Three essays on the Peruvian Labour Market

Juan Manuel del Pozo Segura

Submitted for the degree of Doctor of Philosophy in Economics

University of Sussex

September 2021

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Juan Manuel del Pozo Segura

30th of September 2021

Acknowledgements

This thesis would not have been possible without the help of several people. Among them, I am wholeheartedly thankful to my thesis advisors, Barry Reilly and Matthew Embrey, for their thoughtful guidance and endless patience. Their unconditional support over these years, even amidst the disruptions caused by the COVID pandemic, have turned this challenging academic experience into a great life experience. Other members of the Economics department, especially Andy McKay, Iftikhar Hussain, Sonja Fagernas and Panu Pelkonen, gave me the valuable opportunity to work with them on different research projects. This substantially enhanced my training as a doctoral student and allowed me to explore other areas in Economics besides the one I pursue in this thesis. My fellow PhD colleagues contributed to making this intense period more enjoyable by providing me with a genuine friendship and a welcoming environment to discuss research ideas and life goals. I am also immensely grateful to Efraín Gonzales de Olarte, since this long academic journey would not have been possible without his affectionate support and encouragement during my years at PUCP. Yash holds a very special place througout this process for always encouraging me to challenge myself, for having a sincere interest in my work, and for putting up with my twisted schedule as a PhD student. This thesis is dedicated to my parents for their everlasting love, their hard work to provide me with better life opportunities, and for always believing in me. It is also dedicated to Augusto Saldarriaga, Aníbal Barragán and Diego Patricio who sadly left us too soon because of Covid but their memory and charisma remain intact in me and my family. *¡Un abrazo grande hasta allá!*

UNIVERSITY OF SUSSEX

Juan Manuel del Pozo Segura

Doctor of Philosophy in Economics

Three essays on the Peruvian Labour Market

Abstract

This doctoral thesis studies three phenomena in the labour market which occurred during the period known as the ‘Peruvian Growth Miracle’. From 2005 and until the economic crisis induced by the Covid pandemic, it was characterized by exceptional macroeconomic growth and increased employment and real wages, well above the Latin American average. The first paper decomposes wage differences between Peruvian men and women into parts attributable to differences in labour market attributes and differences in wage structures that women experience in the labour market. Specifically, it studies the evolution across the entire pay distribution of the component associated with gender discrimination experienced by female workers. The effect of those in the informal sector, who are the most affected by this treatment effect, is a central part of the research. The main finding is the presence and the prevalence of ‘sticky floors’, i.e. the confinement of women in low paid jobs, in the informal sector during this period. As important as this, these gaps are primarily explained by the lower reward for their observed characteristics that women obtain in the Peruvian labour market across the wages distribution.

The second and third papers study the labour market impacts of the Venezuelan Exodus, a massive immigration inflow which, according to UNHCR, constituted a forced migratory movement of similar relative magnitude to the Syrian case. The first of these papers studies the impact of this exogenous shock on different labour market outcomes for the natives for both the formal and the informal sector. We use a set of novel econometric techniques that pay attention to aspects of the consistency and inference of the treatment effect estimators, whose discussion remained neglected in most of the literature. Since most of the competition with the natives occurs in low-skilled jobs, a particular emphasis is placed on the effect at the bottom of the skills distribution in the host labour market. On average, the Exodus did not have a statistically significant effect on the outcomes studied. Nevertheless, there is a non-negligible adverse effect for those low-skilled native workers in the urban area more exposed to Venezuelan immigration.

The final chapter focuses on the Venezuelan migrant population. In particular, it investigates how discrimination experienced by these migrants at the hands of Peruvian employees impacts their wages, and the effect that wage inequities have on their perception of workplace discrimination. The study reveals the existence of migrant pay gaps across the wages distribution, and these are more pronounced when we take a more objective measure of discrimination. Most of these gaps are explained by a treatment effect that affects Venezuelan workers based on their nationality, as these have a higher education level than Peruvians and a comparable work experience. The magnitude of this wage-structure effect affects their perception of workplace discrimination. Nonetheless, variables that reflect migrants’ expectations for equality (e.g., education and experience) have a more sizeable effect.

Contents

List of Figures

List of Tables

General introduction

Since the turn of the new millennium, the South-American economy of Peru has performed extraordinarily well in terms of macroeconomic growth. In particular, the period between 2001 and 2019 has been the only time since the beginning of the 20th century that the country experienced uninterrupted GDP per capita growth for more than one decade (see Maddison 2007; World Bank 2021a). Moreover, Peru was one of the few countries that successfully endured the global economic crisis of 2008-09 – the Great Recession. Even though the same period has also been favourable for other Latin American countries, their comparative macroeconomic performance is poor compared to that of Peru (Mendoza 2013). Peru's per capita GDP growth, averaging 4% per year, was more than twice that of the other economies in the region (which averaged 1.4% per year). Over the same time, Peru's annual inflation rate (less than 3%) was consistently the lowest in the region (which averaged 6.8%) per year. Some (e.g. Ross and Peschiera 2015) labelled this exceptional development as the 'Peruvian Growth Miracle'.

This episode started after a decade of recession, hyperinflation and significant deficits in the balance of payments during the 1980s. An essential characteristic of the structural reforms put in place to revert this crisis was a unilaterally substantial trade openness based on primary commodity exports (see Dancourt 1999 for further discussion). The positive terms-of-trade shocks during the 1990s and the favourable external financing conditions played an important role in stabilizing the economy in the subsequent years (Santos and Werner 2015). Nonetheless, the lack of policies for productive diversification aimed at shifting the export basket towards alternative sectors with high value-added, made Peru's growth strategy strongly dependent on the exports of raw minerals and fuels (IMF 2014; United Nations 2019). Throughout the last two decades, the mineral sector has accounted for around 60% of Peru's exports and little more than 10% of the country's tax revenues. Such an economic structure, with undiversified exports and imports mainly comprised of capital goods, is highly vulnerable to fluctuations in terms of trade (Mendoza 2013). More importantly, the weak backward economic linkages inherent in Peru's mining activity, and its low labour intensity relative to countries with similar income levels (Hausmann and Klinger 2008), concentrated its growth benefits within relatively few segments of the population (Gonzales de Olarte 2015; Seminario 2015).

A direct consequence of this growth strategy is the consolidation of a duality in Peru's labour market, with a modern formal sector co-existing with a large low-productive informal sector. This latter sector plays a crucial role in this thesis given its large size compared other economies at similar stages of economic development. More specifically, Peru's informality rate, at 70%, is one of the highest in South America (ILO 2018b). In fact, while the above-mentioned structural reforms focused on improving macroeconomic indicators, they led to a reduction in the proportion of informal workers within the overall economy (Hausmann and Klinger 2008). Nevertheless, the timid attempts of Peru's government to promote formalization in the last years have not been entirely successful (ILO 2014). In addition to the implication that the informal sector represents a loss in tax revenues (Vtyurina 2015), jobs in this sector are precarious, lack access to social benefits and result in pay rates often well below the minimum legal wage. This aspect makes its prevalence troublesome from a policy perspective, as workers in this sector are more vulnerable to poverty and occupational risks (OECD and ILO 2019). Notably, the role of women in its labour market deserves special attention. Compared to countries in Latin America Peru has the highest proportion of women participating in the labour market, which is consistently over 65% (ILO 2018a). Furthermore, along with Mexico, Peru exhibits the widest difference in the proportion of women engaged in the informal sector relative to males (eight percentage points, see ILO 2018b).

The informal labour market has not been the main employment sector for only the Peruvian female workers. It has also been the sector where the vast majority of the approximately 800,000 Venezuelans who suddenly arrived in Peru in 2016 found a job. This exogenous inflow of foreign workers occurred as a consequence of the Venezuelan Exodus, the biggest migration in Latin America's recent history, second only to the Syrian immigration (UNHCR 2019a), and culminated in the progressive collapse of the economic and social conditions in that neighbouring country (Restuccia 2019; Reinhart and Santos 2015). The relevance of this labour supply shock cannot be understated. Previous to this Exodus, Peru had one of the lowest stocks of immigrant workers, comprising less than 0.3% of its population (World Bank 2021b). Hence, an influx equivalent to 2.5% of its population (UNHCR 2020) constitutes an unprecedented event whose effects have had repercussions across different sectors of the economy, such as the housing market, and on internal demand. Nonetheless, importantly for this thesis, is the effect of this Exodus on labour market outcomes of the natives. This Exodus is also unique given the close cultural ties between migrants and natives. However, most Peruvians have been reluctant to welcome new Venezuelans. Indeed, their attitude has been characterized by distrust toward the migrants (Blouin et al. 2019; PUCP 2020), which is comparable to what occurs in countries with a sizeable cultural distance between migrants and natives (Dancygier and Laitin 2014). Most of the literature examining the impact of discrimination on migrant wages has been centred around the migration of culturally distant individuals to countries with high levels of average incomes (e.g., Germany, the United Kingdom). In contrast, it is much rarer to study immigration between countries within the same region where language and culture are shared.

This thesis investigates different aspects of these developments over the last two decades in Peru exploiting a novel set of methodological econometric approaches. Surprisingly, these procedures have been under-analysed in the literature. The first paper examines what part of the gender wage gap is attributed to the differential treatment that females obtain in the labour market. Admittedly, this is not the only paper that provides evidence on gender wage differentials for Peru. However, this chapter advances the previous evidence by decomposing these differentials across the wage distribution to provide deeper policy insights. For instance, the method applied allows to assess the male wage advantages at the bottom of the payment distribution in the informal labour market, and what part of this differential is explained by a wage-structure effect. A large part of this literature has associated this wage-structure component of the gap with labour market discrimination. By studying the yearly changes of this quantile treatment effect, it evaluates if the growth induced by the 'Peruvian Miracle' translated into similar wage increases between men and women, and if it induced more competitive market structures eventuating in a reduction in the degree of female discrimination.

The second chapter exploits a quasi-experimental design represented by the Venezuelan Exodus in Peru to investigate if this sizeable and exogenous shock of foreign workers had a causal effect on the labour market outcomes of the native workers. Namely, it investigates the effect of the Venezuelan influx on hourly wages in the formal and the informal sectors separately. Such an emphasis is justified since changes in these equilibria have more immediate social and political repercussions in the host country. However, immigration also induces adjustments in other labour market outcomes such as informality levels and wage inequality. This study's contribution is also methodological, as it applies an alternative method that circumvents the problems associated with the standard DiD method to estimate the impact of immigration, paying specific attention to the consistency of the estimates obtained and the improved power and sizes of the Wald test used to test the research propositions of key interest. An array of different estimators for the treatment effect of this shock are used. These range from panel data to Synthetic Control Methods (SCM), and allow the robustness of the key results to be determined. Given the demographic and occupational characteristics of the immigrants, the chapter also analyses their effect across different subsamples of natives. Additionally, the study investigates the occupational downgrade that Venezuelan migrants experience in the Peruvian labour market, both in terms of the type of job and the complexity of the tasks that jobs in Peru involve compared to Venezuela. The key finding is that the impact of the Venezuelan influx on the Peruvian labour market is not widespread and is confined to an impact on wages in the informal sector in Lima with a downward adjustment consequent on the increased supply entirely consistent with the predictions of a competitive labour market.

The third and final empirical chapter focuses instead on the wage effects associated with the discrimination that informal Venezuelan workers in Peru are perceived to experience. Our data allow us to study the effect of both subjective and objective discrimination across the wages distribution. On the one hand, it ascribes how self-perceived discrimination is associated with lower migrant wages. On the other hand, it also estimates the magnitude of the gap between Peruvians and Venezuelans. Importantly, a novel decomposition allows unveiling that part of these wage differences which cannot be ascribed to observed characteristics of those who report perceiving discrimination. The chapter also examines how studies the migrants' perception of discrimination is affected by wage inequalities and disparities experienced by Venezuelans. The approach used allow us to quantify Venezuelan awareness of their differential treatment in the labour market on their perception of discrimination, and assess if the variables related to their expectation of equal treatment in the labour market are more important drivers than the objective discrimination. The key finding of this chapter is that unequal wage treatment, as objectively measured by the wage structure effect between Peruvians and Venezuelans, exhibits a statistically significant effect, although modest, on the perception of discrimination. However, non-wage factors like educational level, Venezuelan labour force experience, and the time spent in Peru exert stronger effects that are larger in magnitude on the Venezuelan perception of discrimination.

The thesis also contains a concluding section that brings together the common issues explored in the thesis and discusses the agenda for future research emerging from this research. Each empirical chapter in this thesis is self-contained in nature. As such, it comprises the research questions and explains in detail the relevant context. Also, it provides a detailed description of the methodology applied as well as the data sources used for the empirical work. This latter is crucial, as the high-quality data availability in Peru compares very favourably with that currently available for most other Latin American countries. The availability of such data has made it possible to study topics in this thesis that have important policy relevance, part of which is discussed in more detail in the concluding chapter of this thesis.

Chapter 1

The distribution of the Gender Pay Gap during the 'Peruvian Growth Miracle': an unconditional quantile approach

1.1 Introduction

This study analyzes the gender wage gaps in Peru between 2005 and 2017. This is a period coined by some (Ross and Peschiera 2015) as the 'Peruvian Growth Miracle' given the high (real) growth rates in both per capita GDP (4.5%, more than twice the average for Latin American countries, 1.8%) and hourly-wages in the formal and informal sectors (1.7% and 3.8%, respectively) that Peru experienced during that period.[1] There is a debate in the literature about the behaviour of gender pay gaps during similar growth episodes. Some (Pampel and Tanaka 1986; Boserup 1970) state that these usually experience an inverted U pattern as the economy grows. Thus, the burgeoning industrial sector increases labour demand in male-dominated activities at the initial stages, increasing the wage gap. In later stages, labour demand in the service sector expands, and consequently, job opportunities for women increase, inducing a reduction in the gap. Other empirical studies suggest a more deterministic relationship. On the one hand, Seguino (1997) finds that gender wage gaps in South Korea did not decrease during the 1970-1990 period, characterized by a significant exports boom and a heavy industrialization process. On the other hand, Mitra et al. (2015)) using a panel of 101 countries between 1990 and 2000, find that growth and gender wage gaps have been negatively associated. In recognition of this discussion and the female vulnerability that gender wage gaps imply[2], this study analyzes if the growth of wages during the 'Peruvian Growth Miracle' has been accompanied by a change in gender pay disparities *across* the wage distribution. We are ultimately interested in that part of this differential attributed to the unequal treatment that working women face. In other words, we are concerned with that portion of these gender differences that are attributable to the fact that the labour market systematically pays more to males than females with similar characteristics. Importantly, we conduct the analysis separating the formal and informal sectors, given their inherent labour market differences and the pervasiveness of the latter sector in the Peruvian economy.

Studies that have analyzed gender wage gaps for Peru (and Latin America) share a number of specific characteristics. First, some studies measure differences in mean wages between males and females on the basis of a male dummy estimated by OLS within a Mincerian earnings equation (e.g. Garavito 2011;

[1] An improvement was also found in the education and health sectors (Beteta and Del Pozo 2016), infrastructure (Webb 2013), sanitation services (World Bank 2010) as well as other areas (for a brief survey, see PCM 2013 and INEI 2014b).

[2] As a consequence of this, gender gaps was one of the most referenced aspects by both the Peruvian government (in terms of legal standards, MTPE 2010) and international agencies (CEPAL et al. 2013; OIT and PNUD 2009) during the last decade.

Coppola and Calvo-Gonzalez 2011). These studies portray a partial and imprecise picture of the wage gaps because they ignore disparities at other parts of the wages distribution other than the mean. Second, some other studies also decompose this mean gap using separate gender-specific sub-samples and identify that part attributable to unequal treatment against women (see Castillo 2011; Montes 2007; Yamada et al. 2013 for analysis for Peru and Atal et al. 2009 for Latin American countries). Their policy recommendations are limited because they only focus on a specific statistic of the wages distribution without providing insights on other distribution statistics. Third, to the best of our knowledge, no single study considers the heterogeneity between formal and informal labour markets. Hence, none has explicitly addressed the potential differences in gender pay gaps across these two distinct labour markets.

In order to advance the empirical literature on this topic, this paper applies Recentered Influence Functions (RIF) regressions (Firpo et al. 2007). This estimation method provides the effect of a given covariate at every percentile of the unconditional distribution of the outcome variable for years 2005, 2011 and 2017. This framework permits the analysis of different aspects of the gender wage gap, defined here as the difference between males and females in log hourly wages after controlling for relevant characteristics. Initially, we estimate the magnitude of the gender wage gap not only at the mean but, departing from previous studies, also at different percentiles of its marginal distribution. We are specifically interested in 'sticky floors', the male's advantage occurring at the lower end (typically at the 10th percentile) of the wages distribution, which reflects the tendency of women to be confined and stuck in poorly paid jobs. Equally important are the 'glass ceilings' effects, which are disparities present at the upper end (typically at the 90th percentile) reflecting the limit on women's progression towards high-pay jobs. Secondly, the RIF method enables us to extend the Oaxaca-Blinder (OB) decomposition (Oaxaca 1973 and Blinder 1973) of mean difference to every percentile of the unconditional distribution of log hourly wages. Hence, we decompose the gender wage gaps in the *endowment effect* (that part due to males having better levels of characteristics than females) and the *wage-structure effect* (that part due to unequal treatment against women). This latter component is crucial for policy purposes because it has been typically associated with discrimination in the labour economics literature. Finally, we investigate to what extent these results differ between the formal and informal sectors over the 2005-2017 period. The comparison of these results over the relevant years provides insights on what happens with gender pay disparities during the Peruvian Growth episodes.

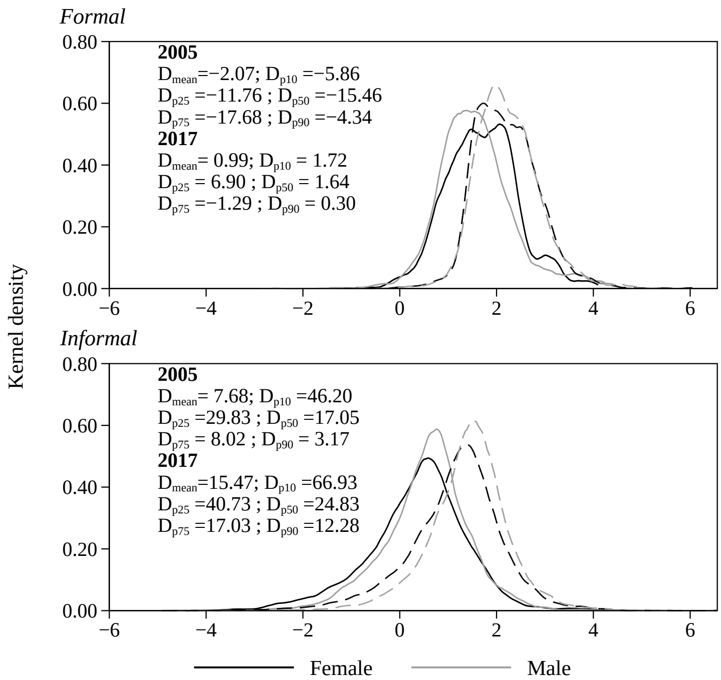
There are four key findings. First, in both sectors, gender pay gaps are statistically significant across the unconditional distribution and favour male workers, even when controlling for the observed characteristics that determine the (log) hourly wages. Second, the gaps are found to be considerably wider in the informal compared to the formal sector. In the latter, for 2017, these estimated effects are around 50% in the 10th percentile and around 15% in the 90th percentile. Third, once we decompose these differences, we find that most of the gender pay gaps are attributable to unequal treatment against women in both formal and informal sectors. So, even though we find inequalities across the wages distribution, these are more striking at the bottom end, reflecting the role of 'sticky floors'. In 2017, the pro-male bias in the informal sector at the bottom and the top deciles were 71% and 21%, respectively. Fourth, women at the bottom of the log hourly-wages distribution in the informal sector face a sizeable sticky floor, one which is sizeable and has not decreased over time. Consequently, this persistence of the treatment effect casts doubt on the overall efficiency of policies put in place by the Peruvian government to alleviate this problem during the 'Peruvian Growth Miracle'.

This chapter is organized as follows. Section 2 provides a concise description of the Peruvian labour market during the 2005-2017 period in terms of inequality, participation and sectoral employment, and outlines government policies implemented to offset gender wage disparities. Section 3 describes the empirical methodology applied, which is the RIF regression, and describes the dataset for this research. Section 4 reports the estimates for the gender wage gaps at different percentiles of the (log.) hourly wages distribution, decomposes these gaps into treatment and endowments effects and presents robustness checks to assess the sensitivity of our results. Section 5discusses the results in terms of their policy implications and outlines potential areas for future research.

1.2 Peru's labour market during the 'Peruvian Growth Miracle'

During the 2005-2017 period, the Peruvian labour market experienced a generalized rightward shift of the log hourly-wage distributions for males and females (Figure 1.1). In the formal sector, females had in 2005 higher wages than males across the distribution. By 2017, this relationship reverted and males earned higher wages at almost every percentile of the wages distribution, although their advantage is modest and reaches a peak of 7% (see Table 1.A1 in the Appendix). In contrast, male to female differences in hourly wages increased in the informal sector around the mean of the distributions and at some of the key order statistics shown, particularly at the lower percentiles.

Figure 1.1 – Wage distributions for males and females by sector, 2005 and 2017

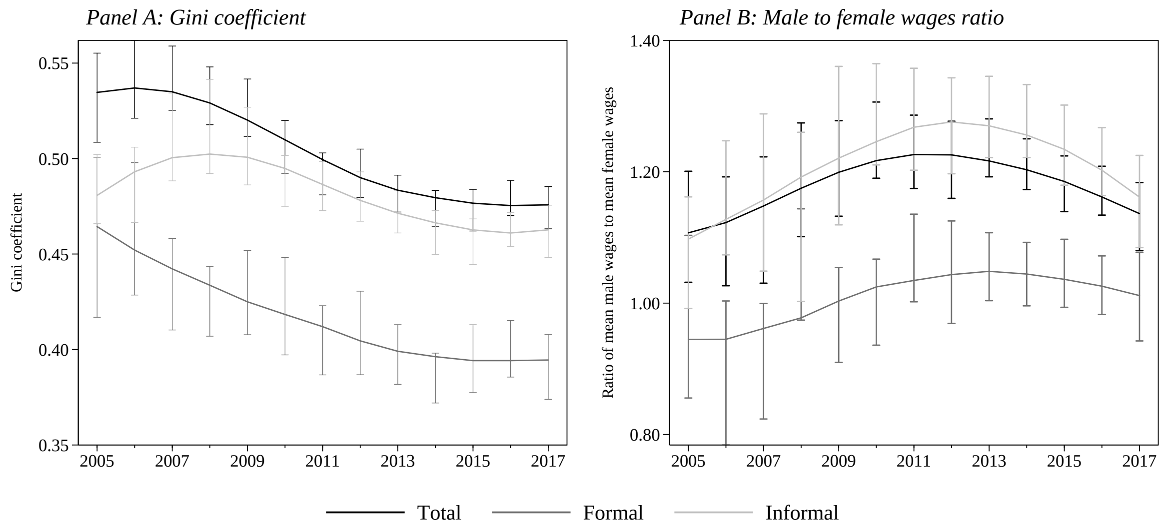


Note: Solid lines correspond to 2005 and dashed lines to 2017. D corresponds to differences between males and females i (observed) hourly wages relative to female's (expressed as percentages), calculated over different points of the distribution. Sample includes individuals between 18 and 65 years. Densities estimated applying sampling weigh
Elaborated by the author based on INEI − National Household Survey (2005−2017)

Consequently, the overall inequality in hourly wages during these 12 years, as measured by the Gini coefficient, has decreased in the formal sector and remained broadly constant in the informal sector (left panel of Figure 1.2). Specifically, the Gini fell from 0.46 to 0.40 in the former sector and remained around 0.47 in the latter. The evolution of the ratio of mean hourly wages (male to female, right panel of the Figure) reflect that, as men in the informal sector have earned wages around 1.1 and 1.2 time higher than their female counterparts throughout the period. Note that the behaviour of the mean gender gap after pooling formal and informal workers (darkest line) resembles that of the informal sector, and its value is similar to what other studies have found for Peru in the last years (MTPE 2014; OIT and PNUD 2009; CEPAL et al. 2013; INEI 2014a).

At the same time, employment rates have remained above 90% throughout and by 2017 represent, respectively, 96% and 94% for males and females of working age (Figure 1.3). The fact that most of the employment in Peru is located in the informal sector can explain these remarkably high rates, given no barriers to entry and the lack of law enforcement for these activities (Aliaga 2010; Freije 2002; Saavedra

Figure 1.2 – Gini index and mean wages ratio by sector, 2005-2017



Note: (Ln.) hourly−wages considered. Sample include individuals between 18 and 65 years. Estimates calculated applying sampling weights and VCE corrected according to survey's complex sample design. Local linear smooth plots shown. Vertical lines correspond to the Taylor−linearized confidence intervals at 95% of significance.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

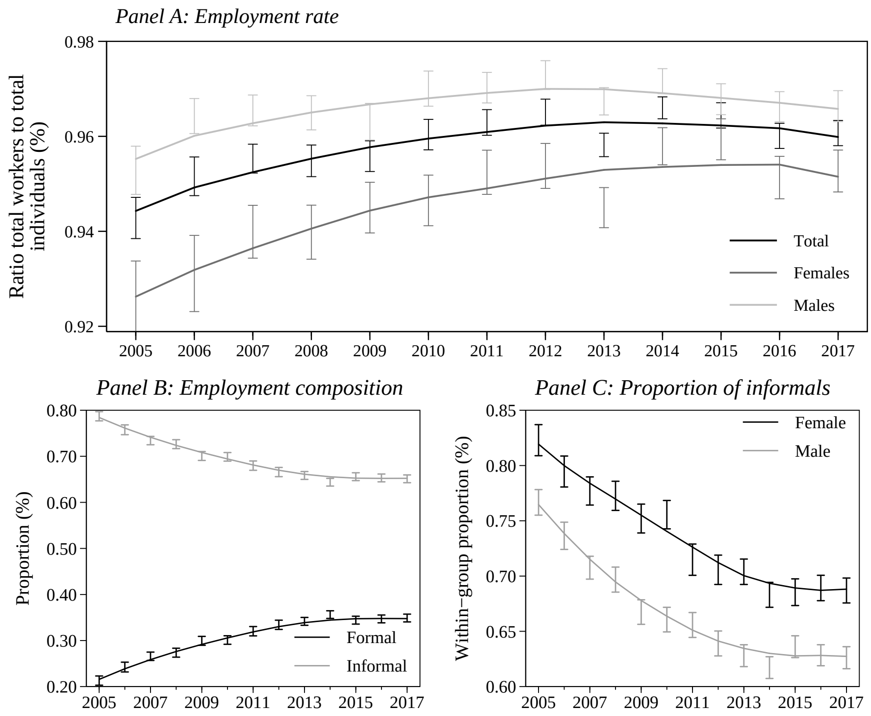
1999). The proportion of informal employees, defined as those not affiliated to a pension system, has been decreasing in a monotonic fashion. Nonetheless, it still represented 72% of the total employment in 2017 (bottom left panel), consistent with ILO (2014) reports. The probability of females engaging in these activities was higher than males over those 13 years (bottom right panel). In fact, women are more likely to work as self-employed and are also more likely to be found in small and medium-sized firms.

This latter reveals a more widespread vulnerability problem for female employment, as a higher relative proportion of these is employed in the informal sector. Lower productivity characterizes jobs in this sector, as well as the absence of institutions such as the minimum wage, social benefits, labour and safety regulations (IADB 2010). In addition, the self-employment occupational category where women are more likely to be engaged is characterized by its low productivity and volatility of incomes (INEI 2014b), while a small firm size in Peru is associated with long hours of work, lack of physical and legal protection, and a higher degree of occupational risks (MTPE 2012). This results in a further over-representation of women in low-paying jobs, which increases inequality, as shown in Figure 1.2.

Data from Peru's National Household Survey for 2015-2017 suggest two key reasons for workers in the informal sector to remain in that sector. Slightly less than half of them are aware of the benefits of being a formal worker but do not consider these sufficiently attractive. A smaller proportion of informal workers (around 35%) would like to be formal but their low earnings do not allow them to afford the bureaucratic and administrative fees to transition to that sector. Nevertheless, there are differences in the importance that males and females ascribe to these two. Disregarding the formal status is more prevalent for men (around 60% in their case), while the low earnings argument is more prevalent for females (around 40%).

The Peruvian government has promoted a set of actions to improve labour market conditions for females during those years. First, the National Employment Policies (2012) comprised specific actions of government agencies to promote females' employment, employability, and entrepreneurship. Second, two National Plans of Equality of Opportunities (2000-2005 and 2006-2010) aimed to assure decent female work through specific instruments. These include equitable labour market legislation and programs to strengthen female's productive capacities. In the end, these two policies resulted in the 2007 Law of Equality of Opportunities between Females and Males (Law N° 28983), which fostered access to employment, training, promotion and safe working conditions for women (MTPE 2015). According to CEPAL et al. (2013), this represented an important advance in legislative design. Two other legal norms stand out. The

Figure 1.3 – Employment participation and composition, 2005-2017



Note: Sample includes individuals between 18 and 65 years. Estimates calculated applying sampling weights and VCE corrected according to the survey's complex sample design. Local linear smooth plots shown. Vertical lines correspond to the Taylor−linearized 95% confidence intervals.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

2007 Supreme Decree N° 027-2007-PCM defines mandatory guidelines in all government institutions related to non-discrimination and the formulation of gender-equality-related public policies. Also, the 2009 Law 294098 provides the right of paid leave for both parents after childbirth. This was done in an attempt to encourage and strengthen a more equitable distribution of housework tasks (MTPE 2010).

The characterizations above suggests the need to move beyond mean log hourly-wages differences to estimate the magnitude of 'glass ceilings' and 'sticky floors'. A significant wage-structure effect could prove that these policies were ineffective in fostering lower discrimination against women. The following section presents the analytical framework applied in order to get address these concerns.

1.3 Methodological framework

This section initially presents the Recentered Influence Function (RIF) regression method and demonstrates how it can extend the OB decomposition to every percentile of the unconditional distribution of the outcome variable. We also briefly discusses here the potential effect that sample selection may exert on the estimates obtained. The dataset used in this study is then described including key changes in the labour market gender composition during the economic growth period of 2005-17.

1.3.1 RIF regressions and OB decomposition

Firpo et al. (2007, 2009); Fortin et al. (2010) outline an OLS-based regression, unconditional quantile regression method, that permits the estimation of the impact of changes in an explanatory factor on the unconditional quantile of the outcome variable. The point of departure for this is the understanding that many common descriptive statistics can be expressed as statistical functionals (i.e. any function of the outcome variable's distribution function). For quantiles, the function $T(F) = F^{-1}(\tau)$, with $F^{-1}(.)$ continuously differentiable at all quantiles τ, gives the value of the outcome variable after inverting the distribution function at a particular quantile τ. The Influence Function (IF), $IF(y;q_\tau)$ where y is the observed outcome variable and q_τ is the distributional statistic of interest (the τ-th quantile), corresponds to the first-order directional derivative of $T(F)$ and assesses the influence of either adding or deleting an individual observation on the quantile of interest without the need to re-calculate the statistic. Adding back the distributional statistic of interest to the IF (since it is centred around zero) yields the Recentered Influence Function (RIF), $RIF(y;\nu) = \nu(F_y) + IF(y;\nu)$.

The expression of the RIF for the τ-th quantile of the unconditional distribution of y is defined as

$$RIF(y,q_\tau) = q_\tau + IF(y,q_\tau) \equiv q_\tau + \frac{\tau - I\{y_i \leq q_\tau\}}{f_y(q_\tau)} \tag{1.1}$$

where $I(.)$ is an indicator function, q_τ is the τth quantile of the unconditional distribution of y, and $f_y(q_\tau)$ is the probability density function of the marginal distribution of the outcome variable evaluated at q_τ. The RIF regression model is then defined as:

$$E[RIF(y,q_\tau)|\mathbf{x}] = \mathbf{x}'\gamma \tag{1.2}$$

where the RIF is assumed to be a linear function of the covariates contained in $\mathbf{x}$. An important property of the RIF is that the mean of the recentered function corresponds to the quantile of interest,and by the law of iterated expectations we have that $q_\tau = E[RIF(y;q_\tau)] = E[E[RIF(y,q_\tau)|\mathbf{x}]] = \mathbf{x}'\gamma$. This expression can be estimated by OLS. Firpo et al. (2009) show that such a regression, with the RIF function for the quantiles of y (wages in this case) replacing y itself as the dependent variable, is in fact an unconditional quantile regression as it estimates the effect of the covariates on the marginal τ-th quantile of y.

For our purposes, the RIF regression allows estimating the unconditional gap at every quantile of the wages. This provides a way to investigate the presence of 'sticky floors' or 'glass ceiling' in the pay distribution. This latter suggests a crucial conceptual difference relative to the conditional regression approach (Koenker and Bassett 1978, 1982; Koenker and Hallock 2001). In the conditional case, the specification of the covariates determines the quantile (given it is conditional on the covariates contained in the specification), but in the unconditional case the quantile is independent of the covariates used.[3] This, in turn, affects the scope of policy recommendations that conditional quantile regression provides. Whereas conditional quantile models do not yield generalizable results to a population or broader policy context (since these regression models do not average up to their unconditional population counterparts), unconditional quantile models provide consistent estimates of the impact of explanatory variables on the unconditional distribution of the outcome variable.

The computation of the components described in Equation 3.1 is required before OLS estimation of Equation 1.2. In order to do so, we can use the analogy principle: compute the sample $\hat{q}_\tau$ and then estimate the density value at point $\hat{f}(\hat{q}_\tau)$ using non-parametric kernel density methods. An estimate of the RIF for each observation is then obtained by plugging the density estimates into Equation 3.1. The sampling variances for the RIF regression estimates can be computed using bootstrapping techniques. However, each RIF variable must be recomputed based on each new sample generated as part of the bootstrapping exercise. For the current application, we use 250 replications for the bootstrapping routine (see below).

If we assume a linear model for wages $w_s = \mathbf{x}_s\beta_s + \varepsilon_s$ for $s = M,F$, where w are the observed wages, $\mathbf{x}$ is a vector of covariates, β a vector of parameters, ε_s are unobservable characteristics and $s = \{Males, Females\}$, we are able to apply the typical Oaxaca (OB) decompositions (Oaxaca 1973) which, under plausible regu-

[3] To the extent that quantile regression implies a nonlinear operator, the law of iterated expectations (LIE) does not hold and hence coefficients estimated under this framework are conditional to that particular model. On the other hand, to the extent that RIF regression amounts to OLS with a transformed dependent variable, properties of linear regression, namely LIE, hold. This then enables the use of linear decompositions.

larity conditions[4], allows for an additive separation into treatment and endowment effects of gender differences in mean wages. However, the application of OLS-based RIF procedures allow extending this standard decomposition from unconditional averages differences to unconditional τth percentile-specific differences

$$\Delta q_\tau = \underbrace{\left(\overline{\mathbf{x}}_M' - \overline{\mathbf{x}}_F'\right)\hat{\beta}_{M,\tau}}_{\text{endowment effect}} + \underbrace{\overline{\mathbf{x}}_F'\left(\hat{\beta}_{M,\tau} - \hat{\beta}_{F,\tau}\right)}_{\text{treatment effect}} \tag{1.3}$$

This extension is possible since a critical assumption of the OB method, namely that the conditional expectation function of wages is linear (in parameters), holds under RIF. The first term, the endowment effect, captures that part of the gap attributed to the more favourable distribution of male relative to female characteristics. The second term, the wage-structure (treatment) effect, reflects the higher returns that males systematically secure in the labour market relative to comparable females. This component has also been defined as discrimination against women.[5] Compared to alternative decomposition methods, mainly Machado and Mata (2005), RIF decomposition is exact and exploits the existence of a residual obtained through numerical integration of the conditional to the marginal distribution. However, it is obvious that Equation 1.3 is subject to the standard 'index problem' and, as specified above, assumes that the male wage structure provides the wage distribution prevailing in the absence of any differential.

Studies for Peru have almost exclusively relied on the classical OB mean approach to decompose gender wage gaps. For instance, Montes (2007) analyzed the urban areas during the 1997-2000 period and found that the small negative treatment effect (i.e. favouring women) in 1997 disappeared by 2000. Castillo (2011) studied wage gaps on a national level for the 2003-2009 period and reported that out of the total gaps, which ranged between 15% and 22%, the treatment effect (favouring males) represents about a half of them. This is a lower estimate than MTPE (2014) reports for 2012: out of the 33% gender gap, 29% is found to be due to a treatment effect. Also, Atal et al. (2009) decompose gaps for 18 Latin American countries and find that the treatment effect accounts for 20% of the gap, whereas the endowment effect represents a vanishing small -2%. This indicates that, on average, females had better characteristics than their male counterparts. In fact, these also report that the magnitude of treatment effect in Peru is the fifth-largest among the set of countries analyzed. At the same time, few studies in Latin America have applied the RIF OB approach to study gender wage gaps. On the one hand, Salardi (2012) analyzes pay disparities for Brazil between 1986 and 2006. She finds that gender pay gaps favoured males at the bottom at the top and of the distribution in the former year. By 2006 this U shaped pattern disappeared, although the gaps remained across the distribution. The decomposition results reported suggest that in both years the gaps were attributable to the presence of treatment effects, mainly at the extremes, and this component has declined in time. On the other hand, Pacheco (2013) studied urban Nicaragua in 2005 and 2009 and reports that in 2005 the treatment effect attributed to being female yielded different magnitudes at different percentiles. By 2009, the treatment component shrank at the lower and upper parts of the wages distribution.

An important assumption of unconditional RIF regressions is that variables included in the estimation are exogenous. However, wages are likely to be characterized by a selection process that can lead to inconsistent estimates. In this particular application, this sample selection is twofold. In the first place, the selection into employment is induced by the job search, which affects participation in the labour market (Gronau 1973; Heckman 1979). In the second place, given the high informality rates, there is also selection

[4] According to Fortin et al. (2010), these are: mutually exclusive groups, outcomes defined according to a defined structural form, feasibility of a simple counterfactual treatment, existence of overlapping support and invariance of conditional distributions (construction of the counterfactual for workers in B that would have prevailed if they were paid like those in A assumes that the conditional wage distribution can be extrapolated).

[5] Note that this decomposition implies the creation of a counterfactual wage, which in this case corresponds to $\overline{\mathbf{x}}_F'\hat{\beta}_{M,\tau}$ and is interpreted as the predicted wage for the females if their characterstics were paid according to the labour market returns for males at the τth percentile of the unconditional distribution.

into the formal and informal sector since we are estimating equations separately for both sectors.[6] For semi-parametric estimators, including polynomials of the Heckman-type selection correction terms into the outcome equations has been suggested (Buchinsky 1998b; Newey 1988). This strategy is also a possibility for RIF regressions, but there are challenges to its implementation. On the one hand, there is the need for variables that impact the probability that the individual is working but does not affect the level of wages. On the other hand, even if a suitable instrument were available, there is a need for an identification strategy for the constant term in the log wage equation because it is conflated with the constant of the higher polynomial selections terms. A way to circumvent this is to specify a RIF regression model with a rich set of covariates. It is plausible that these included observed characteristics are correlated with those unobservables that affect participation, resulting in a sample selection on observables. This can only be assumed but not tested. Similar to Salardi (2012) and Pacheco (2013), the results presented here do not include an explicit correction for selection bias and hence do not claim to reflect causal effects. It should be noted that the literature on selection correction within a RIF framework remains unsettled with no definitive view on the appropriate methodology either within the context of a single or multiple selection correction contexts (as would be the case here). However, the approach adopted here provides a detailed and thorough description of the distribution of gender pay gaps and their evolution over time. Furthermore, it is implausible to conceive that the results reported here are driven exclusively by selection effects. Such selection effects, if present, are likely to be small in magnitude and unlikely to distort the broad concluding narrative outlined here.

1.3.2 Data

The dataset used in this study is derived from three repeated cross-sections from Peru's National Household Survey (ENAHO, according to its initials in Spanish), collected by the National Statistical Office (Institute of Statistics and Informatics) for the years 2005, 2011 and 2017. Three reasons underlie the choice of this dataset. First, this constitutes Peru's government primary source of information for obtaining employment indicators given its representativeness of both formal and informal sectors (see Figure 1.A1 in the Appendix for the number of observations in each sector). Second, it allows conceptually comparable estimates to be obtained through the years since its sampling design has been unaltered since 2005. Third, it includes detailed information on hourly wages (which avoids measurement error problems in the key dependent variable for this study) and individual-level demographic and labour market characteristics.

The availability of this information allows us to estimate three models which successively provide refined estimates of the gender gaps. The first includes a set of demographic variables typically included in Mincerian equations; the second model adds a vector of dummies for the industry where the individual is employed. The third (full) model completes the specification by including a vector of dummies for the individual's occupation.[7] So we have

$$\omega_i = \beta_0 + \delta_1 sex_i + \beta_1 \mathbf{age}_i + \beta_2 \mathbf{educ}_i + \delta_2 urban_i + \mathbf{\Phi}_1 \mathbf{Indust}_i + \mathbf{\Phi}_2 \mathbf{Occup}_i + \upsilon \tag{1.4}$$

where $\omega \equiv \ln\left(\frac{w}{H}\right)$ corresponds to the logarithm of the individual's hourly-wage (for their principal and secondary activity, including monetary wages and in-kind payments), *sex* is a dummy which equals 1 if the individual is male and 0 otherwise, **age** corresponds to a second order polynomial of individual's age (in years), **educ** corresponds to a second order polynomial of individual's schooling (in years), *urban* is a dummy which equals 1 if the individual is located in an urban area and 0 otherwise, **Indust** is a vector

[6] Whether the decisions of entering into the labour market and working in the formal or informal sectors are taken simultaneously or as a two-stage process only adds complexity to the estimation but does not change the fact that sample selection remains a plausible process here.

[7] ENAHO also includes the employment size of the firm where the individual works, which according to some studies (e.g. World Bank 2015 and Távara et al. 2014) has a positive relationship with a firm's productivity levels in Peru. However, we do not use this variable in the regressions since this is only for available for individuals who work in the private sector.

of eight industry dummies and **Occup** is a vector of eight occupation dummies. These latter two follow international classifications in order to facilitate comparability with other studies. On the one hand, the vector of industry dummies is a reduced version of the International Standard Industrial Classification of All Economic Activities (ISIC), Revision 3.[8] On the other hand, the vector of occupation dummies is a reduced version of the International Standard Classification of Occupations (ISCO-08) from the International Labour Organization. [9]

In the same vein as Aktas and Uysal (2012); Albrecht et al. (2003, 2009); Buchinsky (1998a), we restrict the sample to those individuals aged between 18 and 65 years old who are employed. Related to the age interval, the lower bound corresponds to the minimum legal age in Peru[10]and the upper to the legal retirement age. In regard to the employment variable, INEI considers as employed those who had a job, worked more than 15 hours per week and declared a positive wage.[11]

Informality is approximated by the absence of affiliation to a pension system. Admittedly there is not a unique definition of informal status. Nevertheless, according to Freije (2002), p. 2: "Informal workers lack almost every form of social protection [...] No access to the pensions system protection make informal workers unable to retire and force them to work longer perhaps under decreasing productivity of their human capital". Since this description characterizes an important part of informal workers in Peru rather than alternative definitions (e.g. working on a firm without accounting books, not receiving an invoice for their professional services, working less than 40 hours per week, etc.), we choose this as our indicator of informality. Also, we make a distinction in the occupational categories included in both of these sectors. On the one hand, we consider as formal only those white and blue-collar workers, both in the public and private sector, who are in a pensionable job. On the other hand, we consider informal employers, independent or white and blue-collar workers without a pensionable job. We exclude from the analysis family workers and unpaid family workers. Later, we undertake robustness checks to assess whether our results change when considering alternative occupational categories for the formal and informal workers (see section 1.4).

Information in ENAHO was recorded after a probabilistic, stratified and multi-stage sampling to provide representative estimates of the population. Consequently, it is necessary to take this survey design into account by, on the one hand, applying the sampling weights (which reflects the inverse of the probability that the observation is included in the sample)[12] and, on the other hand, adjusting the standard errors for stratification and clustering of the sample. Failure to do the former would result in biased estimates of population parameters, while failure to do the latter would result in artificially lower standard errors and, in turn, in misleadingly high test statistics (Chen and Shen 2015; Kreuter and Valliant 2007; Kolenikov 2010). To the extent that we are primarily estimating RIF regressions and the terms of the Oaxaca Blinder decomposition, and because to our best knowledge, the Taylor-linearization formulas to estimate analytical VCE under survey design are not available for these, we calculate the standard errors and the (percentile-method) confidence intervals of the estimates by bootstrapping which, unlike Jacknife or Balance Repeated Replication techniques, provides consistent VCE estimates in the case of non-smooth statistics such as

[8]The original classification (available at http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=2) considers 17 major groups but we add them into 8: Agriculture, forestry, and fishing; Mining and quarrying; Manufacturing and Public Utilities; Construction; Wholesale and Retail Trade, Hotels and Restaurants; Transport, Storage, and Communication; Finance, Insurance, and Real Estate; Community, Social and Personal Services.

[9]The original classification (available at http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm) considers ten major groups but, provided that the number of females as managers and armed forces who are non-missing was null for both years, we merge these major groups with that of Professionals. In the end, we have eight major groups: Managers, Professionals and Armed forces, Technicians and associates, Clerks, Service and sales workers, Skilled agricultural and fishery workers, Craft and related trades workers, Plant and machine operators and assemblers and Elementary occupations.

[10]The minimum working age (without the need of explicit parental authorization) is 14 years old but this varies between different industries. For example., this minimum age is 15 in agriculture; 16 in mining and 17 in fishing. Therefore, we take the minimum common age.

[11]INEI's definition of unemployment corresponds to those who did not have a job but was looking for it. It also includes those who were not looking for a job but were engaged in productive activities which might not involve a monetary payment.

[12]These sampling weights, included within the dataset, have already been adjusted by INEI to correct for non-representativeness and non-response. There is no need to apply further calibration or poststratification to this variable (Chen and Shen, 2015).

quantiles. We use the Stata routine devised by Kolenikov (2010), which rescales the bootstrap replicates to ensure that the bootstrapping scheme resembles the original sampling in every iteration. In contrast to the more naive paired-bootstrapping, this will provide the correct VCE estimates for survey data.

Before turning to the econometric estimation, it is useful to review how the labour force composition changed between 2005 and 2017 in terms of some key variables (Table 1.1). In the case of formal workers, there is a statistically significant increase in the proportion of workers aged between 18 and 25 years and those aged between 56 and 65 years at the expense of the intermediate age groups. Specifically, the proportion of males in the youngest age bracket increased from 9% in 2005 to 13% in 2017 and from 9% to 15% in the case of the females. Turning to the employment across industries, female participation increased mainly in the wholesale and retail industry (by 7 percentage points), while male participation mainly increased in construction (4 percentage points) industry. Female workers increased by five percentage points their participation in Services and sales and Elementary jobs (both of which are characterized by lower wages) at the expense of their participation in Managerial and professional occupations. A different pattern emerges when we focus on the informal sector. The largest increase in the age composition in this sector occurred for the oldest age bracket for both male and female workers (around 4 percentage points in both cases). Wholesale and retail is the industry where more than half of female workers are allocated while one-third of male workers are employed in agriculture, forestry and fishing industry. Informal female workers increased their participation mainly in the Community, social and personal services industry (4 percentage points) and males increased their participation mainly in two industries: Transport, storage and communication (4 percentage points) and Construction (5 percentage points). More women reallocated in the Service and sales occupations (8 percentage points) in the informal sector, and in comparison males increased their participation mainly in the Machine operators occupations(4 percentage points) over the period.

Table 1.1 – Participation rates by sex and sector, 2005 and 2017

	2005				**2017**				**Diff.**			
	Females		Males		Females		Males		Females		Males	
Formal												
Age Group (Years)												
18 25	9.390	(1.030)	9.113	(0.730)	15.668	(0.782)	13.612	(0.530)	6.279***	(1.293)	4.500***	(0.903)
26 35	38.229	(2.033)	31.854	(1.203)	31.238	(0.882)	29.908	(0.709)	-6.991***	(2.216)	-1.946	(1.396)
36 45	32.267	(1.776)	32.476	(1.238)	28.504	(0.871)	27.254	(0.684)	-3.763*	(1.980)	-5.223***	(1.414)
46 55	15.722	(1.273)	18.684	(0.957)	16.771	(0.651)	18.115	(0.531)	1.049	(1.429)	-0.569	(1.095)
56 65	4.393	(0.811)	7.873	(0.684)	7.819	(0.480)	11.111	(0.411)	3.426***	(0.942)	3.238***	(0.798)
Ethnicity (D)												
Not indigenous	95.825	(0.620)	88.521	(0.781)	95.362	(0.452)	90.156	(0.505)	-0.463	(0.772)	1.635*	(0.937)
Indigenous	4.175	(0.620)	11.479	(0.781)	4.638	(0.452)	9.844	(0.505)	0.463	(0.772)	-1.635*	(0.937)
Industry												
Agric., forest. and fish	2.705	(0.811)	6.053	(0.707)	3.269	(0.305)	7.074	(0.400)	0.564	(0.871)	1.021	(0.832)
Mining and Quarrying	0.897	(0.550)	4.982	(0.774)	0.562	(0.151)	4.158	(0.303)	-0.335	(0.570)	-0.825	(0.834)
Manufacturing	12.776	(1.353)	14.870	(0.963)	10.439	(0.666)	14.825	(0.620)	-2.337	(1.509)	-0.045	(1.153)
Construction	1.607	(0.428)	7.910	(0.657)	3.064	(0.334)	11.998	(0.478)	1.457***	(0.543)	4.088***	(0.813)
Wholesale and Retail	9.094	(1.163)	11.487	(0.883)	16.962	(0.770)	11.432	(0.530)	7.867***	(1.395)	-0.055	(1.039)
Transport, Storage, and Comm.	3.052	(1.318)	4.589	(0.594)	4.742	(0.495)	6.401	(0.370)	1.690	(1.409)	1.812***	(0.701)
Finance, Insurance, and Real Estate	9.887	(1.401)	10.987	(1.176)	12.300	(0.689)	12.843	(0.515)	2.414	(1.567)	1.856	(1.285)
Community, Social and Personal Svs	59.983	(2.375)	39.122	(1.309)	48.662	(1.035)	31.269	(0.720)	-11.321***	(2.604)	-7.853***	(1.505)
Occupation												
Managers, Profess. and Armed forces	44.584	(2.175)	26.712	(1.337)	32.402	(0.982)	19.681	(0.618)	-12.182***	(2.394)	-7.031***	(1.474)
Technicians and related	14.378	(1.521)	12.160	(0.970)	14.051	(0.726)	14.498	(0.544)	-0.328	(1.688)	2.338**	(1.116)
Sales clerks	21.213	(1.887)	11.652	(0.742)	25.080	(0.906)	15.390	(0.576)	3.867*	(2.097)	3.738***	(0.939)
Service and sales workers	5.480	(0.869)	6.577	(0.606)	10.655	(0.649)	8.580	(0.422)	5.175***	(1.085)	2.004***	(0.740)
Skilled agric. and fish. workers	0.000	(0.000)	0.361	(0.160)	0.083	(0.040)	0.479	(0.133)	0.083**	(0.040)	0.118	(0.208)

Continued on next page

Table 1.1 – Participation rates by sex and sector, 2005 and 2017 (*continued from previous page*)

	2005				**2011**				**Diff.**			
	Females		Males		Females		Males		Females		Males	
Craft and related trades workers	4.348	(0.793)	12.504	(0.909)	2.731	(0.356)	10.527	(0.492)	-1.617*	(0.869)	-1.977*	(1.033)
Machine operators	0.648	(0.295)	7.951	(0.634)	0.586	(0.171)	9.364	(0.466)	-0.062	(0.341)	1.412*	(0.787)
Elem. occupations	9.349	(1.423)	22.083	(1.061)	14.412	(0.726)	21.480	(0.637)	5.063***	(1.602)	-0.603	(1.247)
Informal												
Age Group (Years)												
18 25	18.681	(0.613)	22.504	(0.567)	18.225	(0.507)	19.605	(0.425)	-0.457	(0.800)	-2.899***	(0.714)
26 35	28.938	(0.753)	29.422	(0.613)	23.305	(0.532)	24.673	(0.472)	-5.633***	(0.922)	-4.749***	(0.774)
36 45	28.203	(0.742)	24.402	(0.516)	26.383	(0.550)	26.585	(0.485)	-1.820**	(0.925)	2.184***	(0.711)
46 55	15.715	(0.512)	14.706	(0.395)	19.363	(0.428)	16.433	(0.354)	3.648***	(0.668)	1.727***	(0.530)
56 65	8.463	(0.370)	8.966	(0.298)	12.724	(0.371)	12.703	(0.315)	4.262***	(0.525)	3.737***	(0.437)
Ethnicity (D)												
Not indigenous	79.672	(0.739)	77.582	(0.762)	76.540	(0.681)	75.107	(0.674)	-3.132***	(1.039)	-2.475**	(1.071)
Indigenous	20.328	(0.739)	22.418	(0.762)	23.460	(0.681)	24.893	(0.674)	3.132***	(1.039)	2.475**	(1.071)
Industry												
Agric., forest. and fish	14.668	(0.574)	38.186	(0.749)	14.984	(0.425)	35.217	(0.626)	0.315	(0.815)	-2.969**	(1.294)
Mining and Quarrying	0.074	(0.029)	0.815	(0.149)	0.123	(0.043)	1.486	(0.171)	0.049	(0.052)	0.671***	(0.228)
Manufacturing	12.264	(0.576)	11.634	(0.485)	10.410	(0.431)	8.966	(0.373)	-1.854**	(0.720)	-2.668***	(0.638)
Construction	0.735	(0.161)	9.847	(0.404)	0.778	(0.098)	15.451	(0.453)	0.042	(0.188)	5.604***	(0.622)
Wholesale and Retail	50.418	(0.823)	14.962	(0.511)	54.097	(0.654)	12.943	(0.403)	3.679***	(1.072)	-2.018***	(0.684)
Transport, Storage, and Comm.	1.803	(0.220)	12.172	(0.429)	1.715	(0.177)	16.195	(0.449)	-0.088	(0.284)	4.023***	(0.648)
Finance, Insurance, and Real Estate	3.256	(0.325)	3.743	(0.302)	4.045	(0.274)	3.459	(0.236)	0.790*	(0.431)	-0.285	(0.391)
Community, Social and Personal Svs	16.781	(0.636)	8.642	(0.378)	13.849	(0.464)	6.283	(0.293)	-2.933***	(0.798)	-2.359***	(0.487)
Occupation												
Managers, Profess. and Armed forces	6.125	(0.450)	3.034	(0.259)	3.633	(0.263)	2.209	(0.165)	-2.492***	(0.526)	-0.824***	(0.311)
Technicians and related	4.972	(0.391)	5.620	(0.448)	3.955	(0.269)	5.650	(0.276)	-1.017**	(0.478)	0.030	(0.536)
Sales clerks	4.590	(0.393)	1.746	(0.178)	3.996	(0.264)	1.587	(0.140)	-0.594	(0.476)	-0.159	(0.228)
Service and sales workers	31.896	(0.773)	7.329	(0.335)	40.179	(0.629)	7.990	(0.321)	8.283***	(1.002)	0.661	(0.476)
Skilled agric. and fish. workers	9.995	(0.423)	28.379	(0.657)	10.320	(0.349)	26.591	(0.537)	0.324	(0.617)	-1.788	(1.100)
Craft and related trades workers	9.188	(0.506)	14.930	(0.494)	8.741	(0.389)	14.948	(0.465)	-0.448	(0.638)	0.018	(0.714)
Machine operators	1.541	(0.219)	10.349	(0.423)	1.380	(0.171)	14.837	(0.446)	-0.161	(0.278)	4.488***	(0.634)
Elem. occupations	31.692	(0.763)	28.614	(0.648)	27.796	(0.593)	26.189	(0.511)	-3.896***	(0.970)	-2.425***	(0.830)

Note: (Ln.) Sample includes individuals between 18 and 65 years. Estimates calculated applying sampling weights and VCE corrected according to survey's complex sample design. SE in parenthesis.
Elaborated by the author based on INEI - National Household Survey (2005-2017)

1.4 Results

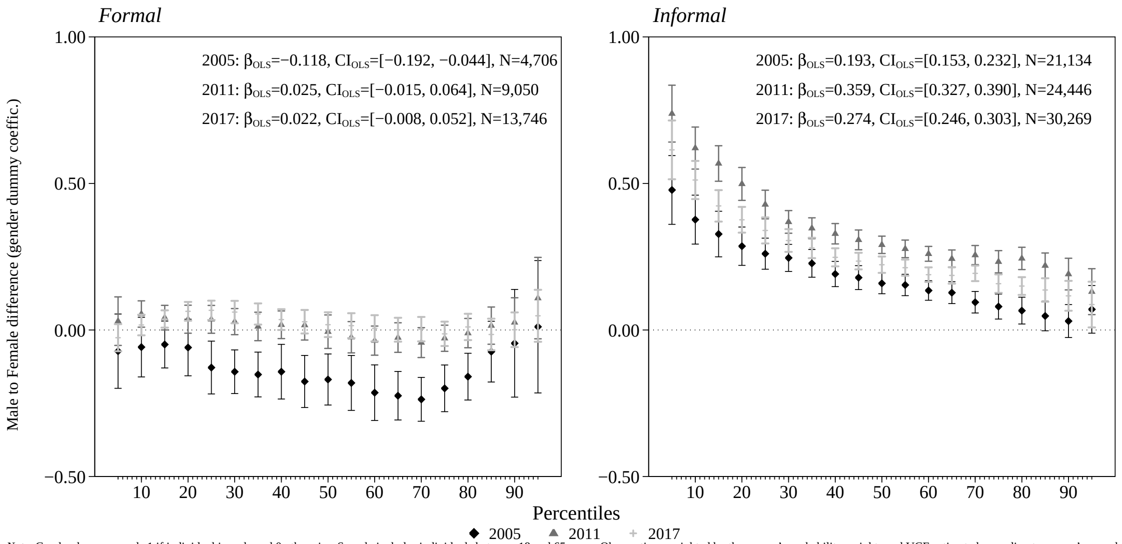
This section applies RIF regression to estimate the magnitude of percentile-specific log hourly-wage gender gaps and to decompose these unconditional gaps in order to identify that part that can be attributed to unequal treatment of women. Finally, we undertake robustness checks to analyze the sensitivity of the results reported here.

1.4.1 Gender Wage Gap Estimates

The observed male to female differences in (log-hourly) wages vary between the formal and informal sector (Figure 1.4 and Table 1.A2 in the Appendix). On the one hand, the main characteristic of the formal labour market in 2005 was that females earned higher wages than their male counterparts, not only on average but also for every percentile of the distribution. Nonetheless, this female advantage appears to have vanished by

2017 since the confidence intervals of the estimates are no longer below the zero line but, instead, around that line. On the other hand, there is preliminary evidence of both ‘sticky floors’ and ‘glass ceilings’ in the informal sector, since across the wages distribution, males earn, statistically, higher wages than females. Importantly, ‘sticky floors’ are deeper than ‘glass ceilings’ since the gaps decrease as we move up across the distribution: for 2017, the male to female difference is 62% at the 5th, 51% at the 10th and 42% at the 15th percentile (bottom end) and amounts to 16% at the 75th and 12% at the 90th percentile (top end). Also, note that there is no evidence of change in these gaps in this sector between the beginning and the end of the period. Gaps in 2011 seem statistically larger than those in 2005.

Figure 1.4 – Raw gaps by sector, 2005, 2011 and 2017

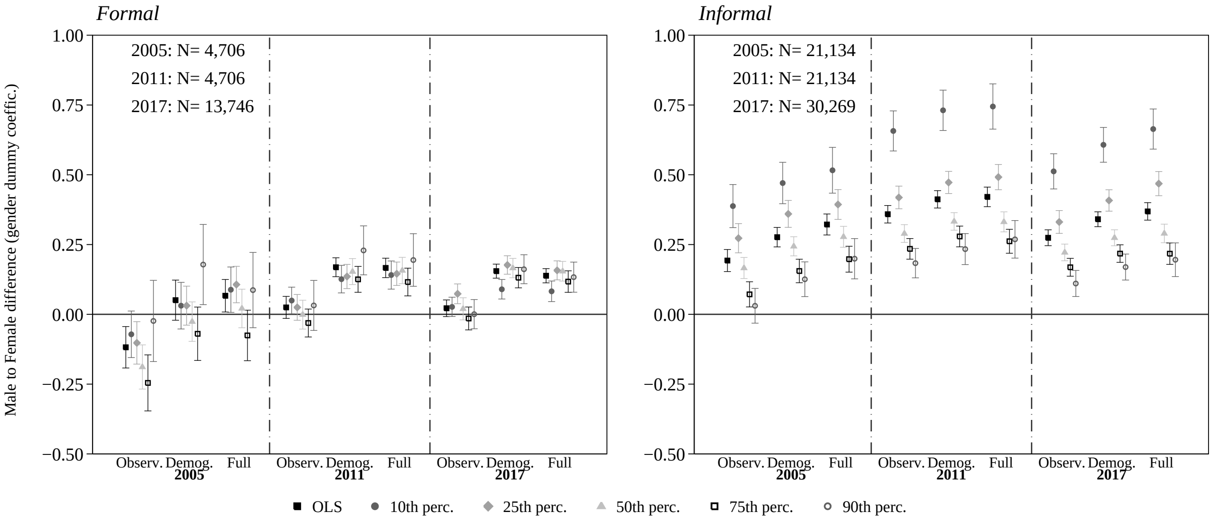


Note: Gender dummy equals 1 if individual is male and 0 otherwise. Sample includes individuals between 18 and 65 years. Observations weighted by the survey’s probability weights and VCE estimated according to survey’s complex sampling design. Vertical lines correspond to the 95% confidence intervals.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

RIF regression allows us to control for individual differences in observed characteristics, which provides more compelling evidence about gender wage gaps. In order to do this, we pool both males and females observations and estimate the gender dummies at the 10th, 25th, 50th, 75th and 90th percentiles of the unconditional wages distribution using RIF regression and at the mean of the unconditional wages distribution using OLS (Figure 1.5). For simplicity and to conserve space, we omit results from the model which only includes demographic variables and the set of industry dummies (nevertheless, they are shown in Table 1.A3 in the Appendix). As before, results differ between sectors but, notably, do not differ across the different models. In the formal sector, gaps in 2011 and 2017 increase once we control for different individual characteristics, and they remain relatively constant, around 15%, across the percentiles of the unconditional distribution. Interestingly, the observed gaps favouring women in 2005 (shown in the previous graph) disappear once we control for observed characteristics.

However, we find statistical evidence of both ‘sticky floors’ and ‘glass ceilings’ in the informal sector. The former is more acute than the latter, given the negative gradient of those cross-percentiles estimates. From the full model for 2017, ‘sticky floors’ manifest themselves as a 66% and 47% gap at the 10th and 25th percentiles, whereas ‘glass ceilings’ manifest themselves as a 21% and 19% gap at the 75th and 90th percentile. Also, the results indicate that gender differences have not contracted relative to 2005. In particular, gaps from the full model in 2005 are respectively 12 and 5 percentage points lower at the 10th and 25th percentile than in 2017. In comparison to 2011, the estimated gaps are 23 and 2 percentage points lower at those same percentiles (although their confidence intervals overlap). The essential message so far is clear: gaps in the informal sector suggest a sticky floor effect and a glass ceiling, being the former more pronounced. Neither has decreased materially between the beginning and the end of the period under analysis.

Figure 1.5 – Modelled gaps by sector, 2005, 2011 and 2017



Note: Sample includes individuals between 18 and 65 years. Gender dummy equals 1 if individual is male and 0 otherwise. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies; the full model adds a set of industry dummies and occupation dummies. Observations weighted by the survey's probability weights and bootstrapped−VCE estimated according to survey's complex sampling design. Vertical lines correspond to the percentile−based bootstrap 95% confidence intervals.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

These results using the pooled dataset imply that labour market returns to observed characteristics between males and females are comparable. Since our ultimate interest lies in estimating the treatment effect component of the gender wage gap, which depends on the between-gender differences in these returns, we now calculate the value of the (adjusted) Wald (cross-model) statistic[13] for the hypothesis that there are no differences between coefficients from the male-only sample and those from the female-only sample. We calculate this statistic for each set of results arising from RIF regressions at 10th, 25th, 50th, 75th and 90th percentiles and from OLS. The set of estimates is shown in the Appendix A for formal sectors (Table 1.A4 - Table 1.A6) and informal (Table 1.A7 - Table 1.A9). For brevity, we omit results for the model including only demographic variables plus the set of industry dummies. The key results suggest that in the formal sector the estimates for the males and females samples differ at every percentile of the unconditional wage distribution for both years (Table 1.2). More specifically, when we consider the demographics only model, we find statistical differences in returns to education between males and females at the percentiles shown. When we consider the model including all the variables, there are significant differences in estimated industry dummies for 2005 and occupation dummies for 2011 and 2017. In this last year, the industry dummy estimates are greater for males at the percentiles below or equal to the median. Occupation dummies are larger for males at percentiles greater than or equal to the median (Table 1.A6). Focusing on the informal sector, the results also show statistical differences (below 5%) between male and female coefficients for both years. In the demographics only model, the vector of schooling and the urban dummy differ consistently between males and females. When considering the full model, the males' and females' dummies for urban areas and occupations are statistically different for the three years shown across the different percentiles. In particular, urban area coefficients are greater for females across the 2005 distribution. In addition, occupation dummies estimates are larger for males at percentiles greater than or equal to the median in both years (Table 1.A8 and Table 1.A9).

[13]The (adjusted) Wald test statistic amounts to an F test (with degrees of freedom equal to $(k, \mathrm{dff} - k + 1)$, with dff the design degrees of freedom) in the case of survey data (Aneshensel 2013).

Table 1.2 – (Adjusted) Wald tests for coefficient differences between males and females by sector, 2005, 2011 and 2017

	Formal						Informal					
	OLS	P. 10	P. 25	P. 50	P. 75	P. 90	OLS	P. 10	P. 25	P. 50	P. 75	P. 90
2005												
Demograph. only												
Regression	2.067*	1.006	6.124***	1.992*	6.963***	5.172***	2.634**	7.793***	2.814***	1.627	1.070	1.439
Shooling (years) (v)	5.185***	0.523	15.487***	4.131**	14.222***	6.008***	0.111	0.225	3.730**	3.751**	1.556	1.150
Age (years) (v)	0.643	0.078	0.286	0.763	1.842	1.054	0.343	3.322**	2.808*	0.604	0.287	2.371*
If indigenous (d)	0.172	5.214**	1.803	0.002	3.899**	9.211***	0.057	0.411	0.200	0.209	0.669	0.466
If hh in urban area (d)	0.284	0.009	0.365	0.703	1.917	1.493	10.377***	31.797***	2.590	0.002	0.792	0.170
Full model												
Regression	2.099***	2.773***	5.349***	3.950***	3.372***	3.928***	7.447***	8.278***	8.649***	5.708***	2.663***	4.185***
Shooling (years) (v)	3.264**	0.267	0.368	0.757	8.083***	3.150**	1.706	2.312*	1.011	1.898	2.617*	0.034
Age (years) (v)	0.321	0.287	0.326	0.262	0.210	0.022	0.870	5.572***	4.886***	0.751	0.949	2.296
If indigenous (d)	0.110	5.262**	1.833	0.001	3.184*	7.069***	0.031	0.000	1.000	0.187	0.246	0.578
If hh in urban area (d)	0.197	0.147	0.206	1.521	1.121	0.164	34.381***	64.188***	29.111***	5.776**	0.003	0.016
Industry (v)	2.356**	4.744***	1.958*	1.933*	1.492	5.619***	1.676	2.979***	0.898	2.263**	2.295**	4.514***
Occupation (v)	0.980	0.887	3.736***	3.249***	0.822	1.614	7.980***	8.301***	5.640***	3.233***	3.588***	3.988***
2011												
Demograph. only												
Regression	1.001	2.544**	5.809***	10.399***	2.429**	3.090***	3.514***	5.524***	11.752***	3.539***	0.318	2.490**
Shooling (years) (v)	1.298	4.087**	13.386***	18.838***	5.184***	4.081**	0.664	2.795*	7.396***	3.182**	0.553	5.231***
Age (years) (v)	1.090	0.119	3.550**	7.716***	0.828	0.029	0.053	0.142	0.027	0.238	0.399	0.910
If indigenous (d)	0.063	6.224**	0.000	0.003	2.141	0.266	0.004	0.161	0.053	2.153	0.452	0.296
If hh in urban area (d)	1.310	1.906	0.072	0.017	0.824	4.838**	12.740***	20.099***	21.542***	2.275	0.071	0.472
Full model												
Regression	2.488***	2.321***	3.551***	7.827***	2.954***	2.044***	9.683***	9.906***	17.497***	8.158***	2.323***	1.666**
Shooling (years) (v)	0.010	1.307	1.192	3.195**	2.314*	1.087	1.542	0.516	2.757*	1.096	0.134	1.668
Age (years) (v)	1.066	0.514	2.175	5.813***	1.048	0.183	0.115	0.394	0.282	0.198	1.315	1.377
If indigenous (d)	0.856	9.110***	0.788	0.245	2.085	0.091	0.029	0.828	0.323	1.276	1.518	0.113
If hh in urban area (d)	3.074*	4.860**	0.486	0.626	1.266	2.710*	45.015***	56.029***	70.613***	15.820***	0.054	1.577
Industry (v)	1.526	1.989*	1.171	1.933*	0.593	1.765*	0.314	0.201	0.662	2.149**	0.799	1.201
Occupation (v)	3.456***	1.876*	3.504***	7.107***	3.936***	1.568	12.462***	11.264***	12.476***	6.343***	4.523***	1.606
2017												
Demograph. only												
Regression	2.088*	1.405	6.309***	17.562***	1.708	4.358***	10.941***	7.868***	13.380***	5.505***	2.268**	2.317**
Shooling (years) (v)	5.537***	2.250	18.125***	39.691***	1.357	11.343***	3.168**	2.577*	1.881	1.589	0.410	3.613**
Age (years) (v)	0.344	0.879	2.473*	0.877	1.021	0.414	1.721	1.667	2.340*	1.304	4.426**	4.082**
If indigenous (d)	1.322	1.019	0.282	1.762	7.666***	1.527	3.964**	5.550**	3.640*	0.010	0.095	0.027
If hh in urban area (d)	0.498	2.350	2.380	0.280	0.004	0.049	44.332***	32.802***	45.131***	11.723***	2.506	0.282
Full model												
Regression	2.225***	2.724***	2.992***	6.703***	2.687***	2.173***	12.881***	14.376***	18.675***	7.166***	7.351***	3.162***
Shooling (years) (v)	3.379**	1.386	3.418**	11.187***	0.708	2.794*	10.136***	6.892***	3.922**	3.357**	1.980	2.575*
Age (years) (v)	0.136	0.340	1.294	0.421	1.205	0.354	1.903	3.712**	7.761***	3.045**	4.254**	3.380**
If indigenous (d)	0.246	2.857*	1.942	1.046	6.495**	1.333	2.558	2.266	2.637	0.068	0.276	0.004
If hh in urban area (d)	0.020	0.030	0.062	0.459	0.016	0.147	132.881***	127.437***	122.771***	45.274***	14.666***	0.149
Industry (v)	1.290	2.747***	2.325**	1.219	0.887	0.913	2.273**	5.431***	3.505***	1.050	10.722***	3.349***
Occupation (v)	4.244***	2.602**	4.174***	6.087***	3.438***	1.886*	6.522***	8.400***	5.887***	1.833*	2.161**	4.349***

it:Note: F statistics of joint-hypotheses tests shown. Sample includes individuals between 18 and 65 years. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. Observations weighted by the survey's probability weights and VCE estimated according to survey's complex sampling design. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level

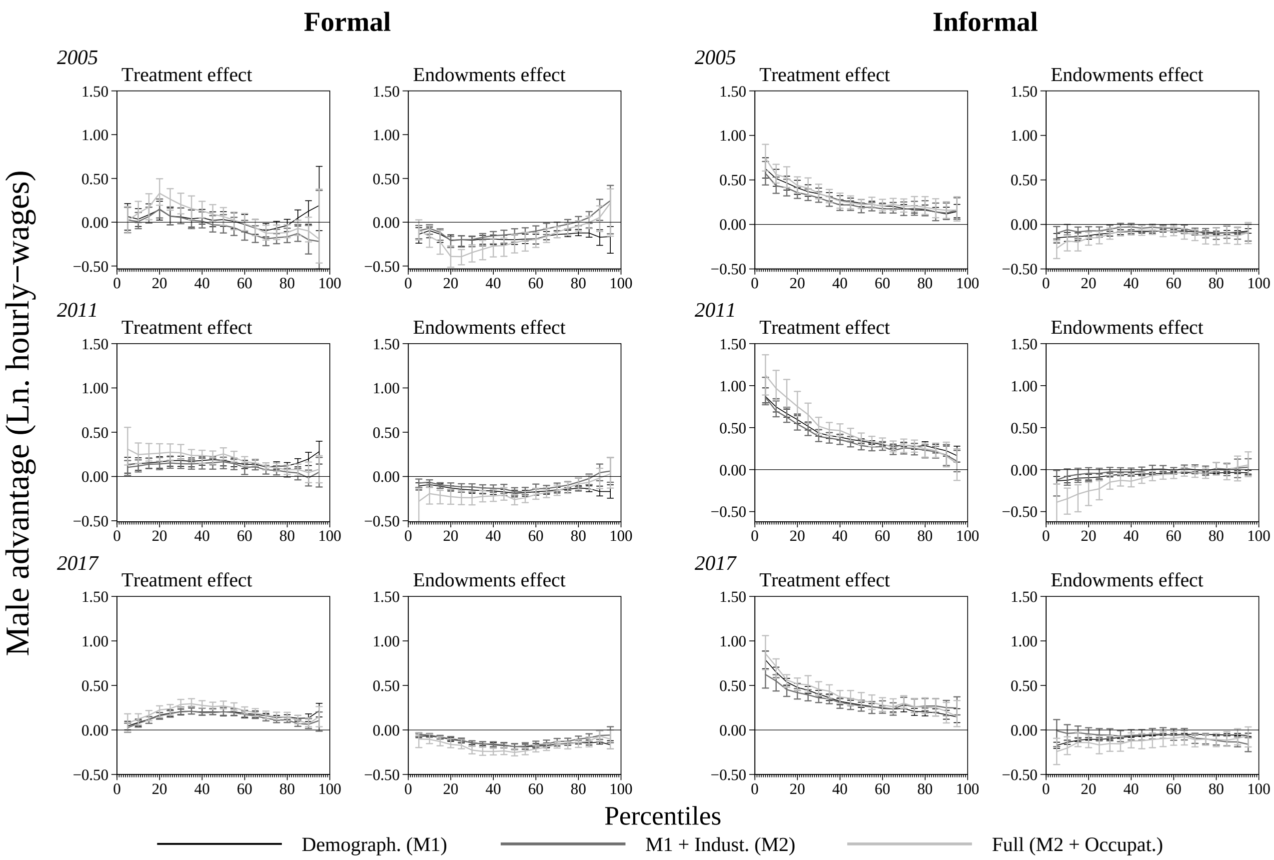
Elaborated by the author based on INEI - National Household Survey (2005-2017)

Thus, the results presented provide a detailed portrayal of the magnitudes of the unconditional wage gaps in Peru. However, they do not provide us with a critical insight for policy: Are the leading cause behind the observed (raw) differences attributed to unequal treatment against women or an endowment effect favouring males? To address this issue, in the following subsection, we will use RIF regressions to decompose the observed unconditional gaps at every percentile of the wage distribution, based on Equation 1.3.

1.4.2 Decomposition of the gender wage gaps

Results for the unconditional decompositions of differences between males and females in log-hourly wages are invariant across the three different models estimated, provided their confidence intervals overlap all over the distribution (see Figure 1.6 and Table 1.A10 in the Appendix). Consequently, results that follow for formal and informal sectors are robust to alternative specifications. As anticipated, regularities differ between these sectors. Focusing on the formal sector, the endowment effects provide evidence that in both 2011 and 2017, women had better level endowments of characteristics than the male workers. Nevertheless, since the treatment effect mirrors the latter but with opposite signs (i.e., the differences in labour market returns favour males relative to females), the gap is statistically insignificant.

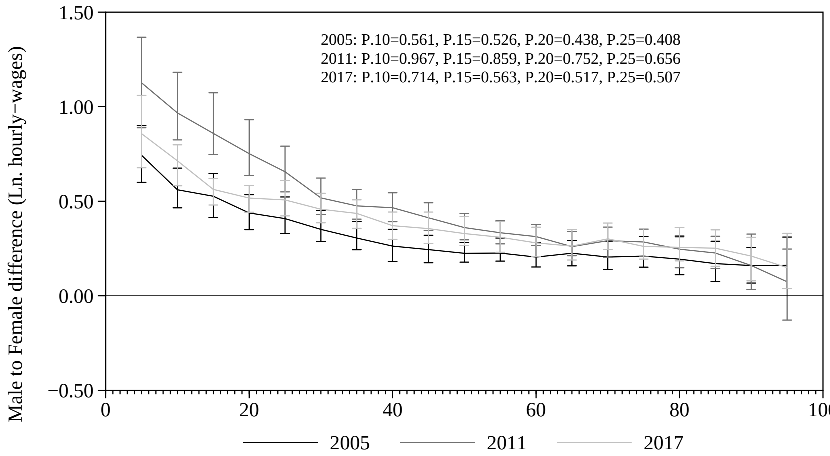
Figure 1.6 – RIF decompositions by sector, 2005, 2011 and 2017



Note: Sample includes individuals between 18 and 65 years. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. Observations weighted by the survey's probability weights and VCE estimated according to survey's complex sampling design. Vertical lines correspond to the percentile−based bootstrap 95% confidence intervals.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

Three key results for policy-making are found in the informal sector. In the first place, in every single year, the treatment effect is positive at every point of the log hourly-wages distribution. In contrast, the endowments effect is close to zero, implying that the gender wage gap can be explained mainly by the treatment effect alone. Put differently, the disadvantage women face in the informal sector is due to an unequal treatment exercised against them. In the second place, 'sticky floors' and 'glass ceilings' exist, but the estimated effects are more sizeable at the bottom of the distribution. Based on the full model, this effect amounts to a gender difference of 56% in 2005, 97% in 2011 and 71% in 2017 at the 10th percentile. At the 90th percentile, it amounts to 16% in 2005 and 2011 and 21% for 2017. In other words, women in the lower parts of the distribution earn the lowest wages rewards and thus are those who suffer the greatest effect of labour market discrimination. Finally, the unequal treatment that women face in the informal sector, mainly at the bottom end of the distribution, has not changed between the beginning and the end of the 'Peruvian Growth Miracle' since confidence intervals of these estimates overlap for 2005 and 2017 (Figure 1.7).

Figure 1.7 – Treatment effect in the informal sector, 2005, 2011 and 2017



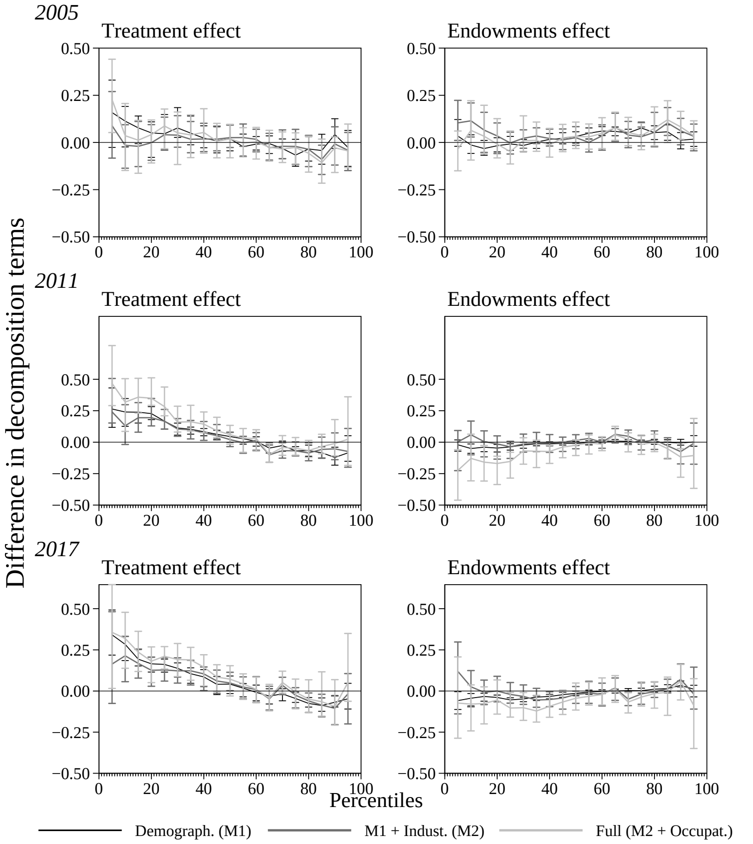
Note: Sample includes individuals between 18 and 65 years. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. Observations weighted by the survey's probability weights and VCE estimated according to survey's complex sampling design. Vertical lines correspond to the percentile−based bootstrap 95% confidence intervals.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

1.4.3 Robustness Checks

These main decomposition results hinge on different assumptions related to the definition of employment in the formal and informal sector and the use of sampling weights. In order to test the sensitivity of our results to these assumptions, we undertake a series of robustness checks. These measure the difference between the estimates of the decomposition terms shown above and those found under alternative definitions of employment and sample design. First, we analyze the differences in the decomposition terms when we consider only those informal workers who are employers and white and blue-collar workers, both private and public. Therefore, we see what happens to the results when we exclude from the sample independent workers. Results (Figure 1.8 and Table 1.A11 in the Appendix) across the distribution suggest that decomposition terms do not vary in 2005 under this alternative definition. However, there are statistical differences in the years 2011 and 2017. In the case of the endowment effect, results suggest that the alternative definition provides estimates that are 10 and 15 percentage points higher in 2005 and 2017, respectively. The difference in the treatment term reveals a more definitive pattern. It is positive at the bottom end of the distribution and turns negative as we move to higher percentiles. In both years, estimates under the alternative informal definition are 30 and 25 percentage points lower at the 10th and 25th percentile, respectively. In contrast, our main estimates are more conservative at the top end, about 10 percentage points lower at the 90th percentile.

We now analyze the differences in the decomposition terms when we leave out from the analysis the public sector workers either in the formal or informal sectors (Figure 1.9 and Table 1.A12 in the Appendix). The number of such workers in the informal sector is relatively small as the majority are concentrated in the formal sector. After excluding these workers, the treatment effects differ mainly from the median of the distribution until approximately the 80th percentile. In turn, the difference in endowment effects emerges in the bottom half of the distribution. In fact, the largest differences were found in 2005. Nevertheless,

Figure 1.8 – Difference in RIF decompositions under alternate informal employment definition, 2005, 2011 and 2017

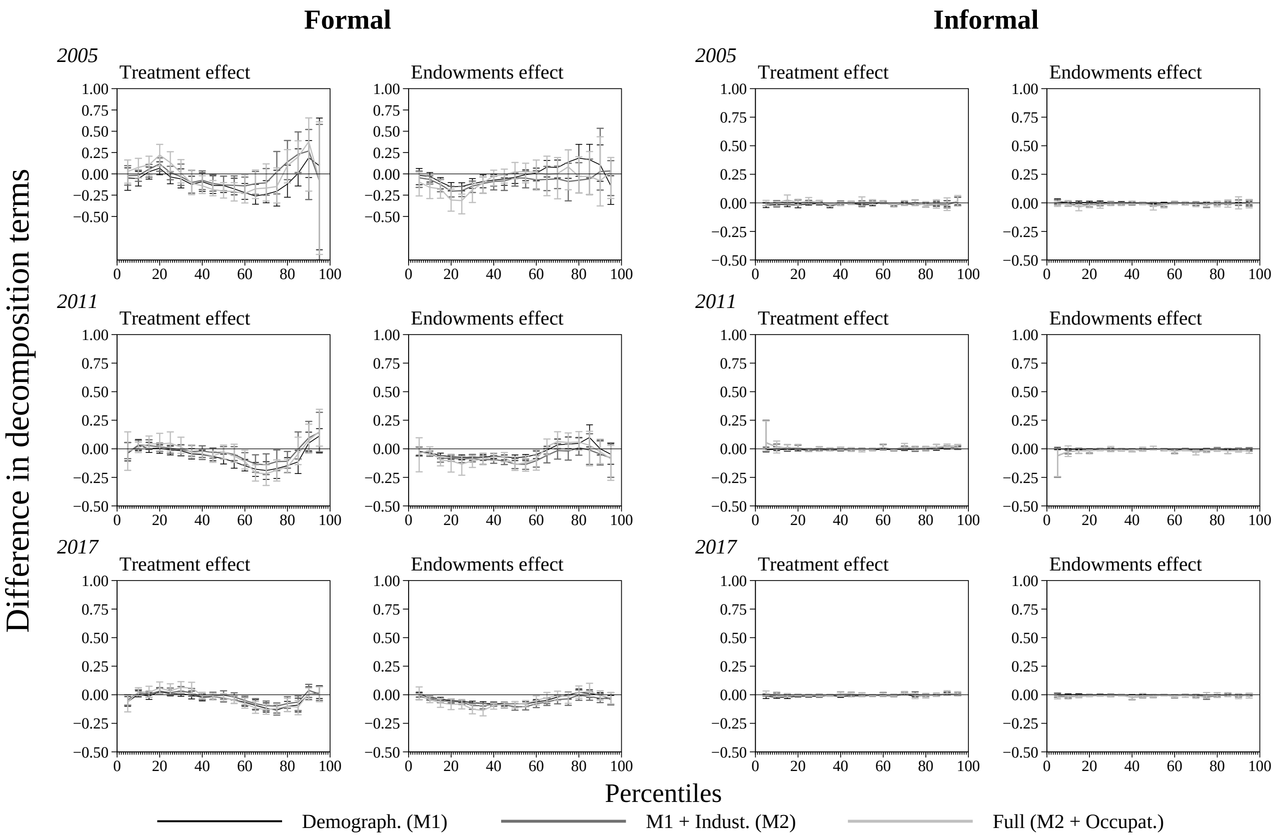


Note: Difference calculated as SRS minus SVY estimates. Difference calculated as the original SVY minus alternate SVY estimates. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) well as ethnicity and urban area dummies. Observations weighted by the survey's probability weights and bootstrape VCE estimated according to the survey's complex sampling design. Vertical lines corresponds to the percentile−base bootstrap 95% confidence intervals.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

the differences are negative, suggesting that our main results are more conservative than those under this alternative definition. This is unsurprising given that public sector workers tend to be relatively well paid in Peru. Finally, we compare our results with those arising after estimating the decompositions without individual sampling weights. In other words, we treat the data as originating from simple random sampling (Figure 1.10 and Table 1.A13 in the Appendix). The results suggest that differences are small in magnitude and not statistically significant.

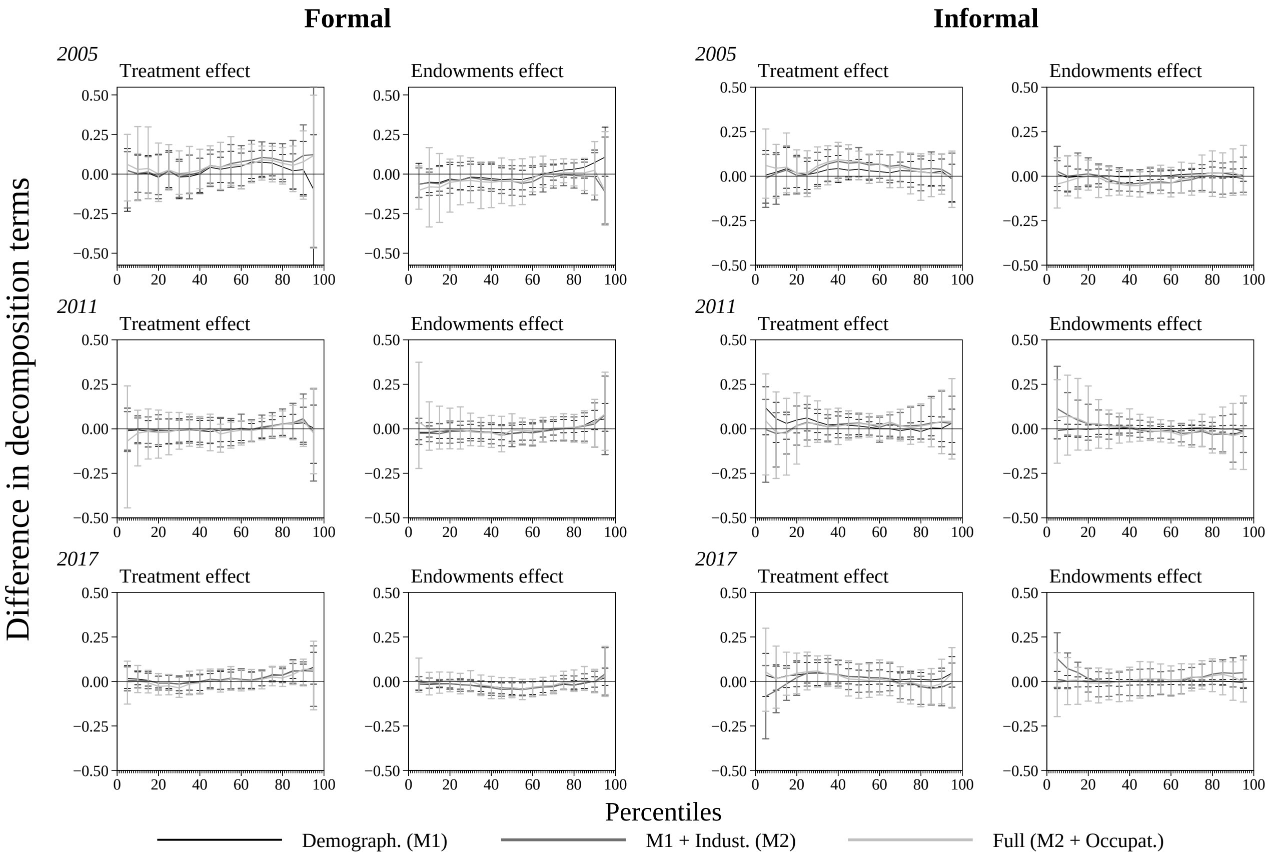
Consequently, we can assert that results presented here are robust across models estimated and, importantly, when comparing to the results that cater for the sample design. Overall, the core estimates reported in the main text are more conservative than those that consider only the private workers in both formal and informal sectors. The sensitivity when excluding independent workers from the informal sample needs to be further assessed. Notwithstanding, it is crucial to consider that our main key policy results still hold even if we assume this latter alternative definition. Specifically, in the informal sector, sticky floor and 'glass ceilings' are present and persisted across the years, with the former phenomenon considerable more pronounced than the latter.

Figure 1.9 – Difference in RIF decompositions under alternate employment definitions, 2005, 2011 and 2017



Note: Difference calculated as SRS minus SVY estimates. Difference calculated as the original SVY minus alternate SVY estimates. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. Observations weighted by the survey's probability weights and bootstraped−VCE estimated according to the survey's complex sampling design. Vertical lines corresponds to the percentile−based bootstrap 95% confidence intervals.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

Figure 1.10 – Difference in RIF decompositions under SRS estimation by sector, 2005, 2011 and 2017



Note: Sample includes individuals between 18 and 65 years. Difference calculated as SRS minus SVY estimates. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. Vertical lines corresponds to the percentile−based bootstrap 95% confidence intervals.
Elaborated by the author based on INEI − National Household Survey (2005−2017)

1.5 Conclusions and policy recommendations

This study used RIF regressions (Firpo et al. 2007, 2009) as the primary tool to extend the Oaxaca-Blinder gender-wage gap decomposition (Oaxaca 1973 and Blinder 1973) to every percentile of the unconditional (log) hourly-wage distribution. For the 2005-2017 period, characterized by the high growth of hourly wages in both the formal and informal sector, this extension allowed us to decompose the unconditional difference between men and women into a part attributable to unequal treatment against women (typically framed as discrimination against women in the labour economics literature) and that part attributable to differences in levels of characteristics (endowments). Hence, this study advances the existing literature for the Latin American region and mainly Peru since previous analyzes have only focused on the mean of the wages distribution and have not considered intrinsic differences in formal and informal labour markets. In addition, it focuses on the pay gap over a period of sharp and rapid economic growth in Peru.

Results suggest that, across different models, gaps are statistically significant across the distribution of wages in the formal and informal labour markets, revealing the existence of 'sticky floors' and 'glass ceilings' in both sectors. These two phenomena have persisted over time. In addition, the magnitudes of these gaps are more striking in the informal sector and are remarkably high at the bottom of the distribution (ranging from 40% to 50% in 2005) and have increased in 2017 (around 50% to 65%), whereas gaps at the top end remain fairly constant (around 20%). This indicates that informal working women at the bottom of the wages distribution face a more significant disadvantage when compared to men. Pay gaps oscillate around 13% in 2011 and 2017 across the wages distribution in the formal sector. The results based on the decomposition of these gaps differ between both sectors. In the formal sector, the treatment effect does not exhibit a clear pattern in 2005, but by 2017 the endowment effects favour females and the treatment effects favour males. A more remarkable result is found in the informal sector, where the gaps at every percentile can be explained almost exclusively by the unequal treatment effect against women. In particular, there is evidence that the magnitude of the sticky floor has persisted over time in this sector.

The decomposition estimates appear robust across different models at every percentile. Importantly, the results presented here are not materially different depending on how the sample design is treated. However, they are more conservative than those under an alternative definition of employment, excluding public sector workers from the sample. Nevertheless, it is important to note that excluding self-employed informal workers does have statistical effects on the results presented here. These differences are greater at the bottom of the distribution. However, even under this robustness check, the main results are that the treatment effect exhibits deeper 'sticky floors' than 'glass ceilings', which is a finding that remains unchanged.

Consequently, from a policy-maker's perspective, the most important take-away is that informal working women at the bottom end of the distribution are those that remain the most vulnerable given their low hourly wages and the fact that they suffer a higher burden of discrimination. The 'Peruvian Growth Miracle' is not one that reduces the gender pay gaps over this period. This result casts doubt on the overall effectiveness of policies put in place by the Peruvian government to alleviate and mitigate this problem. A tentative solution would require a more coordinated effort in terms of increasing the transition of workers from the informal to the formal sector since an essential part in reducing wage inequality in Latin America in the last decades can be attributed to this shift (Messina and Silva 2018).

The empirical analysis reported here is largely descriptive in nature. Nevertheless, this does not diminish the contribution of the empirical exercise to the debate on gender inequality in Peru. In particular, the vulnerable nature of women at the bottom end of the pay distribution in the informal sector is a consistent theme that emerges and one that should animate policy concern. Nevertheless, two key elements need to be incorporated into future research on the gender wage gap in Peru. The first and most important is to explicitly model selection into employment and formal/informal choice and establish if the results stated here are robust to selection correction. The literature on the appropriate way to correct such selection bias within the RIF framework is not a settled issue. No strong consensus has yet emerged on the appropriate empirical strategies to adopt. However, we believe the magnitude and persistence of the gender treatment effects reported at the bottom end of the pay distribution are likely to be invariant to the selection correction procedures, though this awaits further research. The second element is to further analyze to what extent the findings presented here remain invariant to the use of alternative definitions of informal workers. The robustness checks undertaken here have provided some insights on this issue. However, this key definitional concept is not uniquely defined as it varies across both context and countries. Although the use of the non-pensionable nature of the job as the basis for classification of informal and formal sectors offers a meaningful characterization of informality in Peru, other definitions must be considered as a means of investigating the robustness of this definition to alternative definitions and classifications. This remains a central part of an agenda for future research here.

