# Three Essays on Public Policy and Development

**A THESIS**
**SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL**
**OF THE UNIVERSITY OF MINNESOTA**
**BY**

**Julieth Andrea Santamaria Bonilla**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS**
**FOR THE DEGREE OF**
**Ph.D. in Applied Economics**

**Joseph Ritter and Paul Glewwe**

**December, 2022**

# Acknowledgements

The development of this doctoral degree was possible thanks to the funding I received from the Department of Applied Economics at the University of Minnesota through their fellowship program. I am also grateful for the support of the Graduate School of the University of Minnesota for the funding provided through the Doctoral Dissertation Fellowship.

There are many people I would like to thank for their support during my PhD.

First, I would like to thank Joe Ritter, who supported me since he started being my second-year paper advisor. Throughout the PhD, he gave me excellent mentoring in both research and the art of teaching, and always had his door open to have long conversations with me about my research. His support on multiple projects, grants, and applications was second to none.

I owe special thanks to Paul Glewwe, co-advisor, who gave me great learnings as a teacher and pushed me immensely to finish these research papers. His extensive and detailed comments undoubtedly gave this dissertation a twist. I will always be grateful for his willingness and flexibility to address my concerns.

In addition, I would like to greatly thank Marc Bellemare, co-author and mentor, who provided excellent intellectual and emotional support. I appreciate his practicality and honesty. Always open to listening and finding solutions to help me on my path to becoming a researcher. I definitely feel that he believed more in my research than I did myself.

There are other professors I would like to thank: Alan Benson, a member of my committee, for his comments to my dissertation; and Deborah Levinson, with whom I had few conversations, but great learning that marked me professionally. I would also like to thank instructors Terry Hurley and Steve Miller.

On the other hand, my time at the University of Minnesota left me with several friends and colleagues who undoubtedly made my experience different. I would especially like to thank Monika Fellmer and José Casco, my dearest friends, who accompanied me and supported me during this journey. UMN would not have been the same without you. I also thank other friends, colleagues and co-authors I made along the way: Kenn Chua, Khoa Vu, Natalia Ordaz, Soomin Lee, Dojin Park, Ian Luby, Colette Salemi, Maxwell Mkondiwa, and Thomas Durfee.

During my PhD, I had multiple opportunities to work and gain experience in the field. For that reason, I want to extend my enormous thanks to the constant support and mentoring from Julian Cristia and Lucia Hanmer throughout these years, both of whom are excellent human beings and always found opportunities for me to work with them. All these experiences with both made me a better professional. I would like to thank the opportunities to do research with Eliana Rubiano and Pedro Bernal. I am also indebted for the last mile support of Ben Roseth, whom I thank for his enormous emotional support, patience and flexibility that enabled me cross the finish line of this long road.

Finally, I would like to thank my partner, Fernando Lino, whom I met and fell in love with during my studies at UMN. Fernando – I feel very fortunate to have found you. You have been the engine that drove me to question myself at every point of this dissertation, who pushed me to do better quality work, who was there to offer me company in the long nights of work, who has advised me to make better decisions, and who has been there to celebrate my successes and lift me up in my low moments. All these experiences we have shared have been a blessing. This title is definitely yours as well!

# Dedication

Para mis abuelos, Teresa Ramirez y Salomón Bonilla, a quienes siempre tendré en mi mente y corazón a cada paso que doy. Para mi mamá, Mireya Bonilla, a quien le agradezco su apoyo incondicional.

## Abstract

This dissertation studies the effects of three policy decisions on citizen outcomes.

Chapter 1 analyzes the effect of open-door immigration policies on local labor markets. Using the sharp and unprecedented surge of Venezuelan refugees into Colombia, I study the impact on wages and employment in a context where work permits were granted at scale. To identify which labor markets immigrants are entering, I overcome limitations in official records and generate novel evidence of refugee settlement patterns by tracking the geographical distribution of Internet search terms that Venezuelans but not Colombians use. While official records suggest migrants are concentrated in a few cities, the Internet search index shows migrants are located across the country. Using this index, high-frequency labor market data, and a difference-in-differences design, I find precise null effects on employment and wages in the formal and informal sectors. A machine learning approach that compares counterfactual cities with locations most impacted by immigration yields similar results. All in all, the results suggest that open-door policies do not harm labor markets in the host community.

Chapter 2 examines the influence of gender inequality on poverty among Syrian refugees in Jordan between 2013, the year many refugees fled the Syrian conflict and 2018. In 2013, Syrian refugees in Jordan faced numerous constraints due to their poverty, including limited access to labor market opportunities and loss of assets. However, since then, many policies aiming to boost their inclusion into the host community and their access to jobs and services have been introduced. Two waves of Home-Visit surveys, collected by the United Nations High Commissioner for Refugees (UNHCR), are analyzed to track the evolution of poverty among Syrian refugees in Jordan. To compare changes in poverty between female and male-headed households, we use relative comparisons of deciles in the expenditure distribution, and quantile regressions. We find that spending distribution has shifted over time, negatively affecting female-headed households. In 2013, female-headed households below the median had lower expenditure than male-headed households. In 2018, this pattern occurs in all deciles. Relative comparisons also shed light on

the changes overtime of poverty patterns. We find small differences between poverty rates of female and male-headed households using both the standard poverty measure and one that adjust for possible economies of scale. Regardless of the poverty measure, the poverty gender gap has increased over time with female-headed households experiencing poverty more intensely. Female single caregivers remain at the most risk of falling into poverty when compared with other types of households and over time. Our approach can help policymakers design more effective programs of assistance that respond to gender-based differences in vulnerability to poverty and find durable solutions for displaced populations.

Finally, Chapter 3 analyzes the unintended effects of policy-making in the education sector. Choosing the best candidates for a job is not an easy task. Some employers have implemented pre-employment testing as a filtering system to ensure that only qualified candidates are hired. However, this mechanism can also discourage high-qualified workers from applying and harm productivity. This paper assesses the impact of implementing a pre-employment test for teacher candidates aspiring to be employed in Colombian public schools on students' test scores. A national standardized test was used as a requirement to be hired and promoted. Candidates were evaluated on verbal and numerical reasoning as well as on subject-area topics and pedagogy theory. Using a difference-in-differences model that uses school-level and time variation on the share of teachers hired through the pre-employment test process, I evaluate changes on students' tests scores one and eight years after the reform. I find that conditioning hires on teachers' test scores did not affect student outcomes one year or eight after the pre-employment test was put in place. Some models show negative but non-significant and very close to zero effects that begin to appear in years 6 and 7 after the first teachers are hired. I also find a small worsening in the test scores of girls relative to boys, partially explained by the relative decrease in the hiring of female teachers. The latter is possibly associated with a worse test performance of female candidates. The results do not necessarily indicate that pre-employment tests do not work, but they do suggest that policy makers should be careful when implementing these types of reforms because poor test design can prevent the hiring of good candidates.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# ‘When a Stranger Shall Sojourn with Thee’: The Impact of the Venezuelan Exodus on Colombian Labor Markets

## 1.1 Introduction

In light of large waves of displacement worldwide, there is considerable speculation about the effects of migration on host communities. Approximately one in three people around the world believe migrants take jobs that native workers want (Gallup, 2015)[1]. This issue, among other concerns regarding migrants, has been at the center of media attention and political debate for years. With 70.8 million forcibly displaced people worldwide (UNHCR, 2022b) and nearly 201.3 million international migrants of working age (IOM, 2020), policymakers worry about the effects of open-door policies on local labor markets, particularly as the numbers keep rising. The impact of the presence of economic migrants has been widely studied, often yielding mixed results (Basso et al., 2019; Brücker and Jahn, 2011; Venturini and Villosio, 2006). However, economic migration is gradual and predictable, and is driven by local economic conditions (Peri and Yasenov, 2019). For this reason, studying sudden and unexpected inflows of displaced populations in contexts of open-door policies may provide a cleaner source of exogenous labor supply shocks.

This study examines an unexpected migration-induced supply shock to evaluate its effects on wages and employment in the host community. In particular, this paper analyzes the recent and unprecedented surge of migration from Venezuela and the open-door policy Colombia has put in place to draw its conclusions. Three political events in Venezuela are associated with the unleashing of the current migratory crisis that started in July 2016 and peaked in mid-2018. Figures reported by the International Organization for Migration (IOM) and the United Nations High Commissioner for Refugees (UNHCR) estimate that over four million Venezuelans have fled the country since the crisis started, 80% of whom have chosen Latin American countries as their main destination. The referred entities also estimate that Colombia has welcomed around 1.8 million immigrants, which represents about 3.6% of Colombia's population and 7.2% of the labor force[2].

