputation procedures, show similar magnitudes for the total migrants flows (Tribín et al., 2020).

Figure 1 – Average Years of Education: Immigrants and Natives

(a) Over Time **(b)** By Occupations

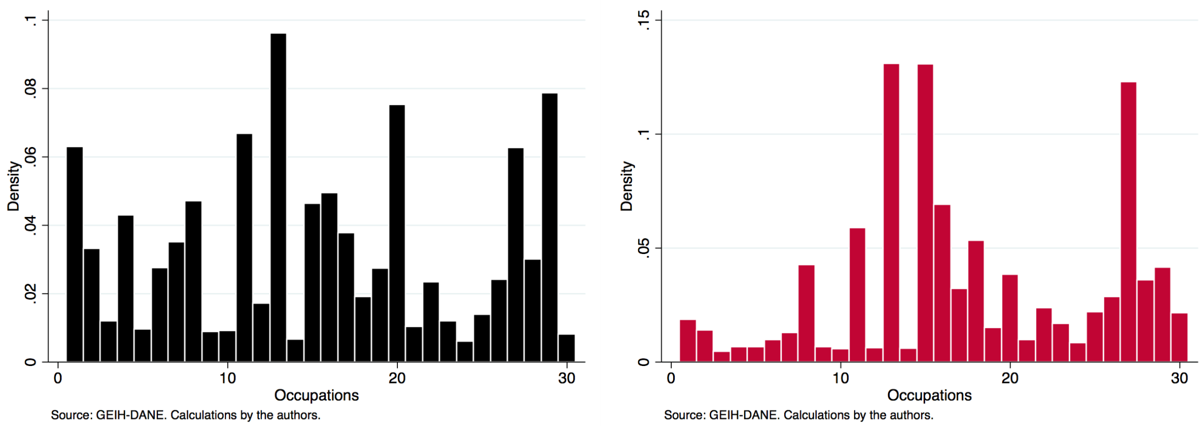


Notes: Panel A shows the average years of education of natives and immigrants aged 25 years and older in each quarter of the period 2015-2019 and the corresponding averages over all quarters. Panel B shows the average years of education of natives and immigrants in each occupation (for the pooled data) and the corresponding averages over all occupations. Observations are weighted by survey expansion weights.

In spite of their higher educational attainment, Venezuelan migrants tend to work in occupations with lower requirements of skills relative to non-migrants. We consider the 30 most representative occupations in the survey (for a description of the occupations see Appendix A.2). Broadly speaking, the average years of education of workers in each occupation tends to decrease with the value of the occupation code: the first quintile of codes is related to high-skilled or "cognitive" occupations, whereas the fifth quintile is related to low-skill or "manual" occupations. Immigrants have higher levels of education than non-migrants across most occupations (see Panel B of Figure 1), but, compared to non-migrants, their occupational allocation is more concentrated in the middle and the right side of the distribution, i.e. into occupations other than those with high skills requirements (Figure 2).

Figure 2 – Occupational Distribution

(a) Natives **(b)** Immigrants



Notes: Figure shows the occupational distribution of Natives (Panel A) and Immigrants (Panel B) in the pooled data. Observations are weighted by survey expansion weights.

We also find there are important and significant residual (controlling for observables) income gaps for immigrants. These gaps are estimated by using Mincerian regressions with the following general form:

$$lnY_{islt} = X_{it}\beta + \phi I_i + D_l + D_t + \varepsilon_{islt} \tag{1}$$

where Y_{islt} denotes a measure of labor income of individual i in occupation s, province l, and quarter t; X_{it} refers to a series of individual controls that include gender, work experience (proxied by the age of individual at time t minus 15 years), work experience squared and in most specifications years of education, plus an indicator of whether the household is in a rural area; D_l and D_t are province and time fixed effects; and I_i is an indicator of whether individual i is migrant, so ϕ captures the migrant premium of interest. Robust standard errors are clustered at the municipality level. Our preferred measure of labor income is one that includes both wages and fringe benefits for salaried workers, and net-profits from personal business in the case of non-salaried or self-employees.[11]

Columns (1)-(3) of Table 1 report the results from estimating equation (1) using our preferred measure of labor income. Before factoring out differences in educational attainment, column (1) shows that immigrants on average perceive a residual labor income 39.8 log points [lp.] (or 49%) lower than non-migrants. Once we compare workers with similar years of education, column (2) shows that the premium decreases to 33.6 lp. (or 40%), a magnitude still considerable.[12]

[11]Fringe benefits include extra-legal premiums, overtime pay, transportation subsidies, etc. Appendix A.3 shows a summary of the different possible measures of labor income that can be obtained from the survey.

[12]Notice that although immigrants have on average more years of education, the premium in the residual income decreases when we control for education. This outcome is due to the composition of the migrants

Table 1 – Immigrant Income Premia

	(1) Log income for all workers	(2) Log income for all workers	(3) Log income all workers	(4) Log income for formal workers	(5) Log income for formal workers	(6) Log income for new hires	(7) Log income for new hires	(8) Log income for all workers	(9) Log income all workers
Migrant	-0.398***	-0.336***	-0.217***	-0.405***	-0.277***	-0.238***	-0.123***	-0.240***	-0.121**
	(0.068)	(0.057)	(0.043)	(0.034)	(0.031)	(0.0464)	(0.0320)	(0.0652)	(0.0441)
Gender	0.421***	0.530***	0.431***	0.337***	0.215***	0.535***	0.403***	0.530***	0.432***
	(0.048)	(0.072)	(0.087)	(0.059)	(0.047)	(0.0806)	(0.0820)	(0.0726)	(0.0875)
Experience	0.057***	0.054***	0.049***	0.066***	0.061***	0.0544***	0.0507***	0.0467***	0.0420***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.00137)	(0.00173)	(0.00250)	(0.00283)
Experience sq.	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Education		0.111***	0.077***	0.112***	0.075***	0.080***	0.056***	0.112***	0.077***
		(0.003)	(0.003)	(0.002)	(0.002)	(0.006)	(0.003)	(0.003)	(0.003)
Local experience								0.008***	0.008***
								(0.002)	(0.001)
Observations	1,502,645	1,502,537	1,502,537	758,321	758,321	408,188	408,188	1,502,537	1,502,537
R-squared	0.186	0.346	0.401	0.359	0.406	0.251	0.306	0.346	0.401
Location FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Occupation FE	NO	NO	YES	NO	YES	NO	YES	NO	YES

Notes: Columns (1)-(3) show the results of regression (1) for all workers using different sets of controls.Columns (4) - (5) show the results of columns (2) - (3) only for salaried workers, and Columns (6) - (7) only for new hires. Columns (8) - (9) controls also for a proxy of local working experience. Observations are weighted by survey expansion weights. Standard errors are clustered by municipalities in parentheses. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

A natural question is whether the latter premium is explained only by the different allocation of immigrants across occupations, or whether non-migrants are also paid more in the same occupations. For this, column (3) of Table 1, reports the premium controlling for occupation fixed-effects. On average, the within-occupation premium is 21.7 lp., decreasing 12 lp. with respect to the premium without controlling for occupation fixed-effects. The fact that the within-occupation premium exists, but it is significatively lower than the premium in column (2), suggests that both explanations play a role. Hence the income-gap for immigrants is consequence not only of the composition effect of more immigrants with similar observables working in occupations with lower remunerations relative to non-migrants, but also of the presence of within-occupations premia. Table D.3 shows that these within-occupation premia, which are going to play a key role in identifying the magnitude of both frictions in our structural model, are heterogeneous across occupations and statistically significant (at 5% level or lower) in the case of 26 of our 30 occupations.

Four additional facts are worth noting. First, the reported income gaps are not only the result of a different allocation of immigrants between the formal (salaried) and the informal (non-salaried) sector,[13] but also of the existence of intra-sectoral premia. For example, consider only the formal sector, where the labor income measure might be more accurate. Columns (4)-(5) of Table 1 use only salaried workers and show that the reported premia remain and their magnitudes are even magnified in the formal sector. This result is not surprising since, in light of our findings, part of the income gaps are due to the additional frictions that immigrants face. Such barriers might be more pronounced in the formal sector, for example, in the case of the recognition of educational degrees, or the difficulties to obtain permits to work legally. We recognize this fact in an alternative specification of our baseline structural model, in which we segment the labor market and its frictions between the formal and informal sectors.

Second, the residual income gaps are apparent even among recently hired workers. One possible explanation of the observed income gaps is that Venezuelan immigrants are not able to find a job that matches with their skills simply because the aggregate labor demand does not keep pace with the aggregate labor supply, so there are not enough jobs to go around for them.[14] Certainly, the rapid influx of Venezuelan immigrants suddenly increased the aggregate labor supply of workers, generating a downward pressure on wages for new hires. Given that immigrants are disproportionately represented among new hires,[15] the observed income gaps in the overall workforce may stem from lower wages that new hires receive.

population across observables, effect that is absorbed by the whole set of controls. Without any control, the premium for migrants in total income is 27 lp., while controlling only for education is 5 lp. higher.

[13] In Colombia, in spite of the fact that the rate of informal work has decreased (at least until the end of 2019), informal workers still account in 2019 for almost half of the labor force.

[14] We thank to an anonymous referee for bringing this remark to our attention, as well as the one mentioned subsequently.

[15] In the pooled data of workers, immigrants from Venezuela represent 2.1% of the total workforce (4.7% for 2019 only), while in the pool of new hires, they represent 4.7% of the hires (10.5% for 2019 only).

However, by restricting the sample to workers who have been in their current job for less than 12 months, serving as a proxy for new hires, Columns (6)-(7) of Table 1 indicate that the income gaps persist, although at a lower magnitudes than the initial ones. The premium without controlling for occupation fixed effects decreases by 29%, from 33.6 lp. to 23.8 lp, while the within-occupation premium falls 43% (from 21.7 lp. to 12.3 lp.). So although the excess of supply of workers and the disproportionate number of new hires who are immigrants could play a role explaining the observed income gaps in the whole set of workers, the persistence of economically significant gaps among newly hired employees suggests the presence of some type of frictions that immigrants face within the labor market.

Third, the residual income gaps remain after controlling for a proxy of working experience only in Colombia. Since it is likely that local working experience is more valued by employers than the experience gained in other countries, the observed income gaps can be due to immigrants having less local working experience compared to natives. In this case, the native income premia would not be a consequence of a specific friction in the labor market (such as discrimination); instead, they would reflect how the market values different workers' attributes. Unfortunately, the Colombian household survey does not provide a direct measure of total working experience, it only inform us about the experience in the current job. So we construct a proxy of local working experience using for natives the same variable of experience as before (based on age); and an imputation for immigrants from any country (regardless they migrated from Venezuela or not) based on the question of whether the immigration was recent (during the last 12 months) or over the previous five years, and the experience in the current job for recent immigrants.[16] We control for this proxy in Columns (7)-(8) of Table 1. The premium in column (2) decreases by 28%, from 33.6 lp. to 24.1 lp., while the within-occupation premium falls even more (44%, from 21.7 lp. to 12.4 lp.). Nonetheless, residual income disparities persist and remain economically significant, indicating that while local work experience is indeed valued by the market, there exist income differentials against immigrants that cannot be explained by this factor alone.[17]

Finally, the residual income gaps are time-variant and correlate with the magnitude of migration. Figure E.2 in Appendix E shows the evolution of the immigrant premium for each cross-section in our data and its 95% confidence interval. The residual income gap is not statistically significant until the end of 2016. Starting in early 2017, and coinciding with the sudden increase in the migration inflows, the immigrant premium becomes statistically significant, and its magnitude begins to increase over time, to start to stabilize around 0.21 lp.

[16] Particularly, for recent foreigner immigrants we impute as their local working experience the experience in their current job, while for foreigner immigrants during the last five years we impute the maximum number between the experience in their current job and a random number between 1 and 5 years. We also use a similar imputation scheme to impute data for returned Colombians from abroad. Admitted not ideal, this imputation strategy is the best we can do given the limitations of our data.

[17] We evaluate how robust are the baseline results of our structural model when we consider local working experience as other observable worker's attribute.

in 2019. The temporal evolution of the residual income gaps suggests a correlation between the rate of immigration and the magnitude of the gaps, indicating the need to consider time-varying frictions in our structural model.

Taken together, our findings point to the possibility that immigrants, relative to natives, might have been facing some type of frictions that prevent them from working in their preferred occupations, and that originate the observed income gaps even within occupations. Our aim next is to identify whether these frictions exist and to quantify their magnitude and impact on the host economy. Existing research on migration highlights several potential difficulties that immigrants face when seeking employment in their destination countries, including challenges in obtaining legal work permits, difficulties in having their educational degrees recognized, discrimination, and limited professional networks, among other factors. In the Colombian case, a recent survey conducted by the official statistics bureau in Colombia (DANE) among more than 8500 Venezuelan immigrants show that 60% of the immigrants in the labor force report difficulties to find a job (DANE, 2021). Among those who report obstacles, the two main reported barriers were not having the required documentation to perform the job (69% of respondents confirm this difficulty) followed by perceiving discrimination due to their nationality (24% of affirmative answers). See Figure E.3 in Appendix E for a comprehensive breakdown of responses to all the considered difficulties.

Hence, we will examine two specific types of frictions that we believe can account for the primary reported challenges. Firstly, we consider discrimination, which, akin to Hsieh et al. (2019), will manifest as a discrepancy between the wages earned by immigrants and their marginal productivities. Secondly, following the indirect approach of Pulido and Święcki (2020), we will force that some immigrants, even after acknowledging their discrimination hurdle, make involuntary occupational choices. These forced allocations can arise due to a diverse set of barriers, including the absence of necessary documentation, the main obstacle reported by immigrants in the survey. It is worth noting that our primary focus will be on the impediments to finding the desired job, so immigrants will be considered as "job takers" instead of entrepreneurs. This is because, given the compelled nature of the Venezuelan exodus to Colombia, where 92% of the immigrants state that their reason for migration is to improve their economic prospects (DANE, 2021), they are unlikely to possess the resources to be entrepreneurs.[18] Although the importance of entrepreneurship in the immigrant population has been established in other migration scenarios (e.g., Azoulay et al., 2022 for the U.S.), it appears to hold less relevance in the context of the Venezuelan exodus in Colombia.

In an economy where workers self-select into different occupations, our two primary empirical findings, namely, differences in the allocation of immigrants across occupations and the presence of residual income gaps, cannot, on their own, provide sufficient evidence for the

[18] For example, the incidence of entrepreneurship among Venezuelan immigrants, proxied by the number of individuals who identify as employers and work in firms with more than five employees, is very small (0.09%), and much lower than that of natives (0.49%).

existence and magnitude of the frictions in question. To see this, consider for example the case of discrimination for immigrants in a given occupation. Under self-selection, discrimination could deter migrant workers with low unobservable abilities to enter such occupation. Only immigrants with high enough abilities to overcome the discrimination hurdle will accept to work; a smaller fraction of the immigrant workforce relative to what would be without discrimination. Since the average quality of immigrants skills will be higher in this small fraction of their workforce, this composition effect could offset the direct effect of the discrimination wedge on the observed migrant premium. Therefore, we need to carefully consider the implications of self-selection for jointly the immigrants' occupational allocations and their earnings gaps, in order to identify the extent of the frictions immigrants face. The next section present a Roy's (1951) type of model which purports to accomplish this objective.

3 Theoretical model

What can our reduced-form findings tell us about the extent of frictions that immigrants face? To answer this question we introduce in this section a simple discrete-time Roy model of occupational choice with two groups of workers (immigrants and natives[19]) and two types of frictions: discrimination and involuntary choices. We first present our model with no frictions. The frictionless economy resembles a particular case of Hsieh et al.'s (2019) general equilibrium model,[20] so we try to keep a similar notation for comparison. Next, we show how to generalize this basic framework to introduce each type of our frictions.