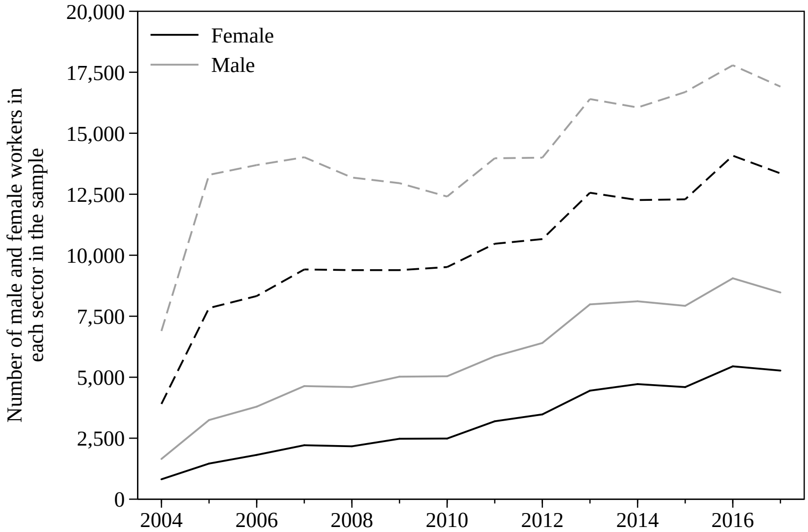
Appendix

Table 1.A1 – Descriptive statistics of hourly-wages for males and females by sector, 2005, 2011 and 2017

	2005				**2011**				**2017**			
	Females		Males		Females		Males		Females		Males	
Formal												
Mean	7.283	(0.460)	7.132	(0.501)	8.690	(0.227)	9.287	(0.259)	11.774	(0.393)	11.890	(0.224)
P10	2.206	(0.098)	2.077	(0.054)	3.065	(0.060)	3.237	(0.046)	4.310	(0.053)	4.385	(0.060)
P25	3.293	(0.116)	2.906	(0.069)	4.219	(0.098)	4.375	(0.069)	5.536	(0.086)	5.919	(0.069)
P50	5.370	(0.183)	4.540	(0.116)	6.621	(0.165)	6.591	(0.105)	8.593	(0.170)	8.735	(0.125)
P75	8.730	(0.327)	7.187	(0.257)	10.509	(0.194)	10.206	(0.164)	13.812	(0.245)	13.635	(0.209)
P90	12.527	(1.955)	11.984	(1.218)	15.655	(0.681)	16.078	(0.470)	20.708	(0.505)	20.770	(0.499)
Gini	0.412	(0.021)	0.477	(0.025)	0.384	(0.010)	0.415	(0.012)	0.396	(0.017)	0.388	(0.008)
Informal												
Mean	2.416	(0.080)	2.601	(0.068)	3.684	(0.098)	4.715	(0.087)	5.035	(0.152)	5.813	(0.075)
P10	0.381	(0.013)	0.557	(0.014)	0.604	(0.015)	1.121	(0.021)	0.898	(0.024)	1.499	(0.025)
P25	0.828	(0.016)	1.075	(0.020)	1.317	(0.030)	2.017	(0.024)	1.858	(0.033)	2.615	(0.042)
P50	1.595	(0.026)	1.866	(0.024)	2.461	(0.032)	3.287	(0.035)	3.456	(0.047)	4.314	(0.042)
P75	2.687	(0.059)	2.903	(0.049)	4.065	(0.072)	5.130	(0.064)	5.659	(0.081)	6.623	(0.069)
P90	4.660	(0.139)	4.807	(0.086)	7.009	(0.151)	8.470	(0.190)	9.289	(0.203)	10.429	(0.198)
Gini	0.517	(0.012)	0.464	(0.012)	0.508	(0.011)	0.464	(0.008)	0.495	(0.014)	0.435	(0.006)

Note: (Ln.) Sample includes individuals between 18 and 65 years. Estimates calculated applying sampling weights and VCE corrected according to survey's complex sample design. SE in parenthesis.
Elaborated by the author based on INEI - National Household Survey (2005-2017)

Figure 1.A1 – Observations count by sex and sector, 2004-2017



Note: Dashed and solid lines refer to informal and formal sectors, respectively. Sample includes working individuals between 18 and 65 years.
Elaborated by the author based on INEI − National Household Survey (2004−2017)

Table 1.A2 – Raw wage gaps by sector, 2005, 2011 and 2017

	OLS		P. 10		P. 25		P. 50		P. 75		P. 90	
	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE
Formal												
2005	-0.118***	(0.038)	-0.058	(0.052)	-0.128***	(0.046)	-0.169***	(0.044)	-0.199***	(0.041)	-0.045	(0.094)
2011	0.025	(0.020)	0.054**	(0.023)	0.036	(0.024)	-0.006	(0.029)	-0.028	(0.023)	0.026	(0.043)
2017	0.022	(0.015)	0.017	(0.018)	0.067***	(0.017)	0.018	(0.022)	-0.013	(0.021)	0.001	(0.030)
Informal												
2005	0.193***	(0.020)	0.376***	(0.043)	0.260***	(0.027)	0.159***	(0.018)	0.080***	(0.022)	0.030	(0.029)
2011	0.359***	(0.016)	0.621***	(0.037)	0.428***	(0.025)	0.291***	(0.015)	0.233***	(0.019)	0.191***	(0.028)
2017	0.274***	(0.015)	0.512***	(0.033)	0.340***	(0.023)	0.223***	(0.014)	0.158***	(0.016)	0.117***	(0.026)

Note: Sample includes working individuals between 18 and 65 years. Gender dummy equals 1 if individual is male and 0 otherwise. Observations weighted by the survey's probability weights and VCE corrected according to survey's complex sample design. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level

Elaborated by the author based on INEI - National Household Survey (2005-2017)

Table 1.A3 – Modelled wage gaps by sector, 2005, 2011 and 2017

	OLS		P. 10		P. 25		P. 50		P. 75		P. 90	
	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE
Formal												
2005 (Obs=4,706)												
Observed	-0.118***	(0.038)	-0.072*	(0.043)	-0.102***	(0.039)	-0.189***	(0.040)	-0.246***	(0.051)	-0.024	(0.074)
Model 1 (Demograph. only)	0.051	(0.037)	0.031	(0.043)	0.031	(0.036)	-0.026	(0.036)	-0.070	(0.049)	0.178**	(0.073)
Model 2 (M1 + Industry)	0.024	(0.033)	0.054	(0.043)	0.058	(0.036)	-0.018	(0.036)	-0.110**	(0.048)	0.036	(0.070)
Model 3 (M2 + Occupation)	0.067**	(0.030)	0.088**	(0.041)	0.107***	(0.033)	0.021	(0.035)	-0.076	(0.046)	0.087	(0.069)
2011 (Obs=9,050)												
Observed	0.025	(0.020)	0.050**	(0.024)	0.025	(0.024)	-0.001	(0.026)	-0.031	(0.026)	0.032	(0.046)
Model 1 (Demograph. only)	0.169***	(0.017)	0.126***	(0.025)	0.136***	(0.022)	0.153***	(0.024)	0.125***	(0.024)	0.229***	(0.045)
Model 2 (M1 + Industry)	0.139***	(0.018)	0.115***	(0.026)	0.117***	(0.022)	0.134***	(0.024)	0.091***	(0.024)	0.150***	(0.047)
Model 3 (M2 + Occupation)	0.166***	(0.018)	0.141***	(0.026)	0.145***	(0.021)	0.157***	(0.024)	0.116***	(0.025)	0.195***	(0.048)
2017 (Obs=13,746)												
Observed	0.022	(0.015)	0.027	(0.018)	0.074***	(0.018)	0.020	(0.020)	-0.015	(0.021)	0.001	(0.027)
Model 1 (Demograph. only)	0.155***	(0.013)	0.090***	(0.018)	0.177***	(0.017)	0.166***	(0.018)	0.131***	(0.019)	0.162***	(0.026)
Model 2 (M1 + Industry)	0.134***	(0.013)	0.083***	(0.019)	0.162***	(0.018)	0.153***	(0.018)	0.110***	(0.019)	0.124***	(0.027)
Model 3 (M2 + Occupation)	0.138***	(0.013)	0.083***	(0.019)	0.158***	(0.017)	0.154***	(0.018)	0.117***	(0.020)	0.133***	(0.028)
Informal												
2005 (Obs=21,134)												
Observed	0.193***	(0.020)	0.388***	(0.039)	0.273***	(0.027)	0.166***	(0.019)	0.072***	(0.023)	0.031	(0.032)
Model 1 (Demograph. only)	0.276***	(0.018)	0.470***	(0.038)	0.360***	(0.025)	0.244***	(0.018)	0.155***	(0.022)	0.126***	(0.032)
Model 2 (M1 + Industry)	0.317***	(0.019)	0.501***	(0.041)	0.386***	(0.027)	0.274***	(0.019)	0.195***	(0.024)	0.195***	(0.037)
Model 3 (M2 + Occupation)	0.322***	(0.019)	0.516***	(0.042)	0.394***	(0.027)	0.278***	(0.019)	0.198***	(0.024)	0.199***	(0.037)
2011 (Obs=24,446)												
Observed	0.359***	(0.016)	0.657***	(0.037)	0.419***	(0.021)	0.289***	(0.016)	0.235***	(0.019)	0.183***	(0.027)
Model 1 (Demograph. only)	0.412***	(0.016)	0.731***	(0.037)	0.473***	(0.020)	0.333***	(0.016)	0.279***	(0.019)	0.234***	(0.028)
Model 2 (M1 + Industry)	0.434***	(0.018)	0.728***	(0.040)	0.485***	(0.023)	0.341***	(0.018)	0.290***	(0.022)	0.312***	(0.034)
Model 3 (M2 + Occupation)	0.421***	(0.018)	0.745***	(0.041)	0.492***	(0.023)	0.331***	(0.018)	0.262***	(0.022)	0.269***	(0.034)
2017 (Obs=30,269)												
Observed	0.274***	(0.015)	0.512***	(0.032)	0.331***	(0.021)	0.222***	(0.015)	0.169***	(0.016)	0.111***	(0.024)
Model 1 (Demograph. only)	0.341***	(0.014)	0.607***	(0.032)	0.408***	(0.020)	0.274***	(0.015)	0.218***	(0.016)	0.169***	(0.024)
Model 2 (M1 + Industry)	0.373***	(0.016)	0.638***	(0.036)	0.447***	(0.022)	0.286***	(0.017)	0.238***	(0.020)	0.237***	(0.031)
Model 3 (M2 + Occupation)	0.369***	(0.016)	0.664***	(0.037)	0.468***	(0.022)	0.290***	(0.017)	0.217***	(0.020)	0.196***	(0.031)

it:Note: Sample includes individuals between 18 and 65 years. Gender dummy equals 1 if individual is male and 0 otherwise. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies; the full model includes adds a set of industry dummies and occupation dummies. Observations weighted by the survey's probability weights and bootstrapped-VCE estimated according to survey's complex sampling design. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Elaborated by the author based on INEI - National Household Survey (2005-2017)

Table 1.A4 – Regression models by gender (demographics only model) in the formal sector, 2005 and 2017

	Females						Males					
	OLS	**RIF**					**OLS**	**RIF**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
2005												
Schooling (years)	-0.103***	0.000	-0.002	-0.058***	-0.138***	-0.261***	0.018	0.055	0.143**	-0.018	-0.070***	-0.080
	(0.026)	(0.034)	(0.022)	(0.015)	(0.025)	(0.070)	(0.038)	(0.078)	(0.069)	(0.039)	(0.022)	(0.063)
Schooling (years)^2	0.009***	0.002*	0.003***	0.007***	0.011***	0.018***	0.004**	0.001	-0.000	0.006***	0.006***	0.007**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)	(0.003)
Age (years)	0.021	0.030**	0.047***	0.060***	0.005	-0.036	0.028*	0.042	0.042	0.046**	0.006	-0.035
	(0.016)	(0.013)	(0.016)	(0.011)	(0.015)	(0.040)	(0.016)	(0.034)	(0.032)	(0.021)	(0.019)	(0.054)
Age (years)^2	-0.000	-0.000*	-0.000**	-0.000***	0.000	0.001	-0.000	-0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
If indigenous (d)	-0.070	-0.145*	-0.035	-0.050	-0.008	-0.033	-0.101*	0.137	0.155	-0.056	-0.203***	-0.406***
	(0.049)	(0.080)	(0.057)	(0.050)	(0.067)	(0.109)	(0.060)	(0.085)	(0.101)	(0.127)	(0.074)	(0.072)
If hh in urban area (d)	0.053	0.060	0.046	-0.055	-0.063	0.222*	0.096**	0.074	0.118**	-0.148*	0.065	0.386***
	(0.060)	(0.076)	(0.066)	(0.060)	(0.067)	(0.124)	(0.047)	(0.121)	(0.056)	(0.077)	(0.068)	(0.075)
Constant	0.587*	-0.461	-0.620*	-0.402	1.435***	2.962***	-0.148	-1.154	-1.975**	-0.363	1.542***	2.462**
	(0.336)	(0.332)	(0.352)	(0.256)	(0.334)	(0.772)	(0.407)	(0.868)	(0.856)	(0.451)	(0.356)	(1.124)
Observations	3245	3245	3245	3245	3245	3245	1461	1461	1461	1461	1461	1461
R^2	0.276	0.049	0.125	0.239	0.206	0.113	0.212	0.061	0.193	0.187	0.110	0.035
Model test	321.789	137.005	194.636	966.730	380.360	64.037	236.795	34.116	161.561	276.861	214.121	56.067
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Educ. test	158.050	107.037	139.193	584.111	295.695	35.189	222.073	27.509	92.033	187.872	105.996	19.738
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
2017												
Schooling (years)	-0.064***	0.044**	0.023	-0.042***	-0.127***	-0.240***	-0.050***	0.060***	0.021	-0.077***	-0.143***	-0.156***
	(0.015)	(0.019)	(0.021)	(0.015)	(0.013)	(0.024)	(0.012)	(0.023)	(0.021)	(0.015)	(0.015)	(0.019)
Schooling (years)^2	0.007***	0.000	0.002**	0.006***	0.010***	0.015***	0.007***	0.000	0.003***	0.009***	0.011***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age (years)	0.035***	0.032***	0.045***	0.043***	0.032***	0.023**	0.029***	0.018**	0.022***	0.044***	0.035***	0.031***
	(0.005)	(0.006)	(0.007)	(0.006)	(0.007)	(0.009)	(0.004)	(0.007)	(0.005)	(0.006)	(0.007)	(0.008)
Age (years)^2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000***	-0.000**	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
If indigenous (d)	-0.091***	-0.001	-0.061	-0.051	-0.007	-0.191***	-0.155***	0.067	-0.025	-0.178**	-0.236***	-0.279***
	(0.027)	(0.028)	(0.041)	(0.042)	(0.045)	(0.046)	(0.044)	(0.058)	(0.061)	(0.081)	(0.058)	(0.048)
If hh in urban area (d)	0.131***	0.144***	0.132***	0.113***	0.130***	0.156***	0.103**	0.049	0.049	0.083*	0.127**	0.169***
	(0.019)	(0.042)	(0.035)	(0.031)	(0.028)	(0.027)	(0.043)	(0.046)	(0.047)	(0.048)	(0.050)	(0.038)
Constant	0.941***	0.061	0.015	0.540***	1.629***	2.735***	0.743***	0.169	0.154	0.212*	1.531***	2.304***
	(0.131)	(0.152)	(0.182)	(0.161)	(0.164)	(0.217)	(0.141)	(0.218)	(0.160)	(0.125)	(0.177)	(0.195)
Observations	8473	8473	8473	8473	8473	8473	5273	5273	5273	5273	5273	5273
R^2	0.258	0.050	0.125	0.190	0.190	0.128	0.322	0.087	0.207	0.300	0.198	0.085
Model test	2671.019	916.302	2568.158	2279.100	1689.871	488.419	2288.968	297.263	2109.318	3354.224	1285.522	340.555
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Educ. test	1381.002	719.692	1102.488	1573.107	1233.623	409.354	1829.973	239.707	879.476	2601.900	1188.097	320.159
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Note: Sample include individuals between 18 and 65 years. Observations weighted by the survey's probability weights, bootstrapped-VCE estimated according to the complex sampling design in parenthesis and p-values of the joint-hypotheses tests in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table 1.A5 – Regression models by gender (full specification) in the formal sector, 2005

	Females						Males					
	OLS	**RIF**					**OLS**	**RIF**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Schooling (years)	-0.074***	0.001	0.012	-0.028*	-0.102***	-0.226***	-0.013	0.037	0.062	-0.045	-0.070***	-0.140**
	(0.021)	(0.035)	(0.023)	(0.015)	(0.023)	(0.059)	(0.038)	(0.084)	(0.072)	(0.042)	(0.023)	(0.066)
Schooling (years)^2	0.006***	0.001	0.000	0.003***	0.009***	0.016***	0.002	-0.001	-0.002	0.004	0.005***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.001)	(0.003)
Age (years)	0.018	0.020*	0.032**	0.040***	0.003	-0.009	0.030*	0.040	0.031	0.036	0.012	0.002
	(0.012)	(0.011)	(0.015)	(0.011)	(0.014)	(0.030)	(0.016)	(0.038)	(0.034)	(0.025)	(0.018)	(0.047)
Age (years)^2	-0.000	-0.000	-0.000*	-0.000**	0.000	0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
If indigenous (d)	-0.085**	-0.142*	-0.048	-0.081*	-0.031	-0.036	-0.107**	0.125	0.128	-0.077	-0.201***	-0.378***
	(0.039)	(0.073)	(0.049)	(0.046)	(0.062)	(0.107)	(0.049)	(0.096)	(0.082)	(0.114)	(0.072)	(0.075)
If hh in urban area (d)	0.035	0.050	0.056	-0.034	-0.069	0.152	0.067	0.103	0.098*	-0.150**	0.031	0.211***
	(0.056)	(0.064)	(0.056)	(0.047)	(0.068)	(0.119)	(0.046)	(0.110)	(0.060)	(0.072)	(0.063)	(0.069)
Industry												
-Mining and Quarrying	1.031***	0.507***	0.520***	0.706***	1.008***	2.056***	1.205***	0.265	0.651***	0.790***	0.972***	3.735***
	(0.100)	(0.136)	(0.119)	(0.129)	(0.123)	(0.262)	(0.303)	(0.346)	(0.250)	(0.199)	(0.183)	(0.524)
-Manufacturing	0.231***	0.183	0.068	0.136*	0.180***	0.259***	0.077	-0.052	0.311	0.129	0.089	0.192
	(0.057)	(0.152)	(0.103)	(0.081)	(0.066)	(0.083)	(0.171)	(0.415)	(0.215)	(0.132)	(0.093)	(0.255)
-Construction	0.133**	-0.005	0.021	0.120**	0.112	0.083	0.447	0.419	0.483	0.649	0.590*	0.593
	(0.060)	(0.193)	(0.126)	(0.060)	(0.081)	(0.110)	(0.339)	(0.367)	(0.418)	(0.453)	(0.346)	(0.702)
-Wholesale and Retail	0.161**	0.264*	0.018	0.025	0.035	0.217*	-0.017	-0.114	0.023	0.003	0.006	-0.315
	(0.063)	(0.140)	(0.103)	(0.070)	(0.069)	(0.118)	(0.209)	(0.481)	(0.288)	(0.206)	(0.110)	(0.236)
-Transport, Storage, and Comm.	-0.088	-0.149	-0.277	-0.216***	-0.066	0.170	0.731*	0.426	0.470	0.533	0.598	1.984*
	(0.123)	(0.195)	(0.174)	(0.081)	(0.111)	(0.239)	(0.424)	(0.383)	(0.471)	(0.420)	(0.385)	(1.104)
-Finance, Insurance, and Real Estate	0.420***	0.258**	0.090	0.109*	0.295***	0.921***	0.294*	0.366	0.440	0.275	0.161	0.452
	(0.130)	(0.124)	(0.105)	(0.061)	(0.101)	(0.336)	(0.165)	(0.343)	(0.281)	(0.174)	(0.122)	(0.301)
-Community, Social and Personal Svs	-0.014	0.241*	0.128*	0.139**	-0.135*	-0.454***	-0.056	0.150	0.456*	0.057	-0.217	-0.623**
	(0.065)	(0.129)	(0.077)	(0.063)	(0.071)	(0.121)	(0.197)	(0.355)	(0.252)	(0.171)	(0.152)	(0.307)
Occupation												
-Managers, Profess. and Armed forces	0.658***	0.556***	0.672***	0.781***	0.570***	0.357**	0.873***	0.599**	1.366***	1.138***	0.504***	0.513*
	(0.092)	(0.089)	(0.074)	(0.082)	(0.091)	(0.182)	(0.164)	(0.250)	(0.208)	(0.132)	(0.154)	(0.308)
-Technicians and associates	0.348***	0.474***	0.575***	0.532***	0.195**	0.206	0.521***	0.421*	1.054***	0.660***	0.155	0.490*
	(0.074)	(0.103)	(0.079)	(0.086)	(0.076)	(0.131)	(0.133)	(0.252)	(0.249)	(0.110)	(0.114)	(0.275)
-Clerks	0.401***	0.447***	0.464***	0.416***	0.308***	0.200*	0.569***	0.472**	1.092***	0.659***	0.205**	0.349**

Continued on next page

Table 1.A5 – Regression models by gender (full specification) in the formal sector, 2005 (*continued from previous page*)

	Females						Males					
	OLS	**RIF**					**OLS**	**RIF**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
	(0.059)	(0.088)	(0.077)	(0.076)	(0.086)	(0.105)	(0.082)	(0.213)	(0.201)	(0.114)	(0.087)	(0.174)
-Service and sales worker	-0.091	0.031	0.180*	0.010	-0.160***	-0.424***	0.004	-0.428	0.275	0.055	0.013	0.358
	(0.074)	(0.124)	(0.103)	(0.080)	(0.060)	(0.124)	(0.174)	(0.451)	(0.331)	(0.125)	(0.085)	(0.267)
-Skilled agric. and fish. workers	0.479	0.882***	0.174	0.699	0.550	0.794						
	(0.393)	(0.127)	(0.494)	(0.491)	(0.429)	(0.759)						
-Craft and related trades workers	0.125**	0.356***	0.327***	0.312***	0.063	-0.235**	0.019	-0.099	0.145	-0.062	-0.060	-0.139
	(0.055)	(0.089)	(0.088)	(0.075)	(0.068)	(0.104)	(0.143)	(0.407)	(0.219)	(0.108)	(0.078)	(0.217)
-Plant and machine operators	0.133**	0.349***	0.346***	0.251***	0.009	-0.322***	0.081	-0.129	-0.122	0.316	-0.148	-0.266
	(0.066)	(0.081)	(0.094)	(0.073)	(0.099)	(0.125)	(0.278)	(0.837)	(0.732)	(0.552)	(0.098)	(0.232)
Constant	0.401*	-0.425	-0.453	-0.216	1.258***	2.115***	-0.034	-0.913	-1.459*	-0.008	1.417***	1.855**
	(0.243)	(0.329)	(0.347)	(0.244)	(0.297)	(0.538)	(0.386)	(0.868)	(0.770)	(0.478)	(0.302)	(0.885)
Observations	3245	3245	3245	3245	3245	3245	1461	1461	1461	1461	1461	1461
R^2	0.414	0.104	0.224	0.354	0.282	0.229	0.367	0.158	0.361	0.295	0.192	0.221
Model test	12086.98	3769.29	14685.49	6.5e+05	10253.88	2.4e+05	20705.70	1.8e+05	14730.23	15613.87	57237.55	4043.31
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Educ. test	62.45	5.15	9.38	52.78	105.45	43.25	9.36	0.23	1.16	6.24	21.84	14.19
	[0.00]	[0.08]	[0.01]	[0.00]	[0.00]	[0.00]	[0.01]	[0.89]	[0.56]	[0.04]	[0.00]	[0.00]
Industry test	145.41	54.62	90.82	108.91	235.53	102.44	23.60	17.58	42.56	100.57	94.57	102.97
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]
Occup. test	102.42	92.60	126.27	184.92	74.26	32.74	72.85	55.39	165.13	89.63	30.61	7.07
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]		[0.00]	[0.00]	[0.00]	[0.00]	[0.31]

Note: Sample include individuals between 18 and 65 years. Observations weighted by the survey's probability weights, bootstrapped-VCE estimated according to the complex sampling design in parenthesis and p-values of the joint-hypotheses tests in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table 1.A6 – Regression models by gender (full specification) in the formal sector, 2017

	Females						Males					
	OLS	**RIF**					**OLS**	**RIF**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Schooling (years)	-0.056***	0.037*	0.026	-0.021	-0.107***	-0.239***	-0.041***	0.073***	0.026	-0.049***	-0.130***	-0.163***
	(0.019)	(0.021)	(0.021)	(0.017)	(0.014)	(0.027)	(0.014)	(0.022)	(0.018)	(0.016)	(0.017)	(0.022)
Schooling (years)^2	0.005***	-0.001	0.000	0.003***	0.008***	0.015***	0.005***	-0.002*	0.001	0.006***	0.009***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age (years)	0.030***	0.028***	0.038***	0.035***	0.028***	0.021**	0.029***	0.020***	0.022***	0.039***	0.033***	0.031***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.010)	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)	(0.010)
Age (years)^2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000***	-0.000**	-0.000***	-0.000***	-0.000***	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
If indigenous (d)	-0.089***	-0.008	-0.064	-0.058	-0.005	-0.174***	-0.118***	0.105*	0.028	-0.154**	-0.216***	-0.258***
	(0.027)	(0.029)	(0.042)	(0.041)	(0.045)	(0.044)	(0.043)	(0.062)	(0.060)	(0.074)	(0.058)	(0.050)
If hh in urban area (d)	0.112***	0.106**	0.105***	0.098***	0.135***	0.150***	0.117***	0.095**	0.092**	0.136***	0.128**	0.128***
	(0.021)	(0.043)	(0.036)	(0.032)	(0.034)	(0.025)	(0.037)	(0.044)	(0.044)	(0.046)	(0.052)	(0.039)
Industry												
-Mining and Quarrying	0.464***	0.335***	0.305***	0.447***	0.494***	0.461***	0.437**	-0.048	-0.010	0.258*	0.479*	1.247**
	(0.046)	(0.053)	(0.033)	(0.049)	(0.062)	(0.079)	(0.176)	(0.083)	(0.090)	(0.139)	(0.274)	(0.577)
-Manufacturing	0.113***	0.126**	0.049	0.074**	0.049	0.114**	0.028	-0.197**	-0.170	0.032	0.139**	0.098
	(0.034)	(0.059)	(0.036)	(0.034)	(0.042)	(0.051)	(0.059)	(0.079)	(0.107)	(0.096)	(0.064)	(0.090)
-Construction	0.180***	0.170**	0.179***	0.232***	0.148***	0.121**	0.000	-0.152	-0.191	-0.025	0.099	0.276**
	(0.044)	(0.076)	(0.046)	(0.057)	(0.043)	(0.047)	(0.045)	(0.102)	(0.121)	(0.076)	(0.073)	(0.132)
-Wholesale and Retail	0.046	0.118*	0.009	-0.002	0.033	0.014	-0.115**	-0.240***	-0.274***	-0.127	0.004	0.126
	(0.037)	(0.062)	(0.048)	(0.045)	(0.031)	(0.041)	(0.058)	(0.078)	(0.091)	(0.094)	(0.055)	(0.079)
-Transport, Storage, and Comm.	0.114***	0.124*	0.051	0.103***	0.075**	0.127	-0.002	-0.109	-0.155	-0.028	0.037	0.141
	(0.040)	(0.072)	(0.056)	(0.037)	(0.034)	(0.083)	(0.051)	(0.104)	(0.107)	(0.084)	(0.065)	(0.103)
-Finance, Insurance, and Real Estate	0.055*	0.150**	-0.011	0.017	-0.017	0.073	-0.009	-0.193**	-0.299***	-0.002	0.093	0.110**
	(0.029)	(0.063)	(0.044)	(0.047)	(0.032)	(0.055)	(0.058)	(0.076)	(0.084)	(0.080)	(0.063)	(0.055)
-Community, Social and Personal Svs	0.016	0.116**	0.028	0.081*	-0.023	-0.141***	-0.051	-0.161**	-0.151*	0.067	0.017	-0.019
	(0.034)	(0.059)	(0.040)	(0.048)	(0.038)	(0.051)	(0.042)	(0.072)	(0.084)	(0.059)	(0.036)	(0.053)
Occupation												
-Managers, Profess. and Armed forces	0.557***	0.399***	0.576***	0.792***	0.541***	0.272***	0.396***	0.306***	0.565***	0.610***	0.265***	-0.007
	(0.024)	(0.050)	(0.045)	(0.029)	(0.038)	(0.063)	(0.045)	(0.053)	(0.042)	(0.045)	(0.096)	(0.114)
-Technicians and associates	0.400***	0.313***	0.464***	0.579***	0.381***	0.222***	0.331***	0.323***	0.609***	0.390***	0.132**	-0.015
	(0.036)	(0.043)	(0.040)	(0.046)	(0.053)	(0.056)	(0.047)	(0.066)	(0.064)	(0.071)	(0.053)	(0.046)
-Clerks	0.321***	0.326***	0.391***	0.431***	0.264***	0.108***	0.264***	0.253***	0.477***	0.351***	0.116**	-0.024

Continued on next page

Table 1.A6 – Regression models by gender (full specification) in the formal sector, 2017 (*continued from previous page*)

	Females						**Males**					
	OLS			**RIF**			**OLS**			**RIF**		
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
	(0.029)	(0.051)	(0.037)	(0.042)	(0.030)	(0.039)	(0.041)	(0.061)	(0.054)	(0.062)	(0.059)	(0.063)
-Service and sales worker	0.067*	-0.007	0.064	0.078*	0.029	0.035	0.029	0.014	0.173***	0.059	-0.037	-0.130*
	(0.038)	(0.084)	(0.061)	(0.043)	(0.035)	(0.043)	(0.047)	(0.090)	(0.062)	(0.071)	(0.052)	(0.067)
-Skilled agric. and fish. workers	0.133	-0.051	0.222	0.205	0.254*	0.065	0.090	0.047	0.636**	0.052	0.016	0.028
	(0.181)	(0.414)	(0.230)	(0.136)	(0.154)	(0.095)	(0.140)	(0.305)	(0.249)	(0.335)	(0.090)	(0.087)
-Craft and related trades workers	0.155***	0.233***	0.281***	0.266***	0.107**	-0.064*	-0.116	-0.275*	0.043	-0.061	-0.102	0.018
	(0.027)	(0.037)	(0.036)	(0.040)	(0.043)	(0.035)	(0.102)	(0.145)	(0.134)	(0.126)	(0.087)	(0.159)
-Plant and machine operators	0.231***	0.273***	0.363***	0.310***	0.120***	0.035	-0.300**	-0.095	-0.121	-0.322***	-0.212***	-0.160
	(0.028)	(0.055)	(0.043)	(0.049)	(0.034)	(0.045)	(0.131)	(0.341)	(0.269)	(0.082)	(0.077)	(0.106)
Constant	0.951***	0.080	0.081	0.580***	1.610***	2.716***	0.792***	0.250	0.285*	0.308**	1.547***	2.285***
	(0.140)	(0.131)	(0.183)	(0.173)	(0.175)	(0.237)	(0.137)	(0.201)	(0.158)	(0.128)	(0.175)	(0.210)
Observations	8473	8473	8473	8473	8473	8473	5273	5273	5273	5273	5273	5273
R^2	0.326	0.084	0.188	0.275	0.229	0.141	0.362	0.128	0.272	0.344	0.209	0.095
Model test	35253.48	57240.70	3.5e+05	82363.14	45464.42	5052.66	1.9e+05	17505.87	1.4e+05	7537.77	6737.19	1753.29
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Educ. test	436.39	43.44	116.46	117.54	353.50	179.21	292.91	78.36	117.86	254.73	93.86	97.59
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Industry test	147.41	198.39	151.45	182.18	167.82	96.08	25.88	30.10	32.50	15.76	31.74	32.19
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.03]	[0.00]	[0.00]
Occup. test	1164.36	134.57	438.42	952.32	401.79	47.47	226.72	120.87	279.38	376.27	31.68	8.27
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]		[0.00]	[0.00]	[0.00]	[0.00]	[0.31]

Note: Sample include individuals between 18 and 65 years. Observations weighted by the survey's probability weights, bootstrapped-VCE estimated according to the complex sampling design in parenthesis and p-values of the joint-hypotheses tests in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table 1.A7 – Regressions models by gender (demographics only model) in the informal sector, 2005 and 2017

	Females						Males					
	OLS	**RIF**					**OLS**	**RIF**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
2005												
Schooling (years)	0.047***	0.162***	0.120***	0.050***	-0.014	-0.058***	0.048***	0.183***	0.081***	0.031***	-0.018	-0.028
	(0.008)	(0.022)	(0.011)	(0.007)	(0.009)	(0.016)	(0.015)	(0.025)	(0.021)	(0.011)	(0.015)	(0.018)
Schooling (years)^2	0.001	-0.007***	-0.004***	0.000	0.005***	0.008***	0.001	-0.008***	-0.001	0.002***	0.005***	0.006***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age (years)	0.051***	0.065***	0.052***	0.049***	0.055***	0.049***	0.044***	0.015	0.025***	0.039***	0.062***	0.074***
	(0.003)	(0.008)	(0.006)	(0.004)	(0.005)	(0.008)	(0.006)	(0.016)	(0.009)	(0.008)	(0.009)	(0.011)
Age (years)^2	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***	-0.000***	-0.000	-0.000**	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
If indigenous (d)	-0.179***	-0.404***	-0.251***	-0.125***	-0.068***	-0.093***	-0.169***	-0.330***	-0.281***	-0.105***	-0.105**	-0.051
	(0.018)	(0.057)	(0.033)	(0.021)	(0.023)	(0.021)	(0.038)	(0.110)	(0.056)	(0.033)	(0.042)	(0.040)
If hh in urban area (d)	0.453***	0.684***	0.694***	0.449***	0.304***	0.234***	0.586***	1.334***	0.797***	0.447***	0.269***	0.253***
	(0.023)	(0.057)	(0.032)	(0.022)	(0.016)	(0.022)	(0.040)	(0.111)	(0.064)	(0.035)	(0.029)	(0.023)
Constant	-1.265***	-2.965***	-2.023***	-1.100***	-0.594***	0.022	-1.531***	-3.153***	-1.894***	-1.147***	-0.912***	-0.561***
	(0.053)	(0.161)	(0.111)	(0.080)	(0.101)	(0.151)	(0.152)	(0.348)	(0.173)	(0.148)	(0.194)	(0.210)
Observations	13296	13296	13296	13296	13296	13296	7838	7838	7838	7838	7838	7838
R^2	0.222	0.083	0.157	0.183	0.131	0.087	0.203	0.109	0.139	0.144	0.123	0.060
Model test	2911.232	1272.204	1800.845	2634.208	1678.086	553.799	2276.907	913.014	2509.572	2027.585	1460.989	409.495
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Educ. test	392.349	69.453	214.370	712.064	323.896	278.876	221.039	53.105	76.774	186.233	385.634	89.607
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
2017												
Schooling (years)	0.002	0.079***	0.052***	0.012	-0.024**	-0.096***	0.017*	0.074***	0.054***	0.014**	-0.016	-0.044***
	(0.008)	(0.021)	(0.011)	(0.009)	(0.010)	(0.009)	(0.009)	(0.021)	(0.013)	(0.007)	(0.010)	(0.015)
Schooling (years)^2	0.003***	-0.002*	-0.001	0.002***	0.004***	0.010***	0.002***	-0.001	-0.000	0.002***	0.004***	0.007***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Age (years)	0.030***	0.033***	0.031***	0.027***	0.034***	0.030***	0.025***	0.016	0.010	0.016***	0.040***	0.058***
	(0.003)	(0.008)	(0.005)	(0.004)	(0.005)	(0.005)	(0.007)	(0.014)	(0.007)	(0.005)	(0.006)	(0.009)
Age (years)^2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000*	-0.000***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
If indigenous (d)	-0.056***	-0.056	-0.055**	-0.037**	-0.076***	-0.122***	0.010	0.135**	0.043	-0.034	-0.064**	-0.113**
	(0.016)	(0.049)	(0.027)	(0.017)	(0.023)	(0.025)	(0.027)	(0.066)	(0.044)	(0.026)	(0.032)	(0.048)
If hh in urban area (d)	0.457***	0.780***	0.664***	0.397***	0.258***	0.230***	0.666***	1.291***	0.993***	0.497***	0.302***	0.249***
	(0.020)	(0.042)	(0.028)	(0.021)	(0.023)	(0.032)	(0.024)	(0.077)	(0.034)	(0.024)	(0.021)	(0.026)
Constant	0.259***	-1.171***	-0.410***	0.436***	0.827***	1.448***	-0.304**	-2.059***	-0.786***	0.260**	0.365***	0.487**
	(0.080)	(0.184)	(0.104)	(0.090)	(0.102)	(0.122)	(0.119)	(0.258)	(0.140)	(0.104)	(0.127)	(0.197)
Observations	16913	16913	16913	16913	16913	16913	13356	13356	13356	13356	13356	13356
R^2	0.161	0.065	0.125	0.117	0.091	0.065	0.161	0.070	0.123	0.113	0.077	0.053
Model test	1386.627	1253.004	1674.052	1485.384	900.294	449.847	3602.882	1119.745	2883.952	3445.113	2045.784	601.757
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Educ. test	437.465	104.388	210.278	354.183	780.618	310.944	887.124	129.340	253.944	300.944	254.415	334.989
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Note: Sample include individuals between 18 and 65 years. Observations weighted by the survey's probability weights, bootstrapped-VCE estimated according to the complex sampling design in parenthesis and p-values of the joint-hypotheses tests in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table 1.A8 – Regression models by gender (full specification) in the informal sector, 2005

	Females						Males					
	OLS			**RIF**			**OLS**			**RIF**		
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Schooling (years)	0.065***	0.146***	0.109***	0.054***	0.023**	0.007	0.079***	0.224***	0.109***	0.059***	0.019	0.006
	(0.007)	(0.019)	(0.010)	(0.008)	(0.010)	(0.017)	(0.014)	(0.022)	(0.020)	(0.012)	(0.015)	(0.019)
Schooling (years)^2	-0.002***	-0.007***	-0.005***	-0.001**	0.001	0.002	-0.002**	-0.011***	-0.004***	-0.001	0.002	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age (years)	0.058***	0.083***	0.067***	0.054***	0.055***	0.047***	0.047***	0.020	0.028***	0.043***	0.065***	0.076***
	(0.003)	(0.007)	(0.007)	(0.004)	(0.006)	(0.008)	(0.006)	(0.016)	(0.008)	(0.008)	(0.009)	(0.011)
Age (years)^2	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***	-0.000***	-0.000	-0.000**	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
If indigenous (d)	-0.168***	-0.374***	-0.231***	-0.119***	-0.071***	-0.092***	-0.176***	-0.375***	-0.297***	-0.101***	-0.093**	-0.045
	(0.020)	(0.057)	(0.036)	(0.022)	(0.023)	(0.020)	(0.034)	(0.109)	(0.051)	(0.032)	(0.041)	(0.040)
If hh in urban area (d)	0.241***	0.266***	0.336***	0.259***	0.193***	0.189***	0.519***	1.300***	0.740***	0.368***	0.191***	0.196***
	(0.030)	(0.081)	(0.047)	(0.023)	(0.021)	(0.031)	(0.038)	(0.105)	(0.057)	(0.036)	(0.033)	(0.023)
Industry												
-Mining and Quarrying	0.701***	0.600***	0.544***	0.657***	0.854***	0.742**	0.751**	2.307***	0.642	0.608	0.209	0.061
	(0.123)	(0.094)	(0.123)	(0.116)	(0.218)	(0.377)	(0.311)	(0.527)	(0.835)	(0.416)	(0.274)	(0.242)
-Manufacturing	0.133***	0.112	0.187***	0.230***	0.183***	-0.150*	0.069	-0.157	0.204	0.233**	0.065	0.055
	(0.037)	(0.080)	(0.065)	(0.045)	(0.065)	(0.088)	(0.088)	(0.167)	(0.163)	(0.095)	(0.137)	(0.244)
-Construction	0.223***	0.137*	0.225***	0.283***	0.361***	0.100	0.311***	0.119	0.396**	0.795***	0.174	-0.403***
	(0.040)	(0.080)	(0.076)	(0.060)	(0.054)	(0.066)	(0.095)	(0.157)	(0.171)	(0.117)	(0.245)	(0.130)
-Wholesale and Retail	0.084*	-0.064	0.028	0.131***	0.167***	0.178***	0.162***	-0.097	0.179	0.182**	0.216***	0.185**
	(0.049)	(0.090)	(0.086)	(0.048)	(0.055)	(0.059)	(0.056)	(0.136)	(0.136)	(0.081)	(0.057)	(0.073)
-Transport, Storage, and Comm.	-0.021	0.034	-0.010	0.063	-0.064	-0.149	0.295**	-0.016	0.297	0.226*	0.249**	0.572**
	(0.042)	(0.087)	(0.075)	(0.058)	(0.052)	(0.105)	(0.122)	(0.195)	(0.189)	(0.119)	(0.117)	(0.291)
-Finance, Insurance, and Real Estate	0.140**	0.102	0.141**	0.221***	0.138	0.181	0.295***	-0.084	0.122	0.406***	0.502***	0.517**
	(0.064)	(0.092)	(0.068)	(0.065)	(0.090)	(0.144)	(0.088)	(0.114)	(0.169)	(0.119)	(0.136)	(0.230)
-Community, Social and Personal Svs	0.123**	0.116	0.128**	0.177***	0.083	0.015	0.168**	-0.166	0.253*	0.295***	0.213***	0.088
	(0.053)	(0.081)	(0.063)	(0.059)	(0.062)	(0.107)	(0.070)	(0.140)	(0.139)	(0.085)	(0.061)	(0.116)
Occupation												
-Managers, Profess. and Armed forces	0.595***	0.105	0.139**	0.304***	0.988***	1.518***	0.424***	0.453***	0.262***	0.299***	0.643***	0.633***
	(0.061)	(0.082)	(0.056)	(0.047)	(0.089)	(0.226)	(0.066)	(0.124)	(0.094)	(0.076)	(0.087)	(0.125)
-Technicians and associates	0.474***	0.140***	0.112**	0.219***	0.694***	0.951***	0.344***	0.099	0.204***	0.168***	0.485***	0.606**
	(0.068)	(0.046)	(0.057)	(0.058)	(0.080)	(0.135)	(0.092)	(0.133)	(0.075)	(0.055)	(0.088)	(0.244)
-Clerks	0.280***	0.004	0.126	0.245***	0.413***	0.560***	0.230***	0.239***	0.175*	0.175**	0.325**	0.248

Continued on next page

Table 1.A8 – Regression models by gender (full specification) in the informal sector, 2005 (*continued from previous page*)

	Females						**Males**					
	OLS			**RIF**			**OLS**			**RIF**		
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
	(0.068)	(0.117)	(0.081)	(0.078)	(0.098)	(0.200)	(0.056)	(0.091)	(0.096)	(0.083)	(0.127)	(0.152)
-Service and sales worker	0.127***	0.085	-0.011	0.065**	0.170***	0.318***	-0.144***	-0.230***	-0.192***	-0.162***	-0.102**	-0.075
	(0.021)	(0.062)	(0.038)	(0.033)	(0.042)	(0.105)	(0.036)	(0.070)	(0.046)	(0.043)	(0.049)	(0.064)
-Skilled agric. and fish. workers	-0.284***	-0.830***	-0.643***	-0.183***	0.014	0.067**	-0.015	-0.172	0.052	-0.039	-0.028	0.027
	(0.046)	(0.114)	(0.068)	(0.045)	(0.037)	(0.029)	(0.072)	(0.215)	(0.134)	(0.076)	(0.052)	(0.072)
-Craft and related trades workers	0.044**	-0.013	0.015	0.017	0.083	0.234***	-0.411***	-1.073***	-0.600***	-0.362***	-0.049	-0.113
	(0.021)	(0.053)	(0.041)	(0.035)	(0.058)	(0.079)	(0.079)	(0.217)	(0.117)	(0.105)	(0.134)	(0.265)
-Plant and machine operators	0.065*	-0.011	0.004	0.043	0.096	0.070	-0.436***	-0.960***	-0.665***	-0.234	-0.206*	-0.167
	(0.039)	(0.050)	(0.066)	(0.072)	(0.065)	(0.091)	(0.119)	(0.287)	(0.227)	(0.185)	(0.113)	(0.245)
Constant	-1.255***	-2.762***	-1.870***	-1.113***	-0.689***	-0.107	-1.637***	-3.004***	-2.000***	-1.323***	-1.109***	-0.711***
	(0.051)	(0.171)	(0.098)	(0.074)	(0.097)	(0.141)	(0.142)	(0.357)	(0.203)	(0.141)	(0.202)	(0.227)
Observations	13296	13296	13296	13296	13296	13296	7838	7838	7838	7838	7838	7838
R^2	0.275	0.109	0.198	0.215	0.186	0.134	0.244	0.131	0.160	0.173	0.156	0.080
Model test	1.9e+08	4293.66	65278.61	1.5e+06	1.9e+05	3.6e+06	82363.46	80963.03	2931.21	9157.24	11287.86	26911.05
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Educ. test	220.02	60.23	123.31	241.52	113.76	108.93	108.95	104.74	38.99	89.54	137.01	43.29
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Industry test	85.20	117.87	39.48	61.93	61.40	65.16	34.70	77.57	10.32	76.65	39.37	72.93
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.17]	[0.00]	[0.00]	[0.00]
Occup. test	207.94	129.58	175.50	55.29	235.78	138.71	98.36	85.96	55.30	52.62	98.90	66.30
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]

Note: Sample include individuals between 18 and 65 years. Observations weighted by the survey's probability weights, bootstrapped-VCE estimated according to the complex sampling design in parenthesis and p-values of the joint-hypotheses tests in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table 1.A9 – Regression models by gender (full specification) in the informal sector, 2017

	Females						**Males**					
	OLS	**RIF**					**OLS**	**RIF**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Schooling (years)	0.032***	0.082***	0.060***	0.028***	0.010	-0.011	0.044***	0.099***	0.074***	0.031***	0.011	-0.006
	(0.008)	(0.021)	(0.010)	(0.008)	(0.010)	(0.011)	(0.009)	(0.023)	(0.014)	(0.008)	(0.010)	(0.014)
Schooling (years)^2	-0.000	-0.004***	-0.002***	-0.000	0.001*	0.003***	-0.000	-0.003*	-0.002**	0.000	0.001**	0.003***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age (years)	0.040***	0.054***	0.047***	0.035***	0.039***	0.034***	0.027***	0.019	0.014*	0.019***	0.041***	0.058***
	(0.003)	(0.008)	(0.005)	(0.004)	(0.005)	(0.006)	(0.007)	(0.016)	(0.007)	(0.005)	(0.007)	(0.009)
Age (years)^2	-0.000***	-0.001***	-0.001***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000**	-0.000***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
If indigenous (d)	-0.033**	-0.006	-0.024	-0.027	-0.065***	-0.095***	0.019	0.112*	0.057	-0.018	-0.044	-0.092**
	(0.016)	(0.049)	(0.026)	(0.016)	(0.021)	(0.023)	(0.026)	(0.060)	(0.046)	(0.027)	(0.034)	(0.046)
If hh in urban area (d)	0.213***	0.240***	0.291***	0.198***	0.125***	0.192***	0.628***	1.380***	0.889***	0.414***	0.242***	0.209***
	(0.022)	(0.054)	(0.032)	(0.021)	(0.022)	(0.040)	(0.028)	(0.088)	(0.035)	(0.026)	(0.023)	(0.034)
Industry												
-Mining and Quarrying	0.416***	0.328***	0.370***	0.407***	0.496***	0.454***	0.328**	0.815**	0.999***	0.254	-0.441***	-0.331***
	(0.067)	(0.070)	(0.060)	(0.066)	(0.111)	(0.155)	(0.135)	(0.339)	(0.217)	(0.298)	(0.076)	(0.129)
-Manufacturing	0.045	0.081	0.100*	0.146**	0.051	-0.170**	-0.256***	-0.764***	-0.149	0.030	-0.138*	-0.223*
	(0.036)	(0.071)	(0.056)	(0.058)	(0.053)	(0.081)	(0.073)	(0.194)	(0.117)	(0.078)	(0.075)	(0.130)
-Construction	0.196***	0.076	0.158***	0.320***	0.318***	0.040	-0.035	-0.427*	0.156	0.109	0.050	-0.139
	(0.030)	(0.051)	(0.048)	(0.047)	(0.037)	(0.063)	(0.111)	(0.228)	(0.141)	(0.134)	(0.138)	(0.130)
-Wholesale and Retail	0.026	-0.055	-0.093	0.032	0.152***	0.057	-0.094***	-0.501***	-0.173**	-0.092*	0.064	0.116*
	(0.028)	(0.081)	(0.062)	(0.043)	(0.029)	(0.067)	(0.034)	(0.086)	(0.084)	(0.051)	(0.047)	(0.064)
-Transport, Storage, and Comm.	-0.052	0.003	-0.128**	-0.064	0.040	-0.037	-0.157**	-0.366	-0.121	-0.112	-0.183**	-0.152
	(0.045)	(0.055)	(0.063)	(0.057)	(0.068)	(0.084)	(0.075)	(0.241)	(0.132)	(0.141)	(0.086)	(0.121)
-Finance, Insurance, and Real Estate	0.115**	0.117**	0.043	0.125**	0.151**	0.109	0.074	-0.398***	0.010	0.116	0.261***	0.249
	(0.052)	(0.051)	(0.058)	(0.051)	(0.062)	(0.113)	(0.065)	(0.134)	(0.108)	(0.073)	(0.098)	(0.154)
-Community, Social and Personal Svs	0.076*	-0.012	0.019	0.125**	0.226***	0.076	-0.042	-0.483***	-0.103	-0.017	0.131**	0.120
	(0.041)	(0.064)	(0.055)	(0.049)	(0.042)	(0.095)	(0.035)	(0.100)	(0.098)	(0.038)	(0.060)	(0.078)
Occupation												
-Managers, Profess. and Armed forces	0.601***	0.192***	0.253***	0.337***	0.748***	1.608***	0.395***	0.211**	0.232***	0.282***	0.470***	0.772***
	(0.070)	(0.062)	(0.053)	(0.053)	(0.077)	(0.184)	(0.067)	(0.087)	(0.064)	(0.068)	(0.077)	(0.178)
-Technicians and associates	0.450***	0.049	0.156***	0.264***	0.569***	1.089***	0.418***	0.093	0.221***	0.226***	0.482***	0.996***
	(0.062)	(0.043)	(0.033)	(0.039)	(0.074)	(0.178)	(0.078)	(0.153)	(0.070)	(0.067)	(0.075)	(0.160)
-Clerks	0.332***	0.277***	0.263***	0.177***	0.283***	0.711***	0.155***	0.223**	0.282***	0.238***	0.139**	-0.081

Continued on next page

Table 1.A9 – Regression models by gender (full specification) in the informal sector, 2017 (*continued from previous page*)

	Females						**Males**					
	OLS	**RIF**					**OLS**	**RIF**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
	(0.048)	(0.047)	(0.050)	(0.061)	(0.087)	(0.149)	(0.033)	(0.090)	(0.041)	(0.048)	(0.060)	(0.094)
-Service and sales worker	0.099**	0.055	0.105*	0.029	0.103**	0.289***	-0.066**	-0.066	-0.091*	-0.016	-0.052*	-0.042
	(0.039)	(0.077)	(0.056)	(0.052)	(0.051)	(0.097)	(0.031)	(0.078)	(0.049)	(0.031)	(0.028)	(0.055)
-Skilled agric. and fish. workers	-0.403***	-1.128***	-0.777***	-0.290***	-0.009	0.086***	-0.228***	-0.277**	-0.571***	-0.366***	-0.112***	0.030
	(0.020)	(0.046)	(0.034)	(0.029)	(0.019)	(0.027)	(0.041)	(0.129)	(0.113)	(0.046)	(0.029)	(0.044)
-Craft and related trades workers	0.057*	-0.077	-0.027	0.007	0.191***	0.259***	-0.294***	-0.818***	-0.573***	-0.291***	0.064	0.175
	(0.031)	(0.058)	(0.028)	(0.032)	(0.039)	(0.082)	(0.081)	(0.234)	(0.113)	(0.087)	(0.067)	(0.135)
-Plant and machine operators	0.017	-0.032	0.011	0.035	0.056	0.023	-0.332***	-0.956***	-0.537***	-0.178	-0.050	0.091
	(0.039)	(0.041)	(0.046)	(0.045)	(0.065)	(0.089)	(0.128)	(0.297)	(0.180)	(0.117)	(0.131)	(0.203)
Constant	0.209***	-0.932***	-0.289***	0.394***	0.632***	1.109***	-0.208	-1.669***	-0.556***	0.344***	0.315**	0.398**
	(0.079)	(0.175)	(0.098)	(0.095)	(0.113)	(0.146)	(0.133)	(0.284)	(0.156)	(0.107)	(0.129)	(0.187)
Observations	16913	16913	16913	16913	16913	16913	13356	13356	13356	13356	13356	13356
R^2	0.233	0.111	0.189	0.173	0.139	0.110	0.198	0.094	0.146	0.135	0.101	0.077
Model test	5.5e+07	5.9e+05	4.2e+05	14377.91	1.2e+06	3359.15	2.9e+07	41286.95	31813.04	48109.46	26827.41	67628.30
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Educ. test	124.47	23.54	80.77	110.69	117.67	56.55	547.04	102.74	149.09	210.33	122.65	188.25
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Industry test	103.79	63.94	144.94	109.33	201.96	32.74	59.20	115.52	100.48	22.58	81.37	36.82
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Occup. test	956.25	868.68	889.22	187.36	168.85	232.04	112.14	39.50	89.12	132.64	143.81	68.48
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]

Note: Sample include individuals between 18 and 65 years. Observations weighted by the survey's probability weights, bootstrapped-VCE estimated according to the complex sampling design in parenthesis and p-values of the joint-hypotheses tests in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table 1.A10 – RIF decompositions by sector, 2005, 2011 and 2017

	2005		**2011**		**2017**	
	Endowment	Treatment	Endowment	Treatment	Endowment	Treatment
Formal						
Model 1 (Demograph.)						
Mean	-0.168†	0.050	-0.150†	0.175†	-0.145†	0.166†
P10	-0.096†	0.035	-0.091†	0.146†	-0.067†	0.084†
P25	-0.200†	0.070	-0.144†	0.179†	-0.122†	0.189†
P50	-0.199†	0.030	-0.189†	0.183†	-0.186†	0.204†
P75	-0.132†	-0.066	-0.144†	0.116†	-0.151†	0.138†
P90	-0.171†	0.127	-0.169†	0.197†	-0.133†	0.135†
Model 2 (M1 + Indust)						
Mean	-0.072†	-0.046	-0.093†	0.118†	-0.118†	0.140†
P10	-0.069†	0.008	-0.064†	0.119†	-0.057†	0.074†
P25	-0.199†	0.069	-0.117†	0.152†	-0.115†	0.182†
P50	-0.134†	-0.035	-0.172†	0.166†	-0.187†	0.204†
P75	-0.022	-0.176†	-0.098†	0.069†	-0.127†	0.114†
P90	0.155†	-0.200†	0.043	-0.015	-0.062†	0.063†
Model 3 (M2 + Occup)						
Mean	-0.165†	0.047	-0.179†	0.204†	-0.176†	0.198†
P10	-0.163†	0.102	-0.193†	0.248†	-0.105†	0.123†
P25	-0.393†	0.263†	-0.238†	0.274†	-0.171†	0.238†
P50	-0.234†	0.065	-0.260†	0.255†	-0.252†	0.269†
P75	-0.073†	-0.126†	-0.126†	0.097†	-0.160†	0.147†
P90	0.055	-0.099	-0.017	0.045	-0.068†	0.069†
Informal						
Model 1 (Demograph.)						
Mean	-0.097†	0.289†	-0.064†	0.422†	-0.082†	0.356†
P10	-0.141†	0.516†	-0.127†	0.747†	-0.139†	0.650†
P25	-0.109†	0.368†	-0.088†	0.516†	-0.106†	0.446†
P50	-0.080†	0.239†	-0.050†	0.340†	-0.059†	0.282†
P75	-0.084†	0.164†	-0.043†	0.275†	-0.051†	0.208†
P90	-0.088†	0.118†	-0.034†	0.225†	-0.056†	0.173†
Model 2 (M1 + Indust)						
Mean	-0.066†	0.259†	-0.021	0.380†	-0.072†	0.346†
P10	-0.060†	0.435†	-0.077	0.697†	-0.038	0.550†
P25	-0.073†	0.332†	-0.045	0.472†	-0.055	0.395†
P50	-0.033	0.192†	0.005	0.286†	-0.049	0.272†
P75	-0.100†	0.179†	-0.020	0.253†	-0.101†	0.259†
P90	-0.098†	0.129†	0.018	0.172†	-0.134†	0.250†
Model 3 (M2 + Occup)						
Mean	-0.124†	0.317†	-0.122†	0.481†	-0.129†	0.404†
P10	-0.185†	0.561†	-0.347†	0.967†	-0.202†	0.714†
P25	-0.148†	0.408†	-0.228†	0.656†	-0.167†	0.507†
P50	-0.065†	0.225†	-0.070†	0.361†	-0.107†	0.329†
P75	-0.130†	0.210†	-0.052	0.285†	-0.104†	0.261†
P90	-0.129†	0.160†	0.031	0.160†	-0.094	0.210†

Note: Sample includes individuals between 18 and 65 years. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. Observations weighted by the survey's probability weights and bootstraped-VCE estimated according to the survey's complex sampling design. † Significant at 95% based on the percentile-based bootstrap confidence intervals.

Elaborated by the author based on INEI - National Household Survey (2005-2017)

Table 1.A11 – Difference in RIF decompositions under alternate informal employment definition, 2005, 2011 and 2017

	2005		**2011**		**2017**	
	Diff. endowm.	Diff. treatm.	Diff. endowm.	Diff. treatm.	Diff. endowm.	Diff. treatm.
Model 1 (Demograph.)						
Mean	0.023†	0.027	-0.014†	0.053†	-0.015†	0.068†
P10	-0.013	0.113	-0.050†	0.240†	-0.044†	0.283†
P25	-0.008	0.046	-0.038†	0.167†	-0.054†	0.162†
P50	0.031	0.020	-0.010	0.045	-0.014†	0.042
P75	0.077†	-0.067	0.002	-0.068†	0.002	-0.045†
P90	0.012	0.042	-0.007	-0.121†	0.034	-0.068†
Model 2 (M1 + Indust)						
Mean	0.045†	0.004	0.003	0.036	-0.007	0.061
P10	0.115†	-0.015	0.059	0.130	0.026	0.214†
P25	-0.003	0.041	-0.037	0.167†	-0.021	0.128†
P50	0.025	0.026	0.014	0.020	-0.021	0.049†
P75	0.031	-0.021	0.002	-0.069	-0.018	-0.025
P90	0.062	-0.008	-0.075†	-0.053	0.071†	-0.105†
Model 3 (M2 + Occup)						
Mean	0.030	0.019	-0.074†	0.113†	-0.069	0.123†
P10	0.064	0.036	-0.129	0.319†	-0.080	0.319†
P25	-0.049	0.087	-0.153†	0.283†	-0.103†	0.211†
P50	0.032	0.019	-0.025	0.060	-0.044	0.072
P75	0.036	-0.026	-0.018	-0.049	-0.037	-0.006
P90	0.083†	-0.028	-0.118†	-0.010	0.060	-0.094

Note: Sample includes individuals between 18 and 65 years. Difference calculated as the original SVY minus alternate SVY estimates. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. Observations weighted by the survey's probability weights and bootstraped-VCE estimated according to the survey's complex sampling design. † Significant based on the percentile-based bootstrap 95% confidence intervals.

Elaborated by the author based on INEI - National Household Survey (2005-2017)

Table 1.A12 – Difference in RIF decompositions under alternate employment definitions (only private workers), 2005, 2011 and 2017

	2005		2011		2017	
	Diff. endowm.	Diff. treatm.	Diff. endowm.	Diff. treatm.	Diff. endowm.	Diff. treatm.
Formal						
Model 1 (Demograph.)						
mean	-0.018	-0.067	-0.029†	-0.058†	-0.037†	-0.037†
p10	-0.028	-0.055	-0.035†	0.029	-0.022†	0.005
p25	-0.146†	-0.036	-0.076†	-0.009	-0.056†	0.010
p50	-0.040	-0.137†	-0.078†	-0.087†	-0.079†	-0.029†
p75	0.139	-0.205	0.042†	-0.173†	-0.001	-0.134†
p90	0.106	0.190	0.000	0.050	-0.002	0.008
Model 2 (M1 + Indust)						
mean	-0.080†	-0.004	-0.064†	-0.023	-0.055†	-0.019†
p10	-0.066†	-0.017	-0.039†	0.033	-0.034†	0.017
p25	-0.198†	0.016	-0.090†	0.006	-0.064†	0.018
p50	-0.046	-0.131†	-0.133†	-0.031	-0.105†	-0.003
p75	-0.090	0.023	-0.020	-0.112†	-0.035	-0.100†
p90	0.027	0.269	-0.046	0.097	-0.031	0.038
Model 3 (M2 + Occup)						
mean	-0.070†	-0.015	-0.052†	-0.035†	-0.048†	-0.026†
p10	-0.158†	0.076	-0.013	0.007	-0.042†	0.025
p25	-0.314†	0.132†	-0.134†	0.050	-0.081†	0.035†
p50	0.024	-0.202†	-0.137†	-0.027	-0.084†	-0.024
p75	0.083	-0.149	0.055	-0.186†	-0.016	-0.119†
p90	-0.082	0.377	-0.025	0.076	-0.006	0.013
Informal						
Model 1 (Demograph.)						
mean	-0.003	-0.009†	-0.002	-0.007	-0.002	-0.006†
p10	0.002	-0.013	-0.005	-0.004	0.001	-0.021†
p25	0.007	-0.011	-0.001	-0.013†	0.001	-0.013†
p50	-0.002	-0.017†	-0.002	-0.003	0.000	-0.010†
p75	-0.002	-0.015	-0.006†	-0.007	-0.004†	-0.012†
p90	-0.009	-0.009	-0.003	0.012	-0.006†	0.007
Model 2 (M1 + Indust)						
mean	-0.011†	-0.000	-0.010†	0.001	-0.007†	-0.001
p10	-0.014	0.003	-0.025†	0.016	-0.015†	-0.006
p25	-0.008	0.004	-0.008	-0.006	-0.005	-0.008†
p50	-0.016	-0.002	-0.002	-0.003	-0.009†	0.000
p75	-0.016†	-0.001	-0.009	-0.003	-0.019†	0.002
p90	-0.001	-0.017	-0.006	0.014†	-0.007	0.008
Model 3 (M2 + Occup)						
mean	-0.014†	0.003	-0.015†	0.006†	-0.011†	0.003
p10	-0.010	-0.000	-0.026	0.017	-0.020†	-0.001
p25	-0.017	0.014	-0.010	-0.004	-0.008†	-0.005
p50	-0.028	0.010	0.002	-0.007	-0.009	-0.000
p75	-0.008	-0.008	-0.015†	0.003	-0.010	-0.006
p90	-0.011	-0.007	-0.017	0.025†	-0.013	0.014

Note: Sample includes individuals between 18 and 65 years. Difference calculated as the original SVY minus alternate SVY estimates. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. Observations weighted by the survey's probability weights and bootstraped-VCE estimated according to the survey's complex sampling design. † Significant based on the percentile-based bootstrap 95% confidence intervals.

Elaborated by the author based on INEI - National Household Survey (2005-2017)

Table 1.A13 – Difference in RIF decompositions SRS and SVY by sector, 2005, 2011 and 2017

	2005		2011		2017	
	Diff. endowm.	Diff. treatm.	Diff. endowm.	Diff. treatm.	Diff. endowm.	Diff. treatm.
Formal						
Model 1 (Demograph.)						
Mean	-0.008	0.025	-0.004	0.001	-0.018	0.019
P10	-0.051	-0.001	-0.020	-0.006	-0.016†	0.013
P25	-0.038	0.020	-0.012	-0.010	-0.019	-0.008
P50	-0.032	0.031	-0.027	-0.008	-0.037†	0.003
P75	0.026	0.070	0.003	0.015	-0.017	0.038†
P90	0.073	0.028	0.050	0.034	0.001	0.059†
Model 2 (M1 + Indust)						
Mean	-0.038	0.055	-0.003	0.000	-0.017	0.018
P10	-0.055	0.004	-0.025	-0.000	-0.008	0.005
P25	-0.038	0.020	-0.013	-0.010	-0.018	-0.010
P50	-0.044	0.043	-0.026	-0.010	-0.043†	0.008
P75	-0.005	0.100	0.001	0.018	-0.013	0.034
P90	-0.016	0.117	0.027	0.057	-0.000	0.061
Model 3 (M2 + Occup)						
Mean	-0.036	0.053	-0.002	-0.001	-0.008	0.009
P10	-0.081	0.029	-0.001	-0.024	-0.004	0.001
P25	-0.044	0.025	-0.004	-0.018	-0.003	-0.024
P50	-0.045	0.044	-0.007	-0.029	-0.034	0.000
P75	0.009	0.087	0.006	0.013	-0.003	0.025
P90	0.024	0.078	0.038	0.046	-0.008	0.068
Informal						
Model 1 (Demograph.)						
Mean	0.006	0.023	-0.000	0.027	0.002	0.025
P10	-0.007	0.023	-0.004	0.057	0.004	0.017
P25	0.002	0.009	0.001	0.061	0.001	0.045
P50	0.002	0.039	0.003	0.016	-0.001	0.025
P75	0.014	0.029	0.005	-0.002	0.001	0.014
P90	0.009	0.028	0.000	0.003	-0.002	0.015
Model 2 (M1 + Indust)						
Mean	-0.008	0.038	0.009	0.018	0.028	-0.001
P10	-0.000	0.016	0.079	-0.026	0.073	-0.051
P25	0.004	0.007	0.025	0.037	-0.001	0.047
P50	-0.036	0.077	-0.014	0.033	-0.002	0.027
P75	-0.002	0.045	-0.014	0.018	0.024	-0.009
P90	-0.002	0.039	-0.034	0.037	0.045	-0.032
Model 3 (M2 + Occup)						
Mean	-0.013	0.043	0.008	0.019	0.006	0.022
P10	-0.029	0.045	0.071	-0.019	0.006	0.015
P25	-0.007	0.018	0.029	0.033	-0.009	0.055
P50	-0.040	0.082	-0.011	0.030	0.014	0.011
P75	0.009	0.035	-0.003	0.007	0.020	-0.005
P90	0.021	0.017	-0.039	0.042	0.026	-0.013

Note: Sample includes individuals between 18 and 65 years. Difference calculated as SRS minus SVY estimates. Demographics only model includes education (2nd degree polinomial), age (2nd degree polinomial) as well as ethnicity and urban area dummies. † Significant based on the 95% confidence intervals.

Elaborated by the author based on INEI - National Household Survey (2005-2017)

Chapter 2

The Effect of the Venezuelan Exodus on Peruvian Labour Market Outcomes

2.1 Introduction

Few policy debates capture the attention of the general public and academics alike as the effect of immigration on labour market outcomes. The empirical evidence that provides insights on this has typically relied on long-lasting immigration processes (Dustmann et al. 2007; Okkerse 2008). Some of these studies (see, e.g. Dustmann et al. 2016; Kerr and Kerr 2011) appear to justify the fears of natives in the EU, USA and Britain who believe that migrants lead to higher natives' unemployment and lower wages (Grigorieff et al. 2020; Blinder 2015; Citrin and Sides 2008). As valuable as these contributions are, such sustained migration inflows, by inducing changes in structural characteristics of the economy (like economic cycles, technology, or the composition of the native labour force), plausibly complicate the identification of causal impacts (Dustmann et al. 2008). An alternative strategy that adds credibility to the results consists of instead studying episodes where the arrival of migrants occur suddenly and exogenously (Kerr and Kerr 2011; Frölich and Sperlich 2019). This is the route that we pursue in this paper.

The Venezuelan Exodus of 2016 provides a unique opportunity to recreate a hypothetical situation of what would have occurred in the Peruvian labour market had this immigration never took place while keeping other factors constant. Such a massive episode has its roots in the acute humanitarian crisis and deterioration of the economic conditions in Venezuela, such as the free fall of its per capita GDP and inflation levels that reached one million percent (see Reinhart and Santos 2015; Restuccia 2019). The 4.5 million who forcibly left their country made it the "largest [Exodus] in the [Latin American] region's recent history" (UNHCR 2019d, p. 10), and second only to the recent Syrian migration crisis. This influx of working migrants transformed Peru overnight into the second-largest recipient of Venezuelans, culminating in the fact that by 2019 these represented 2.5% of Peru's population. This is broadly similar to what Syrian refugees represented in Turkey in 2015 (2.8%, see UNHCR 2020, 2019d and Tumen 2015).

We provide quasi-experimental evidence of the impact that this unprecedented influx had on the Peruvian labour market. We focus on the effect on natives' wages, separating the formal and informal sectors in recognition of the fact that immigrants primarily inserted in the latter sector of the labour market. Departing from most of the literature (see the review in Okkerse 2008), we also study how this event affected wages inequality and the size of the informal sector. We exploit a two-stage Differences-in-Differences (2S-DiD) combined with alternative variance estimators customized to suit our research design. We demonstrate that this strategy improves upon the standard DiD as it removes a usually overlooked source of inconsistency in the Average treatment effect on the treated (ATET) estimator and substantially improves the size and power of the Wald tests. The characteristics of the treatment (at an aggregated level within few units) allow us to also apply panel data and Synthetic Control Methods (Abadie et al. 2010, 2015), that relax some identifying assumptions and confirm the suitability of our main method. Additionally, we study the Venezuelan immigrants' skills and the downgrading they experience in the host labour market.

The results of this paper can be split into two. Venezuelans in Peru experience an occupational downgrade in terms of their jobs and the abilities on the tasks they perform in their work relative to pre-emigration. This occurs despite the fact that most hold an upper education level compared to the native Peruvians. This is similar to what Peri (2016) and Dustmann et al. (2016) finds for gradual immigration processes experienced by Europe and the USA. Estimates for the treatment effects of the Exodus for the native working population comport with the findings of Card (2001), Ottaviano and Peri (2012) and Peri and Yasenov (2017), who do not find evidence of significant wage effects in both the formal and informal sectors. This suggests both that several adjustment margins reduce the wage impact of immigrants and the existence of complementarities between natives and immigrants. However, as theoretically suggested by Dustmann et al. (2008), and similar to what Ceritoglu et al. (2015) find in Turkey, there are large negative effects (around -10%) of the treatment on wages for informal workers in Lima y Callao, the epicentre of the Peruvian economy. Results for the other outcome variables differ across the methods: the exodus caused an increase in informal and inequality in the formal sector between 2% and 4% at the end of the period according to our preferred estimation method, whereas under SCM estimation these changes are not found to be statistically significant.

The contribution of this study is twofold. Firstly, we provide quasi-experimental evidence on the labour market impact of the Venezuelan Exodus, which to date has not received sufficient attention despite its relevance. As Peri (2016); Okkerse (2008); Dustmann et al. (2008) suggest, the literature has mainly focused on developed economies to study the labour market effects of immigration (Card 1990; Hunt 1992; Carrington and de Lima 1996; Peri et al. 2020; Foged and Peri 2015), even in the absence of an exogenous shock. Therefore, we extend the typical scenario to one where the natives are principally engaged in the informal sector and have a lower level of education than the immigrants. Unlike the studies on the Syrian immigration in Turkey (Akgündüz et al. 2015, 2018; Ceritoglu et al. 2015; Del Carpio and Wagner 2015), we separately analyze the impacts for the informal and formal labour market. Secondly, our empirical approach corrects the problems induced by department-year shocks and the small number of clusters and treated regions, which surprisingly remained mostly ignored in the literature. Based on Brewer et al. (2018); Frölich and Sperlich (2019); Angrist and Pischke (2009), and the influential study of Bertrand et al. (2002), these plausibly drive the reported magnitude and statistical significance of previous DiD studies. Compared to the few studies which have analyzed the Venezuelan Exodus under different estimation procedures (Asencios and Castellares 2020; Boruchowicz et al. 2021; Morales and Pierola 2020), we analyze all the areas where these migrants settled and calculate the results for the sub-sample most exposed to direct competition from immigrants. This provides a more complete picture of the effects of the Venezuelan Exodus on the Peruvian labour market.

The chapter is organized as follows. In section 2.2 we review some of the previous literature on the impact of immigration on labour market outcomes. Section 3 describes the data sources for the study's key (formal and informal wages) and ancillary variables (informality levels and formal and informal labour market inequality) in the study. Section 4 describes both the Venezuelan Exodus in Peru and the labour market that these immigrants join. Section 5 discusses the econometric methods and checks the validity of the identifying assumptions. Section 6 presents the results of the main analysis, with the results for the subgroup of low skilled natives presented in Appendix B. Section 7 presents a set of robustness checks in terms of variables definitions and alternative estimation methods. Finally, section 2.8 discusses the results in terms of the Exodus's policy implications, relates this with previous studies, and outlines areas for future research.

2.2 Literature Review

The academic concern about the impact of immigrants on wages dates back to at least Card's (1990) study of the 1978 Mariel Boatlift, which comprised the unexpected arrival of almost 60,000 Cubans to Miami. This study profoundly influenced the direction of research in labour economics because of its novel methodological approach and the fact that the reported absence of effects on native employment and wages challenged standard labour market models (Peri and Yasenov 2017; Angrist and Pischke 2009). Theoretically, the standard framework predicts that, in the short run, immigration will result in a fall of wages for those in the host country because it leads to an increase in labour supply given a downward-sloping labour demand curve (Peri 2016).

However, as soon as we incorporate stylized facts about labour markets, such as complementarity and imperfect substitutability of natives and immigrants or perfect capital mobility[1], the effects of an inflow of unskilled workers are no longer straightforward. Not only this inflow will affect natives' wages differently, with the unskilled ones losing out, but also the average effect on natives wages may be zero or even positive (Dustmann et al. 2008). The way that immigration is defined in these models can also influence our understanding of the effects of immigration. As Card and Peri (2016) show, the adverse effects on native labour market outcomes reported by Borja (2014) are explained by the negative bias mechanically induced by his definition of immigration. Focusing instead on the immigrant-driven supply shocks leads to moderate and even zero effects. Additional adjustment mechanisms, such as changes in the skill mix, industry structure, firms preferences or even task differentiation, induces effects beyond what simpler models predict.

One of the most commonly used empirical method within this debate is the national skill-cell approach. This compares a nation's actual supplies of workers in particular skill groups to those it would have had in the absence of immigration and then uses outside information on the elasticity of substitution among skill groups to compute the relative wage consequences of the supply shock (see Okkerse 2008; Dustmann et al. 2016; Peri 2016). For example, within this basic framework, Borjas (2003) reported negative effects of immigration (around -3%) on the average wage of residents in the USA, with the workers at the bottom and on the top of the socio-economic distribution being the most affected. However, once we modify some of its basic assumptions, different results emerge. For instance, Ottaviano and Peri (2012) drop the perfect substitutability between natives and migrants within skill-defined cells and find that immigration over the 1990-2006 period had small positive effects on the wages of the USA's natives instead. Manacorda et al. (2006) follow a similar approach and finds that immigration was associated with positive (but, again, minor) wage effects for UK natives in the 1975-2005 period. Dustmann et al. (2008) assume that immigrants compete with natives within the same wage percentiles and find negative effects along the lower part of the native distribution (where immigrants mostly locate) and positive wage effects further up the distribution. Peri (2007) finds that immigration is associated with an increase in the average wage of native Californians, the state with the highest percentage of working immigrants.

These studies certainly provide valuable insights. However, a key issue is that these do not recreate a hypothetical situation of what would have occurred if immigration had not taken place (Dustmann et al. 2008). Instead, these empirical approaches rely heavily on theoretical models and, consequently, their results can be understood as a simulation of the impact for given elasticities of substitution (Okkerse 2008). Hence, implausible assumptions about the underlying production model used to derive the elasticity of substitution, such as perfect substitution between natives and immigrants, can explain the adverse effects on wages in some studies following the national skill-cell approach (e.g. Borjas 2003). In turn, unexpected waves of immigrants (or refugees) with little ability to choose their destination provide a more robust basis for analysis relative to episodes where the inflow of migrants is not exogenous but rather more gradual in

[1]Following Dustmann et al. 2008, this is a reasonable assumption for small open economies like Peru. Doing away with this assumption implies that average wages (not only wages of the unskilled) may decrease as a consequence of immigration. As shown in Ottaviano and Peri 2012 the faster capital is able to adjust, the smaller will be the effect on average wages in the economy

nature. Because in this case their arrival is sudden and induced by exogenous political factors rather than a thriving host economy, the unresponsiveness of technology and physical capital to these events allows for a better basis to determine their short-run effects. These impacts can thus be claimed as representing causal effects (Frölich and Sperlich 2019; Peri and Yasenov 2017). The Venezuelan Exodus allows us to explore this quasi-experimental route in the study here.

Quasi-experimental evidence from earlier episodes besides Card (1990) reveals that, in general, these large migration inflows are not damaging for native labour market outcomes. Hunt (1992) examines the repatriation of Algerians to France who arrived in 1962 and by 1968 represented 1.6% of the total French labour force. The study exploits the regional variation in the number of repatriates using instrumental variables (IV) estimation. The author finds that the elasticities of wage and employment relative to the proportion of repatriates are negative although small in magnitude (less than 1%). Carrington and de Lima (1996) studied the impact of the "retornados", whose return to Portugal from Angola and Mozambique in the mid-1970s increased the national labour force by 10% in just three years. Based on their preferred set of results, the authors conclude that this slightly increased unemployment in the host country. Nonetheless, they recognize that the Europe-wide unemployment rate increase over the same period highly influences this effect. Friedberg (2001) studies the impact of the returning Israelis between 1989-1994 by exploiting the variation in immigration across occupations. Their results from IV estimation suggests that immigrants do not adversely impact native labour market outcomes; instead, immigrants work in occupations with low wages, low wage growth and contracting employment. Subsequent applications have improved some of the identification issues in which these studies incur. As discussed below, these are related to the estimation of the standard errors and the selection of control groups. For example, the reassessment of the effects by Peri and Yasenov (2017), using more robust methods to estimate the counterfactual, confirms Card's (1990) findings.[2]

In recent years, much of the quasi-experimental literature on immigration has shifted its focus from developed countries to those developing countries affected directly by the Syrian refugee crisis, particularly Turkey. Most of these studies exploit a standard Differences-in-Differences (DiD) method to identify causal impacts and often provide conflicting conclusions. On the one hand, Ceritoglu et al. (2015) find that this treatment had a negligible impact on wages and large increases in unemployment, accruing mainly among the informal workers and those located at the bottom of the wage distribution. On the other hand, Akgündüz et al. (2015), using regional-level panel data, find that employment rates of natives in various skill groups are mainly unaffected. Akgündüz et al. (2018), using a combination of DiD and Synthetic Control Methods (SCM), report that the refugee influx increased the number of foreign firm entries. Additionally, Balkan and Tumen (2016) use DiD and find that the general level of consumer prices declined by approximately 2.5% due to immigration.

These quasi-experimental studies, however, exhibit several limitations which may threaten the validity of their conclusions offered. For instance, those studying the Syrian episode rely on no more than three pre-treatment years and two post-treatment years. This time window is too small to allow them to support the argument that the identification assumptions of their DiD strategy credibly hold or that it captures the effects of temporal heterogeneity. Also, elasticity estimations from Akgunduz (2015 and 2018) are possibly distorted by the non-comparability of data on the number of immigrants used from 2013 onwards. Beyond data-related issues, methodological problems question the validity of their estimated impacts, as we elaborate in subsection 2.5.1. First, none of these studies accounts for the confounding role of the evolution of the department-year shocks in the point estimates of their causal effects reported[3], which affects the

[2] However, Borjas' (2017) reappraisal of this event instead finds dramatic negative wage effects for natives, as large as 30%. These are subsequently confirmed in Borjas and Monras (2017). Yet, Clemens and Hunt (2017) claim that there are reasons to believe that these are actually spurious: in the first case results are explained by unaccounted change in the composition of the sample used, and in the second by the plausible lack of exogeneity of the instrument used.

[3] Admittedly Balkan and Tumen (2016) and Ceritoglu et al. (2015) equate this to trade volumes; but we have reasons to believe that

consistency of the estimators of interest. Second, except for Ceritoglu et al. (2015)[4], most studies recognize the problem of clustering of the idiosyncratic errors. However, their strategy to correct for this lacks the necessary asymptotic justification, which results in over-rejection of the null hypothesis of zero effect, as we explain below. Thirdly, the selection of the control groups in Balkan and Tumen (2016) and Ceritoglu et al. (2015) is made arbitrarily and, in fact, the share of immigrants in some treated areas is not all that much different from the share in some control areas (see e.g. Figure 2 in the former). The more refined construction of the units in the control group using SCM by Akgündüz et al. (2018) could be thwarted by the poor pre-treatment fit of the estimated counterfactual, which suggests a possible bias in the estimation of their treatment effect.

The few studies that have analyzed the impact of Venezuelan immigration on the Peruvian Labour market take a different econometric approach than the one adopted in this chapter. Under an area analysis approach, Morales and Pierola (2020) estimate the effects of the intensity of Venezuelan settlements and find a 3.2% decrease in monthly earnings for workers with secondary education in the formal services sector. In order to circumvent the lack of data on migrants by region, they use Bartik instrumentation. Besides the complications induced by the presence of a negligible amount of immigrants in most Peruvian provinces, this strategy is potentially subject to the problem related to the construction of these type of instrumental variables, pointed out by Clemens and Hunt (2017) .[5] Asencios and Castellares (2020) and Boruchowicz et al. (2021) focus on Lima y Callao and find a slight reduction in hourly wages for some groups of workers, such as young high school dropouts. The former focuses on the changes in employment and income, using the Heckman (1979) sample selection correction. Nonetheless, their identifying assumptions are unclear, and the fulfilment of bivariate normality of the errors in the Mincerian and the selection equation remains untested. Likewise, despite the fact that their estimating equations are derived assuming that the characteristics of the workers remain unchanged, the possible "time-in-sample bias" (Baltagi 2005) from their 3-month rotating panel is not addressed. Boruchowicz et al. (2021), in turn, use the original Synthetic Control Methods (SCM) of Abadie et al. (2010, 2015). In this chapter, we apply a modified version of the original SCM estimator, which is more robust to their reported deviations from the convex hull assumption, and we also perform statistical inference under a more powerful test statistic (see below).

Even though wages adjustments have more immediate social and political repercussions for the host country (Dustmann et al. 2019), these are only one of the channels through which the labour market might react to the arrival of foreign workers (Dustmann et al. 2008; Akgündüz et al. 2018; Balkan and Tumen 2016). Changes in the income distribution can influence future growth prospects (Banerjee and Duflo 2003), while higher informality levels are associated with lower productivity (La Porta and Shleifer 2014). Hence, we also study the impact of the immigrant (or refugee) influx on hourly wages inequality in the formal and informal sectors, as well as on the level of informality.

2.3 Data

This study combines two different data sources. The first micro-dataset is the National Survey of the Resident Venezuelan Population (ENPOVE) in Peru, collected in 2018 by Peru's National Statistical Office (INEI). It provides demographic and labour market information of Venezuelan migrants for each of the five cities where these settled. These cities are comprised of Lima and Callao, Tumbes, Cusco, Arequipa and La Libertad. This data also records the specific route and time spent by the migrants in every location from their departure point in Venezuela until they arrived at their current location in Peru. This information al-

these are not the only factors explaining the department-year effects.

[4]The statistical significance of some of their results can be explained because the White (1980) robust standard errors used can under-estimate the asymptotic variance matrix, resulting in over-sized Wald tests (Cameron and Miller 2015)

[5]An instrumental variable strategy was also pursued by Akgündüz et al. (2018) for the Turkish case. However, in this case the number of observations is too small given their estimation strategy is valid only asymptotically.

lows assessing some aspects of the identification assumptions for the empirical analysis. Migration patterns from ENPOVE closely follow the pattern of the whole population (compare Figure 2.A3 and Figure 3.A2). In addition, it contains information on the last occupation (or job) held by Venezuelans prior to emigration, which allows an analysis of their occupational downgrading. The second micro-dataset is Peru's National Household Survey (ENAHO), also collected by INEI. This is a set of repeated cross-sections from 2005 to 2019 that constitutes the official source in Peru for labour market indicators, as it is statistically representative of workers in both formal and informal sectors. Its sampling design, unaltered since 2005, provides comparable information of the outcomes and controls across the years. As this provides information for natives, this is the dataset used for the econometric analysis in this chapter. We take the treatment as taking place at the department level, with the treated areas being the five departments (regions) where the five cities above are located (see Figure 3.A1 above).

By taking the yearly ENAHO samples, the aggregated variables at the department level to perform the SCM are calculated with a high level of precision (see the survey's documentation in http://iinei.inei.gob.pe/microdatos/). This contrasts with Boruchowicz et al. (2021), who aggregate the (quarterly) ENAHO sample data at a lower administrative level than what the sample design originally allows. By doing so, the collapsed variables used for their estimation are affected by the "small-area estimation" problem (Tzavidis et al. 2018; Pfeffermann 2002), plausibly inducing a measurement error in their results (see Peri and Yasenov 2017 for discussion in the SCM context). The data used here also allow for a larger number of years compared to the studies undertaken for Turkey (reviewed above) and for more pre-treatment periods than the average of those in the Ferman et al. (2020) review. A sample restriction we impose throughout is to include only those individuals between 18 and 65 years old who are employed.[6]

Hourly wages, the ratio of self-reported total wages to self-reported hours of work (including both primary and secondary jobs), are deflated using the Peruvian Central Bank GDP series, taking 2007 as the base year. We approximate informal employment through the lack of affiliation to a pension system. This characteristic is more intimately linked to a crucial aspect of informality, namely the lack of protection by legal and regulatory frameworks that are associated with poor-quality jobs without social security. This is preferred to alternative definitions based on the firm (e.g. if it does not have accounting books or does not provide a legal invoice to hire workers) or in the individual (e.g. working less than 40 hours per week) (Hussmanns 2001; ILO 2002; Freije 2002).[7] Because INEI considers employed those who worked more than 15 hours per week regardless of the occupational category (INEI 2019b, p. 553), we drop those classified as "unpaid family workers" from our samples.

The demographic (at individual-level) control variables in the regressions are sex (equals 1 if male and 0 otherwise), age (in years), years of schooling[8], area of residence (equals 1 if living in an urban area and 0 otherwise). The industry and occupation control variables follow standard international classifications. The vector of six industry dummies is a reduced version of the International Standard Industrial Classification of All Economic Activities (ISIC), Revision 4: 1) Agriculture, forestry and fishing (A and B and C); 2) Manufacturing and Public Utilities (D and E); 3) Construction (F); 4) Wholesale and Retail Trade, Hotels and Restaurants (G and I); 5) Transport, Storage, and Communication (H and J); 6) Finance, Insurance, and Real Estate and Community, Social and Personal Services (K - U). The vector of seven occupation dummies, originally defined as an adaptation of the International Standard Classification of Occupations (ISCO-2008) from ILO: 1) Managers, Professionals and Armed forces (MGs 1, 2 and 10); 2) Technicians

[6]The lower age bound of 18 years old represents the minimum legal age that a person can offer their work and the upper bound represents the legal retirement age in Peru.

[7]Note how by focusing on the main occupation to classify the worker as informal, we conform to Hussmanns (2004) suggestion of taking jobs as the observation units rather than employed persons given the existence of multiple job-holding workers.

[8]Peru's education system does not have an upper cap on the number of years a student can remain at a given level (which is relevant mainly for university or technical institution students). For those who reported concluding the level, we take the minimum amount of years it took to attain. For those who did not conclude, schooling is the minimum of years from the immediate lower level plus the years studied.

and associates (MG 3) and Skilled agricultural and fishery workers (MG 6); 3) Clerks (MG 4); 4) Service and sales workers (MG 5); 5) Building workers, electricians, artisans and telecommunications (MG 7); 6) Industrial machinery operators, assemblers and drivers (MG 8); 7) Elementary occupations (MG 9).[9] The size of the firm, directly related to the firm's productivity (Távara et al. 2014), is not used since it is absent for those workers in the public sector.

By merging occupation data in the ENAHO and ENPOVE with ONET 25.1 (2019) dataset (from US-Department of Labour) we calculated for every worker indices of cognitive, communication and manual intensity as well as a global occupational complexity index following Ottaviano et al. (2013) (see appendix therein).[10] ONET dataset assigns scores that describe the intensity of distinct abilities ("skills") in every occupation. Using Hardy et al. (2018) cross-walks, we translate these from O*NET-SOC-10 to SOC-10 and from this to ISCO-2008 codes. We then use table 2 in the annex of INEI (2016) to correlate the former codes into INEI nomenclature. We take three types of skills from this database: Cognitive Intensity (10 variables classified as "cognitive and analytical"), Communication Intensity (4 variables capturing written and oral expression and understanding) and Manual Intensity (19 variables capturing dexterity, strength, and coordination). Each of the values in these variables is re-expressed as its corresponding percentile to measure the relative importance of a given skill among other workers (e.g. a task with a score of 0.02 for some skill indicates that only 2 percent of workers were supplying that skill less intensively). We then take the average of these re-scaled values within each of the three types of skills. The (overall) Complexity index, $Complexity = (Cognitive + Communication) / ManualIntensity$, which summarises the intensity of cognitive-communication skills relative to manual skills of each worker's occupation.

For the panel data and SCM estimation, micro-level information for wages, informal status and the controls are averaged within departments (using the sampling weights). The variable for inequality in DiD analysis is the Recentered Influence Function for the Gini coefficient (Firpo et al. 2009; Rios-Avila 2020), which provides estimated coefficients reflecting the marginal impact on the index itself. In the panel data and SCM estimation, inequality is represented by the departmental-level Gini coefficient.

2.4 The Venezuelan Exodus in Peru

2.4.1 The Venezuelan immigration process

The Venezuelan Exodus represents a refugee crisis of comparable magnitude to the 6.6 million Syrian refugees. Due to the socio-economic instability and political turmoil in which Venezuela was plunged since 2013, approximately 4.5 million left the country as of January 2020. About 85% settled in neighbouring Latin America and the Caribbean countries (R4V 2020b). This massive volume of refugees and migrants, and the ensuing humanitarian crisis that this triggered, led some to label it the largest external displacement crisis in Latin America's recent history (UNHCR 2019d).

The Venezuelan case is a prime example of economic mismanagement and reveals a link between domestic debt, financial repression, and external vulnerability (Reinhart and Santos 2015). The collapse in oil prices between 2013 and 2019 explains the 60% collapse of Venezuelan GDP per capita (left panel of Figure 2.A1), while the boom from the year 2000 correlates with the substantial increase in debt (right panel of the same figure). The seigniorage system's failure led to the emergence of hyperinflation in 2013, which was about one million percent in 2018 (see Restuccia 2019). Consequently, between 2014 and 2018, the country's poverty rates doubled, from 48% to 87% (España and Ponce 2018). Food shortages since

[9] The original ISIC Rev 4 classification includes 10 major groups (MG). The original ISCO-08 classification includes 21 sections.

[10] Admittedly, US occupation skills will differ from those in Peru and Venezuela. However, it is reasonable to assume that this difference goes in the same direction for occupations in both countries. There are no reasons to believe that the suddenness of the influx might have led to composition changes in the host country in the short run. Moreover, because we are not interested in the actual value of these indices but, instead, in their change, these differences in levels in both countries relative to the USA are not a primary concern.

2015 led many to shift their consumption basked to only price-controlled food items, which also became increasingly scarce. Comparable shortages in medicines caused an increase in their prices and these items thus became unaffordable for most Venezuelans.[11] Community-based support organizations established by the government did not improve the situation (IACHR 2017). The unprecedented drop in the quality of life explains why Venezuela consistently topped Bloomberg's Economic Misery Index between 2014 and 2018.

Political and social turmoil accompanied this economic collapse. There was government repression of public demonstrators, which resulted in thousands of arbitrary arrests and prosecutions by military courts (HRW 2018). The detainees were subject to torture, abuses with sexual violence against women, as documented by the Inter-American Commission for the Human Rights (IACHR 2017). The insecurity caused by a crime crisis further eroded the social fabric from 2014, as Venezuela turned into the "second most violent country in the world" according to the Venezuelan Observatory of Violence.[12] Homicides became the primary cause of death among adolescents and young people, particularly the poorest in society (IACHR 2017).

These factors led to an unprecedented emigration of Venezuelans who arrived in Peru in 2016, which we take as the beginning of the treatment (Figure 2.A3).[13] Within a brief period, Peru went from being one of the Latin American countries least used to foreign workers (Torales et al. 2003) to the second-largest recipient of Venezuelans worldwide (see Figure 2.A5 based on UNHCR 2020). Since the beginning of this influx until 2019, approximately 830,000 Venezuelans entered the country representing 2.5% of Peru's population. According to the Population Census of 2017, 85% of these settled in five regions. The most important receptor was the metropolitan area of Lima and Callao, the capital city and the epicentre of economic activity in Peru. This region comprises 38% of the total employment and 48% of the total GDP. Tumbes, bordering Ecuador, is their entry post in north of Peru. La Libertad and Arequipa are respectively the second and third recipients of Venezuelan influx, with the latter being the second largest contributor to Peru's GDP. In contrast, Cusco, where the Venezuelans have mostly integrated, is a tourist-centred city and generates mostly services-oriented jobs.[14]

The main emigration route for the Venezuelans (depicted as solid lines in Figure 3.A1) begins in mountainous West Venezuela (brown areas in that map), according toUNHCR (2018b) and OIM (2018b; 2018c; 2019).[15] Consistent with this, our data also reveal that 93% of immigrants reach Peru through Colombia or Ecuador, with Tumbes (in the northern end of Peru) a common entry point (Table 2.A2). However, this migration path is not free of risks for these immigrants, as 40% of the violent and other incidents experienced occurred during this transit (with the most frequent being robbery, physical assault, and death threats) (UNHCR 2019b). Because this route almost exclusively involves land, our data also confirms that most of the trips are made by bus (coinciding with UNHCR 2019b report). This contrasts with the multiple transportation modes that also characterize the Syrian journeys towards Europe (UNHCR 2018a).

In order to help incorporate this unprecedented inflow of Venezuelans into the formal sector, the Peruvian government adopted a series of measures which were developed in three phases (see Table 2.A1 for a further description). The first one, between January 2017 and January 2018, established the Temporary Permit of Permanence (PTP). This provided Venezuelans with one year of legal permanence in Peru and enabled them to work as employees or as self-employed.[16] However, long waiting times to obtain this per-

[11] As of 2016, there was an 82.8% deficiency of calories in the basic food basket, 9 out of 10 homes were food insecure and 70% had lost more than 8.7 Kg of body weight. As of 2017 the estimated shortage of medicines was around 90% an children mortality increased 30% between 2015 and 2016 (IACHR 2017).

[12] According to this source, homicide rates reached 91.8 for every 100,000 inhabitants. According to the Citizen Council for Security and Criminal Justice, Venezuela's capital, Caracas, is considered the most violent in the world, along with seven other Venezuelan cities (IACHR 2017).

[13] Data publicly released by Peru's immigration authority prior to 2016 refers only to those immigrants over 18 with a working visa.

[14] Peru is the largest recipient of Venezuelan asylum seekers, comprising 50% of these requests (R4V 2020a). As of 2019, in Cusco and Arequipa asylum-seeking applications from the Venezuelan population increased sharply (INEI 2019a).

[15] The rugged terrain of the jungles in southern Venezuela makes it an impenetrable territory but one that is also linked to illegal activities, mainly drug trafficking (Van Dun 2016).

[16] By the end of January 2018 operationalized the steps for Venezuelans to apply for the Special Resident immigration status defined

mit resulted in only one in four Venezuelans being PTP holders as of 2018 (see Table 2.A2). Specifically, the applicant had to gather documentation from different government agencies in Peru (including the INTERPOL international exchange token) and wait an additional six months for the immigration Authority to validate the PTP document (see Blouin 2019)[17].

The next two phases then set restrictions for Venezuelans' assimilation into the Peruvian formal labour market. The second phase began in the last week of August 2018, which abolished the free entry of Venezuelans. Two norms eliminated the PTP and made the passport mandatory to enter the country from September 2018 onwards.[18] This requirement constitutes an exceptionally prohibitive measure, as passport processing fees oscillated between 2,000 and 5,000 USD within Venezuela, where the monthly minimum wage does not exceed 5 USD (Blouin et al. 2019). Indeed, the two most prominent entrance peaks in 2018 are explained by Venezuelans trying to circumvent this requirement (see Figure 2.A3). The third and final phase began in mid-2019 when, in addition to a valid passport, Venezuelans were also required to possess a Humanitarian Visa to enter Peru. It could only be obtained at the Peruvian Consulate in Venezuela, Colombia or Ecuador. Nonetheless, the vast number of Visa requests in these locations made it unfeasible for the consular services to process them within a reasonably short time window (UNHCR 2019b). On the 14th of June 2019, the day before this rule came into place, an influx of 8,000 Venezuelans took place (UNHCR 2019c, see Figure 2.A3). Subsequently, the influx fell to a daily average of approximately 1,300 Venezuelans, way below the average of 3,000 entrants experienced before June (Migraciones 2018).[19] It is likely that these measures also led to an increase in the number of those who irregularly entered Peru, although there are no official figures.[20]

2.4.2 The Venezuelan integration into the host economy

A steady per capita GDP growth (see Figure 2.A6) and a high employment rate, as shown in the upper panel in Figure 2.1, characterises the Peruvian economy during this period. The same graph shows that a fundamental difference with the labour market of developed economies, where previous studies of the impact of immigration have focused (e.g. Card 1990; Hunt 1992; Carrington and de Lima 1996), lies in Peru's sizeable informal sector. Even though it has progressively reduced, more than 60% of the employment remains informal. This share is significantly larger than the one that characterizes the Turkish labour market (ILO 2017, 2019), which experienced a comparable inflow of Syrian immigrants (see section 2.2). Average hourly wages in this sector have been less than half than the prevailing in the formal sector (lower right panel of Figure 2.1). This latter reflects differences in labour productivity between sectors and their ability to absorb sector migrants (mainly the informal, as shown below). Consequently, in our regression modelling, we analyze the impacts of the Exodus in both sectors separately.