One of the main challenges of conducting research using unexpected migration waves is to identify where migrants settle. Most refugees remain undocumented for long periods of time, making it difficult to track them through official records. For example, when the Venezuelan exodus started, the Colombian government put in place registration centers to provide migrants with

[1] Analysis using 142 countries (2012-2014)

[2] Calculations by the author using data published by the Ministry of Labor

temporary residence and work permits. However, migrant records are not a reliable source of data to identify where migrants settle. There are several reasons for this lack of reliability. First, the number of Venezuelans in the official registry is about half the amount estimated by IOM and UNHCR. This underestimation can be partially attributed to the voluntary nature of the registration process. Second, the number of records may overestimate the number of migrants located near the border or in the biggest cities; and, third, registering in one city does not necessarily imply that the migrant will enter the labor market in said location.

To identify where migrants settle, I use geographical variation in the Internet search intensity of keywords that Venezuelans are more likely to use compared to Colombians. Those keywords include 'Venezuelans in' given that new migrants are more likely to look for their fellow-countrymen communities; and 'PTP' or 'PEP', which stand for the residence and work permits granted by the government to allow Venezuelan migrants to enter Colombia. Compared to official records, this index suggests that migrants are not as congregated at the borders or in the main cities. For example, while official records indicate that about one third of the migrants have settled in Bogotá, the Internet search index suggests that merely one fifth have entered that market. A similar analysis using the US data also confirms that this type of text analysis of online interactions is promising at identifying where migrants settle, especially considering undocumented migrants. One of the reasons is that it is relatively easy and low-cost to conduct Internet searches given the easy access to free Wi-Fi zones or internet cafes. In addition, Google searches can identify the density of internet hits even in remote areas with a sufficiently identifiable number of searches.

Combining the time and geographical variation in the Internet search index, individual information on wages and employment from the Colombian Labor Market Survey, and a difference-in-differences design, the findings reveal negligible changes in wages in both formal and informal sectors due to migration-induced supply shocks. If anything, there are mild reductions in wages of natives working in occupational labor markets in which migrants are entering disproportionally more, such as elementary occupations, services, and clerical jobs. The results suggests that, in the worst-case scenario, a one-percentage point increase in the migrant labor supply will result in a decrease of up to 5 cents per dollar. Regressions also suggest a precisely estimated zero change in employment.

The results of the model are key to assess the average effect of migration at the country level. However, one might be concerned about the impact that migrants may have had in cities characterized by relatively stronger migratory flows. To assess the impact of migration at the city level, I construct artificial counterfactual (ArCo) cities using an elastic net model to train the pre-treatment data and predict the post treatment period to make inference [3]. Comparing each metropolitan area that had a relatively large influx of migrants with its respective ArCo city, the results are consistent with the country-level effects. That is, there are mild reductions in wages and null effects on employment. Altogether, the findings support the idea that opening borders and allowing immigrants to enter a country freely do not damage the labor prospects of their native population.

This paper makes contributions to two aspects of the literature on migration. On the one hand, although many highlight the advantages of using episodes of forced displacement as a natural experiment to exploit exogenous migration flows (Borjas, 2003; Card, 1990; Del Carpio and Wagner, 2015; Friedberg, 2001; Bahar et al., 2020a), data on the geographical location of new refugees and displaced populations is scarce (Ruiz and Vargas-Silva, 2013). Thus, my first contribution is to introduce the use of text analysis of Internet searches to identify the location of immigrants. This method is especially promising in contexts where surges of immigration are unexpected and where tracking migrants is costly. Similar approaches are currently being developed, such as the use of social media to identify and locate migrants (Martin and Singh, 2018; Palotti et al., 2020). Beyond the purpose of developing better indexes to obtain more precise estimates, these strategies may serve policymakers to improve targeting of aid programs for immigrants and natives.

Second, the literature on migration has largely focused on the effects of south-to-north migration. This type of migration is frequently characterized by the arrival of imperfect substitutes to the local labor force, mainly due to differences in the native languages of immigrants and natives. Peri and Sparber (2009) found that the downward pressure on wages due to the arrival of migrants is partially alleviated by this imperfect substitutability. They find that immigrants specialize in occupations intensive in physical labor, while locals will reallocate to language intensive jobs. But

[3] This design aggregates data at the regional level to construct a high-dimensional panel time-series dataset. Then, it uses an elastic net to construct the artificial counterfactual following the methodology by Carvalho et al. (2018).

what happens when both population groups share the same language and when substitutability between foreign and native workers is larger? With 85% of the displaced population worldwide being hosted in developing countries (**?**) and 79% settling in neighboring countries that may share the same language, this highly relevant question may be answered by analyzing south-to-south migration. Thus, this study provides evidence that, even when workers are more substitutable, there are still null effects on wages and employment caused by migrants.

The remainder of this paper is as follows: in Section 2, I provide a summary of the unfolding of the events that led to the Venezuelan exodus and briefly describe the characteristics of the migrant population. In Section 3, I describe the data and the identification strategy. In Section 4, I provide a summary of the results. In the last section, I provide some final insights and policy implications of this research.

## 1.2 The Venezuelan Exodus

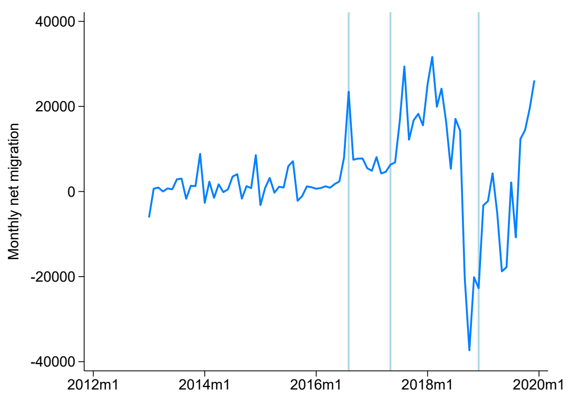
Chavez's presidential term ended with his death in 2013, leaving Venezuela without a strong private sector, with a weakened oil industry, with oil exports committed to repaying debts, with limited cash flow (Hernandez et al., 2016), with crises in its diplomatic relations with the neighboring countries (Romero, 2008), and with an opposition party that was increasingly gaining popularity (Lopez and Watts, 2013). In April 2013, Nicolás Maduro assumed power after being elected president by a narrow margin. Once in power, Maduro displayed further abuse of power, and the economy of Venezuela showed signs of a growing economic crisis.

In 2015, a diplomatic tension between Colombia and Venezuela grew due to an alleged presence of a Colombian armed group in Venezuelan territory. In retaliation, Maduro declared a state of exception and announced the closure of borders. Said closure, which was initially put in effect only in the State of Táchira, banned any transaction of goods and migratory flows. The border re-opened a year later, on July 6, 2016, in response to massive protests in Venezuelan border cities. Figure 1.1 associates the first peak in net migration, defined as inflows minus outflows, to this event.

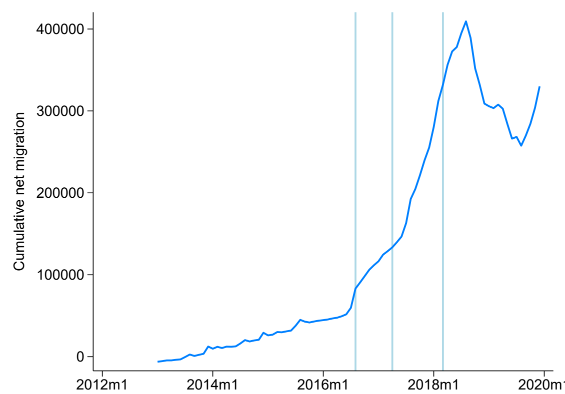
During the year when the border remained closed, inflation, poverty rates, and food insecurity grew dramatically in Venezuela. In 2016, a Venezuelan household survey reported that 93.3% of households considered their earnings as insufficient to cover food expenditures, and 72.7% of individuals reported weight loss in the last year. Poverty rates were skyrocketing, with 81.8% of households in poverty and 51.5% in extreme poverty. Violence was also problematic for many households: 94% of households declared violence was on the rise. (ENCOVI, 2016). Thus, one can associate the economic situation's worsening, increases in violence, lack of access to services and hyperinflation to the start of a massive migratory wave in 2016. The Venezuelan exodus reached its peak in mid-2018. The controversial re-election of Maduro can be associated with this peak.