3.1 Frictionless Economy

A continuum of workers choose an occupation i at each time t from a set of M available occupations, to maximize their contemporaneous utility.[21] There are two groups of workers: immigrants and natives, indexed by $g \in G = \{I, N\}$. Hence, workers can be characterized by the occupational choice i they make and the group g they belong. Workers are endowed by unobservable heterogeneous abilities ϵ_i over occupations, and possess an amount of human capital h_{igt} at time t that is given by:

$$h_{igt} = \overline{h}_{ig} a_{igt}^{\gamma} s_{ig}^{\phi_i} \tag{2}$$

where $\overline{h}_{ig}$ represents permanent differences in human capital or "talent" common to the group g in a given occupation i; γ captures the return to experience, a_{igt}, that we assume is simply

[19]The word "natives" will refer to the group of non-migrants, but it is worth to mention that in the earlier years a significant proportion of immigrants from Venezuela were return migrants (born in Colombia).

[20]The case of a frictionless economy for a single-cohort with no heterogeneity in preferences.

[21]We abstract here from modeling inter-temporal choices given the short period of time we analyze and the lack of panel data to make inference. Turning the model into a dynamic one could greatly complicate the analysis without affecting our main insights.

the age of individual at time t minus 15 years; s_{ig} are the years of education, which we assume are fixed for all individuals prior to the migration period; and finally we let the returns to education to vary across occupations, with magnitude given by ϕ_i. Since both ages and years of education are observables, we collapse both variables in $x_{igt} \equiv a_{igt}^{\gamma} s_{ig}^{\phi_i}$, a term that we refer as the "returns" from observables.[22]

For analytical tractability we borrow from Eaton and Kortum's (2002) model of trade and assume abilities draws ϵ_i are drown from a multivariate Fréchet distribution:

$$F(\epsilon_1, ..., \epsilon_M) = \exp\left[-\sum_i^M \epsilon_i^{-\theta}\right] \tag{3}$$

The parameter θ measures the dispersion of abilities of workers, with a higher value of θ corresponding to a smaller dispersion. The mean parameter of the Fréchet distribution is normalized to 1, but this parameter is isomorphic to $\overline{h}_{ig}$.

Denote y_{igt} the income that a worker receives for her labor supply at time t, equal to the value of her efficiency units of labor. This value is the product of the price per efficiency unit of labor in occupation i at time t, w_{it}, and the amount of efficiency units of labor, which in turn is the product of the worker's human capital, given by equation (2), and the worker's idiosyncratic talent ϵ in her chosen occupation i:

$$y_{igt} = w_{it}\epsilon_i\overline{h}_{ig}x_{igt} \tag{4}$$

For the formulation of workers' utility we allow for general societal preferences for specific occupations, similar to compensating differentials. Thus, workers contemporaneous utility U_{igt} is simply the product of their consumption at time t, c_{igt}, and a parameter z_{igt} that measures the common utility benefit of all members of society from working in occupation i. For identification we normalize the value of this parameter to 1 in a given occupation so the values of z_{igt} are compensating differentials relative to the reference occupation. The worker's problem is thus to choose her occupation at the beginning of period t that maximizes her contemporaneous utility:

$$V_{igt} = \max_i \{U_{igt}\} = \max_i \{z_{igt}c_{igt}\}, \tag{5}$$

where in our static formulation consumption is simply equal to income, given by equation (4).

Finally, since our main interest is to study the effects of frictions for the allocation of heterogeneous workers across multiple occupations, we abstract from firm heterogeneity and instead assume that a representative firm produces final output Y from workers in M multiple

[22] Given the empirical evidence in Section 2 suggesting a role of local working experience in explaining the immigrants' residual income gaps, in a robustness check of our baseline specification, we include our proxy for this variable as an additional observable.

occupations according to a CES technology:

$$Y_t = \left\{ \sum_i^M \left[A_{it} \sum_g^G q_{gt} p_{igt} \mathbb{E}\left(h_{igt}\epsilon_{ig}\right) \right]^{\frac{\sigma-1}{\sigma}} \right\}^{\frac{\sigma}{\sigma-1}}, \tag{6}$$

where A_{it} is the exogenous productivity of occupation i at time t, q_g is the total amount of workers in group g at time t, σ is the elasticity of substitution across occupations, p_{igt} is the share of workers of group g who choose occupation i at time t and $\mathbb{E}\left(h_{igt}\epsilon_{ig}\right)$ is a measure of the average quality of workers of group g who choose occupation i at time t. The latter two terms have an explicit solution given our functional form choice for the abilities draws ϵ_i, as we will show in Section 4. Notice that, the production function in (6) implies that, for a given occupation, an efficiency unit of labor of a native is a perfect substitute to that of an immigrant. As we stated in the introduction, this is in line with the literature that does not reject perfect substitutability between natives and immigrants when observables are controlled for; and with the fact that many arguments in favor of imperfect substitutability apply when immigrants have comparative advantage relative to natives, a setting that is more likely when immigrants and natives differ in their skills or cultural backgrounds, as opposed as our case.[23]

Finally, the competitive equilibrium of the model is shown in Appendix B.1.

3.2 Type I of frictions: Discrimination against immigrants

Our first case of frictions is discrimination against immigrants in the labor market. Similar to Hsieh et al. (2019), discrimination takes the form of an occupation-specific time-invariant wedge between immigrants's marginal products and their wages. This wedge can be derived as a function of discriminatory preferences of employers, following the literature on discrimination (Becker, 1971; Altonji and Blank, 1999). Assuming only immigrants face discrimination, workers' income in equation (4) becomes:

$$y_{igt} = \left(1 - \tau_{ig}\right) w_{it} \epsilon_i \overline{h}_{ig} x_{igt} \tag{7}$$

with $\tau_{ig} = 0$ if $g = N$. In this way, discrimination works as a "tax" on immigrants earnings, where the fraction of income taxed is equal to $\tau_i \in [0, 1]$. The fact that wedges are heterogeneous across occupations generate implications for selection of immigrants workers across occupations and their income gaps, as we will explain in section 4.

[23] Lebow (2022a) finds that in the case of the Venezuelan exodus to Colombia, there is an important degree, although bounded, of immigrant-native substitutability among low educated workers, the more homogeneous segment. His study obtains an elasticity of around 15, comparable in magnitude with a long-run elasticity estimated in the U.S. of around 20, suggesting some degree of imperfect substitution. However, his specification of the labor demand is not within-occupations as here, and labor is not expressed in terms of efficiency units.

3.3 Type II of frictions: Involuntary occupation choices

The second type of frictions can be thought as barriers for immigrants that force some of them to work in occupations different to their preferred ones, even after taking into account the presence of discrimination that wedges imply. These frictions might reflect additional obstacles that immigrants face to find a job in their desired occupations, including lack of professional connections or networks, issues with the recognition of educational degrees or difficulties to obtain permits to work legally. We want to capture the idea that immigrants, relative to natives, have larger probabilities of not getting to work in the occupation they would like even if they have a strong comparative advantage in the occupation. In a richer setting, such type frictions could be rationalized by, for example, more unfavorable job search conditions for immigrants or higher search costs (Liu, 2010; Chassamboulli and Peri, 2015).

Following Pulido and Święcki (2020) we model the extent of these barriers indirectly by assuming that a fraction of workers are forced to make involuntary occupational choices, and by allowing this fraction to be possibly different between immigrants and natives. So we simply assume that at the beginning of each period every worker gets a random draw, such that a worker will be able to choose the occupation they desire with probability $1 - \alpha_g$, and they will be forced to work in any other occupation, selected randomly, with probability α_g. We allow for α_g to be time-variant for immigrants (so we will refer α_g as α_{gt}, keeping in mind $\alpha_{Nt} = \alpha_N \forall\ t$) to reflect the fact that this type of frictions could depend, for instance, on how sluggish their labor market is, which in turn would depend on the size of the immigration rate; or on the introduction of reforms that help to regularize immigrants.[24] The implications of α_{gt} for the allocations of immigrants across occupations and their income gaps over time are outlined in the next section.

4 Inference procedure

In this section we describe how the presence of both types of immigrants' frictions distort their occupational allocations and lead to income gaps, and hence how we can identify the extent of these frictions given data on income gaps and occupational shares for immigrants relative to natives. First, we present the implications of both types of frictions for both occupational allocations and wage premia under our framework. Next, we describe our inference procedure to quantify the extent of frictions from data and comment on our baseline results.

4.1 Occupational shares and wage premia

Denote $\widetilde{w}_{igt} \equiv (1 - \tau_i)\, w_{it}\overline{h}_{ig}x_{igt}z_i$ the overall "reward" that someone from group g with the mean ability obtains by working in occupation i at time t, so the worker's problem at each

[24] For example, in August 2018 the Colombian Government introduced the PEP program, a large scale reform that regularized approximately half a million immigrants (Bahar et al., 2021).

time t is to choose the occupation with the largest value of $\widetilde{w}_{igt}\epsilon_i$. Further, denote $\widetilde{p}_{igt}$ the share of workers of group g at time t that without forced choices, would choose occupation i. Proposition 1 refers to the occupational shares and the average ability of workers in each occupation.

Proposition 1. *The share of workers of group g who work in occupation i p_{igt} is given by:*

$$p_{igt} = (1 - \alpha_{gt})\,\widetilde{p}_{igt} + \alpha_{gt} M^{-1} \tag{8}$$

where $\widetilde{p}_{igt} = \frac{\widetilde{w}^{\theta}_{igt}}{\sum_s \widetilde{w}^{\theta}_{sgt}}$. Further, the geometric average of abilities of the group g in an occupation i at time t is given by:

$$\widehat{\epsilon} = \widetilde{\Gamma}\left(\frac{1}{\widetilde{p}_{igt}}\right)^{\frac{1}{\theta}(1-\delta_{igt})} \tag{9}$$

where $\widehat{x}$ denotes the geometric average, $\widehat{x} \equiv \exp^{\mathbb{E}(\log x)}$; $\delta_{igt} = \frac{\alpha_{gt}}{M p_{igt}}$ is the share of workers within an occupation i who do not voluntary chose such occupation; and $\widetilde{\Gamma} \equiv e^{\frac{\gamma_{em}}{\theta}}$, with γ_{em} the Euler–Mascheroni constant.

Proof. See Appendix B.2. □

Occupational shares of a group given in (8) are a weighted average between the random allocation of the fraction α_{gt} of workers who cannot make voluntary choices, and the allocation of the fraction $(1 - \alpha_{gt})$ of workers that can work in their preferred occupations, given by $\widetilde{p}_{igt}$. In turn, this latter allocation depends on the average reward $\widetilde{w}_{igt}$ relative to their power mean over all occupations ($\sum_s \widetilde{w}^{\theta}_{sgt}$). Thus, in the case in which all desired occupational choices were feasible ($\alpha_{gt} = 0\ \forall t$), occupational allocations would depend only on the relative returns of occupations, so the differences across allocations between natives and immigrants would come only from group-occupation specific factors. That is, the price per efficiency unit of labor in a given occupation (w_{it}), which is common between groups for each occupation, would not cause differences between immigrants' occupational shares relative to natives in a world in which all the desired choices are feasible. Only discrimination wedges, group-specific permanent differences in human capital and group-specific preferences for a given occupation, would cause differences in the allocations of immigrants. These forces as determinants of occupational allocations start to be distorted once involuntary choices are introduced, and lose explanatory power the larger the extent of forced choices.

Self-selection induced by voluntary choices affects the average quality of workers in an occupation. To see this, notice that equation (9) implies that the geometric average of abilities in a given occupation is inversely related to the share of the group working in such occupation, p_{igt} (of which $\widetilde{p}_{igt}$ is direct function). Thus, in occupations where a group has low participation, for instance, because workers are discriminated against, workers will have higher abilities on average (in our example, because they are the ones who can overcome the discrimination hurdle). The presence of this sorting effect is stronger the larger are the allocations due to voluntary choices. In the extreme scenario where all choices were involuntary, $\alpha_{gt} = 1 \; \forall \; g, t$, this selection effect would not be present and their average ability would be the unconditional mean of the draws of abilities, given by $\widetilde{\Gamma}$ in the case of the Fréchet distribution.

The results for both occupational shares and average abilities in Proposition 1 lead to a direct implication in terms of the within-occupation income gaps for immigrants:

Corollary 1. *The income gap for immigrants in occupation i at time t (IG_{it}), defined as the ratio of the geometric average of earnings of immigrants relative to the same average for natives, is given by:*

$$IG_{it} \equiv \frac{\widehat{y}_{iIt}}{\widehat{y}_{iNt}} = (1 - \tau_{iI}) \frac{\overline{h}_{iI} \widehat{x}_{iIt} \left(\widetilde{p}_{iIt}\right)^{\frac{1}{\theta}(\delta_{iIt}-1)}}{\overline{h}_{iN} \widehat{x}_{iNt} \left(\widetilde{p}_{iNt}\right)^{\frac{1}{\theta}(\delta_{iNt}-1)}} \tag{10}$$

Proof. Straightforward from $\exp^{\mathbb{E}(y_{igt})} = (1 - \tau_{ig}) \, w_{it} \overline{h}_{ig} \exp^{\mathbb{E}[x_{igt} \log(\epsilon_i)]}$ and equations (8) and (9). □

Corollary 1 shows that occupational income gaps for immigrants are a combination of the effects of occupation-specific wedges, which reflect discriminatory frictions; gaps in returns on observables and in the permanent components of talent, reflecting the effect of differences in the composition of human capital across groups; and on the relative desired occupational allocations. These allocations are the result of sorting across occupations but are in turn distorted by the extent of frictions that force involuntary choices. Corollary 1 also implies that with information of within-occupations income gaps, occupational allocations, gaps in the "returns" of observables and an assumption about the latent permanent component of talent across groups for each occupation, it is possible to pin down the magnitudes of wedges $(1 - \tau_{ig})$ and probabilities of forced choices α_{gt} from the system of equations that (10) implies. This is the basic idea of our inference procedure, which is described below.

4.2 Inference procedure and results

Our procedure to quantify the extent of occupational misallocation for immigrants relies on finding the magnitudes of the frictions for which the system of equations (10) fits best the data. With an assumption about the innate differences of talent across groups in each occupation, $\frac{\overline{h}_{iI}}{\overline{h}_{iN}}$, a value for the parameter θ, and the definition of $\widetilde{p}_{igt}$ in equation (8), it is possible to find M wedges $(1 - \tau_{iI})$ and $T + 1$ probability of forced choices α_{It} and α_N for which the system of $M \times T$ equations (10) fits best our set of information $\{IG_{it}, p_{igt}, x_{igt}\}$.[25] Similar to

[25]Formally, defining the set of $M + T + 1$ variables $x = \{(1 - \tau_{iI}), \alpha_{It}, \alpha_N\}$ and the function $g(x)$ from $\mathbb{R}^{M+T+1}$ to $\mathbb{R}^{M \times T}$ given by $g(x) = IG_{it} - z_{it}(x)$ where $z_{it}(x)$ is the RHS of equation (10), since $M \times T > M + T + 1$ is not possible to find a exact solution of the overdetermined system of equations $g(x) = 0$. Instead, we solve the minimization problem $\min_x \parallel g(x) \parallel$, where $\parallel \cdot \parallel$ is a vector norm on $\boldsymbol{R}^{M+T+1}$ (we employ the Euclidean norm, i.e. least squares).

Hsieh et al. (2019), for our baseline results we assume that the levels of the latent permanent components are the same across groups and normalize them to 1 ($\bar{h}_{ig} = 1 \; \forall \; i, g$), so immigrants have on average the same permanent components of talent than natives in each occupation,[26] but in our robustness checks we present alternatives to this assumption, modifying our model specification to infer values of $\frac{\bar{h}_{iI}}{\bar{h}_{iN}}$ for each i.