Despite sharing the same (Spanish) language and close historical ties, Peruvians' reactions to Venezuelans who arrived at their country during the Exodus have not been entirely favourable. As analysed in more detail in the next chapter, most Peruvians do not welcome the arrival of more Venezuelans into the country and believe that these immigrants negatively affect their wages and labour market opportunities. Follow up surveys confirm that Peruvians' attitudes towards Venezuelans immigrants deteriorated in the following

in DS Nº 007-2017-IN (see Table 2.A1). This provided them with an immigration card that proved regular residence in Peru.

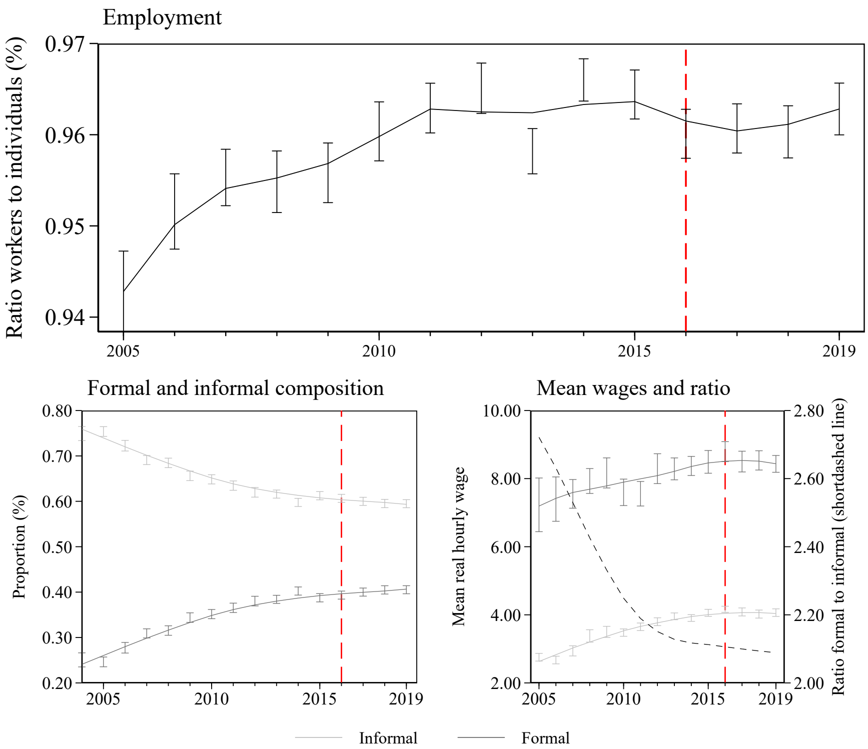
[17] In May 2018, Peru's government gave the opportunity to those Venezuelans who were waiting for their PTP to be processed to obtain the Extraordinary Working Permit. This was two-month working permit, valid until the granting of the PTP.

[18] The first of these norms was given by a Supreme Decree (DS) on 19th August 2018, which tightened their deadline to enter Peru and request their PTP from 31th of December 2018 (as declared in the DS of 23th January 2018) to 31th of October. The second DS was issued five days later.

[19] At the same time, a new pre-admission process was established which implied that, after June 2019, application for refugee status (which would obviate the need for the visa) could only be done once Venezuelans reached the border of the country via an interview (IDEHPUCP et al. 2020).

[20] According to ENPOVE, 90% of Venezuelans in Tumbes regard Peru as their final destination. Data from OIM (2020) allow us to assume that the 63% of Venezuelans who were denied entrance via Tumbes would have tried to enter anyway. The growing number of Venezuelans who subsequently entered Peru did not have a passport (OIM 2018b,c, 2019) provides further support for this.

Figure 2.1 – Employment rate, informality and wages, 2005-2019



Note: Sample restricted to only employed individuals between 18 and 65 years. Vertical dashed lines denote the beginning of the treatme
Source: Author's calculations using ENAHO 2005-2019 data.

years (see PUCP 2020). This occurs despite sharing the same language and their close cultural ties. In fact, in contrast to other massive immigration episodes studied by Card (1990), Carrington and de Lima (1996) or Foged and Peri (2015), Venezuelan immigrants are more educated than the natives. Specifically, around 60% of Venezuelan workers in the treated areas have higher education, whereas a Peruvian worker's most common educational level is secondary (see Table 2.1). Nevertheless, Peruvians believe that Venezuelans are uneducated (see table 1 in the next chapter). Also, while the gender composition is similar among natives and immigrants, there is a sizeable age difference. The Venezuelan population is younger, with almost 70% aged between 18 and 35.

Venezuelans almost exclusively work in the informal labour market (bottom panel in Table 2.1).[21] This coincides with what the descriptive studies of Koechlin et al. (2019), Fuentes (2019), as well as the OIM's Displacement Tracking Matrices (OIM 2018a; 2018b; 2018c; 2019) report, and mirrors the experience of the Syrian immigrants in the Turkish labour market (Akgündüz et al. 2015; Ceritoglu et al. 2015). However, other factors besides the large size of the informal sector in Peru explain this. One is that the long waiting times to obtain the PTP (described above) clashed with their need to immediately earn a minimum income to support themselves and to send remittances.[22] Another reason lies in Peru's tight legislation for hiring foreigners (Geronimi 2004), which offered little incentives to employers.[23] Also, the deficient law enforcement in Peru (Viollaz 2019) allowed natives to hire as informal workers even those Venezuelans with valid

[21] In Table 2.1 we define as informal those workers with a health insurance status. This is different from what we use in the main estimations (affiliation to the pension system) because data in ENPOVE only includes the insurance variable. This might slightly over-estimate the percentage of informal workers, since in ENAHO dataset (which includes both affiliation to pension and to health insurance) 76% of the workers without health insurance are not affiliated to a pension system.

[22] ENPOVE data reveals that 70% of Venezuelan workers send remittances and that every transfer represents between 5% and 25% of their average labour market income.

[23] The law for the hiring of foreign workers (Legislative Decree No. 689 of 1991 and Supreme Decree No. 014-92-TR of 1992) declares that they cannot represent more than 20% of the total number of workers and their wages cannot exceed 30% of the total payroll in the firm. In addition, the Supreme Decree No. 179-2004-EF defines that foreigners without permanent residence in Peru are subject to a special income tax regime of 30%.

work permits (Blouin 2019; IDEHPUCP et al. 2020).[24] This latter factor also encouraged Venezuelans engage in self-employment in the lower productivity occupations (OIM 2018d).

Table 2.1 – Demographic characteristics of Peruvians and Venezuelan immigrants in treated areas (%), 2018

	Peruvians	**Venezuelans**					
		Arequipa	Cuzco	La Libertad	Lima y Callao	Tumbes	*Total*
Education level							
No Level	1.03	0.00	0.00	0.18	0.07	0.00	0.08
Primary	11.67	4.61	8.90	14.32	9.33	5.61	9.37
Secondary	44.57	29.01	28.31	24.43	30.41	60.30	30.33
Technical	19.21	16.85	20.44	20.90	20.45	13.83	20.41
College	20.91	48.40	41.74	38.58	39.02	19.69	39.09
Postgraduate	2.60	1.12	0.61	1.59	0.71	0.56	0.73
Gender							
Female	44.04	43.44	39.93	41.74	42.67	48.62	42.66
Male	55.96	56.56	60.07	58.26	57.33	51.38	57.34
Age group (years)							
18-25	16.20	27.97	34.57	31.01	32.07	34.89	32.02
26-35	26.53	44.96	45.07	42.48	41.30	42.98	41.38
36-45	26.07	19.77	14.15	17.25	18.27	14.74	18.24
46-55	18.41	6.09	5.61	6.61	7.13	5.98	7.10
56-65	12.80	1.21	0.60	2.64	1.24	1.41	1.27
Does not have HI							
Percentage	30.20	95.20	93.98	96.06	94.16	96.65	94.21
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Employed in Venezuela							
Percentage		85.22	81.11	81.37	80.94	69.62	80.97
		(0.01)	(0.02)	(0.02)	(0.01)	(0.04)	(0.01)

Notes: Sample restricted to only those employed between 18 and 65 years. Estimates for Peruvians considering only the treated areas. HI stands for Health Insurance. *Source*: Author's calculations using ENAHO 2018 and ENPOVE data.

In order to explore if their informal-worker status affects their ability to use their higher educational background in the host country, we calculate the occupational distribution and job complexity measures for Venezuelans before and after they migrate, extending the idea in Dustmann et al. (2013) and Foged and Peri (2015). Two key results emerge (Table 2.2). Firstly, Venezuelan immigrants are over-represented in low productivity occupations (e.g., sales and elementary occupations) and have a lower mean complexity index than the natives. This also holds for the earlier phase of immigrants. Secondly, Venezuelans experience a striking downgrading after arriving in Peru. On the one hand, 40% of Venezuelans had managerial and technical jobs in their home country, and only 10% performed this same type of work in Peru. On the other hand, the share of those who perform elementary occupations increases from 5% pre-emigration to 22% post-emigration. The lower panel shows that their mean complexity index is less than half for the jobs held in Venezuela. The shift from occupations intensive in communication and cognitive skills towards more manual jobs mostly explains this result. These regularities also hold for immigrants with secondary and higher education while those who are lower educated do not experience such a downgrade (Table 2.3).[25]

This downgrading occurs even in the absence of the language barrier, which constitutes a hurdle for Syrian immigrants in their host markets (Akgündüz et al. 2018; Konle-Seidl and Bolits 2016). In turn, it is similar to what migrants from outside the European Union experience in Germany, United Kingdom and the United States (Dustmann et al. 2013, 2016), where the immigration process was instead gradual and lengthy.[26]

[24] IDEHPUCP et al. (2020) reports that a large number of PTP holders had this document rejected by potential employers, who would "only accept foreign personnel if they have the immigration card", contravening immigration rules.

[25] The previous evidence of immigrants working in low occupational categories and their downgrading upon arrival also holds within

Table 2.2 – Occupational distribution and skill level, 2018

	Peruvians	**Venezuelans**			
		Earlier		Recent	
		Pre	Post	Pre	Post
Occupation category (%)					
Managers and professional worker	11.29	18.06	3.97	17.07	2.91
Technical workers	17.18	23.04	7.93	21.40	5.36
Clerical services sales workers	29.10	37.90	45.88	39.21	47.80
Craft and trades workers	8.96	8.05	13.20	6.75	14.26
Machine operators	11.78	7.68	7.49	9.38	6.59
Elementary occupations	21.69	5.27	21.53	6.19	23.08
Complexity score					
Communication skill	0.30	0.39	0.28	0.38	0.28
Cognoscitive skill	0.28	0.36	0.24	0.35	0.23
Manual skill	0.54	0.47	0.53	0.50	0.53
Complexity index	3.94	5.82	1.96	4.78	1.88

Notes: Sample restricted to only those employed between 18 and 65 years. Pre and post refer to the job the Venezuelan immigrant had in Venezuela (before migrating) and in Peru, respectively. Score measures follow Ottaviano et al. (2013). *Source*: Author's calculations using ENAHO 2018 data, ENPOVE data, O*NET 25.1 Database and crosswalks by Hardy et al. (2018).

Table 2.3 – Occupational distribution and skill level by education level, 2018

	Primary level					**Secondary level**					**Higher level**				
	Peruv.	*Venezuelans*				*Peruv.*	*Venezuelans*				*Peruv.*	*Venezuelans*			
		Earlier		Recent			Earlier		Recent			Earlier		Recent	
		Pre	Post	Pre	Post		Pre	Post	Pre	Post		Pre	Post	Pre	Post
Occupation category (%)															
Managers and profess	0.03	5.89	6.43	2.88	0.30	0.83	1.49	1.40	1.49	0.89	24.94	25.40	4.76	27.07	4.43
Technical workers	35.61	14.67	1.75	9.42	1.22	15.72	13.98	2.88	17.13	0.99	22.27	27.23	10.59	25.51	8.40
Clerical svs sales workers	19.60	33.56	47.82	42.10	43.53	26.76	48.47	44.06	44.45	43.49	29.43	34.41	46.42	36.16	50.83
Craft and trades workers	7.05	19.26	17.39	12.33	21.17	10.95	14.38	19.38	11.88	17.86	5.05	4.60	10.28	3.30	11.15
Machine operators	7.24	11.18	12.10	20.14	7.26	16.10	12.86	8.99	13.93	9.65	7.49	5.41	6.42	5.34	4.85
Elementary occupations	30.47	15.45	14.51	13.13	26.52	29.63	8.82	23.29	11.12	27.12	10.82	2.95	21.52	2.62	20.33
Complexity															
Communication skill	0.20	0.24	0.28	0.24	0.23	0.21	0.28	0.23	0.29	0.25	0.41	0.45	0.31	0.45	0.30
Cognoscitive skill	0.23	0.27	0.25	0.24	0.20	0.22	0.26	0.20	0.26	0.20	0.36	0.40	0.25	0.41	0.24
Manual skill	0.70	0.61	0.54	0.65	0.58	0.65	0.58	0.59	0.61	0.57	0.43	0.41	0.50	0.42	0.50
Complexity index	0.94	1.69	1.85	1.50	1.32	1.30	2.57	1.16	2.01	1.37	6.61	7.49	2.29	6.75	2.25

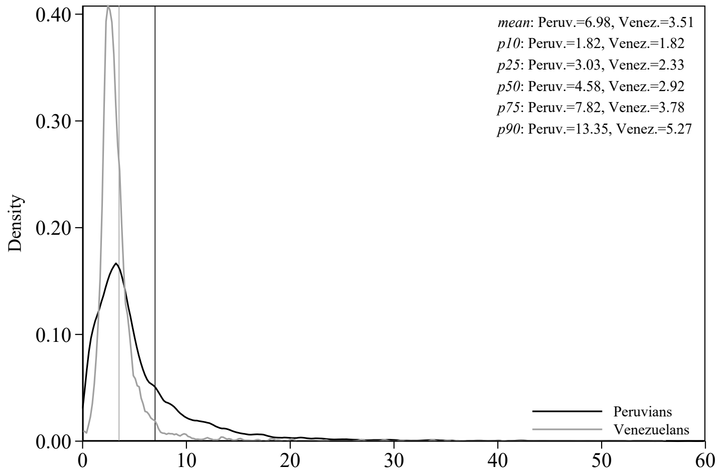
Notes: Sample restricted to only those employed between 18 and 65 years. Pre and post refer to the job the Venezuelan immigrant had in Venezuela (before migrating) and in Peru, respectively. Lower education level includes those with primary and less; intermediate, those with secondary; higher, those with at least some college, technical education and postgraduates. *Source*: Author's calculations using ENAHO 2018 data, ENPOVE data, O*NET 25.1 Database and crosswalks by Hardy et al. (2018).

The absorption of the Venezuelans into the informal sector and consequently into low productivity activities results in a tighter and leftward-shifted hourly wage distribution than that of the natives (Figure 2.2). Their hourly wages at the mean and at different percentiles are lower, and the wage gap increases as we move up the income distribution. Venezuelan wages at the 1st decile are the same as that of Peruvians; in contrast, their hourly wages at the 9th decile (slightly more than 1 GBP in 2018 prices) are less than half that of the natives. In fact, only 60% of Venezuelan workers earn a monthly wage higher than Peru's minimum legal wage as of 2018 (around 170 pounds sterling in 2018 prices). This occurs despite the fact that most migrants had prior work experience in Venezuela (see bottom of Table 2.1). Almost 90% of these work in small-sized firms, characterized by low productivity levels in Peru (Távara et al. 2014) and around 82% work more than the maximum legal number of weekly working hours (Koechlin et al. 2019).

the different treated areas (Table 2.A3).

[26]The evidence of downgrading confirms the suitability of our quasi-experimental approach for analyzing the impacts of the Exodus compared to those based on the national-skill approach (see section 2.2), as these assume that the location of the immigrant in the skills distribution corresponds to their education level (Dustmann et al. 2013).

Figure 2.2 – Hourly wages distribution, 2018



Note: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Vertical lines denote the averages of the distributions. *Source*: Author's calculations using ENAHO 2018 and ENPOVE data.

2.5 Identification strategy

2.5.1 Econometric methods

The analysis of the 2016 Venezuelan Exodus to Peru requires methods based on for quasi-experimental designs. As reported in the previous section, despite its massive size, the sudden influx did not affect in the short run Peru's economic structural characteristics which could also impact the outcomes of interest. These include regional institutional quality, technological progress or its labour demand composition. Furthermore, the origin of this Exodus lies in Venezuela's economic and political factors, which are not related to Peru's economic growth. Due to the availability of repeated cross-sections (RCS) for ten years before and fours years after the treatment, we can exploit the department-level variation of this sharp treatment (see Figure 2.A3 and Figure 3.A2). We are interested in estimating the Average Treatment Effect of the Treated (ATET). In the case of the wages, it tells us how much the average wages changed in the treated departments after being exposed to the influx, relative to a counterfactual situation where these never experienced such shock (see Lee 2016; Frölich and Sperlich 2019). With only one treated department ($s = 1$) and one control ($s = 0$) and one year before ($t = 0$) and after ($t = 1$) the impact, the unconditional ATET in terms of the potential outcomes, where super-indices 1 and 0 denote the potential outcomes with and without treatment, is given by

$$ATET = E\left(Y^1 - Y^0 | s = 1, t = 1\right) = E\left(Y^1 | s = 1, t = 1\right) - E\left(Y^0 | s = 1, t = 1\right) \tag{2.1}$$

The unobservability of the counterfactual (second term) requires its estimation under some identifying assumptions. To provide a more secure basis for informing policy, we apply different methods for its estimation (Abadie and Cattaneo 2018; Samartsidis et al. 2019; Strumpf et al. 2017). Nonetheless, we take as our main estimator that which arises from a two-stage DiD method. As we explain below, this is better suited for analysing the effects of the Venezuelan Exodus.

2.5.1.1 Estimation at individual level

The DiD method estimates the ATET as the difference in average wages between treated and control groups in $t = 1$ after netting out differences in $t = 0$. Conditional on a vector of covariates $\mathbf{X}$ which is unaffected by the treatment, in the simple case this corresponds in the simple case to

$$ATET^{DiD}(\mathbf{X}) = [E(Y|s=1,t=1,\mathbf{X}) - E(Y|s=1,t=0,\mathbf{X})] - [E(Y|s=0,t=1,\mathbf{X}) - E(Y|s=0,t=0,\mathbf{X})] \quad (2.2)$$

where the potential outcomes are now replaced by their observed averages. The counterfactual is identified by adjusting the observed (conditional) average in the post-treatment of the control department for the differences in averages in the pre-treatment.

A first concern is that this strategy is valid only if the parallel trend assumption (PTA) holds, i.e. in the absence of the Exodus, the (conditional) evolution of average outcomes would have been the same in treated and the control departments. This is essentially untestable and cannot be taken for granted (Lee 2016; Angrist and Pischke 2009; Frölich and Sperlich 2019). Two additional concerns become more evident if we use the equivalent linear model to reflect our case where different departments are treated

$$Y_{st} = \alpha_s + \lambda_t + \beta D_{st} + \mathbf{X}_{st}^{'}\gamma + V_{st}; s = 1,...,S; t = 1,...,T \quad (2.3)$$

where s and t indexes the departments and years. The treatment indicator is $D_{st} = 1(treated_s = 1) \times 1(year \geq 2016)$. I.e., this is 1 for the treated departments from 2016 onwards and 0 otherwise, so $\hat{\beta}$ provides an estimate of the ATET. The terms α_s and λ_t represent a full set of unobserved department and time effects, respectively, which capture any correlation between the policy assignment and time-invariant factors and secular time trends. The (column) vector $\mathbf{X}_{st}$ includes controls that vary along the department-year level and V_{st} is an unobserved random variable with zero mean assumed uncorrelated with $\mathbf{X}_{st}$.[27] Important efficiency gains can be grasped with a multilevel model (Frölich and Sperlich 2019), which for a random draw corresponds to (with the variables in lower case to emphasize that data is now at the individual level)

$$y_{ist} = \alpha_s + \lambda_t + \beta D_{st} + \mathbf{x}_{st}^{'}\gamma + \mathbf{z}_{ist}^{'}\delta_{st} + \varepsilon_{ist} \text{ where } \varepsilon_{ist} = v_{st} + u_{ist}; s = 1,...,S; t = 1,...,T; i = 1,...,M_{st} \quad (2.4)$$

where $\mathbf{z}_{ist}$ is a (column) vector of individual-specific covariates, ε_{ist} is the unobserved individual-specific error comprised now of two random variables, v_{st} and u_{ist} are unobserved department-year and idiosyncratic components with zero mean and which is uncorrelated with explanatory variables and independent of each other. We can think of v_{st} as shocks shared by individuals within the same department and year, e.g. department-level business cycles. M_{st} indexes the number of observations in every department-year cell.

Direct estimation of Equation 2.4, as is customarily done in the literature, leads to two types of problems. The first is related to the consistency of $\hat{\beta}$. Since the level at which the sampling of data for DiD estimation takes place is departmental(region)-year, the asymptotic analysis requires taking $M_{st} \to \infty$ while keeping S and T fixed. It turns out that under "large M_{st}" asymptotics, the presence of v_{st} leads to meaningful negative consequences: $\hat{\beta}$ from Equation 2.4 conflates both the effect of the treatment at department-year level (which we are after) and the differences in the evolution of the department-year shocks (Angrist and Pischke 2009; Frölich and Sperlich 2019). Surprisingly, proper handling of the v_{st} component has been largely ignored in the literature (Bertrand et al. 2002).[28] We address this problem by applying a two-stage DiD estimator (2S-DiD) process described in Hansen (2007) and Imbens and Wooldridge (2014), which our data renders feasible as the average cell size is between 217,685 and 337,842 observations. In the first stage, we apply the Pooled OLS estimator with the individual-level data $y_{ist} = \xi_{st} + \mathbf{z}_{ist}^{'}\delta + u_{ist}; s = 1,...,S, t = 1,...,T, i = 1,...,M_{st}$ to estimate the department-by-year fixed effects, $\hat{\xi}_{st}$. In the second stage

[27] Henceforth we do not specify the dimensions of the matrices and vectors here. However, we assume these are conformable for the operations involved.

[28] In order to see this more clearly, ignoring $\mathbf{x}_{st}$ and $\mathbf{z}_{ist}$ in Equation 2.4 and taking only two years and one treated and one control, the estimator of the ATET is $\hat{\beta} = \beta + [(\varepsilon_{11} - \varepsilon_{10}) - (\varepsilon_{01} - \varepsilon_{00})]$. As we average larger and larger samples within departments-years, a weak LLN assures that plim$\hat{\beta}$ eliminates the influence of u_{ist} since $E(u_{ist}) = 0$ but not of v_{st}, yielding plim$\hat{\beta} = \beta + [(v_{11} - v_{10}) - (v_{01} - v_{00})]$ as $M_{st} \to \infty$.

we estimate the causal parameter of interest using a fixed effects (FE) estimator on the following model

$$\hat{\xi}_{st} = \alpha_s + \lambda_t + \beta D_{st} + \mathbf{x}_{st}^{'}\gamma + v_{st}; s = 1, ..., S; t = 1, ..., T \tag{2.5}$$

D_{st} and ξ_{st} are partialled out from the two-way fixed effects (α_s and λ_t) and $\mathbf{x}_{st}$ by a standard Frisch-Waugh-Lovell argument.[29] We also estimate Equation 2.4 including 4 lags and 4 leads of the treatment to exploit the notion of Granger predictability (Angrist and Pischke 2009). This means that βD_{st} is replaced by $\sum_{t=\tau_0}^{2014} \beta_t D_{st} + \sum_{t=2016}^{2019} \beta_t D_{st}$, where $D_{s\tau} = 1\,(treated_s = 1) \times 1\,(year = \tau)$. Joint significance of the $\hat{\beta}$s in the first term *suggests* that PTA is not plausible while the estimated βs from the second term provide information about effects' time-heterogeneity relative to the last pre-treatment year (2015).

The second type of econometric problem concerns the estimation of the standard errors (SE). Typically, studies (e.g., Balkan and Tumen 2016; Akgündüz et al. 2015, 2018; Mora et al. 2019) only control for the arbitrary within-departments correlation of the idiosyncratic errors u_{ist} in Equation 2.4 (i.e., $E\left(u_{ist}u_{js\tau}\right) = \sigma_v^2, i \neq j$). Their estimates of the variance-covariance matrix using the Liang and Zeger (1986) estimator (cluster-robust variance estimate, henceforth CRVE) ignore the policy autocorrelation problem induced by the serial correlation of v_{st} (so that $E\left(\varepsilon_{ist}\varepsilon_{js(t-\ell)}\right) \neq 0, i \neq j, \ell > 0$, see Imbens and Wooldridge 2014; Hansen 2007), leading to a sizeable underestimation of the actual SEs. As Bertrand et al. (2002) demonstrate, this strategy results in a 44% probability of rejecting a true null hypothesis of no effect using a nominal 5% level test. Even if we rule out the cross-sectional correlation of v_{st}, the validity of CRVE requires that $S \to \infty$. Applications reviewed in section 2.2 are instead characterized by a "small-S", which results in a large over-rejection of the null hypothesis of no effect (Cameron et al. 2008; Cameron and Miller 2015; Lee 2016; Wooldridge 2003). This over-rejection further increases as the cluster heterogeneity increases (Carter et al. 2017; Lee and Steigerwald 2018)[30] and as the number of treated departments S_1 decreases (Mackinnon and Webb 2016; Bell and McCaffrey 2002). Moreover, Brewer et al. (2018) note that in typical DiD settings, the real challenge for inference is the power associated with small-S. Their simulations show that this makes it harder to detect effects of realistic magnitude even with a correctly sized test.

In view of these concerns, we assess the statistical significance of estimates from Equation 2.5 using different methods. In order to reduce the degree of over-rejection due to the small-S problem, we apply a finite-sample correction to the residuals to estimate the CRVE, given by $\sqrt{\frac{S}{S-1}}\sqrt{\frac{N-1}{N-K}}$ (where K is the number of right-hand side variables), and evaluate the Wald statistic using the critical values from a T distribution with $S-1$ degrees of freedom (DoF). Additionally, we carry out inference using the Bias-Corrected CRVEs (CRVE2 and CRVE3) of Bell and McCaffrey (2002). These variance-covariance estimators are built upon the heteroskedasticity-consistent HC2 and HC3 variance estimators from MacKinnon and White (1985) and result in Wald statistics with smaller size distortions than those typically applied in the literature (Cameron et al. 2008; Imbens and Kolesar 2012). We report the effective number of clusters S^* from Carter et al. (2017) which adjusts (downward) the observed S to reflect the degree of sample heterogeneity. A sizeable difference between S and S^* suggests that inference should be based on the p-values from the wild cluster bootstrap procedure (Mackinnon and Webb 2016) and also allow us to evaluate the reliability of Wald statistics based on CRVE.[31] For the wild cluster bootstrap, we use both the Rademacher's and the Webb's (2013) distribution. The latter is found to be more appropriate with less than 10 (effective)

[29] Note how in the first step we have imposed homogeneity of the slopes, $\boldsymbol{\delta}_{st} = \boldsymbol{\delta}$ and in the second step all the uncertainty comes through v_{st} because our large M_{gt} ignores the estimation error of ξ_{st}. This partialling out is done for the purpose of avoiding the rank-deficiency problem (see Cameron et al. 2008) and to avoid ending up with intractably large matrices for the calculation of cluster robust SE estimators.

[30] Heterogeneity in this case is defined as a measure that captures violation of the following assumptions: identical number of observations for every region s and identical variance-covariances matrices for the covariates and for the error across s (Carter et al. 2017; Lee and Steigerwald 2018). Note how the estimating approach based on Equation 2.5 satisfies by construction the first assumption, unlike the more typical analyzes undertaken which directly estimates Equation 2.4.

[31] In our application, S^* systematically suggests substantial cluster heterogeneity which rules out the applicability of the alternative inference approach by Donald and Lang (2007). Mackinnon and Webb (2016), in a similar setting as ours, suggest that this is a rather conservative method because it tends to under-reject the null hypothesis.

clusters (see Cameron et al. 2008 for details). We also calculate the p-values that result from evaluating the CRVE2-based Wald statistic using the data-determined DoF from Imbens and Kolesar (2012, henceforth, IK-DoF)[32]. Simulations therein demonstrate that this combination results in slight under-rejection of the null hypothesis of no effect.

As a direct way to deal with the plausible serial correlation of v_{st} in Equation 2.5, we use Hansen's (2007) Feasible GLS (FGLS) estimator assuming that v_{st} follows a (stationary) $AR(1)$ and $AR(2)$ process. A substantial improvement compared to alternative FGLS approaches for panel data (e.g. Kiefer 1980) is that its transformation matrix relies on an iteratively bias-corrected (BC) estimator of the (finite-dimensional) vector $\boldsymbol{\rho}$ that characterizes the nature of the AR process.[33] As we show below, it provides more efficient inference than the usual FE. Indeed, Brewer et al. (2018) simulations show that this bias-corrected FGLS (FGLS-BC) estimator and the CRVE improve the power of the test considerably while providing correctly sized tests. It is also robust to misspecification of the error process, even with small S and with fewer periods than in our application.[34] We follow Brewer et al.'s (2018) concluding recommendation and take this as our preferred estimator in our study.[35]

As shown in section 2.6, this approach for the effects of v_{st} leads to different point estimates and standard errors compared to the quasi-experimental studies reviewed above. It suggests that some of their results reported in the existing literature might be driven by their omission from the empirical analyzes undertaken in these studies.

2.5.1.2 Estimation at aggregated level

In order to assess the robustness of our results from our main estimation strategy used in the previous subsection, we also estimate the ATET using the Synthetic Control Method (SCM, Abadie et al. 2010, 2015; Abadie 2020). Compared to the DiD, SCM estimates the ATET for each of the $i = 1,...,5$ treated departments in the post treatment year $t > T_0, t = 12,...,15$ (the beginning of the treatment is 2016 and so $T_0 = 11$) by

$$ATET_{it}^{SCM} = Y_{it}^{I} - Y_{it}^{N} = Y_{it} - Y_{it}^{N} \tag{2.6}$$

where Y_{jt}^{I} is the potential outcome for department i after being exposed to the intervention, which is always observed. Letting Y_{jt}^{N} be the observed outcomes of the $24-5 \equiv J$ departments consigned to the control group known as the donor pool indexed by $j = 5+1,...,5+J$, the counterfactual for each i is identified by a weighted average of Y_{jt}^{N}s

$$Y_{it}^{N} = \sum_{j=5+1}^{5+J} \omega_j^* Y_{jt}^{N} \tag{2.7}$$

The optimal synthetic control weights, the ω_j^*s, minimize a quadratic distance between units in the donor pool and each i only in terms of pre-treatment observed attributes given by

[32] This measure of DoF is in turn based on Bell and McCaffrey's (2002) DoF, which approximates the finite sample distribution of the CRVE2-based Wald test via a t-distribution with K DoF. This K is a function of the design matrix which equalizes the first two moments of an expression involving the population variance and the estimated variance to the first two moments of a χ_K^2distribution. Imbens and Kolesar (2012) extend this approach to exploit the implicit equicorrelated errors structure present in the clustering case. See their paper for further details.

[33] This bias in Kiefer's approach is induced because under FE, $\boldsymbol{\rho}$ is calculated from residuals $\hat{v}_{st} \approx v_{st} - \bar{v}_s$ which with T fixed do not behave like the underlying errors. The bias can be substantial with even a moderate T (Hansen 2007).

[34] Additionally, under the maintained hypothesis of serially correlated errors, FGLS is not affected by the negative relationship between power of the test and panel length because the first-stage transformation removes the serial correlation from the error terms, unlike the case of FE estimation.

[35] Asymptotically, failure of the $AR(p)$ assumption to construct the FGLS estimator does not affect its consistency. In fact, a weighting matrix based on an incorrect parametrization of the serial correlation process will often still be closer to the optimal GLS weighting matrix than the identity matrix used by standard OLS (Brewer et al. 2018; Hansen 2007). Even modest efficiency gains can be beneficial (Cameron and Miller 2015)

$$\mathbf{W} \equiv (\omega_{5+1}, ..., \omega_{5+J}) = \text{argmin}\,(\mathbf{X}_i - \mathbf{X}_0\mathbf{W})'\,\mathbf{V}\,(\mathbf{X}_i - \mathbf{X}_0\mathbf{W}) \text{ s.t. } \omega_j \geq 0 \text{ and } \sum_{j=5+1}^{5+J} \omega_j = 1 \tag{2.8}$$

where $\mathbf{X}_i$ and $\mathbf{X}_0$ is a column vector and a J-columns matrix, respectively, of pre-treatment characteristics for i and for each of the js in the donor pool, $\mathbf{V}$ is a diagonal symmetric positive semidefinite matrix which reflects the contribution of each of those characteristics in the counterfactual and $\mathbf{W}$ is the vector of weights.[36] The two restrictions on $\mathbf{W}$ ensure that the counterfactual is constructed using only a subset of untreated departments which best resemble the pre-treatment economic characteristics of the treated areas i (i.e. is sparse) and that is based on interpolation of these units.[37]

Compared to the DiD, SCM estimates the causal effect of interest even when Y_{jt}^N follows a more general Latent Factor Model, with a set of factor loadings and common factors replacing the unobserved effects in Equation 2.5 (see Bai 2009).[38] However, for this to be possible, the convex hull condition must hold. This implies that the ω_j^*s allow the weighted average of the outcomes of units in the donor pool to fit perfectly the outcome of each of the i in all pre-treatment periods (see Appendix B in Abadie et al. 2010). In practice, this condition is rarely met (Powell 2018) and Ferman and Pinto (2021) show that in this case the estimated ω_j^* are biased. Consequently, we improve upon the existing empirical evidence for the Venezuelan Exodus by also applying the modified synthetic control estimator introduced by Ferman and Pinto (2021). This demeans the data using information from the pre-treatment period, $\tilde{Y}_{ht} = Y_{ht} - \left(\frac{1}{T_0}\right)\sum_{t=1}^{T_0} Y_{ht}$ with $h = i, 5+1, ..., 5+J$. The authors show that, with imperfect fit, this method can lead to lower bias of the ATET estimator even under a non-zero correlation between treatment assignment and time-varying unobservables.[39]

Following Peri and Yasenov (2017) and Peri et al. (2020), a way to check the lack of statistically significant differences between the synthetic unit and the observed treated unit comes from the estimated coefficients of the saturated panel model given vt

$$y = treated + \delta_{pre}pre + \sum_Q \delta_{post,q}post_q + \pi_{pre\times treated}\,(pre \times treated) + \sum_Q \pi_{post\times treated,q}\,(post_q \times treated) + \varepsilon \tag{2.9}$$

where $treated$ is 1 if the observation is the treated unit and 0 otherwise; pre is 1 if the observation belongs to the pre treatment period 2005-2014 (2015 is the base) and 0 otherwise and $post_p$, with $q = 2016, 2017, 2018, 2019$, is 1 if the observation belongs to every year in the post-treatment and 0 otherwise.

Abadie et al. (2010) propose an exact inferential method based on the distribution of the estimator from (in-space) placebo interventions. Each of these placebo effects are calculated by iteratively taking one department as the treated while placing the remaining J units in the donor pool, and estimating the treatment effects as in Equation 2.6. Under the null hypothesis of no effect, an abnormally large effect for the treated i compared to the placebos signals its statistical significance. However, some large effects might not be actually caused by the intervention but by a poor pre-treatment fitting, distorting the comparison against the placebo runs. Moreover, taking an arbitrary cut-off for the Root Mean Squared Prediction

[36] In our application, each of the columns in $\mathbf{X}_i$ and $\mathbf{X}_0$ can be thought as partitioned in 2: a column vector comprised of pre-treatment averages of each of the $\mathbf{Z}$ economic predictors and a column vector comprised of some observed pre-treatment values for the outcome. This latter controls for an Ashenfelter-type induced selection that can potentially bias the estimator.

[37] The estimation of the covariates weights matrix, $\mathbf{V}$, is important to achieve bias reduction. Based on Kuosmanen et al. (2021) results, we avoid using a nested optimization algorithm for the estimation of $\mathbf{V}$. We instead resort to the regression-based approach to determine the elements of this matrix.

[38] This is a generalization of Equation 2.4 in which the sum of the two scalars $\alpha_s + \lambda_t$ is replaced by the product of two vectors, say $\boldsymbol{\alpha}_s\boldsymbol{\lambda}_t$, each representing a set of fixed factors (factor loadings) and a set of year-specific factors (common factors)(Samartsidis et al. 2019). The DiD cannot identify the ATET in this model because the unobserved group and year effects are no longer linearly related. Direct estimation requires a larger number of departments and years than those in our application to ensure that asymptotic unbiasedness applies (see Bai 2009; Xu 2017).

[39] Ferman and Pinto (2021) show that the method suggested by Doudchenko and Imbens (2016), who measure outcomes in levels but including an intercept in the minimization problem to estimate the $(\omega_{5+1}^*, ..., \omega_{5+J}^*)$ and construct the counterfactual, is equivalent to theirs.

Errors (RMSPE) as a criterion for "good" fitting ends up being completely arbitrary (Abadie 2020; Abadie et al. 2015). Consequently, we resort to p-values from the distribution of ratios of post and pre-treatment RMSPEs, $RMSPE_h^{post} \big/ RMSPE_h^{pre}$ where $h = i, 5+1, ..., 5+J$ from the permutations, calculated as the relative rank of the ratio, $p = Rank \big/ J$.[40] Compared to the graphical analysis in Boruchowicz et al. (2021), by relying on p-values we add transparency to the inference as these penalize the estimated effects by a poor pre-treatment fit. Firpo and Possebom (2018) conclude from their simulations that this p-value is the uniformly more powerful test compared to other inference procedures conventionally applied with SCM. Due to the size of J, we consider an effect as significant only if the p-value for i is the smallest obtained.

As alternative robustness checks (in section 2.7), we also estimate the ATET from the aggregated panel data (Equation 2.3) using the (partialled-out) FE and First Differences (FD) correcting the standard errors using methods in the previous sub-section. These two estimators consistently identify the ATET even if the the unobserved effects (α_s, λ_t) are correlated with the explanatory variables. However, a key assumption is strict exogeneity, $E(v_{st}|D_{sT}, ..., D_{s1}, \mathbf{X}_{sT}, ..., \mathbf{X}_{s1}, \alpha_s, \lambda_t) = 0$ (plus a suitable rank assumption) which implies that the explanatory variables in each t are uncorrelated with v_{st}, $\forall s \neq t$.

In order to safeguard against the possibility these estimators are confounded by individual state trends and a failure of the strict exogeneity assumption, we estimate two additional models. The first is a random-growth model where an additional source of heterogeneity is given by a trend of the unobserved department-specific effect in Equation 2.3, $g_s t$. Likewise, we do not restrict the correlations between $(\alpha_s, \lambda_t, g_s t)$ and the explanatory variables but still retain the assumption strict retain the assumption of strict exogeneity. As shown in Wooldridge (2010), first-differencing the equation with the random-trend leads to

$$\Delta Y_{st} = g_s + \eta_t + \beta \Delta D_{st} + \Delta \mathbf{x}_{it}^{'} \gamma + \Delta u_{it}; s = 1, ..., S; t = 2, ..., T \tag{2.10}$$

where $\eta_t \equiv \lambda_t - \lambda_{t-1}$. The second model is a dynamic AR(1) panel data model which drops the strict exogeneity and only assumes sequential exogeneity (i.e. allows for a feedback of u_{st} to $D_{s,t+p}$, $p = 1, ..., T - t$)

$$Y_{st} = \alpha_s + \lambda_t + \theta Y_{s,t-1} + \beta D_{st} + \mathbf{X}_{st}^{'} \gamma + V_{st}; s = 1, ..., S; t = 2, ..., T \tag{2.11}$$

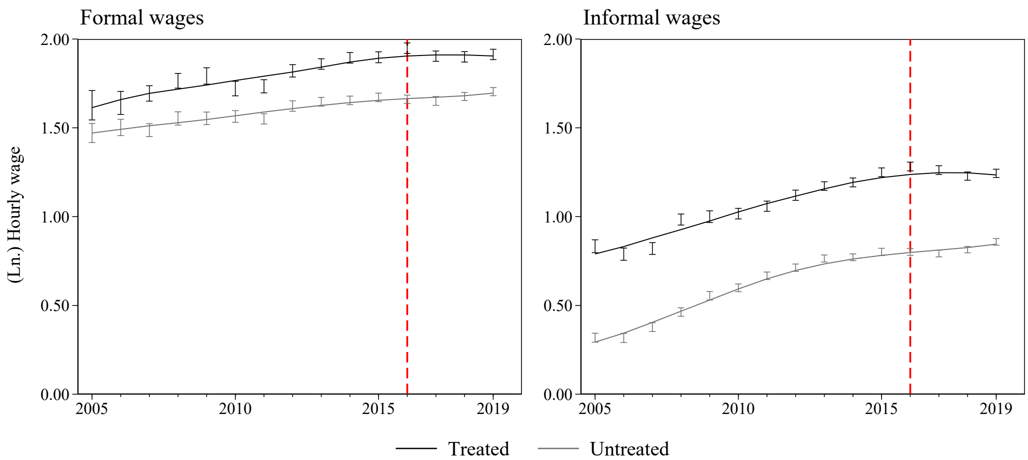
This requires that only one lag of y_{it} is necessary for dynamic completeness once we condition on the unobserved effects and the observed covariates. The coefficient corresponding to the lagged variable is not of direct interest here. However, it proxies for time-varying unobserved factors that affect the evolution of the outcome (Bond 2002) and safeguards the key estimators from the threat of reverse causality (Leszczensky and Wolbring 2019). We estimate this as a two-step difference GMM estimator (Arellano and Bond 1991) incorporating an efficient (cluster-robust) weighting matrix to account for the correlation between differenced errors. We instrument the endogenous term $\Delta Y_{s,t-1}$ by its (internal) lagged levels. A proliferation in the number of instruments is avoided by restricting the instruments to just two lags and further collapsing them (by adding horizontally formerly distinct columns of the Holtz-Eakin et al. 1988 instrument matrix). The failure to do so can result in overfitting of the endogenous variables and a reduction of the power of the Hansen test for overidentifying restrictions (Roodman 2009a,b). We anticipate a good small sample performance in the light of these restrictions as the loss of relevant information for omitting more distant lags tends to be modest (Bond 2002).

[40] For each placebo run, $RMSPE_j^{pre} = \sqrt{\frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{jt} - \sum_k w_k^* Y_{kt}\right)^2}$ and $RMSPE_j^{post} = \sqrt{\frac{1}{T-T_0}\sum_{t=T_0}^{T}\left(Y_{jt} - \sum_k w_k^* Y_{kt}\right)^2}$ where $k = i, 5+1, ..., 5+J$ excluding the j taken as treated

2.5.2 Identifying assumptions checks

We first examine the pre-2016 evolution of the average outcomes. A likely result of the (unconditional) PTA is that the series for treated and untreated departments in the pre-treatment period were also parallel (Angrist and Pischke 2009; Angrist and Krueger 1999). The averages of (log.) hourly wages in the treated areas have been consistently higher than the comparable averages for the untreated group and is reduced after the beginning of the treatment. The monotonic growth in the first ten years is a clear manifestation of the 'Peruvian Growth Miracle' (Ross and Peschiera 2015). Their parallel trends signal the validity of the assumption (Figure 2.3). The graph for the ancillary variables (Figure 2.A7) indicates that immigrants locate in areas with lower informal employment (though with rates above 55%) and that the gap between untreated and treated areas has risen from 10 to 15 percentage points. In turn, Gini coefficients for both formal and informal sectors are not statistically different. However, the (unconditional) PTA holds in both cases despite the non-linear evolution of the inequality in the informal sector (see Frölich and Sperlich 2019). These results suggest that including all non-treated departments as control groups (instead of just an ad-hoc selection) is unlikely to lead to counterfactual identification problems.

Figure 2.3 – Trends in real (log.) hourly wages, 2005-2019



Note: Sample includes employed native individuals between 18 and 65 years. Real hourly wages in 2007 PEN. Vertical dashed lines denote the beginning of the treatment.
Source: Author's calculations using ENAHO 2005-2019 data.

Secondly, we examine if treated departments are different in terms of observed characteristics compared to those untreated units. Because the DiD approach controls for all time-invariant differences between treatment and control groups, we focus on time-varying differences that affect the outcome variables' trends in the pre-treatment.[41] The characteristics in both groups have remained steady and are reasonably similar, suggesting that significant differences in varying unobserved confounders are unlikely. Specifically, the proportion of males and average workers' age and education are close (Table 2.A4). Compared to untreated areas, natives in treated areas have a more extensive representation in Craft and Trades occupations (10% vs 14%, respectively) and less in the Technical jobs (16% vs 28%, respectively). Also, natives in treated areas have a higher representation in the Services industry (around 24% vs 18%) and a lower representation in the Agriculture sector (6% vs 27%). The fact that treated areas are the most significant urban centres in the country explains these findings. We include control for these variables in our econometric modelling to account for these differences.

Thirdly, we analyze patterns of migration of Venezuelans within Peru using ENPOVE data. It is done in recognition that the migration of Venezuelans across Peru in search of better job opportunities can undermine the identification of ATET, as it would lead to spillover effects over the departments that we assume

[41] Because the treatment is at department-level, the exposure of individuals to the treatment is unconfounded (Angrist and Pischke 2009).

to be untreated (Abadie 2020). The first indication that there is no selective migration is the reduced time (a little more than ten days) it takes them to reach their current location in Peru once they leave Venezuela. Such a short time window makes it unlikely for them to integrate into labour markets in the sites along their migration route. This time window is plausibly explained by the transit difficulties they experience in their journey. For example, those heading towards the Andean regions (Arequipa and Cuzco) take almost one month to reach their destination due to the extra time and monetary costs involved and the lower supply of transportation services to those areas. Nevertheless, the transit time outside Peru mainly explains the total transit time of Venezuelans. In fact, it thoroughly explains the transit time for Venezuelans settled in Tumbes.

A second indication that immigrants did not change labour markets within Peru is the low proportion of long journeys prior to getting to their current location. Only 15% took more than one month and 6% more than two months to reach their current location. A further indication is that almost all of them remained in the same department once they arrived in Peru. Indeed, 83% remained in the same district (a smaller administrative area than department), ranging from 75% to 98% across the treated areas. Additionally, the fact that over 90% of Venezuelans plan to remain in Peru reassures us that our study does not suffer from emigration outside the host country either, unlike what Akgündüz et al. (2018) report for the Syrians in Turkey. This evidence suggests the absence of migration of Venezuelans across regions and outside Peru. Hence we believe that spillovers are unlikely to be a concern for identification.

Table 2.4 – Venezuelan immigration routes and timing, 2018

	Treated areas					**Total**
	Arequipa	Cuzco	La Libertad	Lima y Callao	Tumbes	
Total of trips						
Mean	7.63	8.06	7.23	6.49	6.18	6.52
	(0.10)	(0.16)	(0.16)	(0.06)	(0.25)	(0.05)
Trip length outside Venezuela						
Total	1.07	0.94	0.73	0.36	0.85	0.38
	(0.13)	(0.21)	(0.12)	(0.04)	(0.16)	(0.04)
Outside Peru	0.68	0.64	0.57	0.34	0.85	0.35
	(0.09)	(0.14)	(0.09)	(0.04)	(0.16)	(0.04)
Inside Peru	0.39	0.30	0.16	0.02		0.03
	(0.08)	(0.13)	(0.06)	(0.01)		(0.01)
Proportion of journeys						
More than 1 month	0.26	0.22	0.20	0.14	0.35	0.14
	(0.02)	(0.03)	(0.02)	(0.01)	(0.05)	(0.01)
More than 2 months	0.15	0.12	0.11	0.06	0.12	0.06
	(0.01)	(0.02)	(0.01)	(0.00)	(0.03)	(0.00)
Has stayed in the same department						
Proportion	0.95	0.95	0.94	1.00	0.97	1.00
	(0.01)	(0.02)	(0.01)	(0.00)	(0.01)	(0.00)
Has stayed in the same district						
Proportion	0.75	0.68	0.91	0.83	0.97	0.83
	(0.02)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)
Plans to stay in Peru						
Proportion	0.95	0.91	0.93	0.94	0.95	0.94
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)

Notes: Sample restricted to only those employed between 18 and 65 years. Trip length in months. Trip length outside Venezuela refers to length of the trip (in months) from 1st stop outside Venezuela to the current place in Peru. Average length of intermediate stay measures the average time (in monthds) per place where immigrants transit outside Venezuela and Peru, for those whose trip length is more or equal than 2 months. SEs in parenthesis. *Source*: Author's calculations using ENPOVE data.

If the Venezuelan Exodus drives down wages for certain skill groups, native Peruvians in that skill group can also move around different labour markets to gain higher wages (or at least face strong incentives to do so). This, in turn, will dissipate the impacts of immigration throughout the national economy and hence our econometric methods will not be able to identify these effects (Dustmann et al. 2008). For the USA, Card (2001) finds that mobility flows of natives and older immigrants in the USA are minimal, and hence this substitution effect is likely negligible in magnitude. We have reasons to believe that the same holds for Peru. The first reason for this is that the real estate market in Peru is not as developed as those in the USA and Europe, which makes the internal migration of workers more challenging. The second reason is that in 2013-2018 almost half of the departments in Peru had a positive migration rate. Cuzco and Tumbes, two of the treated departments, are part of this group, but their inflow rates are negligible (less than 0.5 for every 1000 workers) compared to the other regions (INEI 2019b, Table 9.3). Finally, following Dustmann et al. (2013), our large regional definitions (regions and not districts) make migratory movements more likely to be internalized and reflected in the econometric results.

2.6 Results

This section presents the ATETs of the Venezuelan Exodus on the key (viz., a formal and informal wages) and ancillary (viz. informal employment levels, formal and informal sector inequality) outcomes using the whole set of employed workers between 18 and 65 years old (as described in section 2.3). For the DiD estimation, both the typical single-stage and the two-stage, we include as controls a set of individual level demographic variables typically included in Mincerian equations (a male dummy, years of age and years of schooling, and an interaction of these two latter, a dummy variable for urban settlement status) in addition to vectors containing industry and occupation dummies.[42]

For the SCM estimation, the variables included as controls at department-year level are real GDP per capita, the percentage of the economically active population in services, the share in manufacturing, along with the economically active population with secondary schooling. Additionally, we include the average years of schooling, the proportion of people aged 18 to 25 and 26 to 35, as well as the proportion of workers in urban areas and a dummy for if the department is on the border frontier with Ecuador and Colombia. Because the pre-treatment outcomes would also allow us to control for factor loadings in the estimation of the counterfactual, we include in the model pre-treatment outcomes for only years 2005, 2010 and 2015. In section 2.7 we apply other estimation methods that relax some of the assumptions assumed here.

Since we anticipate heterogeneous effects on specific subgroups of the native population, we also estimate results for an array of sub-groups. Also, based on the demographic characteristics of Venezuelan workers (from Table 2.1) and the documented downgrade they experience (from Table 2.2), we expect that the most affected group will be those those with at most secondary education and aged between 18 and 35 years old. These are more prone to directly compete in the labour market with Venezuelans immigrants. Hence, we separately analyse this particular subsample (see section 2.8, Table 2.B1 to Table 2.B22). The effects on these "low skilled" workers are mostly indistinguishable from those in the whole sample and so we only refer tangentially to these in what follows.

2.6.1 Key outcomes

We take as a departure point the single-stage DiD with repeated cross-sections. I.e., initially, we (naively) ignore the role of v_{st} in Equation 2.4 and rely on the variance estimators conventionally used in this small-S design. However, the estimated treatment effect (upper panel in Table 2.5) is not statistically significant under any inference method. The lower panel suggests that the effects exhibit temporal heterogeneity instead. Inference based on the CRVE suggests that the treatment had significant effects in the formal sector only at the beginning (in 2016) and in the informal sector only at the end of the period (in 2019).

[42] Additional estimation results from models including only demographic controls and those adding only a vector of industry dummies where the individual is employed were estimated. The results for these are available upon request.

Nonetheless, the reliability of these commonly applied methods is in our case, at best, dubious. The first indication of this is the sizeable difference in the effective number of clusters (around 6) and the observed number of clusters (24), suggesting that inference should be based instead on p-values from the wild cluster bootstrap.[43] A further indication is provided by the statistical significance of the (cluster-robust) F-test on the leads ($D_{s2011},\ldots,D_{s2014}$) from the regression for the informal sector. This latter suggests that the key DiD identification assumption is not fulfilled in this sector. Still, results from this single-stage DiD method suggest that the Exodus caused an increase in native formal hourly wages of 3.2% in 2016 and an increase in wages in the informal sector. Hourly wages in the formal sector for less-skilled workers were not affected (see table Table 2.B1).

Table 2.5 – Single stage DiD: ATET of Venezuelan immigration on natives' (log) formal and informal wages (full set of controls), 2011-2019

	Formal wages					Informal wages				
	$\hat{\beta}$	SE Homosk.	SE CRVE	P-val. Rade	P-val. Webb	$\hat{\beta}$	SE Homosk.	SE CRVE	P-val. Rade	P-val. Webb
Aggregated										
2016-2019	0.009	(0.008)	(0.015)	[0.460]	[0.530]	0.029	(0.008)***	(0.020)	[0.225]	[0.295]
S*	6.208					6.451				
Yearly (Base 2015)										
2016	0.032	(0.017)*	(0.016)*	[0.085]	[0.100]	0.028	(0.017)	(0.020)	[0.365]	[0.350]
2017	0.016	(0.017)	(0.015)	[0.300]	[0.335]	0.020	(0.017)	(0.018)	[0.315]	[0.360]
2018	-0.028	(0.017)	(0.020)	[0.270]	[0.285]	-0.039	(0.017)**	(0.024)	[0.320]	[0.235]
2019	-0.022	(0.017)	(0.020)	[0.395]	[0.435]	-0.066	(0.017)***	(0.025)**	[0.125]	[0.100]
N	148,419					267,006				
S*	6.139					6.519				
F stat. OLS	1.625					3.884**				

Notes: The sample is restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Individual level covariates are age and schooling years in levels and interacted and gender, area, industry and occupation dummies. SE CRVE refers to Cluster Robust Variance Estimator of SEs; P-val. WB Rade and Webb, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights and Webb weights, respectively, using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). F-stat refers to the statistic of the F test for the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Once we take into account the effect of the unobserved department-year specific shocks (v_{st}) by estimating the 2S-DiD (Equation 2.5, using in the first stage the complete set of controls) along with more efficient variance estimators, the effects of the Exodus on wages lose statistical significance (Table 2.6).[44] Two results stand out compared to those from the 1-stage DiD. First, p-values from the wild bootstrap procedure based on the Rademacher distribution (as well as those based on the Webb distribution, not shown in the table), which provide more reliable inference given S^*, are noticeably larger. This is also true for p-values that result from the evaluation of the CRVE2-based Wald statistic using the IK-DoF. Secondly, the lack of significance of Granger F-tests (bottom of the table) provides more credibility to these results as this suggests the fulfilment of the (conditional) parallel trends assumptions. Note how these conclusions are invariant to the number of lags we use to characterize the AR process.

[43] The effective number of clusters and the IK-DoF in the regressions in this study (for both the total and low-skilled only sample) are small enough to suggest that reliable inference can be achieved by methods other than CRVE.

[44] This lack of statistical significance remains invariant to a change in the set of controls used in the first stage (Table 2.A5).

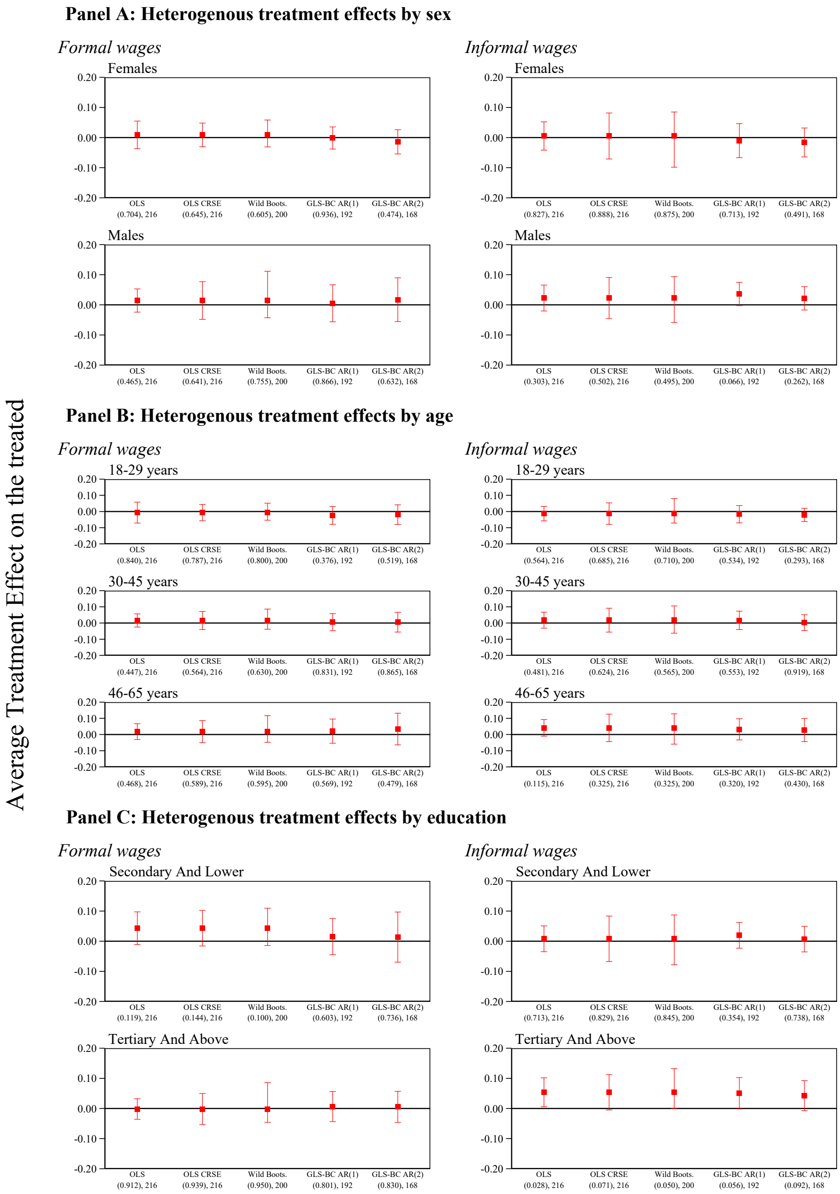
Table 2.6 – 2 stages DiD: ATET of Venezuelan immigration on natives' (log) formal and informal wages (full set of controls), 2011-2019

		Formal wages							Informal wages						
		$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. IK	P-val. WB	N	$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. IK	P-val. WB	N
Aggregated															
2016-2019	FE OLS	0.013	(0.015)	(0.024)	(0.024)	[0.609]	[0.675]	216	0.016	(0.018)	(0.030)	(0.031)	[0.623]	[0.635]	216
	FE GLS-BC AR(1)	0.008	(0.023)	(0.023)				192	0.019	(0.026)	(0.023)				192
	FE GLS-BC AR(2)	0.010	(0.021)	(0.029)				168	0.009	(0.024)	(0.019)				168
	S*	7.835							7.835						
	IK-DoF	6.846							6.846						
Yearly (Base 2015)															
2016	FE OLS	0.027	(0.033)	(0.023)	(0.024)	[0.298]	[0.295]	216	0.029	(0.038)	(0.024)	(0.025)	[0.293]	[0.320]	216
	FE GLS-BC AR(1)	0.026	(0.029)	(0.026)				192	0.026	(0.027)	(0.029)				192
	FE GLS-BC AR(2)	0.027	(0.029)	(0.027)				168	0.026	(0.026)	(0.029)				168
2017	FE OLS	-0.006	(0.033)	(0.025)	(0.025)	[0.827]	[0.835]	216	0.024	(0.038)	(0.025)	(0.025)	[0.366]	[0.390]	216
	FE GLS-BC AR(1)	-0.007	(0.035)	(0.027)				192	0.019	(0.036)	(0.031)				192
	FE GLS-BC AR(2)	-0.006	(0.031)	(0.028)				168	0.020	(0.033)	(0.028)				168
2018	FE OLS	0.011	(0.033)	(0.040)	(0.041)	[0.793]	[0.770]	216	-0.011	(0.038)	(0.032)	(0.033)	[0.746]	[0.795]	216
	FE GLS-BC AR(1)	0.010	(0.037)	(0.044)				192	-0.017	(0.042)	(0.038)				192
	FE GLS-BC AR(2)	0.011	(0.033)	(0.046)				168	-0.016	(0.036)	(0.034)				168
2019	FE OLS	0.004	(0.033)	(0.031)	(0.032)	[0.914]	[0.915]	216	-0.030	(0.038)	(0.028)	(0.028)	[0.322]	[0.290]	216
	FE GLS-BC AR(1)	0.003	(0.038)	(0.034)				192	-0.037	(0.045)	(0.032)				192
	FE GLS-BC AR(2)	0.004	(0.034)	(0.036)				168	-0.035	(0.037)	(0.029)				168
	S*	7.835							7.835						
	IK-DoF	6.846							6.846						
	F stat. FE OLS	1.707							1.080						
	F stat. GLS-BC AR(1)	0.211							0.394						
	F stat. GLS-BC AR(2)	0.008							0.272						

Notes: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender, area, industry and occupation dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the statistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

We now estimate the ATET using the 2S-DiD method restricting the data to different subpopulations in terms of sex, age and education. However, the lack of significance of the treatment effect on hourly wages remains (Figure 2.4). Only within the small group of informal workers with technical, college and postgraduate education, the treatment is found to be statistically significant on the basis of the confidence intervals (at 5% of significance) across different inference procedures. The effect on this group, which represents only 11% of the total workers and 16% of the informal employees, is positive and around 5.2%.[45] The general lack of significance of the effects is also found when separating the low-skilled sub-sample of males and females (Figure 2.B1).

Figure 2.4 – 2 stages DiD: ATET Heterogeneity of Venezuelan immigration on natives' (log) formal and informal wages by gender, sex and education, 2011-2019



Note: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender, area, industry and occupation dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. FE OLS CRVE refers to Cluster Robust Variance Estimator of SEs; Wild Boots., to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications and FE GLS-BC, to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. *Source*: Author's calculations using ENAHO 2011-2019 data.

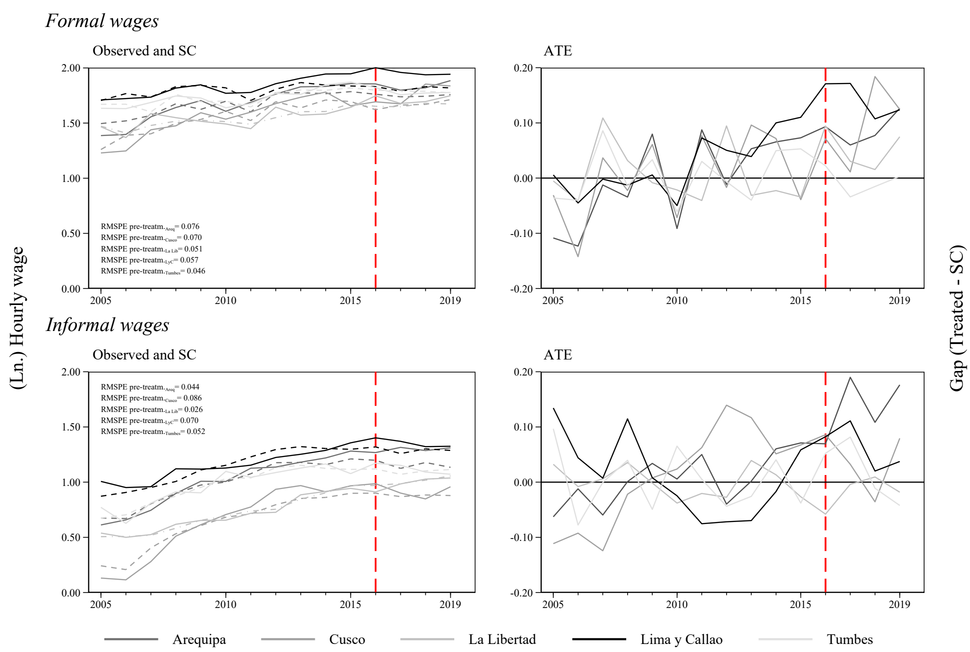
The estimated synthetic control units, which allows estimation of the ATET under the SCM, does a fair job reproducing the pre-treatment trajectories of treated departments. Three different criteria provide sup-

[45]The average department-year cell size in this group is 12,797.6 observations

port for this. The pre-treatment RMSPE values (left panels of Figure 2.5), summarizing the discrepancies in pre-treatment outcomes among the treated departments and their estimated synthetic control, are reasonably low. Also, the averages of the control variables for the treated regions and their corresponding synthetic controls are close to each other (Table 2.A7). The lack of statistical significance of the coefficient from the interaction of the pre-treatment period dummy and the synthetic unit dummy (column labelled "2005-2014" in Table 2.A8, from DiD-like regressions (Equation 2.9), provides further reassurance. This also holds for the analysis on the low-skilled subpopulation (table Figure 2.B2 and Table 2.B6).

Focusing on the estimated ATET (shown in the right side panels in Figure 2.5 following Equation 2.6), the observed effects of the Exodus on hourly wages are positive. Nonetheless, a visual assessment of its statistical significance reveals that, except for Cusco and Lima in the formal sector and Arequipa in the informal sector, these effects are not extreme compared to the distribution of placebo effects (Figure 2.A9). A more thorough assessment, based on the p-values from the distribution of ratios of RMSPEs (shown in Table 2.7), confirms the lack of statistical significance of the effects except for hourly wages in the formal sector in Lima and Callao. In this department, the Exodus increased the wages by 14%, an estimate larger than the ones reported by previous studies. Nevertheless, the pre-treatment fit for Lima and Callao is far from perfect as this standard method from Abadie et al. (2010) assumes. A re-assessment of the effects using the Ferman and Pinto (2021) synthetic control estimator, in order to safeguard from this failure of the convex hull assumption, shows that the effects in all departments, including Lima and Callao, are statistically insignificant (Table 2.8). Although the effects in the low skilled population based on the standard SCM yields non significance in all departments (table Table 2.B7), those obtained using the demeaning procedure are statistically significant in Lima and Callao's informal sector (table Table 2.B10). This result will be further subject to a robustness check in subsection 2.7.2.

Figure 2.5 – SCM: ATET of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years before data aggregation. Real hourly wages in 2007 PEN.
Source: Author's calculations using ENAHO 2005-2019 data.

The estimation of the counterfactuals using the standard SCM for sub-samples across sex, age and education yields statistically insignificant treatment effects, as shown in Table 2.9 (see Figure 2.A11 for the

Table 2.7 – SCM: P-values for the ATET of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019

	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
Formal wages					
Arequipa	0.0893	0.0758	1.2216	7/20	0.3500
Cusco	0.0978	0.0700	1.6671	4/20	0.2000
La Libertad	0.0538	0.0508	1.2342	6/20	0.3000
Lima y Callao	0.1434	0.0574	2.5467	1/20	0.0500
Tumbes	-0.0060	0.0459	0.4746	20/20	1.0000
Informal wages					
Arequipa	0.1360	0.0443	3.2643	2/20	0.1000
Cusco	0.0406	0.0862	0.7343	14/20	0.7000
La Libertad	-0.0176	0.0260	1.1832	10/20	0.5000
Lima y Callao	0.0627	0.0696	1.0391	12/20	0.6000
Tumbes	0.0202	0.0516	1.0293	11/20	0.5500

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

evolution of the TEs). The same conclusion is reached when splitting the sample among low-skilled males and females (see Table 2.B11)

Table 2.8 – SCM demeaned: P-values for the ATET of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019

	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
Formal wages					
Arequipa	0.0168	0.0882	0.3832	18/20	0.9000
Cusco	0.1464	0.0907	1.8068	3/20	0.1500
La Libertad	0.0459	0.0629	0.8288	14/20	0.7000
Lima y Callao	-0.0685	0.0456	1.7697	4/20	0.2000
Tumbes	-0.0455	0.0472	1.2175	9/20	0.4500
Informal wages					
Arequipa	0.0455	0.0487	1.0362	13/20	0.6500
Cusco	0.0170	0.0811	0.6312	16/20	0.8000
La Libertad	0.0023	0.0344	0.9277	13/20	0.6500
Lima y Callao	-0.1428	0.0577	2.7023	2/20	0.1000
Tumbes	0.0086	0.0560	0.9869	10/20	0.5000

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.9 – SCM: P-values for the ATET of Venezuelan immigration on natives' (log) formal and informal wages across subsamples, 2005-2019

	Formal wage					**Informal wage**				
	ATE	RMSPE pre	Ratio post-pre	Rank	p-value	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
Sex										
Arequipa										
Male	0.1705	0.1354	1.2645	4/24	0.1667	0.1336	0.0610	2.2250	1/24	0.0417
Female	0.0922	0.1185	0.9848	9/24	0.3750	0.0837	0.0952	0.9207	6/24	0.2500
Cusco										
Male	0.1266	0.0933	1.6329	2/24	0.0833	0.0716	0.0904	0.9141	12/24	0.5000
Female	0.1220	0.1313	1.1349	5/24	0.2083	-0.0067	0.1042	0.7497	12/24	0.5000
La Libertad										
Female	0.0623	0.1232	0.9058	10/24	0.4167	0.0406	0.0988	0.6837	10/24	0.4167
Male	0.0812	0.0688	1.1921	4/24	0.1667	-0.0059	0.0472	0.2177	20/24	0.8333
Lima y Callao										
Male	0.2212	0.1681	1.3173	2/24	0.0833	0.0497	0.0792	0.7311	14/24	0.5833
Female	0.2066	0.2042	1.0379	9/24	0.3750	0.1460	0.1428	1.0988	5/24	0.2083
Tumbes										
Female	0.0423	0.1072	0.7002	11/24	0.4583	-0.0470	0.1632	0.3857	17/24	0.7083
Male	0.0107	0.0794	0.4046	17/24	0.7083	-0.0110	0.0536	1.3747	4/24	0.1667
Age										
Arequipa										
30-45 years	0.1573	0.1231	1.3506	5/24	0.2083	0.0432	0.0696	0.9565	13/24	0.5417
18-29 years	0.0288	0.1359	0.5320	14/24	0.5833	0.0936	0.1204	1.0004	8/24	0.3333
46-65 years	0.1606	0.1327	1.2182	8/24	0.3333	0.1713	0.1272	1.3968	2/24	0.0833
Cusco										
46-65 years	0.1532	0.1471	1.0575	8/24	0.3333	0.0322	0.1135	0.5900	11/24	0.4583
30-45 years	0.1114	0.0888	1.5210	3/24	0.1250	0.0341	0.0887	0.9593	13/24	0.5417
18-29 years	0.0910	0.1177	1.2686	5/24	0.2083	0.0981	0.1638	0.7266	13/24	0.5417
La Libertad										
30-45 years	0.0526	0.0592	1.0254	6/24	0.2500	0.0105	0.0411	0.5899	18/24	0.7500
46-65 years	0.1005	0.0914	1.1212	9/24	0.3750	-0.0314	0.0741	0.7418	9/24	0.3750
18-29 years	0.0699	0.1244	0.7982	11/24	0.4583	0.0393	0.0785	0.9050	11/24	0.4583
Lima y Callao										
30-45 years	0.2512	0.1792	1.4028	4/24	0.1667	0.0666	0.0865	0.9374	14/24	0.5833
18-29 years	0.1018	0.1779	0.8500	11/24	0.4583	0.0519	0.1748	0.3861	19/24	0.7917
46-65 years	0.2754	0.1760	1.5648	4/24	0.1667	0.1252	0.1613	0.8432	4/24	0.1667
Tumbes										
18-29 years	0.0669	0.1449	0.5339	15/24	0.6250	-0.1200	0.1046	1.2860	4/24	0.1667
46-65 years	0.1069	0.1384	0.9917	10/24	0.4167	0.0272	0.1001	1.0086	4/24	0.1667
30-45 years	-0.0067	0.0860	0.9386	7/24	0.2917	0.0257	0.0529	1.2819	6/24	0.2500
Education										
Arequipa										
High level	0.1063	0.1205	0.9534	9/24	0.3750	0.1296	0.1101	1.3220	2/24	0.0833
Low level	0.1895	0.1558	1.2340	6/24	0.2500	0.1268	0.0919	1.3977	4/24	0.1667
Cusco										
High level	0.0967	0.0888	1.2863	5/24	0.2083	0.1260	0.1444	1.0536	7/24	0.2917
Low level	0.0870	0.1396	1.4555	3/24	0.1250	0.0624	0.0943	0.9535	5/24	0.2083
La Libertad										
Low level	0.0471	0.0674	0.7616	13/24	0.5417	0.0002	0.0408	0.7990	11/24	0.4583
High level	0.0800	0.0668	1.4482	4/24	0.1667	0.0772	0.1104	0.7387	11/24	0.4583
Lima y Callao										
High level	0.1498	0.1994	0.7832	13/24	0.5417	0.1749	0.2538	0.7026	12/24	0.5000
Low level	0.2432	0.1755	1.3959	4/24	0.1667	0.0417	0.0672	0.9335	6/24	0.2500
Tumbes										
Low level	0.0351	0.1074	0.4957	19/24	0.7917	-0.0197	0.0642	1.0685	5/24	0.2083
High level	-0.0270	0.0775	0.5850	16/24	0.6667	-0.0168	0.1770	0.1687	20/24	0.8333

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

2.6.2 Ancillary outcomes

As with the key outcomes, we take as the point of departure the single-stage DiD method using repeated cross-sections (Equation 2.4) to analyze the ancillary outcomes, including as controls the same set of Mincerian demographic and occupational variables as before. The aggregated treatment effect (upper panel in Table 2.10) is, again, not statistically significant for any of the ancillary outcome variables, whereas some dynamic effects, reported in the lower panel, are found to be significant only when using the CRVE estimator. However, the sizeable difference between the observed number of clusters and the effective number (S^*) suggests that this inference is unreliable. Accordingly, testing the null of no effect using p-values from wild bootstrap methods reveals that only the effect on the size of the informal sector is statistically significant. However, this occurs under a violation of the conditional PTA. The same results are found in the low-skilled working sample (Table 2.B12).

Table 2.10 – Single Stage DiD: ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector (full set of controls), 2011-2019

	Informality rate				Formal sector inequality				Informal sector inequality			
	$\hat{\beta}$	SE CRVE	P. WB R	P. WB W	$\hat{\beta}$	SE CRVE	P. WB R	P. WB W	$\hat{\beta}$	SE CRVE	P. WB R	P. WB W
Aggregated												
2016-2019	0.003	(0.008)	[0.740]	[0.740]	0.000	(0.006)	[0.985]	[0.985]	0.014	(0.011)	[0.495]	[0.510]
S*	6.211				6.208				6.453			
Yearly (Base 2015)												
2016	0.016	(0.007)**	[0.140]	[0.160]	0.012	(0.010)	[0.320]	[0.315]	0.013	(0.012)	[0.235]	[0.300]
2017	0.003	(0.009)	[0.780]	[0.765]	0.004	(0.012)	[0.775]	[0.750]	0.030	(0.016)*	[0.420]	[0.440]
2018	0.005	(0.010)	[0.630]	[0.650]	0.016	(0.009)	[0.190]	[0.165]	0.020	(0.018)	[0.400]	[0.415]
2019	0.018	(0.007)**	[0.040]	[0.020]	0.011	(0.008)	[0.215]	[0.130]	0.032	(0.016)*	[0.385]	[0.420]
N	415,425				148,419				267,006			
S*	6.048				6.135				6.513			
F stat. OLS	8.622***				0.705				2.194			

Notes: The sample is restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Individual level covariates are age and schooling in levels and interacted and gender, area, industry and occupation dummies SE CRVE refers to Cluster Robust Variance Estimator of SEs; P. WB Rade and Webb, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights and Webb weights using 199 replications, respectively. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). F-stat refers to the stastistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

When applying the two-step DiD method to control for the confounding effect of the unobserved department-year specific shocks (Equation 2.5), two results stand out. Firstly, the finding from using the single-stage DiD that the treatment increased informality levels by two percentage points in 2019 is confirmed (Table 2.11). The different variance-covariance matrix estimators suggest this is a robust finding. In fact, it is statistically significant (at 5%) according to the p-values constructed from the wild bootstrap methods (which tend to be more conservative compared to alternative methods for inference) and also according to the p-values using the IK-DoF (at 10%, not shown in the table). The Granger predictability test is statistically significant in the estimation using FE although those based on the FGLS-BC are not. This provides some reassurance that the estimated effect under this latter estimation method corresponds to the ATET. This same evidence is also found for low skilled workers (Table 2.B13), but the significance is now marginal when evaluated against the Rademacher-based wild Bootstrap p-values. Second, the immigration shock increased inequality by two percentage Gini points in the formal sector in 2018, and it is statistically significant only in the whole sample and by taking a liberal significance level according to p-values from Rademacher wild bootstrap (10% significance). The inference obtained using the wild bootstrap Webb-based or IK-DoF (not shown) p-values suggest instead that this estimate is not statistically significant.

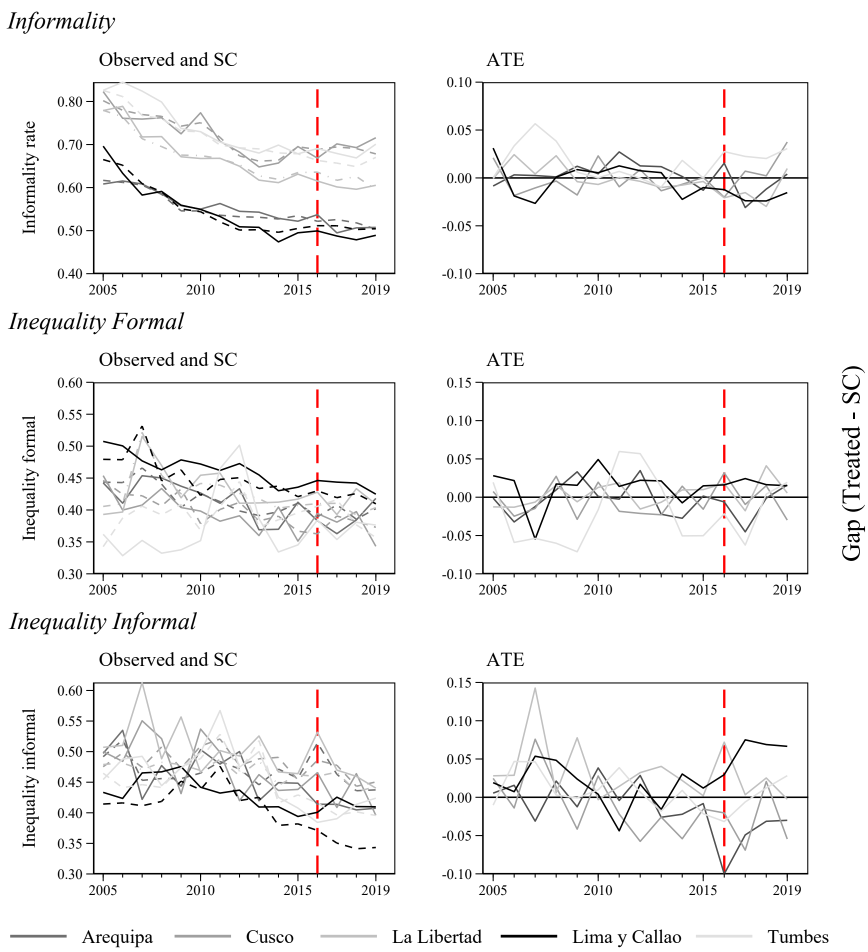
Table 2.11 – 2 stages DiD: ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector (full set of controls), 2011-2019

		Informality rate					Formal sector inequality					Informal sector inequality				
		$\hat{\beta}$	SE CRVE	SE CRVE2	P-val. WB	N	$\hat{\beta}$	SE CRVE	SE CRVE2	P-val. WB	N	$\hat{\beta}$	SE CRVE	SE CRVE2	P-val. WB	N
Aggregated																
2016-2019	FE OLS	-0.002	(0.008)	(0.008)	[0.795]	216	-0.006	(0.007)	(0.007)	[0.425]	216	-0.003	(0.016)	(0.017)	[0.890]	216
	FE GLS-BC AR(1)	0.006	(0.006)			192	0.002	(0.007)			192	0.008	(0.013)			192
	FE GLS-BC AR(2)	0.008	(0.005)			168	0.010	(0.008)			168	0.013	(0.011)			168
	S*	7.835					7.835					7.835				
	IK-DoF	6.846					6.846					6.846				
Yearly (Base 2015)																
2016	FE OLS	0.003	(0.007)	(0.007)	[0.685]	216	0.017	(0.014)	(0.014)	[0.235]	216	0.006	(0.020)	(0.021)	[0.800]	216
	FE GLS-BC AR(1)	0.004	(0.008)			192	0.017	(0.016)			192	0.006	(0.022)			192
	FE GLS-BC AR(2)	0.004	(0.008)			168	0.017	(0.016)			168	0.008	(0.022)			168
2017	FE OLS	-0.003	(0.008)	(0.008)	[0.780]	216	0.003	(0.014)	(0.014)	[0.855]	216	0.010	(0.014)	(0.014)	[0.530]	216
	FE GLS-BC AR(1)	-0.002	(0.008)			192	0.003	(0.015)			192	0.011	(0.015)			192
	FE GLS-BC AR(2)	-0.002	(0.008)			168	0.003	(0.015)			168	0.013	(0.015)			168
2018	FE OLS	-0.003	(0.008)	(0.008)	[0.750]	216	0.023	(0.012)*	(0.012)*	[0.080]	216	0.023	(0.017)	(0.017)	[0.225]	216
	FE GLS-BC AR(1)	-0.002	(0.009)			192	0.023	(0.013)			192	0.023	(0.019)			192
	FE GLS-BC AR(2)	-0.002	(0.009)			168	0.023	(0.014)			168	0.026	(0.019)			168
2019	FE OLS	0.019	(0.009)*	(0.010)*	[0.035]	216	0.022	(0.014)	(0.014)	[0.135]	216	0.015	(0.014)	(0.015)	[0.370]	216
	FE GLS-BC AR(1)	0.020	(0.010)*			192	0.022	(0.015)			192	0.015	(0.016)			192
	FE GLS-BC AR(2)	0.020	(0.010)*			168	0.022	(0.015)			168	0.018	(0.017)			168
	S*	7.835					7.835					7.835				
	IK-DoF	6.846					6.846					6.846				
	F stat. FE OLS	3.143**					0.883					0.608				
	F stat. GLS-BC AR(1)	1.043					0.683					0.237				
	F stat. GLS-BC AR(2)	0.437					0.269					0.081				

Notes: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender, area, industry and occupation dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. SE CRVE refers to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the statistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Analysing the estimated synthetic departments, the pre-treatment trajectories for these counterfactuals and the observed treated departments exhibit a reasonable overlap, before 2016, for the outcome related to the proportion of informal workers (Figure 2.6). However, the fit for the inequality outcomes is relatively noisier in the pre-treatment period. Nevertheless, the pre-treatment differences are not statistically significant according to the DiD-like regressions (Table 2.A12), and the discrepancy of the averages of covariates between the observed treated departments and the estimated synthetic control departments are within reasonable bounds (Table 2.A11). Turning to ATET estimation, the observed effects for informality and the Gini in the formal sector are small. In contrast, the observed effects for inequality in the informal sector are mostly negative except for Lima and Callao (first column in Table 2.12). Nonetheless, the visual assessment of the statistical significance of these effects (Figure 2.A12) do not suggest statistical significance in any treated area. This is also confirmed when taking a more robust inference approach based on the p-values from the permutations (last column in Table 2.12), as the RMSPE ratios for the treated areas are located in the bottom half of the distribution of ratios. This lack of significance is further confirmed when applying the Ferman and Pinto (2021) demeaning procedure (Table 2.A14) to control for the more obvious lack of perfect treatment in the case of these outcomes.

Figure 2.6 – SCM: ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years before data aggregation. Real hourly wages in 2007 PEN. The figures show the causal effect taking each treated region separately. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.12 – SCM: P-values for the ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-201

	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
1 Informality					
Arequipa	-0.0058	0.0115	1.5917	8/20	0.4000
Cusco	0.0067	0.0139	1.5524	9/20	0.4500
La Libertad	-0.0140	0.0112	1.8059	8/20	0.4000
Lima y Callao	-0.0188	0.0165	1.1865	12/20	0.6000
Tumbes	0.0252	0.0240	1.0667	14/20	0.7000
2 Inequality Formal					
Arequipa	-0.0091	0.0210	1.1505	9/20	0.4500
Cusco	0.0019	0.0205	1.1518	9/20	0.4500
La Libertad	0.0117	0.0122	1.9949	4/20	0.2000
Lima y Callao	0.0180	0.0280	0.6561	13/20	0.6500
Tumbes	-0.0151	0.0502	0.6872	14/20	0.7000
3 Inequality Informal					
Arequipa	-0.0527	0.0221	2.7035	4/20	0.2000
Cusco	-0.0308	0.0392	1.1816	12/20	0.6000
La Libertad	0.0245	0.0535	0.7150	15/20	0.7500
Lima y Callao	0.0600	0.0298	2.1036	3/20	0.1500
Tumbes	0.0007	0.0251	0.8872	12/20	0.6000

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

2.7 Robustness checks

2.7.1 Panel data methods

We verify the results above using Panel data estimators from data aggregated at the department-year level for the full set of years (following Equation 2.3). The control variables used are the same for SCM, excluding the dummy if the department is on the borders with Ecuador and Colombia and excluding also the pre-treatment outcomes. Estimation of the ATET using FE, FD and FE-GLS for the key variable confirm the foregoing conclusions (Table 2.A15). Using the whole sample, the Exodus fails to register an effect on wages in both the informal and formal sectors. In fact, for the formal sector, the statistical significance of the cluster robust F-test for treatment dummies in the pre-treatment years suggest a violation of the conditional PTA, so arguably the estimated impacts do not actually correspond to the ATET of interest. For the low skilled population (see Table 2.B3), the Exodus had a negative effect on informal wages in 2019 but only according to the p-values from the wild bootstrap based on the Rademacher distribution for the FE estimator. The p-values from the Webb distribution and those based on the IK-DoF (not shown) are larger than 0.10. However, because we anticipate the population errors to be autocorrelated, results from the FD estimator and the FGLS are more reliable. These confirm the lack of statistical significance.

Panel data estimation of the random growth model by FD and FE (Equation 2.10) and also of the dynamic panel estimated by GMM (Equation 2.11 confirms the absence of statistical significance for the causal effects in terms of formal and informal wages (Table 2.A16 and Table 2.A17).

Turning to the ancillary variables, panel data estimators (Table 2.A18) confirms the previous results on informality and inequality in the formal sector. Point estimates for these two outcomes are significant across different methods (including those which address the small S problem manifested in the S^* and IK-DoF). The estimates suggest that participation in the informal market increased by two percentage points in 2019, and inequality increased by three to four percentage points in the formal sector in the last two years. However, the statistical significance of the F-statistic across these different estimations suggests a violation of the basic identification condition of the PTA. The estimated results for the low skilled sample (Table 2.B14) coincide with those obtained for the treatment effects on informality.

Results from the random growth model confirms the increasing effects on informality in the last year and in the formal sector inequality in the last two years and also that the inflow increased inequality in the informal sector between 3 and 4 percentage points in the last two years. (Table 2.A19). Results from dynamic panel for these ancillary variables (Table 2.A20) suggest the lack of significance of the treatment, yet the large J-test statistic suggests that the exogeneity of instrumentation is extremely weak. Therefore, we settle for the results from the random growth models among these panel data estimators.

2.7.2 Alternative Synthetic Control Methods

Because theory rarely tells us what covariates to include in $\mathbf{X}_0$ and $\mathbf{X}_i$, a researcher would have the opportunity to select specifications that adjust their hypothesis, generating substantial opportunities for subjective researcher bias. In fact, simulations in Ferman et al. (2020) reveal that slight variations in model specification for SCM applications lead to significantly different treatment effects.[46] This would do away with the higher transparency that this method offers compared to alternative estimation methods (Abadie et al. 2015; Abadie 2020). Hence, the first robustness check for SCM is to consider alternative models which, as Ferman et al. (2020) show, tend to avoid the problem of cherry-picking of the results. In contrast to what Abadie (2020) recommends in the finite sample setting, they suggest focusing on specifications that only use pre-treatment outcome lags which progressively increase with T_0 in order to take care of unobserved confounders (especially factor loadings). We fit three distinct models for both key and ancillary outcomes: one where the covariate matrix includes only all the pre-treatment periods, another including only pre-treatment outcomes in odd years (2005, 2007, ..., 2015) and another with only even years (2006, 2008, ...,2014). The RMSPEs for the divergence between the pre-treatment difference between treated units and the control unit (see Table 2.A21 and Table 2.A22) are noticeably lower than those from the standard and the demeaned SCM. The results are similar to those using standard and demeaned SCM estimator for each of the five outcomes. This implies that the treatment had non-significant effects for all the outcomes except for low skilled workers' wages in Lima and Callao. This negative effect detected, however, remains non-negligible, around 9%.

A second robustness check applies the residualization process of the outcomes outlined in Peri and Yasenov (2017), in recognition that the average wages of different demographic groups had different national trends in the period of study, which in turn might introduce confounding factors into the analysis[47]. This adjust for the potential confounding adjusting log wages of individual i in department s and year t for observed demographic characteristics with heterogeneous slopes for every year. In terms of the population, the model we fit is

$$\begin{aligned} y = \theta_0 &+ \sum_{t=1}^{T} \theta_{1t}\left(Female \times Year_t\right) + \sum_{t=1}^{T} \theta_{2t}\left(Age \times Year_t\right) + \sum_{t=1}^{T} \theta_{3t}\left(Schooling \times Year_t\right) + \sum_{t=1}^{T} \theta_{4t}\left(Age \times Schooling \times Year_t\right) \\ &+ \sum_{t=1}^{T} \theta_{5t}\left(Area \times Year_t\right) + \sum_{t=1}^{T}\sum_{m=2}^{M} \theta_{6,mt}\left(Industry_m \times Year_t\right) + \sum_{t=1}^{T}\sum_{p=2}^{P} \theta_{7,pt}\left(Occupation_p \times Year_t\right) + \varepsilon \end{aligned} \tag{2.12}$$

where $Year_\tau = 1\left(year = \tau\right)$ is an indicator function equal to 1 if the observation for year τ, $Female$ is a dummy equal to 1 if the individual is female and 0 otherwise, age is measures the age of the individual in years, $area$ is a dummy equal to 1 if the individual lives in the urban area, and $Industry_m$ and $Occupation_p$ are a set of dummies for the industry and occupation where the individual lives, as described in section 2.3.

[46]More specifically, using 7 different common specifications, for a nominal test size of 5% the probability of detecting a false positive in at least 1 specification can be as high as 14% when $T_0 = 12$. To put matters worse, possibilities for specification searching remain high even with an unrealistically large number of pre-treatment years: with $T_0 = 400$ the probability that at least 1 specification is significant at 5 % is 13%.

[47]Note how this differs from the residualization process outlined in Doudchenko and Imbens (2016) who, prior to choosing the weights and the intercept, they suggest regressing the outcomes of the controls the controls on the pre-treatment variables, calculate the residuals and use these residuals in the approaches above. Instead, here we residualize the series from individual covariates not the aggregated ones.

The residual $\hat{\varepsilon}$ from this regression captures individual variation after wiping out those aggregate trends that concerns us. We implement the SCM on those residuals averaged by department. Results in Table 2.A23 for the most part, coincide with those obtained earlier. As in Peri and Yasenov (2017), because we are now trying to match “residualized" outcomes which are noisier measures than observed average wages, these series exhibit year-to-year fluctuations which results in a more noisy pre-treatment match.

2.7.3 Variables definitions

We consider a number of variations in the construction of the dependent variables used in the estimations. Related to wages, the first alternative uses monthly wages, and not standardized by the number of hours. This is done to avoid possible measurement errors in the number of hours. The second uses a different deflator, namely the Consumer Price Index (CPI) taking as the base year 2007 (obtained from Central Reserve Bank of Peru). Related to informality, we restrict the occupational categories included in this sector. On the one hand, we consider as formal only those white and blue collar workers, both in the public and private sector, that are in pensionable jobs. On the other hand, we consider as informal those employers or white and blue collar workers, both in the public and private sector, without a pensionable job. We exclude from the analysis independent workers in recognition of potential differences in their wage data generation process.

The results for the two-stage DiD using these alternative wages definitions remain unchanged after accounting that inference should not rely on the CRVE estimator based on the value of S^{*} (Table 2.A24). Under the alternative definition of informality, results from Wild Bootstrap again show a lack of statistical significance (Table 2.A25).[48]. Hence, our results are not sensitive when redifining the dependent variables.

2.8 Conclusions

The effects of the exogenous immigration shock represented by the Venezuelan Exodus, the most prominent external displacement crisis in Latin America’s recent history, on the Peruvian labour market are studied. However, key conditions in the host country of Peru differ from the typical setting analyzed in the literature, such as Mariel Boatlifters into Miami, repatriated immigrants from Africa to France and Portugal or the sizeable influx of Syrian refugees into Turkey. Peru is a society not used to receiving immigrant workers, and its native labour force is less educated and mainly employed in the informal sector of the economy. The characteristics of this unique event, which began in 2016 and was triggered not by the buoyant macroeconomic conditions in Peru but by a socio-economic collapse in Venezuela, allows us to exploit quasi-experimental methods to establish causal effects. Unlike possible alternative methods available to analyze immigration effects (mainly the National Approach), our methods do not rely on structural modelling or assumptions about production functions or the elasticity of substitutions. Because the treatment occurred at an aggregate level in few departments, we are able to use a combination of empirical methods that address problems about consistency and inference of the estimators that have been recently emphasized by studies in the literature exploiting immigration as a quasi-experiment. Therefore, the study contributes to the literature on the effects that such a significant migratory movement has had on Latin American country’s labour market. This extends the typical focus on wages in recognition that the adjustments induced by such shocks go beyond the formal labour markets, also impacting informal labour markets.

We first described the Venezuelan immigration and, second, analyzed the labour market effects of this shock under different and alternative estimators in an attempt to determine a causal interpretation of the results. Regarding the former, most of the Venezuelan population is younger than the native population, mainly aged between 18 to 35 years old, and most of these are part of the economically active population.

[48]This robustness of the results shown in section 2.6 also holds for the estimations from the single-stage DiD method (not shown)

They are primarily employed in the informal sector. Their education level is predominantly tertiary (college or technical), a higher level than that of the natives, most of whom have secondary schooling. Despite this, these immigrants suffer a considerable occupational downgrade in Peru, both in terms of the occupational category and job complexity compared to the activity they worked in Venezuela. Specifically, the occupations they enter in the host country are more manual in nature than those involving cognitive and communication abilities. This results in a lower occupational complexity score. This result also applies to Venezuelans who have spent sufficient time in Peru to avail better job opportunities. An important explanation for this lies in the several hurdles progressively implemented by the Peruvian government to restrict the flow of immigrants along with the inefficiency of the country's immigration authorities and its legal standards, which prevented Venezuelans from joining the formal labour market. The earnings of migrants as a consequence are lower than the natives across most of the distribution.