Figure 1.1: Net Migration of Venezuelans in Colombia

Monthly Net Migration




Cumulative Net Migration




*Notes:* Own calculations using data from Migración Colombia. The graph on the left shows monthly net migration, while the graph on the right shows cumulative net migration. As depicted in the graph, the first peak in net migration flows take place in July 2016 after the re-opening of borders. The light blue vertical lines trace the events described in Appendix A1.

This wave of migrants, which included Venezuelans as well as Colombian returnees, entered Colombia by land through three entry points: the Simón Bolivar International Bridge (Villa del Rosario/Cúcuta), the Páez Bridge (Arauca) and the Paraguachón International Bridge (Maicao), the former being the main entry point. Upon reaching the Colombian border, Venezuelans walked, hitchhiked, or took buses to their destination. Colombian authorities allowed Venezuelans to enter freely[4], even accepting expired passports as proof of identification (Ministerio de Relaciones Exteriores, 2019). Nevertheless, lack of appropriate documentation forced many to enter through irregular paths and become undocumented migrants.

The Colombian government implemented a number of policies to facilitate the integration of migrants. For example, they created a special residence permit, called PEP (Permiso Especial de Permanencia) for the exclusive use of Venezuelan nationals. Application to this permit was simple, with zero application fees involved, and with only minor requirements such as not being a convict and having entered through an official border-control site. This virtually allowed all legalized Venezuelans to work and access the healthcare system (World Bank, 2018). This permit was valid for 90 days and could be renewed every 90 days for up to 2 years. Recent initiatives from the bureau of migration include the expedition of new permits through an online renewal process, since most permits have expired or are about to expire.

Given the large numbers of undocumented migrants entering Colombia through irregular routes, at the peak of the crisis (between April and June 2018), the Colombian government set up registration centers at the borders and in large metropolitan areas. The objective was to build an administrative dataset of immigrants (RAMV) that would eventually grant them access to a PEP. Every Venezuelan seeking to temporarily or permanently reside in Colombia was encouraged to register in person with any identification document certifying their country of origin.

Despite these efforts, only one-third of Venezuelans in Colombia are currently documented and hold a residence or work permit, while two-thirds remain undocumented (Migración Colombia, 2020a). In this context, there are two reasons why official records may lack the precision to determine where undocumented migrants settled. First, considering that registration offices put in place for undocumented immigrants were mainly located at the borders and in large cities, the official records may underestimate Venezuelans' presence in other cities of the country. Second,

[4]Venezuelans are allowed to enter Colombia using different types of documents. For example, people living at the border can use a border mobilization card (or Tarjeta de Movilidad Fronteriza) that can be requested online with a fee of about 5 dollars. This card allows Venezuelans to enter the country and stay for up to 7 days. They can also enter as visa holders if they have the sponsorship of a company (Universidad del Rosario, 2020).

most immigrants requesting residence permits often end up in jobs in the formal sector. Given that most formal jobs are located in metropolitan areas, data coming from official registries might overestimate Venezuelans' congregation in these cities.

## 1.3 Data and Descriptive Statistics

### 1.3.1 Labor Market Data

Individual information on wages and employment as well as the characteristics of the native and immigrant population come from the Colombian Great Integrated Household Survey (or GEIH, for its acronym in Spanish) designed by DANE (Departamento Administrativo Nacional de Estadistica). This survey also provides detailed demographic information of Colombian households including age, gender, years of education, municipality or department of residence, type of job, hours worked, labor earnings, and firm size. The GEIH records information across 23 departments, 13 metropolitan areas and 11 intermediate cities [5]. This paper uses the sample of monthly information from January 2013 to December 2019. This repeated cross-sectional dataset reports data on around 52,000 people per month, which, for the time frame of the analysis, sets the number of observations at 3,674,040.

To identify the nonimmigrant population, which is the focus of this study, as well as the characteristics of Venezuelan migrants, this study uses the migration module of the GEIH. Since 2012, this additional module has been applied to all the respondents, recording information of the place of birth, changes in the place of residence in the last year and over the last five years, and the motives for the change of location. Given the relatively recent nature of migration in Colombia, the sample of migrants in this module was not representative of the population before late 2017. This module became publicly available in September 2019 after a sizeable sample of immigrants was achieved (DANE, 2019)[6].

[5] Colombia has 32 departments and 1023 municipalities

[6] At the time of publication, the Bureau of Statistics - DANE - warned that data on employment of migrants should be interpreted carefully and only using moving averages of the last year. For that reason, this study focuses exclusively on the effects of the native population and does not use this data to identify the location of immigrants. Also see https://www.facebook.com/DANEColombia/videos/3366521323562668

Table 1.1: Descriptive statistics of the nonimmigrant and immigrant sample after July 2016

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | | Venezuelan | | Colombian | |
| Variables | Nonimmigrant | | immigrant | | Returnee | |
| | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female (%) | 51.24 | 49.98 | 51.12 | 49.99 | 49.94 | 50.01 |
| Age | 36.78 | 14.17 | 29.29 | 10.31 | 38.84 | 13.25 |
| *Education (%)* | | | | | | |
| None | 3.61 | 18.64 | 1.07 | 10.28 | 4.38 | 20.46 |
| Primary | 21.32 | 40.96 | 8.87 | 28.43 | 32.26 | 46.75 |
| Secondary | 48.13 | 49.97 | 63.16 | 48.24 | 54.36 | 49.82 |
| Higher education | 26.94 | 44.37 | 26.90 | 44.35 | 9.01 | 28.63 |
| *Labor market* | | | | | | |
| Labor Force (%) | 80.26 | 39.81 | 92.10 | 26.98 | 80.32 | 39.77 |
| Employment (%) | 91.47 | 27.93 | 81.57 | 38.77 | 81.12 | 39.14 |
| Informal (Firm size - %) | 66.50 | 47.20 | 88.68 | 31.69 | 88.53 | 31.87 |
| Informal (Social security - %) | 59.75 | 49.04 | 95.77 | 20.13 | 88.10 | 32.39 |
| Usual hours worked | 44.67 | 16.12 | 49.61 | 17.63 | 45.56 | 17.47 |
| Hourly earnings (COP) | 6148.94 | 5649.75 | 4000.97 | 2791.51 | 3705.66 | 2579.73 |
| Observations | 2,592,639 | | 16,008 | | 6,320 | |
| % of the sample | 99.15% | | 0.61% | | 0.24% | |

*Notes:* This table presents descriptive statistics for the sample of analysis. Data come from the Colombian Labor Market Survey - GEIH. The statistics are computing using individuals observed after July 2016, the month where the re-opening of borders took place. A description of the labor market variables is available in Appendix A. The table presents weighted means and standard deviations of selected variables for three groups: Nonimmigrants (columns 1-2), Venezuelan immigrants (Columns 3-4), and Colombian returnees (Columns 5-6).

This paper distinguishes between three groups of population. First, nonimmigrants, that is, individuals who were born in Colombia and who have not changed the municipality of residence in the last five years. This group is the center of analysis in this paper since it is the most susceptible to migration-induced supply shocks. Second, Venezuelan immigrants, who were born and whose place of residence over the last year was Venezuela. Third, Colombian returnees, who were born in Colombia and whose place of residence in the last year was Venezuela. Information on the latter two groups is used to understand the characteristics of the supply shock, while measuring the effects of migration in these groups is beyond the scope of this paper.

Table 1.1 displays descriptive statistics of nonimmigrants, Venezuelan immigrants and Colombian returnees following the migration crisis that started in July 2016. The table shows that sex ratios of immigrants are not different from the ones of the native population. However, the average Venezuelan immigrant is 7.5 and 9.5 years younger than nonimmigrants and returnees, respectively. The latter is consistent with Becker's theory that younger individuals are more likely to migrate because their lifetime-expected benefits are larger, given a greater estimated duration of stay in the host county (Becker 1964).

About 26% of immigrants and nonimmigrants have completed a bachelor's degree, indicating that both groups have similar levels of education. This, along with the fact that Venezuelan immigrants have the same native language as Colombians, suggests that labor supplied by these two groups might be highly substitutable. On the other hand, Colombian returnees seem to be slightly less educated, with only 9% having accessed higher education.