We use data of immigrant and native workers between the ages of 25 and 70 for the period 2015 to 2019, so our inference focuses on workers after they finish schooling but prior to their retirement. Panel A of Table 2 presents the respective stock sizes of immigrant and native workers in each year; while the first row of Panel C in the same Table reveals that the peak of the increase in the workforce due to immigration occurred in 2009 (4.4%). For labor income, we use our preferred measure converted to constant Colombian pesos of 2015. The occupational income gaps are defined as in equation (10), that is, in terms of geometric averages, but our results are similar if we instead use medians to avoid the influence of outliers. To measure the returns on observables x_{igt}, we need also values for the parameters γ and ϕ_i, the returns of experience and education respectively. For these values, we use the Mincerian returns in our pooled data from regressions of log income on years of schooling and age controlling for other observables.[27]

Table 2 – Estimated frictions for baseline specification and parameterization

	$\forall \; t$	2015	2016	2017	2018	2019
A. Number of workers*						
Natives (thousands)		15638	15684	15703	15633	16106
Share in total (%)		99.6	99.3	98.8	97.2	95.8
Immigrants (thousands)		60	114	189	449	709
Share in total (%)		0.4	0.7	1.2	2.8	4.2
B. Magnitudes of estimated frictions						
$Var\left[(1+\tau_{iI})\right]$	0.10	—				
$\alpha_{I,t}$		5.1%	3.8%	5.6%	8.1%	9.2%
α_N	4.7%	—				
C. Total increases due to immigration (%)						
Workforce		0.4	0.7	1.2	2.9	4.4
Output		0.3	0.6	1.0	2.1	3.1

Notes: *Workers in productive ages: between 25 and 70 years old.

[26] In other words, we are assuming that there are no innate talent differences between natives and immigrants. Remember that $\bar{h}_{ig}$ is isomorphic to the mean parameter of the Fréchet distribution, so in our baseline both groups draw from the same distribution of abilities. This, given the similar backgrounds of immigrants relative to natives, is a reasonable assumption.

[27] We use the same set of controls than in equation (1). For estimating γ, homogenous across occupations, we control also for occupation-group fixed effects. For estimating ϕ_i, heterogenous across occupations, we run the Mincerian regressions for each occupation.

For θ, the dispersion of abilities draws, similar to Hsieh et al. (2019) and Bryan and Morten (2019), we use the model's implication for the relative dispersion of wages within occupation-groups. Particularly, those wages should follow a Fréchet distribution with shape parameter equal to θ, and hence a ratio variance to mean equal to:

$$\frac{Variance\,(\theta)}{Mean\,(\theta)} = \frac{\Gamma\left(1-\frac{2}{\theta}\right) - \left[\Gamma\left(1-\frac{1}{\theta}\right)\right]^2}{\Gamma\left(1-\frac{1}{\theta}\right)} \tag{11}$$

for values of $\theta > 2$, where $\Gamma(\cdot)$ is the gamma function. So for each year, we compute the ratio variance to mean of the exponent of the residuals from cross-sectional regressions of log income on the 30×2 occupation-group dummies, and then solve equation (11) for the value of θ. The ratios variance to mean fluctuate between 2.43 and 2.55, so the resulting estimates of θ are on average 2.35. We use this value for θ in our baseline results, but we will explore robustness to setting it as low as 1.5 or as high as 3.5 in the next section. With this parametrization, we solve the system of equations resulting from inserting (8) in (10) using global solvers.[28]

Panel B of Table 2 shows our baseline results for both the variance of the estimated wedges and the probabilities of forced occupational choices, along with the values of the parameters used in the inference procedure. First, regarding wedges, it is worth to say that we focus on the variance of wedges because their dispersion is what really matters for workers misallocation;[29] their mean simply reflects the normalization used for the latent permanent component of talent across groups. We obtain a variance of 0.10, which implies a considerable dispersion of our estimated wedges: their values fluctuate between 0.3 times the median wedge (in the case of teaching professionals and scientists) and 1.7 times the median wedge (in the case of health professionals). Panel A of Figure E.4 in Appendix E displays the estimated wedges in each occupation. The substantial heterogeneity of those values across occupations suggests that the gains from removing discrimination could be sizable.

Second, regarding the fractions of immigrants in each period who are forced to make involuntary choices, α_{It}, we find these fractions tend to grow over time, coinciding with the increase of the immigration rate, rising from 5.1% in 2015 to 9.2% in 2019. Panel B of Figure E.4 displays these shares compared to the obtained value for native workers (α_N, equal to 4.7%). With the exception of 2016, in all years of the Venezuelan exodus the proportions of immigrants making involuntary occupation choices are larger than the obtained for natives. By 2019, the year of the largest migration inflows, this proportion is about twice as large as the one found for natives. Finally, Figure E.5 in Appendix E plots the observed occupational income gaps compared to the predicted by our two sets of estimated frictions using the RHS

[28] Particularly, we use the genetic algorithm with a population size of 2000 individuals, 10 times larger than the suggested by default in Matlab, for example. We verified that independently the initial population chosen, we get always the same solution.

[29] If wedges were distributed log-normal along productivities, there would be a perfect correlation between the magnitude of the variance of wedges and the gains of reallocate workers across occupations; see Hsieh et al. (2019) for a proof, or in the context of firm-level misallocation, Chen and Irarrazabal (2015).

of equation (10), as a graphic representation of the fit of the model. For the relatively small number of parameters inferred, there is a strong positive association between the observed and predicted income gaps, with a relatively high correlation coefficient (0.73).

The magnitudes found for both types of frictions point in the direction that reallocating workers across occupations according to their frictionless choices could imply non-negligible gains on aggregate productivity. To quantify these gains, we first need to solve for the values of the remaining unobservable variables compatible with general equilibrium, and use those values to obtain the responses of the endogenous outcomes (allocations, prices and output) in the counterfactual equilibria. The next section develops a procedure in this direction and presents a series of robustness checks of the results.

5 Counterfactuals and robustness

In this section we show how aggregate productivity and occupational allocations would change when implementing two types of reforms: i) Removing entirely each type of frictions for immigrants; ii) Equalizing immigrants' frictions to those found for natives. We first describe the procedure to obtain those counterfactual equilibria and report its results. Next, we examine how sensitive are our results to the values of the calibrated parameters during the procedure, particularly to θ and σ. Finally, we explore robustness to alternative specifications of our model, that include to allow for differences in innate talent across occupations, to consider time-variant discriminatory wedges and to control for local working experience in the vector of observables

5.1 Counterfactual exercises

To quantify the aggregate implications of our estimated frictions, we need first to solve for the remaining exogenous variables of the model: group-specific preferences for a given occupation, z_{igt}, and the productivities for the representative firm of each occupation i, A_{it}. These variables are kept constant when our counterfactual exercises are performed. Appendix B.3 depicts the procedure to solve for these values in the observed economy jointly with the equilibrium values of the endogenous efficiency wages (w_{it}) and the total output of the economy (Y_t), following several implications of the model. The procedure needs a value of σ, the elasticity of substitution among occupations, a parameter that we make equal to 3 in our baseline results (a common value in the literature).[30] Nevertheless, we explore robustness to setting it as low as 2 or as high as 5 in the next subsection. Figure E.8 in Appendix E displays the resulting average values over years of group-specific preferences and productivities for each occupation.

[30]We also require information on the absolute numbers of both immigrant and native workers, which are presented in Panel A of Table 2. It is worth noting that to compute the labor market frictions in the previous section, we did not rely on the number of workers. Rather, we only needed data on their occupational distribution and the magnitudes of the income premiums

Preferences among occupations do not vary greatly between natives and immigrants, and, as expected, occupations with higher requirements of educational attainment are inferred as more productive for the representative firm.

Once this set of values are found for the observed equilibrium, two counterfactual equilibria can be computed. The first is to remove entirely each set of frictions for immigrants. In terms of our model, it involves setting τ_{iI} and α_{It} equal to zero, and to derive the endogenous response of allocations, efficiency wages and total output in the new equilibrium. This is a drastic reform in which all immigrants choose occupations according to the efficient allocation. Even though this counterfactual is extreme by nature, and hence perhaps unrealistic as policy reform, by removing jointly and separately each set of frictions we get a clear understanding of the relative importance of each one. The second counterfactual is equalizing immigrants' frictions to those found for natives. In spite of our inference procedure is being able to tell us the extent of type II frictions for natives (α_N), one of its identification assumptions is that natives do not face "taxes" in their income. Hence, from our baseline results we cannot directly assess to which counterfactual value of wedges' variance we should reduce our estimated variance. So for this case, we propose a measure of the counterfactual variance of wedges based on estimating a restricted version of the model for different sub-groups of natives.

The procedure to obtain the endogenous variables in the counterfactuals once we have the intended values of frictions for each reform is described in detail in Appendix B.4. Basically, we use a fixed-point algorithm to find total output, occupational allocations and efficiency wages that clear each occupational labor market in each year given the values of the exogenous variables and the intended frictions in the proposed counterfactuals. The annual aggregate gains from the reforms are computed as the percentage change in total output for each year of the counterfactual economy relative to total output in the actual economy. In what follows, we describe our findings for each proposed reform.

Reform I: Removing frictions for immigrants

We first evaluate the counterfactual of removing entirely both types of frictions for immigrants ($\tau_{iI} = \alpha_{I,t} = 0 \ \forall \ i, t$). First rows of Panel A in Table 3 display the results for our baseline parameterization in each year of the studied period. By considering the results for the most recent year (2019), when the participation of immigrants in the Colombian workforce reaches its peak, removing all frictions for immigrants would permanently increase total output by as much as 0.9%. Since in both the counterfactual and the actual economy the amount of workers is the same, the rise in output is the result of the increase in aggregate labor productivity (where labor is measured in efficiency units) due to the improvement in the allocative efficiency of labor among occupations. An inspection of the counterfactual occupational allocation reveals that around 30% of immigrants in each year would reallocate as a consequence of the reform. This magnitude is a lower bound, since we are not able to quantify transitions that do not alter occupational shares.[31] The average counterfactual occupational allocation of immigrants compared to the observed one is displayed in Panel A of Figure 3, where it is evident that immigrants gain participation in occupations with higher skill requirements. Since the reallocation of immigrant workers has general equilibrium implications for efficiency wages, there is also a small response in terms of reallocation of natives: up to 0.4% of their workforce in 2019.[32]

Table 3 – Productivity gains and shares of workers reallocated by reforms removing frictions

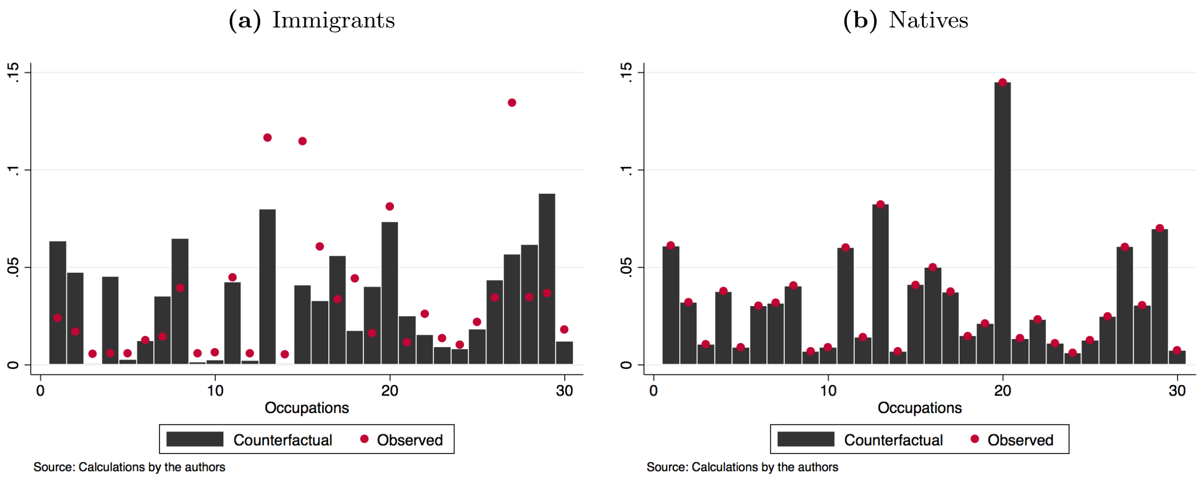
	2015	2016	2017	2018	2019
A. Reform I					
- Productivity gains (%):					
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \; \forall \; i, t$	0.06	0.12	0.21	0.58	0.90
Only type I: $\tau_{iI} = 0 \; \forall \; i$	0.05	0.09	0.16	0.40	0.61
Only type II: $\alpha_{I,t} = 0 \; \forall \; t$	0.01	0.02	0.04	0.14	0.23
- Share of immigrants reallocated (%):					
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \; \forall \; i, t$	31.08	31.18	32.24	30.95	29.41
Only type I: $\tau_{iI} = 0 \; \forall \; i$	30.00	30.40	30.93	29.31	27.92
Only type II: $\alpha_{I,t} = 0 \; \forall \; t$	2.36	1.60	2.13	3.37	3.86
- Share of natives reallocated (%):					
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \; \forall \; i, t$	0.04	0.07	0.11	0.26	0.38
Only type I: $\tau_{iI} = 0 \; \forall \; i$	0.03	0.07	0.11	0.25	0.36
Only type II: $\alpha_{I,t} = 0 \; \forall \; t$	0.00	0.00	0.00	0.02	0.04
B. Reform II					
- Productivity gains (%):	0.02	0.04	0.08	0.24	0.38
- Share of workers reallocated (%):					
Immigrants:	10.21	10.31	10.55	9.69	9.11
Natives:	0.01	0.02	0.04	0.08	0.12

Notes: Reform I refers to the counterfactual of removing type I (discriminatory wedges, τ_{ig}) and type II (involuntary choices, α_{gt}) frictions for immigrants. Reform II refers to the counterfactual of equating both types of frictions for immigrants to the values of natives. Reform II assumes $\alpha_{It} = \alpha_N \forall \; t$ and a counterfactual variance of wedges equal to 0.047, a value derived from estimating the model only for natives with the subpopulations rural-men, urban-men, rural-women and urban-women, assuming only urban-men are not discriminated against. All results are computed using $\theta = 2.35$ and $\sigma = 3$.

[31] The estimated reallocation and the following ones, are computed adding up the total amount of workers in occupations with positive variations in their participation in the total. So we are omitting possible reallocations that do not alter occupational shares (for example, a worker transitioning from sector s and s', whereas another worker makes the opposite transition) which we cannot identify.

[32] Panel B of Figure 3 shows for natives their counterfactual occupational allocation compared to the observed one, but the changes are very small.

Figure 3 – Counterfactual occupational distributions (Reform I)



Notes: Figure shows the occupational distribution of Immigrants (Panel A) and Natives (Panel B) in the pooled data after Reform I is implemented, compared to the actual distributions. For the definition of Reform I, see note in Table 3.

The latter results lead to two questions. First, how significant are the productivity gains resulting from our reform when compared to the overall gains attributable to immigrant inflows? In our model, immigrants contribute to the host country's aggregate output by augmenting the supply of human capital in each occupation, which in turn affects both the values of the exogenous unobservable variables compatible with general equilibrium and the endogenous variables (such as wages per efficiency unit and output). Using our calibrated value of σ, we compute the output in the absence of immigration letting to adjust all other remaining variables. The results are shown in the final row of Table 2. Immigrant workers from Venezuela, who faced the inferred frictions in the Colombian labor market, generated increases in aggregate output ranging from 0.3% in 2015 to 3.1% in 2019. This means that our reform in 2019 alone, for example, which removes all frictions for immigrants, would have led to an additional 0.9% increase in aggregate output due to the higher labor productivity stemming from the improved allocation of workers across different occupations. Stated differently, the reform would have boosted the increase in aggregate output due to immigration by 29% in 2019 (i.e., 0.9% out of 3.1%).

Second, which type of friction is more important for the results of this reform? For this, we next evaluate the counterfactuals of removing each type of frictions separately. By 2019, removing only discriminatory wedges ($\tau_{iI} = 0 \; \forall \; i$) leads 28% of immigrants and 0.4% of natives to reallocate, increasing aggregate productivity by 0.6%. So discriminatory wedges account for around two-thirds of the total gains from the reform of eliminating all frictions. Instead, removing only involuntary occupational choices ($\alpha_{It} = 0 \; \forall \; t$) leads only to 4% of immigrants to reallocate and increases aggregate productivity around 0.2%. As opposed to discriminatory wedges, immigrants who are misallocated by forced choices are randomly assigned among all occupations. Thus, their reallocation does not importantly affect efficiency wages, generating almost no changes in the allocation of natives. To sum up, discriminatory wedges have larger implications in terms of allocative efficiency and involve additional general equilibrium effects, a channel that has a second-round effect in the occupational allocation of the native workforce.