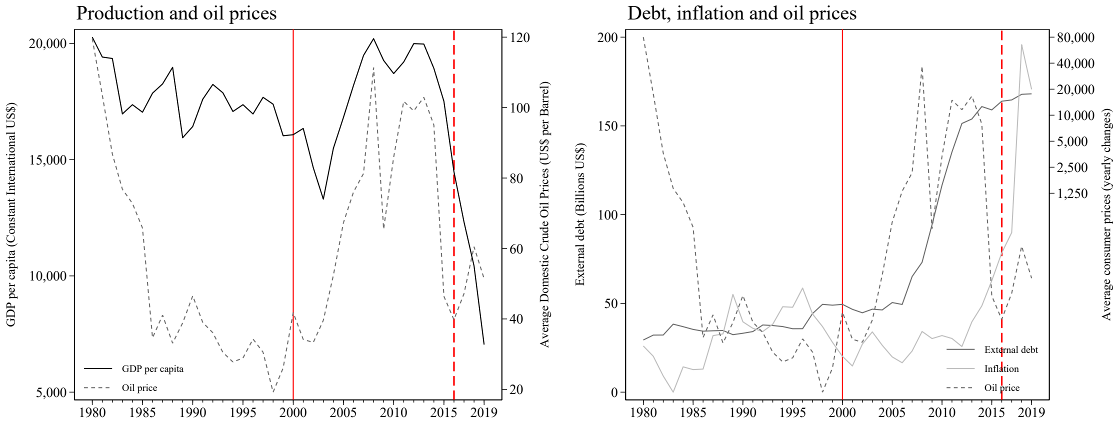
The treatment effects of the Exodus, based on the 2-step Difference in Difference procedure, using both a repeated cross-section and panel data, reveal that the Exodus has impacted neither formal nor informal wages in Peru. The methods applied account for the distortion (under-rejection) in the Wald tests induced by the fact that the number of clusters is smaller than those observed (24) and that the treatment happens in a few number of regions. The procedure also accounts for the unobserved department shocks specific to a given period, which confounds the estimated treatment effect on the treated. Once these attributes are taken into account, our DiD strategy results in statistically insignificant effects of the Exodus on our key outcome variables in the Peruvian labour market. This robustness of this finding is also detected when using the Synthetic Control Method estimator, which constructs the counterfactual for every department based on an optimal convex combination of control units. However, an important exception of this result is the adverse effects that the treatment exerted on the wages for informal workers in Lima y Callao, which suffered a reduction of around 10%. This result remains valid under alternative methods. Results for the ancillary variables, i.e. informality and informal and formal Gini inequality indices, are different depending on the method. On the one hand, inequality in informal and formal labour markets have increased in the last year of the treatment examined here (2019): the former in +2% and the latter by 2% to 4%. However, the SCM procedure does not reveal statistically significant results in any of the treated departments.

The absence of effects in most cases suggests that either the increase in the relative supply of Venezuelans was offset by an increase in relative demand in the affected labour markets or Venezuelans provided some degree of labour complementary for native workers. In the case of Lima y Callao, an effect in line with the predictions of a competitive labour market has been observed. This evidence departs from what previous studies for Peru have found. Even though this result might provide support to the views that immigration has a damaging effect on wages of a specific subsection of the native population, it is necessary to interpret this relative to the labour supply increase that the Exodus induced in Lima and Callao. Such an assessment suggests that wages had an inelastic response to the Exodus. Specifically, a 20% increase in labour supply reduced the wage for informal unskilled workers by 10%. Thus, it is arguable that the impact of this Exodus on the wages of the unskilled natives is, in fact, moderately low. A more elaborate rationalization of this potentially differing nature of the effects obtained is confined to the agenda for future research.

Appendix A: Descriptives and results for the whole sample

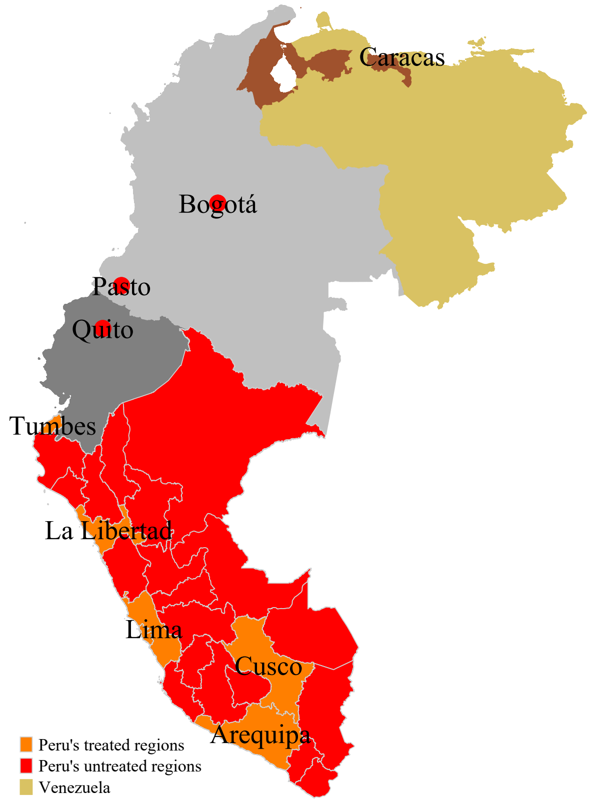
Descriptives

Figure 2.A1 – Macroeconomic indicators in Venezuela, 1980-2019



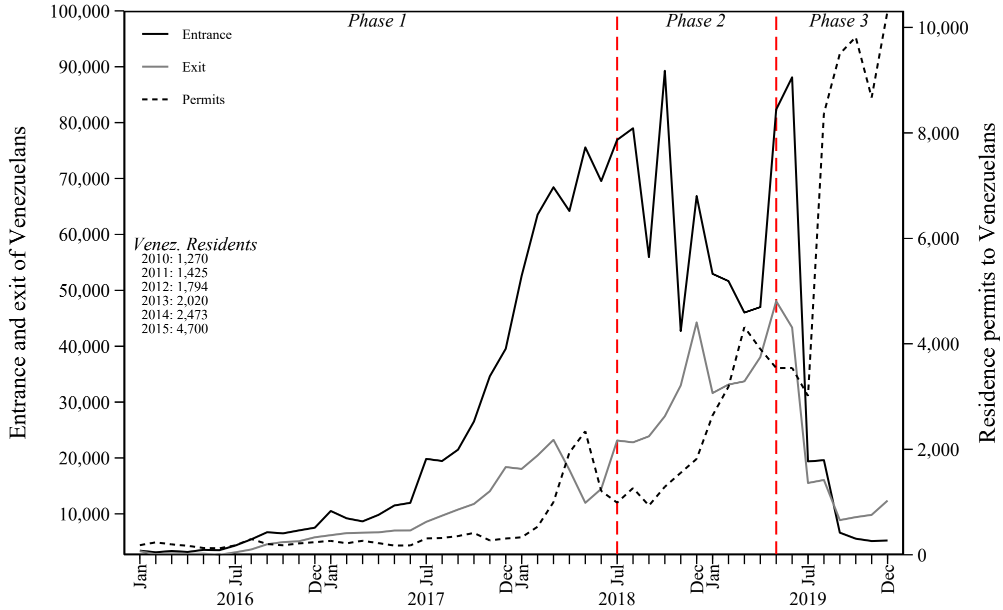
Note: GDP per capita is observed only until 2010; onwards it is imputed. Inflation in logarithm base 10 scale. Solid vertical line is the first year of Hugo Chavez in power and the vertical dashed line is the beginning of the treatment. *Source*: GDP and inflation from IMF's World Economic Outlook (October 2020), oil price from inflationdata.com and External debt from World Bank, International Debt Statistics.

Figure 2.A2 – Treatment areas in Peru and Venezuela



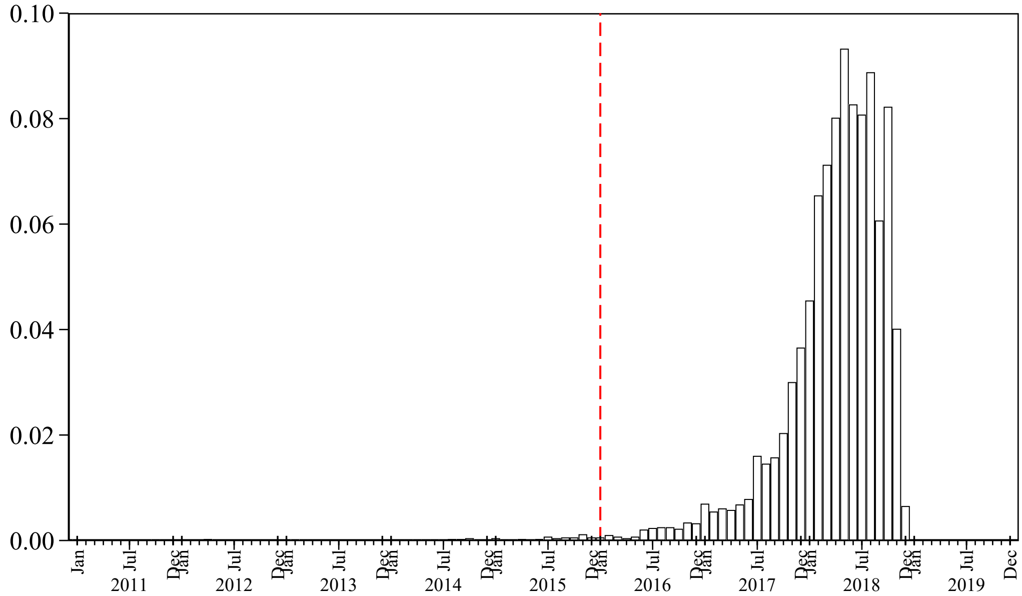
Note: Ecuador and Colombia in dark and light grey, respectively. The 5 states in Venezuela where half of the Venezuelan immigrants in Peru began their emigration journey in brown. Lines correspond to the most commmon routes followed by Venezuelan immigrants according to UNHCR (2018) *Source*: Author's calculations using ENPOVE data.

Figure 2.A3 – Venezuelan immigrants flow in Peru, Jan. 2016 - Dec. 2019



Note: Data includes all the immigrants, not only those with working permit and including underaged. *Source*: Data on Residents from Response for Venezuelans (https://r4v.info/es/situations/platform) and on entrance, exit and residence permits from Superintendencia de Migraciones, Peru.

Figure 2.A4 – Venezuelan immigrant's flow and time spent in Peru (months) as of 2018 in ENPOVE sample



Note: Sample restricted to only those employed between 18 and 65 years. Date of arrival relative to when they entered Peru. The vertical l
the treatment. *Source*: Author's calculations using ENPOVE data.

Table 2.A1 – Legal Regulations set by the Peruvian Government to face with the Venezuelan Exodus

DS	Date	Effect	Requisites	Comments
Phase 1				
DS N° 002-2017-IN: Approval of guidelines for the Temporary Permit of Permanence (PTP) for people of Venezuelan nationality	3 January 2017	Provide the PTP for 1 year (extendable) to Venezuelans who have entered Peru before 2 of February 2017 ("before the publication of this legal norm" is a modification, 48 hours later, of the originally proposed date "before the exclusion of Venezuela from the Mercosur Agreement [5 of August 2017]"; deadline to submit the application: 3 July 2017	1) Have legally entered Peru before 2 of February 2017 2) Being in Peru in an irregular situation as a result of the expiration of residence authorization, or being in a regular situation 3) Have no criminal or judicial record at the national and international level 4) Pay S/. 42	PTP is a document issued by the migration authority which accredits to the Venezuelan citizen their regular migratory situation in Peru and enables them to carry out formal economic activities according to the Peruvian legislation. This was validated gradually within the following 2 months by Peru's Migration Authority
DS Nº 007-2017-IN: Definition of New Migrant Categories	27 March 2017	Migration authority defines the *Special Resident Migration Status* for those foreigners who, having entered Peru, need to regularize their immigration status. This allows the foreigner multiple entrances into Peru and to work as employee or self-employed in the public and or private sectors. Initially, the validity of this document is 1 year but can be extended	-	The Special Resident Status will be given to those Venezuelans who have been granted the PTP as in DS N° 002-2017-IN after 1 year of residence in Peru
DS Nº 023-2017-IN Extension of the approval of Guidelines for the Temporary Permit of Permanence (PTP) for people of Venezuelan nationality in DS N° 002-2017-IN	29 July 2017	Provide the PTP for 1 year (extendable) to Venezuelans who have entered Peru before 31 of July 2017; deadline to submit the application: 1 December 2017	Same as in DS N° 002-2017-IN except: 1) Have legally entered Peru before 31 of July 2017	
DS Nº 001-2018-IN Extension of the approval of Guidelines for the Temporary Permit of Permanence (PTP) for people of Venezuelan nationality in DS N° 023-2017-IN	23 January 2018	Provide the PTP for 1 year (extendable) to Venezuelans who have entered Peru before 31 of December 2018; deadline to submit the application: 30 June 2019	Same as in DS Nº 023-2017-IN except: 1) Have legally entered Peru before 31 of December 2018	

Table 2.A1 – Legal Regulations set by the Peruvian Government to face with the Venezuelan Exodus

DS	Date	Effect	Requisites	Comments
RS No 043-2018-MIGRACIONES	30 January 2018	Defines the procedure for obtaining the Special Resident Migration Status for Venezuelans after the expiration of the PTP	They must meet the following conditions: 1) submitting the application 30 days before the expiration of their PTP 2) have not been absent from Peruvian territory for more than 183 consecutive or alternate days since their last departure as a beneficiary of the PTP, within a period of 365 days, without authorization from Immigration; 3) have no criminal, judicial or police record	The motivation behind this was the fact that the PTP delivered to Venezuelans began to expire. Hence, this set concrete guidelines to obtain the Special Resident migration Status for Venezuelans in Peru, defined in DS Nº 007-2017-IN. The immigration card constitutes the identification document for Venezuelans that proves their regular residence and provides them important legal benefits
RS 0000165-2018-MIGRACIONES Free delivery of the "Extraordinary Work Permit" to Venezuelans whose application for Temporary Permit of Permanence (PTP) is being processed	12 May 2018	Migration Authority adopted the Extraordinary Work Permit for all those who were processing the PTP. This document enabled Venezuelans to perform economic activities as employee or self-employed for 2 months, subject to automatic extension, until obtaining the PTP.	Filling of an online application	This was issued in recognition of the challenges faced by Venezuelans whose PTPs were being processed, mainly because of the number of months it took them getting the required documentation from different government institutions and changing their tourist permit to the PTP, which is usually six months and during this time they cannot carry out economic activities.
RM Nº 176-2018-TR Dispositions for hiring Venezuelans with Temporary Permit of Permanence (PTP) or Extraordinary Work Permit	6 July 2018	Defines that • The quotas restrictions for foreigners within firms are not changed • The term of the contract for the Venezuelan cannot go beyond the validity of their Extraordinary Work Permit or PTP • The loss of validity of the Extraordinary Work Permit or PTP invalidates the contract	-	Establishes dispositions for hiring Venezuelans with PTP or Extraordinary Work Permit
Phase 2				

Table 2.A1 – Legal Regulations set by the Peruvian Government to face with the Venezuelan Exodus

DS	Date	Effect	Requisites	Comments
DS Nº 007-2018-IN Modification of Guidelines for Granting of the Temporary Permit of Permanence (PTP) for people of Venezuelan nationality approved in DS No. 001-2018-IN	19 August 2018	Provide the PTP for 1 year (extendable) to only those Venezuelans who entered Peru before 31 of October 2018; deadline to submit the application: 31 December 2018	Same as in DS Nº 001-2018-IN except: 1) Have legally entered Peru before 31 of October 2018	Reduction of the validity period to request PTP
RS N° 000270-2018-MIGRACIONES	24 August 2018	A valid passport is required for Venezuelans for entrance into Peru from 25 of August 2018, except in "humanitarian" cases that included: applicants for the refugee status, children under 18 years, pregnant women in a state of vulnerability and those over 65-year-old who had deteriorated health due to the emigration trip. This put an effective end to the PTP		The National Coordinator for Human Rights deemed this as illegal because this violated the the right to freedom of movement of Venezuelans. The Superior Court of Justice of Lima voted in favour of this habeas corpus claim on 5 of October 2018, officially dropping this disposition. Yet, 2 months later, on 4 of December, this decision was reversed by Peru's Judicial Power.
DS N° 008-2018-TR Supreme Decree that modifies the Regulation of the Law of Contracting Foreign Workers, approved by Supreme Decree No. 014-1992-TR	13 September 2018	From 13 of October 2018 1) Registration of contracts in the Labour ministry switch to virtual, which overrides the need of issuing an official permit by the corresponding authority 2) There is no need to provide a legal document that states that the employer fulfil with the quota of foreign workers within their firm 3) The government has the authority audit the firm within the next 5 years from the date the contract with the foreigner was terminated in order to verify the fulfilment of this new law		Its purpose was updating regulations on contracting of foreign personnel to address the migration issue and to simplify current regulations. This not only applies to Venezuelans but for all non-Peruvians
Phase 3				

Table 2.A1 – Legal Regulations set by the Peruvian Government to face with the Venezuelan Exodus

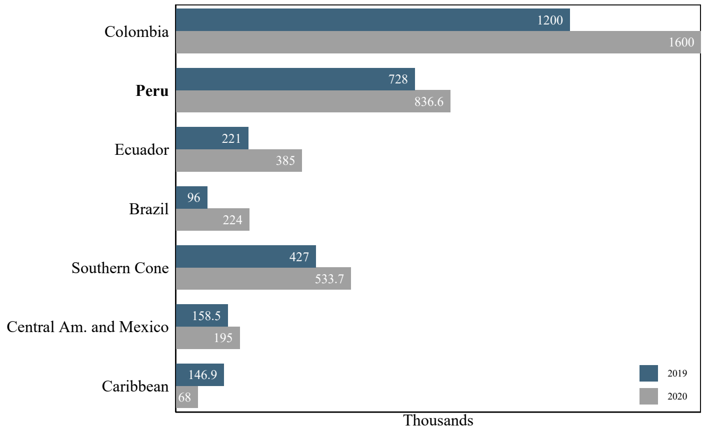
DS	Date	Effect	Requisites	Comments
RS N° 000177-2019-MIGRACIONES Venezuelans can enter the country only with Passport and a particular types of Visa	12 June 2019	The right of Venezuelans to enter Peru without a passport or without a tourist visa is suspended from 15 of June 2019	Venezuelans can enter if they hold passports (either valid or expired) under the following conditions For those with Temporal migratory or Resident migratory status 1) Passport 2) Visa granted by a Peruvian Consular Office For those with Humanitarian Resident migratory status 1) Passport 2) Birth Certificate 3) Humanitarian Visa granted by a Peruvian Consular Office In any case, the migrant must also show a legalized Venezuelan criminal record certificate	

Table 2.A2 – Venezuelans' route characteristics and immigration status, 2018

	Treated areas					Total
	Arequipa	Cuzco	La Libertad	Lima y Callao	Tumbes	
Intermediate countries						
Colombia only	5.57	3.84	3.30	6.82	0.71	6.72
Ecuador and Colombia	86.38	86.22	93.12	84.46	99.04	84.68
Ecuador only	1.86	1.91	1.62	1.89	0.26	1.88
None	2.84	1.14	1.48	5.85	0.00	5.70
Other	3.35	6.89	0.49	0.98	0.00	1.02
Immigration post						
Lima airport	8.06	3.42	2.50	8.74	0.00	8.58
Tacna other	2.17	5.84	0.51	0.16	0.00	0.21
Tumbes	89.77	90.74	96.99	91.10	100.00	91.21
Mode of transportation						
Air and bus	8.28	8.20	5.90	6.41	0.96	6.41
Air only	2.37	0.90	0.08	5.23	0.00	5.08
Bus and foot/sea	7.39	3.88	7.82	2.53	17.48	2.72
Bus only	81.08	86.07	85.53	85.43	79.79	85.37
Other	0.88	0.95	0.66	0.40	1.76	0.42
Document shown at the border						
ID	18.02	34.06	28.21	18.52	49.26	18.80
Other	0.14	0.48	1.07	0.62	0.66	0.63
Passport	81.84	65.46	70.72	80.86	50.08	80.57
Migration status						
Tourist Visa	2.96	4.82	2.74	4.04	4.98	4.01
Work/Student Visa	0.08	1.43	0.00	0.13	0.00	0.13
TPP requester	60.91	54.72	50.51	54.51	51.39	54.50
TPP holder	25.07	14.53	19.63	29.03	8.04	28.72
Refugee	5.55	31.16	18.99	4.85	22.60	5.24
Immigration card	4.12	1.20	3.41	5.92	0.64	5.83
Other	3.58	1.58	7.57	3.98	13.25	4.06

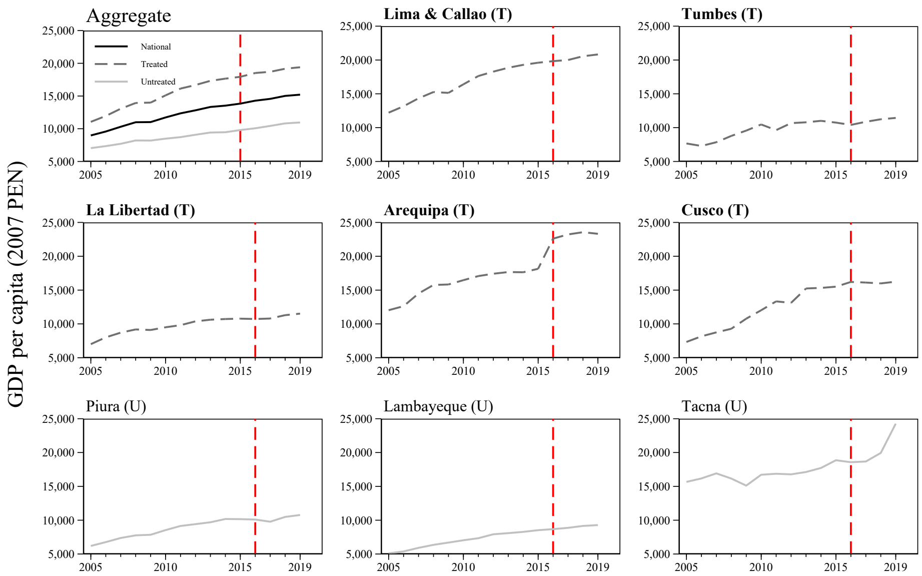
Notes: Sample restricted to only those employed between 18 and 65 years. Refugee status include those who have requested it and those who hold it. Any individual can hold more than 1 migration status. Mode of transportation and intermediate countries involved measure the relative frequency (%) for individuals within every treated location. *Source*: Author's calculations using ENPOVE data.

Figure 2.A5 – Venezuelan immigrants stock across different countries, 2019 and 2020



Source: UNHCR (2019) Venezuela situation. Fact Sheet, April 2019 and UNHCR (2020) Venezuela situation. Fact Sheet, January 2020

Figure 2.A6 – Per capita GDP of treated and selected untreated regions in Peru, 2005-2019



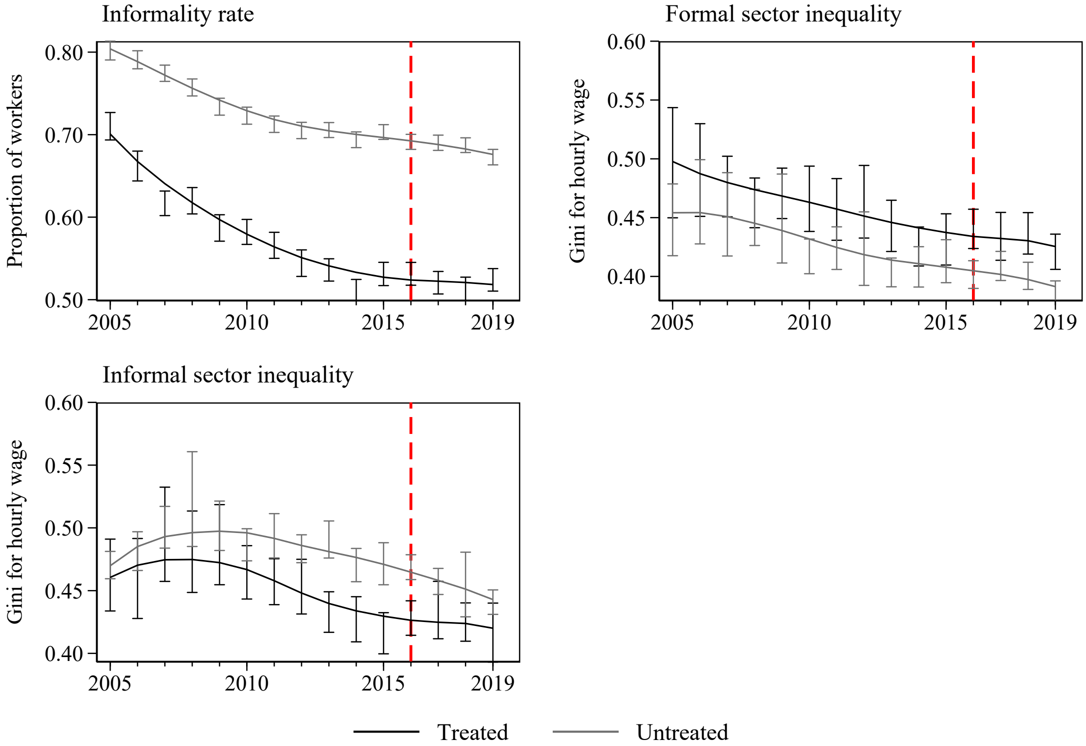
Note: (T) and (U) refer to treated and untreated regions. Vertical dashed lines denote the beginning of the treatment. *Source*: Author's calculations using INEI's official website and INEI (2013).

Table 2.A3 – Occupational distribution and skill level by treated areas, 2018

	Arequipa					**Cuzco**					**La Libertad**					**Lima y Callao**					**Tumbes**				
	Peruv.	*Venezuelans*				*Peruv.*	*Venezuelans*				*Peruv.*	*Venezuelans*				*Peruv.*	*Venezuelans*				*Peruv.*	*Venezuelans*			
		Earlier		Recent			Earlier		Recent			Earlier		Recent			Earlier		Recent			Earlier		Recent	
		Pre	Post	Pre	Post		Pre	Post	Pre	Post		Pre	Post	Pre	Post		Pre	Post	Pre	Post		Pre	Post	Pre	Post
Occupation category (%)																									
Managers and profess	10.36	20.37	3.30	19.83	2.71	7.27	13.61	6.29	16.24	3.11	7.95	14.48	1.76	17.63	1.82	12.43	18.09	3.99	17.03	2.94	7.48	0.00	0.00	11.18	0.63
Technical workers	17.50	26.70	15.36	21.87	5.26	34.90	10.76	14.14	17.13	5.70	19.78	29.73	5.56	24.01	4.63	14.93	22.98	7.89	21.36	5.38	16.69	23.83	12.47	18.37	2.87
Clerical svs sales workers	26.56	33.94	56.99	35.66	54.17	28.21	54.81	45.33	40.53	56.71	26.90	33.93	56.19	38.55	47.75	29.81	37.93	45.72	39.27	47.67	29.99	55.13	56.19	38.73	55.49
Craft and trades workers	7.81	4.82	3.11	8.67	13.43	5.80	9.15	15.95	6.82	10.96	9.27	10.55	11.50	6.98	12.44	9.41	8.05	13.28	6.71	14.33	7.37	0.00	0.00	8.46	11.87
Machine operators	14.07	7.41	8.42	7.93	2.74	7.34	2.06	3.58	12.79	2.13	13.67	7.57	9.44	8.78	5.72	11.63	7.69	7.48	9.39	6.68	13.13	6.81	6.96	12.36	4.66
Elementary occupations	23.71	6.75	12.81	6.03	21.69	16.47	9.61	14.71	6.49	21.40	22.43	3.74	15.55	4.04	27.65	21.79	5.26	21.65	6.23	23.00	25.35	14.23	24.39	10.89	24.48
Complexity																									
Communication skill	0.28	0.41	0.36	0.39	0.29	0.29	0.34	0.31	0.35	0.27	0.27	0.39	0.29	0.39	0.25	0.31	0.39	0.28	0.38	0.28	0.26	0.31	0.28	0.34	0.25
Cognoscitive skill	0.28	0.37	0.29	0.36	0.24	0.29	0.30	0.26	0.32	0.22	0.26	0.37	0.24	0.36	0.21	0.28	0.36	0.24	0.35	0.23	0.25	0.26	0.23	0.31	0.19
Manual skill	0.58	0.45	0.47	0.49	0.52	0.60	0.52	0.47	0.55	0.53	0.60	0.49	0.51	0.49	0.54	0.51	0.46	0.53	0.50	0.53	0.60	0.46	0.58	0.54	0.53
Complexity index	3.14	6.77	2.53	5.36	1.86	2.64	6.57	2.75	3.66	1.69	2.49	3.91	1.85	4.93	1.73	4.45	5.83	1.95	4.77	1.88	2.36	3.36	1.24	3.89	1.51

Notes: Sample restricted to only those employed between 18 and 65 years. Pre and post refer to the job the Venezuelan immigrant had in Venezuela (before migrating) and in Peru, respectively. Score measures follow Ottaviano et al. (2013). *Source*: Author's calculations using ENAHO 2018 data, ENPOVE data, O*NET 25.1 Database and crosswalks by Hardy et al. (2018).

Figure 2.A7 – Trends in informality, formal and informal sector inequality, 2005-2019



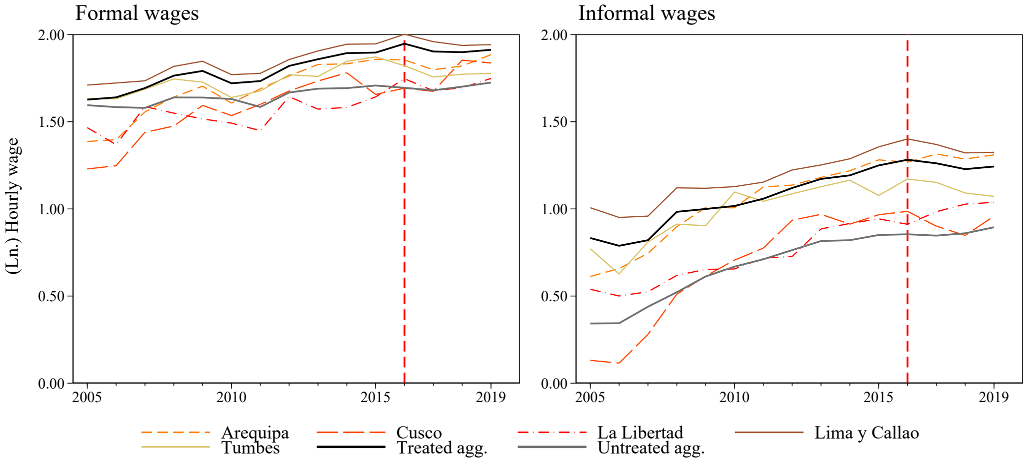
Note: Sample includes employed native individuals between 18 and 65 years. Vertical dashed lines denote the beginning of the treatment. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A4 – Means of covariates by treatment area, 2011-2019

	2011		2012		2013		2014		2015	
Treated regions										
Male (prop.)	0.57	(0.00)	0.57	(0.00)	0.57	(0.00)	0.57	(0.00)	0.57	(0.00)
Age (years)	38.13	(0.13)	38.03	(0.13)	38.18	(0.11)	38.41	(0.12)	38.64	(0.13)
Education (years)	11.08	(0.07)	11.27	(0.07)	11.26	(0.06)	11.26	(0.06)	11.20	(0.06)
Urban area (prop.)	0.93	(0.01)	0.92	(0.01)	0.93	(0.00)	0.93	(0.00)	0.93	(0.00)
Industry (prop.)										
Agriculture	0.06	(0.00)	0.05	(0.00)	0.05	(0.00)	0.05	(0.00)	0.06	(0.00)
Mining	0.03	(0.00)	0.03	(0.00)	0.03	(0.00)	0.03	(0.00)	0.04	(0.00)
Manufacture	0.14	(0.00)	0.15	(0.00)	0.14	(0.00)	0.13	(0.00)	0.13	(0.00)
Construction	0.07	(0.00)	0.08	(0.00)	0.08	(0.00)	0.08	(0.00)	0.08	(0.00)
Retail	0.28	(0.01)	0.28	(0.01)	0.29	(0.00)	0.29	(0.01)	0.27	(0.01)
Transport	0.10	(0.00)	0.10	(0.00)	0.09	(0.00)	0.10	(0.00)	0.10	(0.00)
FIRE	0.09	(0.00)	0.09	(0.00)	0.09	(0.00)	0.09	(0.00)	0.09	(0.00)
Services	0.24	(0.01)	0.23	(0.01)	0.23	(0.01)	0.23	(0.00)	0.23	(0.01)
Occupation (prop.)										
Managers	0.12	(0.00)	0.12	(0.00)	0.12	(0.00)	0.12	(0.00)	0.11	(0.00)
Tehnical	0.16	(0.00)	0.16	(0.00)	0.15	(0.00)	0.15	(0.00)	0.15	(0.00)
Clerical and sales	0.26	(0.01)	0.26	(0.01)	0.28	(0.00)	0.28	(0.00)	0.26	(0.00)
Craft and trades	0.14	(0.00)	0.14	(0.00)	0.14	(0.00)	0.14	(0.00)	0.14	(0.00)
Machine operators	0.09	(0.00)	0.08	(0.00)	0.08	(0.00)	0.08	(0.00)	0.09	(0.00)
Elementary	0.23	(0.01)	0.24	(0.01)	0.23	(0.01)	0.23	(0.00)	0.25	(0.01)
Untreated regions										
Male (prop.)	0.62	(0.00)	0.62	(0.00)	0.62	(0.00)	0.62	(0.00)	0.63	(0.00)
Age (years)	38.37	(0.09)	38.27	(0.09)	38.51	(0.08)	38.66	(0.09)	38.65	(0.09)
Education (years)	9.35	(0.06)	9.55	(0.05)	9.51	(0.05)	9.46	(0.05)	9.53	(0.05)
Urban area (prop.)	0.66	(0.01)	0.66	(0.01)	0.67	(0.01)	0.67	(0.01)	0.68	(0.01)
Industry (prop.)										
Agriculture	0.27	(0.00)	0.26	(0.01)	0.25	(0.00)	0.26	(0.00)	0.26	(0.00)
Mining	0.06	(0.00)	0.06	(0.00)	0.06	(0.00)	0.06	(0.00)	0.07	(0.00)
Manufacture	0.09	(0.00)	0.08	(0.00)	0.09	(0.00)	0.08	(0.00)	0.08	(0.00)
Construction	0.06	(0.00)	0.06	(0.00)	0.07	(0.00)	0.07	(0.00)	0.07	(0.00)
Retail	0.23	(0.00)	0.24	(0.00)	0.24	(0.00)	0.24	(0.00)	0.24	(0.00)
Transport	0.08	(0.00)	0.07	(0.00)	0.08	(0.00)	0.08	(0.00)	0.08	(0.00)
FIRE	0.04	(0.00)	0.04	(0.00)	0.04	(0.00)	0.04	(0.00)	0.03	(0.00)
Services	0.18	(0.00)	0.19	(0.00)	0.19	(0.00)	0.18	(0.00)	0.18	(0.00)
Occupation (prop.)										
Managers	0.09	(0.00)	0.10	(0.00)	0.09	(0.00)	0.09	(0.00)	0.08	(0.00)
Tehnical	0.28	(0.00)	0.28	(0.00)	0.27	(0.00)	0.27	(0.00)	0.27	(0.00)
Clerical and sales	0.20	(0.00)	0.21	(0.00)	0.22	(0.00)	0.21	(0.00)	0.21	(0.00)
Craft and trades	0.10	(0.00)	0.10	(0.00)	0.10	(0.00)	0.09	(0.00)	0.09	(0.00)
Machine operators	0.08	(0.00)	0.08	(0.00)	0.08	(0.00)	0.08	(0.00)	0.08	(0.00)
Elementary	0.25	(0.00)	0.24	(0.00)	0.24	(0.00)	0.26	(0.00)	0.26	(0.00)

Notes: Sample restricted to only those employed natives between 18 and 65 years. SEs in parenthesis. *Source*: Author's calculations using ENAHO 2005-2019 data..

Figure 2.A8 – Trends in (log) hourly wages in the formal and informal sector, 2005-2019



Note: Data aggregated at department and year level. Sample includes employed native individuals between 18 and 65 years. Real hourly wages in 2007 PEN. Vertical dashed lines denote the beginning of the treatment. *Source*: Author's calculations using ENAHO 2005-2019 data.

Results: Key outcomes

Table 2.A5 – 2 stages DiD: ATET of Venezuelan immigration on natives' (log) formal and informal wages (demographic controls only), 2011-2019

		Formal wages							Informal wages						
		$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. IK	P-val. WB	N	$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. IK	P-val. WB	N
Aggregated															
2016-2019	FE OLS	0.016	(0.017)	(0.027)	(0.029)	[0.589]	[0.645]	216	0.015	(0.018)	(0.031)	(0.032)	[0.660]	[0.680]	216
	FE GLS-BC AR(1)	0.013	(0.025)	(0.030)				192	0.014	(0.026)	(0.022)				192
	FE GLS-BC AR(2)	0.012	(0.025)	(0.036)				168	0.007	(0.023)	(0.018)				168
	S*	7.835							7.835						
	IK-DoF	6.846							6.846						
Yearly (Base 2015)															
2016	FE OLS	0.031	(0.035)	(0.032)	(0.033)	[0.379]	[0.395]	216	0.023	(0.039)	(0.023)	(0.024)	[0.371]	[0.380]	216
	FE GLS-BC AR(1)	0.030	(0.031)	(0.035)				192	0.019	(0.027)	(0.028)				192
	FE GLS-BC AR(2)	0.031	(0.033)	(0.037)				168	0.020	(0.026)	(0.028)				168
2017	FE OLS	-0.004	(0.035)	(0.031)	(0.032)	[0.909]	[0.895]	216	0.027	(0.039)	(0.024)	(0.025)	[0.309]	[0.335]	216
	FE GLS-BC AR(1)	-0.005	(0.038)	(0.035)				192	0.022	(0.036)	(0.030)				192
	FE GLS-BC AR(2)	-0.004	(0.035)	(0.037)				168	0.024	(0.032)	(0.027)				168
2018	FE OLS	0.013	(0.035)	(0.050)	(0.051)	[0.804]	[0.815]	216	-0.011	(0.039)	(0.029)	(0.030)	[0.722]	[0.785]	216
	FE GLS-BC AR(1)	0.012	(0.040)	(0.055)				192	-0.018	(0.042)	(0.034)				192
	FE GLS-BC AR(2)	0.013	(0.038)	(0.059)				168	-0.015	(0.035)	(0.030)				168
2019	FE OLS	0.003	(0.035)	(0.043)	(0.044)	[0.952]	[0.955]	216	-0.028	(0.039)	(0.029)	(0.030)	[0.371]	[0.340]	216
	FE GLS-BC AR(1)	0.002	(0.041)	(0.047)				192	-0.036	(0.045)	(0.032)				192
	FE GLS-BC AR(2)	0.002	(0.039)	(0.051)				168	-0.033	(0.037)	(0.029)				168
	S*	7.835							7.835						
	IK-DoF	6.846							6.846						
	F stat. FE OLS	1.707							1.080						
	F stat. GLS-BC AR(1)	0.211							0.394						
	F stat. GLS-BC AR(2)	0.008							0.272						

Notes: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender and area dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the statistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.A6 – SCM: SC Weights for the ATE of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019

	Formal wages					**Informal wages**				
	Areq.	Cusco	La Lib.	L y C.	Tumb.	Areq.	Cusco	La Lib.	L y C.	Tumb.
Amazonas	0	0	0	.137	.218	0	0	0	0	0
Ancash	0	0	0	0	0	0	0	0	0	0
ApurÃmac	0	0	0	0	0	0	0	0	0	0
Ayacucho	0	0	0	0	0	0	0	0	0	0
Cajamarca	0	0	0	0	0	0	0	0	0	0
Huancavelica	0	0	0	0	0	0	.211	0	0	0
HuÃ¡nuco	0	0	0	0	0	0	0	0	0	0
Ica	.053	0	.254	0	0	.042	0	.389	.351	0
JunÃn	0	0	0	0	0	0	0	0	0	0
Lambayeque	0	0	.252	0	0	0	0	.277	0	0
Loreto	0	0	0	0	0	0	0	.282	0	.143
Madre de Dios	0	0	0	.405	.233	.33	0	0	.649	.434
Moquegua	.485	.154	0	.406	0	0	0	0	0	0
Pasco	0	0	0	0	0	0	0	0	0	.101
Piura	0	.239	.054	0	0	0	0	0	0	0
Puno	0	.607	0	0	0	0	.21	0	0	0
San MartÃn	0	0	.198	0	0	0	0	.052	0	0
Tacna	.462	0	0	.051	.428	0	.219	0	0	.322
Ucayali	0	0	.242	0	.122	.628	.36	0	0	0

Notes: In the upper panel the sample for the outcome is restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A7 – SCM: Covariate balance for the ATE of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019

		Arequipa	Cusco	La Lib.	L y C.	Tumbes
Covariates						
GDP pc (thousands)	Obs.	15.924	11.714	9.441	16.380	9.498
	Wei. Formal Wages	31.199	12.254	8.922	27.306	13.050
	Wei. Informal Wages	10.753	8.587	10.393	16.273	15.440
Empl. LF low skill	Obs.	0.536	0.678	0.667	0.551	0.678
	Wei. Formal Wages	0.572	0.681	0.688	0.641	0.674
	Wei. Informal Wages	0.704	0.709	0.666	0.644	0.658
Empl. LF serv.	Obs.	0.644	0.549	0.588	0.734	0.683
	Wei. Formal Wages	0.659	0.515	0.578	0.591	0.613
	Wei. Informal Wages	0.622	0.552	0.595	0.616	0.639
Empl. LF manuf.	Obs.	0.192	0.133	0.195	0.229	0.127
	Wei. Formal Wages	0.152	0.154	0.158	0.126	0.127
	Wei. Informal Wages	0.142	0.133	0.158	0.134	0.120
Schooling (years)	Obs.	11.154	9.211	9.607	11.452	9.883
	Wei. Formal Wages	10.888	9.522	9.752	10.072	9.851
	Wei. Informal Wages	9.705	9.364	10.075	10.312	10.106
Proport. 18-25 yo.	Obs.	0.168	0.154	0.183	0.189	0.177
	Wei. Formal Wages	0.150	0.176	0.179	0.146	0.165
	Wei. Informal Wages	0.170	0.170	0.183	0.167	0.166
Proport. 26-35 yo.	Obs.	0.270	0.264	0.272	0.288	0.293
	Wei. Formal Wages	0.272	0.260	0.271	0.277	0.279
	Wei. Informal Wages	0.282	0.270	0.272	0.279	0.283
Urban population	Obs.	0.881	0.535	0.776	0.979	0.922
	Wei. Formal Wages	0.819	0.598	0.790	0.707	0.720
	Wei. Informal Wages	0.767	0.620	0.799	0.784	0.751
Ecu.-Col. border(d)	Obs.	0.000	0.000	0.000	0.000	1.000
	Wei. Formal Wages	0.000	0.239	0.054	0.137	0.218
	Wei. Informal Wages	0.000	0.000	0.282	0.000	0.143
Lagged Outcomes						
2005	Obs. Formal Wages	1.387	1.229	1.467	1.711	1.634
	Wei. Formal Wages	1.495	1.260	1.472	1.707	1.669
	Obs. Informal Wages	0.612	0.131	0.539	1.007	0.772
	Wei. Informal Wages	0.675	0.242	0.506	0.873	0.675
2010	Obs. Formal Wages	1.607	1.536	1.492	1.771	1.639
	Wei. Formal Wages	1.698	1.608	1.514	1.822	1.698
	Obs. Informal Wages	1.006	0.706	0.655	1.128	1.097
	Wei. Informal Wages	1.001	0.682	0.693	1.153	1.032
2015	Obs. Formal Wages	1.859	1.657	1.643	1.947	1.872
	Wei. Formal Wages	1.786	1.696	1.677	1.839	1.817
	Obs. Informal Wages	1.282	0.967	0.943	1.356	1.077
	Wei. Informal Wages	1.211	0.899	0.970	1.298	1.115

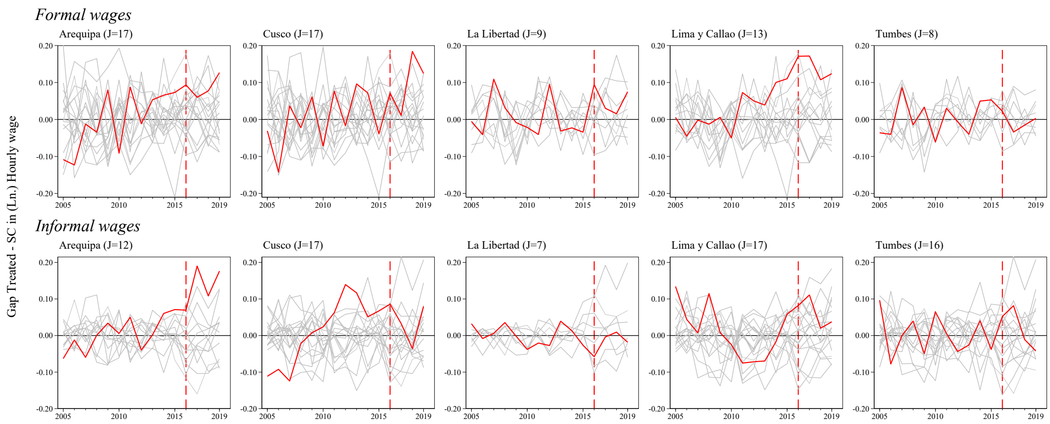
Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. All variables (except the lagged outcomes) are averaged for the 2005-2015 period. LF=Labour force; Ecu-Col border is a dummy that equals 1 if the region neighbours Ecuador and Colombia. Real hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A8 – SCM: DiD Regressions for the ATE of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019

	GLS with AR(1) disturbances					*Baltagi and Wu (1999) estimator*				
	(2005-14)	2016	2017	2018	2019	(2005-14)	2016	2017	2018	2019
Formal Wages										
Arequipa	-0.021	0.031	0.007	0.032	0.088	-0.041	0.039	0.017	0.040	0.093
	(0.124)	(0.125)	(0.167)	(0.194)	(0.214)	(0.109)	(0.116)	(0.143)	(0.155)	(0.162)
Cusco	0.104	0.116	0.061	0.238	0.183	0.084	0.126	0.074	0.252	0.196
	(0.138)	(0.138)	(0.188)	(0.222)	(0.248)	(0.131)	(0.138)	(0.171)	(0.187)	(0.195)
La Libertad	0.031	0.123	0.057	0.042	0.101	0.030	0.123	0.057	0.041	0.101
	(0.102)	(0.119)	(0.135)	(0.139)	(0.141)	(0.079)	(0.092)	(0.104)	(0.108)	(0.109)
Lima y Callao	-0.036	0.077	0.089	0.032	0.053	-0.061	0.079	0.087	0.025	0.042
	(0.086)	(0.089)	(0.113)	(0.125)	(0.133)	(0.071)	(0.081)	(0.094)	(0.098)	(0.099)
Tumbes	-0.028	-0.019	-0.069	-0.047	-0.027	-0.037	-0.020	-0.073	-0.054	-0.035
	(0.094)	(0.102)	(0.123)	(0.132)	(0.136)	(0.074)	(0.087)	(0.099)	(0.102)	(0.103)
Informal Wages										
Arequipa	-0.014	0.002	0.125	0.046	0.117	-0.027	0.011	0.140	0.065	0.138
	(0.119)	(0.119)	(0.165)	(0.199)	(0.227)	(0.119)	(0.121)	(0.156)	(0.177)	(0.190)
Cusco	-0.020	0.023	-0.027	-0.092	0.026	-0.031	0.030	-0.015	-0.077	0.042
	(0.163)	(0.163)	(0.226)	(0.273)	(0.310)	(0.157)	(0.160)	(0.208)	(0.236)	(0.254)
La Libertad	0.040	-0.032	0.022	0.035	0.007	0.039	-0.031	0.024	0.038	0.011
	(0.090)	(0.090)	(0.126)	(0.152)	(0.173)	(0.090)	(0.093)	(0.119)	(0.133)	(0.141)
Lima y Callao	-0.072	0.020	0.045	-0.049	-0.034	-0.064	0.016	0.039	-0.057	-0.042
	(0.091)	(0.091)	(0.126)	(0.152)	(0.172)	(0.086)	(0.088)	(0.114)	(0.128)	(0.138)
Tumbes	0.079	0.090	0.119	0.026	-0.006	0.074	0.093	0.125	0.034	0.003
	(0.134)	(0.135)	(0.184)	(0.218)	(0.244)	(0.117)	(0.120)	(0.154)	(0.173)	(0.184)

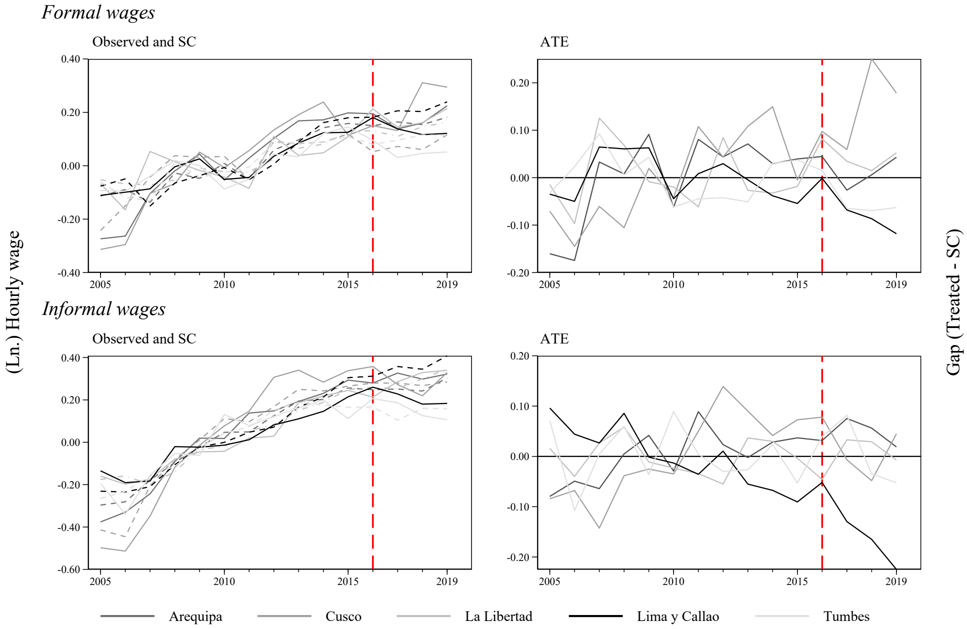
Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. Each row represents a regression of annual observations for the treated areas and its synthetic control between 2005 and 2019, N=30, which includes a dummy variable for the treated region, period dummies (2005-2014, 2016, 2017, 2018, 2019 with 2015 excluded) and their interactions; these interaction coefficients are the coefficients reported. Real hourly wages in 2007 PEN * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.A9 – SCM: Placebos in space of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years before data aggregation. Real hourly wages in 2007 PEN. The dark lines show the gap between the observed and the SC outcomes for the treated regions; the light ones, the gap for only those J regions in the Donor Pool whose RMSPE pre-treatment is lower than 1.5 times that of the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.A10 – SCM demeaned: ATET of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019



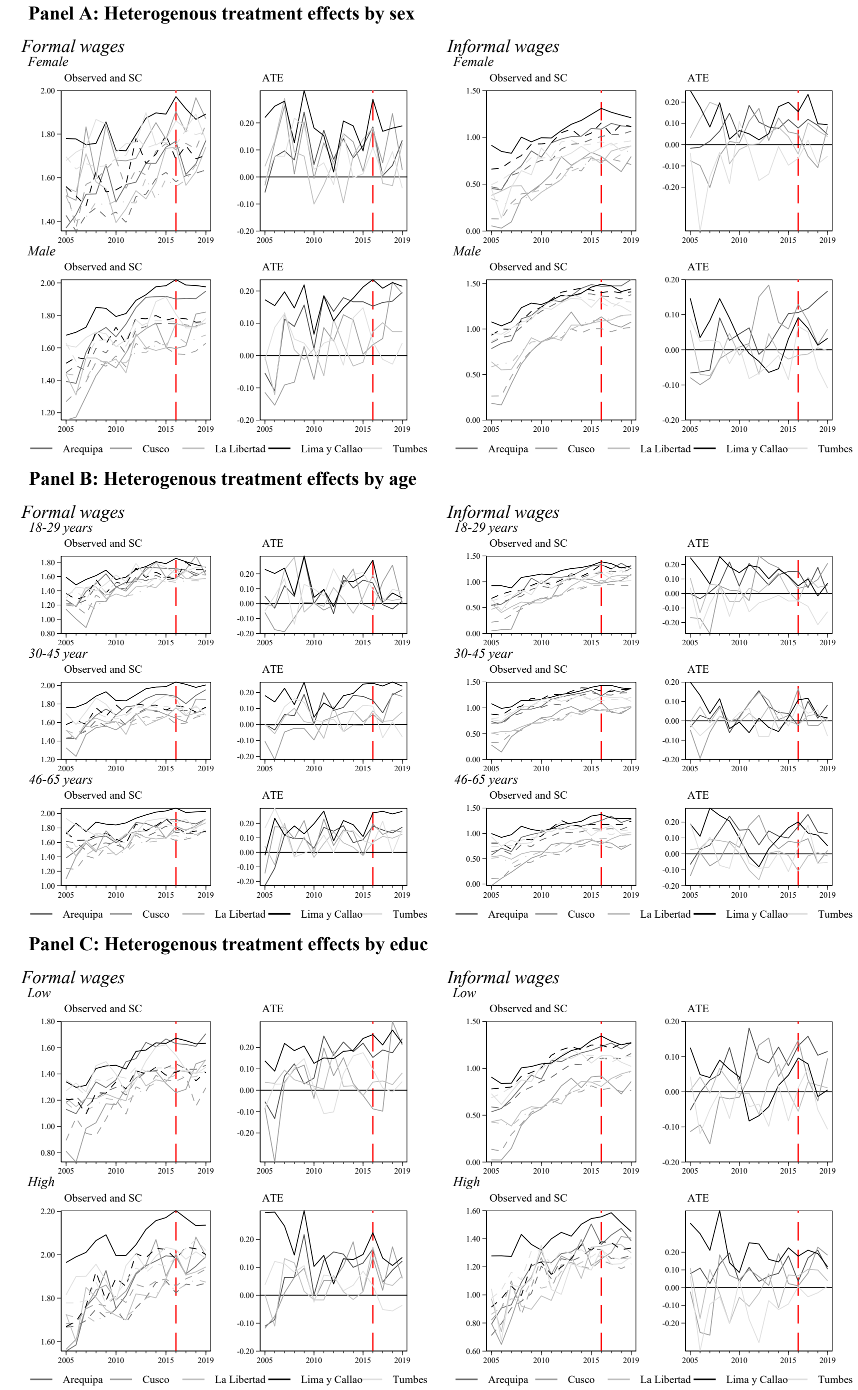
Note: Sample restricted to only those employed between 18 and 65 years before data aggregation. Real hourly wages in 2007 PEN. In all of these, we first applied the demeaning routine suggested by Ferman & Pinto (2019). *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A9 – SCM demeaned: DiD Regressions for the ATET of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019

	GLS with AR(1) disturbances					*Baltagi and Li (1991) estimator*				
	(2005-14)	2016	2017	2018	2019	(2005-14)	2016	2017	2018	2019
Formal Wages										
Arequipa	-0.019	0.013	-0.051	-0.013	0.029	-0.030	0.016	-0.048	-0.012	0.028
	(0.112)	(0.112)	(0.152)	(0.179)	(0.199)	(0.102)	(0.108)	(0.133)	(0.145)	(0.152)
Cusco	0.140	0.115	0.088	0.289	0.227	0.092	0.137	0.118	0.320*	0.254
	(0.126)	(0.127)	(0.173)	(0.204)	(0.228)	(0.122)	(0.130)	(0.160)	(0.173)	(0.180)
La Libertad	0.008	0.093	0.044	0.024	0.061	0.006	0.093	0.043	0.022	0.060
	(0.098)	(0.114)	(0.130)	(0.135)	(0.136)	(0.076)	(0.087)	(0.100)	(0.105)	(0.106)
Lima y Callao	0.019	0.050	-0.019	-0.039	-0.073	0.034	0.044	-0.027	-0.048	-0.081
	(0.086)	(0.086)	(0.116)	(0.136)	(0.150)	(0.075)	(0.081)	(0.098)	(0.106)	(0.109)
Tumbes	-0.016	-0.015	-0.090	-0.091	-0.082	-0.024	-0.014	-0.089	-0.092	-0.085
	(0.088)	(0.090)	(0.116)	(0.130)	(0.139)	(0.071)	(0.078)	(0.094)	(0.100)	(0.103)
Informal Wages										
Arequipa	-0.011	-0.003	0.043	0.026	-0.009	-0.019	0.003	0.052	0.037	0.003
	(0.113)	(0.113)	(0.158)	(0.190)	(0.217)	(0.117)	(0.120)	(0.155)	(0.174)	(0.187)
Cusco	-0.035	0.009	-0.073	-0.111	-0.015	-0.045	0.016	-0.062	-0.097	-0.000
	(0.173)	(0.173)	(0.240)	(0.288)	(0.327)	(0.165)	(0.168)	(0.218)	(0.247)	(0.265)
La Libertad	0.033	-0.041	0.038	0.035	-0.001	0.027	-0.037	0.045	0.043	0.008
	(0.093)	(0.093)	(0.129)	(0.156)	(0.177)	(0.089)	(0.092)	(0.118)	(0.132)	(0.141)
Lima y Callao	0.028	0.034	-0.048	-0.088	-0.151	0.046	0.022	-0.067	-0.110	-0.176
	(0.093)	(0.093)	(0.128)	(0.155)	(0.175)	(0.089)	(0.091)	(0.117)	(0.131)	(0.140)
Tumbes	0.077	0.091	0.133	0.016	-0.002	0.075	0.093	0.136	0.020	0.003
	(0.130)	(0.130)	(0.178)	(0.212)	(0.237)	(0.113)	(0.116)	(0.149)	(0.168)	(0.179)

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. Each row represents a regression of annual observations for the treated areas and its synthetic control between 2005 and 2019, N=30, which includes a dummy variable for the treated region, period dummies (2005-2014, 2016, 2017, 2018, 2019 with 2015 excluded) and their interactions; these interaction coefficients are the coefficients reported. Real hourly wages in 2007 PEN. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.A11 – SCM: ATET of Venezuelan immigration on natives' (log) formal and informal wages across subsamples, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years; the outcome variable is, additionally, restricted to those indicated in the heading of every subgraph before data aggregation. Real hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Results: Ancillary outcomes

Table 2.A10 – SCM: SC Weights for the ATE of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-2019

	Informality rate					**Inequality formal**					**Inequality informal**				
	Areq.	Cusco	La Lib.	L y C.	Tumb.	Areq.	Cusco	La Lib.	L y C.	Tumb.	Areq.	Cusco	La Lib.	L y C.	Tumb.
Amazonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.235
Ancash	0	.094	.585	0	0	0	.014	0	0	0	.488	.763	.287	0	0
ApurÃmac	0	.041	0	0	0	0	0	0	0	.014	0	0	0	0	0
Ayacucho	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cajamarca	0	0	0	0	0	0	0	.101	0	0	0	.21	0	0	0
Huancavelica	0	0	0	0	0	0	.169	0	0	.069	0	0	0	0	0
HuÃ¡nuco	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ica	.319	.109	.104	.3	0	.007	.166	0	.106	.156	.042	0	0	.578	0
JunÃn	0	0	0	0	0	0	0	0	0	.01	0	0	0	0	0
Lambayeque	.337	0	.236	0	.249	.683	0	.243	0	0	0	0	0	0	0
Loreto	0	0	0	0	0	0	0	.048	.17	.357	0	0	0	0	0
Madre de Dios	0	.479	0	.075	.139	0	0	0	0	.168	0	0	0	0	0
Moquegua	.096	0	0	.252	0	.188	.114	.344	.235	0	.47	0	.083	.019	0
Pasco	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Piura	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Puno	0	.276	0	0	0	0	.537	0	0	0	0	0	.496	0	0
San MartÃn	0	0	0	0	0	0	0	.029	0	0	0	0	0	0	0
Tacna	.247	0	0	0	.016	0	0	0	.489	0	0	.027	0	.403	.765
Ucayali	0	0	.076	.373	.596	.122	0	.235	0	.227	0	0	.133	0	0

Notes: In the upper panel the sample for the outcome is restricted to only those employed between 18 and 65 years. Inequality level measured by the Gini index of individual hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A11 – SCM: Covariate balance for the ATE of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-2019

		Arequipa	Cusco	La Lib.	L y C.	Tumbes
Covariates						
GDP pc (thousands)	Obs.	15.924	11.714	9.441	16.380	9.498
	Wei. Inequality Formal	14.536	11.777	20.659	22.111	10.169
	Wei. Inequality Informal	29.657	12.877	11.496	16.576	14.006
	Wei. Informality	15.916	12.618	12.260	20.389	8.680
Empl. LF low skill	Obs.	0.536	0.678	0.667	0.551	0.678
	Wei. Inequality Formal	0.675	0.683	0.666	0.606	0.704
	Wei. Inequality Informal	0.615	0.695	0.686	0.583	0.641
	Wei. Informality	0.620	0.678	0.676	0.634	0.711
Empl. LF serv.	Obs.	0.644	0.549	0.588	0.734	0.683
	Wei. Inequality Formal	0.615	0.484	0.588	0.660	0.588
	Wei. Inequality Informal	0.580	0.524	0.523	0.645	0.641
	Wei. Informality	0.634	0.564	0.576	0.610	0.623
Empl. LF manuf.	Obs.	0.192	0.133	0.195	0.229	0.127
	Wei. Inequality Formal	0.164	0.143	0.156	0.147	0.131
	Wei. Inequality Informal	0.164	0.169	0.159	0.170	0.134
	Wei. Informality	0.167	0.133	0.171	0.162	0.153
Schooling (years)	Obs.	11.154	9.211	9.607	11.452	9.883
	Wei. Inequality Formal	9.832	9.515	9.794	10.647	9.627
	Wei. Inequality Informal	10.177	9.123	9.526	11.017	10.200
	Wei. Informality	10.519	9.768	9.649	10.366	9.608
Proport. 18-25 yo.	Obs.	0.168	0.154	0.183	0.189	0.177
	Wei. Inequality Formal	0.176	0.171	0.160	0.163	0.173
	Wei. Inequality Informal	0.153	0.171	0.175	0.177	0.166
	Wei. Informality	0.176	0.169	0.179	0.165	0.177
Proport. 26-35 yo.	Obs.	0.270	0.264	0.272	0.288	0.293
	Wei. Inequality Formal	0.261	0.259	0.270	0.276	0.282
	Wei. Inequality Informal	0.265	0.264	0.262	0.270	0.276
	Wei. Informality	0.265	0.272	0.262	0.273	0.275
Urban population	Obs.	0.881	0.535	0.776	0.979	0.922
	Wei. Inequality Formal	0.806	0.546	0.732	0.812	0.707
	Wei. Inequality Informal	0.691	0.546	0.582	0.879	0.754
	Wei. Informality	0.849	0.651	0.694	0.809	0.784
Ecu.-Col. border(d)	Obs.	0.000	0.000	0.000	0.000	1.000
	Wei. Inequality Formal	0.000	0.000	0.149	0.170	0.357
	Wei. Inequality Informal	0.000	0.210	0.000	0.000	0.235
	Wei. Informality	0.000	0.000	0.000	0.000	0.000
Lagged Outcomes						
2005	Obs. Inequality Formal	0.442	0.454	0.393	0.507	0.362
	Wei. Inequality Formal	0.444	0.446	0.406	0.479	0.342
	Obs. Inequality Informal	0.497	0.500	0.507	0.433	0.454
	Wei. Inequality Informal	0.492	0.475	0.479	0.414	0.465
	Obs. Informality	0.608	0.823	0.779	0.696	0.826
	Wei. Informality	0.617	0.803	0.779	0.665	0.826
2010	Obs. Inequality Formal	0.427	0.398	0.454	0.472	0.352
	Wei. Inequality Formal	0.426	0.379	0.442	0.423	0.368
	Obs. Inequality Informal	0.503	0.537	0.466	0.442	0.498
	Wei. Inequality Informal	0.465	0.509	0.470	0.438	0.489
	Obs. Informality	0.550	0.774	0.668	0.549	0.730
	Wei. Informality	0.546	0.752	0.674	0.544	0.730
2015	Obs. Inequality Formal	0.412	0.351	0.417	0.436	0.345
	Wei. Inequality Formal	0.411	0.367	0.407	0.421	0.395
	Obs. Inequality Informal	0.448	0.438	0.464	0.394	0.408
	Wei. Inequality Informal	0.457	0.454	0.462	0.382	0.429
	Obs. Informality	0.522	0.695	0.631	0.495	0.679
	Wei. Informality	0.535	0.695	0.634	0.505	0.679

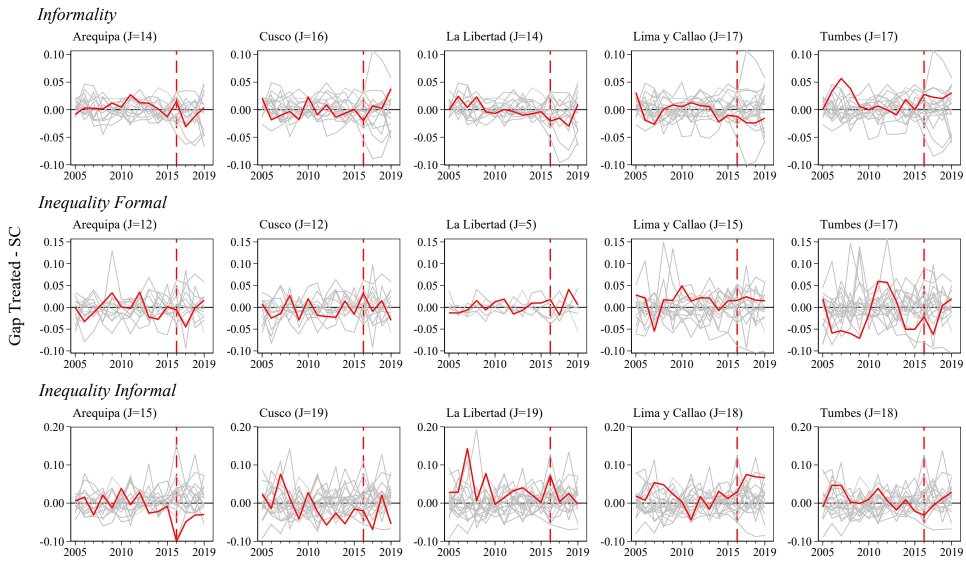
Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. All variables (except the lagged outcomes) are averaged for the 2005-2015 period. LF=Labour force; Ecu-Col border is a dummy that equals 1 if the region neighbours Ecuador and Colombia. Inequality level measured by the Gini index of individual hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A12 – SCM: DiD Regressions for the ATE of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-2019

	GLS with AR(1) disturbances					*Baltagi and Wu (1999) estimator*				
	(2005-14)	2016	2017	2018	2019	(2005-14)	2016	2017	2018	2019
1 Informality										
Arequipa	0.014	0.028	-0.018	0.002	0.017	0.015	0.028	-0.019	0.000	0.016
	(0.023)	(0.023)	(0.032)	(0.038)	(0.043)	(0.021)	(0.021)	(0.028)	(0.031)	(0.033)
Cusco	-0.006	-0.022	0.004	-0.001	0.033	-0.005	-0.022	0.004	-0.001	0.034
	(0.048)	(0.048)	(0.065)	(0.076)	(0.084)	(0.042)	(0.044)	(0.055)	(0.060)	(0.063)
La Libertad	-0.003	-0.017	-0.012	-0.027	0.012	-0.001	-0.019	-0.015	-0.030	0.009
	(0.041)	(0.041)	(0.057)	(0.068)	(0.076)	(0.040)	(0.041)	(0.052)	(0.057)	(0.061)
Lima y Callao	-0.010	-0.004	-0.017	-0.019	-0.011	-0.004	-0.008	-0.023	-0.025	-0.018
	(0.044)	(0.044)	(0.060)	(0.072)	(0.082)	(0.045)	(0.047)	(0.059)	(0.065)	(0.069)
Tumbes	0.018	0.028	0.023	0.022	0.033	0.017	0.029	0.024	0.023	0.034
	(0.036)	(0.036)	(0.050)	(0.060)	(0.068)	(0.033)	(0.034)	(0.044)	(0.050)	(0.053)
2 Inequality formal										
Arequipa	-0.020	-0.013	-0.055	-0.013	0.003	-0.016	-0.014	-0.055	-0.013	0.004
	(0.035)	(0.036)	(0.045)	(0.050)	(0.052)	(0.027)	(0.030)	(0.036)	(0.038)	(0.039)
Cusco	0.017	0.052	0.012	0.037	-0.008	0.013	0.050	0.009	0.033	-0.012
	(0.035)	(0.040)	(0.046)	(0.048)	(0.049)	(0.027)	(0.034)	(0.037)	(0.037)	(0.037)
La Libertad	-0.009	0.009	-0.026	0.033	-0.003	-0.009	0.009	-0.027	0.032	-0.004
	(0.055)	(0.068)	(0.073)	(0.074)	(0.074)	(0.042)	(0.055)	(0.057)	(0.057)	(0.057)
Lima y Callao	-0.010	-0.004	0.003	-0.006	-0.008	-0.008	-0.003	0.004	-0.004	-0.006
	(0.038)	(0.042)	(0.050)	(0.052)	(0.053)	(0.029)	(0.034)	(0.039)	(0.040)	(0.041)
Tumbes	0.019	0.020	-0.024	0.041	0.055	0.028	0.023	-0.018	0.048	0.063
	(0.063)	(0.070)	(0.083)	(0.088)	(0.090)	(0.051)	(0.062)	(0.068)	(0.069)	(0.069)
3 Inequality informal										
Arequipa	0.012	-0.090	-0.039	-0.021	-0.020	0.010	-0.091*	-0.040	-0.022	-0.021
	(0.046)	(0.065)	(0.062)	(0.062)	(0.062)	(0.036)	(0.049)	(0.048)	(0.048)	(0.048)
Cusco	0.010	-0.003	-0.052	0.038	-0.037	0.017	0.006	-0.045	0.045	-0.030
	(0.051)	(0.070)	(0.069)	(0.069)	(0.069)	(0.040)	(0.059)	(0.053)	(0.054)	(0.054)
La Libertad	0.045	0.081	0.007	0.031	0.003	0.045	0.080	0.007	0.031	0.002
	(0.045)	(0.074)	(0.058)	(0.066)	(0.063)	(0.035)	(0.057)	(0.045)	(0.050)	(0.048)
Lima y Callao	0.010	0.021	0.068	0.063	0.061	0.008	0.021	0.067*	0.062*	0.059
	(0.034)	(0.037)	(0.044)	(0.047)	(0.048)	(0.026)	(0.030)	(0.035)	(0.036)	(0.037)
Tumbes	0.035	-0.010	0.015	0.034	0.050	0.035	-0.010	0.016	0.034	0.050
	(0.055)	(0.075)	(0.074)	(0.074)	(0.074)	(0.043)	(0.058)	(0.057)	(0.057)	(0.057)

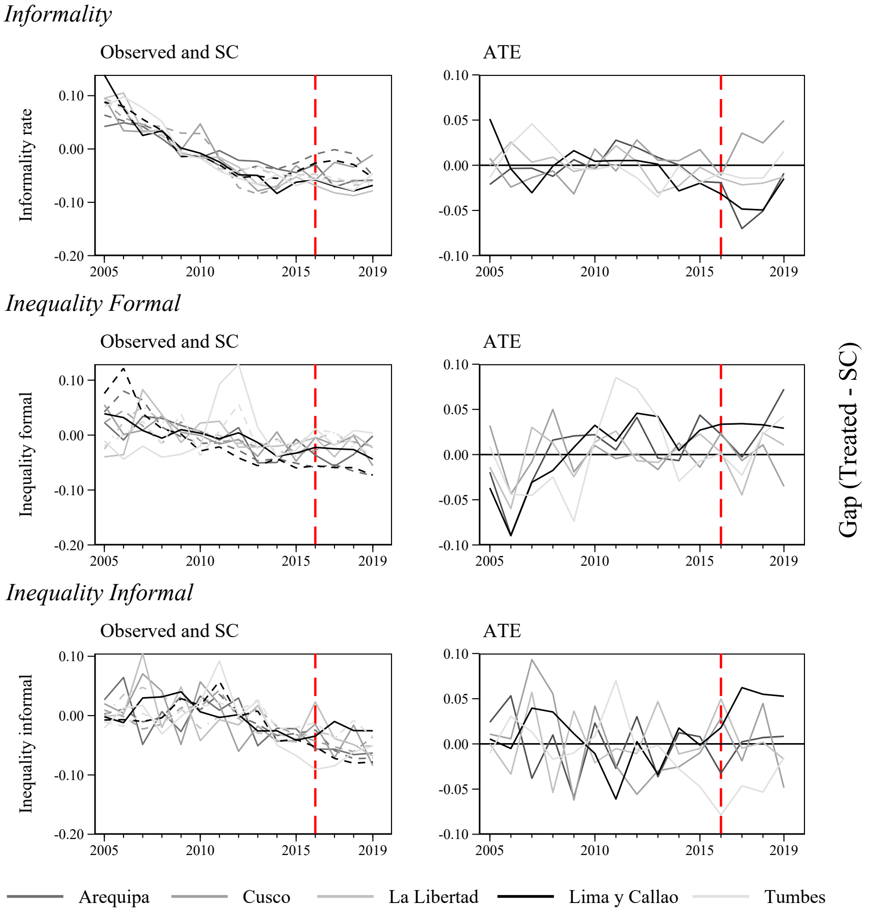
Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. Each row represents a regression of annual observations for the corresponding treated area and its synthetic control between 2005 and 2019, N=30, which includes a dummy variable for the treated region, period dummies (2005-2014, 2016, 2017, 2018, 2019 with 2015 excluded) and their interactions; these interaction coefficients are the coefficients reported. Inequality level measured by the Gini index of individual hourly wages in 2007 PEN. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.A12 – SCM: Placebos in space of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years before data aggregation. Real hourly wages in 2007 PEN. The dark lines show the gap between the observed and tl those J regions in the Donor Pool whose RMSPE pre-treatment is lower than 2 times that of the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.A13 – SCM demeaned: ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-201



Note: Sample restricted to only those employed between 18 and 65 years before data aggregation. Real hourly wages
In all of these, we first applied the demeaning routine suggested by Ferman & Pinto (2019). *Source*: Author's calcula
ENAHO 2005-2019 data.

Table 2.A13 – SCM demeaned: DiD Regressions for the ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-2019

	GLS with AR(1) disturbances					*Baltagi and Wu (1999) estimator*				
	(2005-14)	2016	2017	2018	2019	(2005-14)	2016	2017	2018	2019
1 Informality										
Arequipa	0.018	-0.001	-0.051	-0.031	0.011	0.018	-0.001	-0.051*	-0.031	0.010
	(0.025)	(0.025)	(0.035)	(0.042)	(0.047)	(0.023)	(0.024)	(0.030)	(0.034)	(0.036)
Cusco	-0.013	-0.028	0.019	0.008	0.033	-0.014	-0.027	0.020	0.010	0.035
	(0.051)	(0.051)	(0.069)	(0.081)	(0.090)	(0.045)	(0.048)	(0.059)	(0.065)	(0.068)
La Libertad	-0.019	-0.013	-0.022	-0.021	-0.015	-0.014	-0.016	-0.027	-0.027	-0.021
	(0.041)	(0.041)	(0.056)	(0.067)	(0.075)	(0.038)	(0.040)	(0.050)	(0.056)	(0.059)
Lima y Callao	-0.006	-0.015	-0.034	-0.037	-0.004	0.003	-0.019	-0.041	-0.045	-0.012
	(0.043)	(0.043)	(0.059)	(0.071)	(0.080)	(0.043)	(0.045)	(0.056)	(0.062)	(0.066)
Tumbes	0.018	0.011	0.005	0.006	0.035	0.018	0.011	0.005	0.005	0.035
	(0.037)	(0.037)	(0.051)	(0.061)	(0.069)	(0.034)	(0.034)	(0.045)	(0.050)	(0.054)
2 Inequality formal										
Arequipa	-0.051	-0.021	-0.046	-0.013	0.029	-0.050	-0.021	-0.047	-0.014	0.028
	(0.039)	(0.039)	(0.051)	(0.059)	(0.064)	(0.032)	(0.034)	(0.042)	(0.045)	(0.047)
Cusco	0.021	0.040	0.010	0.028	-0.018	0.019	0.039	0.010	0.027	-0.019
	(0.034)	(0.039)	(0.045)	(0.047)	(0.048)	(0.027)	(0.032)	(0.036)	(0.036)	(0.037)
La Libertad	-0.022	-0.018	-0.065	0.004	-0.009	-0.023	-0.018	-0.065	0.004	-0.010
	(0.047)	(0.055)	(0.063)	(0.065)	(0.066)	(0.037)	(0.043)	(0.049)	(0.050)	(0.051)
Lima y Callao	-0.025	0.009	0.011	0.011	0.008	-0.027	0.009	0.011	0.010	0.007
	(0.040)	(0.041)	(0.054)	(0.063)	(0.069)	(0.035)	(0.038)	(0.046)	(0.050)	(0.052)
Tumbes	-0.002	0.002	-0.024	0.021	0.041	0.008	0.008	-0.016	0.030	0.049
	(0.067)	(0.078)	(0.089)	(0.092)	(0.093)	(0.054)	(0.073)	(0.073)	(0.073)	(0.073)
3 Inequality informal										
Arequipa	-0.011	-0.042	-0.007	-0.002	-0.001	-0.009	-0.040	-0.006	-0.001	0.000
	(0.045)	(0.064)	(0.061)	(0.061)	(0.061)	(0.035)	(0.048)	(0.048)	(0.048)	(0.048)
Cusco	0.010	0.038	-0.010	0.054	-0.039	0.009	0.037	-0.010	0.053	-0.040
	(0.052)	(0.069)	(0.070)	(0.070)	(0.070)	(0.040)	(0.053)	(0.054)	(0.054)	(0.054)
La Libertad	0.007	0.056	0.001	0.008	-0.011	0.004	0.054	-0.001	0.005	-0.013
	(0.055)	(0.077)	(0.073)	(0.074)	(0.074)	(0.043)	(0.054)	(0.058)	(0.058)	(0.058)
Lima y Callao	0.009	0.023	0.069	0.062	0.060	0.007	0.023	0.068*	0.061	0.059
	(0.035)	(0.039)	(0.046)	(0.048)	(0.049)	(0.027)	(0.032)	(0.036)	(0.037)	(0.038)
Tumbes	0.052	-0.031	0.001	-0.006	0.033	0.052	-0.031	0.001	-0.006	0.032
	(0.044)	(0.069)	(0.069)	(0.069)	(0.069)	(0.034)	(0.047)	(0.046)	(0.046)	(0.046)

Table 2.A14 – SCM demeaned: P-values for the ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-2019

	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
1 Informality					
Arequipa	-0.0372	0.0142	3.1236	2/20	0.1000
Cusco	0.0247	0.0176	1.9010	7/20	0.3500
La Libertad	-0.0171	0.0158	1.1094	10/20	0.5000
Lima y Callao	-0.0361	0.0215	1.8007	6/20	0.3000
Tumbes	-0.0054	0.0216	0.6142	19/20	0.9500
2 Inequality Formal					
Arequipa	0.0306	0.0358	1.1422	4/20	0.2000
Cusco	-0.0018	0.0246	0.8998	8/20	0.4000
La Libertad	-0.0018	0.0252	1.0277	5/20	0.2500
Lima y Callao	0.0325	0.0389	0.8366	10/20	0.5000
Tumbes	0.0108	0.0481	0.5555	16/20	0.8000
3 Inequality Informal					
Arequipa	-0.0037	0.0333	0.5146	18/20	0.9000
Cusco	0.0015	0.0456	0.8149	12/20	0.6000
La Libertad	0.0074	0.0323	0.8164	12/20	0.6000
Lima y Callao	0.0470	0.0275	1.8210	5/20	0.2500
Tumbes	-0.0483	0.0296	1.8044	4/20	0.2000

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

Robustness checks

Table 2.A15 – Panel data: ATET of Venezuelan immigration on natives' (log) formal and informal wages (full set of controls), 2005-2019

		Formal wages							**Informal wages**						
		$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. I-K	P-val. WB	N	$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. I-K	P-val. WB	N
Aggregated															
2016-2019	FE OLS	0.048	(0.026)*	(0.033)	(0.035)	[0.199]	[0.200]	360	0.015	(0.020)	(0.026)	(0.028)	[0.606]	[0.655]	360
	FD	0.026	(0.039)	(0.025)	(0.026)	[0.347]	[0.315]	336	0.012	(0.029)	(0.023)	(0.024)	[0.632]	[0.625]	336
	FE GLS AR(1)	0.058		(0.030)*			[0.035]	360	0.009		(0.016)			[0.440]	360
	FE GLS AR(2)	0.055		(0.030)*			[0.035]	360	0.008		(0.017)			[0.440]	360
	S* FE	8.753							8.753						
	S* FD	8.032							8.032						
	IK-DoF FE	7.783							7.783						
	IK-DoF FD	6.955							6.952						
Yearly (Base 2015)															
2016	FE OLS	0.030	(0.058)	(0.030)	(0.031)	[0.368]	[0.380]	360	0.004	(0.045)	(0.032)	(0.033)	[0.906]	[0.930]	360
	FD	0.028	(0.039)	(0.026)	(0.026)	[0.323]	[0.310]	336	0.012	(0.029)	(0.024)	(0.025)	[0.652]	[0.660]	336
	FE GLS AR(1)	0.027		(0.027)			[0.360]	360	0.007		(0.027)			[0.925]	360
	FE GLS AR(2)	0.029		(0.027)			[0.360]	360	0.007		(0.027)			[0.925]	360
2017	FE OLS	0.001	(0.058)	(0.040)	(0.042)	[0.986]	[0.985]	360	0.004	(0.045)	(0.025)	(0.026)	[0.889]	[0.920]	360
	FD	0.001	(0.056)	(0.036)	(0.038)	[0.987]	[0.980]	336	0.022	(0.042)	(0.024)	(0.023)	[0.386]	[0.410]	336
	FE GLS AR(1)	0.001		(0.038)			[0.975]	360	0.016		(0.023)			[0.930]	360
	FE GLS AR(2)	0.002		(0.037)			[0.975]	360	0.016		(0.024)			[0.930]	360
2018	FE OLS	0.002	(0.058)	(0.055)	(0.058)	[0.980]	[1.000]	360	-0.038	(0.045)	(0.024)	(0.024)	[0.158]	[0.190]	360
	FD	0.018	(0.068)	(0.056)	(0.058)	[0.771]	[0.860]	336	-0.026	(0.051)	(0.031)	(0.032)	[0.448]	[0.525]	336
	FE GLS AR(1)	0.011		(0.055)			[0.980]	360	-0.032		(0.028)			[0.420]	360
	FE GLS AR(2)	0.011		(0.055)			[0.980]	360	-0.032		(0.028)			[0.420]	360
2019	FE OLS	0.003	(0.058)	(0.044)	(0.046)	[0.945]	[0.915]	360	-0.034	(0.045)	(0.027)	(0.028)	[0.266]	[0.230]	360
	FD	0.017	(0.079)	(0.047)	(0.049)	[0.733]	[0.715]	336	-0.033	(0.059)	(0.029)	(0.029)	[0.304]	[0.230]	336
	FE GLS AR(1)	0.010		(0.045)			[0.930]	360	-0.037		(0.027)			[0.550]	360
	FE GLS AR(2)	0.011		(0.045)			[0.930]	360	-0.037		(0.027)			[0.550]	360
	S* FE	7.956							7.956						
	S* FD	8.003							8.003						
	IK-DoF FE	6.968							6.968						
	IK-DoF FD	6.992							6.999						
	F stat. FE OLS	2.792**							3.433***						
	F stat. FD	2.043*							3.516***						
	F stat. GLS AR(1)	31.899***							36.794***						
	F stat. GLS AR(2)	33.883***							37.584***						

Notes: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Controls included are gdp per capita, pea employed for low sckilled, pea employed in services, in manufacture, schooling, proportion of population between 18 and 25 and between 26 and 35, percentage of urban population. Equations include 2-way FEs (regions and years). SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the F test cluster robust of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A16 – Panel data random growth: ATET of Venezuelan immigration on natives' (log) formal and informal wages (full set of controls), 2005-2019

		Formal wages							Informal wages						
		$\hat{\beta}$	SE Homosk.	SE CRVE	SE CR2	P-val. I-K	P-val. WB	N	$\hat{\beta}$	SE Homosk.	SE CRVE	SE CR2	P-val. I-K	P-val. WB	N
All															
Aggregated															
2016-2019	FE	0.012	(0.040)	(0.027)	(0.028)	[0.693]	[0.670]	336	0.014	(0.030)	(0.029)	(0.030)	[0.661]	[0.670]	336
	FD	0.036	(0.045)	(0.026)	(0.026)	[0.207]	[0.160]	312	0.008	(0.032)	(0.039)	(0.040)	[0.853]	[0.855]	312
	Eff. G (CSS)	9.229							9.229						
	Eff. DoF (I-K)	7.047							7.047						
Yearly (Base 2015)															
2016	FE	0.006	(0.040)	(0.029)	(0.029)	[0.841]	[0.820]	336	0.010	(0.030)	(0.029)	(0.030)	[0.761]	[0.745]	336
	FD	0.003	(0.122)	(0.029)	(0.029)	[0.922]	[0.910]	312	0.011	(0.089)	(0.029)	(0.030)	[0.725]	[0.725]	312
2017	FE	-0.043	(0.060)	(0.040)	(0.041)	[0.332]	[0.355]	336	0.016	(0.045)	(0.039)	(0.040)	[0.695]	[0.740]	336
	FD	-0.045	(0.251)	(0.043)	(0.045)	[0.346]	[0.335]	312	0.020	(0.183)	(0.040)	(0.041)	[0.639]	[0.700]	312
2018	FE	-0.048	(0.076)	(0.057)	(0.059)	[0.445]	[0.445]	336	-0.033	(0.057)	(0.052)	(0.054)	[0.565]	[0.700]	336
	FD	-0.044	(0.391)	(0.060)	(0.062)	[0.501]	[0.450]	312	-0.028	(0.285)	(0.054)	(0.056)	[0.632]	[0.700]	312
2019	FE	-0.070	(0.091)	(0.055)	(0.056)	[0.253]	[0.265]	336	-0.043	(0.068)	(0.052)	(0.054)	[0.448]	[0.525]	336
	FD	-0.064	(0.540)	(0.058)	(0.061)	[0.325]	[0.275]	312	-0.041	(0.393)	(0.054)	(0.056)	[0.493]	[0.555]	312
	Eff. G (CSS)	7.835							7.835						
	Eff. DoF (I-K)	6.731							6.772						
	F stat. FE	1.694							2.570**						
	F stat. FD	1.460							2.585**						
Low-Skilled															
Aggregated															
2016-2019	FE	-0.053	(0.069)	(0.045)	(0.046)	[0.285]	[0.220]	336	-0.011	(0.046)	(0.048)	(0.051)	[0.835]	[0.830]	336
	FD	-0.046	(0.079)	(0.061)	(0.061)	[0.473]	[0.425]	312	0.015	(0.051)	(0.054)	(0.056)	[0.794]	[0.810]	312
	Eff. G (CSS)	9.229							9.229						
	Eff. DoF (I-K)	7.047							7.047						
Yearly (Base 2015)															
2016	FE	-0.055	(0.071)	(0.046)	(0.047)	[0.276]	[0.210]	336	-0.019	(0.046)	(0.048)	(0.050)	[0.719]	[0.715]	336
	FD	-0.061	(0.219)	(0.049)	(0.050)	[0.258]	[0.200]	312	-0.019	(0.141)	(0.047)	(0.048)	[0.700]	[0.715]	312
2017	FE	-0.076	(0.104)	(0.064)	(0.065)	[0.284]	[0.225]	336	-0.061	(0.069)	(0.068)	(0.070)	[0.416]	[0.520]	336
	FD	-0.080	(0.452)	(0.068)	(0.071)	[0.299]	[0.235]	312	-0.057	(0.291)	(0.069)	(0.071)	[0.446]	[0.510]	312
2018	FE	-0.050	(0.133)	(0.076)	(0.078)	[0.544]	[0.565]	336	-0.087	(0.087)	(0.088)	(0.092)	[0.377]	[0.570]	336
	FD	-0.048	(0.703)	(0.080)	(0.084)	[0.585]	[0.525]	312	-0.085	(0.453)	(0.089)	(0.091)	[0.381]	[0.530]	312
2019	FE	-0.085	(0.159)	(0.086)	(0.087)	[0.364]	[0.295]	336	-0.116	(0.105)	(0.072)	(0.075)	[0.168]	[0.120]	336
	FD	-0.079	(0.971)	(0.089)	(0.094)	[0.430]	[0.335]	312	-0.119	(0.626)	(0.074)	(0.076)	[0.165]	[0.110]	312
	Eff. G (CSS)	7.835							7.835						
	Eff. DoF (I-K)	6.721							6.760						
	F stat. FE	2.149*							1.038						
	F stat. FD	1.755							1.113						

Notes: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Controls included are gdp per capita, pea employed for low sckilled, pea employed in services, in manufacture, schooling, proportion of population between 18 and 25 and between 26 and 35, percentage of urban population. Equations include 2-way FEs (regions and years). SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the F test cluster robust of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A17 – Dynamic Panel data: ATET of Venezuelan immigration on natives' (log) formal and informal wages (full set of controls), 2005-2019

		Formal wages				**Informal wages**			
		2 step GMM		**Iterated GMM**		**2 step GMM**		**Iterated GMM**	
		$\hat{\beta}$	SE CRVE	$\hat{\beta}$	SE CRVE	$\hat{\beta}$	SE CRVE	$\hat{\beta}$	SE CRVE
All									
Aggregated									
Treatment	2016-2019	-0.005	(0.053)	0.017	(0.118)	0.039	(0.041)	0.038	(0.044)
Lagged outcome		0.304*	(0.166)	0.389	(0.233)	0.602***	(0.093)	0.328	(0.337)
AR2		2.418**	[0.016]	1.147	[0.252]	0.053	[0.957]	-0.212	[0.832]
J-test		5.197	[0.268]	5.959	[0.202]	2.370	[0.668]	0.419	[0.981]
Number of instr.		25		25		25		25	
Number of obs.		336		336		336		336	
Yearly (Base 2015)									
Treatment	2016	0.509	(0.408)	0.491	(0.955)	1.838	(2.483)	-0.490	(1.534)
	2017	0.878	(0.731)	0.812	(1.727)	3.607	(4.920)	-1.038	(3.120)
	2018	1.300	(1.046)	1.252	(2.469)	1.947	(2.639)	-1.597	(4.273)
	2019	1.750	(1.378)	1.243	(2.601)	0.335	(0.404)	0.801	(1.983)
Lagged outcome		0.189	(0.426)	0.764	(1.482)	0.317	(0.326)	0.887	(1.962)
AR2		1.819*	[0.069]	0.328	[0.743]	-0.730	[0.465]	-0.384	[0.701]
J-test		0.000	[1.000]	0.000	[1.000]	0.000	[1.000]	0.000	[1.000]
Number of instr.		28		28		28		28	
Number of obs.		336		336		336		336	
Low-Skilled									
Aggregated									
Treatment	2016-2019	-0.083	(0.095)	-0.069	(0.196)	0.024	(0.076)	0.051	(0.157)
Lagged outcome		0.288*	(0.152)	0.338	(0.443)	0.457***	(0.162)	-0.033	(0.271)
AR2		0.814	[0.416]	-0.015	[0.988]	1.659*	[0.097]	1.192	[0.233]
J-test		3.081	[0.544]	0.381	[0.984]	1.913	[0.752]	3.531	[0.473]
Number of instr.		25		25		25		25	
Number of obs.		336		336		336		336	
Yearly (Base 2015)									
Treatment	2016	0.102	(0.442)	0.098	(0.152)	0.824	(0.722)	0.397	(1.068)
	2017	0.501	(0.946)	0.122	(0.217)	0.375	(0.478)	0.664	(1.971)
	2018	0.320	(0.830)	0.298	(0.180)	0.595	(0.739)	0.715	(2.048)
	2019	0.137	(0.771)	0.253	(0.222)	0.750	(0.942)	0.711	(2.068)
Lagged outcome		0.399*	(0.204)	0.816	(0.491)	0.767	(0.498)	-0.023	(0.925)
AR2		-0.664	[0.507]	1.960*	[0.050]	1.301	[0.193]	0.194	[0.846]
J-test		0.000	[1.000]	0.000	[1.000]	0.000	[1.000]	0.000	[1.000]
Number of instr.		28		28		28		28	
Number of obs.		336		336		336		336	

Notes: Sample restricted to only those employed between 18 and 65 years and for low skilled, for those with at most secondary school. Real hourly wages in 2007 PEN. Controls included are gdp per capita, pea employed for low sckilled, pea employed in services, in manufacture, schooling, proportion of population between 18 and 25 and between 26 and 35, percentage of urban population. Equations include 2-way FEs (regions and years). SE CRVE refers to Cluster Robust Variance Estimator of SEs. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A18 – Panel data: ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector (full set of controls), 2005-2019

		Informality rate					**Formal sector inequality**					**Informal sector inequality**				
		β	SE CRVE	SE CRVE2	P-val. WB	N	β	SE CRVE	SE CRVE2	P-val. WB	N	β	SE CRVE	SE CRVE2	P-val. WB	N
Aggregated																
2016-2019	FE OLS	-0.004	(0.011)	(0.011)	[0.730]	360	0.013	(0.013)	(0.013)	[0.315]	360	-0.008	(0.010)	(0.011)	[0.465]	360
	FD	0.008	(0.009)	(0.009)	[0.360]	336	0.031	(0.012)**	(0.012)**	[0.010]	336	0.018	(0.016)	(0.017)	[0.300]	336
	FE GLS AR(1)	0.004	(0.008)		[0.430]	360	0.015	(0.012)		[0.245]	360	-0.005	(0.009)		[0.315]	360
	FE GLS AR(2)	0.003	(0.008)		[0.430]	360	0.015	(0.012)		[0.245]	360	-0.005	(0.009)		[0.315]	360
	S* FE	8.753					8.753					8.753				
	S* FD	8.032					8.032					8.032				
	IK-DoF FE	7.783					7.783					7.783				
	IK-DoF FD	6.954					6.946					6.946				
Yearly (Base 2015)																
2016	FE OLS	0.007	(0.009)	(0.009)	[0.460]	360	0.029	(0.016)*	(0.016)*	[0.060]	360	0.018	(0.019)	(0.020)	[0.365]	360
	FD	0.008	(0.008)	(0.009)	[0.335]	336	0.032	(0.012)**	(0.012)**	[0.005]	336	0.018	(0.016)	(0.017)	[0.305]	336
	FE GLS AR(1)	0.008	(0.008)		[0.430]	360	0.029	(0.015)*		[0.080]	360	0.018	(0.018)		[0.415]	360
	FE GLS AR(2)	0.008	(0.008)		[0.430]	360	0.029	(0.015)*		[0.080]	360	0.018	(0.018)		[0.415]	360
2017	FE OLS	0.001	(0.010)	(0.010)	[0.955]	360	0.003	(0.017)	(0.017)	[0.885]	360	0.011	(0.012)	(0.012)	[0.365]	360
	FD	-0.001	(0.009)	(0.009)	[0.910]	336	0.008	(0.013)	(0.013)	[0.635]	336	0.009	(0.013)	(0.013)	[0.505]	336
	FE GLS AR(1)	-0.000	(0.009)		[0.895]	360	0.004	(0.015)		[0.875]	360	0.010	(0.012)		[0.475]	360
	FE GLS AR(2)	-0.000	(0.009)		[0.895]	360	0.004	(0.015)		[0.875]	360	0.010	(0.012)		[0.475]	360
2018	FE OLS	-0.003	(0.009)	(0.009)	[0.715]	360	0.034	(0.013)**	(0.013)**	[0.010]	360	0.034	(0.015)**	(0.015)**	[0.015]	360
	FD	-0.004	(0.008)	(0.008)	[0.640]	336	0.038	(0.013)***	(0.013)***	[0.015]	336	0.034	(0.015)**	(0.015)**	[0.020]	336
	FE GLS AR(1)	-0.003	(0.008)		[0.760]	360	0.034	(0.013)**		[0.050]	360	0.034	(0.015)**		[0.060]	360
	FE GLS AR(2)	-0.003	(0.008)		[0.760]	360	0.034	(0.013)**		[0.050]	360	0.033	(0.015)**		[0.060]	360
2019	FE OLS	0.024	(0.011)**	(0.011)**	[0.065]	360	0.034	(0.014)**	(0.014)**	[0.020]	360	0.020	(0.013)	(0.013)	[0.190]	360
	FD	0.023	(0.011)*	(0.012)*	[0.100]	336	0.040	(0.018)**	(0.018)**	[0.040]	336	0.024	(0.015)	(0.015)	[0.135]	336
	FE GLS AR(1)	0.024	(0.010)**		[0.050]	360	0.035	(0.014)**		[0.125]	360	0.020	(0.013)		[0.200]	360
	FE GLS AR(2)	0.024	(0.010)**		[0.050]	360	0.035	(0.014)**		[0.125]	360	0.020	(0.013)		[0.200]	360
	S* FE	7.956					7.956					7.956				
	S* FD	8.003					8.003					8.003				
	IK-DoF FE	6.968					6.968					6.968				
	IK-DoF FD	6.990					7.014					7.016				
	F stat. FE OLS	3.409***					1.970*					2.816**				
	F stat. FD	4.305***					2.069*					2.422**				
	F stat. GLS AR(1)	50.246***					24.100***					31.981***				
	F stat. GLS AR(2)	49.659***					24.246***					31.938***				