In terms of access to labor markets, relatively more immigrants are in the labor force. That is, among individuals aged 15-65, 92.1% of immigrants are either employed or searching for a job while the equivalent figure is 80.26% for the average Colombian. On the other hand, unemployment among immigrants and returnees is about 9 percentage points higher than unemployment among Colombians. This suggests that, although relatively more immigrants are willing to work, it is more difficult for them to find a job.

Venezuelan immigrants are mostly employed in informal jobs according to two definitions of informality: the Bureau of Statistics classifies a worker as informal if they work for a firm with fewer than 5 employees; the second definition, more broadly used in the literature of informality, classifies a worker as informal if they do not have a healthcare plan through their employer or do not have a retirement plan. Table 1.1 also suggests that immigrant workers work slightly longer hours for a lower rate per hour. This is indicative of the fact that Venezuelans not only have difficulty finding jobs, but when they do find employment, they access jobs of lower quality.

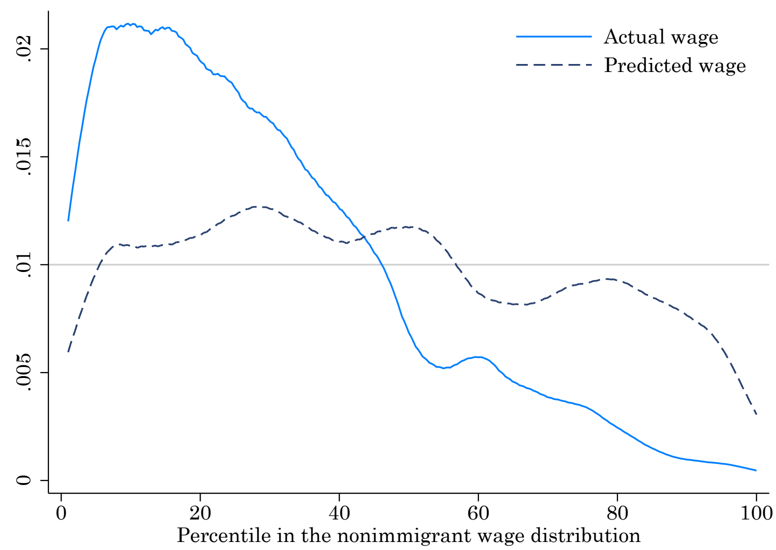
Overall, the descriptive statistics point toward a downgrade in the returns immigrants receive for their skills. 'Downgrading' occurs when immigrants with the same measured skills as natives, say with the same education and experience, receive significantly lower earnings. Dustmann et al.

(2016) and Dustmann et al. (2012) use the cases of Germany and the US to illustrate that downgrading usually takes place in the years immediately after immigrants arrive and fades away in the long run after the assimilation. These authors point to an association between downgrading and the lack of formal requirements or complementary skills, such as fluency in English. In the presence of downgrading, assuming that immigrants compete only with natives in the same observed education-experience cell tends to produce more negative biased estimates.

To show the degree of downgrading of Venezuelan immigrants, Figure 1.2 depicts kernel estimates of the actual and predicted densities of immigrants in the non-immigrant wage distribution. The graph shows where immigrants are currently located and where we would assign them if they received the same return to their education and potential experience as natives. If there were no downgrading, the light blue line and the dashed dark blue line would overlap.

The results are striking. Venezuelan immigrants are predicted to be significantly more concentrated below the 50th percentile and underrepresented in the middle and upper ends of the nonimmigrant wage distribution. Given that Colombians and Venezuelans speak Spanish, we cannot necessarily associate this downgrading with the lack of language skills. However, anecdotal evidence suggests that Venezuelan immigrants face high costs and long waits to validate their past educational credentials (Universidad de Antioquia, 2020). This is perhaps the reason why they are entering more service-related and elementary occupations (see Figure A1).

Figure 1.2: Downgrading of Venezuelan immigrants




*Notes:* This graphs shows evidence of the downgrading of Venezuelan immigrants in Colombia by presenting kernel estimates of the actual (light blue) and predicted (dashed dark blue) density of immigrants in the nonimigrant wage distribution. The prediction can be interpreted as the wages immigrants would receive if their returns to education and experience were the same as that of the nationals. The horizontal line shows as a reference the nonimmigrant wage distribution. The kernel estimates above the horizontal line are where most immigrants are concentrated.

Downgrading of Venezuelans suggests that they exert more pressure at the low end of the nonimmigrant wage distribution. Thus, using Borjas (2003) approach also known as the national skill-cell approach, which relies on the relative density of immigrants in education-experience cells, is not suitable for this context. Instead, using a pure spatial approach that exploits geographical variation of immigrants to evaluate its impact among low-income native workers would provide better estimates. Section 5 will further explain this empirical approach, which will allow me to evaluate the true impact of the Venezuelan migration on Colombian labor markets.

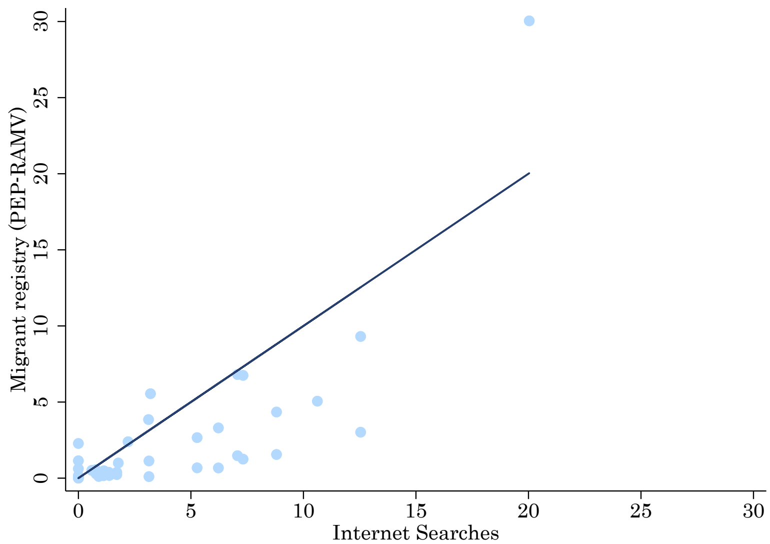
### 1.3.2 Geographical Density of Immigrants

New unexpected influxes of migrants are often thought of in the literature as exogenous shocks to the labor supply, however, tracking the location of new immigrants is challenging. For Colombia, which barely has any experience receiving large waves of immigrants, tracking Venezuelans was the case. As mentioned in Section 2, two initiatives were implemented to this end. On the one hand, the registry of residence permits (*Permiso Especial de Permanencia* - PEP), which collects individual information on those who voluntarily requested the permit. On the other hand, the Census of immigrants (RAMV), which was conducted through encouraging every Venezuelan seeking to temporarily or permanently reside in Colombia to register in person at any registration center. Compared to the former, registration to the RAMV targeted mostly undocumented immigrants who did not have a PEP.

By May 2020, these administrative registries have gathered data on about 784,234 Venezuelans out of 1,809,782 estimated to have settled in the country (Migración Colombia, 2020b). There are multiple reasons why these official records may lack the precision to determine where immigrants settled. First, official records may overestimate immigrants' density at the borders and in large metropolitan areas given that registration offices were mostly located in these regions. Second, given the registration's voluntary nature, only those who had a higher likelihood of getting a formal job might have registered at a higher rate. This may also point to an overestimation of immigrants' density in large cities, where most formal jobs are located. Third, having a record of the place of registration does not necessarily mean the immigrant will enter that geographical labor market. For these reasons, an alternative measure is used in this paper.

To identify where immigrants settle, this paper uses monthly geographical variation in the Internet search intensity of keywords that Venezuelans are more likely to use. Those keywords include 'Venezuelans in' given that new migrants are more likely to look for communities from their own country; or 'PEP', which is the residence and work permit created exclusively for Venezuelan immigrants. For each month after July 2016, when the first peak of positive net migration was recorded, this index assigns 100 to the location with most keyword searches. Every other area is assigned a number between 0 and 100 depending on the geographical density of hits for that keyword. Then, the index is transformed to reflect the share of searches that took place in region r. Thus, each month, the transformed index adds up to one and reflects the percentage of immigrants entering the labor market r at a point in time t. Using the national estimate of Venezuelan immigrants age 15-65 in Colombia, one can compute the number of immigrants entering each labor market. Finally, for each region at each point in time, I compute the size of the immigrant population as percentage of the local population in the labor force[7].