Reform II: Equalizing immigrants' frictions to those found for natives

The aim of our second reform is to reduce frictions for immigrants to a similar level to that for natives. From the point of view of policy analysis, this counterfactual is perhaps more interesting because it provides a calculation of the aggregate productivity gains when immigrants are completely assimilated by the labor market of the host country and face the same frictions as natives. It also provides a crude quantification of the macroeconomic benefits of different programs that help to regularize and to reduce barriers for immigrants, allowing them to compete in the labor market under the same conditions as native workers.

For this counterfactual, we first equalize the values of the fractions of immigrants that are forced to make involuntary choices to the value estimated for natives, i.e. $\alpha_{It} = \alpha_N \; \forall \, t$. Second, we reduce the variance of our discriminatory wedges to a level that reflects the prevalent discrimination in the labor market of natives. Here we face the difficulty that one of our identification assumptions was that natives does not face "taxes" in their income, so their wages reflect their marginal productivities, and thus, from our results we do not have a direct measure of the prevalent discrimination in the labor market for natives.[33] In order to gauge the extent of this discrimination, we re-estimate our model for sub-groups of natives for which one could presumably argue there would be discrimination against them, constraining α_{gt} to our intended value α_N. This out-of-the model inference, although not perfect, will offer us an approximate value of the variance of the wedges for our counterfactual.

For this exercise, we choose as the possible groups facing discrimination women and rural workers. So we re-estimate the constrained model only for native workers using the following four subpopulations: urban-men (UM), rural-men (RM), urban-women (UW), and rural-women (RW). For identification, we assume UM do not face discrimination. With our baseline parameterization, we obtain a variance of wedges equal to 0.03 for RM, 0.08 for UW and 0.10 for WR; a ranking that seems reasonable. Using the average shares of the four groups in the total native population, our computed variances imply a pooled variance of 0.047. This value is close to the obtained in the case of estimating the restricted model only for men and women with women facing discrimination (0.044), or only for urban and rural workers with rural workers facing discrimination (0.040).[34] Therefore, in our counterfactual we shrink im-

[33]This assumption was needed because from the system of equations (10) , $(1 - \tau_{iI})$ would be indistinguishable from $\frac{(1-\tau_{iI})}{(1-\tau_{iN})}$ if we would have assumed wedges for natives too.

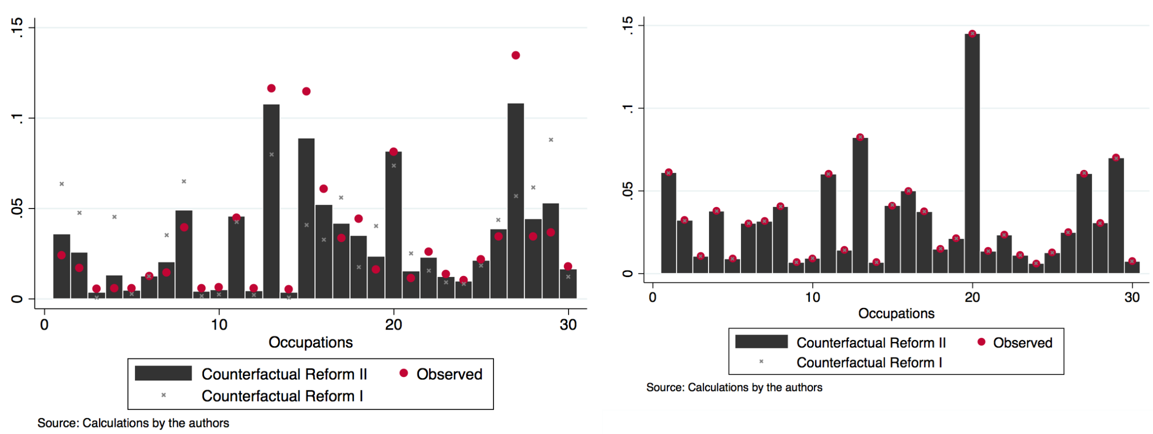
[34]We also estimate a placebo test in which we divide the natives' population into two random groups, to verify that our inference was effectively capturing some type of discrimination instead of measurement error, for instance. In this case, we estimate a variance equal to 0.001.

migrants' wedges until they exhibit a variance equal to 0.047, which corresponds to 47% of our estimated value of 0.100.

Panel B in Table 3 displays the results for our baseline parameterization. By 2019, reducing immigrants' frictions to a similar level to the inferred for natives would lead 9.1% of immigrants and 0.1% of natives to reallocate. Figure 4 shows for each group the average counterfactual occupational allocations, distributions that are half-way between the observed and the counterfactuals with the first reform. As a result, Colombian aggregate labor productivity would permanently increase up to 0.4% due to the assimilation of the new workforce. How large or relevant are these gains in aggregate productivity when compared to the overall gains attributable to immigrant inflows? By comparing the obtained increase in aggregate labor productivity with the growth in aggregate output resulting from immigration displayed in the final row of Table 2, we conclude that this type of reform would have led to a 13% upsurge in the growth of aggregate output attributable to immigration (i.e., 0.4% out of 3.1%). Given that this reform appears to be both realistic and implementable, we consider our findings to be relevant for policymakers who aim to assess the impact of policies intended to expedite the assimilation process of immigrants into the labor market.[35]

Figure 4 – Counterfactual Occupational Distributions (Reform II)

(a) Immigrants **(b)** Natives



Notes: Figure shows the occupational distribution of Immigrants (Panel A) and Natives (Panel B) in the pooled data after Reform II is implemented, compared to the actual distributions and the derived ones from Reform I. For the definitions of Reforms I and II, see note in Table 3.

[35]It is worth noting that without policies aimed at accelerating the assimilation process of immigrants, it could take several years for this process to occur. For instance, Kerr and Kerr (2011) document that empirical studies find that in the US, immigrants converge to native levels in terms of occupation rates within approximately 10 years and in terms of wages within approximately 15 years.

Reforms I and II in the salaried and non-salaried segments of the labor market

Considering the findings presented in Section 2 that suggest a larger income gap for salaried workers, in the Appendix C we re-estimate our model and perform our counterfactual exercises for both the salaried and non-salaried segments of the labor market. The objective of the exercise is to identify differences in the extent of the frictions experienced by immigrants in these two distinct labor market segments, and subsequently evaluate the potential heterogeneous consequences resulting from these differences. In a nutshell, consistent with the wider income gaps that immigrants face in the salaried segment, we conclude that frictions are more pronounced and have greater aggregate implications in the formal sector, than in the informal one. This is coherent with the greater flexibility of the informal sector, where immigrants generally do not face barriers such as the recognition of educational degrees or the need to obtain permits to work legally, among others.

5.2 Robustness to parameterization

We now explore the robustness of our baseline results to alternate values of our calibrated parameters θ and σ. Consider first robustness to changes in the Fréchet shape parameter θ. This parameter is inversely related to the dispersion of abilities draws, and, contrary to σ, affects the values of the estimated frictions. Thus, θ has both a direct and an indirect effect on the aggregate gains of our reforms. The former effect refers to the impact of θ for a given value of our estimated frictions. Since the elasticity of total output to the efficiency loss caused by the variance of wedges is a direct function of θ,[36] the direct effect implies that the loss in aggregate productivity conditional to the extent of the frictions is increasing in θ. Instead, the indirect effect refers to the impact of θ on the aggregate gains via the estimated frictions, particularly the variance of wedges. For a given proportion of involuntary choices, a larger dispersion of abilities draws (smaller values of θ) implies individuals have stronger patterns of comparative advantage across sectors, so in order to rationalize the observed occupational income gaps and allocations more discrimination is needed. By this channel, the variance of wedges, and in turn the aggregate gains of reforms, are thus decreasing in θ. This indirect effect can be attenuated if θ also affects the values obtained for the fractions of involuntary choices.

Columns (2) and (3) of Table 4 show the values for the new estimated frictions and how our counterfactual results for 2019 change when we consider $\theta = 1.5$ and $\theta = 3.5$ respectively; while column (1) redisplays our baseline results from Table 3 for comparison.[37] By implementing

[36] Analytically, this can be shown by assuming a functional form for the joint distribution of productivities and wedges in order to obtain a closed-form expression for aggregate output in terms of the variance of wedges. For example, Hsieh et al. (2019) show that, abstracting from involuntary choices and assuming a joint log-normal distribution and $\sigma \to \infty$, such elasticity is equal to $\frac{1}{2}(\theta - 1)$. Otherwise, only by numerical simulations this could be exemplified.

[37] For the remaining years, a comparison of the time series of the aggregate gains of each reform is displayed

Reform I in 2019, aggregate labor productivity gains rise from 0.73% with $\theta = 1.5$, to 0.90% with our baseline $\theta = 2.35$, and to 0.94% with $\theta = 3.5$. Similarly, the gains from Reform II increase from 0.30% with $\theta = 1.5$, to 0.38% with our baseline $\theta = 2.35$, and to 0.46% with $\theta = 3.5$. Hence, aggregate productivity gains from both reforms are increasing in θ, suggesting that the direct effect of θ is stronger than the indirect effect. However, it is worth to highlight that this latter effect is present, anyway: the procedure infers a larger variance of wedges when θ decreases to 1.5.[38] Overall, our estimated gains from removing immigrants' frictions exhibit only a moderate sensitivity to changes in θ, so our conclusions are not very affected by the calibration of θ.

in Figure E.6 in Appendix E

[38] When θ increases to 3.5, there is almost no change in the variance of wedges. This is because the procedure also infers a simultaneous increase in the extent of involuntary frictions. If we would have restricted the procedure to the same values of α_{gt} than in the baseline, we would have obtained a variance equal to 0.073.

Table 4 – Results for alternative parameterizations

	(1) Baseline	(2) Low θ	(3) High θ	(4) Low σ	(5) High σ
A. Calibrated parameters					
θ	2.35	1.50	3.50	2.35	2.35
σ	3.00	3.00	3.00	2.00	5.00
B. Magnitudes of estimated frictions					
$Var\left[(1+\tau_{iI})\right]$	0.100	0.118	0.105	0.100	0.100
$\alpha_{I,2015}$	5.1%	4.7%	5.1%	5.1%	5.1%
$\alpha_{I,2016}$	3.8%	3.6%	3.8%	3.8%	3.8%
$\alpha_{I,2017}$	5.6%	5.0%	5.8%	5.6%	5.6%
$\alpha_{I,2018}$	8.1%	6.7%	9.6%	8.1%	8.1%
$\alpha_{I,2019}$	9.2%	7.5%	11.0%	9.2%	9.2%
α_N	4.7%	5.5%	2.1%	4.7%	4.7%
C. Results of counterfactual exercises for 2019					
C.1. Reform I					
- Productivity gains (%):					
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \;\forall\; i, t$	0.90	0.73	0.94	0.84	0.96
Only type I: $\tau_{iI} = 0 \;\forall\; i$	0.61	0.45	0.65	0.54	0.67
Only type II: $\alpha_{I,t} = 0 \;\forall\; t$	0.23	0.24	0.21	0.24	0.22
- Share of reallocated workers (%) $[immigrants, natives]$:					
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \;\forall\; i, t$	$[29.4, 0.4]$	$[23.9, 0.1]$	$[39.4, 0.7]$	$[29.3, 0.5]$	$[29.6, 0.3]$
Only type I: $\tau_{iI} = 0 \;\forall\; i$	$[27.9, 0.4]$	$[22.9, 0.1]$	$[36.3, 0.7]$	$[27.8, 0.5]$	$[28.0, 0.2]$
Only type II: $\alpha_{I,t} = 0 \;\forall\; t$	$[3.9, 0.0]$	$[3.1, 0.0]$	$[4.7, 0.1]$	$[3.9, 0.0]$	$[3.9, 0.0]$
C.2. Reform II					
- Productivity gains (%):	0.38	0.30	0.46	0.36	0.39
- Share of reallocated workers (%) $[immigrants, natives]$:	$[9.1, 0.1]$	$[9.1, 0.1]$	$[12.3, 0.2]$	$[9.1, 0.2]$	$[9.2, 0.1]$

Notes: The values of the parameters (θ, σ) for each alternative parameterization are: baseline: $(2.35, 3)$; low θ: $(1.5, 3)$; high θ: $(3.5, 3)$; low σ: $(2.35, 2)$; high σ: $(2.35, 5)$. For the definitions of Reforms I and II, see note in Table 3.

Now consider robustness to changes in the elasticity of substitution σ between occupations. Columns (4) and (5) of Table 4 return to our baseline $\theta = 2.35$ and vary σ, from $\sigma = 2$ in Column (4) to $\sigma = 5$ in Column (5). Since σ does not have any role in the estimation of frictions, the magnitudes of the inferred frictions in both cases are equal to those in our baseline parameterization. However, the same values for frictions have slightly different implications for aggregate labor productivity. Gains from Reform I increase from 0.84% with $\sigma = 2$ to 0.96% with $\sigma = 5$, whereas gains from Reform II rise from 0.36% with $\sigma = 2$ to 0.39% with $\sigma = 5$, so the changes in the gains are marginal. Intuitively, the result that aggregate productivity gains are increasing in σ reflects the fact that with more substitutability across occupations the labor demands for an occupation-group are more sensitive to frictions, and the firm is thus more prone to use misallocated labor, increasing the efficiency loss. To sum up, the gains from our reforms are not very affected by changes in σ, suggesting that our results are also not very affected by the calibration of σ.

5.3 Robustness to specification

Finally we explore robustness to three different model specifications. The first aims to infer simultaneously values for $\frac{\bar{h}_{iI}}{\bar{h}_{iN}}$. One of our identifying assumptions is that there are no innate talent differences between natives and immigrants, i.e. $\bar{h}_{ig} = 1$. This assumption was needed because from the system of equations (10), $(1 - \tau_{iI})$ would be indistinguishable from $\frac{\bar{h}_{iI}}{\bar{h}_{iN}}$ if we would have assumed differences in the permanent component of talent between groups. An alternative specification of the model that allows us to infer values of $\frac{\bar{h}_{iI}}{\bar{h}_{iN}}$ is to assume that discrimination has not always been present for immigrants, but only when their presence was very noticeable to the native public. So we choose an arbitrary threshold for the immigration rate (1%) from which the wedges begin to appear.[39] This means that only from 2017 onwards we assume immigrants face discriminatory wedges. Hence, before 2017, their observed income gaps corrected for selection effects and returns on observables are only consequence of differences in innate talent, whereas from 2017 onwards are consequence both from differences in innate talent and discrimination.

Column (2) in Table 5 shows the estimated frictions and the results for reforms in 2019 for this specification, while Column (1) of the table redisplays our baseline estimates from Table 3 for comparison.[40] With respect to our baseline, the inferred variance of the wedges increases from 0.100 to 0.162, while the extent of type II frictions for immigrants marginally decreases in all years. For natives, instead, the proportion of involuntary choices strongly decreases from 4.7% in the baseline to 0.7%. Figure E.9 in Appendix E shows the inferred values for the relative innate talent differences, which, in spite of exhibiting some heterogeneity across occupations, are on average very close to one (0.97), providing some support for the assumption in our baseline. The presence of differences in innate talent across occupations not only helps the model to have a better fit to the data (see Panel A of Figure E.10 in Appendix E, in which the correlation coefficient between the observed and predicted income gaps is 0.87), a non-surprising result given the lower parsimony of this specification; but also decreases the power of discriminatory wedges to explain the observed income gaps. Hence, even though discriminatory wedges have a larger variance than in our baseline, removing entirely those wedges have a smaller impact on aggregate productivity. Total gains from Reform I decrease from 0.90% in our baseline to 0.52%, a result that is due to the smaller contribution of removing only wedges: while removing only involuntary choices leads to similar gains (0.21%) than in our baseline (0.23%), removing only wedges imply gains of only 0.29%, compared to 0.63% in our baseline. But for Reform II, the consequences for aggregate productivity are pretty

[39]Results are similar choosing one year after or before the selected year.