Notes: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Controls included are gdp per capita, pea employed for low sckilled, pea employed in services, in manufacture, schooling, proportion of population between 18 and 25 and between 26 and 35, percentage of urban population. Equations include 2-way FEs (regions and years). SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the F test cluster robust of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A19 – Panel data random growth: ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector (full set of controls), 2005-2019

		Informality rate					Formal sector inequality					Informal sector inequality				
		β	SE CRVE	SE CR2	P-val. WB	N	β	SE CRVE	SE CR2	P-val. WB	N	β	SE CRVE	SE CR2	P-val. WB	N
All																
Aggregated																
2016-2019	FE	0.009	(0.010)	(0.010)	[0.330]	336	0.031	(0.013)**	(0.013)**	[0.020]	336	0.019	(0.018)	(0.019)	[0.320]	336
	FD	0.013	(0.014)	(0.014)	[0.330]	312	0.045	(0.019)**	(0.020)**	[0.035]	312	0.026	(0.021)	(0.022)	[0.240]	312
	Eff. G (CSS)	9.229					9.229					9.229				
	Eff. DoF (I-K)	7.047					7.047					7.047				
Yearly (Base 2015)																
2016	FE	0.011	(0.009)	(0.009)	[0.230]	336	0.033	(0.013)**	(0.013)**	[0.010]	336	0.020	(0.017)	(0.017)	[0.255]	336
	FD	0.011	(0.009)	(0.009)	[0.195]	312	0.033	(0.013)**	(0.012)**	[0.005]	312	0.019	(0.016)	(0.016)	[0.250]	312
2017	FE	0.005	(0.010)	(0.010)	[0.675]	336	0.010	(0.015)	(0.015)	[0.515]	336	0.013	(0.014)	(0.014)	[0.365]	336
	FD	0.005	(0.010)	(0.010)	[0.620]	312	0.009	(0.014)	(0.014)	[0.515]	312	0.010	(0.014)	(0.014)	[0.465]	312
2018	FE	0.004	(0.009)	(0.009)	[0.665]	336	0.042	(0.016)**	(0.016)**	[0.015]	336	0.041	(0.016)**	(0.015)**	[0.015]	336
	FD	0.004	(0.009)	(0.009)	[0.600]	312	0.041	(0.016)**	(0.016)**	[0.010]	312	0.038	(0.016)**	(0.015)**	[0.020]	312
2019	FE	0.034	(0.012)***	(0.012)***	[0.015]	336	0.045	(0.019)**	(0.018)**	[0.040]	336	0.033	(0.016)**	(0.015)**	[0.050]	336
	FD	0.034	(0.012)***	(0.012)**	[0.010]	312	0.045	(0.021)**	(0.021)**	[0.020]	312	0.030	(0.017)*	(0.017)*	[0.065]	312
	Eff. G (CSS)	7.835					7.835					7.835				
	Eff. DoF (I-K)	6.718					6.767					6.793				
	F stat. FE	2.727**					1.950*					1.825				
	F stat. FD	2.811**					2.126*					1.971*				
Low-Skilled																
Aggregated																
2016-2019	FE	0.013	(0.011)	(0.011)	[0.295]	336	0.035	(0.030)	(0.031)	[0.265]	336	0.002	(0.047)	(0.050)	[0.995]	336
	FD	0.018	(0.017)	(0.018)	[0.335]	312	0.023	(0.039)	(0.040)	[0.615]	312	0.006	(0.066)	(0.070)	[0.930]	312
	Eff. G (CSS)	9.229					9.229					9.229				
	Eff. DoF (I-K)	7.047					7.047					7.047				
Yearly (Base 2015)																
2016	FE	0.015	(0.010)	(0.011)	[0.200]	336	0.036	(0.030)	(0.030)	[0.250]	336	0.000	(0.048)	(0.050)	[0.995]	336
	FD	0.014	(0.009)	(0.010)	[0.175]	312	0.036	(0.030)	(0.030)	[0.250]	312	-0.002	(0.050)	(0.052)	[0.975]	312
2017	FE	0.006	(0.013)	(0.013)	[0.780]	336	0.045	(0.032)	(0.032)	[0.175]	336	-0.029	(0.045)	(0.047)	[0.750]	336
	FD	0.006	(0.014)	(0.014)	[0.775]	312	0.046	(0.034)	(0.034)	[0.175]	312	-0.032	(0.048)	(0.050)	[0.760]	312
2018	FE	0.009	(0.011)	(0.011)	[0.480]	336	0.035	(0.035)	(0.035)	[0.350]	336	0.022	(0.050)	(0.052)	[0.650]	336
	FD	0.009	(0.011)	(0.012)	[0.395]	312	0.038	(0.036)	(0.036)	[0.260]	312	0.018	(0.053)	(0.054)	[0.710]	312
2019	FE	0.033	(0.017)*	(0.017)*	[0.090]	336	0.030	(0.038)	(0.039)	[0.455]	336	-0.012	(0.052)	(0.054)	[0.910]	336
	FD	0.034	(0.019)*	(0.020)*	[0.080]	312	0.033	(0.040)	(0.042)	[0.360]	312	-0.013	(0.055)	(0.058)	[0.910]	312
	Eff. G (CSS)	7.835					7.835					7.835				
	Eff. DoF (I-K)	6.736					6.745					6.769				
	F stat. FE	1.872					3.206**					1.901				
	F stat. FD	1.947*					3.507***					2.123*				

Notes: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Controls included are gdp per capita, pea employed for low sckilled, pea employed in services, in manufacture, schooling, proportion of population between 18 and 25 and between 26 and 35, percentage of urban population. Equations include 2-way FEs (regions and years). SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the F test cluster robust of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A20 – Dynamic Panel data: ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector (full set of controls), 2005-2019

		Informality				**Formal wages**				**Informal wages**			
		2-step GMM		**Iterated GMM**		**2-step GMM**		**Iterated GMM**		**2-step GMM**		**Iterated GMM**	
		$\hat{\beta}$	SE CRVE	$\hat{\beta}$	SE CRVE	$\hat{\beta}$	SE CRVE	$\hat{\beta}$	SE CRVE	$\hat{\beta}$	SE CRVE	$\hat{\beta}$	SE CRVE
All													
Aggregated													
Treatment	2016-2019	0.010	(10.436)	0.017	(0.037)	0.022	(0.044)	0.055	(0.112)	0.025	(0.020)	0.021	(0.162)
Lagged outcome		0.415	(1153.147)	-0.152	(0.672)	0.339	(0.582)	0.488	(4.240)	0.237**	(0.110)	0.030	(0.938)
AR2		0.002	[0.999]	0.278	[0.781]	-0.619	[0.536]	0.031	[0.975]	0.021	[0.984]	0.500	[0.617]
J-test		6.734	[0.241]	3.453	[0.485]	3.366	[0.499]	3.312	[0.507]	0.605	[0.963]	3.896	[0.420]
Number of instr.		25		25		25		25		25		25	
Number of obs.		336		336		336		336		336		336	
Yearly (Base 2015)													
Treatment	2016	-0.282	(0.205)	-0.383	(0.401)	-0.416	(0.270)	-0.416	(2.681)	0.005	(0.023)	-0.266	(0.303)
	2017	-0.549	(0.384)	-0.760	(0.772)	-0.445*	(0.245)	-0.445	(4.304)	0.049	(0.048)	-0.533	(0.566)
	2018	-0.777	(0.544)	-0.742	(0.764)	-1.046	(0.649)	-1.046	(6.634)	0.074	(0.048)	-0.698	(0.630)
	2019	-0.758	(0.550)	-0.716	(0.783)	-1.607	(1.018)	-1.607	(9.291)	0.175	(0.153)	-0.724	(0.615)
Lagged outcome		2.191*	(1.262)	1.852	(2.326)	-1.477	(1.478)	-1.477	(5.870)	-1.567	(1.654)	0.159	(2.305)
AR2		1.685*	[0.092]	0.356	[0.722]	-0.679	[0.497]	-0.211	[0.833]	-0.504	[0.615]	-0.457	[0.647]
J-test		0.000	[1.000]	0.000	[1.000]	0.000	[1.000]	0.000	[1.000]	0.000	[1.000]	0.000	[1.000]
Number of instr.		28		28		28		28		28		28	
Number of obs.		336		336		336		336		336		336	
Low-Skilled													
Aggregated													
Treatment	2016-2019	0.012	(0.012)	0.020	(0.026)	0.012	(0.036)	0.008	(0.050)	-0.005	(0.054)	-0.019	(0.058)
Lagged outcome		0.419	(0.316)	-0.423	(0.890)	-0.002	(0.102)	0.003	(0.088)	0.043	(0.180)	0.159	(0.174)
AR2		1.699*	[0.089]	-0.406	[0.684]	-0.314	[0.753]	-0.247	[0.805]	0.647	[0.517]	0.582	[0.561]
J-test		3.217	[0.522]	1.067	[0.899]	0.220	[0.994]	0.313	[0.989]	1.143	[0.887]	1.878	[0.758]
Number of instr.		25		25		25		25		25		25	
Number of obs.		336		336		336		336		336		336	
Yearly (Base 2015)													
Treatment	2016	-0.320	(0.274)	0.027	(0.129)	-0.022	(0.126)	-0.002	(0.097)	-0.284	(0.320)	-0.171	(0.253)
	2017	-0.659	(0.534)	-0.059	(0.250)	-0.045	(0.223)	0.005	(0.093)	-0.535	(0.569)	-0.418	(0.372)
	2018	-0.905	(0.752)	0.031	(0.351)	-0.110	(0.309)	-0.055	(0.109)	-0.672	(0.786)	-0.432	(0.631)
	2019	-1.271	(0.875)	0.113	(0.475)	-0.138	(0.410)	-0.060	(0.113)	-0.933	(1.059)	-0.583	(0.816)
Lagged outcome		3.201*	(1.858)	-0.265	(1.839)	0.007	(0.113)	-0.078	(0.583)	0.484	(0.560)	0.310	(0.373)
AR2		2.227**	[0.026]	-0.822	[0.411]	0.033	[0.974]	-0.465	[0.642]	2.348**	[0.019]	-0.196	[0.844]
J-test		0.000	[1.000]	0.000	[1.000]	0.000	[1.000]	0.000	[1.000]	0.000	[1.000]	0.000	[1.000]
Number of instr.		28		28		28		28		28		28	
Number of obs.		336		336		336		336		336		336	

Notes: Sample restricted to only those employed between 18 and 65 years and for low skilled, for those with at most secondary school. Real hourly wages in 2007 PEN. Controls included are gdp per capita, pea employed for low sckilled, pea employed in services, in manufacture, schooling, proportion of population between 18 and 25 and between 26 and 35, percentage of urban population. Equations include 2-way FEs (regions and years). SE CRVE refers to Cluster Robust Variance Estimator of SEs. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A21 – SCM matching only on lagged outcomes: ATET of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019

	All					**Low-skilled**				
	ATE	RMSPE pre	Ratio post-pre	Rank	p-value	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
1 Formal Wage										
Arequipa										
All years	0.065	0.003	1.332	10/20	0.500	0.101	0.007	1.257	12/20	0.600
Even years only	0.062	0.005	1.011	11/20	0.550	0.148	0.015	1.218	10/20	0.500
Odd years only	0.069	0.003	1.324	9/20	0.450	0.046	0.012	0.969	15/20	0.750
Cusco										
All years	0.093	0.006	1.508	8/20	0.400	0.069	0.016	1.131	12/20	0.600
Even years only	0.082	0.007	1.302	6/20	0.300	0.075	0.017	1.152	11/20	0.550
Odd years only	0.123	0.006	1.698	5/20	0.250	0.052	0.012	1.693	8/20	0.400
La Libertad										
All years	0.092	0.001	2.743	3/20	0.150	0.113	0.001	4.493	3/20	0.150
Even years only	0.088	0.003	1.770	4/20	0.200	0.108	0.004	1.816	6/20	0.300
Odd years only	0.124	0.002	2.608	2/20	0.100	0.125	0.002	2.741	3/20	0.150
Lima Y Callao										
All years	0.028	0.000	5.274	1/20	0.050	0.054	0.001	2.369	8/20	0.400
Even years only	0.059	0.003	1.267	7/20	0.350	0.123	0.004	2.053	7/20	0.350
Odd years only	0.054	0.000	3.223	1/20	0.050	0.070	0.002	1.939	6/20	0.300
Tumbes										
All years	-0.019	0.001	1.666	7/20	0.350	-0.061	0.004	1.492	12/20	0.600
Even years only	-0.024	0.003	0.545	19/20	0.950	-0.029	0.007	0.697	17/20	0.850
Odd years only	-0.047	0.002	1.142	10/20	0.500	0.025	0.011	0.711	19/20	0.950
2 Informal Wage										
Arequipa										
All years	0.054	0.001	2.514	8/20	0.400	0.061	0.002	2.191	8/20	0.400
Even years only	0.057	0.001	1.880	10/20	0.500	0.084	0.003	1.784	9/20	0.450
Odd years only	0.006	0.001	0.370	19/20	0.950	0.006	0.002	1.623	7/20	0.350
Cusco										
All years	-0.036	0.003	1.132	13/20	0.650	-0.036	0.012	0.761	19/20	0.950
Even years only	-0.026	0.004	0.993	16/20	0.800	-0.006	0.014	0.515	20/20	1.000
Odd years only	-0.040	0.003	1.204	13/20	0.650	-0.036	0.012	0.776	17/20	0.850
La Libertad										
All years	0.021	0.001	1.300	13/20	0.650	0.007	0.001	1.491	12/20	0.600
Even years only	-0.017	0.001	1.005	16/20	0.800	0.008	0.001	1.635	10/20	0.500
Odd years only	0.015	0.001	1.181	15/20	0.750	0.013	0.003	1.041	13/20	0.650
Lima Y Callao										
All years	-0.064	0.001	2.381	8/20	0.400	-0.097	0.000	9.575	1/20	0.050
Even years only	0.006	0.002	1.460	12/20	0.600	-0.043	0.001	2.769	3/20	0.150
Odd years only	-0.075	0.002	2.106	6/20	0.300	-0.090	0.000	5.347	1/20	0.050
Tumbes										
All years	-0.056	0.002	1.637	9/20	0.450	-0.035	0.001	2.775	5/20	0.250
Even years only	-0.078	0.004	1.550	10/20	0.500	-0.106	0.001	3.542	1/20	0.050
Odd years only	-0.036	0.003	1.208	14/20	0.700	-0.013	0.002	1.775	6/20	0.300

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. Estimations under the low skilled heading, in the case of the formal outcome, also restrict to those with at most primary or secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.A22 – SCM matching only on lagged outcomes: ATET of Venezuelan immigration on natives' informality rate and inequality in formal and informal sector, 2005-2019

	All					**Low-skilled**				
	ATE	RMSPE pre	Ratio post-pre	Rank	p-value	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
3 Informality										
Arequipa										
All years	0.021	0.000	6.290	4/20	0.200	-0.007	0.000	2.297	8/20	0.400
Even years only	-0.005	0.000	2.355	10/20	0.500	-0.014	0.000	1.358	12/20	0.600
Odd years only	0.029	0.000	5.061	3/20	0.150	0.013	0.000	1.842	8/20	0.400
Cusco										
All years	0.009	0.000	1.462	15/20	0.750	0.032	0.000	4.113	3/20	0.150
Even years only	0.005	0.000	1.011	17/20	0.850	0.049	0.000	4.981	1/20	0.050
Odd years only	0.015	0.000	1.559	13/20	0.650	0.034	0.000	3.702	3/20	0.150
La Libertad										
All years	-0.013	0.000	2.479	11/20	0.550	-0.009	0.000	1.408	14/20	0.700
Even years only	-0.020	0.000	2.936	7/20	0.350	-0.016	0.000	1.828	8/20	0.400
Odd years only	-0.013	0.000	2.117	11/20	0.550	-0.016	0.000	2.237	4/20	0.200
Lima Y Callao										
All years	-0.008	0.000	1.183	16/20	0.800	0.008	0.000	1.353	16/20	0.800
Even years only	-0.014	0.000	1.333	15/20	0.750	0.016	0.000	1.702	10/20	0.500
Odd years only	-0.008	0.000	1.113	15/20	0.750	0.001	0.000	0.629	18/20	0.900
Tumbes										
All years	-0.012	0.000	2.574	11/20	0.550	0.003	0.000	1.389	15/20	0.750
Even years only	-0.010	0.000	1.775	13/20	0.650	-0.011	0.000	1.106	16/20	0.800
Odd years only	-0.007	0.000	1.465	13/20	0.650	-0.001	0.000	0.865	17/20	0.850
4 Inequality Formal										
Arequipa										
All years	0.006	0.000	1.627	10/20	0.500	0.024	0.000	1.824	8/20	0.400
Even years only	0.017	0.001	1.161	10/20	0.500	0.023	0.001	0.827	13/20	0.650
Odd years only	0.004	0.000	1.123	15/20	0.750	0.023	0.000	1.622	4/20	0.200
Cusco										
All years	0.016	0.000	1.578	12/20	0.600	0.023	0.002	0.688	19/20	0.950
Even years only	-0.011	0.001	0.626	18/20	0.900	-0.015	0.003	0.285	19/20	0.950
Odd years only	0.033	0.001	1.389	10/20	0.500	0.055	0.002	1.176	9/20	0.450
La Libertad										
All years	0.014	0.000	4.192	2/20	0.100	-0.006	0.000	2.625	5/20	0.250
Even years only	-0.002	0.000	1.795	4/20	0.200	-0.000	0.001	0.621	16/20	0.800
Odd years only	0.006	0.000	1.943	6/20	0.300	0.008	0.001	1.206	9/20	0.450
Lima Y Callao										
All years	0.019	0.000	2.467	6/20	0.300	0.020	0.000	4.089	2/20	0.100
Even years only	0.033	0.001	1.248	9/20	0.450	0.019	0.000	2.008	6/20	0.300
Odd years only	0.005	0.000	0.997	16/20	0.800	0.041	0.001	1.731	4/20	0.200
Tumbes										
All years	0.014	0.001	0.871	17/20	0.850	0.053	0.000	2.945	3/20	0.150
Even years only	0.016	0.002	0.864	12/20	0.600	0.065	0.001	2.085	5/20	0.250
Odd years only	0.007	0.003	0.339	18/20	0.900	0.073	0.001	2.884	2/20	0.100
5 Inequality Informal										
Arequipa										
All years	-0.047	0.001	2.397	8/20	0.400	-0.030	0.004	0.668	19/20	0.950
Even years only	-0.045	0.001	1.675	8/20	0.400	-0.021	0.003	0.932	14/20	0.700
Odd years only	-0.056	0.001	2.820	5/20	0.250	-0.034	0.004	0.879	16/20	0.800
Cusco										
All years	0.004	0.001	1.010	18/20	0.900	-0.004	0.002	1.336	15/20	0.750
Even years only	0.004	0.002	0.884	16/20	0.800	-0.016	0.003	0.971	13/20	0.650
Odd years only	0.007	0.001	1.105	12/20	0.600	0.004	0.003	1.260	11/20	0.550
La Libertad										
All years	0.026	0.001	1.384	13/20	0.650	-0.006	0.002	0.920	18/20	0.900
Even years only	0.008	0.001	1.293	12/20	0.600	0.006	0.003	0.621	17/20	0.850
Odd years only	0.022	0.001	1.513	11/20	0.550	0.015	0.003	0.752	15/20	0.750
Lima Y Callao										
All years	0.016	0.000	3.146	5/20	0.250	-0.006	0.000	1.414	11/20	0.550
Even years only	0.014	0.000	1.334	11/20	0.550	-0.049	0.001	2.450	5/20	0.250
Odd years only	0.027	0.000	2.166	9/20	0.450	0.017	0.000	1.159	14/20	0.700
Tumbes										
All years	-0.006	0.000	1.560	13/20	0.650	0.009	0.000	1.534	11/20	0.550
Even years only	-0.014	0.001	1.363	10/20	0.500	0.001	0.001	1.241	9/20	0.450
Odd years only	-0.007	0.000	1.546	10/20	0.500	0.015	0.001	1.444	10/20	0.500

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. Estimations under the low skilled heading, in the case of the formal outcome, also restrict to those with at most primary or secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.A23 – SCM regression adjustment: ATET of Venezuelan immigration on natives' (log) formal and informal wages, 2005-2019

	All					**Low-skilled**				
	ATE	RMSPE pre	Ratio post-pre	Rank	p-value	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
1 Formal Wage										
Arequipa	0.079	0.009	0.854	8/20	0.400	0.112	0.009	1.188	7/20	0.350
Cusco	0.110	0.014	1.023	7/20	0.350	0.119	0.015	1.362	4/20	0.200
La Libertad	0.072	0.003	1.492	3/20	0.150	0.061	0.005	0.893	9/20	0.450
Lima y Callao	0.142	0.012	1.320	4/20	0.200	0.143	0.012	1.316	4/20	0.200
Tumbes	0.058	0.005	0.905	8/20	0.400	0.066	0.009	0.832	10/20	0.500
2 Informal Wage										
Arequipa	0.110	0.004	1.722	3/20	0.150	0.062	0.008	1.158	4/20	0.200
Cusco	0.059	0.005	1.027	9/20	0.450	0.006	0.014	0.490	17/20	0.850
La Libertad	0.039	0.001	1.674	4/20	0.200	0.035	0.004	1.156	5/20	0.250
Lima y Callao	0.017	0.002	0.842	13/20	0.650	-0.008	0.002	1.580	3/20	0.150
Tumbes	-0.038	0.005	0.814	12/20	0.600	-0.059	0.004	1.730	1/20	0.050

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years before data aggregation. Estimations under the low skilled heading, in the case of the formal outcome, also restrict to those with at most primary or secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. M1 to M3 refer to the SC estimator based on residuals aggregated on department-year level that result from regressing the outcome on interactions of year with gender, age, schooling and its interaction, area (M1), including aditionally industry group (M2) and also occupation group (M3). ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.A24 – 2 stages DiD: ATET of Venezuelan immigration on natives' (log) formal and informal wages under alternative wage definitions (full set of controls), 2011-2019

		Monthly wages						**CPI deflactation**					
		$\hat{\beta}$	SE CRVE	SE CR2	P-val. I-K	P-val. WB	N	$\hat{\beta}$	SE CRVE	SE CR2	P-val. I-K	P-val. WB	N
Formal													
Aggregated													
2016-2019	FE OLS	0.020	(0.019)	(0.020)	[0.349]	[0.340]	216	0.013	(0.024)	(0.024)	[0.608]	[0.675]	216
	FE GLS-BC AR(1)	0.019	(0.022)				192	0.008	(0.023)				192
	FE GLS-BC AR(2)	0.021	(0.024)				168	0.010	(0.029)				168
	Eff. G (CSS)	7.835						7.835					
	Eff. DoF (I-K)	6.846						6.846					
Yearly (Base 2015)													
2016	FE OLS	0.030	(0.023)	(0.024)	[0.256]	[0.265]	216	0.027	(0.023)	(0.024)	[0.298]	[0.295]	216
	FE GLS-BC AR(1)	0.030	(0.026)				192	0.026	(0.026)				192
	FE GLS-BC AR(2)	0.030	(0.026)				168	0.027	(0.027)				168
2017	FE OLS	0.001	(0.022)	(0.023)	[0.981]	[0.970]	216	-0.006	(0.025)	(0.025)	[0.827]	[0.835]	216
	FE GLS-BC AR(1)	0.000	(0.024)				192	-0.007	(0.027)				192
	FE GLS-BC AR(2)	0.001	(0.024)				168	-0.006	(0.028)				168
2018	FE OLS	0.037	(0.036)	(0.037)	[0.353]	[0.275]	216	0.011	(0.040)	(0.041)	[0.793]	[0.770]	216
	FE GLS-BC AR(1)	0.037	(0.040)				192	0.010	(0.044)				192
	FE GLS-BC AR(2)	0.037	(0.040)				168	0.011	(0.046)				168
2019	FE OLS	0.002	(0.029)	(0.029)	[0.943]	[0.940]	216	0.004	(0.031)	(0.032)	[0.914]	[0.915]	216
	FE GLS-BC AR(1)	0.002	(0.032)				192	0.003	(0.034)				192
	FE GLS-BC AR(2)	0.002	(0.032)				168	0.004	(0.036)				168
	Eff. G (CSS)	7.835						7.835					
	Eff. DoF (I-K)	6.846						6.846					
	F stat. FE OLS	0.396						1.707					
	F stat. GLS-BC AR(1)	0.419						0.210					
	F stat. GLS-BC AR(2)	0.619						0.008					
Informal													
Aggregated													
2016-2019	FE OLS	0.043	(0.025)	(0.025)*	[0.130]	[0.085]	216	0.016	(0.030)	(0.031)	[0.623]	[0.635]	216
	FE GLS-BC AR(1)	0.029	(0.038)				192	0.019	(0.023)				192
	FE GLS-BC AR(2)	0.023	(0.038)				168	0.009	(0.019)				168
	Eff. G (CSS)	7.835						7.835					
	Eff. DoF (I-K)	6.846						6.846					
Yearly (Base 2015)													
2016	FE OLS	0.032	(0.035)	(0.037)	[0.412]	[0.410]	216	0.029	(0.024)	(0.025)	[0.293]	[0.320]	216
	FE GLS-BC AR(1)	0.030	(0.042)				192	0.026	(0.029)				192
	FE GLS-BC AR(2)	0.029	(0.043)				168	0.026	(0.029)				168
2017	FE OLS	0.041	(0.036)	(0.037)	[0.309]	[0.355]	216	0.024	(0.025)	(0.025)	[0.366]	[0.390]	216
	FE GLS-BC AR(1)	0.037	(0.047)				192	0.019	(0.031)				192
	FE GLS-BC AR(2)	0.036	(0.045)				168	0.020	(0.028)				168
2018	FE OLS	0.038	(0.041)	(0.043)	[0.403]	[0.470]	216	-0.011	(0.032)	(0.033)	[0.745]	[0.795]	216
	FE GLS-BC AR(1)	0.033	(0.055)				192	-0.017	(0.038)				192
	FE GLS-BC AR(2)	0.032	(0.052)				168	-0.016	(0.034)				168
2019	FE OLS	0.014	(0.036)	(0.037)	[0.722]	[0.740]	216	-0.030	(0.028)	(0.028)	[0.322]	[0.290]	216
	FE GLS-BC AR(1)	0.008	(0.048)				192	-0.037	(0.032)				192
	FE GLS-BC AR(2)	0.007	(0.046)				168	-0.035	(0.029)				168
	Eff. G (CSS)	7.835						7.835					
	Eff. DoF (I-K)	6.846						6.846					
	F stat. FE OLS	0.383						1.078					
	F stat. GLS-BC AR(1)	0.292						0.394					
	F stat. GLS-BC AR(2)	0.129						0.259					

Notes: Sample restricted to only those employed between 18 and 65 years. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender, area, industry and occupation dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CR2, to Bell and McCaffrey (2002) bias corrected CRVE CR2 and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. Eff. G (CSS) refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the statistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.A25 – 2 stages DiD: ATET of Venezuelan immigration on natives' outcomes under alternative definition of informality (full set of controls), 2011-2019

		Informal wage					**Informality**					**Informal sector inequality**				
		$\hat{\beta}$	SE CRVE	SE CR2	P-val. WB	N	$\hat{\beta}$	SE CRVE	SE CR2	P-val. WB	N	$\hat{\beta}$	SE CRVE	SE CR2	P-val. WB	N
Aggregated																
2016-2019	FE OLS	0.008	(0.020)	(0.020)	[0.715]	216	-0.004	(0.012)	(0.012)	[0.690]	216	-0.013	(0.024)	(0.025)	[0.570]	216
	FE GLS-BC AR(1)	0.028	(0.025)			192	0.001	(0.010)			192	-0.001	(0.022)			192
	FE GLS-BC AR(2)	0.012	(0.022)			168	0.009	(0.012)			168	0.009	(0.018)			168
	Eff. G (CSS)	7.835					7.835					7.835				
	Eff. DoF (I-K)	6.846					6.846					6.846				
Yearly (Base 2015)																
2016	FE OLS	0.042	(0.035)	(0.036)	[0.215]	216	0.003	(0.009)	(0.009)	[0.710]	216	0.007	(0.034)	(0.034)	[0.880]	216
	FE GLS-BC AR(1)	0.043	(0.039)			192	0.004	(0.010)			192	0.007	(0.037)			192
	FE GLS-BC AR(2)	0.043	(0.040)			168	0.004	(0.010)			168	0.008	(0.038)			168
2017	FE OLS	0.013	(0.028)	(0.028)	[0.660]	216	-0.003	(0.013)	(0.013)	[0.885]	216	0.020	(0.025)	(0.025)	[0.410]	216
	FE GLS-BC AR(1)	0.014	(0.031)			192	-0.002	(0.014)			192	0.020	(0.027)			192
	FE GLS-BC AR(2)	0.013	(0.031)			168	-0.002	(0.014)			168	0.022	(0.028)			168
2018	FE OLS	0.007	(0.029)	(0.029)	[0.820]	216	-0.011	(0.013)	(0.013)	[0.405]	216	0.008	(0.032)	(0.032)	[0.880]	216
	FE GLS-BC AR(1)	0.009	(0.033)			192	-0.010	(0.015)			192	0.008	(0.035)			192
	FE GLS-BC AR(2)	0.007	(0.034)			168	-0.010	(0.015)			168	0.010	(0.035)			168
2019	FE OLS	0.049	(0.032)	(0.032)	[0.130]	216	0.011	(0.011)	(0.011)	[0.315]	216	-0.004	(0.038)	(0.039)	[0.910]	216
	FE GLS-BC AR(1)	0.050	(0.037)			192	0.012	(0.012)			192	-0.004	(0.042)			192
	FE GLS-BC AR(2)	0.049	(0.037)			168	0.012	(0.012)			168	-0.002	(0.044)			168
	Eff. G (CSS)	7.835					7.835					7.835				
	Eff. DoF (I-K)	6.846					6.846					6.846				
	F stat. FE OLS	2.534*					5.403***					0.653				
	F stat. GLS-BC AR(1)	2.774*					3.664**					0.675				
	F stat. GLS-BC AR(2)	2.114					0.478					0.323				

Notes: Sample restricted to only those employed between 18 and 65 years with at most primary or secondary school level. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender, area, industry and occupation dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CR2, to Bell and McCaffrey (2002) bias corrected CRVE CR2 and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. Eff. G (CSS) refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the statistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Appendix B: Results on Low skilled population

Key outcomes

Table 2.B1 – Single stage DiD: ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages (full set of controls), 2011-2019

	Formal wages					Informal wages				
	$\hat{\beta}$	SE Homosk.	SE CRVE	P-val. Rade	P-val. Webb	$\hat{\beta}$	SE Homosk.	SE CRVE	P-val. Rade	P-val. Webb
Aggregated										
2016-2019	0.031	(0.013)**	(0.022)	[0.225]	[0.215]	-0.019	(0.015)	(0.021)	[0.415]	[0.405]
S*	6.088					6.089				
Yearly (Base 2015)										
2016	0.016	(0.026)	(0.018)	[0.415]	[0.435]	0.006	(0.029)	(0.033)	[0.860]	[0.830]
2017	0.034	(0.027)	(0.031)	[0.365]	[0.360]	-0.025	(0.030)	(0.031)	[0.415]	[0.370]
2018	0.022	(0.026)	(0.023)	[0.340]	[0.345]	-0.092	(0.031)***	(0.037)**	[0.170]	[0.145]
2019	0.020	(0.026)	(0.030)	[0.570]	[0.525]	-0.100	(0.030)***	(0.038)**	[0.170]	[0.190]
N	61,297					72,779				
S*	6.197					6.068				
F stat. OLS	1.562					2.891**				

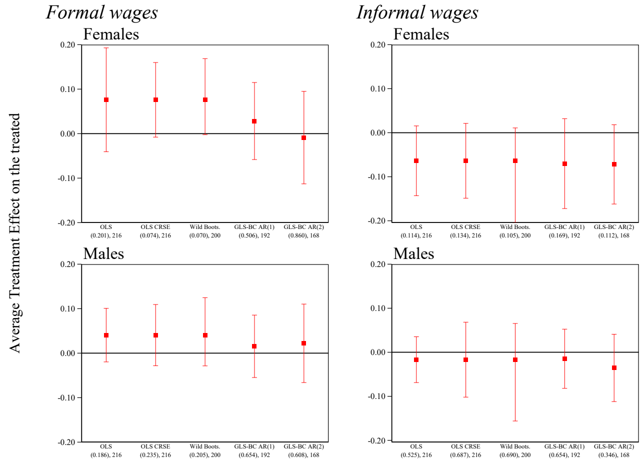
Notes: The sample for the formal outcome is restricted to those with at most primary or secondary school level and the sample for the informal outcome is additionally restricted to those between 18 and 35 years. Real hourly wages in 2007 PEN. Individual level covariates included are age and schooling years in levels and interacted and gender, area, industry and occupation dummies. SE CRVE refers to Cluster Robust Variance Estimator of SEs; P-val. WB Rade and Webb, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights and Webb weights, respectively, using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). F-stat refers to the statistic of the F test for the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.B2 – 2 stages DiD: ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages (full set of controls), 2011-2019

		Formal wages							Informal wages						
		$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. IK	P-val. WB	N	$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. IK	P-val. WB	N
Aggregated															
2016-2019	FE OLS	0.043	(0.025)*	(0.026)	(0.027)	[0.156]	[0.100]	216	-0.041	(0.024)*	(0.033)	(0.034)	[0.276]	[0.250]	216
	FE GLS-BC AR(1)	0.015	(0.036)	(0.029)				192	-0.040	(0.035)	(0.034)				192
	FE GLS-BC AR(2)	0.014	(0.035)	(0.040)				168	-0.054	(0.030)*	(0.033)				168
	S*	7.835							7.835						
	IK-DoF	6.846							6.846						
Yearly (Base 2015)															
2016	FE OLS	-0.006	(0.054)	(0.027)	(0.028)	[0.830]	[0.835]	216	-0.000	(0.050)	(0.042)	(0.044)	[0.995]	[0.995]	216
	FE GLS-BC AR(1)	-0.007	(0.052)	(0.040)				192	-0.001	(0.047)	(0.033)				192
	FE GLS-BC AR(2)	-0.007	(0.052)	(0.031)				168	-0.000	(0.046)	(0.046)				168
2017	FE OLS	-0.021	(0.054)	(0.042)	(0.043)	[0.631]	[0.595]	216	-0.052	(0.050)	(0.052)	(0.054)	[0.366]	[0.455]	216
	FE GLS-BC AR(1)	-0.022	(0.059)	(0.031)				192	-0.053	(0.054)	(0.046)				192
	FE GLS-BC AR(2)	-0.023	(0.056)	(0.047)				168	-0.052	(0.051)	(0.057)				168
2018	FE OLS	0.042	(0.054)	(0.049)	(0.051)	[0.437]	[0.395]	216	-0.049	(0.050)	(0.055)	(0.058)	[0.422]	[0.565]	216
	FE GLS-BC AR(1)	0.041	(0.061)	(0.047)				192	-0.050	(0.057)	(0.057)				192
	FE GLS-BC AR(2)	0.041	(0.057)	(0.056)				168	-0.049	(0.051)	(0.061)				168
2019	FE OLS	0.033	(0.054)	(0.057)	(0.058)	[0.587]	[0.600]	216	-0.055	(0.050)	(0.048)	(0.049)	[0.305]	[0.285]	216
	FE GLS-BC AR(1)	0.032	(0.061)	(0.056)				192	-0.055	(0.057)	(0.061)				192
	FE GLS-BC AR(2)	0.032	(0.058)	(0.064)				168	-0.054	(0.051)	(0.053)				168
	S*	7.835							7.835						
	IK-DoF	6.846							6.846						
	F stat. FE OLS	2.680*							0.968						
	F stat. GLS-BC AR(1)	0.512							0.475						
	F stat. GLS-BC AR(2)	0.779							0.270						

Notes: Sample restricted to only those employed between 18 and 65 years with at most primary or secondary school level; additionally, regressions for the informal sector take only those between 18 and 35 years. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender, area, industry and occupation dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the statistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Figure 2.B1 – 2 stages DiD: ATET Heterogeneity of Venezuelan immigration on low skilled natives' (log) formal and informal wages by gender, 2011-2019



Note: Sample restricted to only those employed between 18 and 65 years with at most primary or secondary school level; additionally, regressions for the informal sector take only those between 18 and 35 years. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender, area, industry and occupation dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. FE OLS CRVE refers to Cluster Robust Variance Estimator of SEs; Wild Boots., to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications and FE GLS-BC, to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.B3 – Panel data: Average Treatment Effect of Venezuelan immigration on low skilled natives' (log) formal and informal wages (full set of controls), 2005-201911-2019

		Formal wages							Informal wages						
		$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. I-K	P-val. WB	N	$\hat{\beta}$	SE Homosk.	SE CRVE	SE CRVE2	P-val. I-K	P-val. WB	N
Aggregated															
2016-2019	FE OLS	0.056	(0.037)	(0.035)	(0.036)	[0.163]	[0.145]	360	-0.009	(0.025)	(0.030)	(0.032)	[0.781]	[0.810]	360
	FD	-0.029	(0.068)	(0.040)	(0.041)	[0.508]	[0.460]	336	-0.011	(0.044)	(0.039)	(0.041)	[0.800]	[0.810]	336
	FE GLS AR(1)	0.058		(0.035)			[0.095]	360	-0.015		(0.020)			[0.725]	360
	FE GLS AR(2)	0.054		(0.035)			[0.095]	360	-0.014		(0.020)			[0.725]	360
	S* FE	8.753							8.753						
	S* FD	8.032							8.032						
	IK-DoF FE	7.783							7.783						
	IK-DoF FD	6.951							6.948						
Yearly (Base 2015)															
2016	FE OLS	-0.016	(0.084)	(0.040)	(0.041)	[0.709]	[0.700]	360	-0.015	(0.057)	(0.044)	(0.045)	[0.743]	[0.745]	360
	FD	-0.027	(0.068)	(0.042)	(0.042)	[0.542]	[0.470]	336	-0.011	(0.045)	(0.039)	(0.040)	[0.791]	[0.790]	336
	FE GLS AR(1)	-0.021		(0.039)			[0.750]	360	-0.014		(0.041)			[0.800]	360
	FE GLS AR(2)	-0.019		(0.038)			[0.750]	360	-0.015		(0.042)			[0.800]	360
2017	FE OLS	-0.025	(0.085)	(0.053)	(0.054)	[0.660]	[0.640]	360	-0.060	(0.057)	(0.049)	(0.050)	[0.266]	[0.360]	360
	FD	-0.019	(0.097)	(0.056)	(0.057)	[0.754]	[0.750]	336	-0.044	(0.063)	(0.048)	(0.048)	[0.386]	[0.425]	336
	FE GLS AR(1)	-0.022		(0.052)			[0.640]	360	-0.052		(0.047)			[0.370]	360
	FE GLS AR(2)	-0.019		(0.051)			[0.640]	360	-0.052		(0.047)			[0.370]	360
2018	FE OLS	0.027	(0.085)	(0.071)	(0.073)	[0.718]	[0.720]	360	-0.063	(0.057)	(0.051)	(0.053)	[0.269]	[0.330]	360
	FD	0.036	(0.119)	(0.070)	(0.072)	[0.635]	[0.690]	336	-0.062	(0.078)	(0.061)	(0.062)	[0.352]	[0.455]	336
	FE GLS AR(1)	0.030		(0.069)			[0.705]	360	-0.064		(0.054)			[0.405]	360
	FE GLS AR(2)	0.032		(0.069)			[0.705]	360	-0.063		(0.054)			[0.405]	360
2019	FE OLS	0.027	(0.085)	(0.067)	(0.069)	[0.712]	[0.735]	360	-0.069	(0.057)	(0.037)*	(0.037)*	[0.107]	[0.095]	360
	FD	0.029	(0.137)	(0.069)	(0.071)	[0.697]	[0.705]	336	-0.082	(0.090)	(0.044)*	(0.045)*	[0.107]	[0.120]	336
	FE GLS AR(1)	0.026		(0.064)			[0.630]	360	-0.076		(0.039)*			[0.390]	360
	FE GLS AR(2)	0.028		(0.065)			[0.630]	360	-0.076		(0.039)*			[0.390]	360
	S* FE	7.956							7.956						
	S* FD	8.003							8.003						
	IK-DoF FE	6.968							6.968						
	IK-DoF FD	7.005							7.009						
	F stat. FE OLS	3.175**							0.928						
	F stat. FD	1.600							0.863						
	F stat. GLS AR(1)	40.871***							9.760						
	F stat. GLS AR(2)	42.134***							9.769						

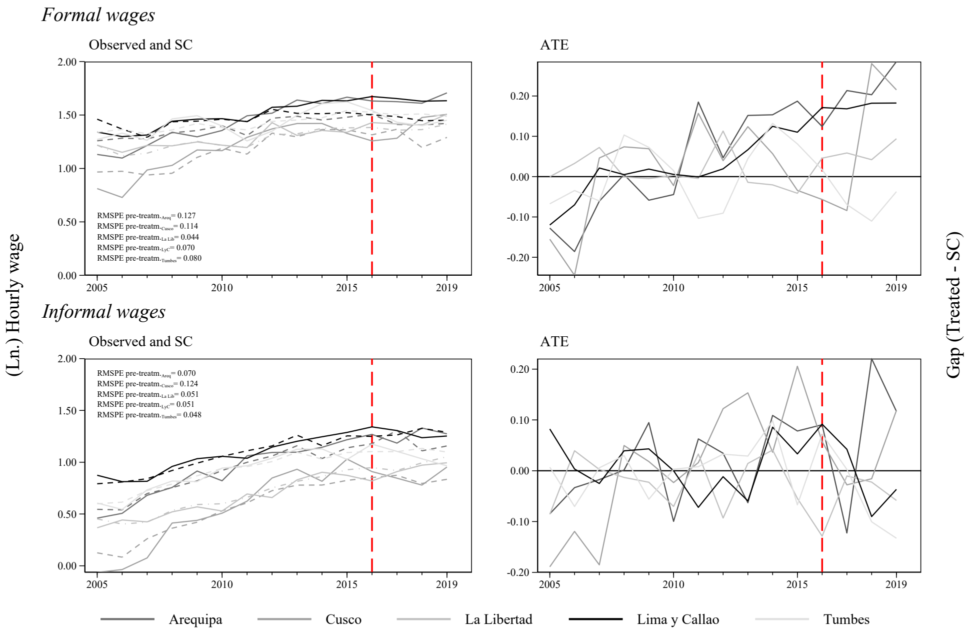
Notes: Sample restricted to only those employed between 18 and 65 years with at most secondary school. Real hourly wages in 2007 PEN. Controls included are gdp per capita, pea employed for low sckilled, pea employed in services, in manufacture, schooling, proportion of population between 18 and 25 and between 26 and 35, percentage of urban population. Equations include 2-way FEs (regions and years). SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the F test cluster robust of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B4 – SCM: SC weights for the ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019

	Formal wages					**Informal wages**				
	Areq.	Cusco	La Lib.	L y C.	Tumb.	Areq.	Cusco	La Lib.	L y C.	Tumb.
Amazonas	0	0	0	0	0	0	0	0	0	0
Ancash	0	0	0	0	0	0	0	0	0	0
ApurÃmac	0	0	0	0	0	0	0	0	0	0
Ayacucho	0	0	.055	0	0	0	0	0	0	0
Cajamarca	0	0	.102	0	0	0	0	.111	0	0
Huancavelica	0	0	0	0	0	0	.308	0	0	0
HuÃ¡nuco	0	0	0	0	0	0	0	0	0	0
Ica	.147	0	.12	0	.24	.077	0	.355	.35	0
JunÃn	0	0	0	0	0	0	0	0	0	0
Lambayeque	0	0	.28	0	0	0	0	.341	0	0
Loreto	0	0	0	0	0	0	0	0	0	.245
Madre de Dios	0	0	0	.382	.414	.043	0	0	.65	.477
Moquegua	.52	.041	0	0	0	.568	.132	0	0	0
Pasco	0	.116	0	0	0	0	0	0	0	0
Piura	0	0	0	0	0	0	0	0	0	0
Puno	0	.484	0	0	0	0	.149	0	0	0
San MartÃn	0	0	.06	0	.209	0	0	0	0	.068
Tacna	.333	.193	.131	.524	0	.312	0	0	0	.21
Ucayali	0	.167	.252	.095	.138	0	.411	.193	0	0

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years and with at most secondary school level and, in the case of the informal outcome, the sample is additionally restricted to those between 18 and 35 years before data aggregation. Real hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.B2 – SCM: ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years with at most secondary school level; additionally, estimations for the informal sector take only those between 18 and 35 years before data aggregation. Real hourly wages in 2007 PEN.
Source: Author's calculations using ENAHO 2005-2019 data.

Table 2.B5 – SCM: Covariate balance for the ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019

		Arequipa	Cusco	La Lib.	L y C.	Tumbes
Covariates						
GDP pc (thousands)	Obs.	15.924	11.714	9.441	16.380	9.498
	Wei. Formal Wages	32.129	10.727	9.137	15.830	12.693
	Wei. Informal Wages	33.654	11.722	10.012	16.275	13.723
Empl. LF low skill	Obs.	0.536	0.678	0.667	0.551	0.678
	Wei. Formal Wages	0.569	0.675	0.691	0.638	0.679
	Wei. Informal Wages	0.572	0.725	0.669	0.644	0.681
Empl. LF serv.	Obs.	0.644	0.549	0.588	0.734	0.683
	Wei. Formal Wages	0.644	0.549	0.588	0.676	0.583
	Wei. Informal Wages	0.643	0.511	0.586	0.616	0.629
Empl. LF manuf.	Obs.	0.192	0.133	0.195	0.229	0.127
	Wei. Formal Wages	0.156	0.148	0.157	0.131	0.133
	Wei. Informal Wages	0.152	0.126	0.172	0.134	0.116
Schooling (years)	Obs.	11.154	9.211	9.607	11.452	9.883
	Wei. Formal Wages	10.910	9.776	9.617	10.369	9.875
	Wei. Informal Wages	10.850	9.061	9.918	10.311	9.883
Proport. 18-25 yo.	Obs.	0.168	0.154	0.183	0.189	0.177
	Wei. Formal Wages	0.150	0.175	0.177	0.166	0.167
	Wei. Informal Wages	0.146	0.162	0.181	0.167	0.166
Proport. 26-35 yo.	Obs.	0.270	0.264	0.272	0.288	0.293
	Wei. Formal Wages	0.270	0.267	0.270	0.281	0.281
	Wei. Informal Wages	0.271	0.268	0.266	0.279	0.287
Urban population	Obs.	0.881	0.535	0.776	0.979	0.922
	Wei. Formal Wages	0.819	0.637	0.746	0.801	0.754
	Wei. Informal Wages	0.807	0.565	0.784	0.784	0.735
Ecu.-Col. border(d)	Obs.	0.000	0.000	0.000	0.000	1.000
	Wei. Formal Wages	0.000	0.000	0.102	0.000	0.000
	Wei. Informal Wages	0.000	0.000	0.111	0.000	0.245
Lagged Outcomes						
2005	Obs. Formal Wages	1.131	0.812	1.218	1.341	1.271
	Wei. Formal Wages	1.259	0.966	1.218	1.459	1.337
	Obs. Informal Wages	0.460	-0.064	0.367	0.874	0.606
	Wei. Informal Wages	0.545	0.125	0.453	0.792	0.599
2010	Obs. Formal Wages	1.355	1.166	1.218	1.466	1.397
	Wei. Formal Wages	1.400	1.187	1.215	1.459	1.381
	Obs. Informal Wages	0.821	0.508	0.529	1.057	0.943
	Wei. Informal Wages	0.920	0.531	0.599	1.057	0.940
2015	Obs. Formal Wages	1.667	1.328	1.339	1.633	1.623
	Wei. Formal Wages	1.480	1.361	1.380	1.521	1.540
	Obs. Informal Wages	1.217	1.030	0.873	1.288	1.026
	Wei. Informal Wages	1.139	0.824	0.928	1.254	1.093

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 and at most secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. All variables (except the lagged outcomes) are averaged for the 2005-2015 period. LF=Labour force; Ecu-Col border is a dummy that equals 1 if the region neighbours Ecuador and Colombia. Real hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B6 – SCM: DiD Regressions for the ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019

	GLS with AR(1) disturbances					*Baltagi and Wu (1999) estimator*				
	(2005-14)	2016	2017	2018	2019	(2005-14)	2016	2017	2018	2019
Formal Wages										
Arequipa	-0.050	-0.048	0.054	0.054	0.147	-0.089	-0.029	0.082	0.085	0.175
	(0.118)	(0.118)	(0.161)	(0.190)	(0.212)	(0.106)	(0.111)	(0.139)	(0.153)	(0.161)
Cusco	0.083	-0.015	-0.037	0.334	0.274	0.072	-0.009	-0.028	0.343*	0.283
	(0.155)	(0.155)	(0.213)	(0.254)	(0.285)	(0.138)	(0.142)	(0.181)	(0.203)	(0.217)
La Libertad	0.036	0.079	0.087	0.068	0.119	0.042	0.079	0.088	0.070	0.122
	(0.108)	(0.115)	(0.141)	(0.152)	(0.158)	(0.084)	(0.095)	(0.111)	(0.117)	(0.119)
Lima y Callao	-0.012	0.082	0.095	0.121	0.131	-0.039	0.090	0.104	0.126	0.132
	(0.102)	(0.103)	(0.136)	(0.155)	(0.168)	(0.084)	(0.089)	(0.110)	(0.119)	(0.124)
Tumbes	-0.003	-0.033	-0.103	-0.134	-0.055	-0.029	-0.034	-0.109	-0.146	-0.071
	(0.121)	(0.128)	(0.158)	(0.172)	(0.179)	(0.096)	(0.108)	(0.126)	(0.132)	(0.135)
Informal Wages										
Arequipa	0.022	0.020	-0.185	0.164	0.067	0.001	0.034	-0.164	0.190	0.094
	(0.162)	(0.162)	(0.223)	(0.266)	(0.300)	(0.148)	(0.151)	(0.194)	(0.219)	(0.234)
Cusco	-0.175	-0.143	-0.222	-0.205	-0.066	-0.187	-0.135	-0.210	-0.190	-0.050
	(0.192)	(0.192)	(0.266)	(0.320)	(0.363)	(0.179)	(0.182)	(0.237)	(0.269)	(0.290)
La Libertad	0.091	-0.071	0.051	0.043	0.010	0.075	-0.061	0.066	0.059	0.027
	(0.121)	(0.121)	(0.167)	(0.201)	(0.228)	(0.117)	(0.121)	(0.153)	(0.170)	(0.180)
Lima y Callao	0.051	0.060	0.012	-0.119	-0.064	0.040	0.068	0.026	-0.102	-0.044
	(0.105)	(0.105)	(0.145)	(0.173)	(0.196)	(0.094)	(0.096)	(0.124)	(0.141)	(0.151)
Tumbes	0.162	0.143	0.077	-0.023	-0.052	0.146	0.154	0.095	-0.000	-0.027
	(0.122)	(0.122)	(0.169)	(0.203)	(0.229)	(0.114)	(0.116)	(0.150)	(0.169)	(0.181)

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with at most secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. Each row represents a regression of annual observations for the treated areas and its synthetic control between 2005 and 2019, N=30, which includes a dummy variable for the treated region, period dummies (2005-2014, 2016, 2017, 2018, 2019 with 2015 excluded) and their interactions; these interaction coefficients are the coefficients reported. Real hourly wages in 2007 PEN * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B7 – SCM: P-values for the ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019

	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
Formal wages					
Arequipa	0.2066	0.1272	1.6851	6/20	0.3000
Cusco	0.0887	0.1140	1.6136	3/20	0.1500
La Libertad	0.0602	0.0442	1.4387	8/20	0.4000
Lima y Callao	0.1761	0.0692	2.5464	2/20	0.1000
Tumbes	-0.0499	0.0800	0.8531	17/20	0.8500
Informal wages					
Arequipa	0.0764	0.0704	2.0763	4/20	0.2000
Cusco	0.0331	0.1243	0.5453	20/20	1.0000
La Libertad	-0.0549	0.0514	1.3948	7/20	0.3500
Lima y Callao	0.0018	0.0505	1.3872	6/20	0.3000
Tumbes	-0.0397	0.0479	1.8919	5/20	0.2500

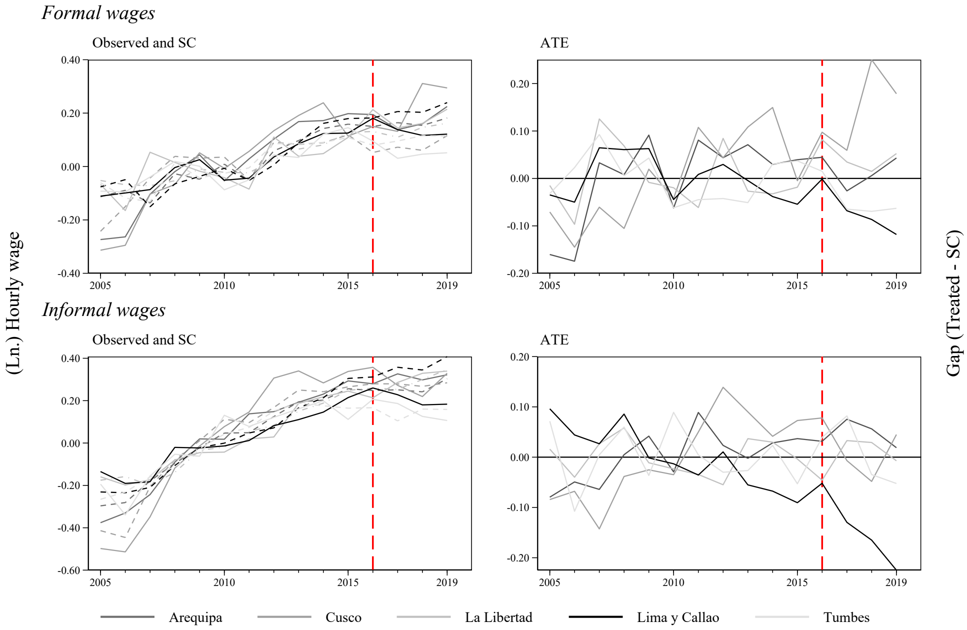
Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with at most secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B8 – SCM demeaned: Covariate balance for the ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019

		Arequipa	Cusco	La Lib.	L y C.	Tumbes
Covariates						
GDP pc (thousands)	Obs.	15.924	11.714	9.441	16.380	9.498
	Wei. Formal Wages	19.221	13.785	9.671	15.474	10.113
	Wei. Informal Wages	19.386	11.317	10.267	15.787	13.548
Empl. LF low skill	Obs.	0.536	0.678	0.667	0.551	0.678
	Wei. Formal Wages	0.574	0.691	0.671	0.577	0.686
	Wei. Informal Wages	0.581	0.685	0.675	0.581	0.676
Empl. LF serv.	Obs.	0.644	0.549	0.588	0.734	0.683
	Wei. Formal Wages	0.596	0.507	0.574	0.594	0.596
	Wei. Informal Wages	0.632	0.485	0.584	0.626	0.642
Empl. LF manuf.	Obs.	0.192	0.133	0.195	0.229	0.127
	Wei. Formal Wages	0.184	0.137	0.169	0.188	0.140
	Wei. Informal Wages	0.170	0.122	0.168	0.177	0.121
Schooling (years)	Obs.	11.154	9.211	9.607	11.452	9.883
	Wei. Formal Wages	11.087	9.394	9.768	11.117	9.844
	Wei. Informal Wages	10.975	9.318	9.790	11.058	9.947
Proport. 18-25 yo.	Obs.	0.168	0.154	0.183	0.189	0.177
	Wei. Formal Wages	0.177	0.160	0.181	0.184	0.178
	Wei. Informal Wages	0.173	0.172	0.181	0.180	0.169
Proport. 26-35 yo.	Obs.	0.270	0.264	0.272	0.288	0.293
	Wei. Formal Wages	0.266	0.272	0.262	0.265	0.283
	Wei. Informal Wages	0.269	0.260	0.270	0.268	0.285
Urban population	Obs.	0.881	0.535	0.776	0.979	0.922
	Wei. Formal Wages	0.881	0.600	0.745	0.895	0.754
	Wei. Informal Wages	0.862	0.544	0.780	0.887	0.747
Ecu.-Col. border(d)	Obs.	0.000	0.000	0.000	0.000	1.000
	Wei. Formal Wages	0.000	0.088	0.163	0.000	0.507
	Wei. Informal Wages	0.000	0.000	0.353	0.000	0.356
Lagged Outcomes						
2005	Obs. Formal Wages	-0.265	-0.345	-0.045	-0.130	-0.143
	Wei. Formal Wages	-0.116	-0.273	-0.044	-0.115	-0.081
	Obs. Informal Wages	-0.427	-0.572	-0.252	-0.169	-0.280
	Wei. Informal Wages	-0.273	-0.397	-0.221	-0.240	-0.289
2010	Obs. Formal Wages	-0.041	0.010	-0.045	-0.005	-0.017
	Wei. Formal Wages	-0.008	0.046	-0.030	-0.011	0.009
	Obs. Informal Wages	-0.066	-0.000	-0.090	0.014	0.057
	Wei. Informal Wages	0.003	0.018	-0.057	-0.016	0.047
2015	Obs. Formal Wages	0.271	0.171	0.076	0.161	0.209
	Wei. Formal Wages	0.191	0.165	0.128	0.201	0.112
	Obs. Informal Wages	0.331	0.522	0.254	0.245	0.141
	Wei. Informal Wages	0.290	0.347	0.290	0.300	0.223

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with at most secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. All variables (except the lagged outcomes) are averaged for the 2005-2015 period. LF=Labour force; Ecu-Col border is a dummy that equals 1 if the region neighbours Ecuador and Colombia. Real hourly wages in 2007 PEN. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.B3 – SCM demeaned: ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years before data aggregation. Real hourly wages in 2007 PEN. In all of these, we first applied the demeaning routine suggested by Ferman & Pinto (2019). *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B9 – SCM demeaned: DiD Regressions for the ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019

	GLS with AR(1) disturbances					*Baltagi and Wu (1999) estimator*				
	(2005-14)	2016	2017	2018	2019	(2005-14)	2016	2017	2018	2019
Formal Wages										
Arequipa	-0.053	-0.067	-0.059	-0.071	-0.014	-0.066	-0.061	-0.051	-0.063	-0.008
	(0.122)	(0.122)	(0.167)	(0.198)	(0.223)	(0.112)	(0.116)	(0.146)	(0.161)	(0.170)
Cusco	0.079	-0.048	-0.031	0.328	0.263	0.043	-0.035	-0.016	0.341	0.272
	(0.172)	(0.172)	(0.233)	(0.273)	(0.302)	(0.163)	(0.177)	(0.214)	(0.229)	(0.236)
La Libertad	0.049	0.069	0.068	0.071	0.106	0.051	0.069	0.069	0.072	0.107
	(0.101)	(0.110)	(0.132)	(0.141)	(0.145)	(0.079)	(0.089)	(0.104)	(0.109)	(0.110)
Lima y Callao	0.008	-0.002	-0.042	-0.060	-0.113	0.018	-0.007	-0.049	-0.068	-0.121
	(0.111)	(0.111)	(0.151)	(0.177)	(0.196)	(0.096)	(0.102)	(0.126)	(0.137)	(0.143)
Tumbes	-0.057	-0.100	-0.198	-0.208	-0.171	-0.075	-0.099	-0.201	-0.215	-0.181
	(0.125)	(0.130)	(0.164)	(0.181)	(0.190)	(0.100)	(0.112)	(0.132)	(0.139)	(0.141)
Informal Wages										
Arequipa	0.021	0.072	-0.055	0.113	-0.035	0.003	0.083	-0.039	0.132	-0.016
	(0.152)	(0.152)	(0.210)	(0.251)	(0.282)	(0.141)	(0.144)	(0.185)	(0.208)	(0.222)
Cusco	-0.129	-0.069	-0.194	-0.196	-0.031	-0.146	-0.058	-0.177	-0.176	-0.010
	(0.190)	(0.190)	(0.264)	(0.316)	(0.358)	(0.181)	(0.185)	(0.239)	(0.269)	(0.288)
La Libertad	0.105	-0.011	0.093	0.066	0.037	0.088	-0.000	0.110	0.086	0.059
	(0.126)	(0.126)	(0.175)	(0.210)	(0.237)	(0.120)	(0.124)	(0.157)	(0.176)	(0.186)
Lima y Callao	0.063	0.076	0.006	-0.101	-0.193	0.065	0.076	0.005	-0.102	-0.193
	(0.112)	(0.112)	(0.154)	(0.185)	(0.209)	(0.102)	(0.104)	(0.134)	(0.151)	(0.161)
Tumbes	0.157	0.146	0.087	-0.027	-0.067	0.144	0.155	0.102	-0.009	-0.047
	(0.122)	(0.123)	(0.170)	(0.204)	(0.230)	(0.115)	(0.118)	(0.152)	(0.171)	(0.183)

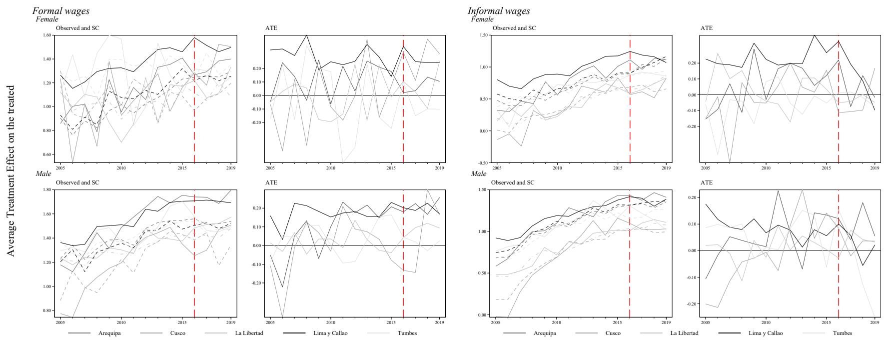
Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with at most secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. Each row represents a regression of annual observations for the treated areas and its synthetic control between 2005 and 2019, N=30, which includes a dummy variable for the treated region, period dummies (2005-2014, 2016, 2017, 2018, 2019 with 2015 excluded) and their interactions; these interaction coefficients are the coefficients reported. Real hourly wages in 2007 PEN. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B10 – SCM demeaned: P-values for the ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019

	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
Formal wages					
Arequipa	0.0117	0.1133	0.1970	20/20	1.0000
Cusco	0.1099	0.1507	1.2960	6/20	0.3000
La Libertad	0.0318	0.0498	0.7168	15/20	0.7500
Lima y Callao	-0.0875	0.0565	1.6900	4/20	0.2000
Tumbes	-0.1022	0.0861	1.3160	7/20	0.3500
Informal wages					
Arequipa	0.0470	0.0843	1.0269	12/20	0.6000
Cusco	0.0376	0.1454	0.5652	19/20	0.9500
La Libertad	0.0022	0.0520	0.7189	16/20	0.8000
Lima y Callao	-0.1048	0.0358	4.0720	1/20	0.0500
Tumbes	-0.0536	0.0492	2.0996	2/20	0.1000

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with secondary school and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.B4 – SCM: ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years with at most secondary school level; additionally, estimations for the informal sector take only those between 18 and 35 years before data aggregation. Real hourly wages in 2007 PEN.
Source: Author's calculations using ENAHO 2005-2019 data.

Table 2.B11 – SCM: P-values for the ATET of Venezuelan immigration on low skilled natives' (log) formal and informal wages across subsamples, 2005-2019

	Formal wage					Informal wage				
	ATE	RMSPE pre	Ratio post-pre	Rank	p-value	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
Sex										
Arequipa										
Female	0.0742	0.1766	0.4989	18/24	0.7500	0.0621	0.1502	0.8745	9/24	0.3750
Male	0.2070	0.1604	1.3098	6/24	0.2500	0.0992	0.1024	1.1132	9/24	0.3750
Cusco										
Female	0.2889	0.2353	1.3142	2/24	0.0833	-0.0370	0.1621	0.7693	13/24	0.5417
Male	0.0467	0.1492	1.3170	4/24	0.1667	0.0344	0.1318	0.3846	18/24	0.7500
La Libertad										
Female	0.0719	0.1467	0.9904	7/24	0.2917	-0.0349	0.1216	0.3358	20/24	0.8333
Male	0.0877	0.0539	1.7068	1/24	0.0417	0.0308	0.0514	0.9499	11/24	0.4583
Lima y Callao										
Female	0.2747	0.2978	0.9382	8/24	0.3333	0.1304	0.2407	0.8536	10/24	0.4167
Male	0.1965	0.1768	1.1176	5/24	0.2083	0.0308	0.0966	0.6740	15/24	0.6250
Tumbes										
Female	-0.0627	0.2782	0.4113	18/24	0.7500	-0.0157	0.1427	0.2844	20/24	0.8333
Male	0.0154	0.0966	0.3161	20/24	0.8333	-0.0444	0.0893	1.8186	2/24	0.0833

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with at most secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

Ancillary outcomes

Table 2.B12 – Single stage DiD: ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector (full set of controls), 2011-2019

	Informality rate				**Formal sector inequality**				**Informal sector inequality**			
	$\hat{\beta}$	SE CRVE	P. WB R	P. WB W	$\hat{\beta}$	SE CRVE	P. WB R	P. WB W	$\hat{\beta}$	SE CRVE	P. WB R	P. WB W
Aggregated												
2016-2019	0.005	(0.010)	[0.590]	[0.605]	0.023	(0.011)**	[0.105]	[0.165]	0.013	(0.009)	[0.295]	[0.325]
S*	6.233				6.087				6.477			
Yearly (Base 2015)												
2016	0.018	(0.007)**	[0.120]	[0.110]	0.028	(0.016)*	[0.295]	[0.275]	0.010	(0.015)	[0.535]	[0.510]
2017	0.004	(0.011)	[0.700]	[0.700]	0.016	(0.017)	[0.375]	[0.435]	0.005	(0.012)	[0.700]	[0.650]
2018	0.001	(0.010)	[0.935]	[0.950]	0.046	(0.024)*	[0.365]	[0.380]	0.009	(0.023)	[0.850]	[0.820]
2019	0.017	(0.010)*	[0.105]	[0.080]	0.002	(0.015)	[0.935]	[0.930]	0.049	(0.017)***	[0.155]	[0.160]
N	280,589				61,297				219,292			
S*	6.241				6.187				6.390			
F stat. OLS	6.486***				2.976**				1.878			

Notes: The sample is restricted to only with at most primary or secondary school level. Real hourly wages in 2007 PEN. Individual level covariates are age and schooling in levels and interacted and gender, area, industry and occupation dummies SE CRVE refers to Cluster Robust Variance Estimator of SEs; P. WB Rade and Webb, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights and Webb weights using 199 replications, respectively. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). F-stat refers to the stastistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.B13 – 2 stages DiD: ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector (full set of controls), 2011-2019

		Informality rate					Formal sector inequality					Informal sector inequality				
		$\hat{\beta}$	SE CRVE	SE CRVE2	P-val. WB	N	$\hat{\beta}$	SE CRVE	SE CRVE2	P-val. WB	N	$\hat{\beta}$	SE CRVE	SE CRVE2	P-val. WB	N
Aggregated																
2016-2019	FE OLS	0.003	(0.011)	(0.012)	[0.855]	216	0.026	(0.012)**	(0.012)**	[0.065]	216	0.001	(0.019)	(0.020)	[0.955]	216
	FE GLS-BC AR(1)	0.011	(0.008)			192	0.027	(0.018)			192	0.010	(0.012)			192
	FE GLS-BC AR(2)	0.010	(0.008)			168	0.035	(0.019)*			168	0.011	(0.012)			168
	S*	7.835					7.835					7.835				
	IK-DoF	6.846					6.846					6.846				
Yearly (Base 2015)																
2016	FE OLS	0.009	(0.010)	(0.010)	[0.415]	216	0.040	(0.034)	(0.034)	[0.250]	216	0.005	(0.023)	(0.024)	[0.845]	216
	FE GLS-BC AR(1)	0.009	(0.011)			192	0.040	(0.037)			192	0.005	(0.025)			192
	FE GLS-BC AR(2)	0.009	(0.011)			168	0.040	(0.037)			168	0.006	(0.026)			168
2017	FE OLS	0.004	(0.010)	(0.010)	[0.785]	216	0.042	(0.031)	(0.030)	[0.170]	216	0.004	(0.015)	(0.015)	[0.780]	216
	FE GLS-BC AR(1)	0.005	(0.011)			192	0.042	(0.034)			192	0.004	(0.016)			192
	FE GLS-BC AR(2)	0.005	(0.011)			168	0.042	(0.034)			168	0.006	(0.018)			168
2018	FE OLS	-0.001	(0.009)	(0.009)	[0.880]	216	0.027	(0.031)	(0.031)	[0.420]	216	0.020	(0.019)	(0.019)	[0.295]	216
	FE GLS-BC AR(1)	-0.000	(0.010)			192	0.027	(0.034)			192	0.020	(0.021)			192
	FE GLS-BC AR(2)	-0.000	(0.009)			168	0.027	(0.034)			168	0.022	(0.023)			168
2019	FE OLS	0.021	(0.013)	(0.014)	[0.095]	216	0.028	(0.031)	(0.031)	[0.380]	216	0.026	(0.017)	(0.018)	[0.185]	216
	FE GLS-BC AR(1)	0.022	(0.015)			192	0.028	(0.034)			192	0.026	(0.019)			192
	FE GLS-BC AR(2)	0.023	(0.014)			168	0.028	(0.034)			168	0.028	(0.022)			168
	S*	7.835					7.835					7.835				
	IK-DoF	6.846					6.846					6.846				
	F stat. FE OLS	1.473					1.722					0.648				
	F stat. GLS-BC AR(1)	0.261					1.878					0.299				
	F stat. GLS-BC AR(2)	0.213					0.004					0.353				

Notes: Sample restricted to only those employed between 18 and 65 years with at most primary or secondary school level. Real hourly wages in 2007 PEN. Using a 2-step estimation procedure presented in Hansen (2007), individual level covariates (age and schooling in levels and interacted and gender, area, industry and occupation dummies) are partialled out first and then the treatment effects are estimated from a 2-way FEs (regions and years) model. SE CRVE refers to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the statistic from the F test of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment using the cluster-robust VCE. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.B14 – Panel data: Average Treatment Effect of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector (full set of controls), 2005-2019

		Informality rate					**Formal sector inequality**					**Informal sector inequality**				
		β	SE CRVE	SE CRVE2	P-val. WB	N	β	SE CRVE	SE CRVE2	P-val. WB	N	β	SE CRVE	SE CRVE2	P-val. WB	N
Aggregated																
2016-2019	FE OLS	0.003	(0.013)	(0.013)	[0.805]	360	0.042	(0.013)***	(0.013)***	[0.000]	360	-0.015	(0.010)	(0.010)	[0.130]	360
	FD	0.012	(0.010)	(0.010)	[0.290]	336	0.033	(0.026)	(0.027)	[0.240]	336	-0.000	(0.042)	(0.045)	[1.000]	336
	FE GLS AR(1)	0.009	(0.009)		[0.685]	360	0.042	(0.013)***		[0.000]	360	-0.015	(0.010)		[0.175]	360
	FE GLS AR(2)	0.009	(0.009)		[0.685]	360	0.042	(0.013)***		[0.000]	360	-0.016	(0.010)		[0.175]	360
	S* FE	8.753					8.753					8.753				
	S* FD	8.032					8.032					8.032				
	IK-DoF FE	7.783					7.783					7.783				
	IK-DoF FD	6.954					6.946					6.945				
Yearly (Base 2015)																
2016	FE OLS	0.011	(0.011)	(0.011)	[0.320]	360	0.037	(0.027)	(0.027)	[0.170]	360	-0.001	(0.045)	(0.046)	[0.975]	360
	FD	0.012	(0.010)	(0.010)	[0.285]	336	0.034	(0.026)	(0.027)	[0.220]	336	-0.000	(0.044)	(0.045)	[0.990]	336
	FE GLS AR(1)	0.012	(0.010)		[0.275]	360	0.037	(0.027)		[0.235]	360	-0.001	(0.044)		[0.980]	360
	FE GLS AR(2)	0.012	(0.010)		[0.275]	360	0.037	(0.027)		[0.235]	360	-0.001	(0.044)		[0.980]	360
2017	FE OLS	0.004	(0.013)	(0.013)	[0.815]	360	0.040	(0.024)	(0.024)	[0.085]	360	-0.029	(0.035)	(0.037)	[0.630]	360
	FD	0.001	(0.013)	(0.013)	[0.985]	336	0.042	(0.024)	(0.024)*	[0.095]	336	-0.031	(0.035)	(0.036)	[0.550]	336
	FE GLS AR(1)	0.003	(0.012)		[0.835]	360	0.040	(0.023)*		[0.155]	360	-0.029	(0.035)		[0.535]	360
	FE GLS AR(2)	0.003	(0.012)		[0.835]	360	0.040	(0.023)*		[0.155]	360	-0.030	(0.035)		[0.535]	360
2018	FE OLS	0.003	(0.010)	(0.010)	[0.685]	360	0.032	(0.026)	(0.026)	[0.225]	360	0.017	(0.039)	(0.040)	[0.690]	360
	FD	0.002	(0.010)	(0.010)	[0.870]	336	0.030	(0.025)	(0.025)	[0.240]	336	0.019	(0.038)	(0.038)	[0.640]	336
	FE GLS AR(1)	0.004	(0.009)		[0.740]	360	0.032	(0.026)		[0.320]	360	0.017	(0.039)		[0.695]	360
	FE GLS AR(2)	0.004	(0.009)		[0.740]	360	0.032	(0.026)		[0.320]	360	0.017	(0.039)		[0.695]	360
2019	FE OLS	0.025	(0.014)*	(0.014)*	[0.100]	360	0.026	(0.027)	(0.028)	[0.400]	360	-0.020	(0.033)	(0.034)	[0.710]	360
	FD	0.023	(0.016)	(0.016)	[0.200]	336	0.023	(0.022)	(0.022)	[0.340]	336	-0.016	(0.032)	(0.033)	[0.790]	336
	FE GLS AR(1)	0.026	(0.014)*		[0.130]	360	0.026	(0.027)		[0.395]	360	-0.021	(0.032)		[0.660]	360
	FE GLS AR(2)	0.026	(0.014)*		[0.130]	360	0.025	(0.027)		[0.395]	360	-0.021	(0.032)		[0.660]	360
	S* FE	7.956					7.956					7.956				
	S* FD	8.003					8.003					8.003				
	IK-DoF FE	6.968					6.968					6.968				
	IK-DoF FD	6.992					7.015					7.018				
	F stat. FE OLS	1.149					2.856**					5.650***				
	F stat. FD	2.423**					4.387***					3.281***				
	F stat. GLS AR(1)	22.258**					34.637***					67.781***				
	F stat. GLS AR(2)	22.189**					34.576***					64.671***				

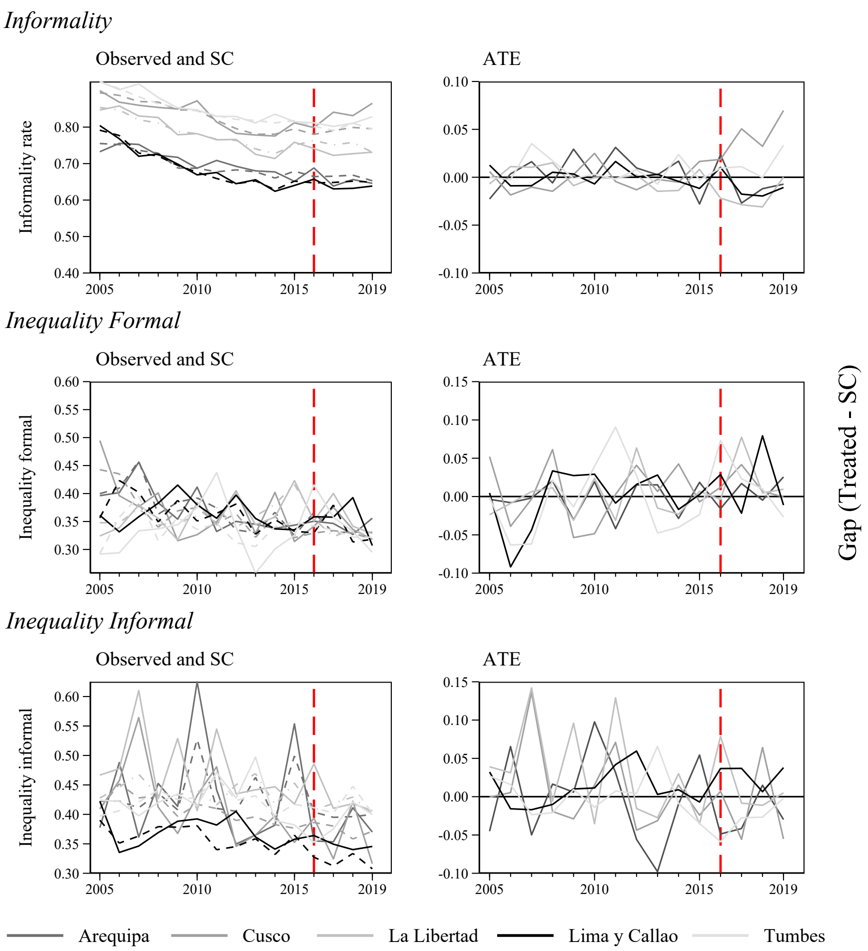
Notes: Sample restricted to only those employed between 18 and 65 years with at most secondary school. Real hourly wages in 2007 PEN. Controls included are gdp per capita, pea employed for low sckilled, pea employed in services, in manufacture, schooling, proportion of population between 18 and 25 and between 26 and 35, percentage of urban population. Equations include 2-way FEs (regions and years). SE Homosk. refers to homoskedasticity-only SEs; SE CRVE, to Cluster Robust Variance Estimator of SEs; SE CRVE2, to Bell and McCaffrey (2002) bias corrected CRVE2; P-val IK, to the p-value taking the effective DoF from Imbens and Kolesar (2012) (IK-DoF) taking CRVE2 SEs and P-val. WB, to the p-value of the Wild Cluster Boostrap taking the Rademacher weights using 199 replications. S* refers to the effective number of clusters from Carter, Schnepel and Steigerwald (2013). FE GLS-BC refers to the FE GLS Bias-corrected estimator of Hansen (2007) using SE CRVE. F-stat refers to the F test cluster robust of the set of interactions of the dummy for treatment qualification and the dummy for the years previous to the treatment. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B15 – SCM: SC weights for the ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019

	Informality rate					**Inequality formal**					**Inequality informal**				
	Areq.	Cusco	La Lib.	L y C.	Tumb.	Areq.	Cusco	La Lib.	L y C.	Tumb.	Areq.	Cusco	La Lib.	L y C.	Tumb.
Amazonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ancash	0	0	.422	0	0	.182	0	.176	0	0	0	.162	0	0	0
ApurÃmac	0	.249	0	0	0	0	.02	0	0	0	0	0	0	0	0
Ayacucho	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cajamarca	0	0	0	0	.056	0	.183	0	0	0	.536	0	0	0	0
Huancavelica	0	0	0	0	0	0	0	0	0	0	0	.225	0	0	0
HuÃ¡nuco	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ica	.259	0	.094	.261	0	.015	.316	0	.358	.075	0	0	.112	.489	.099
JunÃn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lambayeque	.355	0	.037	0	.094	.227	.146	.087	0	0	0	0	0	0	0
Loreto	0	0	0	0	.629	0	0	0	.333	.108	0	0	0	0	.336
Madre de Dios	0	.292	.147	.149	.15	0	0	.284	0	.163	0	.079	0	0	.309
Moquegua	.053	.193	0	.176	0	0	0	0	0	0	0	.083	.069	0	.047
Pasco	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Piura	0	0	0	0	0	0	.334	0	0	0	0	0	.207	0	.209
Puno	0	.266	0	0	0	0	0	.149	0	0	0	0	.609	0	0
San MartÃn	0	0	.145	0	0	0	0	0	0	0	0	0	0	0	0
Tacna	.139	0	.155	0	0	.216	0	0	0	0	.464	0	.003	.511	0
Ucayali	.194	0	0	.414	.071	.36	0	.304	.309	.654	0	.452	0	0	0

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years and with those with at most secondary school level before data aggregation. Inequality level measured by the Gini index of individual hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Figure 2.B5 – SCM: ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years with at most secondary school level before data aggregation. Real hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B16 – SCM: Covariate balance for the ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019

		Arequipa	Cusco	La Lib.	L y C.	Tumbes
Covariates						
GDP pc (thousands)	Obs.	15.924	11.714	9.441	16.380	9.498
	Wei. Inequality Formal	10.693	10.035	10.865	10.371	9.498
	Wei. Inequality Informal	11.303	12.152	9.807	16.119	13.253
	Wei. Informality	12.700	16.280	13.608	17.772	8.909
Empl. LF low skill	Obs.	0.536	0.678	0.667	0.551	0.678
	Wei. Inequality Formal	0.680	0.678	0.699	0.676	0.708
	Wei. Inequality Informal	0.687	0.719	0.680	0.585	0.689
	Wei. Informality	0.651	0.678	0.668	0.650	0.726
Empl. LF serv.	Obs.	0.644	0.549	0.588	0.734	0.683
	Wei. Inequality Formal	0.628	0.554	0.587	0.605	0.618
	Wei. Inequality Informal	0.550	0.547	0.514	0.659	0.604
	Wei. Informality	0.626	0.534	0.581	0.613	0.597
Empl. LF manuf.	Obs.	0.192	0.133	0.195	0.229	0.127
	Wei. Inequality Formal	0.160	0.167	0.146	0.155	0.147
	Wei. Inequality Informal	0.153	0.133	0.158	0.166	0.130
	Wei. Informality	0.167	0.123	0.150	0.157	0.124
Schooling (years)	Obs.	11.154	9.211	9.607	11.452	9.883
	Wei. Inequality Formal	9.837	9.543	9.587	10.042	9.682
	Wei. Inequality Informal	9.222	9.197	9.604	10.997	9.707
	Wei. Informality	10.219	9.572	9.776	10.227	9.390
Proport. 18-25 yo.	Obs.	0.168	0.154	0.183	0.189	0.177
	Wei. Inequality Formal	0.177	0.181	0.173	0.180	0.174
	Wei. Inequality Informal	0.163	0.164	0.180	0.176	0.173
	Wei. Informality	0.179	0.154	0.171	0.167	0.176
Proport. 26-35 yo.	Obs.	0.270	0.264	0.272	0.288	0.293
	Wei. Inequality Formal	0.270	0.265	0.273	0.280	0.282
	Wei. Inequality Informal	0.274	0.271	0.260	0.271	0.281
	Wei. Informality	0.267	0.266	0.272	0.275	0.287
Urban population	Obs.	0.881	0.535	0.776	0.979	0.922
	Wei. Inequality Formal	0.775	0.725	0.692	0.788	0.769
	Wei. Inequality Informal	0.573	0.621	0.612	0.878	0.736
	Wei. Informality	0.835	0.583	0.698	0.801	0.686
Ecu.-Col. border(d)	Obs.	0.000	0.000	0.000	0.000	1.000
	Wei. Inequality Formal	0.000	0.518	0.000	0.333	0.108
	Wei. Inequality Informal	0.536	0.000	0.207	0.000	0.545
	Wei. Informality	0.000	0.000	0.000	0.000	0.685
Lagged Outcomes						
2005	Obs. Inequality Formal	0.396	0.495	0.324	0.361	0.292
	Wei. Inequality Formal	0.400	0.443	0.348	0.356	0.293
	Obs. Inequality Informal	0.378	0.409	0.467	0.422	0.424
	Wei. Inequality Informal	0.423	0.409	0.428	0.390	0.397
	Obs. Informality	0.733	0.900	0.847	0.804	0.925
	Wei. Informality	0.755	0.894	0.854	0.792	0.925
2010	Obs. Inequality Formal	0.412	0.326	0.384	0.380	0.375
	Wei. Inequality Formal	0.392	0.375	0.360	0.351	0.336
	Obs. Inequality Informal	0.625	0.432	0.406	0.392	0.415
	Wei. Inequality Informal	0.527	0.417	0.441	0.381	0.429
	Obs. Informality	0.687	0.872	0.782	0.669	0.847
	Wei. Informality	0.687	0.847	0.782	0.676	0.844
2015	Obs. Inequality Formal	0.343	0.314	0.423	0.335	0.325
	Wei. Inequality Formal	0.325	0.321	0.416	0.332	0.349
	Obs. Inequality Informal	0.554	0.352	0.418	0.358	0.382
	Wei. Inequality Informal	0.499	0.376	0.453	0.365	0.417
	Obs. Informality	0.651	0.812	0.759	0.640	0.815
	Wei. Informality	0.679	0.795	0.750	0.652	0.818

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years and at most secondary school level. All variables (except the lagged outcomes) are averaged for the 2005-2015 period. LF=Labour force; Ecu-Col border is a dummy that equals 1 if the region neighbours Ecuador and Colombia. Inequality level measured by the Gini index of individual hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B17 – SCM: DiD Regressions for the ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019

	GLS with AR(1) disturbances					*Baltagi and Wu (1999) estimator*				
	(2005-14)	2016	2017	2018	2019	(2005-14)	2016	2017	2018	2019
Formal Wages										
Arequipa	-0.050	-0.048	0.054	0.054	0.147	-0.089	-0.029	0.082	0.085	0.175
	(0.118)	(0.118)	(0.161)	(0.190)	(0.212)	(0.106)	(0.111)	(0.139)	(0.153)	(0.161)
Cusco	0.083	-0.015	-0.037	0.334	0.274	0.072	-0.009	-0.028	0.343*	0.283
	(0.155)	(0.155)	(0.213)	(0.254)	(0.285)	(0.138)	(0.142)	(0.181)	(0.203)	(0.217)
La Libertad	0.036	0.079	0.087	0.068	0.119	0.042	0.079	0.088	0.070	0.122
	(0.108)	(0.115)	(0.141)	(0.152)	(0.158)	(0.084)	(0.095)	(0.111)	(0.117)	(0.119)
Lima y Callao	-0.012	0.082	0.095	0.121	0.131	-0.039	0.090	0.104	0.126	0.132
	(0.102)	(0.103)	(0.136)	(0.155)	(0.168)	(0.084)	(0.089)	(0.110)	(0.119)	(0.124)
Tumbes	-0.003	-0.033	-0.103	-0.134	-0.055	-0.029	-0.034	-0.109	-0.146	-0.071
	(0.121)	(0.128)	(0.158)	(0.172)	(0.179)	(0.096)	(0.108)	(0.126)	(0.132)	(0.135)
Informal Wages										
Arequipa	0.022	0.020	-0.185	0.164	0.067	0.001	0.034	-0.164	0.190	0.094
	(0.162)	(0.162)	(0.223)	(0.266)	(0.300)	(0.148)	(0.151)	(0.194)	(0.219)	(0.234)
Cusco	-0.175	-0.143	-0.222	-0.205	-0.066	-0.187	-0.135	-0.210	-0.190	-0.050
	(0.192)	(0.192)	(0.266)	(0.320)	(0.363)	(0.179)	(0.182)	(0.237)	(0.269)	(0.290)
La Libertad	0.091	-0.071	0.051	0.043	0.010	0.075	-0.061	0.066	0.059	0.027
	(0.121)	(0.121)	(0.167)	(0.201)	(0.228)	(0.117)	(0.121)	(0.153)	(0.170)	(0.180)
Lima y Callao	0.051	0.060	0.012	-0.119	-0.064	0.040	0.068	0.026	-0.102	-0.044
	(0.105)	(0.105)	(0.145)	(0.173)	(0.196)	(0.094)	(0.096)	(0.124)	(0.141)	(0.151)
Tumbes	0.162	0.143	0.077	-0.023	-0.052	0.146	0.154	0.095	-0.000	-0.027
	(0.122)	(0.122)	(0.169)	(0.203)	(0.229)	(0.114)	(0.116)	(0.150)	(0.169)	(0.181)

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with at most secondary school level and, in the case of the informal outcome, additionally restrict the sample to those between 18 and 35 years before data aggregation. Each row represents a regression of annual observations for the treated areas and its synthetic control between 2005 and 2019, N=30, which includes a dummy variable for the treated region, period dummies (2005-2014, 2016, 2017, 2018, 2019 with 2015 excluded) and their interactions; these interaction coefficients are the coefficients reported. Real hourly wages in 2007 PEN * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B18 – SCM: P-values for the ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019

	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
1 Informality					
Arequipa	-0.0056	0.0186	1.0303	14/20	0.7000
Cusco	0.0428	0.0129	3.6468	3/20	0.1500
La Libertad	-0.0203	0.0099	2.4121	4/20	0.2000
Lima y Callao	-0.0094	0.0087	1.7318	4/20	0.2000
Tumbes	0.0135	0.0143	1.2823	10/20	0.5000
2 Inequality Formal					
Arequipa	0.0055	0.0217	0.8037	9/20	0.4500
Cusco	0.0150	0.0391	0.5631	16/20	0.8000
La Libertad	0.0220	0.0272	1.4389	2/20	0.1000
Lima y Callao	0.0188	0.0363	1.2098	6/20	0.3000
Tumbes	0.0199	0.0462	0.8875	10/20	0.5000
3 Inequality Informal					
Arequipa	-0.0263	0.0567	0.6368	15/20	0.7500
Cusco	-0.0099	0.0512	0.9916	12/20	0.6000
La Libertad	0.0160	0.0694	0.5777	14/20	0.7000
Lima y Callao	0.0295	0.0259	1.2525	8/20	0.4000
Tumbes	-0.0288	0.0268	1.3115	5/20	0.2500

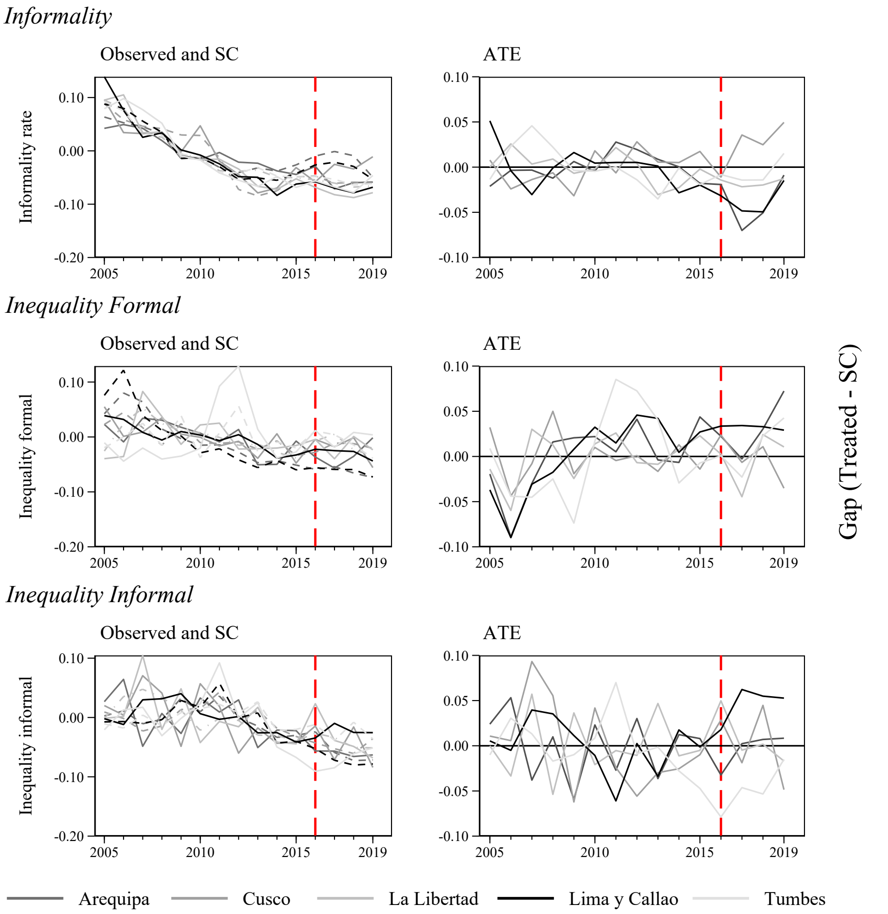
Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with at most secondary school level before data aggregation. ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2005-2019 data.

Table 2.B19 – SCM demeaned: SC weights for the ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019

	Informality rate					**Inequality**					**Occupational complexity**				
	Areq.	Cusco	La Lib.	L y C.	Tumb.	Areq.	Cusco	La Lib.	L y C.	Tumb.	Areq.	Cusco	La Lib.	L y C.	Tumb.
Amazonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ancash	0	0	.073	0	0	.049	0	.292	0	0	0	.319	0	0	0
ApurÃmac	0	.153	0	0	0	0	0	0	0	0	0	.164	0	0	0
Ayacucho	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cajamarca	0	0	0	0	0	0	.045	0	0	0	.014	0	0	0	0
Huancavelica	0	0	0	0	0	0	.128	0	0	0	0	.167	0	0	0
HuÃ¡nuco	0	.345	.118	0	0	0	.243	0	0	0	0	0	0	0	0
Ica	.292	.182	.321	.584	.091	.641	0	.139	.867	0	0	.039	.18	.664	.128
JunÃn	0	0	0	0	0	0	0	0	0	0	0	0	.321	0	0
Lambayeque	.469	0	.18	0	.099	0	0	.029	0	0	0	0	0	0	0
Loreto	0	0	0	0	.117	0	0	0	0	.194	0	0	0	0	.035
Madre de Dios	0	.289	0	0	.221	0	0	.191	0	.021	0	.21	0	0	.001
Moquegua	.14	.031	0	.213	0	.275	.012	0	.101	.066	0	.101	0	0	0
Pasco	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Piura	0	0	.079	0	.114	0	.127	0	0	0	0	0	.098	0	.233
Puno	0	0	0	0	0	0	0	.088	0	0	0	0	0	0	0
San MartÃn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tacna	.1	0	0	0	.055	0	.443	0	0	.192	.986	0	0	.336	.256
Ucayali	0	0	.228	.203	.304	.036	0	.261	.032	.528	0	0	.402	0	.346

Notes: Estimations restrict the sample to only those employed between 18 and 65 years with at most secondary school level before data aggregation. Inequality level measured by the Gini index of individual hourly wages in 2007 PEN. For all these estimations, we first demeaned the data following the routine suggested by Ferman & Pinto (2019).*Source*: Author's calculations using ENAHO 2011-2019 data.

Figure 2.B6 – SCM demeaned: ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019



Note: Sample restricted to only those employed between 18 and 65 years before data aggregation. Real hourly wages
In all of these, we first applied the demeaning routine suggested by Ferman & Pinto (2019). *Source*: Author's calcula
ENAHO 2005-2019 data.