Figure 1.3: Comparison of Official Records and Internet Searches




*Notes:* This graph compares the geographical distribution of immigrants obtained from two sources: official records (PEP-RAMV) and Internet searches. Official records are computed using publicly available data on PEP-RAMV. I aggregate individual records to obtain number of people in each region. Then, I compute the contribution of that region to the pool of immigrants. In the case of the Internet Search index, the geographical distribution of immigrants is computed using the procedure explained in Appendix B. If both sources of data provided a similar distribution of immigrants, all the light blue dots would be placed along the dark blue line. The kernel estimates above the horizontal line are where most immigrants are concentrated.

Figure 1.3 compares immigrants' geographical density drawn from the official registry with the one based on Internet searches. If the geographical distribution of immigrants according to both measures was the same, all dots would be located along the dark blue line. However, this is not the case. According to official records, Bogota hosts nearly one-third of the immigrants. In contrast, the Internet search index shows that about one-fifth are located in the capital city. Such a large difference has substantial implications in determining the share of immigrants that other regions in the country may host. While official records put a greater weight on large metropolitan areas and border cities (see A5), the internet search index indicates that immigrants are located in other areas

[7]See Appendix A2 for more details regarding this transformation

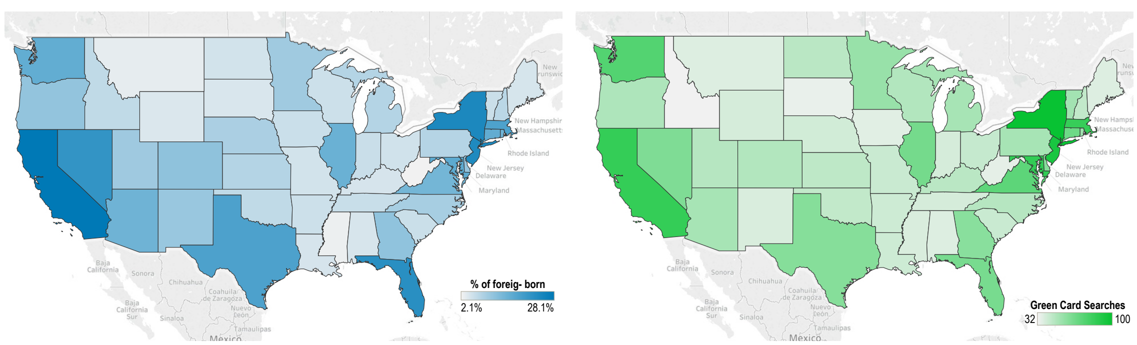
too (see A4).

To determine whether the geographical intensity of Internet searches is a good measure to approximate the density of immigrants, we can apply the same algorithm in another context. The long history of immigration to the United States, along with their experience studying and collecting data on immigrants, makes this country the right subject for a robustness check. Figure 1.4 compares the number of foreign-born individuals in US states as a share of the total population (in blue) and the Internet search intensity index for the pair of words ‘Green Card’ (in green). The geographical distribution of immigrants according to both, the official records in the US and the search index is very similar. The correlation between these two measures is 0.87. Thus, in a country with relatively more accurate data on the geographical density of immigrants, the Internet search measure seems to capture most of the variation.

Figure 1.4: Application of the Internet Search Index to the US Case

Share of Foreign Born

Internet Searches




*Notes:* This graph compares the regional distribution of two measures. The graph on the left (in blue), uses information from IPUMS international to compute the share of foreign-born population in each state. The graph on the right shows the distribution of the Internet search intensity for the pair of words ‘Green Card’. The correlation of these two distributions is 0.87, which imply, a very good overlap between both distributions.

Thus, the Internet search index for keywords that immigrants are more likely use has applications beyond the case of Colombia. This measure can be particularly useful in contexts of forced displacement or unexpected migration waves where tracking immigrants becomes a challenging and costly task. The development of this type of measure might allow policymakers to get timely updates on the movements of displaced populations and allocate humanitarian aid accordingly.

## 1.4 Empirical Strategy

Dustmann et al. (2016) classify studies of migration into three categories. First, the national skill-cell approach, which exploits the relative density of immigrants in school-experience cells. Given that this approach often includes education fixed effects, it usually estimates the impact of immigration on more experienced local workers relative to workers with less experience. The main problem with the approach is the assumption that immigrants enter markets appropriate to their school-experience classification. In other words, it assumes no downgrading of immigrants' skills set. The second design is a pure spatial approach that exploits geographical variation on the density of immigrants. In the presence of downgrading, this approach estimates the total effect of immigration for local workers in certain education-experience group. For this reason, this model is more appropriate to study new waves of migration. Finally, the third approach is a mixture approach that exploits the relative density of immigrants in school-experience-location cells. This approach shares the disadvantages of skill cell-approach.

This paper studies the sudden and unprecedented nature of the Venezuelan exodus to evaluate its effect on the labor markets of the host community using a pure-spatial approach. Using individual data on wages and employment, a difference-in-differences design exploits the geographical variation in exposure to migration-induced supply shocks and the timing of events [8]. The analysis focuses on the host community that has secondary completed or below. The evidence of downgrading of immigrant credentials shown in Figure 1.2 along with the low share of the host and immigrant community accessing higher education, support the idea that this segment of the population is more likely to be affected. The model is as follows:

$$Y_{i,rt} = \beta M_{Ven,rt} + \eta_1 M_{Ret,rt} + \eta_1 M_{Int,rt} + \phi' X_{\{i\}} + \alpha_r + \tau_t + \kappa t + u_{i,rt} \tag{1.1}$$

[8] There are 30 departments collected in the data plus 13 metropolitan areas. From here on, the word 'regions' will refer to both, departments and metropolitan areas. Thus, 43 regions are having different levels of exposure to the supply shock.

where $Y_{i,rt}$ is the outcome of analysis for person $i$, observed only once in location $r$ at time $t$; $T_0$ is a dummy that takes the value of 1 after July 2016, i.e., after the reopening of borders[9]; $m_{Ven,rt}$ is the size of the migration shock in $r$ at time $t$; $m_{Ret,rt}$ controls for the migration shock caused by Colombian returnees; $m_{Int,rt}$ controls for migration shocks that the entry/exit of locals might generate in certain regions; $X_{\{i\}}$ is a vector of individual characteristics that include potential experience, potential experience squared, years of schooling, gender, a dummy for whether the individual is a part-time worker, and the logarithm of the aggregate labor force.; $\tau_t$ is a set of monthly dummies that control for seasonal patterns, $\alpha_r$ are regional fixed effects, and $t$ is a time trend. The coefficient of interest, $\beta_1$, is an estimate of the average impact that Venezuelan immigrants generate in Colombia.

The use of a difference-in-differences model imposes the assumption of *parallel trends*. That is, had not the migration shock happened, trends on wages and employment would have been similar across regions. Using an event study[10], I evaluate whether there are differences in wages and employment between regions with a high influx of immigrants and those that received relatively few. In particular, I classify all regions into two groups: those that received an immigration shock above the median and those below the median. The event study shows the differences between those two groups every quarter. The omitted category is the second quarter of 2016, right before the reopening of borders.

The results of the event study, depicted in Figure 1.5, show that there are no significant differences in trends between the regions that received a high influx of immigrants with those that did not. Joint coefficient tests with clustered as well as bootstrapped standard errors confirm that differences between regions are not significantly different from zero before the re-opening of borders for both, wages and employment in the formal and informal sectors. Thus, the events study provide

[9] Recall that I referred to this event as the first wave of migration.

[10] To analyze the dynamic impact of migration and the parallel trends assumption, an event study is conducted using the following specification:

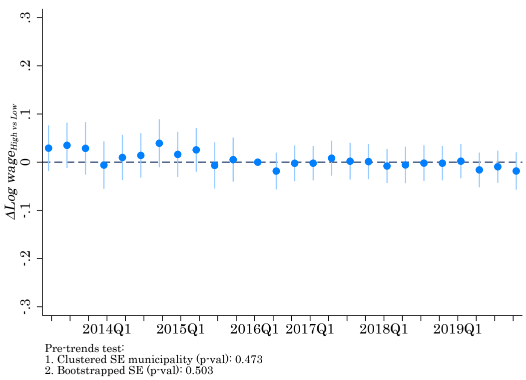
$$Y_{i,rt} = \alpha_r + \tau_t + \sum_{t=2013q1}^{2019q4} \gamma_t \ \tau_t \mathbb{1}[M_{Ven,r} \geq \tilde{M}] + u_{i,rt}$$

, where $\mathbb{1}[M_{Ven,r}|t \geq \tilde{M}]$ represents a dummy that takes the value of 1 if the average migration shock in the post-treatment period in region $r$ is above the median $\tilde{M}$, and 0 otherwise. $\alpha_r$ and $\tau_t$ are region and time fixed effects, respectively.