[40]For the remaining years, a comparison of the time series of the aggregate gains of each reform is displayed in Figure E.7 in Appendix E

similar.[41] This is because the smaller gains from shrinking wedges are compensated with larger gains from equating α_{It} to α_N, given that α_N is now inferred lower. Aggregate gains from Reform II are now 0.31%, close to our baseline (0.38%). To sum up, this specification has non-negligible effects for the gains from Reform I, but delivers similar results for the gains from Reform II.

Table 5 – Results for alternative specifications

	(1) Baseline	(2) Inferring $\frac{\overline{h}_{iI}}{\overline{h}_{iN}}$	(3) Time-variant τ_{itI}	(4) Local working exp.
A. Magnitudes of estimated frictions				
$Var\left[(1+\tau_{iI})\right]$	0.100	0.162	0.142	0.098
$Var\left[\frac{\overline{h}_{iI}}{\overline{h}_{iN}}\right]$	–	0.164	–	–
$Var\left[(1+\tau_{iI,2015})\right]$	–	–	0.106	–
$Var\left[(1+\tau_{iI,2016})\right]$	–	–	0.170	–
$Var\left[(1+\tau_{iI,2017})\right]$	–	–	0.149	–
$Var\left[(1+\tau_{iI,2018})\right]$	–	–	0.142	–
$Var\left[(1+\tau_{iI,2019})\right]$	–	–	0.145	–
$\alpha_{I,2015}$	5.1%	4.5%	–	5.1%
$\alpha_{I,2016}$	3.8%	2.9%	–	3.8%
$\alpha_{I,2017}$	5.6%	5.2%	–	5.5%
$\alpha_{I,2018}$	8.1%	7.7%	–	8.1%
$\alpha_{I,2019}$	9.2%	9.0%	–	9.2%
α_N	4.7%	0.7%	–	4.9%
B. Results of counterfactual exercises for 2019				
B.1. Reform I				
- Productivity gains (%):				
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \ \forall \ i, t$	0.90	0.52	1.28	0.89
Only type I: $\tau_{iI} = 0 \ \forall \ i$	0.61	0.29	1.28	0.60
Only type II: $\alpha_{I,t} = 0 \ \forall \ t$	0.23	0.21	–	0.23
- Share of reallocated workers (%) $[immigrants, natives]$:				
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \ \forall \ i, t$	$[29.4, 0.4]$	$[21.4, 0.3]$	$[51.9, 0.6]$	$[29.4, 0.4]$
Only type I: $\tau_{iI} = 0 \ \forall \ i$	$[27.9, 0.4]$	$[19.0, 0.2]$	$[51.9, 0.6]$	$[27.9, 0.4]$
Only type II: $\alpha_{I,t} = 0 \ \forall \ t$	$[3.9, 0.0]$	$[3.8, 0.0]$	–	$[3.8, 0.0]$
B.2. Reform II				
- Productivity gains:	0.38	0.31	0.53	0.36
- Share of reallocated workers (%) $[immigrants, natives]$:	$[9.1, 0.1]$	$[7.6, 0.1]$	$[12.9, 0.2]$	$[9.1, 0.1]$

Notes: For the definitions of Reforms I and II, see note in Table 3.

Our second alternative specification allows us to consider time-variant discriminatory wedges. So we return to our assumption $\overline{h}_{ig} = 1$ and infer $(1 - \tau_{iIt}) \ \forall \ i, t$. Here we face the difficulty that the system in (10) has $M \times T$ equations, and, with $M \times T$ wedges to in-

[41] To make comparable the reduction in the variance of wedges between this specification and our baseline, we shrink the obtained variance in the same proportion than in our baseline Reform II, i.e. 47%.

fer, we must remove type II frictions to prevent the system from being underdetermined. By doing so, the system becomes exactly determined and wedges are simply obtained by solving equation (10) for each pair occupation-year. Hence, the model is able to perfectly fit the data (see Panel B of Figure E.10 in Appendix E), but its specification is in turn much less parsimonious than in our baseline. Column (3) in Table 5 shows the results for this specification. The inferred variance of discriminatory wedges, although it differs across years, is relatively stable from 2017 onwards, when immigration rates are higher and then occupational income gaps are more precisely estimated; and overall have on average a higher level (0.142) than in our baseline (0.100). Since time-variant wedges now have a larger explanatory power of the observed income gaps, the aggregate gains from both reforms are somewhat larger relative to our baseline results. Removing all frictions increases 1.28% in aggregate labor productivity (0.90% in our baseline), and leads to approximately half of the immigrants to reallocate in 2019 (30% in our baseline). Regarding Reform II, reducing the variance of wedges in the same proportion than in our baseline leads to a 0.53% aggregate gain and 13% of immigrants to reallocate, relative to 0.38% and 9% in our baseline, respectively. Therefore, the consequences for aggregate productivity are, although slightly larger, not very distant from our baseline results. This is due to the fact that the larger variances of the wedges are compensating the absence of type II frictions, while the cross-sectional variation in the inferred variances is playing an analogue role to the variation in time in α_{It}. In conclusion, we find this formulation of frictions delivers results that are not very different than our more parsimonious baseline specification.

The third specification includes local working experience in the set of our observable controls, given the empirical evidence in Section 2 suggesting a role of this variable in explaining immigrants' residual income gaps. This is, we return to our baseline specification with time-invariant wedges and the assumption $\overline{h}_{ig} = 1$, and include the proxy of local working experience constructed in Section 2 in the vector of returns from observables x_{igt}.[42] Column (3) in Table 5 shows the results for this specification. The variance of the discriminatory wedges shows a slight decrease, from 0.100 to 0.098, while the probabilities of involuntary choices remain nearly unchanged. These modest differences in the inferred frictions result in slight changes to the gains in aggregate labor productivity due to our reforms. Specifically, removing all frictions for immigrants would now increase aggregate labor productivity by 0.89%, as compared to 0.90% in our baseline, and reducing the frictions faced by immigrants to the levels experienced by natives would now increase aggregate productivity by 0.36%, rather than 0.38%. The low sensitivity of our results, despite observing non-negligible reductions in the computed income gaps when controlling for local working experience (columns 8-9 in Table 1), underscores the importance of incorporating workers' allocations across sectors in our model. Specifically, in this specification, we maintain observed workers' allocations while only reducing their residual

[42] In this way, denoting by $\widetilde{a}$ the years of local working experience, we have now $x_{igt} \equiv a_{igt}^{\gamma} \widetilde{a}_{igt}^{\widetilde{\gamma}} s_{ig}^{\phi_i}$, and the parameters γ, $\widetilde{\gamma}$ and ϕ_i are re-estimated in a similar way as in the baseline.

income gaps. Consequently, the inferred frictions and their aggregate consequences closely resemble those in the baseline scenario.

6 Conclusions

Due to the additional obstacles that immigrants often face in the labor market of the host country before assimilation, their labor can experience higher levels of misallocation compared to native workers. In situations where there is a sudden and substantial influx of immigrants, this misallocation could have significant macroeconomic implications for the host country's economy, particularly on its aggregate labor productivity. We document the considerable impact of this phenomenon in the case of the "Venezuelan exodus" to the Colombian economy during the period of 2015-2019.

Our research first presents reduced-form evidence which signals a higher degree of occupational misallocation among Venezuelan migrants compared to non-migrants. Specifically, the observed occupational allocation of immigrants, given their educational attainment, and their significant residual income gaps within-occupations, point towards a greater level of labor misallocation for immigrants. When we analyze these findings using a structural model of occupational choice, which takes into account workers who self-select across occupations due to unobservable skills, we can decompose how much of the observed residual income gaps are explained by sorting, and how much of the unexplained part is due to prevalent frictions experienced by immigrants. By using the model, we are able also to infer the cost of these frictions to Colombian aggregate productivity. Our results indicate that those frictions not only have a direct impact on the occupational allocation of immigrants, but also trigger general equilibrium effects that have consequences on the allocation of natives. By eliminating all frictions for immigrants, Colombian aggregate labor productivity could permanently increase by approximately 0.9%.

There are several avenues for future research. For tractability, our model abstracts from capital or the use of other inputs, so our implications for aggregate productivity are limited to the effect of labor misallocation only. But it is possible that this misallocation also generates inefficiencies in the use of other factors across occupations, magnifying the effects on the aggregate TFP. Further, in a model with capital, dynamic considerations could also start to matter. The study of dynamic inefficiencies induced by static factor misallocations, in a world where the losses in aggregate productivity are faced by generations that would not be necessarily the same as those who would benefit of immigrants assimilation (if the process spreads over a long period of time), is a fruitful road for future research. Finally, we made particular choices about the functional form of the talent distribution (Fréchet) and the specification of frictions, collecting previous ways in the literature to generate occupational misallocation of self-selecting workers while keeping the problem analytically tractable. There is room for further exploration of the consequences of moving towards more general specifications.

References

Adamopoulos, T., Brandt, L., Leight, J., and Restuccia, D. (2022). Misallocation, selection and productivity: A quantitative analysis with panel data from China. *Econometrica*, 90(3):1261–1282.

Aleksynska, M. and Tritah, A. (2015). The Heterogeneity Of Immigrants, Host Countries' Income And Productivity: A Channel Accounting Approach. *Economic Inquiry*, 53(1):150–172.

Altonji, J. G. and Blank, R. M. (1999). Race and gender in the labor market. volume 3 of *Handbook of Labor Economics*, chapter 48, pages 3143 – 3259. Elsevier.

Azoulay, P., Jones, B. F., Kim, J. D., and Miranda, J. (2022). Immigration and Entrepreneurship in the United States. *American Economic Review: Insights*, 4(1):71–88.

Bahar, D., Ibáñez, A. M., and Rozo, S. V. (2021). Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees. *Journal of Development Economics*, 151(C).

Becker, G. (1971). *The economics of discrimination*. The University of Chicago Press, 2nd edition.

Birinci, S., Leibovici, F., and See, K. (2021). Immigrant misallocation. Working Papers 2021-004, Federal Reserve Bank of St. Louis.

Bonilla-Mejía, L., Morales, L. F., Hermida-Giraldo, D., and Flórez, L. A. (2020). The labor market of immigrants and non-immigrants: Evidence from the Venezuelan refugee crisis. Borradores de Economia 1119, Banco de la Republica.

Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature*, 32(4):1667–1717.

Borjas, G. J., Chiswick, B. R., and Elsner, B. (2019). *Foundations of migration economics*. IZA Prize in Labor Economics. Oxford University Press.

Borjas, G. J., Grogger, J., and Hanson, G. H. (2011). Substitution between immigrants, natives, and skill groups. NBER Working Papers 17461, National Bureau of Economic Research, Inc.

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

Bryan, G. and Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from Indonesia. *Journal of Political Economy*, 127(5):2229–2268.

Burzynski, M. and Gola, P. (2019). Mexican migration to the United States: Selection, assignment, and welfare. Working Paper Series 2019-10, LISER.

Caruso, G., Gomez Canon, C., and Mueller, V. (2019). Spillover effects of the Venezuelan crisis: Migration impacts in Colombia. *Oxford Economic Papers*. gpz072.

Chassamboulli, A. and Peri, G. (2015). The labor market effects of reducing the number of illegal immigrants. *Review of Economic Dynamics*, 18(4):792 – 821.

Chen, K. and Irarrazabal, A. (2015). The role of allocative efficiency in a decade of recovery. *Review of Economic Dynamics*, 18(3):523 – 550.

Chiswick, B. R. and Miller, P. W. (2011). Educational mismatch: Are high-skilled immigrants really working at high-skilled jobs and the price they pay if they aren't? In Chiswick, B. R., editor, *High-skilled immigration in a global labor market*, pages 109–154. AEI Press, Washington, DC.

DANE (2021). Encuesta pulso de la migración (survey of migrants' perceptions). recovered from https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/encuesta-pulso-de-la-migracion-epm.

Dustmann, C., Schönberg, U., and Stuhler, J. (2016). The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives*, 30(4):31–56.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Gutiérrez S., A. (2017). Venezuela's economic and social development in the era of Chavism. *Latin American Policy*, 8(2):160–188.

Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics*, 6(1):735–770.

Hornung, E. (2014). Immigration and the diffusion of technology: The huguenot diaspora in prussia. *American Economic Review*, 104(1):84–122.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and U.S. economic growth. *Econometrica*, 87(5):1439–1474.

Ingwersen, K. and Thomsen, S. L. (2021). The immigrant-native wage gap in Germany revisited. *The Journal of Economic Inequality*, 19(4):825–854.

Isphording, I. E. and Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *Journal of Economic Behavior and Organization*, 105:30–50.

Joona, P. A., Gupta, N. D., and Wadensjö, E. (2014). Overeducation among immigrants in Sweden: Incidence, wage effects and state dependence. *IZA Journal of Migration*, 3(1):9.

Kerr, S. P. and Kerr, W. R. (2011). Economic impacts of immigration: A survey. *Finnish Economic Papers*, 24(1):1–32.

Knight, B. and Tribín-Uribe, A. M. (2023). Immigration and violent crime: Evidence from the Colombia-Venezuela border. *Journal of Development Economics*, 162(103039):1–15.

Lagakos, D. and Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2):948–80.

Lebow, J. (2022a). Immigration and occupational downgrading in Colombia. mimeo, Duke University.

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

Lebow, J., Moreno Medina, J., Coral, H., and Mousa, S. (2023). Migrant Exposure and Anti-Migrant Sentiment: The Case of the Venezuelan Exodus. Available at SSRN: https://ssrn.com/abstract=3660641 or http://dx.doi.org/10.2139/ssrn.3660641.

Lewis, E. (2013). Immigration and production technology. *Annual Review of Economics*, 5(1):165–191.

Liu, X. (2010). On the macroeconomic and welfare effects of illegal immigration. *Journal of Economic Dynamics and Control*, 34(12):2547 – 2567.

Mauricia, J. (2019). Venezuelan economic crisis: Crossing Latin American and Caribbean borders. *Migration and Development*, 8(3):437–447.

McDonald, J. T. and Worswick, C. (2015). High-skilled immigration in a globalized labor market. In Chiswick, B. R. and Miller, P. W., editors, *Handbook of the economics of international migration*, volume 1 of *Handbook of the Economics of International Migration*, chapter 11, pages 537 – 583. North-Holland.

Nathan, M. (2014). The wider economic impacts of high-skilled migrants: A survey of the literature for receiving countries. *IZA Journal of Migration and Development*, 3(1):1–20.

Nielsen, C. (2011). Immigrant over-education: Evidence from Denmark. *Journal of Population Economics*, 24(2):499–520.

O'Neil, S. (2019). A Venezuelan refugee crisis. In Stares, P. B., editor, *Preparing for the next foreign policy crisis: What the United States should do*. Center for Preventive Action.

Orefice, G. and Peri, G. (2020). Immigration and worker-firm matching. Working Paper 26860, National Bureau of Economic Research.

Oreopoulos, P. (2011). Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. *American Economic Journal: Economic Policy*, 3(4):148–71.

Ortega, F. and Peri, G. (2014). Openness and income: The roles of trade and migration. *Journal of International Economics*, 92(2):231–251.

Peñaloza, L. (2022). Living with the neighbors: the effect of Venezuelan forced migration on the labor market in Colombia. *Journal for Labour Market Research*, 56(14).

Peri, G. (2012). The effect of immigration on productivity: Evidence from U.S. states. *The Review of Economics and Statistics*, 94(1):348–358.

Peri, G. and Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3):135–169.

Pulido, J. and Święcki, T. (2020). Barriers to mobility or sorting? Sources and aggregate implications of income gaps across sectors in Indonesia. mimeo, University of British Columbia.

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1 – 10.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146.

Rozo, S. V. and Vargas, J. F. (2021). Brothers or invaders? how crisis-driven migrants shape voting behavior. *Journal of Development Economics*, 150:102636.

Rydgren, J. (2004). Mechanisms of exclusion: Ethnic discrimination in the Swedish labour market. *Journal of Ethnic and Migration Studies*, 30(4):697–716.

Santamaria, J. (2022). 'When a stranger shall sojourn with thee': The impact of the Venezuelan exodus on Colombian labor markets. Documentos de trabajo 020046, Alianza EFI.