Table 2.B20 – SCM demeaned: Covariate balance for the ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019

		Arequipa	Cusco	La Lib.	L y C.	Tumbes
Covariates						
GDP pc (thousands)	Obs.	15.924	11.714	9.441	16.380	9.498
	Wei. Inequality Formal	23.708	11.236	12.117	18.366	11.992
	Wei. Inequality Informal	16.594	15.120	8.993	15.898	11.063
	Wei. Informality	15.980	11.326	10.164	20.467	10.790
Empl. LF low skill	Obs.	0.536	0.678	0.667	0.551	0.678
	Wei. Inequality Formal	0.581	0.682	0.680	0.579	0.690
	Wei. Inequality Informal	0.594	0.695	0.673	0.582	0.668
	Wei. Informality	0.633	0.693	0.672	0.602	0.692
Empl. LF serv.	Obs.	0.644	0.549	0.588	0.734	0.683
	Wei. Inequality Formal	0.597	0.568	0.584	0.597	0.636
	Wei. Inequality Informal	0.716	0.521	0.582	0.637	0.633
	Wei. Informality	0.619	0.525	0.581	0.603	0.618
Empl. LF manuf.	Obs.	0.192	0.133	0.195	0.229	0.127
	Wei. Inequality Formal	0.177	0.128	0.156	0.184	0.146
	Wei. Inequality Informal	0.145	0.124	0.159	0.174	0.156
	Wei. Informality	0.169	0.121	0.165	0.175	0.144
Schooling (years)	Obs.	11.154	9.211	9.607	11.452	9.883
	Wei. Inequality Formal	10.909	9.354	9.771	11.040	9.849
	Wei. Inequality Informal	10.838	9.254	9.873	11.038	9.962
	Wei. Informality	10.343	9.230	9.811	10.740	9.756
Proport. 18-25 yo.	Obs.	0.168	0.154	0.183	0.189	0.177
	Wei. Inequality Formal	0.168	0.168	0.174	0.178	0.172
	Wei. Inequality Informal	0.169	0.157	0.183	0.179	0.179
	Wei. Informality	0.177	0.163	0.181	0.170	0.176
Proport. 26-35 yo.	Obs.	0.270	0.264	0.272	0.288	0.293
	Wei. Inequality Formal	0.266	0.271	0.271	0.266	0.282
	Wei. Inequality Informal	0.277	0.265	0.269	0.269	0.274
	Wei. Informality	0.262	0.271	0.267	0.269	0.277
Urban population	Obs.	0.881	0.535	0.776	0.979	0.922
	Wei. Inequality Formal	0.843	0.627	0.708	0.879	0.775
	Wei. Inequality Informal	0.853	0.554	0.758	0.884	0.808
	Wei. Informality	0.839	0.588	0.764	0.846	0.773
Ecu.-Col. border(d)	Obs.	0.000	0.000	0.000	0.000	1.000
	Wei. Inequality Formal	0.000	0.172	0.000	0.000	0.194
	Wei. Inequality Informal	0.014	0.000	0.098	0.000	0.268
	Wei. Informality	0.000	0.000	0.079	0.000	0.231
Lagged Outcomes						
2005	Obs. Inequality Formal	0.022	0.121	-0.037	-0.003	-0.039
	Wei. Inequality Formal	0.081	0.062	0.001	0.104	-0.020
	Obs. Inequality Informal	-0.060	-0.008	-0.006	0.050	-0.002
	Wei. Inequality Informal	0.019	-0.006	-0.006	0.031	0.009
	Obs. Informality	0.025	0.067	0.060	0.111	0.066
	Wei. Informality	0.051	0.061	0.058	0.094	0.068
2010	Obs. Inequality Formal	0.038	-0.047	0.023	0.015	0.045
	Wei. Inequality Formal	-0.025	-0.041	0.008	-0.030	-0.007
	Obs. Inequality Informal	0.188	0.015	-0.067	0.019	-0.011
	Wei. Inequality Informal	0.063	0.021	-0.015	0.003	0.016
	Obs. Informality	-0.020	0.039	-0.005	-0.024	-0.012
	Wei. Informality	-0.005	0.019	-0.007	-0.012	-0.010
2015	Obs. Inequality Formal	-0.030	-0.059	0.062	-0.029	-0.005
	Wei. Inequality Formal	-0.051	-0.052	0.014	-0.059	-0.032
	Obs. Inequality Informal	0.116	-0.064	-0.055	-0.015	-0.043
	Wei. Inequality Informal	0.064	-0.025	-0.039	-0.020	-0.012
	Obs. Informality	-0.056	-0.022	-0.028	-0.053	-0.044
	Wei. Informality	-0.027	-0.035	-0.029	-0.030	-0.036

Notes: Estimations restrict the sample for the outcome to only those employed between 18 and 65 years with at most primary or secondary school level before data aggregation. All variables (except the lagged outcomes) are averaged for the 2005-2015 period. LF=Labour force; Ecu-Col border is a dummy that equals 1 if the region neighbours Ecuador and Colombia. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). Inequality level measured by the Gini index of individual hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.B21 – SCM demeaned: DiD Regressions for the ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019

	GLS with AR(1) disturbances					*Baltagi and Li (1991) estimator*				
	(2005-14)	2016	2017	2018	2019	(2005-14)	2016	2017	2018	2019
1 Informality										
Arequipa	0.041	0.051*	0.006	0.018	0.027	0.038	0.052**	0.008	0.020	0.029
	(0.028)	(0.028)	(0.038)	(0.044)	(0.049)	(0.024)	(0.024)	(0.031)	(0.034)	(0.037)
Cusco	-0.020	-0.014	0.032	0.014	0.058	-0.018	-0.014	0.031	0.013	0.056
	(0.033)	(0.033)	(0.045)	(0.054)	(0.060)	(0.030)	(0.032)	(0.040)	(0.044)	(0.046)
La Libertad	-0.026	-0.014	-0.031	-0.029	-0.009	-0.019	-0.018	-0.037	-0.036	-0.017
	(0.030)	(0.030)	(0.042)	(0.050)	(0.057)	(0.029)	(0.030)	(0.038)	(0.043)	(0.045)
Lima y Callao	0.013	0.010	-0.022	-0.022	0.007	0.017	0.008	-0.025	-0.025	0.003
	(0.040)	(0.041)	(0.056)	(0.066)	(0.075)	(0.040)	(0.042)	(0.052)	(0.058)	(0.061)
Tumbes	0.036	0.014	0.010	0.013	0.036	0.030	0.017	0.016	0.020	0.044
	(0.028)	(0.028)	(0.039)	(0.047)	(0.053)	(0.027)	(0.027)	(0.035)	(0.039)	(0.042)
2 Inequality formal										
Arequipa	-0.026	0.031	0.022	0.024	0.060	-0.026	0.030	0.021	0.023	0.059
	(0.060)	(0.064)	(0.078)	(0.084)	(0.087)	(0.047)	(0.053)	(0.062)	(0.065)	(0.066)
Cusco	0.029	-0.000	0.074	0.019	0.036	0.022	-0.001	0.072	0.016	0.033
	(0.071)	(0.078)	(0.094)	(0.100)	(0.102)	(0.056)	(0.066)	(0.074)	(0.077)	(0.078)
La Libertad	-0.051	-0.034	0.047	-0.014	-0.013	-0.051**	-0.034	0.047	-0.014	-0.013
	(0.031)	(0.048)	(0.042)	(0.043)	(0.043)	(0.024)	(0.037)	(0.032)	(0.033)	(0.033)
Lima y Callao	-0.024	0.055	0.039	0.116	0.028	-0.028	0.054	0.035	0.110	0.021
	(0.063)	(0.066)	(0.083)	(0.090)	(0.095)	(0.051)	(0.059)	(0.068)	(0.071)	(0.071)
Tumbes	-0.038	0.057	0.003	-0.002	-0.039	-0.036	0.057	0.004	-0.001	-0.038
	(0.061)	(0.069)	(0.081)	(0.084)	(0.086)	(0.047)	(0.056)	(0.063)	(0.065)	(0.066)
3 Inequality informal										
Arequipa	-0.057	-0.097	-0.087	-0.093	-0.069	-0.057	-0.097	-0.087	-0.093	-0.069
	(0.094)	(0.129)	(0.127)	(0.127)	(0.127)	(0.073)	(0.103)	(0.098)	(0.099)	(0.099)
Cusco	0.042	0.036	-0.030	0.082	-0.035	0.044	0.038	-0.029	0.084	-0.034
	(0.071)	(0.092)	(0.096)	(0.097)	(0.097)	(0.056)	(0.076)	(0.075)	(0.075)	(0.075)
La Libertad	0.015	0.065	-0.013	0.005	-0.012	0.018	0.069	-0.010	0.008	-0.008
	(0.074)	(0.114)	(0.098)	(0.103)	(0.101)	(0.059)	(0.080)	(0.080)	(0.080)	(0.080)
Lima y Callao	-0.006	0.019	0.020	0.006	0.022	-0.005	0.020	0.022	0.007	0.024
	(0.036)	(0.044)	(0.049)	(0.049)	(0.050)	(0.028)	(0.037)	(0.038)	(0.038)	(0.038)
Tumbes	0.037	-0.005	0.016	0.002	0.049	0.033	-0.008	0.014	-0.001	0.046
	(0.039)	(0.056)	(0.052)	(0.053)	(0.053)	(0.031)	(0.040)	(0.041)	(0.042)	(0.042)

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years with at most secondary school level before data aggregation. Each row represents a regression of annual observations for the corresponding treated area and its synthetic control between 2005 and 2019, N=30, which includes a dummy variable for the treated region, period dummies (2005-2014, 2016, 2017, 2018, 2019 with 2015 excluded) and their interactions; these interaction coefficients are the coefficients reported. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). Inequality level measured by the Gini index of individual hourly wages in 2007 PEN. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO 2011-2019 data.

Table 2.B22 – SCM demeaned: P-values for the ATET of Venezuelan immigration on low skilled natives' informality rate and inequality in formal and informal sector, 2005-2019

	ATE	RMSPE pre	Ratio post-pre	Rank	p-value
1 Informality					
Arequipa	-0.0072	0.0180	1.0307	11/20	0.5500
Cusco	0.0371	0.0100	4.5484	1/20	0.0500
La Libertad	-0.0170	0.0181	1.0771	10/20	0.5000
Lima y Callao	-0.0271	0.0129	2.4071	3/20	0.1500
Tumbes	0.0066	0.0148	0.7789	14/20	0.7000
2 Inequality Formal					
Arequipa	0.0537	0.0467	1.1933	5/20	0.2500
Cusco	0.0138	0.0527	0.5715	15/20	0.7500
La Libertad	0.0434	0.0316	1.6743	2/20	0.1000
Lima y Callao	0.0769	0.0719	1.1678	5/20	0.2500
Tumbes	0.0370	0.0522	0.9542	6/20	0.3000
3 Inequality Informal					
Arequipa	-0.0345	0.0649	0.5563	15/20	0.7500
Cusco	-0.0251	0.0614	0.8938	7/20	0.3500
La Libertad	-0.0019	0.0509	0.6327	12/20	0.6000
Lima y Callao	0.0231	0.0219	1.0940	7/20	0.3500
Tumbes	-0.0179	0.0408	0.6726	11/20	0.5500

Notes: Estimations under the total heading restrict the sample for the outcome to only those employed between 18 and 65 years with at most primary or secondary school level before data aggregation. In all of these, we first demeaned the data following the routine by Ferman & Pinto (2019). ATE shows to the average TE for the post-treatment periods; RMSPE, the root mean square prediction error for the pre-treatment period (2005-2015); ratio, the ratio of the RMSPE for the pre-treatment relative to the RMSPE for the post-treatment (2016-2019); rank, the relative position of the ratio among the J+1 units and p-value, the probability that a random draw from the donor pool takes a lower than the value for the corresponding treated region. *Source*: Author's calculations using ENAHO 2011-2019 data.

Chapter 3

A Dual Approach to the Effect of Discrimination Against Venezuelans in Peru

3.1 Introduction

The opportunity provided by a large and sudden influx of foreign workers into a country has been exploited by labour economists in the past to study how exogenous labour supply shocks affects outcomes for natives. Notable examples of this include Card (1990), Hunt (1992) or Carrington and de Lima (1996). In contrast, immigrants' outcomes have been less explored, with the role of discrimination particularly neglected. While this is explained by the fact that changes in the native wage dominate the political discussion (Dustmann et al. 2019), the inherent complexities in defining and measuring discrimination usually hampers its study (Blank et al. 2004; Heckman 1998; Altonji and Blank 1999). However, compelling evidence suggests that discrimination against migrants, whether perceived by them or factually exercised against them, has a critical impact on different areas. It includes not only the immigrants' well-being (Schmitt et al. 2014; Nandi et al. 2016; Safi 2010) but also social cohesion (de Vroome et al. 2014; Auer and Ruedin 2019), employment and labour productivity (Di Stasio et al., 2021; Oreopoulos, 2011; Baert, 2017; Ensher et al., 2001) within the host society.

The emphasis placed by the studies around the effects of the Venezuelan Exodus to Peru in 2016 illustrates the above. This migratory episode, of a magnitude second only to the Syrian refugee crisis, turned Peru overnight into the second-largest recipient of Venezuelans in the world (UNHCR 2020, 2019d). The quasi-experimental evidence in Del Pozo Segura (2021), Boruchowicz et al. (2021) and Morales and Pierola (2020) focus on the effect of this migrant shock across a comprehensive array of labour market outcomes for native workers and also for low-skilled nationals. In turn, the study of how this migrant population, which represents a non-negligible 2.5% of Peru's population, has inserted into the host labour market remains under-explored. This is an important concern, as the natives' reaction to this migrants influx has been unfavourable (PUCP 2020, 2019; UNHCR 2019b). For instance, Carroll et al. (2020) and Mougenot et al. (2021) document the progressive deterioration of their mental health since their arrival to Peru, some of which is attributed to self-perceived discrimination. This concern adds to the importance of directly analysing how discrimination by Peruvians has affected Venezuelans' wages and their perceptions of discrimination and further disentangle the mechanisms behind those impacts.

In view of this, firstly, we analyze how Peruvians' attitudes affect the wages of Venezuelan migrants in terms of both self-perceived and objectively measured discrimination. In order to estimate the relevant perceived and actual treatment effects, we apply a novel decomposition technique suggested by Firpo et al. (2018). Secondly, we study how their self-perceived discrimination varies with the treatment effect once we control for characteristics that psychological theories, such as relative deprivation (Dion 1986; Dion and Kawakami 1996), social identity (Tajfel and Turner 1986) and prototypic conceptions (Inman et al. 1998; Inman and Baron 1996) regard as important predictors of migrants' expectations of equal treatment and their awareness of social inequalities (Banerjee 2008). We do this using two representative surveys conducted by Peru's National Statistical Office, including the 2018 Venezuelan Resident Population Survey (ENPOVE). This latter provides detailed and unique information on the labour market activity of Venezuelans after they arrived in 2016.

The results of this chapter can be split into three. Firstly, the proportion of Venezuelans that perceive discrimination is as large as that prevalent in countries with negligible shares of migrant workers and is more extensive than what migrants report in OECD countries. This supports theories that explain negative attitudes towards migrants linked to their origin, regardless of their individual characteristics. Secondly, we find evidence of a treatment differential among Venezuelans regarding perceived discrimination and more significant treatments between Venezuelans and Peruvians. Concerning the latter, these treatments are found to widen sharply across the unconditional wage distribution. Thirdly, the perception of discrimination of Venezuelans is actually influenced by the objective discrimination experienced. In other words, they are aware of the level of unequal treatment they encounter, which contrasts with some of the previous literature. However, the magnitude is not large, as variables that reflect migrants' expectations for equality (e.g., education and experience) have a more sizeable effect.

Our paper contributes to the existing literature in three ways. The first is by filling the literature gap related to short-run labour market outcomes for immigrants who arrive in conditions similar to a natural experiment. Indeed, as shown below, most of the studies analysing the migrants-natives pay gap as well as the determinants of migrants' self-perceived discrimination are framed within long and progressive migration processes and are undertaken for developed economies. Instead, compared to the Syrian migration into Turkey (Tumen 2015) or the Cuban inflow into Miami (Card 1990), the Venezuelan Exodus is unique in that these migrants inserted primarily into the informal economy, are more educated than the natives and do not face a language barrier nor a significant distance relative to their host society in terms of religion and culture. The second contribution is to further the studies that have analyzed migrant wage gaps or perceptions of discrimination across the unconditional distribution of wages (e.g. Biddle (2013); Banerjee (2008); Chiswick et al. (2008); Auer et al. (2017)). This provides a more detailed portrait of how discrimination affects Venezuelans. Compared with evidence arising from classic decomposition methods (DiNardo et al. 1996; Oaxaca 1973; Blinder 1973), the method we apply is robust to departures from linearity of the population conditional expectation (see Barsky et al. 2002) and relies on weaker identifying assumptions than alternative methods (e.g. Machado and Mata 2005). This latter is essential for policy purposes, as the size of the wage-structure relative to the composition component suggest different strategies for tackling down the pay gap of migrants relative to natives.

The text is organized as follows. section 3.2 briefly describes the Venezuelans' insertion into the Peruvian labour market regarding their perceived discrimination and how this relates to their observed labour market characteristics and wages. section 3.3 discusses the challenges involved in identifying the migrant labour market discrimination and reviews the findings of studies focused on the native-migrant wage gap and the impact of perceived discrimination on their wages. section 3.4 describes the sources of data used for the regression. section 3.5 discusses the econometric methods employed, both for estimation and decomposition. section 3.6 presents estimations of the impact of discrimination on the migrant's wages under the two methodological approaches adopted. Finally, section 3.7 discusses the results in terms of policy implications and points areas for future research.

3.2 Venezuelan integration into the host labour market

As explained in the previous chapter, the Venezuelan Exodus represents a crisis of similar magnitude to that represented by the Syrian crisis, and which was labelled by the United Nation's High Commissioner for the Refugees as the "largest external displacement crisis in Latin America's recent history" (UNHCR 2019d). As of January 2020, approximately 4.5 million Venezuelans had left their country due to socio-economic uncertainty and political turmoil (see Reinhart and Santos 2015; Restuccia 2019; IAHCR 2017; HRW 2018). Over 80% of migrants remained in Latin America and the Caribbean (R4V 2020b). Among these, Peru, one of the fastest-growing countries in the region (Ross and Peschiera 2015) and one of the least used to receiving migrants (Torales et al. 2003), played a unique role by turning overnight into one of the leading destinations for Venezuelans. By January 2020 (see Figure 3.A2), it was both the second-largest recipient of Venezuelan migrants who comprise around 2.5% of the total native population and the largest recipient of requests for asylum-seeking Venezuelans (receiving half of these applications, R4V 2020a). Note that the former proportion is only second to Colombia (3.5%, see UNHCR 2019d) and is similar to the Syrian in Turkey in 2015 (2.1%, see Tumen 2015).

Given Peru's historically high informal employment rates (around 65% by the time the influx began, INEI 2020), the government adopted a series of measures (see table A1 in chapter two for a detailed description) to absorb these new migrants (80% of which are of working age) into the formal sector. Nevertheless, the most important of these, known as the Temporary Permit of Permanence (PTP), was unsuccessful because it implied long waiting times for Venezuelan applicants who needed to enter the labour market (even informally) upon arrival in order to send remittances back home (see below).[1] Concurrently, most Peruvian employers hired Venezuelans as informal workers because of the poorer opportunities provided by the comparatively tight Peruvian labour legislation to hire foreigners (Geronimi 2004).[2] In addition, the deficient law enforcement (Viollaz 2019) allowed them to ignore the working permits issued for Venezuelan job applicants (e.g. the Extraordinary Working Permit or the official Refugee-seeker Card), even if this contravened Peru's immigration rules (IDEHPUCP et al. 2020). These factors were coupled with an additional restriction mandating that, from August 2018, only those with a valid passport could enter Peru. This represented an insurmountable hurdle for most of them since its processing fee within Venezuela, where the basic monthly wage is lower than five USD, oscillated between 2,000 and 5,000 USD (see Blouin 2019).[3] Thus, around 95% of Venezuelan workers ended up in the informal sector, characterized by an absence of regulations and low firm productivity (La Porta and Shleifer 2014; ILO 2015; Pages 2010). Overall, their situation mirrors what Syrian migrants faced when arriving in Turkey (Akgündüz et al. 2015; Ceritoglu et al. 2015).

This sudden influx of a vast amount of Venezuelans in the labour market generated an adverse reaction from Peruvians, despite that the close historical and cultural ties of the migrants with the host country would have a priori suggested otherwise (Dancygier and Laitin 2014). Although the average native feels indifferent towards these immigrants (column 1 in Table 3.1), most do not welcome the arrival of more Venezuelans into the country. Slightly more than half of the natives defends the idea that the government

[1] This is due to long time it takes to obtain, firstly, the documentation from different government institutions (including the INTERPOL international exchange token) and, secondly, the confirmation of the change from tourist-permit holder to PTP holder from Peru's Migration Authority, requiring an additional 6 months (Blouin 2019).

[2] These norms are stated in the Legislative Decree No. 689 of 1991, Law for the hiring of foreign workers, and the Supreme Decree No. 014-92-TR of 1992. The Supreme Decree No. 179-2004-EF defines the special income tax regime, which applies for foreigner both if self-employed (based on their net income) or employees (based on their total income). On the one hand, there are these cannot represent more than 20% of the total number of workers and their wages cannot exceed 30% of the total payroll in the firm. On the other hand, foreigners who do not have permanent residence in Peru are subject to a special income tax regime of 30%.

[3] Further restrictions, implemented after this study's time span, made their legal entry into Peru even more difficult. From 15 of June 2019, Venezuelans must also have a Visa or a Humanitarian Visa in addition to a valid passport, which could only be requested at the very same Consulates (in Venezuela, Colombia and Ecuador) dealing with an excessive amount of applications from Venezuelans trying to enter to Peru (UNHCR 2019b). Additionally, application for refugee status (which would obviate the need for the Humanitarian Visa) could only be made once they reach the border via an interview, during which they could not enter Peru (IDEHPUCP et al. 2020).

should impose a quota on Venezuelan workers, and one in four think instead that they should be banned (columns 2-5).[4] In line with what Gorodzeisky and Semyonov 2019; Wallace and Figueroa 2012; Fasani et al. 2019; ILO 2019 find in Europe and the USA, respectively, natives' opinions are explained by a belief that these immigrants both harm the economy and take jobs that "legitimately belong to Peruvians" (see columns 6-8 in Table 3.1). In contrast to what prevails in the OECD and higher-income countries among those higher-educated (Grigorieff et al. 2020), the table also reveals that the majority of highly-educated Peruvians regard immigrants as a threat to their jobs. This occurs notwithstanding international evidence suggesting that they are the least affected by immigration (Dustmann et al. 2008).[5]

Table 3.1 – Peruvians' attitudes and beliefs towards Venezuelan immigrants

	Unlikeable	**Ideal immigration policy**			**Beliefs about Venezuelans**				
		CoJ	Quota	Ban	Harm econ.	Take jobs	Uneduc.	Unexp.	Delinq.
Sex									
Male	2.10	47.71	50.80	47.43	74.65	76.20	60.68	47.12	55.97
	(0.02)	(2.39)	(1.42)	(1.91)	(1.24)	(1.21)	(1.39)	(1.42)	(1.42)
Female	2.09	52.29	49.20	52.57	75.84	78.32	62.47	50.51	56.41
	(0.02)	(2.39)	(1.42)	(1.91)	(1.19)	(1.15)	(1.35)	(1.39)	(1.38)
Age (Years)									
18-25	2.02	25.46	23.23	16.89	75.27	76.91	57.45	44.91	57.82
	(0.03)	(2.09)	(1.20)	(1.44)	(1.84)	(1.80)	(2.11)	(2.12)	(2.11)
26-35	2.09	23.39	26.05	24.38	75.88	78.30	65.43	49.52	56.27
	(0.03)	(2.03)	(1.24)	(1.65)	(1.72)	(1.65)	(1.91)	(2.01)	(1.99)
36-45	2.08	22.94	21.54	21.00	73.54	76.82	57.85	49.64	56.57
	(0.03)	(2.01)	(1.17)	(1.56)	(1.89)	(1.80)	(2.11)	(2.14)	(2.12)
46-65	2.17	28.21	29.18	37.74	75.94	77.07	64.04	50.50	54.76
	(0.03)	(2.16)	(1.29)	(1.86)	(1.51)	(1.49)	(1.70)	(1.77)	(1.76)
Education Level									
Primary	2.23	19.50	16.00	34.80	76.91	78.61	75.72	62.82	53.99
	(0.05)	(1.90)	(1.04)	(1.83)	(1.74)	(1.69)	(1.77)	(1.99)	(2.06)
Secondary	2.14	47.71	42.20	46.84	75.92	77.26	61.59	47.99	57.48
	(0.02)	(2.39)	(1.40)	(1.91)	(1.28)	(1.25)	(1.46)	(1.50)	(1.48)
Technical	2.03	16.28	20.90	11.75	76.85	81.86	50.84	43.91	59.43
	(0.03)	(1.77)	(1.15)	(1.23)	(2.06)	(1.88)	(2.45)	(2.43)	(2.40)
College	1.97	16.51	20.90	6.61	69.21	70.48	51.91	35.62	52.42
	(0.03)	(1.78)	(1.15)	(0.95)	(2.33)	(2.30)	(2.52)	(2.42)	(2.52)
Total	2.10	18.05	51.51	28.20	75.26	77.28	61.60	48.85	56.20
	(0.02)	(0.78)	(1.02)	(0.92)	(0.86)	(0.84)	(0.97)	(1.00)	(0.99)

Notes: In the first column, rows shows the average of a measure of 'how unlikeable are the Venezuelan immigrants' that goes from 1 (totally like them) to 4 (totally dislike them). In columns 2 to 5, rows show the relative frequency (%) among those who consider as ideal the different immigration policies for Venezuelan immigrants. In columns 6 to 8, rows show the proportion (%) who share different beliefs about Venezuelan immigrants in Peru within the corresponding group. CoJ = Conditional on jobs; Unex. = lack labour market experience; Uneduc. = lack enough schooling; Delinq. = are engaged in criminal activities. SEs in parenthesis. *Source*: Author's calculations using IOP-PUCP data (2019).

A peculiar feature of the Exodus is that in contrast to the studies undertaken by Card (2001), Carrington and de Lima (1996) or Akgündüz et al. (2015), Venezuelan immigrants are more educated than the natives. Specifically, almost 60% of Venezuelan workers have higher education (left panel in Table 3.2), whereas almost half of Peruvian workers (47%) have only secondary education. However, these believe that Venezuelans are mostly uneducated ("Uneduc." heading in Table 3.1). This misperception extends to their past labour market experience (next to the last column of Table 3.1), despite that four out of five Venezuelans have had previous job experience (Table 3.2), and to their criminality (last column in the table), despite that there is not evidence of excessively high crime rates in the regions where Venezuelans have located (Bahar et al. 2020a). A follow-up round of the IOP-PUCP survey (PUCP 2020) suggests that Peruvians still hold these beliefs. Nonetheless, there is no evidence of a reversion to a "type" of Venezuelan immigrant

[4] Similar evidence is provided by a survey conducted by IEP (2019) during the same year: 73% opposes Venezuelan immigration, and 66% agrees with the further immigration restrictions imposed in 2019 (see table table A1 in Chapter Two). The IOP-PUCP survey conducted one year later reveals that the negative attitude of natives towards Venezuelans had intensified relative to the survey reported here (see PUCP 2020).

[5] Along these lines, IEP (2019) reports that 33% of Peruvians directly blame Venezuelans for the loss of jobs of Peruvians, and 75% of those who oppose Venezuelan immigration lies on the belief that they will harm the economy.

that actually matches their negative beliefs (see bottom panel of Figure 3.A3 in Appendix). This natives' mismatch between beliefs and facts about immigrants, however, is not exclusive to Peru as it coincides with has been documented for developed economies with a higher proportion of immigrants such as the USA, Great Britain and Continental Europe (Grigorieff et al. 2020; Blinder 2015; Citrin and Sides 2008). These studies confirm that migrants from less developed countries are often perceived more negatively regardless of their individual characteristics (Kustov 2019).

Given the foregoing, it is not surprising to find that Venezuelans have been more likely to experience discrimination in the host country compared to natives (Figure 3.1). Around 20% of the sample report that they perceived unequal treatment within the labour market, with only 3% of Peruvians reporting the same (see section 3.4 for further details on this data).[6] This ENPOVE estimate for Venezuelans, although conservative when compared to what Freier and Pérez (2021) report[7], is similar to what migrant workers report in the Japanese labour market (20%-25%, see CHRET 2017 p.28), characterized by a small share of migrants in the total working population (Morita 2017; ILO 2019). The proportion of Venezuelans self-perceiving discrimination in Peru is larger than what is found in Switzerland (17%, Auer and Ruedin 2019), New Zealand (15%, Daldy et al. 2013), Britain and EU-14 countries (16%-17%, Fernández-Reino 2020), and most of the states in the USA (Hopkins et al. 2016). It is also more extensive than what is reported by indigenous Peruvians[8], a group that historically has been subject to stark social exclusion and marginalization in terms of access to labour, education and health markets in Peru (Valdivia et al. 2007; Gushiken and Campos 2015).

The likelihood of perceiving labour market discrimination is directly related to the education level of the Venezuelan worker (right panel in Table 3.2). Those with a college education are statistically more prone to perceive unequal treatment than those with only secondary or primary education. This is partly explained by the fact that more educated workers are more aware of social inequities and of their advantage relative to less-skilled native workers; it also increases their expectations for career success (Banerjee 2008). In turn, Venezuelan males are less likely to feel discriminated against. In addition, they are typically younger than Peruvians, with 70% in the 18-35 age category while only 30% of native workers fall in the same bracket. There is no statistical association between being older and perceiving discrimination for Venezuelan workers. However, only the eldest Peruvians are less prone to perceive unequal treatment.[9]

Further, an inspection of the right panel in Table 3.2 suggests that two types of occupation (further explained in section 3.4) comprise 70% of the Venezuelan workers: Elementary occupations (i.e., unskilled and characterized by low wages) and Services and sales workers (mostly in retail establishments). These are also the two occupational groups where workers are most likely to report discrimination. For Venezuelans, the proportion of those who self-report experiencing workplace discrimination is statistically higher than

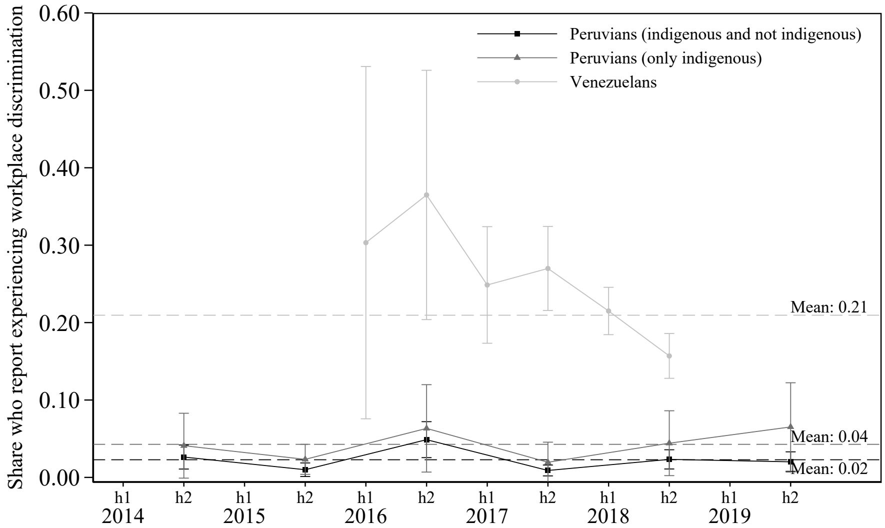
[6] The phrasing of the question in ENAHO and ENPOVE is not the same. In the former, only the head of household is asked if they have perceived discrimination in several instances (out of which within the labour market is one of them) and the reason for this (e.g., race, language, family background or origin). In ENPOVE everybody (over five years old) is asked if they have perceived discrimination (out of which within the labour market is one of them) for being Venezuelan. However, we have reason to believe that the mechanisms behind the nationality-based discrimination of Venezuelans are comparable to those behind race and the ethnic background-based discrimination of Peruvians.

[7] Their estimated proportion, between 68% and 95%, is plausibly explained by the differences in size, scope and timing of their research. Here we focus on quantitative data from ENPOVE, which leads to N=6,125. Freier and Pérez (2021) use instead qualitative data, leading to N=115. Also, we on focus on workplace discrimination (see section 3.4), whereas they consider all types of discrimination reported by Venezuelans, including that experienced in public places. Additionally, ENPOVE was recorded at the end of 2018, whereas theirs extends between 2018 and 2020. A series of events during this period (outside our time frame) contributed to a sharp increase in hostility from Peruvian natives.

[8] We determined the membership to this group based on the mother tongue spoken by the head of the household during their childhood, namely Quechua, Aymara or another native language. According to Benavides and Valdivia (2004), this criterion is more likely to reflect differences in beliefs and practices associated with specific ethnic groups. Even though the reported mother tongue may change with migration from rural to urban areas, this will unlikely affect the prevalence of cultural patterns underlying the exclusion of these specific groups.

[9] An important caveat is that ENAHO only records data (for natives) on self-perceived workplace discrimination for a subset of heads of households from the total sample. Following Figure 3.1, the number of Peruvians who were discriminated against is very small (N=27), contributing to the lack of statistical significance in terms of the estimated differences. Nevertheless, this restriction on the data is only imposed for this table. The full sample of Peruvians is used in the main regression analysis undertaken subsequently.

Figure 3.1 – Perception of discrimination for Peruvians in treated areas and Venezuelan immigrant



Note: Sample restricted to only those employed in the informal sector between 18 and 65 years. h1 and h2 refers to the first and second half of the year. Informality approximated by the lack of health insurance in their occupation. For Peruvians, data points are calculated only for head of househdols in the corresponding year while for Venezuelans are calculated for every individual given their corresponding half of year of arrival to Peru. Horizontal dashed lines refers to the average for the whole period. Vertical lines correspond to the 90% confidence intervals adjusting for clustering. *Source*: Author's calculations using ENAHO and ENPOVE (2018) data.

those who don't within the Services and sales occupations. Their allocation across these groups has changed little since arriving in 2016 (see bottom panel of Figure 3.A2 in the Appendix). Interestingly, the (observed) proportion of those in Managerial and Technical occupational fell from 11% in early 2017 to 4.5% two years later. The total distribution for Peruvians, in contrast, is more evenly spread across the categories shown, with one in four performing as technical workers. As described in Del Pozo Segura (2021), the occupational complexity of Venezuelans is lower than Peruvians' and is even lower for those who reported being discriminated against. Moreover, a lengthier stay in Peru and prior working experience in Venezuela are positively associated with their higher tendency to report experiencing workplace discrimination. The higher expectation of a more equitable treatment from employers as migrants increase their time in the host country is a plausible explanation for this (Banerjee 2008).

Certainly, the fact that a larger share of Venezuelans perceive discrimination in the host labour market does not necessarily mean the existence of mechanisms penalising their outcomes based on their nationality. However, it transpires that in their case, there is an actual relationship between perceiving unequal treatment and earning a lower (log) hourly wage (Table 3.3). The direction of these gaps and their magnitudes, between four and six percentage points favouring those immigrants who did not experience self-perceived discrimination, as well as its statistical significance, remains consistent across the selected statistics reported in the table. It is zero only at the uppermost part of the wage distribution. This same pattern is also found when analysing the (log) hourly wages from their primary occupation (see left panel of Figure 3.A4 in Appendix).

Venezuelans do engage more time in labour market activities than their Peruvian counterparts, especially those Venezuelans who self-perceived discrimination. The lower hourly wage that Venezuelans earn explains this, particularly those perceiving unequal treatment. In turn, this results in this group working more hours. However, working in the informal labour market gives them enough flexibility to undertake practically any desired number of work hours per month. As the histogram of this variable shows (right panel of Figure 3.A4 in Appendix), 73% work more than what is legally allowed per week (48 hours[10]) and, remarkably, 15%, 13% and 8% of them work 60, 72 and even 84 hours per week, respectively. Despite this, 40% of Venezuelan workers earn a monthly wage lower than the minimum legal in Peru as of 2018 (around

[10]Note that almost all of those who work more than the maximum number of hours do so while working in their only job, since only 5% of Venezuelans report having a second job.

170 GBP in 2018 prices). As mentioned above, this occurs even though on average they have higher education, experience and the absence of a language barrier, which usually provide migrants with the basis for relative success in a labour market (Chiswick and Miller 1999).

Table 3.2 – Characterization for Peruvians in treated areas and Venezuelans immigrants and perception of discrimination

				Peruvians			Venezuelans		
	Peruv.	Venez.	Diff.	Discriminat.	Not discrim.	Diff.	Discriminat.	Not discrim.	Diff.
If male	0.62	0.57	0.05***	0.59	0.59	-0.00	0.54	0.58	-0.04***
	(0.01)	(0.01)		(0.10)	(0.01)		(0.01)	(0.01)	
Age group (years)									
18-25	0.17	0.32	-0.15***	0.11	0.11	-0.00	0.32	0.32	-0.01
	(0.01)	(0.01)		(0.06)	(0.01)		(0.01)	(0.01)	
26-35	0.24	0.42	-0.19***	0.19	0.20	-0.01	0.44	0.42	0.02
	(0.01)	(0.01)		(0.07)	(0.01)		(0.02)	(0.01)	
36-45	0.24	0.18	0.06***	0.30	0.26	0.04	0.16	0.18	-0.02
	(0.01)	(0.01)		(0.08)	(0.01)		(0.01)	(0.01)	
46-55	0.22	0.06	0.15***	0.41	0.24	0.17*	0.07	0.06	0.01
	(0.01)	(0.00)		(0.10)	(0.01)		(0.01)	(0.00)	
56-65	0.14	0.01	0.13***	0.00	0.19	-0.19***	0.01	0.01	-0.00
	(0.01)	(0.00)		(0.00)	(0.01)		(0.00)	(0.00)	
Education level									
Primary	0.18	0.10	0.08***	0.41	0.25	0.15	0.08	0.11	-0.03***
	(0.01)	(0.01)		(0.10)	(0.01)		(0.01)	(0.01)	
Secondary	0.49	0.31	0.18***	0.30	0.48	-0.18*	0.29	0.32	-0.03*
	(0.01)	(0.01)		(0.10)	(0.01)		(0.02)	(0.01)	
Technical	0.17	0.19	-0.03***	0.11	0.15	-0.04	0.20	0.19	0.01
	(0.01)	(0.01)		(0.06)	(0.01)		(0.01)	(0.01)	
College	0.16	0.39	-0.23***	0.19	0.12	0.06	0.43	0.38	0.04***
	(0.01)	(0.01)		(0.08)	(0.01)		(0.02)	(0.01)	
Occupation groups									
Managerial	0.04	0.01	0.03***	0.00	0.03	-0.03***	0.01	0.01	-0.00
	(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)	
Technical	0.19	0.05	0.13***	0.22	0.25	-0.03	0.04	0.06	-0.02**
	(0.01)	(0.00)		(0.08)	(0.01)		(0.01)	(0.00)	
Clerical Workers	0.04	0.05	-0.01***	0.04	0.02	0.02	0.05	0.05	-0.01
	(0.00)	(0.00)		(0.04)	(0.00)		(0.01)	(0.00)	
Services And Sales	0.22	0.32	-0.11***	0.30	0.22	0.08	0.35	0.32	0.04**
	(0.01)	(0.01)		(0.10)	(0.01)		(0.01)	(0.01)	
Craft And Trades	0.13	0.13	0.00	0.04	0.11	-0.08**	0.11	0.13	-0.02*
	(0.01)	(0.00)		(0.04)	(0.01)		(0.01)	(0.00)	
Machine Operators	0.12	0.06	0.06***	0.11	0.09	0.02	0.07	0.05	0.01*
	(0.00)	(0.00)		(0.06)	(0.01)		(0.01)	(0.00)	
Elementary Occup.	0.27	0.37	-0.10***	0.30	0.28	0.02	0.37	0.37	-0.01
	(0.01)	(0.01)		(0.09)	(0.01)		(0.01)	(0.01)	
Occup. complexity (mean)	2.62	1.79	0.84***	1.55	1.98	-0.43	1.65	1.82	-0.17*
	(0.11)	(0.05)		(0.35)	(0.15)		(0.09)	(0.05)	
If employed in Venezuela		0.81					0.83	0.80	0.03**
		(0.01)					(0.01)	(0.01)	
Time in Peru (mean)		7.77					8.51	7.58	0.94***
		(0.09)					(0.20)	(0.10)	
N	4,989	6,125	11,114	27	1,254	1,281	1,256	4,869	6,125

Notes: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those regions who arrived to Peru from January 2016 onwards. Informality approximated by the lack of health insurance in their occupation. Discriminated/Non discriminated refers to the perception of discrimination and Diff. refers to the difference in the proportions across these two groups. Time in Peru measured in months. Managerial occupation includes professionals and technical occupations includes Skilled agricultural, forestry and fishery workers. SEs in parenthesis adjusting for clustering. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO (2018) and ENPOVE (2018) data.

Table 3.3 – Labour outcomes statistics for Peruvians in treated areas and Venezuelan immigrant workers by perception of discrimination

	Peruvians					**Venezuelans**				
	Discriminated		Not discriminated		Diff.	Discriminated		Not discriminated		Diff.
(Log) Hourly Wages										
Mean	1.11	(0.16)	1.27	(0.03)	-0.16	1.03	(0.02)	1.09	(0.01)	-0.06***
p10	0.07	(0.31)	0.22	(0.05)	-0.16	0.51	(0.02)	0.57	(0.01)	-0.06**
p25	0.72	(0.25)	0.83	(0.03)	-0.10	0.77	(0.01)	0.81	(0.01)	-0.04**
p50	1.16	(0.15)	1.31	(0.02)	-0.15	1.01	(0.02)	1.04	(0.01)	-0.04**
p75	1.66	(0.26)	1.77	(0.02)	-0.11	1.27	(0.01)	1.31	(0.01)	-0.04**
p90	2.22	(0.28)	2.21	(0.04)	0.01	1.63	(0.03)	1.63	(0.02)	0.00
(Log) Monthly Wages										
Mean	6.42	(0.23)	6.46	(0.03)	-0.04	6.58	(0.01)	6.59	(0.01)	-0.00
p10	5.00	(0.33)	5.20	(0.08)	-0.21	6.09	(0.01)	6.09	(0.01)	0.00
p25	5.86	(0.38)	5.96	(0.04)	-0.10	6.37	(0.02)	6.42	(0.02)	-0.05**
p50	6.60	(0.22)	6.59	(0.03)	0.01	6.60	(0.01)	6.60	(0.00)	0.00
p75	7.18	(0.22)	7.06	(0.03)	0.11	6.79	(0.01)	6.78	(0.00)	0.01
p90	7.68	(0.25)	7.48	(0.03)	0.20	7.07	(0.03)	7.01	(0.01)	0.05
Total Monthly Hours										
Mean	238.51	(18.81)	206.54	(2.59)	31.97*	272.81	(2.31)	258.95	(1.46)	13.85***
p10	104.29	(26.07)	78.21	(4.35)	26.07	173.81	(6.52)	156.43	(4.35)	17.38***
p25	186.85	(28.24)	139.05	(6.52)	47.80	217.26	(4.35)	208.57	(0.00)	8.69***
p50	269.40	(23.90)	208.57	(0.00)	60.83**	260.71	(3.26)	260.71	(0.00)	0.00
p75	304.17	(21.73)	260.71	(2.17)	43.45**	312.86	(2.17)	312.86	(0.00)	0.00
p90	356.31	(26.07)	330.24	(7.60)	26.07	365.00	(1.09)	365.00	(0.00)	0.00
N	27		1,254		1,281	1,256		4,869		6,125

Notes: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those regions who arrived to Peru from January 2016 onwards. Discriminated/Non discriminated refers to the perception of discrimination. (Log) hourly and monthly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. p = percentile. SEs in parenthesis adjust for clustering. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO (2018) and ENPOVE (2018) data.

3.3 Literature review

The seminal economic characterization of labour market discrimination is attributable to Becker's (1971) taste-based theories. In this framework, the subjective and non-pecuniary costs induced for hiring members of a minority group represents a penalty in the profit function of the employer.[11] In the short run, this results in a wage gap between discriminated and non-discriminated workers, which increases with the number of prejudiced employers and the intensity of their discriminatory preferences. As firms enter the market, this gap will dissipate as discriminating employers are forced to leave the industry given the additional economic cost induced by their preferences. However, this theoretical prediction does not correspond with the fact that wage gaps remain in the labour market. Further, psychological and sociological research demonstrating the importance of personal and dynamic intergroup processes explaining this process (Fiske 1998; Pager and Shepherd 2008), suggests that other factors are responsible for these gaps.

An alternative explanation for unequal treatment is motivated by statistical discrimination theory. This approach attributes the wage gap to the profit-maximizing employer's lack of information about the skills and turnover of the applicants, leading to discrimination based on observable characteristics (being the migrant condition a relevant one) given the direct and indirect costs of hiring a wrong candidate (Altonji and Blank 1999). In this literature, two different strands focus on how prior beliefs influence the hiring and paying decisions about the productivity of members of a group (Arrow 1973; Phelps 1972) and in the role of employers' information about individual productivity, where it is assumed that the variation in human capital levels is greater within the minority (migrants) group (Aigner and Cain 1977). The former characterization predicts that employers' stereotypes are self-confirming and lead to a low-level equilibrium for immigrants reflecting their lower expected returns to human capital. In contrast, the latter predicts that members of the minority group end up being less productive since employers face difficulties determining their productivity.

In both approaches, employer discrimination is viewed as a rational response due to preferences or perceived differences (in terms of averages and dispersion) in human capital. Distinguishing between the two is usually tricky. Consequently, the observed wage differentials (between natives and migrants) in the informal Peruvian labour market, where the entrance is almost unrestricted (Viollaz 2019), might signal both a taste for Peruvian employers' discrimination and statistical discrimination. The natives' unfavourable opinion of the productivity of migrants backs this latter (Table 3.1). Although these two theories have been the most popular, other plausible explanations are the non-random sorting and segregation of migrant workers, which explains Venezuelans' allocation into occupational categories with lower productivity (shown in Table 3.2, see Elliott and Lindley 2008 and Peri and Sparber 2009 for evidence the UK and the USA) and their downgrade (see Del Pozo Segura 2021). Also, it is likely that their education and working experience acquired back home are less valued than domestically acquired human capital (Friedberg 2000). Furthermore, discrimination may be unintentional and outside of the discriminator's awareness (Bertrand et al. 2005; Dovidio and Gaertner 2010). See Lang and Lehmann (2012) for other alternative theories.[12]

Empirical studies within these two frameworks have focused on the impact of a migrant on unemployment, both using observational data and field experiments. Among the former, Auer et al. (2017) find that an immigrant's longer unemployment spell in Switzerland is mainly due to discrimination by employers.

[11] Becker (1971) also discussed the consequences of employees discrimination and consumer discrimination. The former refers to the prejudice of workers in the majority group against those in the minority group (migrants) and hence avoid working with these; the latter, to prejudiced consumers in the majority group who attain less utility by purchasing from a member of this minority group. The consumer-based channel may impair the earnings ability of the self-employed members of the minority group. The former consumer-based channel may impair the earnings ability of the self-employed members of the minority group. See Altonji and Blank (1999).

[12] Note that a source of wage differentials emphasized by Freeman (1980) and explored by Kampelmann and Rycx (2016) is that institutional factors, mainly collective bargaining, can reduce wage discrimination against foreigners. However, in our case, most Venezuelans do not have access to unions by being absorbed within the informal sector. Likewise, an explanation based on employers' lexicographic search (see Bertrand and Mullainathan 2004), whereby the ethnic origin of the applicant as embedded in their last name is the primary driver of their hiring decision (Daldy et al. 2013. This latter is not tenable in our case because the cultural origin of Venezuelans is comparable to that of Peruvians.

Across Western Europe, Kogan (2006) provides evidence that in the 90s, the probability of migrants experiencing unemployment was lower in countries with a larger share of unskilled and low-skilled jobs and more flexible labour markets. Additionally, Koopmans (2015) finds that sociocultural factors were the main drivers behind the lower labour market participation and unemployment of predominantly Muslim immigrant groups in 2010. Evidence arising from field experiments using correspondence studies[13] provide support for theories other than statistical discrimination. Relative to native applicants, discrimination against those from an ethnic minority or a migrant background starts from the recruitment even after providing detailed information on their experience in the host country and relevant work skills. Di Stasio and Heath (2019) find that minorities with non-Western origin or background (Pakistan, Bangladesh, Sub-Saharan Africa and Latin America) are less likely to be shortlisted by UK employers, and Di Stasio et al. (2021) find that Muslim applicants, especially males from Africa and the Middle East, face a severe recruitment disadvantage in European countries with different approaches towards cultural and religious rights to minorities (Germany, the Netherlands, Norway, Spain and the UK). For North America, Oreopoulos (2011) finds significant discrimination towards applicants with non-English names in various occupations in Toronto, while Widner and Chicoine (2011) report a similar effect in the United States around 2008 for those whose names are associated with Arab or Middle Eastern origin. See Baert (2017) for a review of similar studies.

Another set of studies for developed economies focuses instead on the outcomes for employed, mainly using non-experimental data. After conditioning on a set of individual and firm-level characteristics, these reveal a widespread native-migrant wage gap primarily for migrants from non-English speaking countries. Brunow and Jost (2019) find that the mean daily wage gap between German workers and migrants ranges between 7.9% to 27%. In contrast, Chiswick et al. (2008) find that the observed positive hourly wage gap between English speaking USA natives and migrants across the pay distribution, especially from the 2nd decile, remains after discounting the effect of differences in endowments. Their results for Australia mimic those from the USA and also confirm Kidd's (1993) findings; namely, that the mean wage gap between natives and non-English speaking migrants ranges from 2% to 19% in the own and paid employment sectors, and this is explained in most cases by the wage-structure effect. Siebers and van Gastel (2015) find that the under-utilization of migrants' skills and the application of socio-ideological labour control widens the migrant earnings gap in the Dutch public sector, while Dickens and McKnight (2008) find that in the UK, the mean wage gap is around 30% for men and 15% for women at arrival, and it takes ten years for the wages of an average migrant to converge to the level of natives (except for Asian men). In fact, across countries in the European Community, Adsera and Chiswick (2007) find a partial natives-migrant mean wage gap of around 40%.

The prevalence of mean hourly wage gaps between natives and migrants is also reported by ILO (2020) across 49 countries around 2015 under a unified methodological approach. For Europe, the gap is as high as 30% in Southern Europe, characterized by a large proportion of immigrants (around 40%) in unskilled occupations, and less than 2% in the UK and Switzerland. For Latin America, in Argentina and Chile, which experienced a large influx of migrants from within the region during the 1990s, the mean gaps are 18% and 2%, respectively, favouring natives. Nonetheless, consistent with what is reported for Peru (Table 3.3), in both countries the migrant wage gaps in the informal sector favour migrants (7% and 17% higher than natives in Argentina and Chile). The gap in the informal sector favours migrants even more in Mexico (23%). A decomposition of the gaps across the hourly wage distribution suggests they are explained mainly by differential returns to migrants in the labour market in most countries.

Nevertheless, the presence of both unconditional or conditional wage gaps and the magnitude of the wage-structure effects should not be automatically equated with wage discrimination. This is because the latter, as Heckman (1998) argued, is defined relative to a hypothetical ceteris paribus experiment where

[13]Which involves sending resumes or cover letters of inexistent applicants fixing a set of relevant occupational characteristic but randomly varying their nationality (or background) traits.

an employer pays a different wage to two otherwise identical individuals who only differ in their migrant status. Nonetheless, because a worker's unobservable characteristics (mainly cognitive and non-cognitive ability), as well as relevant aspects of his observed attributes relevant for the employer, cannot be included in a regression[14], a crucial part of observed gaps can be actually attributed to differences between migrant and native skills and not to labour market discrimination (see, e.g. Neal and Johnson 1996 or Altonji and Blank 1999, p. 3161). In fact, an approach that tries to circumvent this problem using audit studies can detect discrimination when none exists. [15]. Hence, the discriminatory effect (of race or gender) at a randomly selected firm does not necessarily provide an accurate measure of the discrimination in the market as a whole (Heckman 1998).

A complementary approach to measure discrimination, avoiding some of these problems, is to directly study how a worker's perceptions are associated with observed disparities in outcomes among migrants and natives (Blank et al. 2004). Although subjective, perceived discrimination has a significant impact on a migrant's well-being, mental health and objective chances in life (Schmitt et al. 2014; Nandi et al. 2016; Safi 2010; Mougenot et al. 2021), their integration into the host society (de Vroome et al. 2014; Auer and Ruedin 2019) and their productivity (Ensher et al. 2001). Despite their potential correlation, we do not anticipate that this measure aligns perfectly with the actual incidence of discrimination for at least two reasons. Firstly, as explained by Kobrynowicz and Branscombe (1997) and Crosby (1984), self-protective and situational factors as well as psychological characteristics (such as self-esteem, assertiveness and depression) influence the interpretation of prejudicial events and explain deviations from perceived and actual discrimination. A migrant's socio-economic background, location, and even socio-political context can also influence this perception (Fernández-Reino 2020; Hopkins et al. 2016). Secondly, the potential misreporting can go in both directions (Biddle 2013). It is expected (as it occurs in Table 3.2) that the more qualified Venezuelans declare experiencing discrimination given their downgrading and ensuing low wages (see Del Pozo Segura 2021). In contrast, those in the groups that are usually the more poorly paid in the labour market (i.e., females, less educated, younger) may not realize that they are being discriminated against. Nevertheless, we argue that considering direct reports of discrimination by affected individuals is likely to be informative and are also likely to represent lower-bound estimates of the actual occurrence of discrimination (Blank et al. 2004).

The literature on measures of perception of workplace discrimination is scant (see Blank et al. 2004 and Wrench 2007). We follow the route of studies which analyse the migrant's perception of workplace discrimination. Daldy et al. (2013) report that in New Zealand, female immigrants from South East Asia (and New Zealander women born in Pacific Islands) are 60% more prone to report workplace discrimination than New Zealanders of European heritage. This likelihood increases with the years after arrival and the education of female workers. On the other hand, Auer and Ruedin (2019) study how variables related to different aspects of the migration process and integration impact the reporting of discrimination by immigrants in Switzerland. They find that migrants from South America and Asia have a 30% and 40% lower (ceteris paribus) odds, respectively, to report discrimination than European immigrants. Additionally, they find that attachment to the host country and positive past experiences during the migration process play a substantial role in decreasing the likelihood of reporting discrimination.

[14]Data on the quality of the education and workers' ability are usually absent, as well as socio-cultural variables that allow for more successful integration into the labour market for the migrant (Koopmans 2015). The heterogeneity in the types of skills achieved within the same level of education does not match those required in the particular job of the worker. This is particularly relevant for Venezuelans because most of the activities they perform in the informal sector lead to their occupational downgrading, which implies that their cognitive skills marginally contribute to their marginal productivity (see Del Pozo Segura 2021).

[15]On the one hand, correspondence studies do not necessarily apply in more realistic scenarios in which employers are confronted with native and migrant candidates who simultaneously differ along many dimensions and where rejection might be for reasons other than his skill (see Koopmans 2015). On the other hand, sending resumés and applications to the companies requires detailed pre-knowledge of what characteristics are essential for that particular job, which is unfeasible for most applications (see Blank et al. 2004; Heckman and Siegelman 1993). Moreover, discrimination is also in the recruitment and promotion processes as well as in day-to-day interaction with the employees, aspects which are not necessarily observed in cross-sectional data as the one here (Daldy et al. 2013).

Two other studies explicitly examine the relationship between a migrant's self-perceived discrimination and income disparities. This latter is estimated as the wage-structure component between migrants and natives from an Oaxaca-Blinder type decomposition.[16]On the one hand, Biddle (2013) finds that the estimated employment discrimination in the Australian labour market is (partially) positively associated with being overseas-born. The estimated employment and wage discrimination effects are higher for females and decline with the years in the host country.[17] The probability of reporting discrimination by the employer is correlated with the estimated wage discrimination measures. Nonetheless, the small estimated coefficient for wage discrimination suggests that other sources of discrimination are of greater importance.[18] On the other hand, Banerjee (2008) finds that the estimated wage discrimination, either unconditionally or after controlling for observed attributes, does not significantly affect self-perceived discrimination for migrant workers in Canada. This hints that these are either unaware of the degree of inequality or unable to accept that income discrimination could affect them personally.

3.4 Data

We use information from two micro-datasets. The first source is the National Survey of the Resident Venezuelan Population (ENPOVE), collected in the last quarter of 2018 by Peru's National Statistical Office (INEI). This provides information on Venezuelan demographic and labour market variables, including their past working experience in Venezuela, and is statistically representative at the national level and for each of the five cities where these located from 2016 (according to the Population Census of 2017). These cities are Lima and Callao, Tumbes, Cusco, Arequipa and La Libertad (see Figure 3.A1).[19] Importantly, ENPOVE also provides information about the perception of discrimination in different spheres of a migrant's life in Peru, including the workplace, public areas or government institutions. It is worth noting that this type of information is pretty uncommon in surveys of immigrants (Blank et al. 2004). The focus for the analysis here relates to perceived discrimination that occurred only in the workplace.[20] The second dataset is Peru's National Household Survey (ENAHO), also collected by INEI following a similar sampling design as ENPOVE. It constitutes the Peruvian government's primary source of labour market indicators, given its statistical representativeness at the national and regional levels. It includes accurate information on wages and hours of work as well as detailed information on Peruvians' demographic and employment characteristics for those in formal and informal sectors. Its information is comparable across the years (from 2005 onwards), and we use the 2018 ENAHO dataset to match the year recorded in ENPOVE.

By combining these two data sources, we approach the issue of the migrant wage disadvantage in two complementary ways. The first (in subsection 3.6.1), based only on data for Venezuelans from ENPOVE,

[16]In these two studies, estimation of the wage-structure effect implies inputting labour market returns from the sample of natives to characteristics of the migrants, and subtracting from this the migrants' observed wage, as in Banerjee (2008). Biddle (2013) takes an alternative route, by estimating employment (wage) discrimination as the difference between the own workers' predicted employment (wage) conditional on a set of observed attributes and their observed value.

[17]The former outcome is measured by a dummy that is equal to 1 if the worker thinks he was unsuccessful because the employer discriminated against them him (and 0 otherwise); for the latter, by a dummy that is 1 if the worker thinks that he has experienced discrimination by his current employer

[18]The estimations of the conditional expectation of self-reported discrimination in job applications also suggest that this outcome is positively correlated with being a female and being born overseas. At the same time, it is negatively associated with the time the worker has been in Australia's labour market.

[19]These comprise 85% of the Venezuelan population, located mainly in Lima and Callao, the capital of the country and the centre of economic activity in Peru. This represents 38% of the employment and 48% of the total GDP as of 2018. Tumbes, bordering Ecuador, is their entry point to Peru. La Libertad and Arequipa are the second and third recipients of the Venezuelan population, respectively. The latter being the second largest contributor to Peru's GDP (INEI 2019a). Cuzco is a tourist city and generates jobs in mostly services-oriented activities.

[20]Question p702 in the ENPOVE survey asks the individual if he has felt discriminated in (1) the workplace, (2) the Educational Institution, (3) the health facility, (4) Judicial Institutions, (5) street / public places, (6) In public transport, (7) Immigration offices, (8) In the offices of the Chancellery, (9) In your community/neighbourhood and (10) Another place. We focus on (1), which is similar in concept and time frame to what Daldy et al. (2013); Banerjee (2008); Auer and Ruedin (2019) exploit. In fact, the time frame inquired by the question ("last two years") captures the whole labour market experiences of the Venezuelans, as these mainly arrived in 2016.

provides evidence on how the migrants' self-perceived workplace discrimination affects their wages, conditional on the inclusion of a wide range of characteristics. The second approach (in subsection 3.6.2) appends both datasets to study how the individual's nationality (i.e., Peruvian or Venezuelan) affects wages, again controlling for an array of other covariates. This also allows us to obtain a direct measure of objective discrimination (discussed in section 3.5.), which is a key determinant in the probability that a Venezuelan worker perceives discrimination (in subsection 3.6.3). Importantly, the similar sampling design and variables harmonization (in terms of their definition, its categories, and their units of measurement) of ENPOVE and ENAHO, as well as their same level of coverage for the topics in the survey, makes those datasets comparable and justifies its use in a pooled regression framework.

Throughout the paper, we restrict the sample to workers in the informal sector[21] between 18 and 65 years old, living in the five regions where the Venezuelan migrants settled (for the ENPOVE data, we additionally restricted to include only those who arrived in Peru in 2016). We further restrict the sample to include only those who declare a positive hourly wage. Hourly wages are expressed in real terms (using the GDP deflator from Peru's Central Bank, taking 2007 as the reference year). It is constructed as the ratio of self-reported individual's wage, summing what they obtained for their primary, secondary and extra activities (including monetary and in-kind payments) to the self-reported number of hours that they usually work in all their occupations. However, it is worth noting that only 5% of Venezuelans hold two jobs in this sample. We define informal workers as those employed without access to contributory health insurance [22]. This variable reflects the vulnerability experienced by informal workers as these are not protected by legal and regulatory frameworks (ILO 2002; ILC 2002; Hussmanns 2001). This criterion is valid for Peru, as in the Latin American region "informal workers lack almost every form of social protection [...] Restricted access to health, unemployment and injuries insurance, make informal workers too exposed to the normal risks of work" (Freije 2002, p. 2). [23]

In both ENAHO and ENPOVE, the demographic control variables are sex, a dummy equal to 1 if the individual is male and 0 otherwise, age, the worker's age in years, and education level (the maximum education level attained). The control variables that capture an individual's industry and occupation follow international classifications to facilitate comparability across other studies. The vector of six industry dummies is a reduced version of the International Standard Industrial Classification of All Economic Activities (ISIC), Revision 4: 1) Agriculture, forestry and fishing (A and B and C); 2) Manufacturing and Public Utilities (D and E); 3) Construction (F); 4) Wholesale and Retail Trade, Hotels and Restaurants (G and I); 5) Transport, Storage, and Communication (H and J); 6) Finance, Insurance, and Real Estate and Community, Social and Personal Services (K - U). The vector of seven occupation dummies groups the codes in the data, defined initially based on INEI's adaptation of the International Standard Classification of Occupations (ISCO-2008) from ILO: 1) Managers, Professionals and Armed forces (MGs 1, 2 and 10); 2) Technicians and associates (MG 3) and Skilled agricultural and fishery workers (MG 6); 3) Clerks (MG 4); 4) Service and sales workers (MG 5); 5) Building workers, electricians, artisans and telecommunications (MG 7); 6) Industrial machinery operators, assemblers and drivers (MG 8); 7) Elementary occupations (MG 9).

By merging occupation data in ENAHO and ENPOVE with the US-Department of Labour ONET 25.1 (2019) dataset, we calculate scores of the intensity of distinct types of abilities ("skills") required by differ-

[21] INEI defines a worker as those engaged in any economic activity (including those dependent and independent workers who did not work at the time of the interview but had a contract) who worked more than 15 hours per week (INEI 2019b, p. 553). We follow this definition but drop unpaid family workers from our sample.

[22] An alternative variable, lack of access to a pension system of the worker, is included in the ENAHO dataset but is absent in ENPOVE. Hence, our criterion, based on health access, also make the information in both datasets comparable. In the ENAHO sample, both variables are strongly correlated (Pearson $\chi^2_1 = 1,100$, p-value=0.000).

[23] Hussmanns (2004) suggests taking jobs rather than employed persons as the observation unit, as some workers can have multiple jobs. To conform with his suggestion, we take as observation unit the worker, and their informality status is based on their primary occupation.

ent occupations[24]. An important caveat is that these are based on the Standard Occupation Classification (SOC). In contrast, occupation data from ENAHO and ENPOVE are an adaptation from ISCO-2008 nomenclature, officially named National Code of Occupations 2015[25]. Following Ottaviano et al. (2013) (see appendix therein), we take only three types of skills from this database: Cognitive Intensity (comprised by ten variables classified as "cognitive and analytical"), Communication Intensity (4 variables capturing written and oral expression and understanding) and Manual Intensity (19 variables capturing dexterity, strength, and coordination). Once merged with ENAHO and ENPOVE, we re-scale each value to equal the percentile score in that year, which measures the relative importance of a given skill among workers. This ranges between 0 and 1. A task with a score of 0.02 for some skill indicates that only 2 percent of workers in 2018 were supplying that skill less intensively). We take the average of the variables involved in each type of skill to create three indices, as well as a Complexity index summarizing the intensity of a task in cognitive-communication skills relative to manual skills: Complexity score = (Cognitive Intensity + Communication Intensity)/Manual Intensity.

3.5 Econometric methods

3.5.1 RIF regression method

Because of its intimate link to the population regression function, applied labour research focusing on (log) hourly wages has strongly relied on OLS estimation of linear conditional mean functions. Despite its weaker consistency conditions than alternative methods (see Cameron and Trivedi 2005; White 2000), the marginal impacts provided by this method might conceal important heterogeneity of the impacts across the distribution of the outcome (Angrist and Pischke 2009). The literature extensively moved towards the conditional quantile estimator (Koenker and Bassett 1978 and Koenker 2005) to obtain a more complete picture of the joint distribution of the outcome and covariates than OLS. Nonetheless, an essential drawback of this method is that its coefficients capture the covariates' effects on the quantiles of the distribution of the outcome defined by the covariates. Only exceptionally these will coincide with the effects at the unconditional quantiles of the outcome, which we are actually after (Porter 2015).[26] Methods to transition from the conditional to the unconditional quantiles, such as those by Machado and Mata (2005) and Melly (2005), require estimation of all conditional quantiles to pin down a specific unconditional quantile[27] and strongly rely on simulations and numerical integration (Angrist and Pischke 2009).

In this study, we apply the quantile estimator from Firpo et al. (2009). This computationally simpler method circumvents these inherent problems of conditional quantile regression and, importantly, provides the covariates' effects on the unconditional distribution of the outcome. At its core, it builds upon the influence function (IF) concept. Letting F_ω be the unconditional (marginal) distribution of (log) hourly wages (ω) and τ-th be the percentile of the distribution F_ω, $IF(\omega; q_\tau, F_\omega)$ represents the influence of an

[24]Admittedly, US occupation skills will differ from those in Peru and Venezuela; however, it is reasonably to assume that this difference goes in the same direction for occupations in both countries. This is reasonable because there are no reasons to believe that suddenness of the influx might have lead to changes in the composition in the host country in the short run. Moreover, because we are not interested in the actual value of these indices but, instead, in their change, these difference in levels in both countries relative to USA are not important.

[25]We use table 2 in annex of INEI (2016) which provides correspondences to relate the former in codes equivalent to the latter. Then we use table 1 in that annex to translate these into ISCO-2008 codes. We match these with the O*NET dataset, using the cross-walks in Hardy et al. (2018) which translate these into O*NET-SOC-10 to SOC-10 and from this to ISCO-2008 codes.

[26]This does not imply that conditional quantile regression is inappropriate in all applications. Porter (2015) discusses cases, other than estimating treatment effects with controls, where this method provides the effects of interest. See Wenz (2019) for a discussion on the widespread misinterpretation of the estimated coefficients from conditional quantile in part of the applied literature.

[27]From a theoretical point of view, by being an M-estimator, more stringent conditions for consistency and asymptotic normality are required for conditional quantile estimation compared to OLS (see Newey and McFadden 1994 for further discussion). Also, since the objective (also known as check) function is not differentiable, gradient optimization methods are not applicable. Instead, linear programming is required to estimate the coefficients.

individual observation on the τ-th quantile. The Recentered version of the Influence Function (RIF)[28] is expressed as

$$RIF(\omega; q_\tau, F_\omega) = q_\tau + IF(\omega; q_\tau, F_\omega) = q_\tau + \frac{\tau - 1(\omega < q_\tau)}{f_\omega(q_\tau)} \tag{3.1}$$

where $1(.)$ is an indicator function and $f_\omega(q_\tau)$ is the probability density function of ω evaluated at q_t. The expectation of Equation 3.1 conditional on $\mathbf{z}$, a (column) vector of $K+1$ exogenous explanatory variables, can be approximated by a linear (in parameters) regression model

$$E[RIF(\omega; q_\tau, F_\omega)|\mathbf{z}] = \mathbf{z}'\gamma_\tau \tag{3.2}$$

Because the unconditional expectation of the RIF function in Equation 3.1 equals the statistic q_τ (since $\int IF(\omega; q_\tau, F_\omega) \mathrm{d}F(\omega) = 0$), the Law of Iterated Expectations assures that $q_\tau = E[RIF(\omega; q_\tau, F_\omega)] = E[E[RIF(\omega; q_\tau, F_\omega)|\mathbf{z}]] = \mathbf{z}'\gamma_\tau$ (with the outer expectation with respect to the distribution $\mathbf{z}$). Hence, estimation of Equation 3.2 by OLS provides the vector of coefficients $\hat{\gamma}_\tau$ capturing the effect of the covariates on the τth percentile of the unconditional distribution of ω, given a marginal increase in the average of a variable d in $\mathbf{z}$.[29]

An advantage of OLS estimation of the linear model in Equation 3.2 is that, under plausible assumptions, it can consistently estimate the unconditional quantile partial effect even if the population parameters are random. This is akin to consistent estimation by OLS of the (unconditional) partial effects from $E(\omega|\mathbf{z})$ even if the coefficients on $\mathbf{z}$ depend on an unobserved component[30], and contrasts with the inability of the conditional quantile estimator to provide consistent partial effects in this case (see Wooldridge 2010). However, it is necessary to emphasize that the linearity of the RIF model in Equation 3.2 serves as a convenient approximation to a highly non-linear functional, such as $E[RIF(\omega; q_\tau, F_\omega)|\mathbf{z}] = g(\mathbf{z}, \gamma_\tau)$ with $g(.)$ nonlinear in γ_τ, rather than an assumption about the conditional expectation function (this rationale carries over to the decomposition methods below). This assumed linearity, commonplace in the literature (Cameron and Trivedi 2005), implies a zero expected approximation error (Firpo et al. 2018). Consequently, under the quadratic loss function induced by the OLS estimation of the RIF model, this reduced form can be interpreted as the best linear prediction of an (unknown) nonlinear function (White 1980).

Operationalizing this estimator requires estimating the quantile of the distribution of F_ω and the corresponding probability density function by non parametric Kernel density estimation methods, so that $\hat{f}_\omega(\hat{q}_\tau) = \frac{1}{N \times b} \sum_{i=1}^{N} K\left(\frac{\omega_i - \hat{q}_\tau}{b}\right)$. In contrast to the assumptions invoked for the Kernel distribution K[31], those about the bandwidth b are critical since the estimated RIF coefficients are a local approximation for

[28] This recentering is not fundamental if we are only interested in the marginal effects. However, adding back the IF to the statistic allows for identification of the intercept and hence enables the computation of Oaxaca-Blinder decompositions at various quantiles (Firpo et al. 2009, p. 954).

[29] More generally, Firpo et al. (2009) show that rearranging Equation 3.1 leads to $E\left(\frac{\partial E(RIF(\omega; q_\tau, F_\omega)|\mathbf{z})}{\partial d}\right) = \frac{1}{f_\omega(q_\tau)} \int \frac{\partial Pr(\omega > q_\tau|\mathbf{z})}{\partial d} \mathrm{d}F(x)$ where $d \in \mathbf{z}$. Hence, the unconditional quantile partial effect for $d \in \mathbf{z}$ (continuous) involves estimating the average partial effect $E\left(\frac{\partial Pr(\omega > q_\tau|\mathbf{z})}{\partial d}\right)$ and then dividing these (unconditional) probability effects by $f_\omega(q_\tau)$ in order to locally invert them back into the (unconditional) quantile effects. Different methods can be used to estimate $\frac{\partial Pr(\omega > q_\tau|\mathbf{z})}{\partial d}$, and in our subsequent empirical analysis we use the LPM. The estimated coefficients under this method coincide with that from OLS in Equation 3.2 due to the assumed linearity of the RIF model.

[30] For a random draw, let the coefficient of a (continuous) variable $d \in \mathbf{z}$ to also depend on individual-specific unobserved heterogeneity v, as in $E[RIF(\omega; q_\tau, F_\omega)|\mathbf{z}_i, v_i] = (\delta_\tau + \xi_1 v_i) d + \mathbf{z}_i'\gamma_\tau + \xi_2 v_i \equiv \delta_{i,\tau} d + \mathbf{z}_i'\gamma_\tau + \xi_2 v_i$. Because (see next subsection) we are assuming mean independence of v and d, along with $E(v) = 0$, then by the LIE $E(RIF(\omega; q_\tau, F_\omega)|\mathbf{z}) = E(E(RIF(\omega; q_\tau, F_\omega)|\mathbf{z}, v)|\mathbf{z})$. Hence OLS provides a consistent estimator of δ_τ which also equals the (unconditional) quantile partial effect $\frac{\partial E(RIF(\omega; q_\tau, F_\omega)|\mathbf{z}, v)}{\partial d}$ averaged across the population distribution of v. This result can be generalized assuming dependency on v of all coefficients in $\mathbf{z}$ and can also accommodate for d discrete.

[31] As shown in Pagan and Ullah (1999), the Epanechnikov density, characterized by its parabolic shape with bounded support, minimises the Mean Integrated Square Error (MISE). However, evidence therein indicates that differences between the MISE attained by alternative kernels and this optimal one are small.

the effect of changes in the distribution of a covariate on the quantile of interest (Fortin et al. 2011).[32] The main results in this paper take as bandwidth the one following the plug-in method by Silverman (1986) and the Epanechnikov Kernel density.[33] Following Porter (2015), we run a sensitivity analysis by altering the kernel and bandwidth when estimating the densities embedded in the unconditional quantile regression estimator (shown in subsection 3.6.1). An advantage of our application is that, unlike in Firpo et al. (2018) or DiNardo et al. (1996), our focus is on the informal labour market which is characterized by a widespread lack of compliance regarding minimum wages (see section 3.2). This ensures that the problem of a heap in the distribution of the outcome at the minimum wage, which can reduce the precision of the local approximations, and a high dependence on the kernel smoothing factor, is not encountered here.

We estimate RIF equations in Equation 3.2 for each five percentiles beginning from the 10th (i.e. = 0.10, 0.15, ..., 0.90). The model is expressed as $\mathbf{z}'\gamma_\tau \equiv \delta_\tau d + \mathbf{x}'\beta_\tau$, where $\mathbf{x}$ is a (column) vector of K variables containing the intercept and exogenous demographic and labour market characteristics (using splines for the effect of age) and d is a dummy variable. In subsection 3.6.1, d is 1 if the Venezuelan worker self-declares perceiving workplace discrimination (and is 0 otherwise), whereas in subsection 3.6.2 d is 1 if the worker is Peruvian and 0 if Venezuelan. Hence, $\hat{\delta}_\tau$ is the estimated partial effect of individual perception of discrimination or nationality-based discrimination at the percentile τ of the unconditional distribution of (log) hourly wages. To reflect that computation for the RIF involves in turn estimation of a kernel density, we estimate the variances of the estimator using a (clustered) bootstrap (Firpo et al. 2009; Rios-Avila 2020) at the level where the sampling of the data takes place (Abadie et al. 2017).

3.5.2 Decomposition method

Compared to the observed gap at the τth percentile in (log) hourly wages between those workers belonging to the two groups defined by d, $\hat{\delta}_\tau$ from Equation 3.2 is an adjusted gap, as it nets out the effect that (observed) demographic and labour market characteristics have on the outcome. Nonetheless, this is not the only adjustment that is of interest. We also want to estimate what part of differences in ω cannot be attributed to differences in observed characteristics among Venezuelans who perceived unequal treatment and those who did not, and subsequently between Peruvians and Venezuelans. Much of the literature has relied on index number decomposition methods to answer these questions. Essentially, these approaches estimate a counterfactual quantity to separate the observed gap between two groups at any given point in the distribution, Δ_O, into a part attributed to differences in observed characteristics, the composition effect (Δ_X), and the wage-structure effect (Δ_S), attributed to differences in their labour market payment structures (Fortin et al. 2011). This section focuses on estimating this latter component for the τth quantile of the unconditional distribution of ω.

The (two-fold) Oaxaca-Blinder decomposition (OB, see Oaxaca 1973; Blinder 1973) is the most common method for estimating Δ_S at the mean of the unconditional distribution. A powerful reason for this lies in its simplicity (requiring only sample means and OLS estimates from a linear regression model[34]) and the doubly-robust property (see, e.g., Wooldridge 2010) of its estimator for the counterfactual (Kline 2011).[35] Nonetheless, this robustness does not carry over for other distributional statistics. Furthermore,

[32] Given the equation for the unconditional effects from RIF regression, $E\left(\frac{\partial E(RIF(\omega;q_\tau,F_\omega)|\mathbf{z})}{\partial x}\right) = \frac{1}{f_\omega(q_\tau)}\int \frac{\partial Pr(\omega>q_\tau|\mathbf{z})}{\partial x}\mathrm{d}F(x)$, small discrepancies in $\hat{f}_\omega(\hat{q}_\tau)$ can translate in large differences in the estimated effects. For instance, the difference between $\hat{f}_\omega(\hat{q}_\tau) = 0.050$ and $\hat{f}_\omega(\hat{q}_\tau) = 0.025$ will lead to twice the estimated effect.

[33] Silverman's method is reasonable given the log-linear model adopted in our paper. Still, we take a robust variation of the original formula that involves using a fraction of the interquantile range in case this is less than the standard deviation of the estimated distribution. Several other bandwidth methods are outlined in Pagan and Ullah (1999).

[34] Non-parametric methods can also be used to estimate the conditional expectations in the OB decomposition. However, more computationally involving procedures are needed in this case to split the observed gap into its two aggregate components (Fortin et al. 2011).

[35] Kline (2011) proves this by showing that the standard OB decomposition reweights the observations in d=0 to match the covariate distribution of the group $d = 1$, with weights estimated as linear function of the covariates. Because of this, the method provides a

the OB decomposition based on a linear regression model yields biased estimates of Δ_S when the underlying conditional expectation of ω is non-linear (Barsky et al. 2002). Methods extending this decomposition to quantiles (see, e.g., Machado and Mata 2005; Melly 2005; Albrecht et al. 2003; Chernozhukov et al. 2013), and those which use non-parametric reweighting to allow for non-linearities (e.g. DiNardo et al. 1996, Barsky et al. 2002) are unable to provide a path-independent and "exact" detailed decomposition (see Fortin et al. 2011). This detailed decomposition is relevant as it further explains the mechanisms driving each of the two components of the observed gap.

In order to circumvent these limitations and to obtain a more refined estimation of the composition and wage-structure component for the τth unconditional quantile ω, Δ_X^τ and Δ_S^τ, we apply the recent Firpo et al.'s (2018) reweighted decomposition. It extends the original OB method using RIF regressions. As a starting point, the authors assume the existence of an (unknown) joint distribution function $F_{\omega,\mathbf{x},d}$ that describes relationships between ω, the observed (exogenous) characteristics $\mathbf{x}$, and d, which identifies the membership to any of two (exclusive) groups. For a given worker, ω_0 and ω_1 are the (log) hourly wages that *would be* paid if an individual was exogenously assigned to group 0 or group 1. However, because of the fundamental problem of causal inference (Holland 1986), we observe only one of these possible outcomes according to a (Rubin) potential-outcome model

$$\omega = \omega_1 d + \omega_0 (1-d) \tag{3.3}$$

where $d = 1$ if the individual reports perceiving discrimination and 0 otherwise.

From observed data on the triplet $(\omega, \mathbf{x}, d)$ we can non-parametrically identify the distribution of ω for those with $d=1$, $F_{\omega|d=1}$, and those with $d=0$, $F_{\omega|d=0}$, and hence we can define observed (log) hourly wages gap at the τth quantile as $\Delta_O^\tau = q\left(F_{\omega|d=1}\right) - q\left(F_{\omega|d=0}\right)$, with $q(.)$ a scalar function (the quantile operator) that estimates the τth quantile of $F_{\omega|d}$. By the law of total probability, $F_{\omega|d=\ell} = \int F_{\omega|\mathbf{x},d=\ell}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=\ell}$ for $\ell = 0,1$, this observed gap can be re-expressed in terms of the conditional distributions of ω with respect to $\mathbf{x}$ for group each ℓ group, as in

$$\Delta_O^\tau = q\left(F_{\omega|d=1}\right) - q\left(F_{\omega|d=0}\right) = q\left(\int F_{\omega|\mathbf{x},d=1}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1}\right) - q\left(\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=0}\right) \tag{3.4}$$

where $F_{\omega|\mathbf{x},d=\ell}(w|\mathbf{x}=\mathbf{x}_0) \equiv Pr(w \geq \omega|\mathbf{x}=\mathbf{x}_0, d=\ell)\,; \ell = 0,1$ with w a dummy argument.

The decomposition of the observed difference into the two aggregated terms requires a counterfactual *distribution*. For the case where $d=1$ denotes the Venezuelan worker who self-perceived discrimination, we are interested in the counterfactual that simulates the wage distribution that results if workers with observed and unobserved characteristics in the group $d=1$ were paid as those in groups $d=0$, i.e. $\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1}$.[36] We can impose some structure in the wage determination process for the two groups by assuming that it depends on the worker's observable ($\mathbf{x}_i$ with support X) and unobservable characteristics ($\varepsilon_i \in \mathbb{R}^m$), as in

$$\omega_{\ell i} = m_\ell(\mathbf{x}_i, \varepsilon_i)\,; \ell = 0,1 \tag{3.5}$$

where $m_\ell(.)$ is a real-valued mapping $m_\ell : \mathsf{X} \times \mathbb{R}^m \to \mathbb{R}^+$. Adding and subtracting this counterfactual in Equation 3.4 leads to

minimum mean squared error approximation to the population weights, and under actual linearity of those weights, the double-robust property under two independent sets of assumptions ensues.

[36] We are making the "simple counterfactual treatment assumption", which means that we can pay workers from a given group according to the structure of the other. This rules out other possible counterfactual wage structures (assumption 3 of Fortin et al. 2011), including those which involve general equilibrium effects or spill-over effects. Note how we use the term "structure" to denote the unexplained part of the wage gap.

$$\Delta_O^\tau = \left[q\left(\int F_{\omega|\mathbf{x},d=1}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1} \right) - q\left(\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1} \right) \right] + \left[q\left(\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1} \right) - q\left(\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=0} \right) \right] \tag{3.6}$$

where the second term in brackets is that part of the observed gap explained by differences in distributions between the two groups at the τth percentile, (Δ_X^τ). The assumption in Equation 3.5 suggests that, conditional on $\mathbf{x}$, the distribution of wages depends on both the conditional distribution of ε and the wage structure $m_\ell(.)$ (Fortin et al. 2011). Hence, for the first term in Equation 3.6 to *actually* reflect what we are after, the part of Δ_O^τ that only represents differences in wage structure functions $m_1(.)$ and $m_0(.)$ at the τth percentile, we need to fix the distribution of observables and unobservables as the one prevailing for group 1.[37]

Two identifying assumptions, commonplace in the treatment effect literature (see, e.g., Angrist and Pischke 2009 or Wooldridge 2010), achieve this. The first is common support, which implies that any realization of $\mathbf{x}$ in X is observed for both groups: $p(\mathbf{x}_0) \equiv Pr(d=1|\mathbf{x}=\mathbf{x}_0) < 1, \forall \mathbf{x}_0 \in \mathsf{X}$, where $p(\mathbf{x}_0)$ is the propensity score (Rosenbaum and Rubin 1983). In our application, this overlap assumption is satisfied by ensuring that the support of the observed covariates included in the equation for $d=1$ is the same as for $d=0$.[38] The second assumption is unconfoundness, which states that the distribution of unobserved explanatory factors is the same across the two groups once we condition on a set of observable characteristics $\mathbf{x}$: $\varepsilon \perp d|\mathbf{x}$, $\forall \mathbf{x}_0 \in \mathsf{X}$. Even though this assumption strengthens the mean independence assumption to encompass full conditional independence of the errors, it is less restrictive than the assumptions in the classic OB decomposition as it is agnostic about the dependence of the potential outcomes on $\mathbf{x}$ (Kline 2011). Firpo et al. (2018) show that this strong ignorability ensures that no difference in ω will be systematically attributed to differences in distributions of errors. Consequently, Δ_S^τ can be interpreted as a quantile-treatment effect on the treated, with Δ_X^τ being analogous to a selection bias component (Fortin et al. 2011).[39]

Arguably, the array of control variables available for the analysis render the 'ignorability' assumption plausible as these are highly correlated with unobservables. Indeed, none of the control variables can be affected by the variable d, a situation that would potentially induce an endogeneity problem in our estimators. Due to the very high proportion of the Venezuelans who are employed, 95% (see section 3.2), we do not have reasons to believe that they self-select into the labour market based on unobservables that impact both the probability to work and their wage (as in Heckman 1979). This mitigates the problem of finding variables that provide the exclusion restrictions necessary to solve this incidental truncation problem.

Conditional on the strong ignorability assumption, the counterfactual distribution $\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1}$ can be identified non-parametrically from the data, avoiding the need for an a priori (linear) functional assumption for the wage structure. Firpo et al. (2018) use a reweighting function, $\omega_c(d,\mathbf{x})$, that transforms the features of the marginal distribution of characteristics of group $d=0$ to make it similar to those in group

[37] The observed distribution of ω for group ℓ in terms of the corresponding structural form equals $F_{\omega|\mathbf{x};d=\ell}(.|\mathbf{x}=\mathbf{x}_0) = Pr(w \geq m_\ell(\mathbf{x},\varepsilon)|\mathbf{x}=\mathbf{x}_0, d=\ell)$, with w a dummy argument. Then, the simple counterfactual treatment assumption will conflate differences in the wage structure, which is what we want Δ_S^τ to uniquely reflect, and differences in the conditional distributions of ε, $Pr\left(m_\ell^{-1}(\mathbf{x},w) \geq \varepsilon|\mathbf{x}=\mathbf{x}_0; d=\ell\right)$, because $\int F_{\omega|\mathbf{x},d=1}(.|\mathbf{x}=\mathbf{x}_0, d=1)\,dF_{\mathbf{x}|d=0}(x) - \int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0, d=0)\,dF_{\mathbf{x}|d=0}(x) = \int [Pr(w \geq m_1(\mathbf{x},\varepsilon)|\mathbf{x}=\mathbf{x}_0, d=1) - Pr(w \geq m_0(\mathbf{x},\varepsilon)|\mathbf{x}=\mathbf{x}_0, d=0)]\,dF_{\mathbf{x}|d=0}(x)$. So in general $q\left(\int F_{\omega|\mathbf{x},d=1}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=0}\right) - q\left(\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=0}\right) \neq \Delta_S^\nu$.

[38] The wage of immigrants depends on the length of their stay since they arrived in Peru and their work experience in Venezuela. These two variables are not defined for natives. Hence, to fulfil this condition, we restrict the set of conditioning variables in $\mathbf{x}$ for the decomposition of Venezuelans vs Peruvians.

[39] This is because through Equation 3.3 $\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1} = \int F_{\omega_0|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1}$ and by ignorability $\int F_{\omega_0|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1} = \int F_{\omega_0|\mathbf{x}}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1}$ and so the wage structure effect equals $\Delta_S^\tau = q\left(\int F_{\omega_1|\mathbf{x}}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1}\right) - q\left(\int F_{\omega_0|\mathbf{x}}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1}\right) = q\left(F_{\omega_1|d=1}\right) - q\left(F_{\omega_0|d=1}\right)$ which is the treatment effect on the treated for the τth quantile (Fortin et al. 2011).

$d=1$. This is given by $\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1} \approx \int F_{\omega|\mathbf{x},d=0}\,\omega(d,\mathbf{x})\,\mathrm{d}F_{\mathbf{x}|d=0}$. The high-dimensional problem of estimating the conditional probability of $\mathbf{x}$ is circumvented by the Bayes rules. This makes the reweighting function depend instead on the conditional odds of the treatment, as follows

$$\omega_c(d,\mathbf{x}) = \frac{dF_{\mathbf{x}|d=1}}{dF_{\mathbf{x}|d=0}} = \frac{dF(d=0)}{dF(d=1)} \times \frac{dF(d=1|\mathbf{x})}{dF(d=0|\mathbf{x})} = \frac{1-\rho}{\rho}\frac{p(\mathbf{x})}{1-p(\mathbf{x})} \tag{3.7}$$

where ρ is the proportion of the population with $d=1$. Compared to pure reweighting approaches such as those adopted by DiNardo et al. (1996) and Barsky et al. (2002), this method also leads to efficiency gains (Fortin et al. 2011). However, as is well known (see, e.g. Lee 2016), reweighting methods have two potential shortcomings. The first is that poor compliance of the common support assumption renders$\hat{p}(\mathbf{x})$ very close to 0 or 1, making $\hat{\omega}_c(d,\mathbf{x})$ numerically unstable. In our application, we do not find this problem. A second concern is that, by critically relying on $p(\mathbf{x})$, a misspecified treatment model would not correct for the bias induced by selection into the treatment even if this latter is actually based on observables.

Therefore, we estimate models containing a rich set of interactions of covariates which result in their balance across $d=1$ and $d=0$ groups according to standardized measures defined in Austin (2009) and to the Imai and Ratkovic's (2014) overidentification-restrictions test.[40] The first set of models is based on a standard interaction of covariates, including polynomials of continuous variables in a logit model estimated by ML. The second set is based on the best-fitting (using the Akaike criterion information) treatment logit model among a set of candidate models constructed by sequentially interacting the controls up to 4th-degree polynomials (via Stata's bfit command, see Cattaneo et al. 2013). The last model is chosen by the Least Absolute Shrinkage and Selection Operator (LASSO) logit method, which performs supervised model selection, setting to zero some regression coefficients (after including polynomials, splines, and interactions of the covariates in the selection model) depending on their contribution to a penalty term the objective function (see Hastie et al. 2009 and Hastie et al. 2015 for details).[41] The tuning parameter that controls the shrinkage's magnitude is based on the Cross-Validation method (Tibshirani 1996). Based on the evidence from Busso et al. (2014), we normalize the resulting $\hat{\omega}_c(d,\mathbf{x})$ from all these models to improve their finite sample behaviour.

The assumed linearity of the conditional RIF in Equation 3.2[42] implies a linear wage structure in Equation 3.5 which provides a straightforward interpretation of the components of the decomposition. Letting the superindices indicate the sample $\ell=0,1$ (in terms of d) where the corresponding element was estimated, the classic OB expresses the τth percentile for the counterfactual distribution $q\left(\int F_{\omega|\mathbf{x},d=0}(.|\mathbf{x}=\mathbf{x}_0)\,\mathrm{d}F_{\mathbf{x}|d=1}\right)$ (in Equation 3.6) as $\overline{\mathbf{x}}^{d=1'}\hat{\beta}_\tau^{d=0}$. Then, it leads to the following decomposition of the hourly wage gap for the τth unconditional percentile

$$\hat{\Delta}_O^\tau = \hat{\Delta}_S^\tau + \hat{\Delta}_X^\tau = \overline{\mathbf{x}}^{d=1'}\left(\hat{\beta}_\tau^{d=1} - \hat{\beta}_\tau^{d=0}\right) + \left(\overline{\mathbf{x}}^{d=1} - \overline{\mathbf{x}}^{d=0}\right)'\hat{\beta}_\tau^{d=0} \tag{3.8}$$

where $\overline{\mathbf{x}}^{d=\ell}$ is a (column) vector of K means and $\hat{\beta}_\tau^{d=\ell}$ is a (column) vector of K least squares estimates taking as dependent variable a consistent estimator of Equation 3.1, $\widehat{RIF}()$. The estimated reweighting

[40] This test exploits the dual property of $p(\mathbf{x})$ as a conditional probability of treatment assignment and as a covariate balancing score. Hence, after stacking the propensity score's moment conditions and the corresponding covariate balancing moment conditions (which provide the over-identification), the value of the objective function evaluated at the efficient GMM estimates provides the basis for a Hansen J's statistic.

[41] Our primary interest in this "first stage" of the decomposition is estimating a model that allows for covariates balance, not selecting the covariates for the linear RIF model for ω. This latter, instead, is informed by economic theory. Hence, there is no risk of "regularisation bias", arising from specifying an outcome model excluding variables with non-zero coefficients from the LASSO, or incorrect inference of the estimated coefficients at this first stage. Methods that offer "principled" variable selection (e.g. double-lasso, Belloni et al. 2014) are not necessary for our purposes.

[42] Avoiding this linearity assumption would require non-parametric identification of $m_\ell(\mathbf{x}_i,\varepsilon_i), \ell=0,1$, which in turn demand stronger assumptions than ignorability, such as complete unconditional independence of $\mathbf{z}$ and ε and strict monotonicity. This is implausible in most applications (Fortin et al. 2011).

factors $\hat{\omega}_c(d,\mathbf{x})$ allow us to express that counterfactual, instead, as $\overline{\mathbf{x}}^{d=c'}\hat{\beta}_\tau^{d=c}$. These are calculated from the sub-sample for $d=0$ reweighted to have the same distribution of characteristics as the sub-sample for $d=1$, with $\hat{\beta}_\tau^{d=c}$ the vector of (weighted) least squares estimates expressed as

$$\hat{\beta}_\tau^{d=c} = \left(\sum_{i\in d=\ell} \hat{\omega}_{c,i}(d,\mathbf{x})\,\mathbf{x}_i\mathbf{x}_i'\right)^{-1} \sum_{i\in d=\ell} \hat{\omega}_{c,i}(d,\mathbf{x})\,\mathbf{x}_i\widehat{RIF}(.)_i \tag{3.9}$$

This, in principle, leads to $\hat{\Delta}_O^\tau = \hat{\Delta}_S^\tau + \hat{\Delta}_X^\tau = \left[\overline{\mathbf{x}}^{d=1'}\hat{\beta}_\tau^{d=1} - \overline{\mathbf{x}}^{d=c'}\hat{\beta}_\tau^{d=c}\right] + \left[\overline{\mathbf{x}}^{d=c'}\hat{\beta}_\tau^{d=c} - \overline{\mathbf{x}}^{d=0'}\hat{\beta}_\tau^{d=0}\right]$.[43]

Nevertheless, because the linear RIF model is only an approximation to a highly non-linear statistic and the reweighting process might not exactly identify the counterfactual distribution in Equation 3.6, we can obtain a neater estimate of the wage-structure effect by making explicit the role of these errors. This results in the following expression which we estimate

$$\hat{\Delta}_O^\tau = \underbrace{\left[\overline{\mathbf{x}}^{d=1'}\left(\hat{\beta}_\tau^{d=1} - \hat{\beta}_\tau^{d=c}\right) + \left(\overline{\mathbf{x}}^{d=1'} - \overline{\mathbf{x}}^{d=c'}\right)\hat{\beta}_\tau^{d=c}\right]}_{\hat{\Delta}_S^\tau} + \underbrace{\left[\left(\overline{\mathbf{x}}^{d=c} - \overline{\mathbf{x}}^{d=0}\right)'\hat{\beta}_\tau^{d=0} + \overline{\mathbf{x}}^{d=c'}\left(\hat{\beta}_\tau^{d=c} - \hat{\beta}_\tau^{d=0}\right)\right]}_{\hat{\Delta}_X^\tau} \tag{3.10}$$

On the one hand, the first term within $\hat{\Delta}_S^\tau$ is the focus of the decomposition. This corresponds to the pure wage-structure effect ($\hat{\Delta}_{S,p}^\tau$) and reflects that part of the wage-structure effect $\hat{\Delta}_S^\tau$ left after netting out, via the re-weighting, the imbalances between the two observed groups. The second term in $\hat{\Delta}_S^\tau$ is the re-weighting error ($\hat{\Delta}_{S,e}^\tau$) which encapsulates these imbalances in the distribution of characteristics among the groups with $d=1$ and $d=0$. Note how $\hat{\Delta}_S^\tau$ from the classic OB (first term in the right hand side Equation 3.8) embeds this reweighting error in the simple difference between $\hat{\beta}_\tau^{d=1}$ and $\hat{\beta}_\tau^{d=0}$.[44] The negligible size of $\Delta_{S,e}^\tau$ and its lack of statistical significance in our application (see section 3.6) confirm that the propensity score models selected are adequate for identifying the counterfactual. On the other hand, the leading term in $\hat{\Delta}_X^\tau$ is the pure composition effect ($\hat{\Delta}_{X,p}^\tau$). Its associated misspecification error ($\hat{\Delta}_{X,e}^\tau$) indirectly tests whether our linear model for the outcome is valid (Fortin et al. 2011; Rios-Avila 2020; Firpo et al. 2018).[45]

Additionally, the detailed components of the pure effects, $\hat{\Delta}_{S,p}^\tau$ and $\hat{\Delta}_{X,p}^\tau$, can be computed in the same way as for the OB decomposition. Letting the intercept be the first term, indexed by 1, in the subvector $\mathbf{x}\in\mathbb{R}^K$ of $\mathbf{z}$, the individual (detailed) contributions are

$$\hat{\Delta}_{S,p}^\tau = \left(\hat{\beta}_{\tau,1}^{d=1} - \hat{\beta}_{\tau,1}^{d=c}\right) + \sum_{k=2}^{K} \overline{x}_k^{d=1}\left(\hat{\beta}_{\tau,k}^{d=1} - \hat{\beta}_{\tau,k}^{d=c}\right) \text{ and } \hat{\Delta}_{X,p}^\tau = \sum_{k=1}^{K}\left(\overline{x}_k^{d=c} - \overline{x}_k^{d=0}\right)\hat{\beta}_{\tau,k}^{d=0} \tag{3.11}$$

As in the case of the mean, the components in $\hat{\Delta}_S^\tau$ will be subject to the problem of the omitted group[46]. The aggregate decomposition in Equation 3.10 and the detailed is estimated by the routine in Rios-Avila (2020).