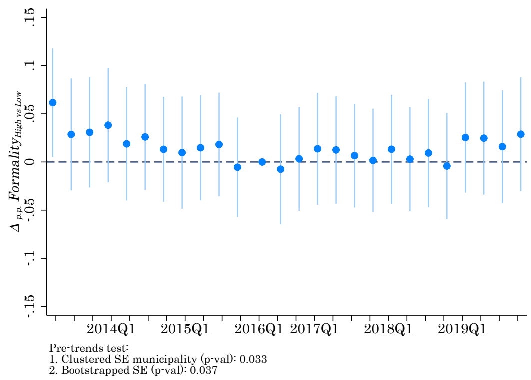
evidence supporting that trends in the growth of wages and employment were similar regardless of the shock the region received.

Figure 1.5: Event Study Comparing Regions with Migration Shocks above and below the Median

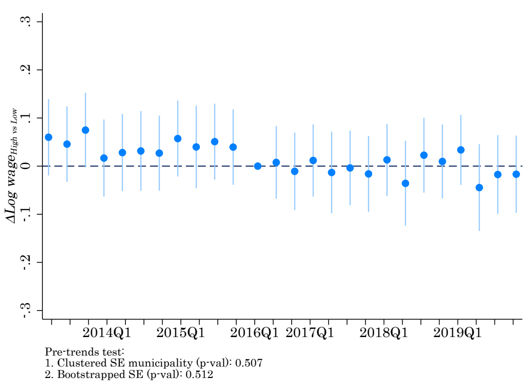
Panel A. Formal Log Wages



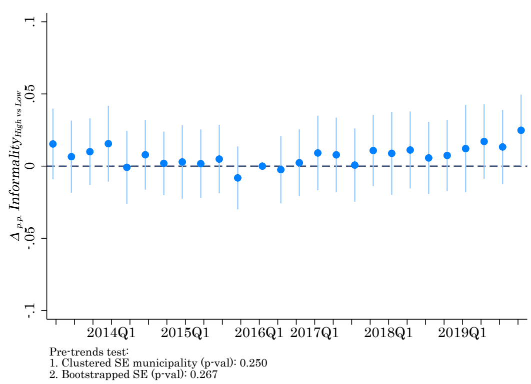

Panel B. Formal Employment




Panel C. Informal Log Wages




Panel D. Informal Employment




*Notes:*These set of graphs present quarterly estimates coming from an event study that compares the respective outcome in regions with high and low inflows of immigrants. Each point corresponds to one coefficient that measures this difference at each quarter. The vertical line is a 95% confidence interval. The black horizontal dash line is set at 0. The point of reference, $T_0$, is where the blue dot without a confidence interval is located: in the second quarter of 2016, right before the re-opening of borders. Equality in pre-treatment trends are tested using a joint coefficient test of the coefficients before the second quarter of 2016. The first tests uses clustered standard errors at the region-quarter level, while the second tests uses bootstrapped standard errors. Panels A and C display the results of the event study in formal and informal wages, respectively. Similarly, Panels B and D depict estimated coefficients in formal and informal employment, respectively.

The pure-spatial approach estimates the national average of the local wage and employment elasticities to immigration. But one might also be interested in computing the size of the effect in region r. In particular, one would be interested in knowing whether the region with the highest influx of immigrants had effects consistent with the national impact. The main challenge is that we do not observe what would have happened with region r in the absence of the migration-induced supply shock.

To compute city-level effects, I construct artificial counterfactual cities for each of the 5 metropolitan areas with the highest influx of immigrants following the methodology by Carvalho et al. (2018). Each of these cities is compared with its artificial counterfactual in the post-treatment period to make inference. This methodology requires high dimensionality in the data. For this reason, data is aggregated at the regional level to construct a time-series panel with monthly frequency. The dataset used in this methodology has $r_{low}+1$ regions, the donors and the treated unit, observed $T$ times.

The Artificial Counterfactual (ArCo) methodology follows two steps:

1. Suppose $y_t$ is the outcome we are evaluating, say wages or employment. For the unit of analysis $r$, we do not observe what would have happened had it now received a high influx of immigrants, that is, we do not observe $y_t^0$ for $t \geq T_0$, where $T_0$ is the time the unit started to receive the migration-induced labor shock. However, we can predict that scenario using the first $T_0 - 1$ observations of the time series and the following model: $y_t^0 = \mathcal{M}_t + \varepsilon_t$, where $\mathcal{M}$ is an elastic net in this paper but could be any measurable mapping. In other words, this step computes an artificial treated region in the pre-treatment period using the pre-treatment observations of the donor units.

2. Once we have computed the first step, we can compute the counterfactual in the post treatment period $\hat{\mathcal{M}}_t$ for $T_0, ..., T$.

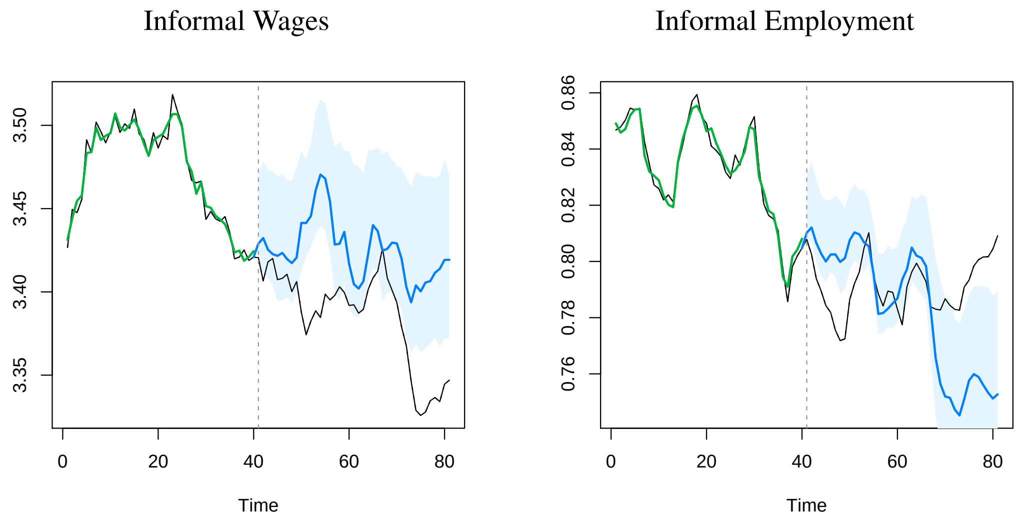
The ArCo estimator is then

$$\hat{\Delta}_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} (y_t - \hat{y}_t^0) \tag{1.2}$$

Although this methodology is similar to the Synthetic Control methodology pioneered by Abadie and Gardeazabal (2003), the ArCo method offers two key advantages for the study of labor outcomes. First, it does not rely on a convex combination of donor observations to construct the counterfactual, which can lead to biased estimators (Ferman and Pinto, 2016). Second, it exploits time variation of outcomes, which increases the number of observations from which the counterfactual is build. On the contrary, the synthetic control usually removes the time-series dynamics because it uses time averages of the observed donors. Thus, the number of observations used to compute weights in a pure cross-sectional setting is very small.

Figure 1.6 illustrates how this methodology fits the data in Cúcuta, the largest entry point from Venezuela to Colombia. The figure depicts the actual data in black for both, informal wages (on the left) and informal employment (on the right). The green lines portray the fitting of the elastic net on the training data. Note that the training data is that of the pre-treatment period or the monthly observations before the re-opening of borders. The vertical dashed line visually marks this point in time. The figure shows that the elastic net model is able to fit the data almost perfectly. Using the results of the training data, the counterfactual is constructed and depicted in light blue along with its confidence interval at the 95%. The average difference between the light blue line and the dark blue line in the post-treatment period is the average treatment effect ($\hat{\Delta}_T$). Preliminary results from the visual inspection of this figure point towards a decrease in informal wages and an increase in informal employment of nationals in the city of Cúcuta.