Tribín, A., Adhvaryu, A., Anzola, C., Ávila, O., Bonilla, L., Castro, J., ..., and Velásquez, S. (2020). Migración desde Venezuela en Colombia: caracterización del fenómeno y análisis de los efectos macroeconómicos. *Revista ESPE - Ensayos sobre Política Económica*, (97):1–74.

Vera, L. (2015). Venezuela 1999–2014: Macro-policy, oil governance and economic performance. *Comparative Economic Studies*, 57(3):539–568.

Weichselbaumer, D. (2017). Discrimination against migrant job applicants in Austria: An experimental study. *German Economic Review*, 18(2):237–265.

Appendix

A Data

A.1 Database

Our dataset comes from the Colombian Wide-scale Integrated Household Survey (GEIH by its acronym in Spanish) produced by the National Administrative Department of Statistics (DANE by its acronym in Spanish), the official statistics bureau in Colombia. The GEIH is the largest monthly statistical operation in the country, with around 21 thousand face-to-face surveys per month in the 23 main metropolitan areas and a rural aggregate. For each household interviewed, the survey provides individual information regarding working conditions (employment status, economic activity, occupation, earnings, expenditures and affiliation to social security) and socio-demographic characteristics such as gender, age, marital status, education, living conditions among other relevant variables.

Since mid-2013 the survey included an additional questionnaire about migration, in which respondents are asked whether they lived in the country five years ago and twelve months ago. Respondents with negative answers in any of those two questions are also asked about the country of precedence. These two questions allow us to identify immigrants and recent immigrants from Venezuela; the former category is what we employ to define an immigrant. For our analysis, we consider only individuals in working age, and exclude those who report being unemployed or outside the labor force (students, retirees, etc.). Table D.1 provides for each year the original sample sizes from GEIH and the resulting ones used in our analysis excluding both unemployed and inactive people. The table also shows the share of the sample in each year that pertains to immigrants and non-migrants for both panels, and some demographic characteristics of both populations such as share of males, average age and average years of schooling.

A.2 Occupations

We reclassified DANE's original set of 99 occupations into 30 new categories, applying the following three rules. First, no reclassification was made for an occupation with a high frequency of workers. Second, occupations with low concentration of workers were added to the most similar job (e.g. firemen, policemen and soldiers were included in security officers) taking as baseline the International Standard Classification of Occupations. Third, few low-frequency occupations with no similar jobs were dropped (e.g. clergy members, athletes). Table D.2 lists the 30 occupational groupings we use. Table D.2 also reports the average years of education in each occupation for both immigrants and non-migrants. Notice that the average years of education tend to be decreasing with the occupational code.

A.3 Labor Income

GEIH includes at least 15 questions related to labor and non-labor earnings. We construct six different labor income measures aggregating the answers to different questions related to labor earnings. Table D.4 lists the 6 different income measures we use, and describes the differences among them. For the main results in this paper we use the measure "Incomesuma". Our results are robust to the choice of any of these income measures.

B Derivations, Proofs and Additional Procedures

B.1 Equilibrium Definitions

The competitive equilibrium of the model for each period t consists on a set of individual occupational choices, total efficiency units of labor of each group in each occupation H_{igt}, total output Y_t and an efficiency wage w_{it} in each occupation such that: i) each individual chooses the occupation that maximizes V_{igt} according to equation (5), ii) the representative firm hires H_{igt} in each occupation to maximize profits; iii) total output Y_t is given by the production function in equation (6); and iv) w_{it} clears each occupational labor market. Therefore, the solution for efficiency wages w_{it} in general equilibrium can be obtained from the following conditions:

1. The definition of the total supply of efficiency units of labor of each group in each occupation, H_{igt}^{supply}, which aggregates individual choices:

$$H_{igt}^{supply} = q_{gt} p_{igt} \mathbb{E}\left(h_{igt}\epsilon_{ig}\right) \tag{B.1}$$

2. The definition of the total demand of efficiency units of labor of each group in each occupation, H_{igt}^{demand}, given by firm profit maximization:

$$H_{igt}^{demand} = A_{it}^{\sigma-1} w_{it}^{-\sigma} Y_t \tag{B.2}$$

3. Total output given by the production function in equation (6), which in equilibrium is also equal to aggregate wages plus total revenues from τ:

$$Y_t = \sum_i \sum_g w_{it} \mathbb{E}\left(h_{igt}\epsilon_{ig}\right) \tag{B.3}$$

4. w_{it} is the value that clears each occupational labor market:

$$H_{igt}^{supply} = H_{igt}^{demand} \tag{B.4}$$

B.2 Proof of Proposition 1

Proof. First, consider the occupational choices when every worker can pick their occupation voluntarily, i.e. with only type I frictions. In this case, the proof essentially mirrors that of Hsieh et al. (2019), so we outline the main steps. The worker's problem at each time t is to choose the occupation with the largest value of $\widetilde{w}_{igt}\epsilon_i$. Hence, $\widetilde{p}_{igt}$ is given by:

$$\begin{aligned}
\widetilde{p}_{igt} &= \Pr\left[\widetilde{w}_{igt}\epsilon_i > \widetilde{w}_{sgt}\epsilon_s\right] \forall s \neq i \\
&= \Pr\left[\epsilon_s < \widetilde{w}_{igt}\epsilon_g / \widetilde{w}_{sgt}\right] \forall s \neq i \\
&= \int F_i\left(\frac{\widetilde{w}_{igt}}{\widetilde{w}_{1gt}}\epsilon_1, \frac{\widetilde{w}_{igt}}{\widetilde{w}_{2gt}}\epsilon_2, \ldots, \epsilon_i, \ldots, \frac{\widetilde{w}_{igt}}{\widetilde{w}_{Ogt}}\epsilon_O\right) d\epsilon
\end{aligned} \tag{B.5}$$

where $F_i(\cdot)$ is the derivative of the cdf funcion given in (3) with respect to its i-th argument. Given the arguments in (B.5) such derivative is:

$$F_i\left(\frac{\widetilde{w}_{igt}}{\widetilde{w}_{1gt}}\epsilon_1, \frac{\widetilde{w}_{igt}}{\widetilde{w}_{2gt}}\epsilon_2, \ldots, \epsilon_i, \ldots, \frac{\widetilde{w}_{igt}}{\widetilde{w}_{Ogt}}\epsilon_O\right) = \theta\epsilon_i^{-\theta-1} \cdot \exp\left[\sum_s \left(\frac{\widetilde{w}_{igt}}{\widetilde{w}_{sgt}}\right)^{-\theta} \epsilon^{-\theta}\right]$$

Notice that $\frac{dF(\epsilon)}{d\epsilon} = \sum_s \left(\frac{\widetilde{w}_{igt}}{\widetilde{w}_{sgt}}\right)^{-\theta} \theta\epsilon_s^{-\theta-1} \exp\left[\sum_s \left(\frac{\widetilde{w}_{igt}}{\widetilde{w}_{sgt}}\right)^{-\theta} \epsilon^{-\theta}\right]$ so evaluating the integral in (B.5) gives:

$$\widetilde{p}_{igt} = \frac{1}{\sum_s \left(\frac{\widetilde{w}_{igt}}{\widetilde{w}_{sgt}}\right)^{-\theta}} \cdot \int dF(\epsilon) = \frac{\widetilde{w}_{igt}^{\theta}}{\sum_s \widetilde{w}_{sgt}^{\theta}} \tag{B.6}$$

Introducing type II frictions implies that a fraction α_{gt} of workers cannot choose their preferred occupation and simply end up randomly allocated in any other occupation. So the share of workers from group g who ends in occupation i at time t, p_{igt} is given by:

$$p_{igt} = (1 - \alpha_{gt})\,\widetilde{p}_{igt} + \alpha_{gt}\left(\frac{1}{M}\right) \tag{B.7}$$

equations (B.6) and (B.7) constitute equation (8) in the text, the first part of Proposition 1.

Now, for the second part, the geometric average of abilities of the group g in an occupation j is given by:

$$\begin{aligned}\exp^{\mathbb{E}[\log(\epsilon_i)]} &= \exp^{\{(1-\delta_{igt})\mathbb{E}[\log(\epsilon_i|\text{choose } i)]+\delta_{igt}\mathbb{E}(\log\epsilon_i)\}} \\ &= \left[\exp^{\mathbb{E}[\log(\epsilon_i|\text{choose } i)]}\right]^{1-\delta_{igt}} \left[\exp^{\mathbb{E}(\log\epsilon_i)}\right]^{\delta_{igt}} \end{aligned}\tag{B.8}$$

with δ_{igt} defined as in the text. Since ϵ_i is distributed Fréchet with parameter θ, $\log\epsilon_i$ is distributed Gumbel with parameter $1/\theta$. Thus, $\mathbb{E}(\log\epsilon_i)$, the unconditional mean of a Gumbel is equal to $\frac{\gamma_{em}}{\theta}$, where $\gamma_{em} \approx 0.5772$ is the Euler–Mascheroni constant; and hence $\exp^{\mathbb{E}(\log\epsilon_i)}$, equal to the unconditional geometric mean of the Fréchet distribution, is given by $\widetilde{\Gamma} \equiv e^{\frac{\gamma_{em}}{\theta}}$. To obtain an expression for $e^{\mathbb{E}\log[\epsilon_i|\text{choose } i]}$, the geometric mean of abilities for individuals who can work in their occupation choice, we proceed as in Hsieh et al. (2019). Denote with stars the variables in the chosen occupation, e.g. ϵ^* denotes the ability in the chosen occupation. Properties of the Fréchet distribution imply that the distribution $G(\epsilon)$ of ϵ^*, the extreme value of ϵ, is also the following Fréchet:

$$G(\epsilon) \equiv \Pr[\epsilon^* < \epsilon] \equiv \exp\left[\sum_s \left(\frac{\widetilde{w}_{gt}^*}{\widetilde{w}_{sgt}}\right)^{-\theta} \epsilon^{-\theta}\right],$$

Denoting $\widetilde{p}_{gt}^* = \frac{\widetilde{w}_{gt}^*}{\sum_s \widetilde{w}_{sgt}^{\theta}}$ according with the definition of $\widetilde{p}_{igt}$ in (B.6), we obtain:

$$\begin{aligned}\mathbb{E}[\epsilon^*] &= \int_0^\infty \epsilon^* dG(\epsilon^*) \\ &= \int_0^\infty \theta\left(\frac{1}{\widetilde{p}_{gt}^*}\right)\epsilon^{*(-\theta-1)} e^{-\left(\frac{1}{\widetilde{p}_{gt}^*}\right)\epsilon^{*-\theta}} d\epsilon^*.\end{aligned}$$

The Gamma function is $\Gamma(\alpha) \equiv \int_0^\infty x^{\alpha-1}e^{-x}dx$. Using the change of variable $x \equiv \frac{1}{p^*}\epsilon^{*-\theta}$, it is possible to show that:

$$\mathbb{E}[\epsilon^*] = \left(\frac{1}{\widetilde{p}_{gt}^*}\right)^{1/\theta} \int_0^\infty x^{-\frac{1}{\theta}} e^{-x} dx = \bar{\Gamma}\left(\frac{1}{\widetilde{p}_{gt}^*}\right)^{1/\theta} \tag{B.9}$$

where $\bar{\Gamma} \equiv \Gamma\left(1 - \frac{1}{\theta}\right)$ is the unconditional mean of the Fréchet distribution with parameter θ. Substituting this result for occupation i in (B.8) and using the unconditional geometric mean of the Fréchet distribution, we obtain:

$$\begin{aligned}
\exp^{\mathbb{E}[\log(\epsilon_i)]} &= \left[\widetilde{\Gamma}\left(\frac{1}{\widetilde{p}_{igt}}\right)^{1/\theta}\right]^{1-\delta_{igt}} \left[\widetilde{\Gamma}\right]^{\delta_{igt}} \\
&= \widetilde{\Gamma}\left(\frac{1}{\widetilde{p}_{igt}}\right)^{\frac{1}{\theta}(1-\delta_{igt})}
\end{aligned} \tag{B.10}$$

(B.10) is equal to equation (9) in the text, the second part of Proposition 1. □

B.3 Procedure to Derive z_{igt}, A_{it}, w_{it} and Y_t in the Observed Economy

In the initial equilibrium, the values of $\widehat{y}_{igt}$, $\widehat{x}_{igt}$, p_{igt} and q_{gt} are observables, τ_{ig} and α_{gt} are our estimated frictions, and σ and θ are calibrated parameters. With this set of information, the following steps describe how to derive the values of z_{igt}, A_{it}, w_{it} and Y_t.

1. With values of p_{igt} and α_{gt}, use equation (8) and the definition of δ_{igt} to derive $\widetilde{p}_{igt}$ and δ_{igt}.
2. With values of $\widehat{y}_{iNt}$, $\widehat{x}_{iNt}$, $\widetilde{p}_{iNt}$, δ_{iNt} and θ, use $\widehat{y}_{iNt} = w_{it}\overline{h}_{iN}\widehat{x}_{iNt}\left(\widetilde{p}_{iNt}\right)^{\frac{1}{\theta}(\delta_{iNt}-1)}$, the geometric average of the income of natives, to derive $w_{it}\overline{h}_{iN}$.
3. Using the identifying assumption $\overline{h}_{ig} = 1$ made for the baseline results (or in the case of the robustness exercises where $\frac{\overline{h}_{iI}}{\overline{h}_{iN}}$ is estimated, the normalization $\overline{h}_{iN} = 1$), w_{it} is equal to $w_{it}\overline{h}_{iN}$.
4. For natives, with values $\widehat{x}_{1Nt}$, w_{1t} and θ, use the normalization $z_{1Nt} = 1$ to derive $\widetilde{w}_{1Nt}^{\theta}$ from the definition of $\widetilde{w}_{igt}$ in the text. For immigrants, use the normalization $z_{1It} = 1$, the assumption $\overline{h}_{1I} = 1$ (or the estimated value of $\frac{\overline{h}_{1I}}{\overline{h}_{1N}}$ in the robustness exercises) to derive $\widetilde{w}_{1It}^{\theta}$ with the values $\widehat{x}_{1It}$, τ_{1I}, $w_{1t}\overline{h}_{1I}$ and θ.
5. Denote $m_{gt} = \sum_s \widetilde{w}_{sgt}^{\theta}$, so $\widetilde{p}_{igt} = \frac{\widetilde{w}_{igt}^{\theta}}{m_{gt}}$. With the values of $\widetilde{p}_{1gt}$ and $\widetilde{w}_{1gt}^{\theta}$, derive m_{gt}.
6. For $i \neq 1$, with the values of m_{gt}, $\widetilde{p}_{iNt}$, τ_{ig}, w_{it}, $\overline{h}_{ig}$, $\widehat{x}_{iNt}$ use the definition of $\widetilde{w}_{igt}$ and $\widetilde{p}_{igt} = \frac{\widetilde{w}_{igt}^{\theta}}{m_{gt}}$ to derive z_{igt}.
7. Properties of the Fréchet distribution imply that if the geometric average of ϵ_{it} is given by equation (9), its arithmetic average is then $\bar{\Gamma}\left[(1-\delta_{igt})\left(\frac{1}{\widetilde{p}_{igt}}\right)^{\frac{1}{\theta}} + \delta_{igt}\right]$, with $\bar{\Gamma}$ defined as in equation (B.9). Thus, compute the supply of efficiency units of labor for each occupation-group H_{igt}^{supply} from equation (B.1) using the definition of the arithmetic average and the values of δ_{igt}, θ, $\widetilde{p}_{igt}$, p_{igt} and q_{gt} and the assumption $\overline{h}_{ig} = 1$ (or the estimated values of $\frac{\overline{h}_{iI}}{\overline{h}_{iN}}$ in the robustness exercises).
8. The aggregate supply of efficiency units of labor for each occupation, H_{it}^{supply}, is simply the sum of the supply for each occupation-group, $H_{it}^{supply} = H_{iIt}^{supply} + H_{iNt}^{supply}$
9. In general equilibrium, total output is equal to aggregate wages (discounting discriminatory taxes for immigrants) plus total revenues from τ, and hence: $Y_t = \sum_i w_{it}H_{it}^{supply}$
10. Finally, using the fact that each occupational labor market must be clear, so $H_{igt}^{demand} = H_{igt}^{supply}$, from the expression of the total demand of efficiency units of labor given in (B.2), derive A_{it} with values of H_{igt}^{supply}, Y_t, w_{it} and σ.