[43] That is, $\hat{q}\left(F_{\omega|d=\ell}\right) = E\left[RIF\left(\omega_{d=\ell};q_\tau,F_{\omega|d=\ell}\right)\right] = \overline{\mathbf{x}}^{d=\ell'}\hat{\beta}_\tau^{d=\ell}$ for $\ell=0,1$ and $\hat{q}\left(F_{\omega_{d=0}|d=1}\right) = E\left[RIF\left(\omega_{d=1};q_\tau,F_{\omega|d=0}\right)\right] = \overline{\mathbf{x}}^{d=c'}\hat{\beta}_\tau^{d=c}$. As shown in Firpo et al. (2018), these estimators are consistent and asymptotically normal under additional assumptions following the strong ignorability.

[44] As discussed in Firpo et al. (2018), in the classic OB mean-decomposition OLS estimates may depend on the distribution of covariates when the conditional expectation of ω is non-linear as OLS minimizes a specification error that itself depends on the distribution of $\mathbf{X}$ (White 1980). In addition, changing the distribution of $\mathbf{x}$ changes the F_ω, leading to a change in $RIF(\omega;q_\tau,F_\omega)$ and, via Equation 3.2, it in turns affects the estimated RIF coefficient. This problem is addressed by using instead $\hat{\beta}_\tau^c$ from the reweighted sample.

[45] Empirically, if the (true) conditional expectation was actually linear, then $\hat{\Delta}_{X,e}^\tau$ will be 0 $\operatorname{plim}\left(\hat{\beta}_\tau^C\right) = \operatorname{plim}\left(\hat{\beta}_\tau^{d=0}\right) = \beta_\tau$. In the classic OB decomposition, this happens only under the actual linearity of the RIF statistic and the zero conditional mean on linear wage structures (Firpo et al. 2018).

[46] Because the contribution of each covariate to Δ_S^q is sensitive to the choice of the base group in the regression model, the elements of the detailed decomposition can be viewed as arbitrary, and there is no general solution to this omitted group problem. In cases where the omitted group has a particular economic meaning, the elements of the detailed decomposition are more interpretable (see Fortin et al. 2011; Firpo et al. 2018).

3.5.3 Probability model method

We now describe the empirical strategy to answer the second research question, i.e. how an objective measure of wage inequity affects the Venezuelan workers' perception of discrimination. Letting Δ_S be the wage-structure effect, a measure of objective discrimination in (log) hourly wages, and $\mathbf{r}$ a (column) vector of R control variables (which includes a constant), we use an index model for a binary response. This is expressed as:

$$Pr(d=1|\Delta_S,\mathbf{r}) = G\left(\rho\Delta_S + \mathbf{r}'\boldsymbol{\theta}\right) \tag{3.12}$$

and can be motivated from a latent model as $d^* = \rho\Delta_S + \mathbf{r}'\boldsymbol{\theta} + \varepsilon$ where $d=1$ if $d^*>0$ and 0 otherwise. Assuming that ε is independent of the $\mathbf{r}$ and Δ_S and follows a standard normal distribution, it can be shown that G is the cumulative normal distribution for a probit model, estimated by maximum likelihood method (Wooldridge 2010). Based on section 3.3, and in special Banerjee (2008), $\mathbf{r}$ includes the duration of migrant's stay in Peru, education, occupation complexity score and gender.[47]

Every estimated equation includes the vector $\mathbf{r}$ and the wage-structure Δ_s estimated at the mean and the 17 percentiles (mentioned above) of the unconditional distribution of ω. The wage-structure at the mean Δ_S is the difference between the predicted Venezuelan wages if they were paid as natives in the Peruvian labour market and their observed (log) hourly wage (Banerjee 2008; Biddle 2013). The prediction for Venezuelans in this case equals the inner product of the coefficients from a classic Mincer OLS regression on the Peruvian sample (see Table 3.A5 in Appendix) and the observed attributes of Venezuelan workers. The wage-structure at a given percentile τ, Δ_S^τ, is the difference between predicted wages for Venezuelans and the RIF for the τth percentile (Equation 3.1) of their observed (log) hourly wage. In this case, the prediction equals for Venezuelans the inner product of the coefficients from a Mincer RIF-regression for the τth percentile (using Equation 3.2) on the Peruvian sample and the observed attributes of Venezuelan workers. The right-hand side variables in these Mincer regressions include gender, age group (in splines), education groups, industry group, occupation and dummies for the regions where the Venezuelans settled. Occupation is modelled using occupational categories dummies and, alternatively, using the occupational complexity score. Note how these estimated Δ_S are not based on the re-weighting procedure described above. However, the magnitude of these treatment estimates following the classic OB decomposition does not appear to be statistically different from those using the more sophisticated re-weighted procedures (see subsection 3.6.2). This indicates that it is unlikely that replacing the former with the latter in our estimation will significantly change our results.

A potential problem with the probit model is that the departures from some of its assumptions (mainly but not exclusively the normal distribution of ε) lead to inconsistent estimators, converging instead to a pseudo-true value whose interpretation is not straightforward (Cameron and Trivedi 2005). This consideration, in fact, appears to be absent from previous studies which also address the determinants of perception of discrimination (e.g. Biddle 2013; Banerjee 2008; Auer and Ruedin 2019; Daldy et al. 2013). Consequently, we test underlying population assumptions (correct functional form, homoscedasticity and normality) behind this estimators' consistency and asymptotic normality using the framework in Chesher and Irish (1987). It is based on the outer-product-of-the-gradient versions of the LM test:

$$LM = \mathbf{i}'\mathbf{R}\left(\mathbf{R}'\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{i} \underset{H_0}{\sim} \chi^2_P \tag{3.13}$$

[47] An alternative model by Auer and Ruedin (2019) includes variables related to drivers, attachment and acceptance of the immigrant to the host country. We do not follow this conceptualization since significant predictors in their study, for Switzerland, do not apply in the current case. For instance, as discussed in section 3.2, Venezuelans do not face a language barrier, while the Peruvian legislation does not allow foreigners, including Venezuelans, as congressmen. Also, their measure of "relationship with the immigration authority" is not defined in our case as the sudden emergence of the Exodus made it unfeasible for the Peruvian government to provide an articulated system of legal and bureaucratic counselling to these migrants.

where $\mathbf{i}$ is a $n \times 1$ vector of ones (n being the number of Venezuelan workers) and $\mathbf{R}$ is an $n \times (1+R+P)$ matrix whose columns are comprised by score contributions of each of the covariates in the model (computed by multiplying Δ_S and $\mathbf{r}$, without the intercept, by the generalised residual), the generalised (or pseudo) residual itself and a set of P additional terms (also interacted with the generalized residual) which depend on the specific test. A test for the correct functional form in Equation 3.12, akin got the RESET test (Ramsey 1969), includes in P the quadratic, cubic and quartic of the standardised probit index (and so $P = 3$ for this test). For a general test of heteroskedasticity, the P terms are the original explanatory variables in the model ($P = R$). For testing normality, we focus on skewness and excess of kurtosis in the generalised residuals ($P = 2$). Each test is chi-square distributed with P degrees of freedom under the corresponding null hypothesis (no incorrect functional form, no heteroskedasticity and no normality, respectively).

3.5.4 Summary

The preceding methods described in this section are used in the following three separate exercises. First, we complement the mean analysis by using unconditional quantile regressions to estimate the impact of the perception of workplace discrimination on the hourly wages of Venezuelans in Peru based on an intercept shift. This is then developed using an extension of the Oaxaca-Blinder decomposition outlined by Firpo et al. (2018), incorporating re-weighting using separate sub-samples for those Venezuelans who perceive discrimination and those who do not. This enables the estimation of the treatment effects across the unconditional log wage distribution. Second, we replicate this analysis using separate samples for Peruvians and Venezuelans in the informal sector to estimate the magnitude of the pay gap between 'natives' and 'migrants' at the mean and across the unconditional pay distribution. Finally, we integrate both these strands by examining, using a probit model, the factors that determine whether or not Venezuelans perceive they are subject to discrimination. This model includes, inter alia, a simple individual-level treatment effect of the wage differential between what a Venezuelan earns in the Peruvian informal labour market and what would be earned if the individual was rewarded according to a Peruvian pay schedule.

3.6 Empirical results

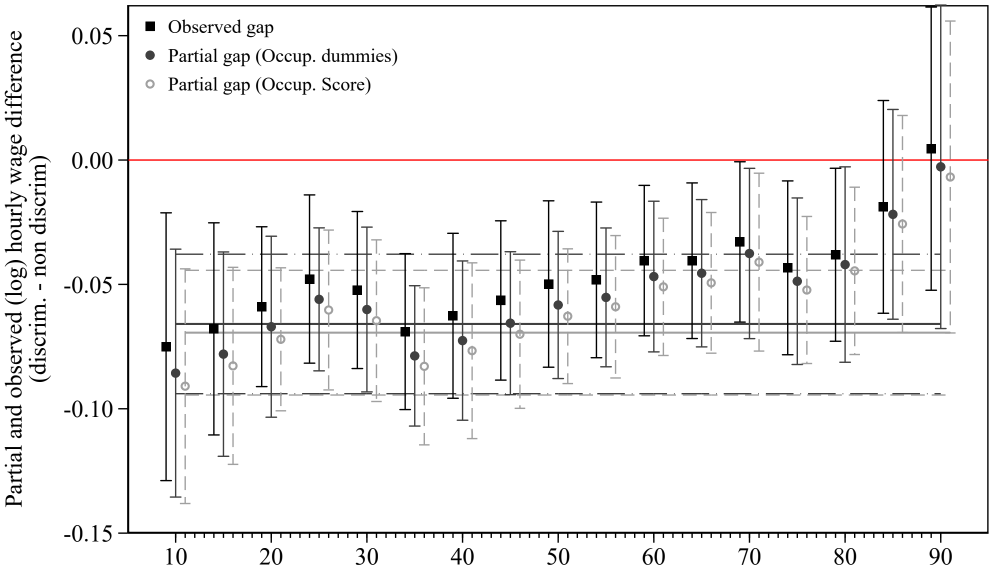
3.6.1 Perception of unequal treatment and the wage gap

The following subsection reports the results restricting the sample to only Venezuelans and taking as the treatment and individual's self-perception of workplace discrimination. This variable d is equal to 1 for those Venezuelans who self-perceived workplace discrimination and 0 for those who did not. Firstly, we show the observed and partial gaps (or ceteris paribus) gaps across the Venezuelan unconditional distribution of the (log) hourly wages, defined as F_ω. Attention then turns to the results for the (re-weighted) OB decomposition in terms of their aggregate components and detailed effects.

3.6.1.1 Estimation

The extension of the analysis in Table 3.3 confirms that the perception of discrimination is associated with an actual (statistically significant) negative effect on (log) hourly wages for most of the percentiles of the distribution F_ω (Figure 3.2). Specifically, the observed gaps are around 5% across different percentiles from the first to the eighth decile. Adjusting these raw gaps using RIF regression (Equation 3.2) to control for demographic variables (gender, age, education) and industry and occupation (either as occupation dummies or as a complexity score) suggest a slightly higher penalty at every percentile of F_ω. The point estimates for these adjusted penalties are between 6% and 9% across the first half of the unconditional distribution of ω. These reduce to 5% at the 80th percentile, losing statistical significance onwards. In general, however, these effects are fairly homogeneous across F_{ω} (except beyond the 8th decile) and are primarily in comfort with the mean penalty of approximately 7% (as displayed by the horizontal lines in these figures).

Figure 3.2 – Observed and adjusted perceived unequal treatment using (log) hourly wage gaps for Venezuelan immigrants



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years old who arrived after January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows the observed and part effect of perceived discrimination on the unconditional distribution of the dependent variable from different models (shown in table 4) at 17 percentiles (beginning at the 10th in steps of 5, circles) and at the mean (solid horizontal line). Vertical lines correspond to the bootstrapped 95% confidence intervals adjusting for clustering.
Source: Author's calculations using ENPOVE (2018) data.

Further details of the ceteris paribus effects of the observed characteristic at different percentiles of F_{ω} are provided in Table 3.4. The reported goodness of fit measures are around 9% for the effects at the mean (OLS) and median (50th percentile) and fall to about 5% at the bottom and top deciles. The magnitude of the estimated effects for demographic variables does not depend on the variable we use for occupation. In contrast to what is found for the sample of Peruvians (see Table 3.A5), the partial gender gap across the distribution of ω is no longer statistically significant, except at the very top (i.e., 9th decile) where being a female is associated with a partial penalty between 5% and 6%. An additional year for those Venezuelans aged 36-45 is associated with increasing their hourly wage by 1%. In turn, an additional year for Venezuelans in the eldest bracket (56 to 65) enhances their wages for those in the upper half of F_{ω}. The estimated effects for education yield the anticipated effects, although the penalty associated with having secondary education relative to college education widens from -3% at the 10th percentile of F_{ω} to -7% at the 75th. In comparison, the penalty for primary education exhibits the opposite pattern, reducing from -14% at the 10th percentile to -7% at the 75th percentile.

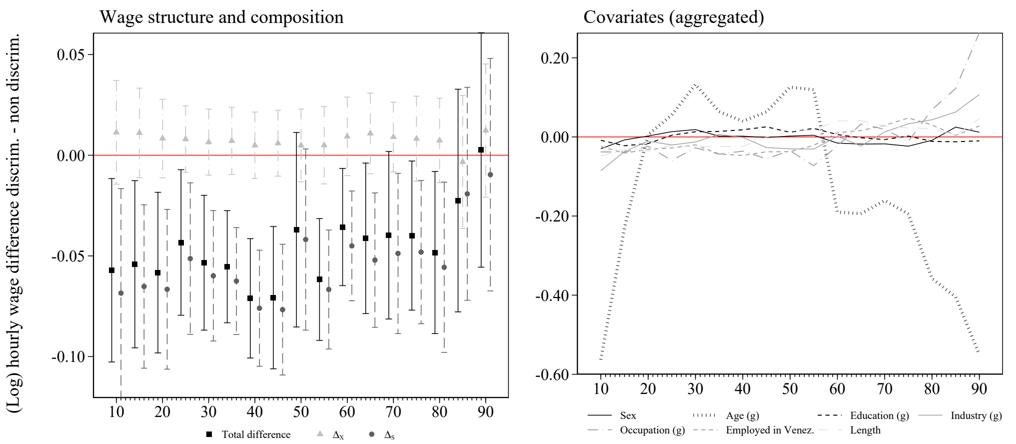
The effect of labour market characteristics based on the same table suggests that being employed in the Wholesale and Retail industry, the leading industrial sector for Venezuelans (see section 3.2), has a negative partial effect on F_{ω} compared to Venezuelans in the manufacturing industry. It reaches -5% at the mean and is between -11% and -8% for the lower percentiles. Even larger penalties are found for Venezuelans employed in the Transport, Storage and Communication sector, with the differential effect as large as almost -20% at the bottom of F_{ω}. Plausibly, this is because Venezuelans in this industry mainly provide transport services for passengers and goods (e.g., as couriers) using digital platforms. In Peru, this type of work is characterized by poor working conditions, a lack of regulation and even the lack of verification regarding basic legal requirements such as driving license, and police permits, resulting in wages below the specified legal minimum (see OIT 2021). The largest occupational penalties relative to those in technical jobs are for those employed as Service and Sales workers and in Elementary occupations. These are, precisely, the two occupational categories with the largest share of Venezuelan workers (see table Table 3.2). The relative pay penalties are 13% and 10% at the median but widen to 43% and 37% at the upper parts of F_ω, respectively. Their work experience acquired back in Venezuela does impact their wages. The reduced room for wage bargaining by being employed in the informal economy partly explains this. Nonetheless, time spent in Peru has a significant cumulative effect. Eight months after their arrival, the average worker secures a (partial) premium of 8% compared to those who arrived more recently.

3.6.1.2 Decomposition

In order to complement these findings on self-perceived discrimination adjusted gaps, we now estimate Δ_S, which represents that part of the observed gap corresponding to the differential labour-market payment received by those who perceive discrimination relative to those who did not. Due to our maintained assumption of selection on observables, Δ_S can be interpreted here as a quantile-estimate of the treatment effect on the non-treated. In other words, it represents how much lower the wages are for those who experienced discrimination relative to a counterfactual situation had these workers never experienced such treatment.

A first approach for estimating this distributional treatment effect is given by the classic (unweighted) OB RIF-decomposition following Equation 3.8. This reveals that most of the observed differences in ω between those two groups can be almost exclusively attributed to the wage structure effect (left panel of Figure 3.3). I.e., differences in unobserved payments in the labour market for those who perceive unequal treatment and those who did not mainly explain the gap. The components of the detailed decomposition (right panel of Figure 3.3) reveal that, at first glance, age (ticked lines) has an inverted U pattern across the wages distributions. However, an examination of its explained and unexplained components (in Table 3.A3) reveal they are estimated rather noisily, with none of the subcomponents found to be statistically significant across the percentiles of F_ω.

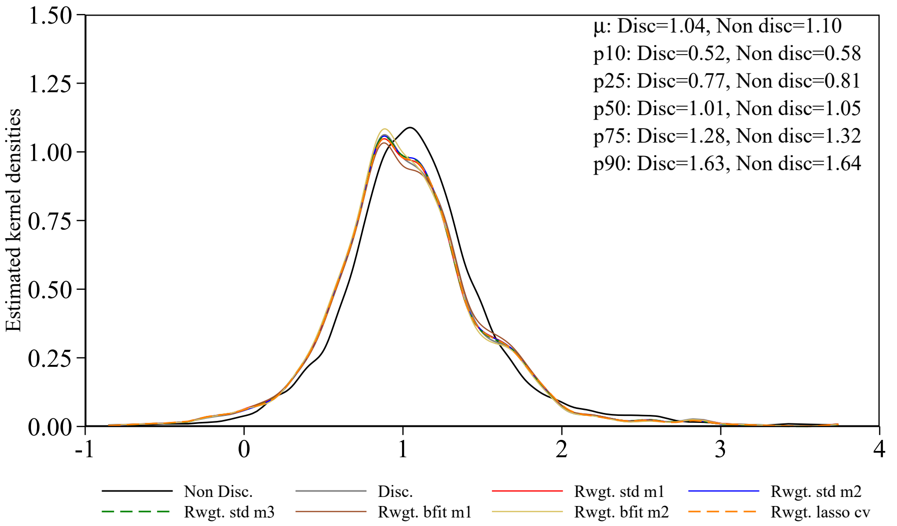
Figure 3.3 – Aggregate Unweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by self-perceived discrimination



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows the decomposition components, without reweighting, at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the treatment model and the decomposition. (g) means that the component of the detailed decomposition refers to the effect after grouping the corresponding dummies (see table 2 for the categories of the control variables). Vertical lines correspond to the bootstrapped SEs adjusting for clustering.
Source: Author's calculations using ENPOVE (2018) data.

In order to assess the validity of the local RIF approximation, which is mostly a practical issue (Fortin et al. 2011), we compare these results with the re-weighted decomposition. One of the main elements behind the Firpo et al. (2018) approach are the estimated weights which allow us to transform the features of the marginal distribution of the outcome variable (i.e., the log hourly wages) into features of the counterfactual distribution (as in Equation 3.7). The estimated counterfactual densities (Figure 3.4) convey how the approach takes the probability density for the outcome variable for those who perceived discrimination and renders it more similar to the probability density for those that did not. The six different treatment models that we estimate here are broadly similar to each other. Furthermore, there is no evidence of features that can potentially induce problems in the interpretation of empirical results, such as 'cliffs' or 'peaks' associated with, say, minimum wages at the bottom end of the distribution, or evidence of censoring of the distribution at higher percentiles.

Figure 3.4 – (Log) hourly wage distribution for Venezuelan immigrant workers by perception of discrimination and for reweighted sample



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years old who arrived after January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows kernel estimates of the unconditional distribution of the dependent variable by perception of discrimination and of the reweighted samples from different treatment models (see text for the covariates included in the treatment models). The procedure reweights those who did perceived discrimination as if they did not.
Source: Author's calculations using ENPOVE (2018) data.

The reweighted decomposition results (see the upper panel of Figure 3.5), following Equation 3.10, point in the same direction. The differences in log hourly wages between those who perceived workplace discrimination and those who did not are practically equal to the overall wage structure effect. In fact, this unexplained effect, which is statistically different from zero, equals the pure unexplained effect $\Delta_{S,p}$ (bottom part). This confirms that our previous estimates of Δ_S reflect the unequal treatment. The non-statistically significance of the reweighting errors $\Delta_{S,e}$ at every percentile of the unconditional distribution of ω provides further validation. In addition, the non-significance of the specification error, $\Delta_{X,e}$, suggests that the linear RIF-regression model provides a good fit even in the presence of high non-linearities of the population statistic or a random coefficients model (middle panel). Following Equation 3.11 and reported in Table 3.5, the detailed decomposition suggests that across the percentiles of F_ω the time (in months) since their arrival in Peru and education levels are the two leading (statistically significant) drivers of the composition effect. Their estimated signs mean that their superior knowledge of the labour market and their higher skills of those who reported experiencing discrimination provide them with a wage advantage compared to those who did not perceive workplace discrimination. However, this is counteracted by the wage-structure effect that acts against them. These results are invariant to whether or not an individual's occupation is parametrized using occupational dummies or occupational complexity score (Figure 3.A5 and Table 3.A4 in Appendix).

3.6.1.3 Robustness checks

Results from the reweighted OB decomposition make two implicit assumptions, as discussed in section 3.5. The first one relates to the choice of the treatment model that provides the propensity scores and hence the weights $\omega_c(d, \mathbf{x})$ used in Equation 3.7. The modelling above uses a rich set of interactions and polynomials of demographic and labour market variables, chosen a priori and successfully balanced the covariates across the two groups $d = 1$ and $d = 0$. The second assumption corresponds to the bandwidth and the assumed kernel functions used to estimate $f_{\omega}(.)$ in Equation 3.1. We have taken a standard approach in terms of the bandwidth by following Silverman's bandwidth rule. However, the validity of the local approximation that the RIF relies on is ultimately an empirical question. Consequently, to confirm that these assumptions do not drive our results, we re-estimate the decompositions above under alternative assumptions. Throughout, we take occupation dummies instead of occupation complexity scores for the log hourly wage model.

A key finding is that changing the treatment model does not change our results (see Figure 3.A6). Put differently, taking a sequential or more computer-driven approach for the treatment assignment model specification (see section 3.5) has no material effect on the key conclusions reported above. This is not surprising since, as depicted in Figure 3.4, the reweighting of the distribution under different approaches provides a similar counterfactual density of ω. For the robustness of the bandwidth and kernel density, we take three different alternatives instead. The first is Silverman's rule taking the canonical bandwidth[48] from the Epanechnikov kernel function. The second is the Sheather-Jones plug-in estimator of the bandwidth again with the Epanechnikov function. The last one is the bandwidth from the normal approximation with the Gaussian (normal) kernel. The results from these variations are indistinguishable from our main findings (see Figure 3.A7).

An alternative type of robustness check implies taking a different definition of the outcomes and employment. On the one hand, we now take the (log.) hourly wages arising only from the main occupation. This results in slightly larger (negative) gaps across F_{ω}. However, these are not statistically different from the results explained above (Figure 3.A8). On the other hand, we restrict the definition of worker to only those employees in blue-collar and white-collar occupations, excluding independent and employers. This reduces the number of observations from 6,1.25 to 4,659. Although the basic treatment assignment model has a more critical effect on our estimates (Figure 3.A9), it conveys the same essential message: the wage-structure effect is the most important factor behind the observed gaps. It also yields a loss of statistical significance of the observed gaps in the first two deciles of the wage distribution. A statistical explanation for this corresponds to the reduction of the sample size implied by this robustness check. However, a more economic-focused interpretation is that the wages of blue-collar and white-collar Venezuelan workers at the very bottom of the distribution are low enough to be reduced by experiencing discrimination (assuming that there is a reasonably close correspondence between the perception and the actual occurrence of discrimination).

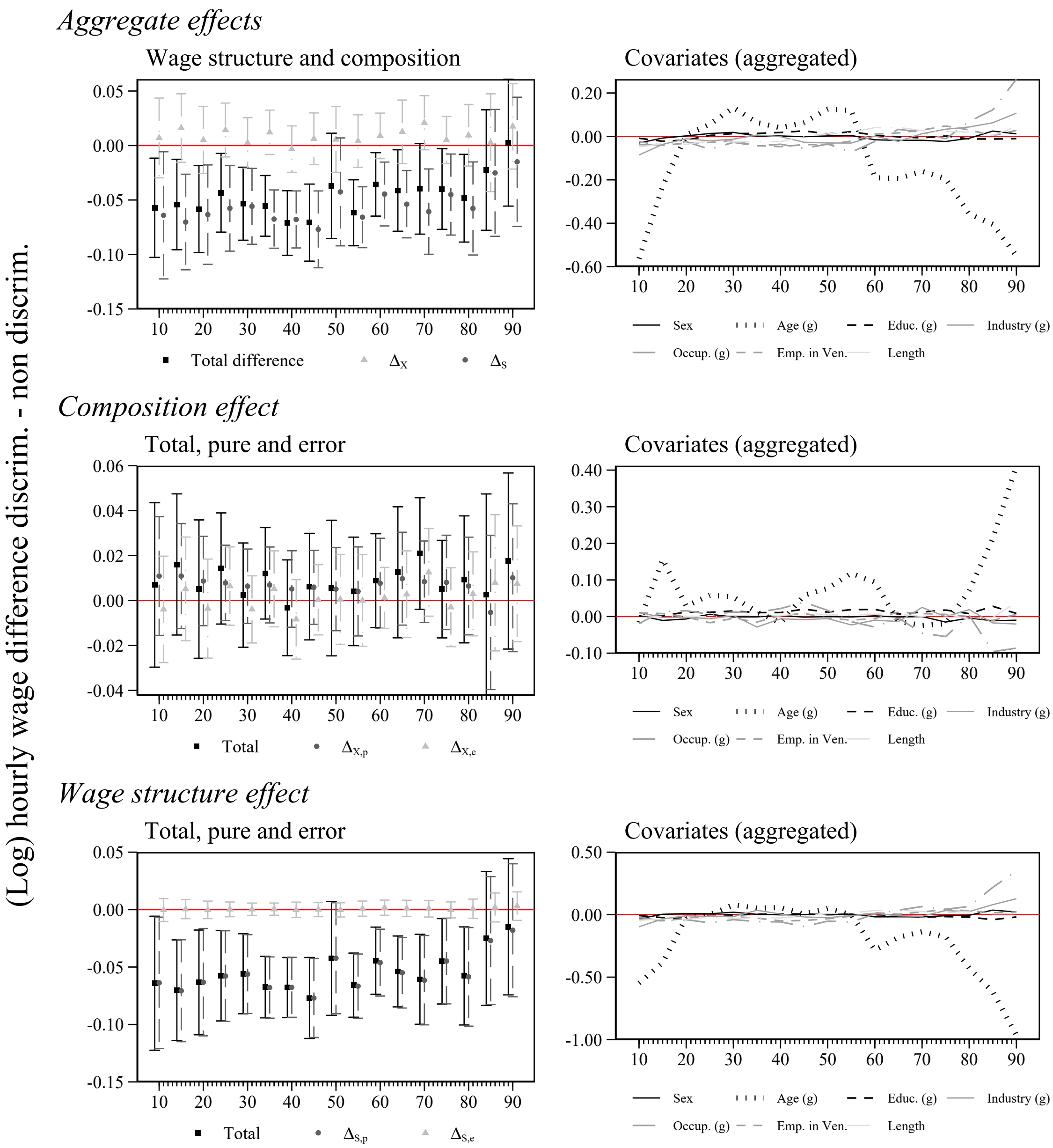
[48] This is given by $\left(R(K)/\left(\sigma_K^2\right)^2\right)^{\frac{1}{5}}$ where $R(K)$ is a measure of the curvature of the chosen Kernel and σ_K^2 is its variance (Pagan and Ullah 1999).

Table 3.4 – Adjusted perceived unequal treatment using Mincerian equations for (log) hourly wages for Venezuelan immigrants

	Occupation dummies						**Occupational complexity score**					
	Mean	p10	p25	p50	p75	p90	Mean	p10	p25	p50	p75	p90
If perceived discrimination (d)	-0.066***	-0.086***	-0.056***	-0.058***	-0.049***	-0.003	-0.069***	-0.091***	-0.060***	-0.063***	-0.052***	-0.007
	(0.014)	(0.027)	(0.016)	(0.014)	(0.017)	(0.033)	(0.014)	(0.027)	(0.018)	(0.016)	(0.015)	(0.031)
If is male (d)	-0.021	0.038*	0.005	0.005	-0.008	-0.064*	-0.015	0.025	0.006	0.011	0.004	-0.053
	(0.015)	(0.022)	(0.014)	(0.012)	(0.016)	(0.033)	(0.015)	(0.021)	(0.014)	(0.013)	(0.015)	(0.032)
Age splines (years)												
18-25	0.002	0.008	0.006	0.005	0.001	-0.013	0.002	0.007	0.006	0.006	0.002	-0.013
	(0.005)	(0.008)	(0.005)	(0.005)	(0.005)	(0.009)	(0.005)	(0.007)	(0.005)	(0.005)	(0.005)	(0.008)
26-35	0.002	-0.001	0.003	0.004	0.003	0.007	0.003	-0.002	0.003	0.004	0.003	0.008
	(0.003)	(0.004)	(0.002)	(0.002)	(0.003)	(0.005)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.006)
36-45	-0.010***	-0.020***	-0.011***	-0.009***	-0.005	0.003	-0.010***	-0.021***	-0.011***	-0.010***	-0.005	0.003
	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)	(0.008)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)	(0.008)
46-55	0.004	0.011	0.004	0.000	0.003	-0.007	0.004	0.010	0.004	0.001	0.004	-0.006
	(0.007)	(0.009)	(0.005)	(0.005)	(0.007)	(0.012)	(0.007)	(0.008)	(0.005)	(0.005)	(0.007)	(0.012)
56-65	0.038**	0.007	0.013	0.026**	0.038**	0.083**	0.038**	0.007	0.013	0.027**	0.038**	0.082**
	(0.017)	(0.018)	(0.011)	(0.011)	(0.017)	(0.039)	(0.017)	(0.017)	(0.010)	(0.012)	(0.019)	(0.036)
Education level (base: College)												
Primary	-0.100***	-0.140***	-0.092***	-0.096***	-0.067***	-0.065	-0.112***	-0.161***	-0.102***	-0.107***	-0.078***	-0.083*
	(0.020)	(0.053)	(0.025)	(0.021)	(0.024)	(0.049)	(0.023)	(0.052)	(0.026)	(0.021)	(0.022)	(0.049)
Secondary	-0.053***	-0.012	-0.042***	-0.063***	-0.060***	-0.089***	-0.063***	-0.032	-0.052***	-0.073***	-0.068***	-0.104***
	(0.016)	(0.027)	(0.014)	(0.015)	(0.019)	(0.032)	(0.017)	(0.028)	(0.013)	(0.016)	(0.019)	(0.032)
Technical	-0.017	-0.003	0.001	-0.014	-0.034*	-0.052	-0.022	-0.013	-0.003	-0.018	-0.038**	-0.060*
	(0.017)	(0.029)	(0.015)	(0.014)	(0.019)	(0.037)	(0.017)	(0.029)	(0.014)	(0.015)	(0.019)	(0.036)
Industry (base: Manufacture)												
Agriculture	0.259	0.074	0.086	0.071	0.139	0.464	0.279	0.112	0.076	0.058	0.159	0.526
	(0.188)	(0.126)	(0.113)	(0.106)	(0.157)	(0.428)	(0.183)	(0.122)	(0.111)	(0.107)	(0.171)	(0.403)
Construction	0.273***	0.149***	0.147***	0.241***	0.358***	0.582***	0.276***	0.167***	0.143***	0.233***	0.359***	0.584***
	(0.032)	(0.046)	(0.026)	(0.033)	(0.044)	(0.086)	(0.029)	(0.039)	(0.023)	(0.028)	(0.041)	(0.088)
Wholesale and retail	-0.052*	-0.101***	-0.080***	-0.036	-0.041	-0.001	-0.079***	-0.060*	-0.107***	-0.086***	-0.087***	-0.034
	(0.027)	(0.037)	(0.024)	(0.026)	(0.037)	(0.061)	(0.020)	(0.036)	(0.022)	(0.024)	(0.030)	(0.044)
Transp., storage, and comm.	-0.110***	-0.213***	-0.142***	-0.072**	-0.079*	0.034	-0.114***	-0.204***	-0.156***	-0.089***	-0.082**	0.053
	(0.035)	(0.066)	(0.035)	(0.034)	(0.046)	(0.081)	(0.033)	(0.062)	(0.036)	(0.034)	(0.040)	(0.065)
FIRE and Services	0.072**	0.022	0.000	0.046	0.036	0.197***	0.081***	0.085**	-0.003	0.033	0.032	0.212***
	(0.032)	(0.039)	(0.026)	(0.028)	(0.036)	(0.073)	(0.023)	(0.038)	(0.023)	(0.027)	(0.031)	(0.059)
Occupation (base: Technical)												
Managers and professionals	0.072	0.065	0.051	0.087*	0.069	-0.037						
	(0.067)	(0.043)	(0.045)	(0.045)	(0.077)	(0.156)						
Clerical workers	-0.123***	-0.019	0.016	-0.012	-0.156***	-0.429***						
	(0.037)	(0.046)	(0.032)	(0.031)	(0.052)	(0.121)						
Service and sales workers	-0.179***	-0.074*	-0.084***	-0.140***	-0.223***	-0.359***						
	(0.038)	(0.041)	(0.029)	(0.028)	(0.041)	(0.114)						
Craft and trades workers	-0.129***	-0.145***	-0.019	-0.045	-0.135***	-0.316***						
	(0.039)	(0.042)	(0.031)	(0.031)	(0.043)	(0.118)						
Machine operators	-0.156***	-0.185***	-0.074**	-0.098**	-0.151***	-0.283**						
	(0.045)	(0.060)	(0.036)	(0.041)	(0.051)	(0.120)						
Elementary occupations	-0.145***	-0.113***	-0.058**	-0.104***	-0.159***	-0.320***						
	(0.041)	(0.037)	(0.027)	(0.025)	(0.044)	(0.120)						
Occup. complexity score							0.005	-0.005	-0.003	0.002	0.007***	0.008
							(0.004)	(0.003)	(0.002)	(0.001)	(0.002)	(0.006)
If employed in Venezuela (d)	0.003	-0.033	-0.014	0.001	0.009	0.008	0.005	-0.031	-0.012	0.004	0.011	0.009
	(0.017)	(0.028)	(0.016)	(0.016)	(0.020)	(0.039)	(0.016)	(0.029)	(0.017)	(0.016)	(0.022)	(0.038)
Length of stay in Peru	0.011***	0.010***	0.009***	0.011***	0.012***	0.013***	0.011***	0.011***	0.009***	0.011***	0.012***	0.013***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)
Constant	1.153***	0.546***	0.745***	0.992***	1.369***	2.112***	1.001***	0.460**	0.708***	0.909***	1.189***	1.776***
	(0.112)	(0.176)	(0.120)	(0.120)	(0.139)	(0.239)	(0.116)	(0.179)	(0.118)	(0.109)	(0.122)	(0.203)
N	6125	6125	6125	6125	6125	6125	6125	6125	6125	6125	6125	6125
R^2	0.098	0.050	0.076	0.092	0.076	0.045	0.091	0.048	0.072	0.084	0.069	0.039
Model χ^2 test	1025.455	141.974	490.049	687.844	469.782	190.413	774.906	179.833	371.371	547.790	375.508	195.170
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Demog. vars χ^2 test	54.248	36.714	50.113	65.594	23.445	23.017	60.772	49.261	73.271	77.276	30.288	25.885
	[0.000]	[0.000]	[0.000]	[0.000]	[0.005]	[0.006]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]
Industry vars χ^2 test	175.587	55.499	119.066	144.204	100.077	68.562	219.068	70.279	143.656	181.618	134.308	79.198
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Occupation vars χ^2 test	42.672	24.280	46.171	70.483	65.420	16.761	1.789	2.390	1.995	1.409	11.600	1.721
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.010]	[0.181]	[0.122]	[0.158]	[0.235]	[0.001]	[0.190]

Notes: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years old who arrived after January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The table shows the partial effects on the unconditional distribution of the dependent variable at selected statistics from 2 different models. College includes also Postgraduate level. Agriculture industry includes also forestry, fishing and Mining and quarrying. Wholesale and Retail industry includes also Hotels and Restaurants and FIRE and Services Industry includes communication, social and personal services. Technical workers occupations includes Skilled agricultural, forestry and fishery workers. Occupational complexity score following Ottaviano et al. (2013). Department dummies not shown. Bootstraped SEs (in parenthesis) and p-values of the χ^2 tests (in brackets) adjusting for clustering. (d)=dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENPOVE (2018) data.

Figure 3.5 – Aggregate Reweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by self-perceived discrimination



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages inclu
> tion and total Wage
structure effects. The outcome model includes occupational dummies, see the text for details on the treatment model and the decomposition. (g) means that the component of tl
the corresponding dummies (see table 2 for the categories of the control variables). Vertical lines correspond to the 95% bootstrapped confidence intervals adjusting for cluster
Source: Author's calculations using ENPOVE (2018) data.

Table 3.5 – Detailed Reweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by self-perceived discrimination

	Mean		p10		p25		p50		p75		p90	
Overall												
Discriminated	1.032***	(0.017)	0.515***	(0.022)	0.772***	(0.020)	1.007***	(0.024)	1.271***	(0.016)	1.632***	(0.031)
Non discriminated	1.089***	(0.011)	0.572***	(0.018)	0.815***	(0.012)	1.044***	(0.010)	1.311***	(0.013)	1.629***	(0.021)
Disc. rwgt. as non	1.023***	(0.016)	0.508***	(0.031)	0.758***	(0.022)	1.002***	(0.025)	1.266***	(0.016)	1.614***	(0.030)
Obs. difference	-0.057***	(0.014)	-0.057**	(0.023)	-0.043**	(0.018)	-0.037	(0.025)	-0.040**	(0.019)	0.003	(0.030)
Total explained	0.009	(0.010)	0.007	(0.019)	0.014	(0.013)	0.006	(0.015)	0.005	(0.011)	0.018	(0.020)
Total unexplained	-0.066***	(0.014)	-0.064**	(0.030)	-0.058***	(0.020)	-0.043*	(0.025)	-0.045**	(0.019)	-0.015	(0.030)
Explained Component												
Specification error	-0.001	(0.006)	-0.004	(0.012)	0.006	(0.009)	0.000	(0.010)	-0.003	(0.009)	0.007	(0.013)
Pure explained	0.010	(0.010)	0.011	(0.013)	0.008	(0.009)	0.005	(0.009)	0.008	(0.011)	0.010	(0.017)
Sex	0.001	(0.002)	0.001	(0.002)	-0.001	(0.002)	-0.000	(0.001)	0.002	(0.002)	0.002	(0.003)
Age (g)	0.000	(0.002)	0.000	(0.003)	0.001	(0.002)	0.001	(0.002)	0.000	(0.002)	-0.002	(0.003)
Education (g)	0.007**	(0.003)	0.005	(0.005)	0.005*	(0.003)	0.004	(0.003)	0.006*	(0.003)	0.007*	(0.004)
Industry (g)	-0.011*	(0.006)	-0.008	(0.005)	-0.007*	(0.004)	-0.010**	(0.005)	-0.012	(0.008)	-0.015	(0.015)
Occupation (g)	-0.003	(0.003)	-0.004	(0.004)	-0.003	(0.002)	-0.003	(0.003)	-0.006	(0.004)	-0.002	(0.006)
Employed in Venez.	-0.000	(0.001)	-0.002	(0.002)	-0.001	(0.001)	-0.001	(0.001)	0.001	(0.001)	0.001	(0.002)
Length in Peru	0.010**	(0.004)	0.004	(0.005)	0.006*	(0.003)	0.009***	(0.003)	0.012***	(0.004)	0.014***	(0.005)
Region	0.006	(0.004)	0.015*	(0.009)	0.008	(0.005)	0.005	(0.004)	0.004	(0.005)	0.005	(0.005)
Unexplained Component												
Reweighting error	0.000	(0.003)	-0.000	(0.005)	0.000	(0.003)	-0.000	(0.003)	-0.000	(0.004)	0.003	(0.006)
Pure unexplained	-0.066***	(0.015)	-0.064**	(0.029)	-0.058***	(0.020)	-0.042*	(0.025)	-0.045**	(0.019)	-0.018	(0.030)
Sex	-0.003	(0.021)	-0.033	(0.030)	0.007	(0.022)	0.002	(0.020)	-0.008	(0.023)	0.022	(0.043)
Age (g)	-0.257	(0.232)	-0.546	(0.414)	-0.002	(0.263)	0.050	(0.267)	-0.174	(0.286)	-0.964*	(0.543)
Education (g)	-0.011	(0.022)	-0.007	(0.033)	-0.008	(0.024)	-0.001	(0.026)	-0.015	(0.024)	-0.019	(0.049)
Industry (g)	0.003	(0.044)	-0.097	(0.069)	-0.014	(0.048)	-0.024	(0.062)	0.033	(0.069)	0.127	(0.112)
Occupation (g)	0.046	(0.080)	-0.019	(0.092)	-0.066	(0.062)	-0.052	(0.073)	0.044	(0.105)	0.350**	(0.162)
Employed in Venez.	0.003	(0.028)	-0.034	(0.047)	-0.011	(0.040)	-0.033	(0.034)	0.042	(0.036)	0.022	(0.069)
Length in Peru	0.002	(0.024)	-0.045	(0.043)	-0.021	(0.028)	0.008	(0.021)	0.016	(0.028)	0.007	(0.053)
Region	0.007	(0.015)	0.026	(0.029)	0.007	(0.022)	0.000	(0.013)	-0.011	(0.019)	0.003	(0.033)
Constant	0.146	(0.255)	0.691	(0.456)	0.049	(0.285)	0.007	(0.302)	0.029	(0.331)	0.433	(0.563)

Notes: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years old who arrived after January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The table show the components of the detailed decomposition, reweighting those who perceived discrimination as if they did not, at selected statistics of the unconditional distribution of the dependent variable using coefficients from RIF regression. The outcome model includes occupational dummies, see the text for details on the treatment model. (g) means that the component of the detailed decomposition refers to the effect after grouping the corresponding dummies (see table 2 for the categories of the control variables). Bootstraped SEs (in parenthesis) adjust for clustering. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENPOVE (2018) data.

3.6.2 Actual unequal treatment and the wage gap

The evidence in the previous subsection suggests the unconditional gaps between those who perceived workplace discrimination and those who did not are almost exclusively explained by the wage structure component. However, this does not inform the more critical question about whether or not the labour market differentially pays the observed characteristics of Venezuelans compared to Peruvian workers. This actual differential may be larger than that based on the perception of discrimination in the previous subsection. To assess this, we now merge the Peruvian and Venezuelan data and construct a treatment variable to re-do the decomposition analysis. The treatment assignment is now based on the worker's nationality, with $d = 1$ being Peruvians and $d = 0$ otherwise. As argued in section 3.3, this is a more direct approach to actual discrimination.

The estimated wage-structure effect for Venezuelans (based on the procedure described in section 3.4) suggest that if their characteristics were rewarded commensurate with those of a comparable Peruvian worker (upper panel Figure 3.6 with occupation modelled as a set of dummies), the resulting simulated distribution would exhibit higher means and percentiles, and also a smaller standard deviation (0.49 for the observed and 0.27 for the simulated). For example, the observed average of ω for Venezuelans would be 0.29 log points higher than that observed (it rises from 1.08 to 1.37). In addition, the first decile would be 0.50 log points larger (rising from 0.55 to 1.10), and the difference would progressively reduce, contracting up to 0.44 log points at the last decile (rising from 1.63 to 1.67). The effect of differential rewards between Peruvians and Venezuelans operates stronger on males than females (lower panel of Figure 3.6). For these, the simulated mean is 0.4 log points larger than the observed while that at the first and last decile is 89 and 11 percentage points, respectively. Similar conclusions hold when we model occupation using the complexity score (Figure 3.A10). The sign and magnitude of the estimated coefficients based on the Peruvian sample on which these results are based (Table 3.A5 in Appendix) provide a sense of validity to these direct discrimination measures.

The differences between the simulated and observed distributions suggests a sizeable unequal treatment for Venezuelan workers relative to their Peruvian counterparts. In order to analyze how much of the observed gap is explained by the differential labour market payment based on the nationality of the workers, we use the classic (unweighted) OB RIF-decomposition following Equation 3.8[49]. Examining these gaps and their decomposition across quantiles suggests two essential facts (Figure 3.7). First, the observed (total) differences in ω between immigrants and natives monotonically increase as we move towards the upper parts of the unconditional log hourly wage distribution. At the 10th percentile, the gap is negative, with Peruvians' hourly wages 20% lower than Venezuelans. At the 20th percentile, the difference stops being statistically different from zero. From that percentile onwards, the gap turns positive with Peruvians earning a wage 80% higher at the 90th percentile. Secondly, similar to what we find when analysing self-perceived workplace discrimination, most of these differences are attributed to a wage-structure effect. Thus, the Δ_X component is not statistically significant for most of the distribution (except at the top quartile). These same results hold even if we decompose the observed gap using the reweighted method (Figure 3.A11).

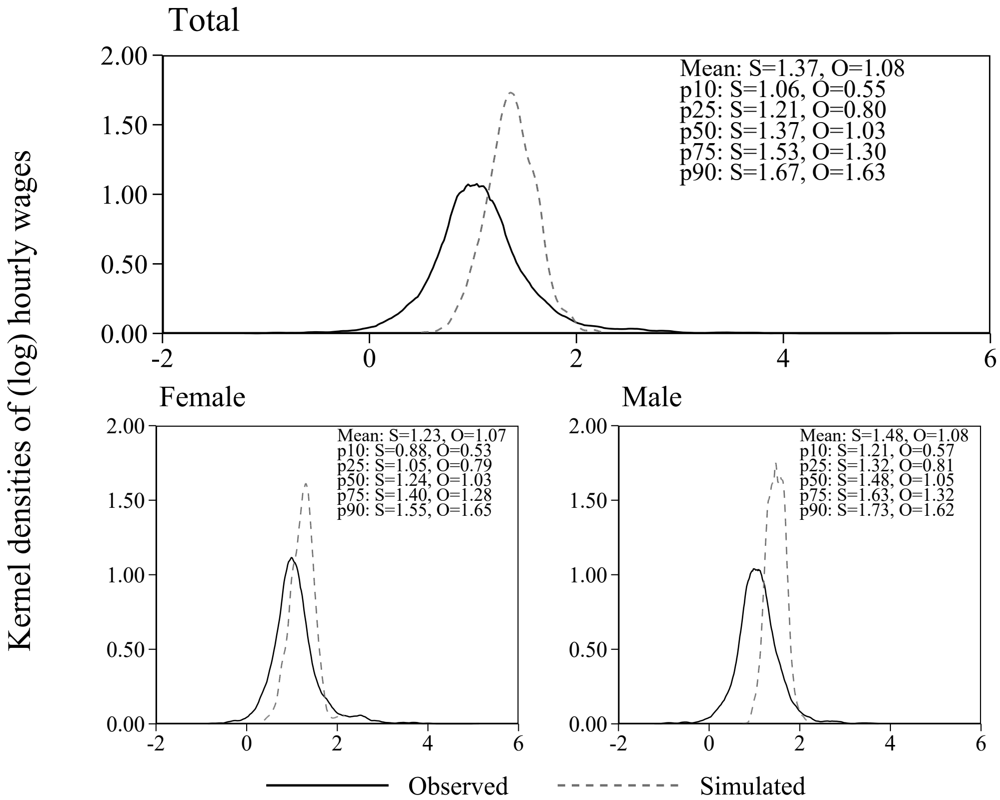
In contrast to the detailed decomposition for self-perceived discrimination, the contribution of some key factors behind the wage-structure effect is statistically different from zero (Table 3.6)[50]. Specifically, Peruvians obtain higher returns for being male than Venezuelans across the distribution, and this magnitude is not negligible. Indeed, this difference ranges between 10% (which occurs at the median) to 20% (at the very bottom of the unconditional distribution). Interestingly, Peruvians, whose educational endowments are

---

[49]Note how in this case, the difference corresponds to $\hat{\Delta}_O^\tau = \hat{\Delta}_S^\tau + \hat{\Delta}_X^\tau = \overline{\mathbf{x}}^{d=0'} \left(\hat{\boldsymbol{\beta}}_\tau^{d=1} - \hat{\boldsymbol{\beta}}_\tau^{d=0}\right) + \left(\overline{\mathbf{x}}^{d=1} - \overline{\mathbf{x}}^{d=0}\right)' \hat{\boldsymbol{\beta}}_\tau^{d=1}$. Hence, Δ_S^τ now identifies a quantile-treatment effect on the non-treated, i.e. how much higher would be the wages of Venezuelans if their attributes were rewarded as Peruvians.

[50]Note how, unlike Equation 3.11, in this case the detailed decompositions are given by $\hat{\Delta}_S^\tau = \left(\hat{\beta}_{\tau,1}^{d=1} - \hat{\beta}_{\tau,1}^{d=0}\right) + \sum_{k=2}^K \overline{x}_k^{d=0} \left(\hat{\beta}_{\tau,k}^{d=1} - \hat{\beta}_{\tau,k}^{d=0}\right)$ and $\hat{\Delta}_X^\tau = \sum_{k=1}^K \left(\overline{x}_k^{d=1} - \overline{x}_k^{d=0}\right) \hat{\beta}_{\tau,k}^{d=1}$

Figure 3.6 – Observed and simulated distributions of (log.) hourly wages for Venezuelan immigrant workers in Peru

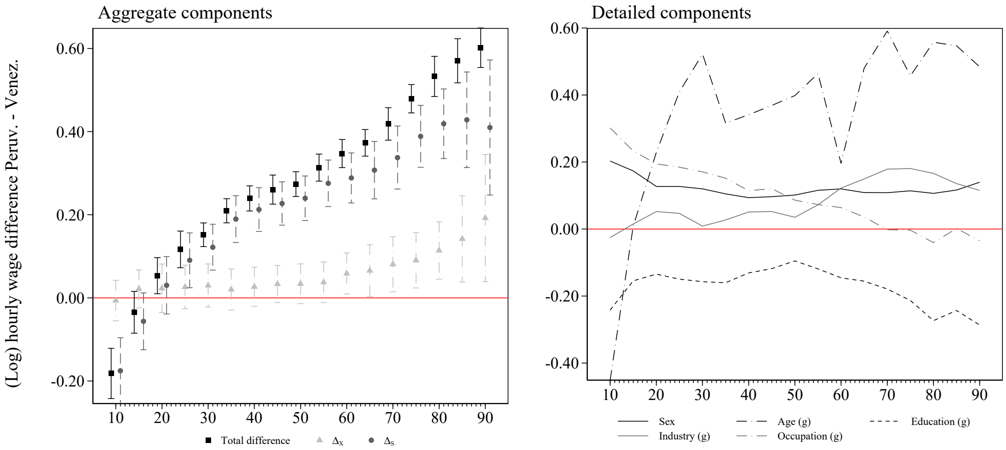


Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years and those who arrived to Peru from January 2016 onwards. Informality approximated by the lack of health insurance in their occupation. Simulated distribution based on coefficients from Mincer equations of (log) hourly wages including income from the main and secondary occupations on Peruvian workers (shown in table A6) which include occupational dummies. Real hourly wages in 2007 PEN.
Source: Author's calculations using ENAHO (2018) and ENPOVE data (2018).

lower, also receive lower education returns than Venezuelans. This is evidence that, effectively, employers are aware of the lower skills of their natives, and hence are paid accordingly. In addition, differences in occupations play a relevant role in Δ_X and, not surprisingly, favour Peruvians across the unconditional distribution of the (log) hourly wages.

We analyze the robustness of these results by running the same checks as in the previous subsection. Again, changing the bandwidth and kernel density provides the same results as our main findings (see Figure 3.A12). This robustness is also found when we take as the outcome the (log.) hourly wages arising only from the main occupation (Figure 3.A13). The last check, restricting the type of workers who are included in the sample for estimation, results in a more significant wage gap favouring Peruvians throughout the unconditional distribution of ω (Figure 3.A14). The disadvantage of Peruvians at the very bottom of the distribution disappears under this redefinition. Instead, the former have a wage advantage in 10 log points compared to Venezuelans. However, the main message in the OB decomposition remains: the wage gap's main driver is the differential treatment that Venezuelans face in the labour market.

Figure 3.7 – Aggregate Unweighted OB decomposition of (log) hourly wages by Peruvians and Venezuelan workers



Note: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those regions who arrived to Peru from January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows the decomposition components, without reweighting, at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the decomposition. Vertical lines correspond to the bootstrapped SEs adjusting for clustering. *Source*: Author's calculations using ENAHO and ENPOVE (2018) data.

Table 3.6 – Detailed unweighted OB decomposition of (log) hourly wages by Peruvians and Venezuelan workers

	Mean		p10		p25		p50		p75		p90	
Overall												
Peruvian	1.322***	(0.013)	0.373***	(0.026)	0.920***	(0.019)	1.334***	(0.012)	1.776***	(0.015)	2.231***	(0.017)
Venezuelans	1.077***	(0.011)	0.555***	(0.019)	0.804***	(0.011)	1.061***	(0.009)	1.297***	(0.009)	1.629***	(0.019)
Obs. difference	0.245***	(0.017)	-0.182***	(0.031)	0.117***	(0.022)	0.274***	(0.015)	0.479***	(0.017)	0.602***	(0.024)
Total explained	0.086**	(0.041)	-0.006	(0.025)	0.026	(0.027)	0.034	(0.024)	0.090***	(0.034)	0.192**	(0.078)
Total unexplained	0.159***	(0.046)	-0.175***	(0.040)	0.090***	(0.034)	0.240***	(0.027)	0.389***	(0.038)	0.410***	(0.083)
Explained Component												
Sex	-0.001	(0.001)	0.002*	(0.001)	0.001	(0.001)	0.001	(0.001)	0.000	(0.001)	-0.002	(0.002)
Age (g)	0.013	(0.012)	-0.032**	(0.014)	-0.007	(0.009)	0.003	(0.010)	0.028**	(0.014)	0.062***	(0.022)
Education (g)	-0.018***	(0.004)	-0.012**	(0.006)	-0.017***	(0.003)	-0.022***	(0.004)	-0.017***	(0.004)	-0.020***	(0.007)
Industry (g)	0.055	(0.041)	0.020	(0.022)	0.028	(0.023)	0.020	(0.022)	0.035	(0.031)	0.089	(0.073)
Occupation (g)	0.031***	(0.006)	0.011**	(0.005)	0.015***	(0.005)	0.027***	(0.005)	0.039***	(0.007)	0.057***	(0.016)
Region (g)	0.007	(0.004)	0.005	(0.006)	0.008	(0.005)	0.005	(0.005)	0.005	(0.004)	0.007	(0.005)
Unexplained Component												
Sex	0.144***	(0.020)	0.201***	(0.040)	0.127***	(0.024)	0.101***	(0.020)	0.114***	(0.021)	0.142***	(0.033)
Age (g)	0.320	(0.219)	-0.420	(0.458)	0.417	(0.294)	0.396	(0.259)	0.429*	(0.260)	0.422	(0.359)
Education (g)	-0.187***	(0.034)	-0.230***	(0.060)	-0.132***	(0.036)	-0.073**	(0.033)	-0.196***	(0.049)	-0.266***	(0.064)
Industry (g)	0.038	(0.058)	-0.045	(0.079)	0.019	(0.054)	0.015	(0.050)	0.145**	(0.067)	0.027	(0.115)
Occupation (g)	0.063	(0.044)	0.291***	(0.063)	0.170***	(0.042)	0.059	(0.038)	-0.042	(0.051)	-0.092	(0.098)
Region (g)	-0.036**	(0.015)	-0.088***	(0.031)	-0.034*	(0.017)	-0.003	(0.014)	-0.014	(0.020)	-0.077***	(0.029)
Constant	-0.182	(0.240)	0.115	(0.473)	-0.476	(0.310)	-0.254	(0.275)	-0.048	(0.280)	0.253	(0.397)

Notes: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those regions who arrived to Peru from January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The table show the components of the detailed decomposition, reweighting those who perceived discrimination as if they did not, at selected statistics of the unconditional distribution of the dependent variable using coefficients from RIF regression. The outcome model includes occupational dummies, see the text for details on the treatment model. (g) means that the component of the detailed decomposition refers to the effect after grouping the corresponding dummies (see table 2 for the categories of the outcome variables). Bootstraped SEs (in parenthesis) adjust for clustering. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO and ENPOVE (2018) data.

3.6.3 Impact of wage-structure on perception of unequal treatment

The above partial (or ceteris paribus) gaps and the wage-structure components allow us to disentangle the factors driving the observed gaps that are attributable to discrimination experienced by Venezuelans. Attention now turns to an explanation of the factors that drive the Venezuelans' perception of workplace discrimination. We are interested in the impact of the nationality-based wage-structure effect, Δ_S (expressed again in logs), once we control for factors that can be regarded as relevant predictors of migrants' expectations of equal treatment and their awareness of inequalities (see section 3.5 for the construction of Δ_S).

As discussed in section 3.5, a potential issue, usually overlooked by studies who analyze the determinants of this perception is that the violation of assumptions behind conditional maximum likelihood estimation results in estimators that converge instead to a pseudo-true value. The p-values from the tests for heteroskedasticity, normality and functional form (bottom of the Table 3.7) confirm that all key assumptions are comfortably satisfied in our probit regressions. The coefficient of interest, the unequal treatment measure, has a positive partial effect across the different estimated models and always retains the same level of statistical significance throughout (Table 3.7). This holds regardless of how we estimate the wage-structure effect, suggesting that perceptions of discrimination correlate with wage disparity and confirm that employees correctly are underpaid. From the model with all covariates included, doubling the wage-structure effect increases the probability of perceiving discrimination by ($\ln 2 \times 0.040 =$) 2.8 percentage points relative to a mean of 20% (see Table 3.2). Even though its numerical magnitude is small, it is still larger than what is found by Biddle (2013) and Banerjee (2008); in this latter case, the same change to Δ_s results in a 0.01 percentage points increase in the probability.[51]

In contrast, some other covariates in the table exert larger effects. Being a female, for example, increases the probability of perceiving discrimination by four percentage points. Those with higher education are also more prone to perceive workplace discrimination. A Venezuelan with some college or a technical degree has an increased perception of discrimination of almost five percentages points relative to someone with just primary education. This indicates that the inequality exerted against the average Venezuelan worker makes the migrant more aware of attitudes of discrimination against him. Thus, if they find that their qualifications are undervalued, they may believe that discrimination is at play (Banerjee 2008). One extra month in Peru also has a very moderate effect on the perception of discrimination. However, it becomes non-negligible when we consider that the average Venezuelan, who has been eight months in Peru, has a probability of perceiving unequal treatment five percentage points larger than when he arrived. Interestingly, while age is not an statistically significant determinant of self-perceiving discrimination, previous work experience in Venezuela is a statistically significant determinant. This asymmetry can be ascribed to the fact that Venezuelan workers fix their expectations based on their work experience acquired in Venezuela.

We now analyze the impact of nationality-based wage-structure effect Δ_S at different percentiles of the unconditional (log) hourly wages distribution. The plot of the point estimates Δ_S for different quantiles suggest that, on average, the Venezuelan workers react differently to disparities at different points of the unconditional (log) hourly wage distribution. The estimated effects are more pronounced at the middle of the distribution than at the bottom (Figure 3.2). This provides indication that it is the inequality exerted against the median Venezuelan worker that the migrant is more aware of. In turn, inequality exerted against those more disadvantaged is less relevant for the worker's perception. Possibly, these workers justify this specific wage structure in terms of lower qualifications, experience or other unobservable attributes for those in the bottom part of the distribution. However, discrimination from the 70th percentile onwards stops being perceived. This latter is consistent with the results above, as workplace discrimination in terms of self-perception lacks statistical significance at this upper part of the distribution

[51] Admittedly, including this estimated variable Δ_S in the model can affect the estimated variance-covariance of the estimators (see Wooldridge 2010). Nonetheless, the confirmation provided by the different tests that the assumptions of the probit model are fulfilled suggests that the estimated variances also accurately estimate the population variance even under this generated regression case.

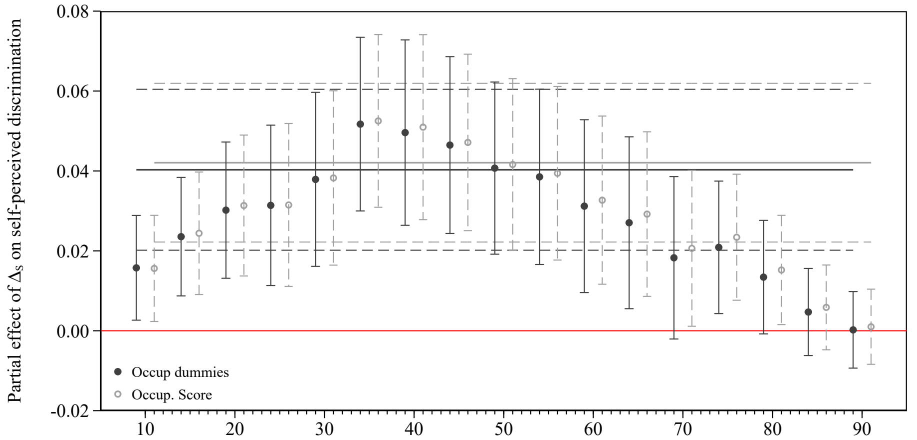
Table 3.7 – Impact of mean wage structure effect on perception of discrimination of Venezuelan immigrant workers (marginal effects)

	Peruvian model using occup. dummies				**Peruvian model using occup. complexity**			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Δ_S (wage structure)	0.042***	0.045***	0.039***	0.040***	0.043***	0.046***	0.040***	0.042***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
If the individual is male (d)	-0.040***	-0.042***	-0.037***	-0.040***	-0.041***	-0.042***	-0.038***	-0.041***
	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)
Age splines (years)								
18-25		-0.005	-0.006	-0.006		-0.005	-0.006	-0.006
		(0.004)	(0.004)	(0.004)		(0.004)	(0.004)	(0.004)
26-35		-0.000	-0.000	-0.000		-0.000	-0.000	-0.000
		(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
36-45		-0.000	-0.000	-0.000		-0.000	-0.000	-0.000
		(0.003)	(0.003)	(0.003)		(0.003)	(0.003)	(0.003)
46-55		-0.001	-0.001	-0.001		-0.001	-0.001	-0.001
		(0.005)	(0.005)	(0.005)		(0.005)	(0.005)	(0.005)
56-65		-0.002	-0.002	-0.002		-0.002	-0.002	-0.002
		(0.011)	(0.011)	(0.011)		(0.011)	(0.011)	(0.011)
Education level (base: primary)								
Secondary educ.			0.031	0.031			0.031	0.031
			(0.021)	(0.021)			(0.022)	(0.021)
Technical educ.			0.044**	0.045**			0.044**	0.044**
			(0.019)	(0.019)			(0.019)	(0.019)
College and PG educ.			0.045**	0.048**			0.044**	0.046**
			(0.019)	(0.019)			(0.019)	(0.019)
Occup. complexity score				-0.005				-0.005*
				(0.003)				(0.003)
Time in Peru	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
If employed in Venezuela (d)	0.034**	0.041***	0.039***	0.040***	0.033**	0.041***	0.039***	0.040***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
N	6125	6125	6125	6125	6125	6125	6125	6125
R^2	0.010	0.011	0.012	0.013	0.010	0.011	0.012	0.013
Model χ^2 test	44.448	49.126	60.882	66.970	45.156	49.985	61.227	67.716
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Normality test	0.756	0.407	1.356	2.643	0.580	0.378	1.240	2.363
	[0.685]	[0.816]	[0.508]	[0.267]	[0.748]	[0.828]	[0.538]	[0.307]
Heterosk. test	6.609	9.612	12.464	17.822	6.808	9.515	12.316	17.425
	[0.158]	[0.383]	[0.409]	[0.164]	[0.146]	[0.391]	[0.421]	[0.181]
RESET test	0.777	0.495	1.363	4.139	0.631	0.408	1.313	4.139
	[0.855]	[0.920]	[0.714]	[0.247]	[0.889]	[0.939]	[0.726]	[0.247]

Notes: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those regions who arrived to Peru from January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. Δ_S refers to the estimate of the mean wage structure effect for the Venezuelan workers, estimated as the difference between the simulated log hourly wage (using coefficients from the Mincer regression on log hourly wages on the Peruvian sample, shown in table A6) and the observed (log) hourly wage of Venezuelans. SEs (in parenthesis) and p-values of the tests (in brackets) adjusting for clustering. (d)=dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO (2018) and ENPOVE (2018) data.

The partial estimated effects of the other variables are similar to those reported in the previous table based on the mean treatment effect (Figure 3.8). Females are less aware of inequalities at this part of the distribution. The diagnostics for these regressions confirm that assumptions for maximum likelihood estimation are satisfied here. However, a feature of the estimations in this subsection is the prevalence of a low pseudo-R^2 which, in fact, is a shared characteristic with previous studies (Biddle 2013; Banerjee 2008; Daldy et al. 2013). Even though this is not as important as the statistical and economic significance of the explanatory variables (Wooldridge 2010), this just confirms the difficulty in disentangling the reasons for discrimination. As suggested by Fernández-Reino (2020), multiple factors might be at play at the same time and hence involves the study of social attitudes and stereotypes about certain groups. These are relevant issues but beyond the scope of the current economic analysis undertaken here.

Figure 3.8 – Impact of wage structure effect on perception of discrimination of Venezuelan immigrant workers (marginal effects)



Note: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those regions who arrived to Peru from January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows the partial effect of the wage structure effect, Δ_S, calculated at 17 percentiles (beginning at the 10th n steps of 5), calculated as the difference between the simulated log hourly wage (using coefficients from the Mincer regression on log hourly wages on the Peruvian sample, shown in table A6) and the RIF of the corresponding statistic of the observed (log) hourly wage of Venezuelans/ Vertical lines correspond to the bootstrapped 95% confidence intervals adjusting for clustering.
Source: Author's calculations using ENPOVE (2018) data.

3.7 Conclusions

The Venezuelan Exodus into Peru is unique since, until then, foreigners were practically absent from their labour market. These migrants also are more educated than Peruvians and most of them have prior job experience in Venezuela. Our data provides the opportunity to investigate how discrimination experienced by Venezuelans affects their wages under two complementary approaches. We show that the percentage of Venezuelans who self-perceive workplace discrimination (20%) is comparable to that of Japan, one of the countries with the lowest proportion of migrant workers globally. In contrast, this estimate is considerably larger than what is reported in the OECD and the USA or even among indigenous Peruvians, one of the groups more (if not the most) discriminated in Peru. This is somewhat unexpected because Venezuelans are a population that, unlike a large proportion of migrants into the countries listed above, have close cultural ties with Peruvians, do not face a language constraint and possess no big cultural distance to ignite such a reaction from Peruvians. This evidence indirectly confirms Kustov's (2019) view that negative attitudes of the natives are closely linked to the origin of migrants, with those from less developed countries perceived more negatively regardless of their personal characteristics. At the same time, it refutes views that negative attitudes are generally against less skilled migrants (Hainmueller et al. 2015).

Subjective (self-perceived) discrimination is associated with an observed (log) hourly wage penalty of around 5%. The disentanglement of the composition effect behind the gaps suggest that being more educated and having stayed longer in Peru plays a favourable role in the wages of those who actually perceived discrimination. However this advantage is completely counteracted by the wage-structure effect that plays against them. The Venezuelans who self-perceived discrimination, especially those with lower wages, face larger wage penalties for being employed in Wholesale and Transport Industries and in Services and sales jobs, the most common industries and occupation for these migrants characterized by unskilled tasks. In terms of objective (nationality-based) discrimination, Venezuelans earn lower wages across the lower fifth of the percentiles, but from that point onwards the wage disadvantage that Venezuelans experience relative to Peruvians increases as we move to the upper parts of the unconditional wage distribution. The advantage of Peruvians can be explained by the concentration of Venezuelan workers in low skill occupational categories (and conversely, by an under-representation in top occupational categories), which is associated to their the lower complexity of the tasks in their occupations and their occupational downgrading (see Del Pozo Segura 2021).

Finally, the perception of discrimination of Venezuelans is actually influenced by the objective discrimination they experience. They are more sensitive to the level of wage discrimination (relative to the Peruvians) at the middle of the income distribution. In turn, they tend not to be influenced by the inequality at the bottom end of the wages distribution. This contrasts with the lack of responsiveness to the wage structure effect that Banerjee (2008) finds for migrant workers in Canada, and confirms that negative attitudes towards certain minorities result in more discrimination against these, as suggested by Fernández-Reino (2020). The magnitude of these effects is small but still are larger than what previous studies, like Biddle (2013), found. Specifically, doubling the size of inequality increases the probability of perceiving discrimination by 2.8 percentage points relative to a mean of 20%. Factors related to their expectations for equitable treatment have a larger role in this perception. The level of education is known to increase expectations for career success and awareness of social inequalities. Similarly, lengthier stay in Peru increases the probability of perceiving self-discrimination.

From a policy perspective, this study goes beyond simple summary measures of migrant wages gaps and is able to disentangle its causes. Since the wage-structure effect is the main driver of observed wage gaps under the two approaches, reducing this disadvantage of Venezuelan migrants in Peru is intimately related to a change in the negative attitudes of the employers against them. Admittedly, this cannot be changed in the short run, as it involves dealing with stereotypes and perceptions ingrained in the migrant population. Nonetheless, a more effective way to reduce the impact of this component goes through mechanisms that provide incentives for their integration away from occupations where they are currently concentrated. This implies a downgrade of their abilities and greater concentration into other activities where their skills can be considered a more valuable asset. As ENPOVE data reveals, about 40% of Venezuelans in Peru have Social Sciences, Commercial and Law degrees, and 20% have Engineering backgrounds. It is highly likely that their professional skills will be more valued in occupations and industries that match their training. A sensible first step in that direction lies moving away from the permits system, due to its failure in Peru (and also in Colombia, as assessed by Bahar et al. 2020b), and considering citizenship rights for the Venezuelans instead. This can eliminate barriers for an important proportion of qualified immigrants, who previously lacked documentation, to apply for formal jobs. Colombia, a country comparable in terms of the size informal sector and absorption of Venezuelan migrants, has been the first country to follow this route since the beginning of 2021, providing Venezuelans with citizenship rights over health and education.

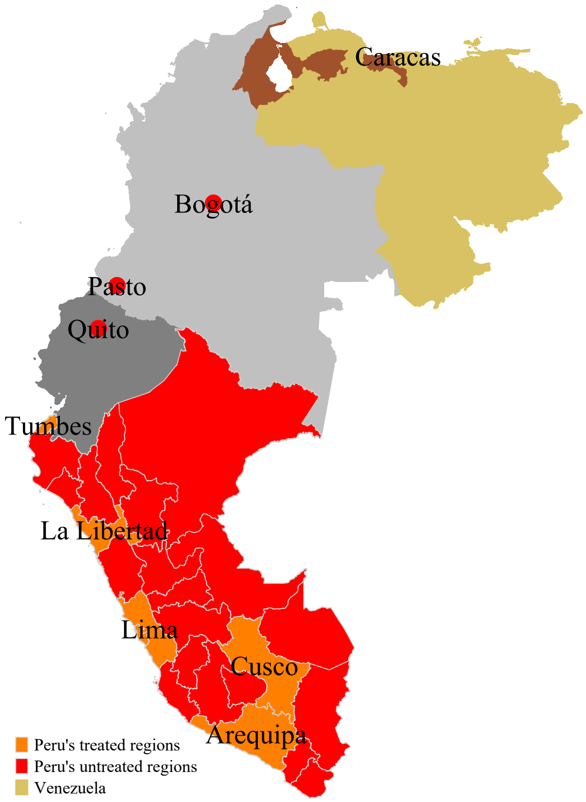
The Venezuelans' observed labour market skills, approximated here by education, are higher than those of Peruvian natives. However, a further possibility not captured in our data is that these higher skills are not necessarily useful for the elementary occupations and service and sales occupations within the informal sector that Venezuelans generally perform. By the very nature of the tasks involved in these activities in Peru (relatively unqualified and requiring face-to-face attention to customers), the set of skills acquired in formal (university) education do not necessarily match the more manual and less cognitive-related tasks required. However, given that this is Peru's first experience with a large influx of migrants, there is scope for designing and introducing policies that can promote training and equal opportunity for the upward mobility of migrant workers. These can set a precedent for possible future inflows of migrant workers, given its recent history of rapid economic growth. However, we are aware that the sizeable informal size of Peru's economy is an essential structural characteristic that contributes to the permanence of these sizeable wage gaps. This complicates putting in place well-designed minimum wages with broad legal coverage for those sectors and occupations in which migrants are chiefly employed, as suggested by ILO (2020). In any event, a broader strategy that also includes adopting fair and effective labour migration policies is required. This needs to be designed in conjunction with policies that address decent working conditions and ensure greater coherence across employment, education and training.

Additional work needs to disentangle the role of other psychological factors behind workplace discrimination, which, to date have remained unmeasured. In fact, we are cannot unpick the effect that experiences and sensitivity to certain events have on the perceptions of discrimination of Venezuelan workers. As with other studies, we leave the influence of context and the interaction between context and individual characteristics to future research, given these are more complex relationships and require inter-disciplinary thinking beyond simply economics. Even though analysing migrants' perception represents an advantage relative to other studies in this field, the role of types of skills and human capital development of migrants are as relevant as their perceptions. Hence, an approximation of these unobserved (or hard to observe) attributes which are relevant in the labour market is needed in further studies. From a methodological point of view, further work needs to be done assessing the effect in the statistical significance induced by including a generated regressor (the wage-structure) in the model, as this might affect the estimated variance-covariance matrix. Also, the role of how alternative counterfactuals impact on the results from the re-weighted decompositions needs to be further examined, as this study focused its robustness checks in terms of definition of the outcome and employment, as well as characteristics of the Kernel estimation.

Appendix

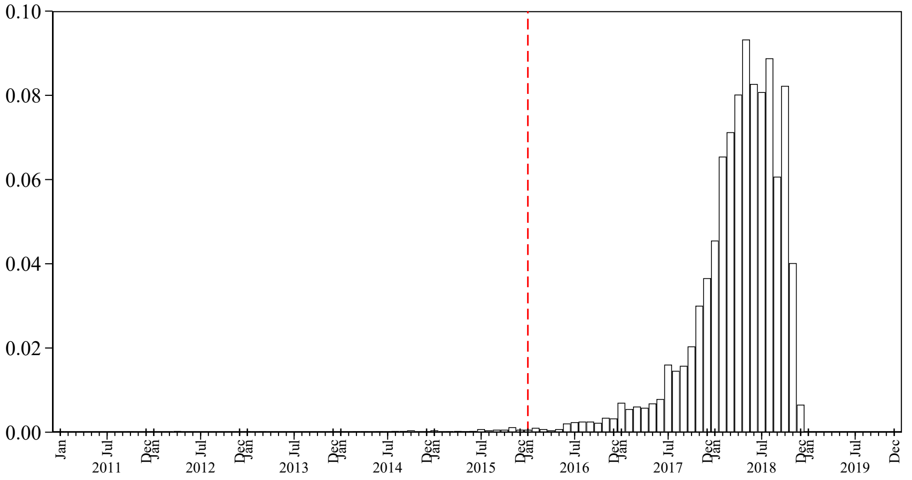
Descriptives

Figure 3.A1 – Treatment areas in Peru and Venezuela



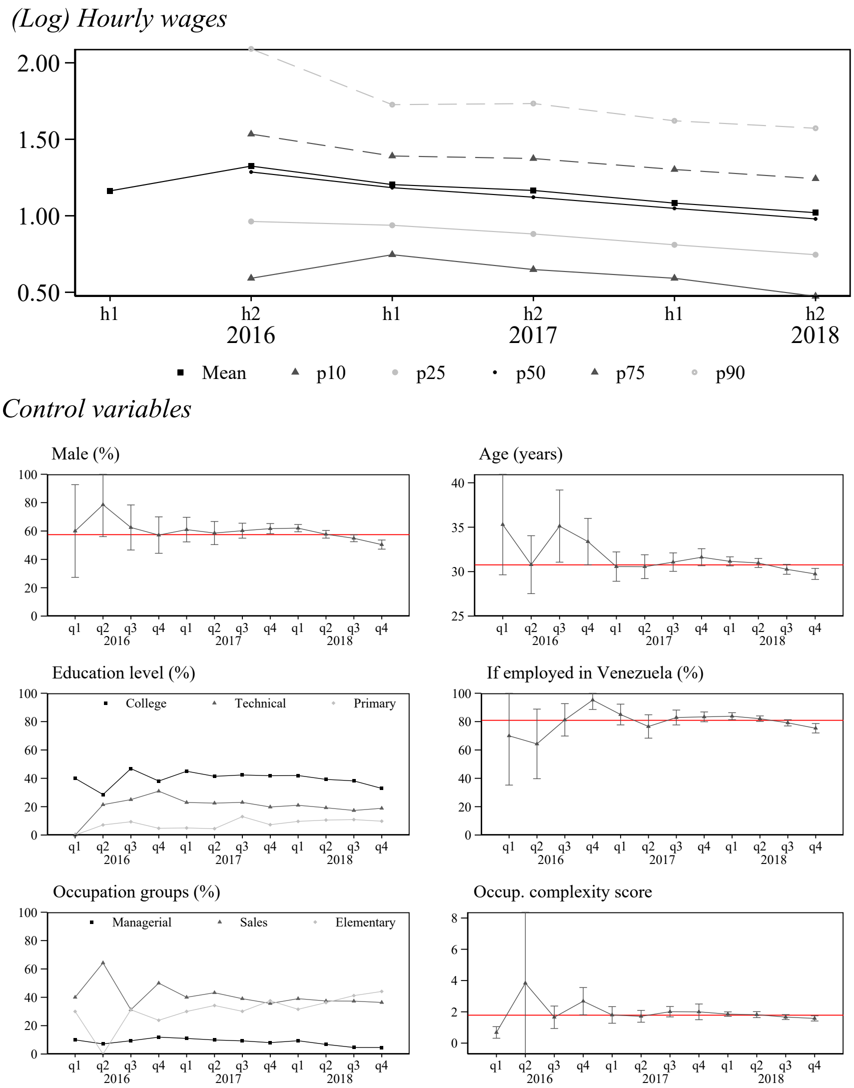
Note: Ecuador and Colombia in dark and light grey, respectively. The 5 states in Venezuela where half of the Venezuelan immigrants in Peru began their emigration journey in brown. Lines correspond to the most commmon routes followed by Venezuelan immigrants according to UNHCR (2018) *Source*: Author's calculations using ENPOVE data.

Figure 3.A2 – Venezuelan immigrant's flow and time spent in Peru (months) as of 2018 in ENPOVE sample



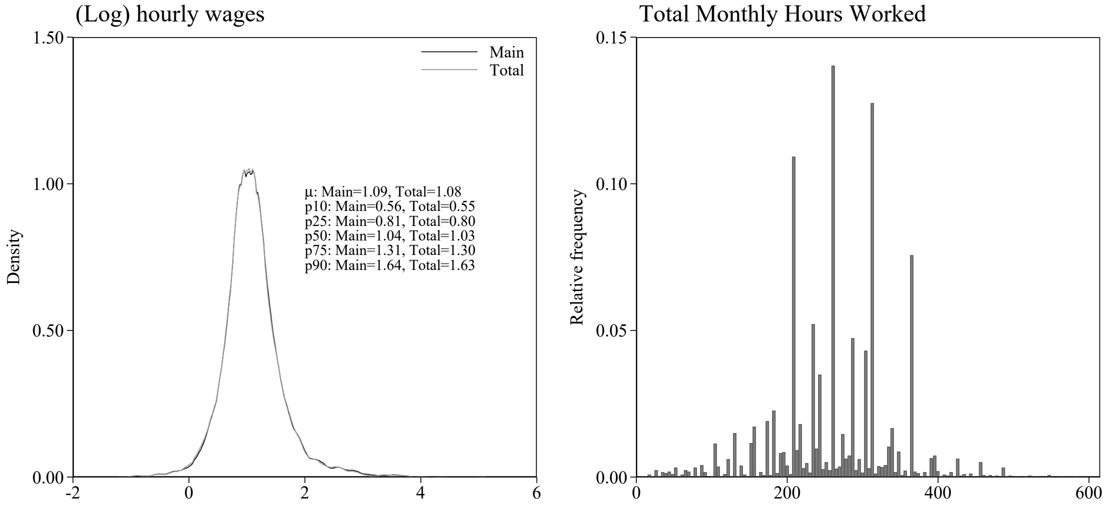
Note: Sample restricted to only those employed between 18 and 65 years. Date of arrival relative to when they entered Peru. The vertical line shows the b
of the treatment. *Source*: Author's calculations using ENPOVE data.

Figure 3.A3 – Outcome and control variables characterization for Venezuelans immigrant workers by quarter of arrival to Peru



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016 onwards. q1-q4 (h1,h2) refers to the quarter (half) of the year when they entered Peru. (Log.) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The horizontal red line shows the average in the sample. Managerial occupation includes technical workers and Skilled agricultural, forestry and fishery workers. Vertical lines correspond to the 95% confidence intervals adjust for clustering.
Source: Author's calculations using ENPOVE (2018) data.

Figure 3.A4 – Variables distribution for Venezuelan immigrant workers



Note: Sample restricted to only those Venezuelans employed between 18 and 65 years who arrived to Peru from January 2016. Total labour hourly wages and monthly hours worked include the main and secondary occupations. Real hourly wages in 2007 PEN. *Source*: Author's calculations using ENPOVE data (2018).

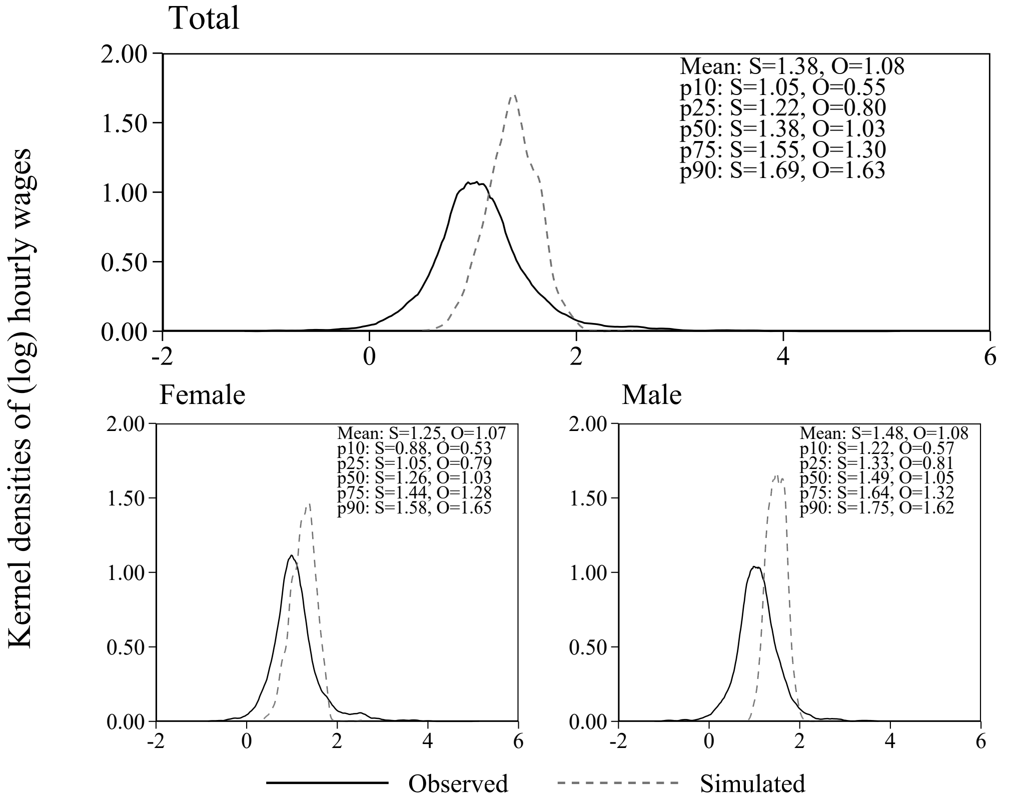
Estimations and Decompositions

Table 3.A5 – Mincer equations of (log.) hourly wages for informal Peruvian workers in treated areas

	Occupational dummies						**Occupational complexity**					
	Total		Male		Female		Total		Male		Female	
If the individual is male (d)	0.221***	(0.026)					0.243***	(0.025)				
Age splines (years)												
18-25	0.016*	(0.009)	0.021*	(0.012)	0.011	(0.014)	0.018**	(0.009)	0.020*	(0.012)	0.014	(0.014)
26-35	0.007	(0.004)	0.003	(0.005)	0.014*	(0.008)	0.008*	(0.004)	0.005	(0.005)	0.013	(0.008)
36-45	0.006	(0.005)	0.009	(0.006)	0.003	(0.008)	0.006	(0.005)	0.008	(0.006)	0.004	(0.008)
46-55	-0.010*	(0.006)	-0.015**	(0.007)	-0.002	(0.009)	-0.010*	(0.006)	-0.015**	(0.007)	-0.003	(0.009)
56-65	-0.008	(0.008)	-0.004	(0.010)	-0.009	(0.013)	-0.008	(0.008)	-0.005	(0.010)	-0.008	(0.013)
Education level (base: College)												
Primary	-0.445***	(0.050)	-0.410***	(0.066)	-0.504***	(0.075)	-0.474***	(0.048)	-0.424***	(0.063)	-0.564***	(0.072)
Secondary	-0.272***	(0.040)	-0.279***	(0.053)	-0.267***	(0.062)	-0.299***	(0.038)	-0.288***	(0.050)	-0.326***	(0.060)
Technical	-0.137***	(0.042)	-0.124**	(0.059)	-0.165***	(0.062)	-0.153***	(0.041)	-0.127**	(0.056)	-0.201***	(0.062)
Industry (base: Manufacture)												
Agriculture	-0.039	(0.055)	-0.157**	(0.061)	0.086	(0.108)	-0.024	(0.049)	-0.176***	(0.055)	0.299***	(0.090)
Construction	0.335***	(0.049)	0.244***	(0.051)	0.361*	(0.188)	0.317***	(0.048)	0.220***	(0.049)	0.409*	(0.216)
Wholesale and retail	0.105**	(0.048)	0.030	(0.056)	0.116	(0.093)	0.040	(0.042)	-0.038	(0.049)	0.214***	(0.077)
Transp., storage, and comm.	-0.003	(0.054)	-0.079	(0.064)	0.043	(0.106)	-0.005	(0.043)	-0.100**	(0.045)	0.165*	(0.094)
FIRE and Services	0.201***	(0.048)	0.116*	(0.064)	0.217**	(0.092)	0.192***	(0.045)	0.074	(0.059)	0.392***	(0.078)
Occupation (base: Technical)												
Managers and professionals	0.285***	(0.063)	0.226***	(0.083)	0.362***	(0.096)						
Clerical workers	-0.095*	(0.056)	-0.196**	(0.080)	-0.049	(0.075)						
Service and sales workers	-0.143***	(0.051)	-0.090	(0.070)	-0.186**	(0.074)						
Craft and trades workers	-0.058	(0.053)	-0.005	(0.059)	-0.348***	(0.123)						
Machine operators	-0.041	(0.058)	-0.028	(0.068)	-0.180	(0.177)						
Elementary occupations	-0.098**	(0.043)	-0.094*	(0.053)	-0.124*	(0.066)						
Score Complexity							0.009***	(0.002)	0.008***	(0.002)	0.011***	(0.002)
Constant	0.998***	(0.213)	1.154***	(0.280)	1.129***	(0.346)	0.889***	(0.210)	1.138***	(0.277)	0.827**	(0.333)
N	4989		3103		1886		4989		3103		1886	
R^2	0.121		0.105		0.124		0.118		0.102		0.118	
Model F test	28.035	[0.000]	16.773	[0.000]	10.754	[0.000]	31.269	[0.000]	18.644	[0.000]	12.595	[0.000]
Demographic vars F test	21.809	[0.000]	9.009	[0.000]	6.861	[0.000]	28.340	[0.000]	10.231	[0.000]	8.966	[0.000]
Industry vars F test	17.882	[0.000]	14.011	[0.000]	2.626	[0.023]	19.060	[0.000]	15.941	[0.000]	6.948	[0.000]
Occupation vars F test	9.757	[0.000]	5.847	[0.000]	6.487	[0.000]	34.281	[0.000]	13.611	[0.000]	21.478	[0.000]

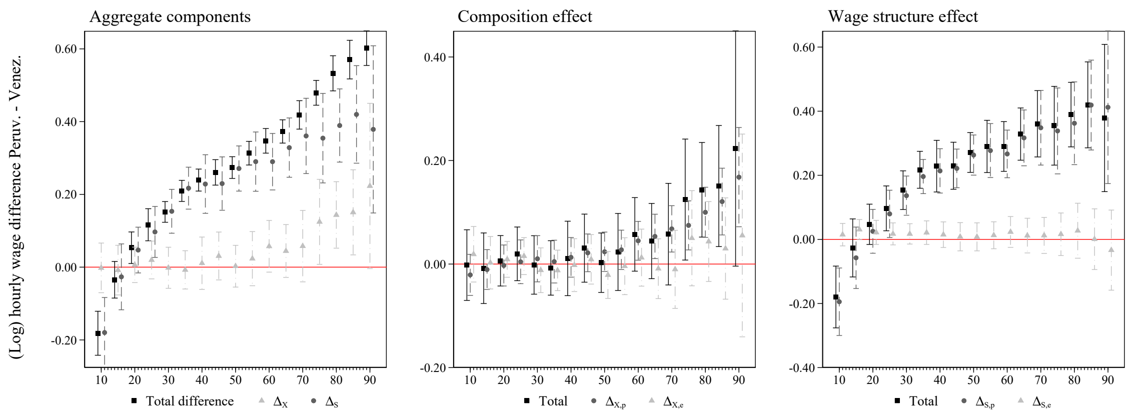
Notes: Sample restricted to only those Peruvians employed in the informal sector between 18 and 65 years in the regions exposed to Venezuelan immigration and without health insurance in their occupation. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. Wholesale and Retail industry includes Hotels and Restaurants and FIRE and Services Industry includes communication, social and personal services. Technical workers occupations includes Skilled agricultural, forestry and fishery workers. Occupational complexity score follows Ottaviano et al. (2013). Department dummies not shown. SEs (in parenthesis) and p-values of the F tests (in brackets) adjustomg for clustering. (d)=dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO (2018) data.

Figure 3.A10 – Observed and counterfactual distributions of (log.) hourly wages for Venezuelan immigrant workers in Peru taking occupational score, 2018



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years and those who arrived to Peru from January 2016 onwards. Informality approximated by the lack of health insurance in their occupation. Simulated distribution based on coefficients from Mincer equations of (log) hourly wages including income from the main and secondary occupations on Peruvian workers (shown in table A6) which include occupational complexity score following Ottaviano et al. (2013). Real hourly wages in 2007 PEN. *Source*: Author's calculations using ENAHO (2018) and ENPOVE data (2018).

Figure 3.A11 – Aggregate Reweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by self-perceived discrimination (occupational complexity score)



Note: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those who arrived to Peru from January 2016 onwards (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph dissagregates each component of the reweighted OB decomposition at selected statistics, with $\Delta_{X,p}$ and $\Delta_{S,p}$ referring to the pure component of the Composition and total Wage structure effects and $\Delta_{X,e}$ and $\Delta_{S,e}$ to the specification and reweighting errors. The outcome model includes occupational dummies, see the text for details on the treatment model and the decomposition. Vertical lines correspond to the bootstrapped SEs (in parenthesis) adjusting for clustering.
Source: Author's calculations using ENAHO and ENPOVE (2018) data.