Figure 1.6: Fitting of the Artificial Counterfactual to the Data in Cúcuta, the Main Entry City.




*Notes:* Both graphs display the fitting of the ArCo methodology to the data using as example the results for Cúcuta, the main border city. The graph on the left shows the results in informal wages and the one on the right in informal employment. The black line depicts the observed data. The green line shows how the elastic net fits the observed data in the pre-treatment period. The light blue line depicts the prediction of the model in the post-treatment period along with a 95% confidence interval. The vertical dashed line signals the moment in time when borders re-opened. Results for other regions are shown in Figures A2 and A3 in Appendix C.

## 1.5 Results

Table 1.2 shows the main results for the *difference-in-differences* model described in equation 1. Panel A presents effects on wages and Panel B on employment. Each column presents results coming from a linear regression that uses the sample of Colombians who have not changed their place of residence in the last year. $M_{Ven,rt}$ measures the supply shock from the influx of Venezuelan immigrants while $M_{Ret,rt}$ measures the equivalent shock from Colombian returnees. Although Table A4 presents evidence that the arrival of immigrants is not instigating Colombians to out-migrate to other cities within the country[11], this and subsequent estimations in this paper control for $M_{Int,rt}$. The literature shows that failing to control for this effect would bias the estimates towards zero

[11]Table A4 shows estimates of a linear regression of the flow of Venezuelans into city $r$ at time $t$ on the flow of internal migrants. The regression includes region and time fixed effects. The regression shows that for an increase in 1 percentage point in the number of Venezuelan immigrants as a share of the labor force, there is a displacement of 0.01 percentage points. This effect is small and not significant.

(Card, 2001; Borjas, 2003).

Table 1.2: Effects on Wages and Employment

| | (1)<br>Formal | (2)<br>Informal | (3)<br>Both |
|---|---|---|---|
| *Panel A. Wages* | | | |
| $M_{Ven,rt}$ | -0.000 | -0.006*** | -0.005*** |
| | (0.001) | (0.002) | (0.002) |
| Observations | 276,332 | 689,062 | 965,394 |
| R-squared | 0.049 | 0.091 | 0.133 |
| *Panel B. Employment* | | | |
| $M_{Ven,rt}$ | 0.002* | 0.001 | 0.001* |
| | (0.001) | (0.000) | (0.000) |
| Observations | 404,347 | 927,731 | 1,223,911 |
| R-squared | 0.143 | 0.053 | 0.038 |

*Notes:* This table presents the results of the estimation of Equation 1 using a difference-in-differences design. These regressions use the sample of Colombians who have not changed their city of residence in the past year and who have completed secondary or less. Panel A displays the effects on wages while Panel B shows the effects on employment. A description of the labor market outcomes is available in Appendix A. Column (1) presents the results in the formal sector, Column (2) in the informal sector, and Column (3) in both. Standard errors, in parentheses, are clustered at the region-month/year level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

There is a negligible reduction in wages of nationals cause by the arrival of Venezuelan Immigrants. If the number of immigrants as a share of the labor force were to increase one percentage point, the reduction would be of 0.6% in the informal sector. Colombian returnees on the other hand generate a larger impact. There is a reduction of 5.3% in wages in the informal sector of the nonimmigrant population if the share of returnees were to increase 1% percentage point. There are no effects of immigration on formal wages. The latter can be partially attributed to the fact that immigrants enter at a lower rate into formal markets, in spite of the relatively easy access to work permit. Thus, the downward pressure on wages is larger in informal than formal markers.

These close to zero effects are not abnormal in the literature of immigration that use a pure-spatial approach. Card (2001); Boustan et al. (2010); Dustmann et al. (2012)provide examples with similar effects in the United States and the United Kingdom. Part of the rational for the lack of effects on wages is related to rigidities in the short term. The presence of institutional constraints such as minimum wages and unions makes it challenging for wages to be elastic downwards. This is especially true in the low tail of the nonimmigrant wage distribution, which is the focus of this paper.

Turning to the effects on employment, the results suggest null effects. If anything, for every increase of one percentage point of the immigrant population there is an increase in formal employment of 0.2 percentage points. However, this effect is only significant at the 10% level of confidence. There is no significant effect on informal employment. The large number of observations imply this is a very precise null effect. This effect is not uncommon in the literature either. On the contrary, papers that find adverse effects on employment caused by migration-induced supply shocks usually fail to define the appropriate labor market that immigrants enter (for instance, those that use the pure national education-experience cell approach).

Two main reasons might be driving the results on employment: First, there is a relatively sizable segment in the literature of migration that shows immigrants are risk takers. This characteristic makes them more prone to be entrepreneurs and innovators (Bahar et al., 2020b; Bernstein et al., 2018). If immigrants are becoming entrepreneurs upon arrival or becoming self-employed, employment of the nationals would remain unaffected. Second, the time frame of the Venezuelan exodus is not long enough to let the market react to the shock. This is a plausible explanation given that the peak was achieved in May 2018, and this paper uses data until December 2019. Future research to evaluate the medium and long-term effects of this migration wave is thus needed.

The city-level effects, which compares selected cities with their own *Artificial Counterfactual city*, draw similar results. Figures A2 and A3 depict the fitting of the elastic net in 6 selected cities/regions on the training data (in green), and the prediction on the post-treatment period (in blue). It shows that each artificial city fits the data perfectly in the pre-treatment period.

The comparison between the observed data and the artificial counterfactual identifies the treatment effect from the migration-induced supply shock. Table **??** summarizes the results. Columns (1) and (4) show the size of the impact on wages and employment while columns (2)-(3) and (5)-(6) display their confidence intervals at the 5% level of confidence. Panel A displays results on wages while panel B on employment. In terms of wages, the cities most hit by the immigration wave are the cities of Cúcuta and Bucaramanga in the informal sector, and the department of Atlántico in the formal sector. Although statistically significant, these effects are still small. In Cúcuta, the most important border city, an increase in the immigrant share of 1 percentage point leads to a decrease in hourly earnings of up to 4%. For all other cities the effect is also small or nonsignificant.

## 1.6 Heterogeneous Effects

Although the national and city-level effects are negligible, one might think that the arrival of migrants has affected specific groups of the population. To test this hypothesis, I first assess whether women are more affected by this shock. Using a variation to the model of Equation 1 in which a triple interaction is included, I find that formal and informal wages are relatively less affected in women (see Table A2). Remembering that the average effect is zero, this result reinforces the tendency to zero. Table A3 shows that the effects on employment are not different for women and men in the informal sector. However, the results in the formal sector indicate that women benefit relatively more from the migration shock by increasing their holding of formal jobs.

One can also think that younger generations are the most affected by the influx of immigrants due to their shorter experience and lack of credentials. Using a triple difference model for categories of variables, evidence is found that, compared to workers over 55 years of age, the youth are slightly more affected in terms of wages. The size of this effect is, however, small. For an increase in 1 percentage point on the share of immigrants, individuals aged 15-24 perceive a reduction in informal wages of 0.7% while an increase in formal wages of 0.2%.

As can be seen in figure A1, the majority of Venezuelan immigrants enter service and sales occupations, administrative jobs, and elementary occupations. For this reason, tables A3 and A2 present results that focus on these sectors. Each row of panel C comes from a linear model that distributes Venezuelans in the national territory according to the distribution shown in Figure A1. The final result is a regression that assumes there is a migrant-induced supply shock for each occupation-region-month cell. The results show a significant 5.3% reduction in hourly wages for those Colombians working in elementary occupations in the informal sector. The drop in wages is smaller for sales, service and clerical support workers. In terms of employment, the estimates suggest that migration has no impact on these sectors. In other words, labor demand is capable of absorbing the shock quickly.

## 1.7 Final Remarks

Large immigration waves cause concern among policymakers and nationals. According to Gallup (2012-2014), 1 in 3 people think immigrants take jobs that nationals want. Colombia has not been an exception to this rule. The recent influx of Venezuelans to Colombia has been very substantial for a country the size of Colombia.