B.4 Procedure to Derive p_{igt}, w_{it} and Y_t for the Counterfactual Economy

In the counterfactual equilibrium, the values of q_{gt}, $\widehat{x}_{igt}$, $\overline{h}_{ig}$, z_{igt}, A_{it}, σ and θ are the same as in the observed economy and τ^c_{ig} and α^c_{gt} are our intended frictions (hereafter superscript c denotes counterfactual values). With this set of information, the following fixed-point algorithm describes how to derive the values of p^c_{igt}, w^c_{it} and Y^c_t for the counterfactual economy.

1. Guess a value of w^c_{it}. Start with $w^c_{it} = w_{it}$.
2. With values of w^c_{it}, τ^c_{ig} $\widehat{x}_{igt}$, $\overline{h}_{ig}$, z_{igt} and θ, derive $\widetilde{p}^c_{igt}$ from its definition.
3. With values of α^c_{gt} and $\widetilde{p}^c_{igt}$ use equation (8) and the definition of δ_{igt} to derive p^c_{igt} and δ^c_{igt}.
4. Use the definition of the arithmetic average of ϵ_{it} in step 7 of Appendix (B.3), and the definition of the supply of efficiency units of labor for each occupation-group from equation (B.1), to compute $H^{supply,c}_{igt}$ with the values of δ^c_{igt}, θ, $\widetilde{p}^c_{igt}$, p^c_{igt} and q_{gt}.
5. The aggregate supply of efficiency units of labor for each occupation, $H^{supply,c}_{it}$, is the sum of the supply for each occupation-group, $H^{supply,c}_{it} = H^{supply,c}_{iIt} + H^{supply,c}_{iNt}$
6. With values of p^c_{igt}, A_{it}, σ, $H^{supply,c}_{igt}$, q_{gt} use the CES production function in equation (6) to derive Y^c_t.
7. From the expression of the total demand of efficiency units of labor given in (B.3), and using the fact $H^{demand}_{igt} = H^{supply}_{igt}$, compute the efficiency wages $w^{c'}_{it}$ compatible with Y^c_t, A_{it}, σ, $H^{supply,c}_{it}$.
8. Substitute w^c_{it} by $w^{c'}_{it}$ in step 1 an repeat steps 2-7 until $w^{c'}_{it} \approx w^c_{it}$.

C Frictions and Reforms for Salaried and Non-Salaried Workers

In this Appendix we proceed to re-estimate our model and perform our counterfactual exercises for both the salaried and non-salaried segments of the labor market, given the findings presented in Section 2 that suggest larger income gaps for salaried workers. Table C.1 shows the results, in which column (1) reproduces the baseline results for comparison. Column (2) presents the estimated frictions (Panel A) and the gains in aggregate productivity from our counterfactual exercises for 2019 (Panel B) assuming that the economy is populated solely by salaried workers, a setting that serves as a proxy of the formal labor market segment; while column (3) replicates the simulation using only non-salaried workers, our proxy for informality.

Consistent with the wider income gaps that immigrants face in the salaried segment, the variance of discriminatory wedges is larger in this segment, 0.24 compared to 0.07 in the informal sector. Furthermore, in comparison to the baseline, the probabilities of involuntary choices α_{gt} decline for both natives and immigrants in both segments, mainly due to the inability of the workforce to reallocate across segments.[43] For salaried workers, the mix of a larger dispersion in discriminatory barriers and lower α_{gt} lead to lower productivity gains from our Reform I with respect to those in the baseline (i.e. the lower α_{gt} dominate the effect). However, because α_{gt} are reduced for both native and immigrant workers, there are larger productivity gains from our Reform II with respect the baseline, given the larger

[43] In the baseline scenario, when frictions are removed, workers can reallocate between formal and informal labor market segments, and thus the estimated value of α_{gt} reflects involuntary occupation choices across the entire economy. However, in the exercises presented in this appendix, workers can only switch between occupations within their respective segments. As a result, the model infers a lower α_{gt}.

dispersion in discriminatory barriers. Note that in the case of Reform I, suppressing only discriminatory barriers has a smaller impact on productivity than in the baseline, even though the discriminatory barriers are more widespread among salaried workers. This is because immigrant workers are underrepresented in the formal sector compared to the entire economy, so the effect of a specific set of barriers is less pronounced. When we apply the estimated barriers from the formal sector to all workers in the economy (results displayed in Panel C), we find that removing only discriminatory barriers now has a greater impact on overall productivity.

On the other hand, in the informal sector we observe fewer frictions for immigrants, including a smaller dispersion in discriminatory wedges and lower α_{gt}. Therefore, both Reforms I and II have a lower impact on aggregate productivity than in the baseline. Furthermore, it is important to note that due to the higher concentration of immigrants in the informal sector, the computation of gains from our reforms by imposing the inferred barriers of the informal sector onto the entire economy (Panel C) reveals lower impacts on aggregate productivity, as opposed to the results seen in the salaried segment.

Table C.1 – Results for salaried and non-salaried segments

	(1) Baseline	(2) Salaried	(3) Non-salaried
A. Magnitudes of estimated frictions			
$Var\left[(1+\tau_{iI})\right]$	0.100	0.237	0.068
$\alpha_{I,2015}$	5.1%	0.9%	0.8%
$\alpha_{I,2016}$	3.8%	0.7%	1.9%
$\alpha_{I,2017}$	5.6%	1.6%	0.7%
$\alpha_{I,2018}$	8.1%	3.5%	1.9%
$\alpha_{I,2019}$	9.2%	2.9%	2.9%
α_N	4.7%	0.1%	0.1%
B. Productivity gains (%) of counterfactual exercises for 2019			
B.1. Reform I			
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \ \forall\ i, t$	0.90	0.69	0.49
Only type I: $\tau_{iI} = 0 \ \forall\ i$	0.61	0.60	0.40
Only type II: $\alpha_{I,t} = 0 \ \forall\ t$	0.23	0.07	0.08
B.2. Reform II			
- Productivity gains (%):	0.38	0.56	0.20
C. Productivity gains (%) of counterfactual exercises for 2019, by imposing the estimated frictions to the full population			
C.1. Reform I			
Both types: $\tau_{iI} = \alpha_{I,t} = 0 \ \forall\ i, t$	0.90	0.79	0.48
Only type I: $\tau_{iI} = 0 \ \forall\ i$	0.61	0.70	0.39
Only type II: $\alpha_{I,t} = 0 \ \forall\ t$	0.23	0.21	0.08
C.2. Reform II			
- Productivity gains (%):	0.38	0.63	0.19

Notes: For the definitions of Reforms I and II, see note in Table 3.

D Additional Tables

Table D.1 – Sample statistics by year

	2015	2016	2017	2018	2019
Panel A: Database using only employees					
Sample size	308.632	302.263	294.083	292.922	308.404
Non-Migrants					
% Non-Migrants	99.5	99.2	98.5	96.8	95.3
% Male	54.9	54.8	55.0	55.2	55.4
Average Age	39.5	39.7	40.1	40.5	41.0
Average Years of Schooling	9.2	9.3	9.3	9.5	9.7
Migrants					
% Migrants	0.4	0.7	1.4	3.1	4.6
% Male	61.2	62.9	59.4	60.2	59.3
Average Age	36.0	35.3	34.5	33.5	33.3
Average Years of Schooling	8.4	8.5	9.4	9.9	10
Panel B: Original Database					
Sample size	787.044	778.238	767.867	762.753	756.063
Non-Migrants					
% Share in the total	99.6	99.2	98.7	97.4	96.0
% Male	47.0	47.1	47.1	47.1	47.0
Average Age	31.2	31.4	31.7	32.0	32.3
Average Years of Schooling	6.8	6.9	7.0	7.1	7.2
Migrants					
% Share in the total	0.4	0.7	1.3	2.6	3.9
% Male	49.8	50.4	49.4	49.5	48.1
Average Age	29.0	28.3	27.7	27.6	27.3
Average Years of Schooling	6.9	6.8	7.4	8.0	8.0

Notes: Panel A refers to the original database excluding unemployed and people outside the labor force. Source: GEIH - DANE.

Table D.2 – Occupation categories

Code	Occupation description	Average years of education	
		Non-migrants	Migrants
1	Administrative professionals and those related to architecture	10.1	10.2
2	Security officers (captains, pilots, inspectors, etc.)	16.9	17.6
3	Health professionals	16.3	15.3
4	Teaching professionals and scientists (chemists, biologists, etc.)	13.4	13.2
5	Authors, composers and photographers	13.9	13
6	Managing directors	12.5	12.4
7	Unclassified administrative workers	10.6	10.7
8	Office accounting workers	11.4	12.4
9	Office assistants	10.3	11.2
10	Office transport workers	8.4	9.4
11	Merchants, owners of wholesale and retail trade	12.8	12.5
12	Head of sales, insurance & real estate agents, brokers and related	9.4	10.4
13	Salesmen	9.3	11.4
14	Hospitality workers	8.3	10.7
15	Cooks, waiters and related	7.4	9.1
16	Personal service workers (helpers)	7.6	8.9
17	Cleaners, security guards and laundry workers	10.3	10.7
18	Hairdressers, beauticians and related workers	11.3	11.4
19	Unclassified service workers	5.0	5.9
20	Agricultural workers	6.8	6.3
21	Metal and factory workers	8.7	9.3
22	Tailors, upholsterers and related workers	8.1	9.5
23	Shoemakers, carpenters and related workers	8.8	9.3
24	Mineral and stone processing operators	8.6	9.5
25	Machinery installers, watchmakers and related workers	9.1	9.7
26	Electricians, operators and handicraft workers	7.4	8.2
27	Construction workers	8.5	9.7
28	Machine operators, packers and related workers	9.4	9.2
29	Vehicle drivers	6.8	9.2
30	Pawns, garbage collectors, polishers and related	6.8	9.2

Table D.3 – Migrant premia by occupation

Code	Short description for occupation	All workers		Only salaried workers	
		Premium	Std. error	Premium	Std. error
1	Administrative professionals and related	-0.431***	0.103	-0.331***	0.102
2	Security officers	-0.644***	0.029	-0.583***	0.021
3	Health professionals	-0.205*	0.105	-0.262**	0.106
4	Professors, chemists, biologists and related	-0.657***	0.103	-0.611***	0.093
5	Authors, composers and photographers	-0.178**	0.073	0.061	0.173
6	Managing directors	-0.197	0.211	-0.261	0.265
7	Unclassified administrative workers	-0.384***	0.046	-0.361***	0.035
8	Office accounting workers	-0.362***	0.109	-0.323***	0.057
9	Office assistants	-0.130	0.081	-0.176	0.114
10	Office transport workers	-0.272***	0.046	-0.341***	0.033
11	Merchants	-0.158***	0.029	-0.276***	0.050
12	Sales Workers	-0.484***	0.033	-0.541***	0.041
13	Salesmen	-0.325***	0.078	-0.322***	0.026
14	Hospitality workers	-0.179	0.106	-0.157***	0.040
15	Cooks, waiters and related	-0.114*	0.063	-0.153***	0.050
16	Personal service workers (helpers)	-0.124**	0.047	-0.135***	0.038
17	Cleaners, security guards and laundry	-0.433***	0.071	-0.356***	0.085
18	Hairdressers, beauticians and related	-0.025	0.037	-0.292***	0.034
19	Unclassified service workers	-0.535***	0.067	-0.605***	0.098
20	Agricultural workers	-0.033***	0.011	-0.174***	0.010
21	Metal and factory workers	-0.183*	0.093	-0.272**	0.099
22	Taylors, upholders and related workers	-0.179***	0.049	-0.418***	0.053
23	Shoemakers, carpenters and related workers	-0.119***	0.037	-0.226***	0.056
24	Mineral and stone processing operators	-0.372***	0.055	-0.368***	0.067
25	Machinery installers, watchmakers and related	-0.236***	0.055	-0.439***	0.021
26	Electricians, operators and handicraft workers	-0.262***	0.043	-0.384***	0.027
27	Building workers	-0.119***	0.032	-0.199***	0.027
28	Machine operators, packers and related	-0.455***	0.037	-0.443***	0.062
29	Vehicle drivers	-0.218***	0.048	-0.361***	0.053
30	Pawns in general, garbage collector and others	-0.293***	0.054	-0.427***	0.090

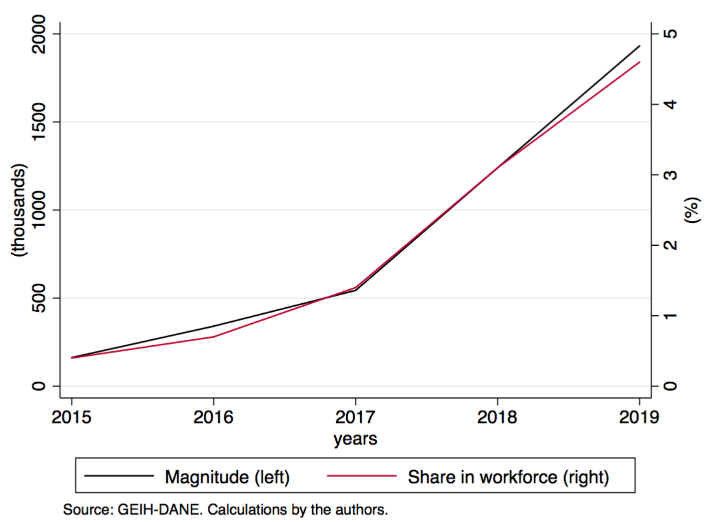
Notes: Columns "premium" and "std. errors" display the coefficient estimates and standard errors respectively of ϕ in the regression (1) for each occupation. Standard errors clustered by municipalities. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table D.4 – Labor income measures

No.	Name	Definition
1	Income	Answer to the question "How much did you earned the last month?"
2	Incomesum	"Income" plus reported extra earnings such as bonuses, subsidies, etc.
3	Incomewoa	"Income" minus extra earnings for those who included them in "Income"
4	Incomesuma	"Incomesum" minus extra earnings for those with missing "Income"
5	Incomeadj10	"Income" plus earnings related to labor/business activities for non-salaried individuals, trimmed in 10k Colombian pesos (3 USD).
6	Incomeadj1	"Income" plus earnings related to labor/business activities for non-salaried individuals, trimmed in 1k Colombian pesos (0.3 USD).