Table 3.A1 – Adjusted perceived unequal treatment using Mincerian equations for (log) hourly wage for Venezuelan Immigrants by perception of unequal treatment (occupational dummies)

	Did not perceive discrimination						**Did perceive discrimination**					
	Mean	p10	p25	p50	p75	p90	Mean	p10	p25	p50	p75	p90
If is male (d)	-0.020	0.038	-0.003	-0.000	-0.000	-0.070*	-0.027	-0.016	0.022	0.004	-0.043	-0.051
	(0.016)	(0.026)	(0.015)	(0.014)	(0.021)	(0.040)	(0.032)	(0.043)	(0.034)	(0.029)	(0.033)	(0.062)
Age splines (years)												
18-25	0.004	0.013	0.007	0.003	0.004	-0.008	-0.003	-0.012	0.009	0.009	-0.004	-0.031
	(0.005)	(0.009)	(0.006)	(0.005)	(0.005)	(0.010)	(0.009)	(0.013)	(0.009)	(0.010)	(0.011)	(0.020)
26-35	0.002	-0.001	0.001	0.001	0.002	0.006	0.005	0.004	0.006	0.006	0.004	0.015
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.006)	(0.005)	(0.010)	(0.005)	(0.006)	(0.007)	(0.012)
36-45	-0.008*	-0.020***	-0.009**	-0.005	-0.002	0.006	-0.015**	-0.023*	-0.019***	-0.019**	-0.019**	-0.009
	(0.004)	(0.007)	(0.004)	(0.004)	(0.004)	(0.008)	(0.007)	(0.013)	(0.007)	(0.008)	(0.007)	(0.012)
46-55	0.002	0.006	-0.002	-0.000	-0.003	-0.014	0.011	0.027	0.025**	0.010	0.016	0.019
	(0.008)	(0.010)	(0.006)	(0.006)	(0.007)	(0.014)	(0.010)	(0.022)	(0.011)	(0.014)	(0.014)	(0.024)
56-65	0.042**	0.016	0.023*	0.026**	0.053***	0.091**	0.025	-0.026	-0.022	0.006	0.007	0.050
	(0.018)	(0.023)	(0.012)	(0.013)	(0.020)	(0.046)	(0.039)	(0.068)	(0.039)	(0.029)	(0.035)	(0.079)
Education level (base: College)												
Primary	-0.083***	-0.117**	-0.080***	-0.090***	-0.049*	-0.047	-0.195***	-0.136	-0.125**	-0.117*	-0.159**	-0.176
	(0.025)	(0.050)	(0.026)	(0.023)	(0.027)	(0.062)	(0.061)	(0.108)	(0.055)	(0.070)	(0.064)	(0.110)
Secondary	-0.055***	-0.009	-0.044***	-0.061***	-0.060***	-0.084**	-0.047	-0.051	-0.061	-0.031	-0.052	-0.114
	(0.020)	(0.030)	(0.015)	(0.018)	(0.019)	(0.040)	(0.039)	(0.057)	(0.039)	(0.039)	(0.042)	(0.071)
Technical	-0.023	-0.003	-0.005	-0.007	-0.046**	-0.059	0.010	0.007	0.040	-0.006	-0.017	-0.026
	(0.020)	(0.031)	(0.017)	(0.018)	(0.020)	(0.045)	(0.029)	(0.052)	(0.039)	(0.040)	(0.047)	(0.073)
Industry (base: Manufacture)												
Agriculture	0.088	0.021	0.013	-0.006	0.028	0.050	0.827**	0.321**	0.207	0.309**	0.515	1.786**
	(0.152)	(0.155)	(0.122)	(0.113)	(0.142)	(0.286)	(0.361)	(0.149)	(0.132)	(0.151)	(0.345)	(0.807)
Construction	0.248***	0.179***	0.145***	0.237***	0.344***	0.501***	0.420***	0.078	0.119*	0.311***	0.571***	0.974***
	(0.034)	(0.048)	(0.029)	(0.034)	(0.051)	(0.087)	(0.075)	(0.109)	(0.068)	(0.075)	(0.122)	(0.277)
Wholesale and retail	-0.048*	-0.068*	-0.081***	-0.030	-0.044	-0.020	-0.075	-0.166**	-0.089*	-0.066	-0.029	0.070
	(0.027)	(0.039)	(0.026)	(0.031)	(0.041)	(0.069)	(0.056)	(0.076)	(0.048)	(0.061)	(0.080)	(0.112)
Transp., storage, and comm.	-0.113***	-0.194***	-0.136***	-0.062	-0.119**	-0.018	-0.101	-0.287**	-0.200**	-0.185**	0.016	0.186
	(0.043)	(0.064)	(0.043)	(0.041)	(0.054)	(0.094)	(0.081)	(0.115)	(0.084)	(0.085)	(0.101)	(0.156)
FIRE and Services	0.072***	0.046	0.006	0.034	0.028	0.178**	0.066	-0.021	-0.023	0.035	0.081	0.266*
	(0.027)	(0.041)	(0.032)	(0.029)	(0.036)	(0.076)	(0.061)	(0.081)	(0.051)	(0.069)	(0.087)	(0.140)
Occupation (base: Technical)												
Managers and professionals	0.062	-0.002	0.034	0.116**	0.081	-0.075	0.110	0.007	0.098	0.114	0.047	0.107
	(0.079)	(0.069)	(0.053)	(0.055)	(0.084)	(0.211)	(0.136)	(0.077)	(0.070)	(0.120)	(0.197)	(0.378)
Clerical workers	-0.120***	-0.003	0.031	0.003	-0.147***	-0.463***	-0.152	-0.064	-0.060	-0.041	-0.211*	-0.297
	(0.038)	(0.054)	(0.042)	(0.039)	(0.052)	(0.133)	(0.109)	(0.108)	(0.084)	(0.111)	(0.119)	(0.216)
Service and sales workers	-0.188***	-0.067*	-0.073**	-0.130***	-0.229***	-0.411***	-0.150	-0.103	-0.128*	-0.161*	-0.250***	-0.141
	(0.040)	(0.040)	(0.035)	(0.032)	(0.046)	(0.128)	(0.092)	(0.081)	(0.068)	(0.083)	(0.093)	(0.174)
Craft and trades workers	-0.134***	-0.138***	-0.012	0.000	-0.143***	-0.381***	-0.121	-0.126	-0.093	-0.122	-0.144	-0.054
	(0.039)	(0.043)	(0.036)	(0.035)	(0.047)	(0.118)	(0.093)	(0.107)	(0.083)	(0.077)	(0.107)	(0.203)
Machine operators	-0.161***	-0.146**	-0.098**	-0.093**	-0.144**	-0.321**	-0.148	-0.315**	-0.077	-0.048	-0.166	-0.158
	(0.049)	(0.067)	(0.047)	(0.042)	(0.058)	(0.134)	(0.104)	(0.139)	(0.085)	(0.096)	(0.121)	(0.236)
Elementary occupations	-0.157***	-0.111***	-0.042	-0.090***	-0.169***	-0.386***	-0.111	-0.135*	-0.113*	-0.110	-0.153	-0.048
	(0.042)	(0.042)	(0.028)	(0.030)	(0.047)	(0.127)	(0.092)	(0.080)	(0.068)	(0.074)	(0.101)	(0.190)
If employed in Venezuela (d)	0.003	-0.033	-0.011	0.006	0.000	0.001	-0.002	-0.060	-0.044	-0.038	0.058	0.035
	(0.018)	(0.031)	(0.019)	(0.018)	(0.021)	(0.041)	(0.032)	(0.055)	(0.044)	(0.037)	(0.041)	(0.070)
Length of stay in Peru	0.011***	0.011***	0.010***	0.010***	0.011***	0.012***	0.011***	0.004	0.006**	0.010***	0.013***	0.016***
	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)	(0.004)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
Constant	1.116***	0.395*	0.725***	0.998***	1.328***	2.070***	1.194***	1.107***	0.723***	0.909***	1.420***	2.162***
	(0.128)	(0.212)	(0.157)	(0.126)	(0.145)	(0.245)	(0.230)	(0.320)	(0.244)	(0.269)	(0.259)	(0.534)
N	4869	4869	4869	4869	4869	4869	1256	1256	1256	1256	1256	1256
R^2	0.094	0.051	0.077	0.085	0.073	0.043	0.129	0.068	0.074	0.087	0.101	0.080
Model χ^2 test	747.941	190.826	499.402	501.377	369.871	153.507	154.744	80.335	230.596	212.537	101.891	69.833
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Demog. vars χ^2 test	35.474	23.618	39.889	30.411	22.347	20.763	16.601	5.204	23.060	13.932	17.531	10.518
	[0.000]	[0.005]	[0.000]	[0.000]	[0.008]	[0.014]	[0.055]	[0.816]	[0.006]	[0.125]	[0.041]	[0.310]
Industry vars χ^2 test	132.272	53.120	106.465	105.063	63.517	48.970	70.664	21.693	23.204	53.695	29.732	18.318
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]	[0.003]
Occupation vars χ^2 test	37.830	15.529	27.881	63.036	54.983	16.387	8.467	8.613	29.023	11.676	14.517	7.875
	[0.000]	[0.017]	[0.000]	[0.000]	[0.000]	[0.012]	[0.206]	[0.197]	[0.000]	[0.070]	[0.024]	[0.247]

Notes: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years old who arrived after January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The table shows the partial effects on the unconditional distribution of the dependent variable from the model including occupational dummies at selected statistics. College includes also Postgraduate level. Agriculture industry includes also forestry, fishing and Mining and quarrying. Wholesale and Retail industry includes also Hotels and Restaurants and FIRE and Services Industry includes communication, social and personal services. Technical occupations includes Skilled agricultural, forestry and fishery workers. Department dummies not shown. Bootstraped SEs (in parenthesis) and p-values of the χ^2 tests (in brackets) adjusting for clustering. (d)=dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENPOVE (2018) data.

Table 3.A2 – Adjusted perceived unequal treatment using Mincerian equations for (log) hourly wage for Venezuelan Immigrants by perception of unequal treatment (occupational complexity score)

	Did not perceive discrimination						Did perceive discrimination					
	Mean	p10	p25	p50	p75	p90	Mean	p10	p25	p50	p75	p90
If is male (d)	-0.014	0.025	-0.002	0.008	0.011	-0.060	-0.024	-0.029	0.023	0.012	-0.026	-0.040
	(0.015)	(0.025)	(0.013)	(0.014)	(0.018)	(0.044)	(0.032)	(0.044)	(0.034)	(0.029)	(0.034)	(0.061)
Age splines (years)												
18-25	0.004	0.012	0.007	0.004	0.005	-0.006	-0.003	-0.012	0.009	0.009	-0.002	-0.029
	(0.005)	(0.009)	(0.006)	(0.005)	(0.005)	(0.009)	(0.009)	(0.012)	(0.010)	(0.010)	(0.011)	(0.020)
26-35	0.002	-0.002	0.002	0.002	0.003	0.006	0.005	0.003	0.006	0.006	0.004	0.014
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.007)	(0.006)	(0.009)	(0.005)	(0.005)	(0.007)	(0.011)
36-45	-0.008**	-0.021***	-0.010**	-0.005	-0.003	0.006	-0.014*	-0.023*	-0.019***	-0.018***	-0.017**	-0.006
	(0.004)	(0.007)	(0.004)	(0.003)	(0.004)	(0.008)	(0.008)	(0.014)	(0.007)	(0.006)	(0.008)	(0.012)
46-55	0.003	0.005	-0.002	0.001	-0.002	-0.013	0.010	0.025	0.024**	0.009	0.015	0.018
	(0.008)	(0.010)	(0.006)	(0.005)	(0.007)	(0.013)	(0.011)	(0.021)	(0.012)	(0.011)	(0.013)	(0.027)
56-65	0.041**	0.016	0.023*	0.026**	0.052***	0.090*	0.025	-0.025	-0.022	0.004	0.008	0.050
	(0.019)	(0.022)	(0.013)	(0.012)	(0.020)	(0.046)	(0.047)	(0.074)	(0.047)	(0.033)	(0.034)	(0.083)
Education level (base: College)												
Primary	-0.094***	-0.134***	-0.089***	-0.099***	-0.061**	-0.068	-0.213***	-0.163	-0.144**	-0.136**	-0.172***	-0.182*
	(0.026)	(0.051)	(0.029)	(0.025)	(0.030)	(0.070)	(0.060)	(0.105)	(0.060)	(0.066)	(0.052)	(0.107)
Secondary	-0.065***	-0.026	-0.053***	-0.071***	-0.071***	-0.104**	-0.056	-0.071	-0.072*	-0.038	-0.052	-0.112*
	(0.018)	(0.032)	(0.015)	(0.019)	(0.020)	(0.044)	(0.035)	(0.056)	(0.040)	(0.038)	(0.042)	(0.067)
Technical	-0.027	-0.012	-0.008	-0.010	-0.050**	-0.067	0.002	0.000	0.034	-0.011	-0.022	-0.034
	(0.020)	(0.035)	(0.017)	(0.019)	(0.020)	(0.046)	(0.031)	(0.050)	(0.039)	(0.038)	(0.046)	(0.067)
Industry (base: Manufacture)												
Agriculture	0.111	0.063	0.005	-0.038	0.055	0.137	0.838**	0.301**	0.205	0.332**	0.525	1.789**
	(0.164)	(0.151)	(0.131)	(0.100)	(0.158)	(0.271)	(0.339)	(0.144)	(0.128)	(0.148)	(0.327)	(0.773)
Construction	0.250***	0.193***	0.147***	0.221***	0.344***	0.505***	0.427***	0.100	0.114*	0.310***	0.572***	0.984***
	(0.029)	(0.047)	(0.030)	(0.030)	(0.052)	(0.092)	(0.072)	(0.083)	(0.067)	(0.073)	(0.123)	(0.252)
Wholesale and retail	-0.079***	-0.033	-0.101***	-0.097***	-0.091***	-0.053	-0.081*	-0.130**	-0.106**	-0.089*	-0.079	0.048
	(0.021)	(0.041)	(0.026)	(0.025)	(0.033)	(0.050)	(0.045)	(0.052)	(0.047)	(0.052)	(0.069)	(0.086)
Transp., storage, and comm.	-0.116***	-0.171**	-0.160***	-0.099***	-0.112***	0.018	-0.105	-0.352***	-0.183***	-0.142**	0.006	0.135
	(0.038)	(0.070)	(0.040)	(0.038)	(0.042)	(0.075)	(0.065)	(0.120)	(0.071)	(0.072)	(0.087)	(0.119)
FIRE and Services	0.078***	0.098**	0.009	0.004	0.028	0.202***	0.090*	0.036	-0.009	0.051	0.070	0.258**
	(0.025)	(0.042)	(0.029)	(0.029)	(0.033)	(0.063)	(0.052)	(0.059)	(0.048)	(0.060)	(0.078)	(0.109)
Occup. complexity score	0.007*	-0.003	-0.003	0.002	0.007***	0.010	-0.004	-0.012	-0.006	0.001	0.005	-0.003
	(0.004)	(0.002)	(0.002)	(0.001)	(0.003)	(0.007)	(0.007)	(0.014)	(0.008)	(0.005)	(0.007)	(0.010)
If employed in Venezuela (d)	0.005	-0.032	-0.009	0.010	0.002	0.003	0.002	-0.058	-0.040	-0.035	0.061	0.041
	(0.018)	(0.027)	(0.020)	(0.019)	(0.022)	(0.041)	(0.029)	(0.051)	(0.043)	(0.040)	(0.042)	(0.066)
Length of stay in Peru	0.011***	0.012***	0.010***	0.010***	0.012***	0.012***	0.011***	0.004	0.006**	0.010***	0.013***	0.015***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.004)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
Constant	0.951***	0.312	0.697***	0.940***	1.140***	1.667***	1.056***	0.987***	0.631***	0.790***	1.210***	2.010***
	(0.127)	(0.231)	(0.136)	(0.123)	(0.131)	(0.209)	(0.207)	(0.285)	(0.233)	(0.233)	(0.267)	(0.483)
N	4869	4869	4869	4869	4869	4869	1256	1256	1256	1256	1256	1256
R^2	0.087	0.048	0.074	0.075	0.065	0.036	0.123	0.066	0.071	0.079	0.092	0.076
Model χ^2 test	559.155	180.639	423.036	368.532	236.929	153.261	139.023	41.442	95.199	147.013	102.519	64.143
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.005]	[0.000]	[0.000]	[0.000]	[0.000]
Demog. vars χ^2 test	40.979	29.515	54.002	30.476	28.642	19.064	18.100	8.518	21.499	18.718	17.287	10.380
	[0.000]	[0.001]	[0.000]	[0.000]	[0.001]	[0.025]	[0.034]	[0.483]	[0.011]	[0.028]	[0.044]	[0.321]
Industry vars χ^2 test	184.416	67.361	113.844	141.074	73.260	59.121	76.603	27.542	37.712	78.147	37.611	23.063
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Occupation vars χ^2 test	2.753	1.343	1.628	1.814	7.960	2.320	0.269	0.786	0.544	0.037	0.511	0.115
	[0.097]	[0.246]	[0.202]	[0.178]	[0.005]	[0.128]	[0.604]	[0.375]	[0.461]	[0.847]	[0.475]	[0.735]

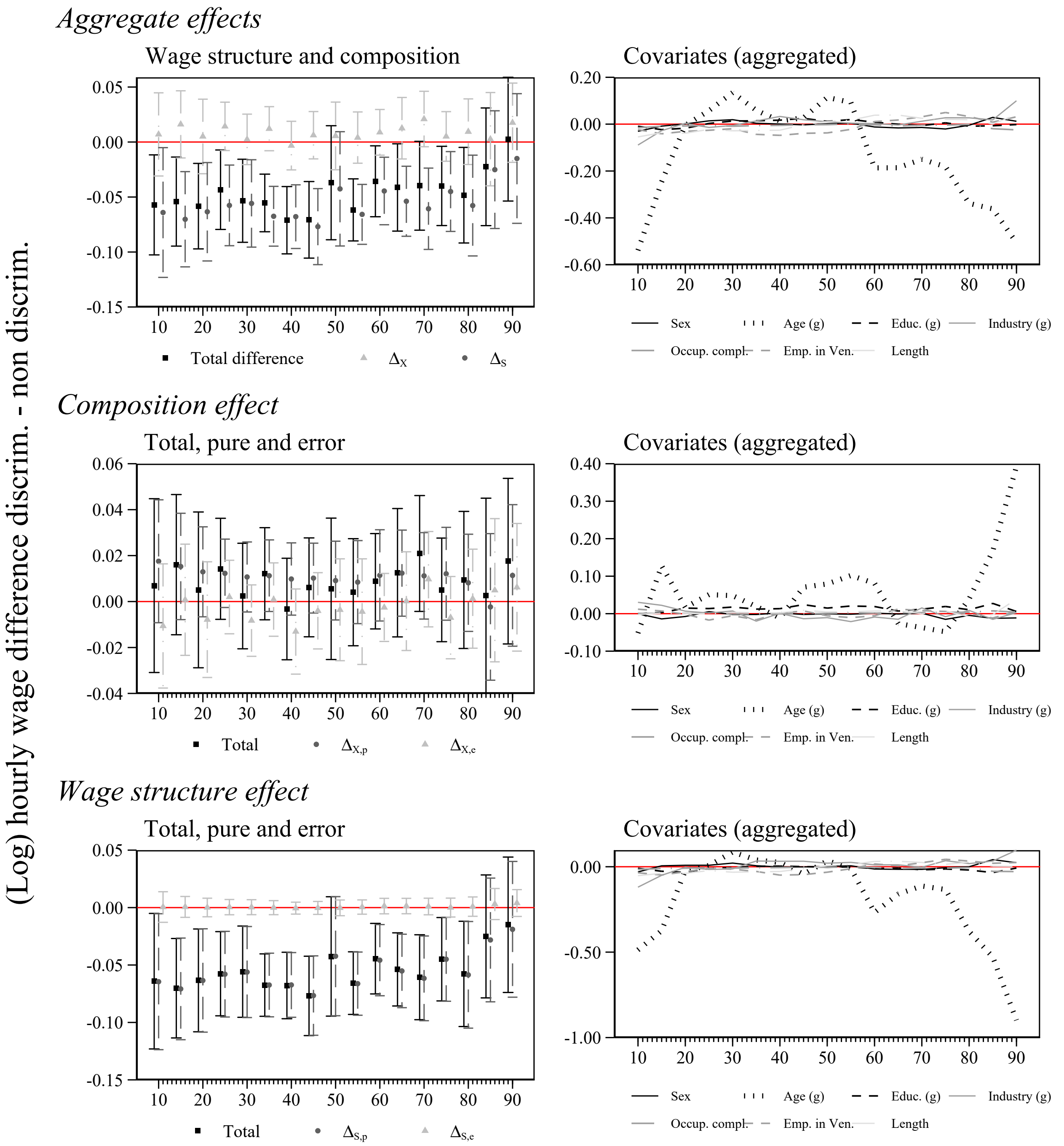
Notes: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years old who arrived after January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The table shows the partial effects on the unconditional distribution of the dependent variable from the model including occupational complexity score (following Ottaviano et al. 2013) at selected statistics. College includes also Postgraduate level. Agriculture industry includes also forestry, fishing and Mining and quarrying. Wholesale and Retail industry includes also Hotels and Restaurants and FIRE and Services Industry includes communication, social and personal services. Department dummies not shown. Bootstraped SEs (in parenthesis) and p-values of the χ^2 tests (in brackets) adjusting for clustering. (d)=dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENPOVE (2018) data.

Table 3.A3 – Detailed Unweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by perceived unequal treatment (occupational dummies)

	Mean		p10		p25		p50		p75		p90	
Overall												
Discriminated	1.032***	(0.017)	0.515***	(0.022)	0.772***	(0.020)	1.007***	(0.024)	1.271***	(0.016)	1.632***	(0.031)
Non discriminated	1.089***	(0.011)	0.572***	(0.018)	0.815***	(0.012)	1.044***	(0.010)	1.311***	(0.013)	1.629***	(0.021)
Obs. difference	-0.057***	(0.014)	-0.057**	(0.023)	-0.043**	(0.018)	-0.037	(0.025)	-0.040**	(0.019)	0.003	(0.030)
Total explained	0.010	(0.010)	0.011	(0.013)	0.008	(0.008)	0.005	(0.009)	0.008	(0.011)	0.012	(0.017)
Total unexplained	-0.067***	(0.015)	-0.068***	(0.027)	-0.051***	(0.019)	-0.042*	(0.023)	-0.048***	(0.018)	-0.010	(0.029)
Explained Component												
Sex	0.001	(0.001)	0.001	(0.002)	-0.001	(0.001)	-0.000	(0.001)	0.002	(0.001)	0.002	(0.002)
Age (g)	-0.000	(0.002)	-0.000	(0.003)	-0.000	(0.002)	-0.000	(0.002)	-0.000	(0.002)	-0.001	(0.003)
Education (g)	0.007**	(0.003)	0.005	(0.004)	0.006**	(0.003)	0.004*	(0.003)	0.006**	(0.003)	0.008*	(0.004)
Industry (g)	-0.011**	(0.005)	-0.008*	(0.004)	-0.007**	(0.003)	-0.010**	(0.004)	-0.012*	(0.006)	-0.015	(0.013)
Occupation (g)	-0.003	(0.003)	-0.004	(0.003)	-0.003	(0.002)	-0.003	(0.003)	-0.006*	(0.003)	-0.003	(0.005)
Employed in Venez.	-0.000	(0.001)	-0.002	(0.002)	-0.001	(0.002)	-0.001	(0.001)	0.002	(0.001)	0.001	(0.003)
Length in Peru	0.010**	(0.004)	0.004	(0.005)	0.006*	(0.003)	0.010***	(0.003)	0.012***	(0.004)	0.015***	(0.005)
Region	0.007*	(0.004)	0.016*	(0.008)	0.008*	(0.005)	0.006	(0.004)	0.004	(0.004)	0.006	(0.005)
Unexplained Component												
Sex	-0.004	(0.021)	-0.031	(0.029)	0.014	(0.020)	0.002	(0.019)	-0.025	(0.023)	0.010	(0.044)
Age (g)	-0.164	(0.221)	-0.563	(0.368)	0.056	(0.236)	0.126	(0.249)	-0.194	(0.265)	-0.549	(0.483)
Education (g)	-0.003	(0.021)	-0.014	(0.031)	-0.001	(0.022)	0.007	(0.025)	-0.003	(0.023)	-0.018	(0.047)
Industry (g)	0.000	(0.046)	-0.077	(0.065)	-0.014	(0.048)	-0.020	(0.063)	0.045	(0.063)	0.122	(0.106)
Occupation (g)	0.030	(0.084)	-0.033	(0.084)	-0.057	(0.062)	-0.032	(0.071)	-0.005	(0.093)	0.267	(0.164)
Employed in Venez.	-0.004	(0.026)	-0.022	(0.046)	-0.026	(0.037)	-0.036	(0.034)	0.046	(0.034)	0.027	(0.056)
Length in Peru	-0.002	(0.025)	-0.052	(0.038)	-0.027	(0.025)	0.003	(0.023)	0.011	(0.027)	0.031	(0.049)
Region	0.002	(0.015)	0.016	(0.028)	0.005	(0.019)	-0.006	(0.013)	-0.014	(0.019)	-0.002	(0.032)
Constant	0.078	(0.254)	0.708*	(0.398)	-0.001	(0.270)	-0.087	(0.289)	0.092	(0.306)	0.102	(0.536)

Notes: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years old who arrived after January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The table show the components of the detailed decomposition, without reweighting, at selected statistics of the unconditional distribution of the dependent variable using coefficients from RIF regression. The outcome model includes occupational dummies, see the text for details on the treatment model. (g) means that the component of the detailed decomposition refers to the effect after grouping the corresponding dummies (see table 2 for the categories of the control variables). Bootstraped SEs (in parenthesis) adjusting for clustering. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENPOVE (2018) data.

Figure 3.A5 – Aggregate Reweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by self-perceived discrimination (occupational complexity score)



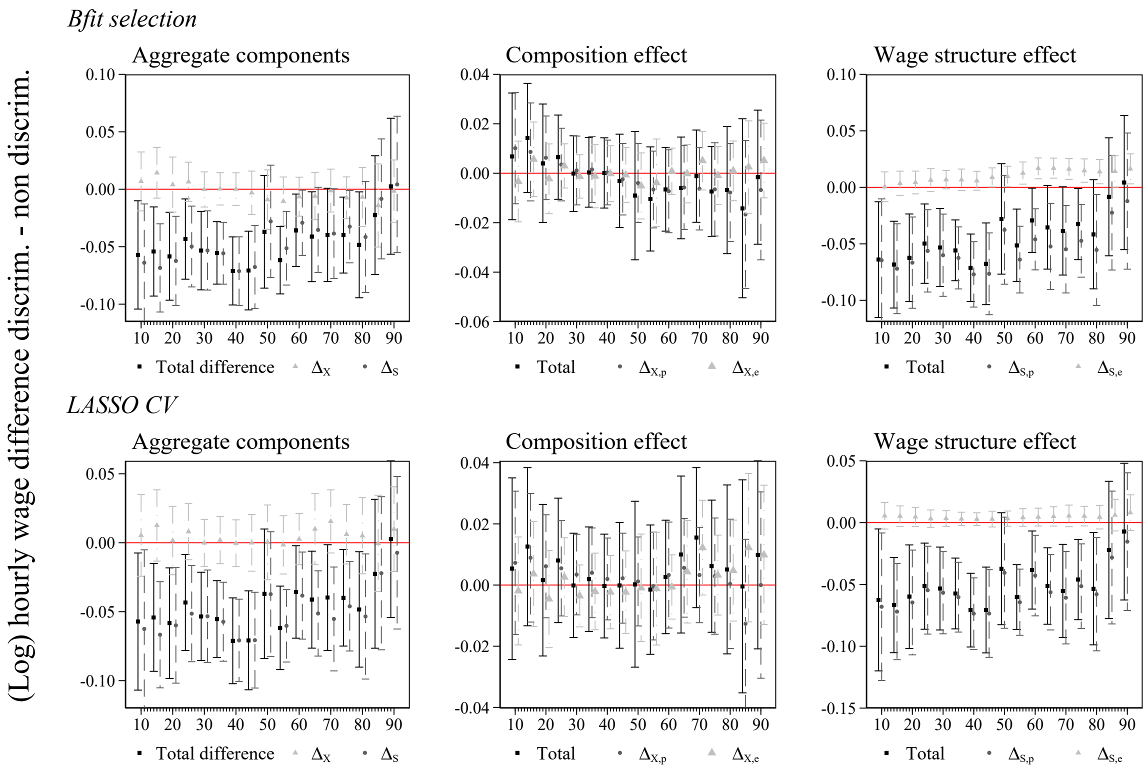
Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages inclu approximated by the lack of health insurance in their occupation. The graph shows the components of the reweighted OB decomposition at selected statistics, with Δ_X and Δ_S b The outcome model includes occupational complexity score (following Ottaviano et al. 2013), see the text for details on the treatment model and the decomposition. (g) means to the effect after grouping the corresponding dummies (see table 2 for the categories of the control variables). Vertical lines correspond to the 95% bootstrapped confidence in
Source: Author's calculations using ENPOVE (2018) data.

Table 3.A4 – Detailed Reweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by perceived unequal treatment (occupational complexity score)

	Mean		p10		p25		p50		p75		p90	
Overall												
Discriminated	1.032***	(0.016)	0.515***	(0.023)	0.772***	(0.019)	1.007***	(0.027)	1.271***	(0.016)	1.632***	(0.031)
Non discriminated	1.089***	(0.011)	0.572***	(0.018)	0.815***	(0.012)	1.044***	(0.009)	1.311***	(0.012)	1.629***	(0.020)
Disc. rwgt. as non	1.023***	(0.016)	0.508***	(0.031)	0.758***	(0.019)	1.002***	(0.026)	1.266***	(0.016)	1.614***	(0.031)
Obs. difference	-0.057***	(0.014)	-0.057**	(0.023)	-0.043**	(0.018)	-0.037	(0.026)	-0.040**	(0.018)	0.003	(0.029)
Total explained	0.009	(0.011)	0.007	(0.019)	0.014	(0.011)	0.006	(0.016)	0.005	(0.012)	0.018	(0.018)
Total unexplained	-0.066***	(0.014)	-0.064**	(0.030)	-0.058***	(0.019)	-0.043	(0.027)	-0.045**	(0.019)	-0.015	(0.030)
Explained Component												
Specification error	-0.004	(0.007)	-0.011	(0.014)	0.002	(0.008)	-0.004	(0.011)	-0.007	(0.009)	0.006	(0.014)
Pure explained	0.014	(0.010)	0.018	(0.014)	0.012	(0.008)	0.009	(0.009)	0.012	(0.010)	0.011	(0.016)
Sex	0.001	(0.001)	0.001	(0.002)	-0.001	(0.002)	-0.000	(0.001)	0.001	(0.001)	0.002	(0.003)
Age (g)	0.000	(0.002)	0.000	(0.003)	0.001	(0.002)	0.000	(0.002)	0.000	(0.002)	-0.002	(0.003)
Education (g)	0.008**	(0.003)	0.006	(0.005)	0.006**	(0.003)	0.005*	(0.003)	0.006*	(0.003)	0.007	(0.005)
Industry (g)	-0.011*	(0.006)	-0.009	(0.005)	-0.007*	(0.004)	-0.010*	(0.005)	-0.013*	(0.007)	-0.017	(0.013)
Occupation	0.000	(0.001)	0.001	(0.002)	0.000	(0.001)	-0.000	(0.001)	-0.000	(0.001)	0.000	(0.001)
Employed in Venez.	0.000	(0.001)	-0.001	(0.002)	-0.001	(0.001)	-0.001	(0.001)	0.002	(0.001)	0.001	(0.002)
Length in Peru	0.010***	(0.004)	0.004	(0.005)	0.006*	(0.003)	0.010***	(0.003)	0.012***	(0.004)	0.014***	(0.005)
Region	0.006	(0.004)	0.015	(0.009)	0.008*	(0.005)	0.005	(0.004)	0.005	(0.004)	0.006	(0.005)
Unexplained Component												
Reweighting error	0.001	(0.004)	0.000	(0.007)	0.000	(0.003)	-0.000	(0.003)	0.000	(0.004)	0.004	(0.006)
Pure unexplained	-0.067***	(0.014)	-0.065**	(0.030)	-0.058***	(0.019)	-0.042	(0.026)	-0.045**	(0.019)	-0.019	(0.030)
Sex	-0.004	(0.020)	-0.030	(0.030)	0.009	(0.024)	0.002	(0.019)	-0.005	(0.022)	0.024	(0.040)
Age (g)	-0.238	(0.226)	-0.486	(0.387)	0.008	(0.266)	0.036	(0.260)	-0.133	(0.274)	-0.900*	(0.545)
Education (g)	-0.013	(0.022)	-0.009	(0.034)	-0.011	(0.025)	-0.003	(0.024)	-0.013	(0.024)	-0.007	(0.049)
Industry (g)	0.014	(0.042)	-0.119**	(0.056)	-0.015	(0.044)	0.021	(0.049)	0.035	(0.061)	0.097	(0.074)
Occupation	-0.020	(0.016)	-0.017	(0.025)	-0.006	(0.014)	-0.003	(0.011)	-0.008	(0.019)	-0.029	(0.027)
Employed in Venez.	0.005	(0.027)	-0.035	(0.050)	-0.009	(0.043)	-0.032	(0.035)	0.043	(0.037)	0.024	(0.071)
Length in Peru	-0.001	(0.024)	-0.052	(0.044)	-0.022	(0.027)	0.006	(0.021)	0.014	(0.028)	0.002	(0.054)
Region	0.007	(0.016)	0.028	(0.030)	0.008	(0.019)	0.001	(0.014)	-0.010	(0.018)	0.003	(0.030)
Constant	0.182	(0.240)	0.656*	(0.395)	-0.019	(0.279)	-0.071	(0.281)	0.031	(0.294)	0.768	(0.543)

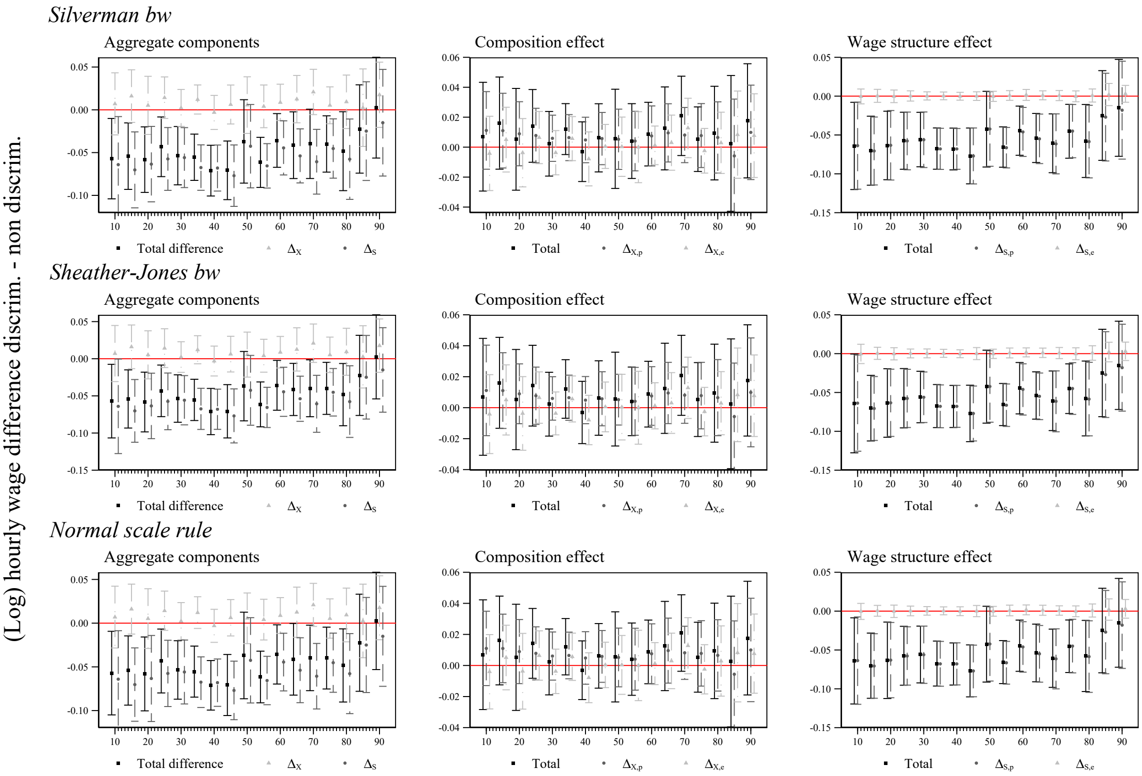
Notes: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years old who arrived after January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The table show the components of the detailed decomposition, reweighting those who perceived discrimination as if they did not, at selected statistics of the unconditional distribution of the dependent variable using coefficients from RIF regression. The outcome model includes occupational complexity score (following Ottaviano et al. 2013), see the text for details on the treatment model. (g) means that the component of the detailed decomposition refers to the effect after grouping the corresponding dummies (see table 2 for the categories of the control variables). Bootstraped SEs (in parenthesis) adjust for clustering. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENPOVE (2018) data.

Figure 3.A6 – Aggregate Reweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by self-perceived discrimination under different treatment models



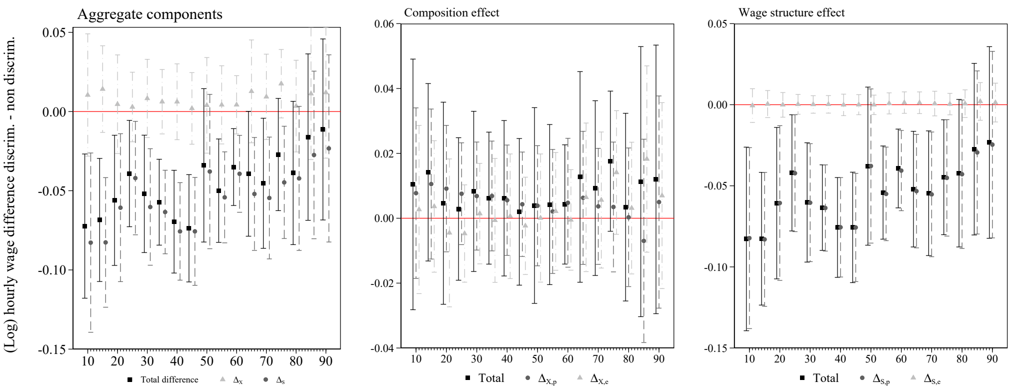
Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows the components of the reweighted OB decomposition at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the treatment models and the decomposition procedure.
Vertical lines correspond to the bootstrapped SEs adjusting for clustering. *Source*: Author's calculations using ENPOVE (2018) data.

Figure 3.A7 – Aggregate Reweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by self-perceived discrimination under different bandwidths



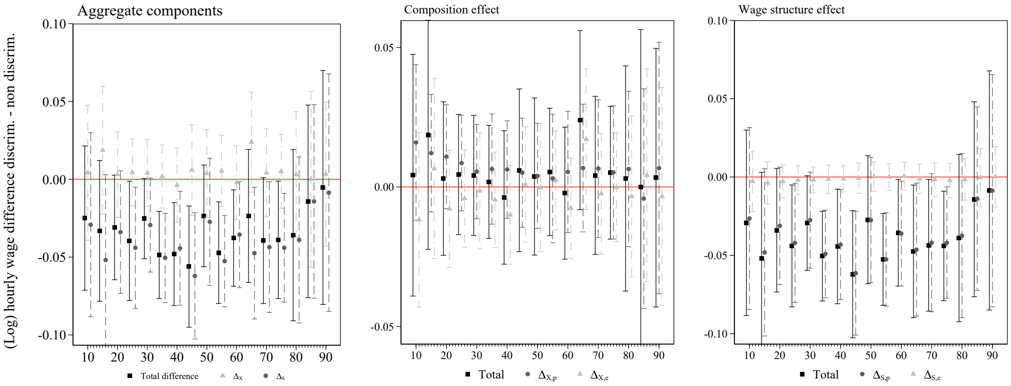
Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows the components of the reweighted OB decomposition at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the treatment model and the decomposition. Vertical lines correspond to the bootstrapped SEs adjusting for clustering.
Source: Author's calculations using ENPOVE (2018) data.

Figure 3.A8 – Aggregate Reweighted OB decomposition of (log) hourly wages (from the main occupation only) for Venezuelan immigrants by self-perceived discrimination



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages include income from the main occupation only. Informality approximated by the lack of health insurance in their occupation. The graph shows the components of the reweighted OB decomposition at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the treatment model and the decomposition. Vertical lines correspond to the bootstrapped SEs adjusting for clustering. *Source*: Author's calculations using ENPOVE (2018) data.

Figure 3.A9 – Aggregate Reweighted OB decomposition of (log) hourly wages for Venezuelan immigrants by self-perceived discrimination under restricted definition of employment



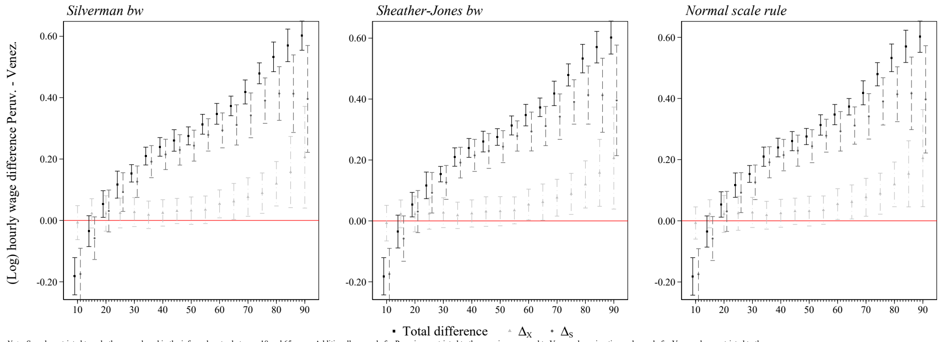
Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows the components of the reweighted OB decomposition at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the treatment model and the decomposition. Vertical lines correspond to the bootstrapped SEs adjusting for clustering. *Source*: Author's calculations using ENPOVE (2018) data.

Table 3.A6 – Impact of wage structure effect on perception of discrimination of Venezuelan immigrant workers at selected statistics (marginal effects)

	Peruvian Mincer model using occup. dummies						**Peruvian Mincer model using occup. complexity**					
	Mean	p10	p25	p50	p75	p90	Mean	p10	p25	p50	p75	p90
Δ_S (wage structure)	0.040***	0.016**	0.031***	0.041***	0.021**	0.000	0.042***	0.016**	0.031***	0.042***	0.023***	0.001
	(0.010)	(0.007)	(0.010)	(0.011)	(0.008)	(0.005)	(0.010)	(0.007)	(0.010)	(0.011)	(0.008)	(0.005)
If the individual is male (d)	-0.040***	-0.037***	-0.038***	-0.038***	-0.035***	-0.031***	-0.041***	-0.037***	-0.037***	-0.039***	-0.036***	-0.032***
	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Age splines (years)												
18-25	-0.006	-0.005	-0.006	-0.006	-0.005	-0.005	-0.006	-0.005	-0.006	-0.006	-0.006	-0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
26-35	-0.000	-0.000	0.000	0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
36-45	-0.000	0.000	-0.000	-0.000	0.000	0.000	-0.000	0.000	-0.000	-0.000	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
46-55	-0.001	-0.002	-0.001	-0.001	-0.002	-0.002	-0.001	-0.002	-0.001	-0.001	-0.002	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
56-65	-0.002	-0.003	-0.003	-0.003	-0.003	-0.004	-0.002	-0.003	-0.003	-0.003	-0.003	-0.004
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Education level (base: primary)												
Secondary educ.	0.031	0.033	0.032	0.033	0.033	0.035*	0.031	0.033	0.032	0.033	0.033	0.035*
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Technical educ.	0.045**	0.047**	0.048**	0.047**	0.048**	0.053***	0.044**	0.048**	0.048**	0.047**	0.046**	0.052***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
College and PG educ.	0.048**	0.055***	0.053***	0.054***	0.053***	0.060***	0.046**	0.055***	0.053***	0.052***	0.051***	0.060***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.018)
Occup. complexity score	-0.005	-0.004*	-0.005*	-0.004*	-0.004	-0.004	-0.005*	-0.004*	-0.005*	-0.005*	-0.005*	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Time in Peru	0.006***	0.006***	0.006***	0.006***	0.006***	0.005***	0.006***	0.006***	0.006***	0.006***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
If employed in Venezuela (d)	0.040***	0.039***	0.039***	0.039***	0.040***	0.040***	0.040***	0.039***	0.039***	0.039***	0.040***	0.040***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
N	6125	6125	6125	6125	6125	6125	6125	6125	6125	6125	6125	6125
R2	0.013	0.012	0.012	0.013	0.012	0.011	0.013	0.012	0.012	0.013	0.012	0.011
Model χ^2 test	66.970	65.028	66.471	68.324	64.544	64.589	67.716	64.897	66.393	68.578	65.222	64.515
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Normality test	2.643	2.928	3.000	3.085	3.684	3.952	2.363	3.052	2.910	2.958	3.366	3.961
	[0.267]	[0.231]	[0.223]	[0.214]	[0.159]	[0.139]	[0.307]	[0.217]	[0.233]	[0.228]	[0.186]	[0.138]
Heterosk. test	17.822	21.457	22.043	17.141	19.755	17.799	17.425	22.169	22.964	16.584	18.517	17.890
	[0.164]	[0.064]	[0.055]	[0.193]	[0.101]	[0.165]	[0.181]	[0.053]	[0.042]	[0.219]	[0.139]	[0.162]
RESET test	4.139	3.346	3.943	3.231	5.811	4.364	4.139	3.342	3.711	3.214	6.102	4.409
	[0.247]	[0.341]	[0.268]	[0.357]	[0.121]	[0.225]	[0.247]	[0.342]	[0.294]	[0.360]	[0.107]	[0.221]

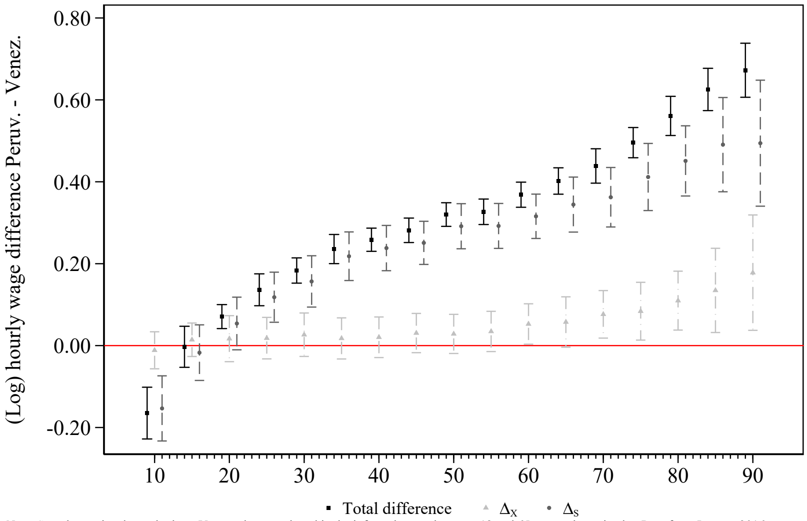
Notes: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those regions who arrived to Peru from January 2016 onwards. (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. Δ_S refers to the estimate of the mean wage structure effect for the Venezuelan workers, estimated as the difference between the simulated log hourly wage (using coefficients from the Mincer regression on log hourly wages on the Peruvian sample, shown in table A6) and the RIF of the corresponding statistic of the observed (log) hourly wage of Venezuelans. SEs (in parenthesis) and p-values of the tests (in brackets) adjusting for clustering. (d)=dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. *Source*: Author's calculations using ENAHO (2018) and ENPOVE (2018) data.

Figure 3.A12 – Aggregate Unweighted OB decomposition of (log) hourly wages by Peruvians and Venezuelan workers under different bandwidths



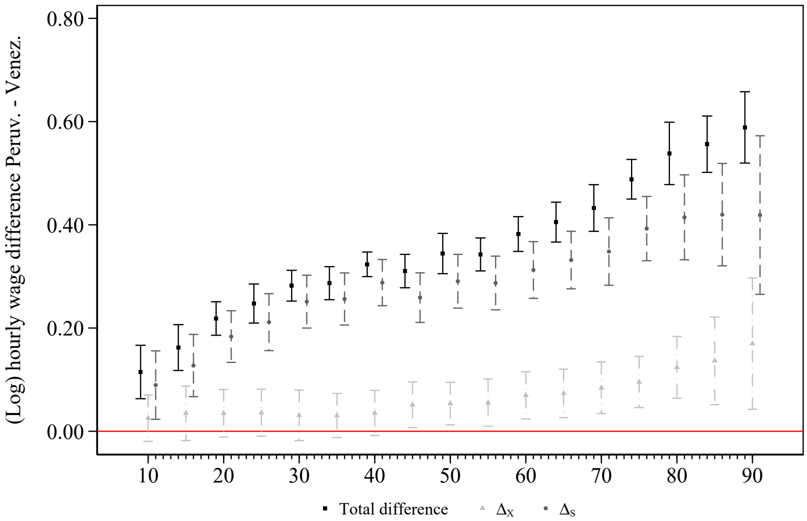
Note: Sample restricted to only those employed in the informal sector between 18 and 65 years. Additionally, sample for Peruvians restricted to those regions exposed to Venezuelan migration and sample for Venezuelans restricted to those who arrived to Peru from January 2016 onwards (Log) hourly wages include income from the main and secondary occupations. Informality approximated by the lack of health insurance in their occupation. The graph shows the components of the unweighted OB decomposition at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the decomposition. Vertical lines correspond to the bootstrapped SEs (in parenthesis) adjusting for clustering. *Source*: Author's calculations using ENAHO and ENPOVE (2018) data.

Figure 3.A13 – Aggregate Unweighted OB decomposition of (log) hourly wages (from the main occupation only) by Peruvians and Venezuelan workers under different bandwidths



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages include income from the main occupation only. Informality approximated by the lack of health insurance in their occupation. The graph shows the components of the reweighted OB decomposition at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the decomposition. Vertical lines correspond to the bootstrapped SEs adjusting for clustering. *Source*: Author's calculations using ENAHO and ENPOVE (2018) data.

Figure 3.A14 – Aggregate Reweighted OB decomposition of (log) hourly wages by Peruvians and Venezuelan under restricted definition of employment



Note: Sample restricted to only those Venezuelans employed in the informal sector between 18 and 65 years who arrived to Peru from January 2016. (Log) hourly wages include income from the main occupation only. Informality approximated by the lack of health insurance in their occupation. The graph shows the components of the reweighted OB decomposition at selected statistics, with Δ_X and Δ_S being the total Composition and total Wage structure effects. The outcome model includes occupational dummies, see the text for details on the decomposition. Vertical lines correspond to the bootstrapped SEs adjusting for clustering. *Source*: Author's calculations using ENAHO and ENPOVE (2018) data.

Concluding remarks

This thesis sheds light on three issues relevant to the Peruvian labour market over the last two decades. The confluence of two unique phenomena characterizes this period and provide a backdrop against which the analysis is set. First, the sustained growth in per capita income (GDP), which became known as the 'Peruvian Growth Miracle'. Peru's economic growth rate between 2001 and 2019 reached a historical peak and reflected a doubling in the average per capita GDP growth rates compared to Latin American countries. This period witnessed a general increase in real wages. Nonetheless, throughout this period, the informal sector remained the most dominant labour market segment, and is sizeable compared to the other countries in the region. The second phenomenon during those years was the Venezuelan Exodus, which surged as a consequence of the deterioration in the socio-economic conditions in Venezuela. The sizeable influx of workers who arrived in Peru, traditionally one of the countries with the lowest proportion of immigrants in its labour market, transformed the country into the second-largest recipient of Venezuelan migrants and refugees. The distinctive nature of this Exodus compared to similar episodes in other countries rests on a number of distinct factors. First, the significant proportion of Peru's population these Venezuelan migrants comprised, 2.5%, is comparable to what the Syrian immigrants represented in Turkey. Second, the schooling levels of these migrants are higher than the average of the natives. Third, in contrast to other similar episodes in the recent past, both natives and migrants share a common language and have close cultural ties.

The first empirical chapter examines if the exceptional macroeconomic growth era in Peru was accompanied by a more equitable treatment for female workers across the unconditional pay distribution. We find that males maintained a wage advantage across the whole pay distribution throughout the selected years. These gender disparities were found to be even deeper in the informal sector. Notably, these gaps are primarily attributed to the differential treatment females experienced in the Peruvian labour market and not because they have less attractive observed characteristics than their male counterparts. This chapter's most policy-relevant finding is the existence and persistence of 'sticky floors' in the informal sector. That is, women at the bottom of the wage distribution experience a double penalty. On the one hand, their low wages and the lack of social protection that characterizes jobs in this sector (such as minimum wages, health and safety regulations, and access to pension systems) render them particularly vulnerable to situations of entering poverty. On the other hand, their already low wages are lower than those of males with similar labour market attributes.

The other two chapters study two distinct aspects of the Exodus. The first aspect concerns the impact of this large inflow (or labour supply shock) on a set of labour market outcomes for Peruvians. Using a combination of different quasi-experimental methods, we find that this shock did indeed affect the different outcomes analysed. More specifically, this exogenous increase in labour supply did not lead to statistically significant changes overall in hourly wages for the natives in the formal and the informal sectors. The Exodus also had non-statistically significant effects on the size of the informal sector and, for the most part, on the wage inequality for both formal and informal sectors.

However, a non-negligible effect is found on the wages for those low skilled informal workers in a particular metropolitan area of Peru. This subgroup of workers residing in the capital city of Lima, and who are most likely to compete with Venezuelan workers directly, experienced a reduction in their wages of around 10%. The policy-relevant issue, in this case, relates to the adverse effect that workers in this sector of the economy experienced. This result might give support to those that regard immigration as damaging for natives' wages. Nonetheless, as explained in the conclusions to chapter two, the relative size of the increase in the labour supply labour is larger than the effect on wages in this city. Specifically, a 20% increase in labour supply reduced the wage for informal unskilled workers by 10%, suggesting an inelastic response. Thus, it is arguable that the impact of this Exodus on the wages of the unskilled natives is moderately low.

The final chapter is concerned with the effect of discrimination on the wages of Venezuelan immigrants and the role that the wage inequity they experience exerts on their perception of workplace discrimination. We find a migrant pay gap across the wage distribution. The wage gap increases in the upper part of the wages distribution, and the gap is largely explained by a treatment effect against the Venezuelan workers. The treatment effects at different percentiles are statistically significant determinants of a Venezuelan migrant's probability of perceiving workplace discrimination. Nonetheless, its effect is negligible compared to other variables that reflect an expectation of equal treatment. These include education, previous work experience, or time spent in Peru. In this case, the most critical policy issue pertains to the existence of wage gaps that harm migrants in the informal sector, most of which cannot be explained by differences in the levels of endowments between Peruvians and Venezuelans. This evidence, along with the natives' generalized prejudice towards Venezuelan migrants, plausibly suggests the prevalence of this treatment effect against migrants has its roots in discrimination. In fact, as most of these migrants are also employed in the informal labour market, there are no effective mechanisms that ensure fair payment for the work they undertake or that permit them to report abuse in the workplace. As studies reviewed suggest, this can have a detrimental effect on their psychological wellness, mainly because they are aware of and directly experience these inequities. In addition, Venezuelans experience occupational downgrading and are confined to lower occupational categories, which results in the underutilization of their human capital. The inevitable implication of this is low monthly wages, as almost 60% of migrants earn wages below the minimum legal monthly wage. Furthermore, the long hours of weekly work that these migrants engage in can negatively affect their health and productivity.

On the basis of these conclusions, we now discuss some policy recommendations. The first chapter provides evidence that macroeconomic growth cannot be expected to automatically reduce inequalities in the labour market, such as those gender-related ones. In fact, this seems to be what the Peruvian government anticipated, as during this period it primarily focused on nominally improving the legislation promoting gender equality. Nonetheless, due to the lack of implementation of comprehensive policies to address the problem of unequal treatment of female workers, Peru remains at the bottom of international rankings measuring labour market quality (WEF 2020). In the face of this, some strategies can be pursued to ameliorate partially this problem. Due to the relatively high percentage of Peruvian women in the labour force, a share larger than other Latin American countries (ILO 2018a), an essential part of these strategies should focus on human capital accumulation and ensuring their access to high-quality jobs. Policies like these imply promoting female incorporation into higher-paid occupations and careers, emphasising areas with strong labour demand. An alternative strategy consists of strengthening the ability of the Labour Inspection Authority to enforce Law No 30709 prohibiting gender wage discrimination. Further steps would involve, for example, promoting access to credit for women who own small and medium enterprises (Morrison 2021). Some other experiences put in place by countries in the Latin America region, such as extending the coverage of minimum wages to domestic workers, done in Brazil, and combining collective regulation with social security, as Uruguay put in place, can also be helpful (ILO 2016).

On the basis of the third chapter, a way that Venezuelans' low wages can be increased is by creating incentives to migrate to occupations that match their professional skills. As emphasized above, this is a migrant population with higher education levels than the Peruvians, and yet are in jobs where they experience a downgrade, requiring mostly manual skills instead of cognitive and analytical abilities. An alternative way to increase their wages would be through re-orienting their abilities to match those required in their current occupational sector. Nonetheless, this is more likely to occur if they integrate into the formal labour market, as employers in this sector are more sensitive to the regulations that would foster equal treatment of migrants compared to natives. Further, there is the need to tackle the unexplained part of the migrant pay gaps, which finds more traction in the formal sector. This includes implementing the principle of "equal remuneration for work of equal value" and, importantly, fostering their rights to a fair payment (ILO 2020). This transition towards the formal labour market, however, cannot be seen as a magic bullet. This is because the relatively small size of the formal sector in Peru implies that only a small proportion of workers will be able to transit to this sector.

Nevertheless, the scope of what these types of policies can influence to achieve better results is constrained due to the historically large size of Peru's informal sector. Hence, a first-order policy that this thesis suggests is the integration of the informal and the formal sectors to achieve a reduction of the size of the former sector. This is the fundamental problem that Peruvian policy-makers need to address as a matter of urgency in the coming years. A crucial finding in this thesis is that workers in the informal labour market are more vulnerable to lower wages and to experiencing discrimination based on gender and nationality. In fact, the effects of a large informal sector resonate beyond the labour market itself, as it has important repercussions for reducing inequality and poverty. The labour market ultimately decides who are to be the winners and losers, with the latter group vulnerable to poverty. Specifically, a more significant informal sector is positively associated with higher inequality and poverty (Messina and Silva 2018) and lower productivity (IADB 2010). Simultaneously, economic growth is associated with a more rapid decline in self-employment, which is the most prevalent occupational category in the informal sector in Peru (La Porta and Shleifer 2014). Hence, as the large informal sector limits what can be done, we cannot comprehensively foster labour protection policies unless the formal sector assumes a role of primacy within the country. For migrants and females, the discrimination they experience is likely to be significantly reduced if they were in the more protected and regulated formal labour market. The effectiveness of the policies related to higher transparency and more vigorous legal enforcement to tackle the problems mentioned above would be more effective in the formal labour market. Certainly, Peru's historically weak governance and institutions make more challenging the goal of reducing the wage gaps and providing protection compared to more developed countries. Still, this is an essential step towards a situation that allows policies addressing these concerns to have an actual effect.

Further work needs to be done to complement the findings in this thesis and to test their robustness. For the first chapter, applying the more recent reweighted Oaxaca-Blinder decomposition for the gender wage gaps would provide a robustness check to verify that the sizeable unobserved component can be attributed to a 'pure' treatment effect. Also, netting out the effect of changes in the composition of observed attributes of workers in the labour market during the period under analysis could provide a more refined measure of changes in the magnitude of the gaps. Likewise, correcting for sample selection of women into work would verify whether the proportion attributed to the unobserved components in the wage gaps as reported in the chapter corresponds to what is reported here. The second chapter could also investigate what is driving the results only in Lima and Callao and the nature of the labour markets in the other regions. This would allow us the opportunity to unravel the mechanisms through which the other regions in Peru have been unaffected by the Exodus and to verify if the increase in labour supply was offset by the increase in labour demand in those regions. The role of complementarities in these labour markets also requires further investigation. From a technical point of view, a more precise separation of the sample where Venezuelans have settled without incurring a 'small area estimation problem', can be adopted via recent Machine Learning techniques.

The work developed in the last chapter would benefit from exploring alternative counterfactual distributions within the reweighted decomposition techniques. In addition, it would be helpful to assess how the results change when applying sampling weights to the estimated regressions and under the inclusion of other characteristics of the migrants, such as the educational experience of Venezuelans. Finally, as a robustness check, it would be useful to model the probability of perceiving discrimination as a function of treatment effects estimated using the reweighting decomposition procedure. This type of exercise would provide some reassurance on the importance of the treatment effect in determining an individual's perception of discrimination.

Bibliography

Abadie, A. (2020). Using synthetic controls: feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*.

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Working Paper 24003.

Abadie, A. and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 10:465–503.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490).

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2).

Adsera, A. and Chiswick, B. R. (2007). Are there gender and country of origin differences in immigrant labor market outcomes across european destinations. *Journal of Population Economics*, 20(3):495–526.

Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, 30:175–187.

Akgündüz, Y., van den Berg, M., and Hassink, W. (2015). The impact of refugee crises on host labor markets: The case of the syrian refugee crisis in turkey. Discussion Paper Series 8841.

Akgündüz, Y., van den Berg, M., and Hassink, W. (2018). The impact of the syrian refugee crisis on firm entry and performance in turkey. Policy Research Working Paper 8323, The World Bank.

Aktas, A. and Uysal, G. (2012). *Explaining the Gender Wage Gap in Turkey Using the Wage Structure Survey*. BETAM working paper series 005.

Albrecht, J., Björklund, A., and Vroman, S. (2003). Is there a glass ceiling in Sweden? *Journal of Labor Economics*, 21(1).

Albrecht, J., VanVuuren, A., and Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to The Netherlands. *Labour Economics*, 16(4).

Aliaga, L. (2010). *Informal Economy Budget Analysis in Peru and Metropolitan Lima*. Wiego Working Paper No 13. Cambridge, MA: Harvard Kennedy School.

Altonji, J. G. and Blank, R. M. (1999). Race and gender in the labor market. In Ashenfelter, O. and Card, D., editors, *Handbook of Labour Economics*, volume 3, chapter 48, pages 3143–3259. Elsevier.

Aneshensel, C. (2013). *Theory-based data analysis for social sciences*. California: Sage editions.

Angrist, J. and Krueger, A. (1999). Empirical strategies in labor economics. In *Handbook of Labor Economics*, volume 3, pages 1277–1366. Elsevier, 1 edition.

Angrist, J. D. and Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, N.J. : Princeton University Press.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297.

Arrow, K. J. (1973). The theory of discrimination. In Ashenfelter, O. and Rees, A., editors, *Discrimination in Labor Markets*. Princeton University Press, Princeton.

Asencios, R. and Castellares, R. (2020). Impacto de la inmigracion venezolana en el empleo y los salarios: el caso peruano. Working paper.

Atal, J., Ñopo, H., and Winder, N. (2009). *New Century, Old Disparities Gender and Ethnic Wage Gaps in Latin America*. American Development Bank, Research Department.

Auer, D., Bonoli, G., and Fossati, F. (2017). Why do immigrants have longer periods of unemployment? swiss evidence. *International Migration*, 55(1):157–174.

Auer, D. and Ruedin, D. (2019). Who feels disadvantaged? reporting discrimination in surveys. In Steiner, I. and Wanner, P., editors, *Migrants andExpats: The Swiss Migration and Mobility Nexus*, pages 221–242. Springer.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baselinecovariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28:3083–3107.

Baert, S. (2017). Hiring discrimination: An overview of (almost) all correspondence experiments since 20058. Discussion Paper Series 1073.

Bahar, D., Dooley, M., and Selee, A. (2020a). Inmigrantes venezolanos, crimen y percepciones falsas: Un análisis delos datos en colombia, perú y chile. Technical report, Washington, DC.

Bahar, D., Ibáñez, A. M., and Rozo, S. V. (2020b). Give me your tired and your poor:impact of a large-scale amnesty program for undocumented refugee. Discussion Papers Series 13743.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77:1229–1279.

Balkan, B. and Tumen, S. (2016). Immigration and prices: Quasi-experimental evidence from syrian refugees in turkey. Discussion Paper Series 9642.

Baltagi, B. (2005). *Econometric Analysis of Panel Data*. Chichester : John Wiley & Sons, 3rd edition.

Banerjee, A. V. and Duflo, E. (2003). Inequality and growth: What can the data say? *Journal of Economic Growth*, 8:267–299.

Banerjee, R. (2008). An examination of factors affecting perception of workplace discrimination. *Journal of Labor Research*, 29:380–401.

Barsky, R., Bound, J., Charles, K. K., and Lupton, J. P. (2002). Accounting for the black-white wealth gap: A nonparametric approach. *Journal of the American Statistical Association*, 97(459):663–673.

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

Bell, R. M. and McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28(2):169–179.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The review of Economic Studies*, 81(2):608–650.

Benavides, M. and Valdivia, M. (2004). Metas del milenio y la brecha étnica en el perú. Technical report, Lima.

Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, 95(2):94–98.

Bertrand, M., Duflo, E., and Mullainathan, S. (2002). How much should we trust differences-in-differences estimates? Working paper 8323.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013.

Beteta, E. and Del Pozo, J. M. (2016). Aplicación de una metodología para el análisis de las desigualdades socioeconómicas en acceso a servicios de salud y educación en Perú en 2005-2012. In Francke, P. and Rodriguez, J., editors, *Políticas públicas en un contexto de crecimiento*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.

Biddle, N. (2013). Comparing self perceived and observed labour market discrimination in australia. *Economic Papers*, 32(3):383–394.

Blank, R. M., Dabady, M., and Citro, C. F., editors (2004). *Measuring racial discrimination.* Washington, D.C.: National Academies Press.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural variables. *Journal of Human Resources*, 8:436–455.

Blinder, S. (2015). Imagined immigration: The impact of different meanings of 'immigrants' in public opinion and policy debates. *Political Studies*, 63:80–100.

Blouin, C. (2019). Estudio sobre el perfil socio económico de la población venezolana y sus comunidades de acogida: una mirada hacia la inclusión. Technical report.

Blouin, C., Mamani, F., Ruiz, C., Briceño, A., Freier, L. F., Scull, C., and Berganza, I. (2019). Documento preparatorio para audiencia tematica ante la comisión interamericana de derechos humanos. Technical report.

Bond, S. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, (1):141–162.

Borjas, G. (2017). The wage impact of the marielitos: A reappraisal. *Immigration and Labor Markets*, 70(5).

Borjas, G. J. (2003). The labor demand curve is downward sloping: reexamining the impactof immigration on the labor market. *Quarterly Journal of Economics*, 118(4):1335–1374.

Borjas, G. J. (2014). *Immigration Economics*. Cambridge, Massachusetts, London, England: Harvard University Press.

Borjas, G. J. and Monras, J. (2017). The labour market consequences of refugee supply shocks. *Economic Policy*, 32(91):361–413.

Boruchowicz, C., Martinelli, C., and Parker, S. W. (2021). Economic consequences of mass migration: the venezuelanexodus in peru. Working paper.

Boserup, E. (1970). *Woman's Role in Economic Development*. London: George Allen & Unwin.

Brewer, M., Crossley, T. F., and Joyce, R. (2018). Inference with difference-in-differences revisited. *Journal of Econometric Methods*, 7(1).

Brunow, S. and Jost, O. (2019). Wages of migrant and native employees in germany: new light on an old issue. Discussion Paper 201910.

Buchinsky, M. (1998a). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics*, 13:1–30.

Buchinsky, M. (1998b). Recent advances in quantile regression models a practical guideline for empirical research. *The Journal of Human Resources*, 33(1):88–126.

Busso, M., DiNardo, J., and McCrary, J. (2014). New evidence on the finite sample properties of propensity scorereweighting and matching estimators. *Review of Economics and Statistics*, 96:885–897.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.

Cameron, C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Industrial & Labor Relations Review*, 43(2):245–257.

Card, D. (2001). Immigrant inflows, nativeoutflows, and the local labor market: Impacts of higher immigration. *Journal of Labor Economics*, 19(1).

Card, D. and Peri, G. (2016). Immigration economics: A review. Working paper.

Carrington, W. and de Lima, P. (1996). The impact of 1970s repatriates from africa on the portuguese labor market. *Industrial and Labour Relations Review*, 49(2):330–347.

Carroll, H., Luzes, M., Freier, L. F., and Bird, M. D. (2020). The migration journey and mental health: Evidence from venezuelan forced migration. *SSM - Population Health*, 10.

Carter, A. V., Schnepel, K. T., and G. Steigerwald, D. (2017). Asymptotic behavior of a t-test robust to clusterheterogeneity. *The Review of Economics and Statistics*, 99(4):698–709.

Castillo, W. (2011). *Discriminación salarial por género en el Perú : 2003-2009*. Thesis for the degree of magister in Applied Economics. Universidad De Chile - Facultad De Ciencias Físicas y Matemáticas.

Cattaneo, M. D., Drukker, D. M., and Holland, A. D. (2013). Estimation of multivalued treatment effectsunder conditional independence. *The Stata Journal*, 13(3):407–450.

CEPAL, FAO, ONU, PNUD, and OIT (2013). *Trabajo decente e igualdad de género. Políticas para mejorar el acceso y la calidad del empleo de las mujeres en América Latina y el Caribe*. Santiago de Chile: Comisión Económica para América Latina y el Caribe (CEPAL).

Ceritoglu, E., Yunculer, H. B., Torun, H., and Tumen, S. (2015). The impact of syrian refugees on natives' labor market outcomes in turkey: Evidence from a quasi-experimental design. Discussion Paper Series 9438.

Chen, H. and Shen, Q. R. (2015). *Variance Estimation for Survey-Weighted Data Using Bootstrap Resampling Methods: 2013 Methods-of-Payment Survey Questionnaire*. Working paper. Currency Department. Bank of Canada.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Chesher, A. and Irish, M. (1987). Residual analysis in the grouped and censored normal linear model. *Journal of Econometrics*, 34:33–61.

Chiswick, B. R., Le, A. T., and Miller, P. W. (2008). How immigrants fare across the earnings distribution in australia and the united states. *Industrial Labor Relations*, 61:353–373.

Chiswick, B. R. and Miller, P. W. (1999). Language skills and earnings among legalized aliens. *Journal of Population Economics*, 12(1):63–89. Special Issue on Illegal Migration.

CHRET (2017). Analytical report of the foreign residents survey. revised edition. Research project commissioned by ministry of justice,.

Citrin, J. and Sides, J. (2008). Immigration and the imagined communityin europe and the united states. *Political Studies*, 56:33–56.

Clemens, M. A. and Hunt, J. (2017). The labor market effects of refugee waves: reconciling conflicting results. Working Paper 23433.

Coppola, A. and Calvo-Gonzalez, O. (2011). *Higher Wages, Lower Pay. Public vs. Private Sector Compensation in Peru*. Policy Research Working Paper 5858. The World Bank Latin America and the Caribbean Region.

Crosby, F. (1984). The denial of personal discrimination. *American Behavioral Scientist*, 27:371–386.

Daldy, B., Poot, J., and Roskruge, M. (2013). Perception of workplace discrimination among immigrants and native born new zealanders. Iza dp no. 7504.

Dancourt, O. (1999). Neoliberal reforms and macroeconomic policy in peru. *CEPAL Review*, (67):51–73.

Dancygier, R. M. and Laitin, D. D. (2014). Immigration into europe: Economic discrimination, violence, and public policy. *Annual Review of Political Science*, 17:43–64.

de Vroome, T., Verkuyten, M., and Martinovic, B. (2014). Host national identification of immigrants in the netherlands. *International Migration Review*, 48(1):1–27.

Del Carpio, X. V. and Wagner, M. (2015). The impact of syrian refugees on the turkish labor market. Policy Research Working Paper 7402.

Del Pozo Segura, J. M. (2021). The effect of the venezuelan exodus on peruvian labour market outcomes. Ph.D. dissertation chapter.

Di Stasio, V. and Heath, A. (2019). Are employers in britain discriminating against ethnic minorities? Technical report.

Di Stasio, V., Lancee, B., Veit, S., and Yemane, R. (2021). Muslim by default or religious discrimination? results from a cross-national field experiment on hiring discrimination. Technical report.

Dickens, R. and McKnight, A. (2008). Assimilation of migrants into the british labour market. STICERD Research Paper CASE/133.

DiNardo, J., Fortin, N., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5).

Dion, K. and Kawakami, K. (1996). Ethnicity and perceived discrimination in toronto: Another look at the personal/groupdiscrimination discrepancy. *Canadian Journal of Behavioural Science*, 28(3):203–213.

Dion, K. L. (1986). Responses to perceived discrimination and relative deprivation. In Olson, J. M., Herman, C. P., and Zanna, M. P., editors, *Relative deprivation and social comparison: the Ontario symposium*, volume 4, pages 159–179. Hillsdale: Erlbaum.

Donald, S. G. and Lang, K. (2007). Inference with difference-in-differences and otherpanel data. *The Review of Economics and Statistics*, 89(2):221–233.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression,difference-in-differences and synthetic control methods: A synthesis. Working Paper 22791.

Dovidio, J. F. and Gaertner, S. L. (2010). Intergroup bias. In Fiske, S. T., Gilbert, D. T., and Lindzey, G., editors, *The Handbook of Social Psychology*, pages 1084–1121. New York: Wiley, 5th edition.

Dustmann, C., Frattini, T., and Glitz, A. (2007). The impact of migration: a review of the economic evidence. Final report. File No.102 Contract No 36/2006.

Dustmann, C., Frattini, T., and Preston, I. (2013). The effect of immigration along the distribution of wages. *Review of Economic Studies*, 80:145–173.

Dustmann, C., Glitz, A., and Frattini, T. (2008). The labour market impact of immigration. Discussion Paper Series CDP No 11/08.

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.

Dustmann, C., Vasiljeva, K., and Damm, A. (2019). Refugee migration and electoral outcomes. *Review of Economic Studies*, 86:2035–2091.

Elliott, R. J. R. and Lindley, J. K. (2008). Immigrant wage differentials, ethnicity andoccupational segregation. *Journal of the Royal Statistical Society Series A*, 171(3):645–671.

Ensher, E. A., Grant-Vallone, E. J., and Donaldson, S. I. (2001). Effects of perceived discrimination on job satisfaction, organizational commitment, organizational citizenship behavior, and grievances. *Human Resource Development Quarterly*, 12(1):53–72.

España, L. P. and Ponce, M. G. (2018). Encuesta sobre condiciones de vida en venezuela. In *IIES-UCAB*.

Fasani, F., Mastrobuoni, G., Owens, E., and Pinotti, P. (2019). *Does Immigration Increase Crime?* Cambridge University Press.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pre-treatment fit. *Quantitative Economics*, Forthcoming.

Ferman, B., Pinto, C., and Possebom, V. (2020). Cherry picking with synthetic controls. *Journal of Policy Analysis and Management*, 39(2):510–532.

Fernández-Reino, M. (2020). Migrants and discrimination in the uk. Technical report.

Firpo, S., Fortin, N., and Lemieux, T. (2007). *Decomposing Wage Distributions using Recentered Influence Function Regressions*. Mimeo, University of British Columbia.

Firpo, S., Fortin, N., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Firpo, S. and Possebom, V. (2018). Synthetic control method: inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, 6(2).

Firpo, S. P., Fortin, N. M., and Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics*, 6(28).

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In Gilbert, D. T., Fiske, S. T., and Lindzey, G., editors, *The Handbook of Social Psychology*, volume 2, pages 357–411. New York: McGraw-Hill, 4th edition.

Foged, M. and Peri, G. (2015). Immigrants' effect on native workers:new analysis on longitudinal data. DP 8961.

Fortin, N., Lemieux, T., and Firpo, S. (2010). *Decomposition methods in economics*. National Bureau Of Economic Research .Working Paper 16045.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labour Economics*, volume 4, pages 1–102. North-Holland.

Freeman, R. (1980). Unionism and the dispersion of wages. *Industrial and Labor Relations Review*, 34(1):3–23.

Freier, L. and Pérez, L. (2021). Nationality-based criminalisation of south-south migration: the experience of venezuelan forced migrants in peru. *European Journal on Criminal Policy and Research*, 27:113–133.

Freije, S. (2002). *Informal Employment in Latin America and the Caribbean: Causes, Consequences and Policy Recommendations*. Inter-American Development Bank. Sustainable Development Department, Social Development Division, Labor Markets Policy Briefs Series.

Friedberg, R. (2000). You can't take it with you? immigrant assimilation and the portability of human capital. *Journal of Labor Economics*, 18(2):221–51.

Friedberg, R. M. (2001). The impact of mass migration on the israeli labor market. *The Quarterly Journal of Economics*, 116(4):1373–1408.

Frölich, M. and Sperlich, S. (2019). *Impact EvaluationTreatment Effects and Causal Analysis*. Cambridge University Press.

Fuentes, R. (2019). Situación del mercado laboral juvenil a nivel nacional y regionalen el contexto de la migración venezolana. Technical report.

Garavito, C. (2011). Desigualdad en los ingresos: género y lengua materna. In Iguíñiz, J. and León, J., editors, *Desigualdad distributiva en el Perú: dimensiones*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.

Geronimi, E. (2004). Admisión, contratación y protección de trabajadores migrantes. panorama de la legislación y la práctica nacionales de argentina, bolivia, brasil, chile, colombia, ecuador, españa, perú, portugal y uruguay. Estudios sobre migraciones internacionales.

Gonzales de Olarte, E. (2015). *Una economía incompleta. Perú 1950-2007*. LIma : PUCP.

Gorodzeisky, A. and Semyonov, M. (2019). Unwelcome immigrants: Sources of opposition to different immigrant groups among europeans. *Frontiers in Sociology*, 12.

Grigorieff, A., Roth, C., and Ubfal, D. (2020). Does information change attitudes towards immigrants? representative evidence from survey experiments. *Demography*, 57(3):1117–1143.

Gronau, R. (1973). *Wage comparisons - A selectivity Bias*. National Bureau Of Economic Research. Working Paper 13.

Gushiken, R. and Campos, M. Á. (2015). Línea de base de brechas sociales por origen étnico en el perú. Technical report.

Hainmueller, J., Hiscox, M. J., and Margalit, Y. (2015). Do concerns about labor market competition shape attitudes toward immigration? new evidence. *Journal of International Economics*, 97(1):193–207.

Hansen, C. B. (2007). Generalized least squares inference in paneland multilevel models with serial correlationand fixed effects. *Journal of Econometrics*, 140:670–694.

Hardy, W., Keister, R., and Lewandowski, P. (2018). Educational upgrading, structural change and the task composition of jobs in europe. *Economics Of Transition and Institutional Change*, 26(2).

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC.

Hausmann, R. and Klinger, B. (2008). Growth diagnostic: Peru. Working Paper PE-P1074, Center for International Development. Harvard University.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Heckman, J. J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, 12(2):101–116.

Heckman, J. J. and Siegelman, P. (1993). The urban institute audit studies: Their methods and findings. In Fix, M. and Struyck, R. J., editors, *Clear and Convincing Evidence: Measurement of Discrimination in America*. Urban Institute Press, Lanham.

Holland, P. W. (1986). *Statistics and causal inference*. Journal of the American Statistical Association.

Holtz-Eakin, D., Newey, W., and S. Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56(6):1371–1395.

Hopkins, D. J., Mummolo, J., Esses, V. M., Kaiser, C. R., Marrow, H. B., and McDermott, M. (2016). Out of context: The absence of geographic variation in us immigrants' perceptions of discrimination. *Politics, Groups, and Identities*, 4(3):363–392.

HRW (2018). The venezuelan exodus. the need for international protection and the regions response. Technical report.

Hunt, J. (1992). The impact of the 1962 repatriates from algeria on the french labor market. *Industrial & Labor Relations Review*, 45(3):556–572.

Hussmanns, R. (2001). Informal sector and informal employment: elements of a conceptual framework. Technical report, New Delhi.

Hussmanns, R. (2004). Statistical definition of informal employment: Guidelines endorsed by the seventeenth international conference of labour statisticians (2003). Technical report, New Delhi.

IACHR (2017). Situation of human rights in venezuela. Technical report.

IADB (2010). *The age of productivity. Transforming economies from the bottom up*. Inter-American Development Bank.

IAHCR (2017). Situation of human rights in venezuela. Technical report.

IDEHPUCP, GPMH, CEAS, CNDH, and Clínica jurídica para migrantes y refugiados (2020). Informe alternativo al comité de protección de los derechos de todos los trabajadores migratorios y de sus familiares. online, https://cdn01.pucp.education/idehpucp/wp-content/uploads/2020/12/21213437/Versi

IEP (2019). Conocimiento y actitudes hacia la migración venezolana. encuesta nacional urbano rural. Informe de opinión.

ILC (2002). Decent work and the informal economy. 90th session. report vi, Geneva.

ILO (2002). Effect to be given to resolutions adopted by the international labour conference at its 90th session (2002). 285th session, Geneva,.

ILO (2014). Trends in informal employment in peru: 2004 - 2012. Technical report, International Labour Organization.

ILO (2015). Recent experiences of formalization in latin america and the caribbean. Notes on formalization.

ILO (2016). Closing the gender pay gap: A review of the issues, policymechanisms and international evidence. Technical report, Geneva : International Labour Organization. Gender, Equality and Diversity Branch.

ILO (2017). Labour 2017 overview latina america and caribbean. Technical report.

ILO (2018a). *Labour Overview 2018*. Lima: Regional Office for Latin America and the Caribbean.

ILO (2018b). *Women and men in the informal economy: A Statistical Picture*. Geneva : International Labour Organization, 3rd edition.

ILO (2019). 2018 labour overview of latin america and the caribbean. Technical report.

ILO (2020). The migrant pay gap: Understanding wage differences between migrants and nationals. Technical report.

Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society, Series B*, 76:243–263.

Imbens, G. and Wooldridge, J. (2014). Cemmap lecture notes. Technical report.

Imbens, G. W. and Kolesar, M. (2012). Robust standard errors in small samples: ome practical advice. Working Paper 18478.

IMF (2014). Peru. selected issues paper. IMF country report 14/22, Washington D.C.: International Monetary Fund.

INEI (2014a). *Perú: Brechas de género, 2001-2013. Avances hacia la igualdad entre hombres y mujeres*. Lima: Instituto Nacional de Estadística e Informática (INEI).

INEI (2014b). *Perú: Evolución de los Indicadores de los Objetivos de Desarrollo del Milenio al 2012*. Lima: Instituto Nacional de Estadística e Informática (INEI).

INEI (2016). *Clasificador Nacional de Ocupaciones 2015*. Lima : Instituto Nacional de Estadística e Informática.

INEI (2019a). *Condiciones de Vida de la Población Venezolana que reside en Perú*. Lima : Instituto Nacional de Estadística e Informática.

INEI (2019b). *Perú: Evolución de los Indicadores de Empleo e Ingreso por Departamento, 2007-2018*. INEI.

INEI (2020). *Producción y empleo informal en el Perú. Cuenta satélite de laeconomía informal 2007-2019*. Instituto Nacional de Estadística e Informática.

Inman, M. L. and Baron, R. S. (1996). Influence of prototypes on perceptions of prejudice. *Journal of Personality and Social Psychology*, 70(4):727–739.

Inman, M. l., Huerta, J., and Oh, S. (1998). Perceiving discrimination: the role of prototypes and norm violation. *Social Cognition*, 16(4):418–450.

Kampelmann, S. and Rycx, F. (2016). Wage discrimination against immigrants: measurement with firm-level productivity data. *IZA Journal of Migration*, 5(15).

Kerr, S. and Kerr, W. (2011). Economic impacts of immigration: a survey. *Finnish Economic Papers*, 24(1):1–31.

Kidd, M. (1993). Immigrant wage differentials and the role of self-employment in australia. *Australian Economic Papers*, 32(60):92–115.

Kiefer, N. M. (1980). Estimation of fixed effect models for time series of cross-sections with arbitraryintertemporal covariance. *Journal of Econometrics*, 14:195–202.

Kline, P. (2011). Oaxaca-blinder as a reweighting estimator. *American Economic Review*, 101(3):532–37.

Kobrynowicz, D. and Branscombe, N. R. (1997). Who considers themselves victims of discrimination? individual difference predictors of perceived gender discrimination in women and men. *Psychology of Women Quarterly*, 21(3):347–363.

Koechlin, J., Solórzano, X., Larco, G., and Fernández-Maldonado, E. (2019). Impacto de la inmigración venezolana en elmercado laboral de tres ciudades: Lima, arequipa y piura. Technical report.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge university press.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50(1):43–61.

Koenker, R. and Hallock, K. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4):143 – 156.

Kogan, I. (2006). Labor markets and economic incorporation among recent immigrants in europe. *Social Forces*, 85(2):697–721.

Kolenikov, S. (2010). Resampling variance estimation for complex survey data. *The Stata Journal*, 10(2):165–199.

Konle-Seidl, R. and Bolits, G. (2016). Labour market integration of refugees: Strategies and good practices. Technical report.

Koopmans, R. (2015). Does assimilation work? sociocultural determinants of labour market participation of european muslims. *Journal of Ethnic and Migration Studies*, 42(2):197–216.

Kreuter, F. and Valliant, R. (2007). A survey on survey statistics. what is done and can be done in stata. *The Stata Journal*, 7(1):1–21.

Kuosmanen, T., Zhou, X., Eskelinen, J., and Malo, P. (2021). Design flaw of the synthetic control method. MPRA Paper 106390.

Kustov, A. (2019). Is there a backlash against immigration from richer countries? international hierarchy and the limits of group threat. *Political Psychology*, 40(5):973–1000.

La Porta, R. and Shleifer, A. (2014). Informality and developmen. *Journal of Economic Perspectives*, 28(3):109–126.

Lang, K. and Lehmann, J.-Y. K. (2012). Racial discrimination in the labor market: Theory and empirics. *Journal of Economic Literature*, 50(4):959–1006.

Lee, C. H. and Steigerwald, D. G. (2018). Inference for clustered data. *The Stata Journal*, 18(2).

Lee, M.-J. (2016). *Matching, Regression Discontinuity, Difference in Differences, and Beyond*. New York, NY : Oxford University Press.

Leszczensky, L. and Wolbring, T. (2019). How to deal with reverse causality using panel data? recommendations for researchers based on a simulation study. *Sociological Methods & Research*.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Machado, J. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4):445–465.

Mackinnon, J. G. and Webb, M. (2016). Wild bootstrap inference for wildly different cluster sizes. *Journal of applied econometrics*, 32(2):233–254.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.

Maddison, A. (2007). World population, gdp and per capita gdp 1-2003 ad. Technical report, Groningen Growth and Development Centre.

Manacorda, M., Manning, A., and Wadsworth, J. (2006). The impact of immigration on the structure of male wages: Theory and evidence from britain. CEP Discussion Paper 754.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12:577 – 590.

Mendoza, W. (2013). Peruvian miracle: good luck or good policies? Working Paper 371, Pontificia Universidad Católica del Peru.

Messina, J. and Silva, J. (2018). *Wage Inequality in Latin America : Understanding the Past to Prepare for the Future*. Washington, DC: World Bank.

Migraciones (2018). Actualidad migratoria. Technical Report 006. Year 2.

Mitra, A., Bang, J., and Biswas, A. (2015). Gender equality and economic growth: Is it equality of opportunity or equality of outcomes? *Feminist Economics*, 21(1).

Montes, J. (2007). *Evolution of the gender wage gap in Peru, 1997-2000*. Thesis for the degree of Master of Science submitted to the Office of Graduate Studies of Texas A&M University.

Mora, C., Bustillos, J., and Jin, G. (2019). The short-term impact of venezuelan migration on the labor market outcomes of natives in peru and colombia. Master Project.

Morales, F. and Pierola, M. D. (2020). Venezuelan migration in peru: Short-term adjustments in the labor market. Working Paper Series IDB-WP-1146.

Morita, L. (2017). Why japan isn t more attractive to highly skilled migrants. *Cogent Social Sciences*.

Morrison, A. (2021). La participación laboral de lasmujeres en el perú: 10 mensajesclave y 6 recomendaciones de política. Nota técnica IDB-TN-2147, Banco Interamericano del Desarrollo. División de Género y Diversidad.

Mougenot, B., Amaya, E., Mezones-Holguin, E., Rodriguez-Morales, A. J., and Cabieses, B. (2021). Immigration, perceived discrimination and mental health: evidence from venezuelan population living in peru. *Globalization and Health*, 17(8).

MTPE (2010). *Informe Anual 2009. La Mujer en el Mercado Laboral Peruano*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

MTPE (2012). *Políticas Nacionales de Empleo*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

MTPE (2014). *Informe Anual 2012. La Mujer en el Mercado Laboral Peruano*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

MTPE (2015). *Informe Anual 2014. La Mujer en el Mercado Laboral Peruano*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

Nandi, A., Luthra, R. R., and BenzevalI, M. (2016). Ethnic and racial harassment and mental health: identifying sources of resilience. Working Paper Series 2016-14.

Neal, D. A. and Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5):869–95.

Newey, W. (1988). *Two step series estimation of sample selection model*. Presented at the 1988 European meeting of the Econometric Society.

Newey, W. K. and McFadden, D. (1994). Handbook of econometrics. volume 4, chapter Large Sample Estimation and Hypothesis Testing, pages 2111–2245. Amsterdam: North Holland.

Oaxaca, R. (1973). Male-female differentials in urban labor markets. *International Economic Review*, 14(693-709).

OECD and ILO (2019). *Tackling Vulnerability in the Informal Economy*. Paris: OECD Publishing. Development Centre Studies.

OIM (2018a). Monitoreo de flujo de la población venezolana en el perú. reporte 2. Technical report.

OIM (2018b). Monitoreo de flujo de la población venezolana en el perú. reporte 3. Technical report.

OIM (2018c). Monitoreo de flujo de la población venezolana en el perú. reporte 4. Technical report.

OIM (2018d). Regional action plan. Technical report.

OIM (2019). Monitoreo de flujo de la población venezolana en el perú. ronda 5. Technical report.

OIM (2020). Monitoreo de flujo de la población venezolana en el perú. reporte 7. Technical report.

OIT (2021). El aporte de laspersonas refugiadas ymigrantes venezolanas en los servicios esenciales de repartode alimentos y medicina durante lapandemia de la covid-19 argentina, colombia y perú. Technical report.

OIT and PNUD (2009). *Trabajo y Familia: Hacia nuevas formas de conciliación con corresponsabilidad social*. Santiago de Chile: Oficina Internacional del Trabajo.

Okkerse, L. (2008). How to measure labour marketeffects of immigration: a review. *Journal of Economic Surveys*, 22(1):1–30.

Oreopoulos, P. (2011). Why do skilled immigrants struggle in the labor market. *American Economic Journal: Economic Policy*, 3(4):148–171.

Ottaviano, G. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1):152–197.

Ottaviano, G., Peri, G., and Wright, G. (2013). Immigration, offshoring, and american jobs. *American Economic Review*, 103(5):1925–1959.

Pacheco, E. (2013). *Gender Wage Gap in Urban Nicaragua: Evidence from Decomposition Analysis*. Master in Economic Development and Growth. Lund University.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press.

Pager, D. and Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, 34:181–209.

Pages, C., editor (2010). *La era de la productividad. Cómo trans formar las economías desde sus cimientos*. Banco Interamericano de Desarrollo.

Pampel, F. C. and Tanaka, K. (1986). Economic development and female labor force participation: A reconsideration. *Social Forces*, 64(3):599–619,.

PCM (2013). *Perú: tercer informe nacional de cumplimiento de los Objetivos de Desarrollo del Milenio.* Lima: Presidencia del Consejo de Ministros (PCM), Sistema de Naciones Unidas en el Perú.

Peri, G. (2007). Immigrants' complementarities and native wages: evidence from california. Working Paper 12956.

Peri, G. (2016). Immigrants, productivity, and labor markets. *Journal of Economic Perspectives*, 30(4):3–30.

Peri, G., Rury, D., and Wiltshire, J. C. (2020). The economic impact of migrants from hurricane maria. Discussion Paper 13049.

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

Peri, G. and Yasenov, V. (2017). The labor market effects of a refugee wave: Synthetic control method meets the mariel boatlift. Discussion Paper 10605.

Pfeffermann, D. (2002). Small area estimation: New developments and directions. *International Statistical Review*, 70(1):125–143.

Phelps, E. (1972). The statistical theory of racism and sexism. *American Economic Review*, 62:659–61.

Porter, S. (2015). Quantile regression: Analyzing changes in distributions instead of means. In Paulsen, M. B., editor, *Higher Education: Handbook of Theory and Research: Volume 30*. Cham: Springer International Publishing.

Powell, D. (2018). Imperfect synthetic controls: Did the massachusetts health care reform save lives. Working Paper 1246.

PUCP (2019). Creencias y actitudes hacia los inmigrantes venezolanos en el peru. Technical report.

PUCP (2020). Cambios en las actitudes hacia los inmigrantes venezolanos en lima-callao 2018-2019. Bulletin 166.

R4V (2020a). Refugee and migrant response plan for refugees and migrants from venezuela 2020. peru. Technical report.

R4V (2020b). Situation report january march. Technical report.

Ramsey, J. B. (1969). Tests of specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society, B*, 31:350–71.

Reinhart, C. and Santos, M. A. (2015). From financial repression to external distress: the case of venezuela. Working Paper 21333.

Restuccia, D. (2019). The monetary and fiscal historyof venezuela, 1960-2016. Working paper.

Rios-Avila, F. (2020). Recentered influence functions (rifs) in stata. rif regression and rif decomposition. *The Stata Journal*, 20(1):51–94.

Roodman, D. (2009a). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1).

Roodman, R. (2009b). How to do xtabond2: An introduction to difference and system gmm in stata. *The Stata Journal*, 9(1):86–136.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Ross, K. and Peschiera, J. (2015). Explaining the peruvian growth miracle. In Santos, A. and Werner, A., editors, *Peru: Staying the Course of Economic Success*, chapter Explaining the Peruvian Growth Miracle. Washington D.C. : IMF.

Saavedra, J. (1999). *La dinámica del mercado de trabajo en el Perú antes y después de las reformas estructurales*. Serie Reformas Económicas, No 27. Comisión Económica para América Latina y el Caribe (CEPAL).

Safi, M. (2010). Immigrants' life satisfaction in europe: Between assimilation and discrimination. *European Sociological Review*, 26(2):159–176.

Salardi, P. (2012). *Wage Disparities and Occupational Intensity by Gender and Race in Brazil: An Empirical Analysis Using Quantile Decomposition techniques*. University of Sussex, Job market paper.

Samartsidis, P., Seaman, S. R., Presanis, A. M., Hickman, M., and De Angelis, D. (2019). Assessing the causal effect of binary interventions from observational panel data with few treated units. *Statistical Science*, 34(3):486–503.

Santos, A. and Werner, A., editors (2015). *Peru: Staying the course of economic success*. Washington D.C. : IMF.

Schmitt, M. T., Branscombe, N. R., Postmes, T., and Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: A meta-analytic review. *Psychological Bulletin*, 140(4):921–948.

Seguino, S. (1997). Gender wage inequality and export-led growth in south korea. *Journal of Development Studies*, 34(2):102–132.

Seminario, B. (2015). *El desarrollo de la economía peruana en la era moderna: precios, población, demanda y producción desde 1700*. Lima: Universidad del Pacífico.

Siebers, H. and van Gastel, J. (2015). Why migrants earn less: in search of the factors producing the ethno-migrant pay gap in a dutch public organization. *Work, Employment and Society*, 29(3):371–391.

Silverman, B. W. (1986). *Density Estimationjiir Statistics and Data Analysis*. New York: Chapman and Hall.

Strumpf, E., Harper, S., and Kaufman, J. S. (2017). Fixed effects and difference-in-differences. In Oakes, J. M. and Kaufman, J. S., editors, *Methods in Social Epidemiology*. San Francisco, CA : Jossey Bass.

Tajfel, H. and Turner, J. C. (1986). The social identity theory of inter-group behavior. In Worchel, S. and Austin, L. W., editors, *Psychology of intergroup relations*, pages 2–24. Nelson-Hall, Chicago.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288.

Torales, P., González, M. E., , and Vichich, N. P. (2003). Migraciones laborales en sudamérica: la comunidad andina. Estudios sobre migraciones internacionales.

Tumen, S. (2015). The use of natural experiments in migration research. Iza world of labor. 191–201.

Tzavidis, N., Zhang, L.-C., Luna, A., Schmid, T., and Rojas-Perilla, N. (2018). From start to finish: a framework for the production of small area official statistics. *Journal of the Royal Statistical Society. Series A*, 181(4):927–979.

Távara, J., González de Olarte, E., and Del Pozo, J. M. (2014). Heterogeneidad estructural y articulación productiva en el Perú: evolución y estrategias. In Infante, R. and Chacaltana, J., editors, *Hacia un desarrollo inclusivo. El caso del Perú*. Santiago de Chile : Comisión Económica para América Latina y el Caribe (CEPAL).

UNHCR (2018a). Desperate journeys: refugees and migrants arriving in europe and at europe's borders. january - december 2018. Technical report.

UNHCR (2018b). Venezuela situation as of june 2018. situational update. Technical report.

UNHCR (2019a). Global trends forced displacement in 2019. Technical report, United Nations High Comissionate for the Refugees.

UNHCR (2019b). Protection monitoring: Venezuela situation. january - june 2019. Technical report.

UNHCR (2019c). Unhcr scales up response as record number of venezuelans arrive in peru. `https://www.unhcr.org/news/press/2019/6/5d05310d4/unhcr-scales-response-record-number-venezuelans-arrive-peru.html#_ga=2.221975642.339970819.1617661338-2145690065.1617661338`. Accessed: 15 May 2021.

UNHCR (2019d). Update: Durable solutions for syrian refugees. january - february 2019. Technical report.

UNHCR (2020). Venezuela situation. fact sheet. as of january 2020. Technical report.

United Nations, editor (2019). *Commodity Dependence: A Twenty-year Perspective*. New York: United Nations.

Valdivia, N., Benavides, M., and Torero, M. (2007). Investigación, políticas y desarrollo en el perú. chapter Exclusión, identidad étnica y políticas de inclusión social en el Perú: el caso de lapoblación indígena y la población afrodescendiente. Lima: GRADE.

Van Dun, M. (2016). Cocaine flows and the state in peru's amazonian borderlands. *Journal of Latin American Studies*, 48(3).

Viollaz, M. (2019). Multiple violations of labor market regulations: Patterns in the peruvian labor market and the impact of enforcement. Documento de Trabajo 244.

Vtyurina, S. (2015). The role of fiscal policies in peru'stransformations. In Santos, A. and Werner, A., editors, *Peru: Staying the Course of Economic Success*, chapter Explaining the Peruvian Growth Miracle. Washington D.C. : IMF.

Wallace, M. and Figueroa, R. (2012). Determinants of perceived immigrant job threat in the american states. *Sociological Perspectives*, 55(4):583–612.

Webb, R. (2013). *Conexión y despegue rural*. Lima: Fondo Editorial de la Universidad San Martín de Porres.

WEF (2020). Global gender gap report. Technical report, Geneva : World Economic Forum:.

Wenz, S. E. (2019). What quantile regression does and doesn't do: A commentary on petscher and logan (2014). *Child Development*, 90(4):1442–1452.

White, H. (1980). Using least squares to approximate unknown regression functions. *International Economic Review*, 21:149–170.

White, H. (2000). *Asymptotic Theory for Econometricians*. California: Academic Press Inc.

Widner, D. and Chicoine, S. (2011). It's all in the name: Employment discrimination against arab americans. *Sociological Forum*, 26(4):806–821.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass. : MIT Press.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *AEA Papers and Proceedings - Issues*.

World Bank (2010). *Peru - Recent economic development in infrastructure investing in infrastructure as an engine for growth : spending more, faster and spending better. Vols. I and II*. Washington, D.C.: The World Bank.

World Bank (2015). *Peru: Building on Success. Boasting productivity for faster growth*. Washington, D.C.: The World Bank.

World Bank (2021a). World development indicators. gdp per capita growth (annual [data file]. Technical report, Retrieved from https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG?locations=PE.

World Bank (2021b). World development indicators. international migrant stock (population) [data file]. Technical report, Retrieved from https://data.worldbank.org/indicator/SM.POP.TOTL.ZS?locations=PE.

Wrench, J. (2007). The measurement of employment discrimination by research: problems of comparability labour market discrimination of international migrants: concepts, data sources and issues.

Xu, Y. (2017). Generalized synthetic control method: Causal inference withinteractive fixed eects models. *Political Analysis*, 25:57–76.

Yamada, G., Lavado, P., and Velarde, L. (2013). *Habilidades No Cognitivas y Brecha de Género Salarial en el Perú*. Banco Central de Reserva del Perú. DT. N° 2013-014. Working Paper series.