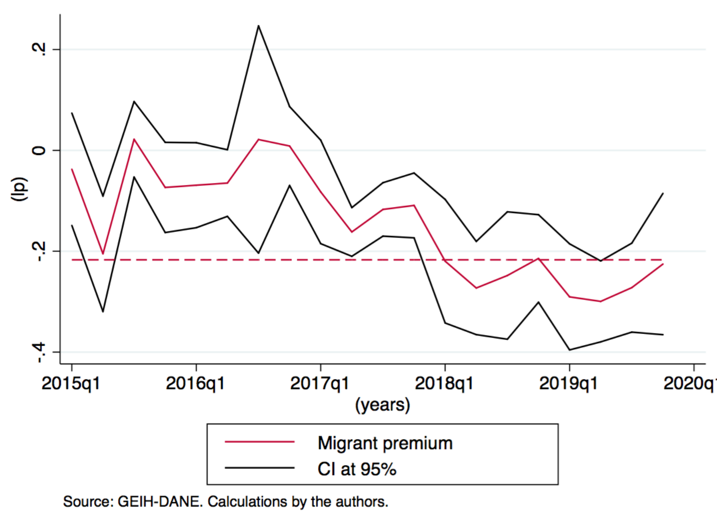
E Additional Figures

Figure E.1 – Migrants Inflow and Share in Colombian Workforce



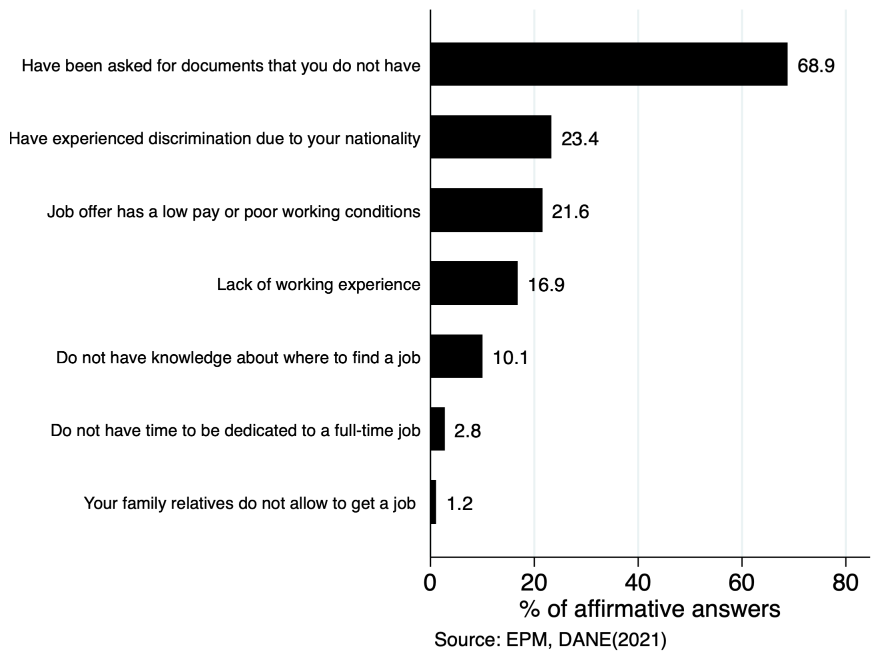
Notes: Figure shows the evolution of the inflow of immigrants from Venezuela to Colombia over 2015-2019 and how much they represent in terms of the Colombian workforce. Observations are weighted by survey expansion weights.

Figure E.2 – Migrant Premium over Time



Notes: Figure shows the evolution of the migrant premium, computed using regression (1) for all workers controlling for occupational fixed-effects, for each quarter of the period 2015-2019, and its corresponding 95% confidence interval. The dashed line corresponds to the migrant premium in the pooled data controlling also for time fixed-effects (column (3) of Table 1). Standard errors are clustered by municipalities, and observations are weighted by survey expansion weights.

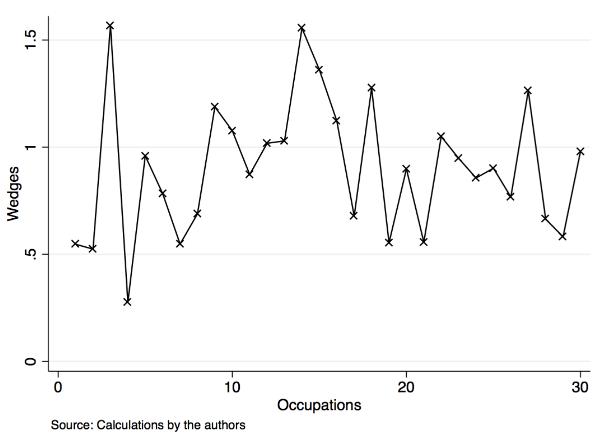
Figure E.3 – Percentage of immigrants facing obstacles to find a job, by type of barrier



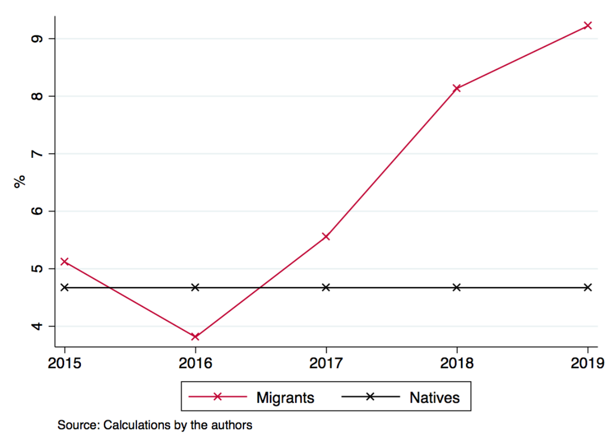
Notes: Figure reproduces the percentage of immigrants who reported facing the displayed obstacles when seeking employment, as revealed by the Survey of Migrants' Perceptions (*Encuesta Pulso de la Migración* in Spanish - EPM) from DANE (2021).

Figure E.4 – Estimated Frictions under the Baseline Parametrization

(a) Discriminatory wedges $(1 + \tau_{iI})$

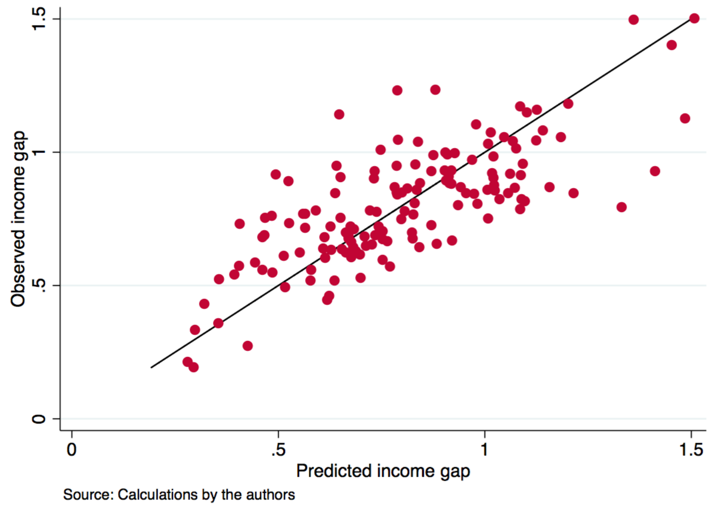


(b) Fractions of involuntary choices α_{gt}



Notes: Figure shows in Panel A the inferred values of the discriminatory wedges for immigrants in each occupation, $(1 + \tau_{iI})$, which have a variance equal to 0.100, and in Panel B the inferred values of the fractions of involuntary choices for immigrants (α_{It}) and for natives (α_N). Both frictions are inferred using $\theta = 2.35$.

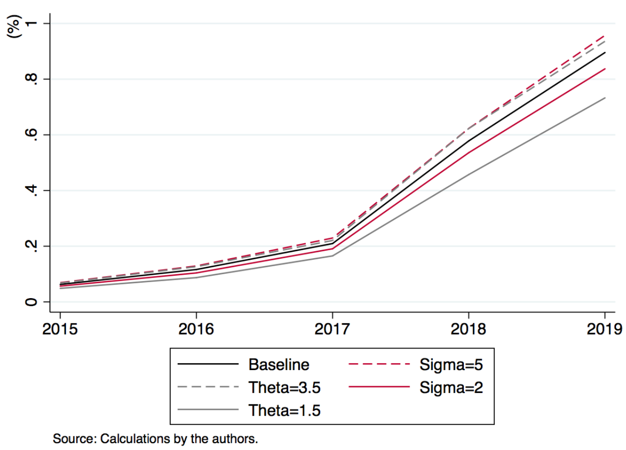
Figure E.5 – Model Fit under Estimated Frictions



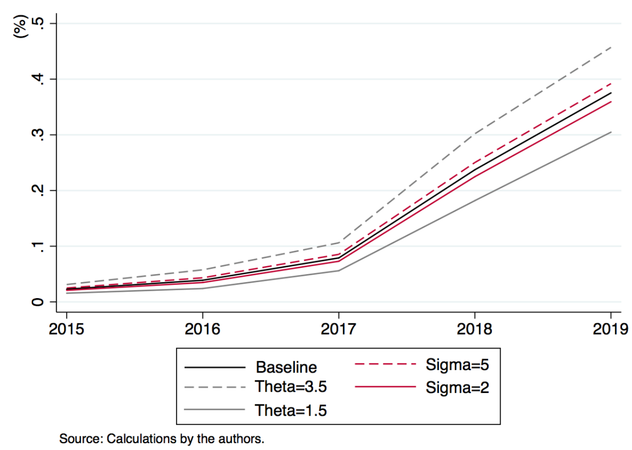
Notes: Figure shows the observed income gaps with predicted income gaps under our set of estimated frictions $\{(1-\tau_{iI}), \alpha_{It}, \alpha_N\}$ using the RHS of equation 10 , and a 45° line.

Figure E.6 – Gains from Reforms by Year: Robustness to Parameterization

(a) Reform I



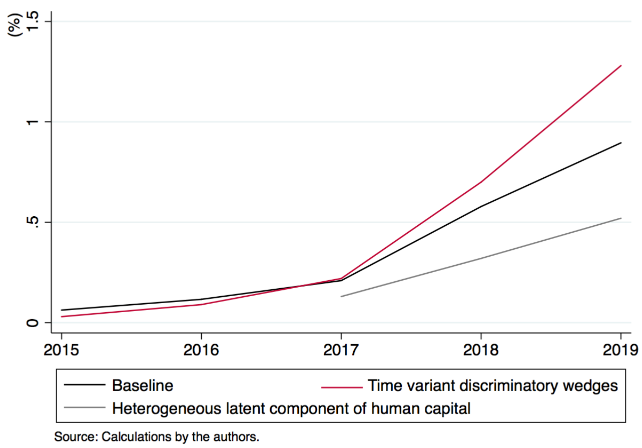
(b) Reform II



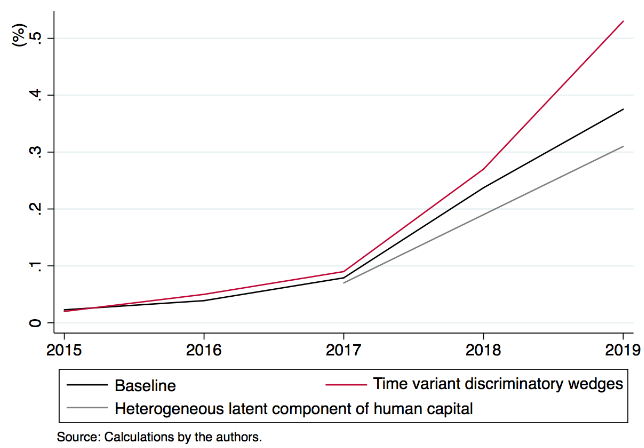
Notes: Figure shows in Panel A the gains from the complete Reform I ($\tau_{iI} = \alpha_{I,t} = 0 \ \forall \ i, t$) for each year of the "Venezuelan exodus" depending on the values θ and σused (baseline parameterization uses $\theta = 2.35$ and $\sigma = 3$) and in Panel B the same comparison for Reform II.

Figure E.7 – Gains from Reforms by Year: Robustness to Specification

(a) Reform I



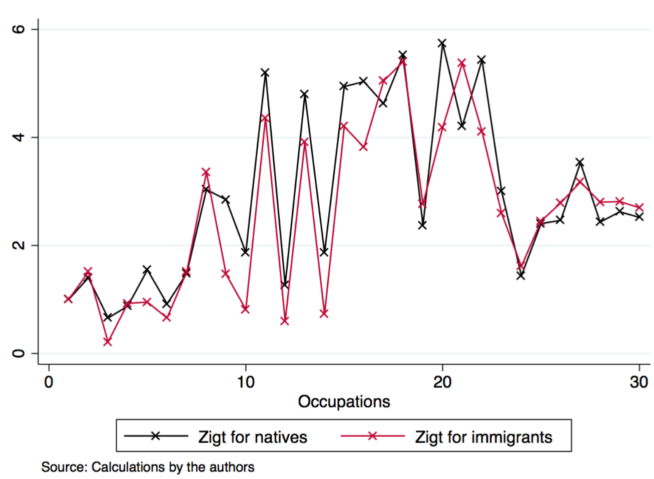
(b) Reform II



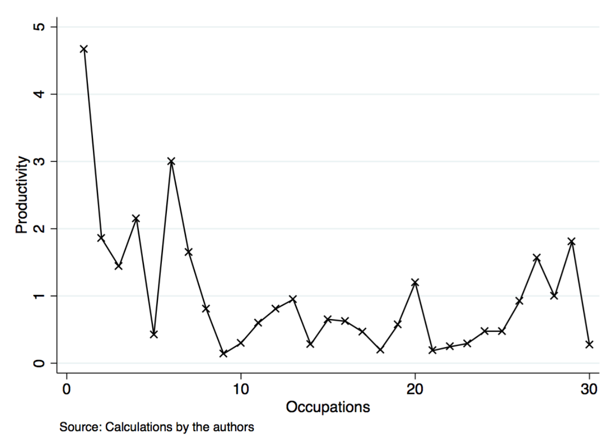
Notes: Figure shows in Panel A the gains from the complete Reform I ($\tau_{iI} = \alpha_{I,t} = 0 \; \forall \; i, t$) for each year of the "Venezuelan exodus" depending on the model specification chosen, and in Panel B the same comparison for Reform II.

Figure E.8 – Average Group-specific Preferences and Productivities in the Baseline

(a) Group-specific Preferences $\bar{z}_{ig}$



(b) Productivities $\bar{A}_t$



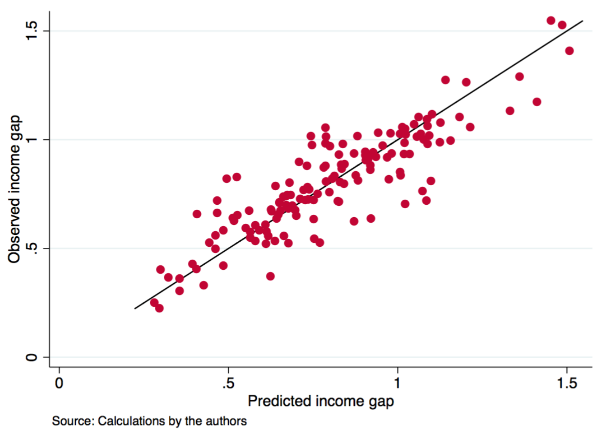
Notes: Figure shows in Panel A the averages over time of the inferred values of the group specific preferences, z_{igt}, which are normalized to $z_{1gt} = 1$, and in Panel B the averages of the productivities A_{igt}. Both measures use the baseline parameterization $\theta = 2.35$ and $\sigma = 3$.

Figure E.9 – Innate Talent Differences $\frac{\overline{h}_{iI}}{\overline{h}_{iN}}$ in a Specification with Wedges Starting in 2017.

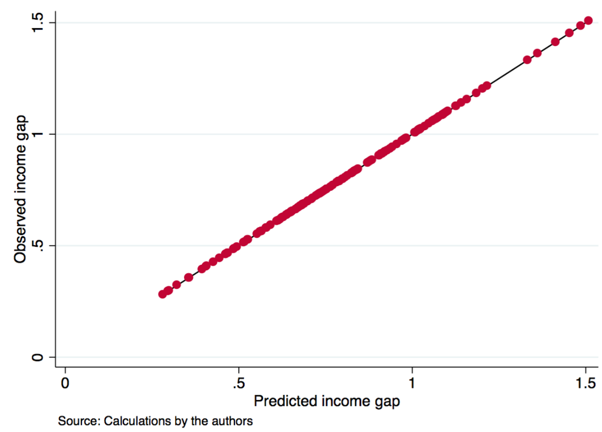
Notes: Figure shows the inferred permanent components of latent human capital of immigrants relative to natives $\frac{\overline{h}_{iI}}{\overline{h}_{iN}}$ in a specification where discriminatory wedges start in 2017 (whose results are reported in Column (2) of Table 5), and the average over occupations in the dashed line. The inference uses the baseline parameterization $\theta = 2.35$ and $\sigma = 3$.

Figure E.10 – Model Fit under Alternative Specifications

(a) Inferring $\frac{\overline{h}_{iI}}{\overline{h}_{iN}}$



(b) Time-variant $(1 + \tau_{itI})$



Notes: Figure shows in Panel A the model fit under the specification with differences in innate talent between natives and immigrants ($\frac{\overline{h}_{iI}}{\overline{h}_{iN}}$); and in Panel B under time-variant discriminatory wedges $(1 + \tau_{itI})$.