In this paper, I studied the effects on labor markets of a massive and unprecedented immigration flow from Venezuela to Colombia. In this context, I take advantage of the timing of events, the size of the event, and an open-borders policy to evaluate how the surge of immigration has affected wages and employment in the formal and informal sectors for Colombians. But evaluating if there is an effect of immigration on wages has been difficult in the past for two main reasons: first, it is difficult to find large and unexpected (exogenous) waves of immigration, luckily, the venezuelan exodus enters into this category; and two, it is difficult to identify where migrants settle once they cross the border in the presence of documented and undocumented migrants.

To address the latter, I use the Internet search intensity of keywords that Venezuelans, not Colombians, are more likely to use. The results show that the Internet search index is more parsimonious than official records in contexts where immigration is relatively a new phenomenon. While the official records overestimate the presence of immigrants in large and border cities, the Internet search index shows immigrants are spread across the national territory. This measure seems to work in other contexts too. For example, I show evidence of the high correlation between the Internet searches of the words 'Green Card' and the percentage of foreign-born in the US States.

Combining the time and geographical variation in the Internet search index, individual information on wages and employment from the Colombian Labor Market Survey, and a difference-in-differences design, the findings reveal negligible changes in wages in both formal and informal sectors due to migration-induced supply shocks. If anything, there are mild reductions in wages of natives working in occupational labor markets in which migrants are entering disproportionally more, such as elementary occupations, services, and clerical jobs. The results suggests that, in the worst-case scenario, a one-percentage point increase in the migrant labor supply will result in a decrease of up to 5 cents per dollar. Regressions also suggest a precisely estimated zero change in employment.

Using artificial counterfactual (ArCo) cities to compute city-level effects, the results are consistent with those of the difference-in-differences model for most cities. That is, there are mild reductions in wages and null effects on employment. Altogether, the findings support the idea that opening borders and allowing immigrants to enter a country freely do not damage the labor prospects of their native population. For the main entry point, Cúcuta, there are slightly larger adverse effects.

In summary, this paper finds that the open borders policy that Colombia implemented has not generated adverse effects on the labor outcomes of Colombians. Other upcoming papers point in the same direction (see for example Bahar et al. (2020a); Tribín-Uribe et al. (2020); Morales-Zurita et al. (2020)). This paper adds to a growing wave of literature that is accumulating evidence supporting open-door policies. This branch of literature has also found that, contrary to the believe of some, immigrants do not generate increases in crime or victimize nationals (Tribín-Uribe and Knight, 2020). On the contrary, in the long term, the literature has found evidence of the benefits that the arrival of immigrants has in terms of entrepreneurship and innovation. Newly produced research have also found that places where immigrants go reduce unemployment because they cause increases in housing value and increases in consumption (Howard, 2019; Tribín-Uribe et al., 2020).

# Chapter 2

# The Impact of Protracted Displacement on Syrian Refugees in Jordan: the Evolution of Household Composition and Per Capita Expenditure

Coauthored with
Lucia Hanmer (lhanmer@worldbank.org)
Eliana Rubiano (erubiano@worldbank.org)

## 2.1 Introduction

Forced displacement causes abrupt changes in living conditions. Immediately after displacement, forcibly displaced people are often in a highly vulnerable situation due to the loss of assets, the disruption of family structures, the loss of support networks, and their limited access to labor markets and services. These changes occur within existing structures of gender inequality, and while men, women and children are all likely to be adversely affected, these changes may have different consequences for women and girls relative to men and boys. However, one might also expect that the impacts of displacement may be mitigated over time as displaced people establish new homes and networks, find work, and integrate into their new communities. Empirical research on income poverty rates among displaced populations is growing, yet the impact of gender inequality on poverty among displaced populations is rarely discussed, nor is much known about the dynamics of poverty in situations of displacement, largely because of lack of data. This study seeks to fill this gap.

Understanding how gender inequality affects spending capacity among displaced populations is important as displacement changes social structures and access to opportunities in ways that differ for women and men. For example, during displacement, household composition changes due to separation, conflict-related mortality, or widowhood. This shock often results in the formation of new female-headed households. Evidence for sub-Saharan Africa shows that female-headed households that have previously been highly dependent on male income – notably, widow-headed households, – are particularly prone to poverty, although other types of female-headed households may not be more likely to be poor than male-headed households (Brown et al., 2019). For this reason, while it is important to analyze differences in poverty risk between households with female and male heads, it is also important to understand how different routes to female headship, and the changes in the household composition more broadly, impact poverty risks and expenditure levels.

This paper explores level and distributional changes in expenditure among Syrian refugee households in Jordan between 2013 and 2018 using a gender lens. By the end of 2018, the UNHCR had registered more than 671,000 Syrians who had fled their homes and and settled in Jordan.[1] At

[1] UNHCR's data portal records 671,551 registered Syrian refugees residing in Jordan on 13 January 2019. See UNHCR (2022c) available at https://data.unhcr.org/en/situations/syria/location/36.

the end of the time period covered by this research, the UNHCR's 2019 Vulnerability Assessment Framework (VAF) population study[2] (Brown et al., 2019) revealed that 78 percent of the population of registered refugees lived below the Jordanian poverty line, and between 2017 and 2019, there has been only a 2-percentage point reduction of highly vulnerable cases.[3] The UNHCR (2018c) also reports that close to 37% of the registered Syrian refugees in Jordan are separated from at least one family member. Household structures have often been fluid. Many Syrian families sent members ahead to Jordan to settle and sometimes, after the reunification of the rest of the family in Jordan, male family members travelled on to Turkey or to Europe (UNHCR, 2018b). These separations likely affected families emotionally and economically. However, little is known about the changes in economic well-being over time or how gender inequality has shaped poverty and expenditure outcomes.

We use two waves of data from a unique household survey, the United Nations High Commissioner for Refugees (UNHCR) Home-Visits survey for 2013-14 and 2017-18. First, we create comparable measures of expenditure per capita over time. Then, we assess whether certain types of households are more likely to be below the median of the per capita expenditure distribution in each time period. The UNHCR registers households as having a male or female principal applicant (PA), and we use this classification as equivalent to male or female headship used in the poverty literature. We find that there is no significant difference in the share of female- versus male-PA households below the median in either 2013-14 or 2017-18.

However, there are important differences between male and female PA households that affect the probability of being poor. As is evidenced in other countries (see for example Brown et al. (2019)'s study of 43 African countries), we find that Syrian refugee female-PA households are smaller in size than male-PA households. Once economies of scale in consumption are considered we find that female-PA households are more likely to be poor than male-PA households, and that this gap has widened over time. Compared to other household heads, single caregivers (the vast majority of whom are women) exhibit some of the highest risks of being poor and this risk has

[2]The Vulnerability Assessment Framework for 2019 is based on analysis of a data collected from a random, representative sample of registered refugees in October and November of 2018. The sample consisted of 2,248 Syrian refugee households, which comprised 3,712 cases, or household units, and over 10,400 individuals and was fielded in October and November of 2018.

[3]A case is roughly equivalent to a household unit.

increased over time.

We also find that the distribution of expenditure changed over time, in a way that negatively affected female-headed households. The gap between male- and female-headed households grew over time, causing a deeper gap to be experienced among the poorest households. The factors that affected this trend cannot be explained by the demographic variables of the household members or by the type of family arrangement, but are possibly explained by factors exogenous to the household.

Our contributions to the existing literature are thus twofold: first, we present results on the evolution of poverty among refugees; and second, we show that identification of how poverty rates differ between households and their trends over time is highly sensitive to the choice of poverty measurement in the context of displacement. Policy design should consider these differences and consider the size and structure of the family, as well as its income and assets.

The paper proceeds as follows: in section 2, we describe the relevant context in Jordan, and section 3 provides information about the dataset. Section 4 presents our empirical strategy, and section 5 documents changes in demographic characteristics, family composition, and their association with changes in poverty for Syrian refugees living in Jordan between 2013 and 2018. Section 5 also explores the link between gender inequality and poverty by examining differences between male- and female-PA households and households with different demographic compositions, illustrating the sensitivity of the results to assumptions made about economies of scale. Finally, we consider the policy implications of our results.

## 2.2 Context

The war in Syria has unleashed an unprecedented humanitarian crisis. Since the civil war started in 2011, 5.3 million Syrians out of a population of 21 million in have fled Syria (UNHCR, 2019a). Most found asylum in neighboring countries. From the onset of the crisis Jordan has opened its doors to refugees, accepting large numbers of people who fled across borders into the country;

while some register with the UNHCR, many others do not. As of June 2022, the UNHCR reported that there are 5.7million registered Syrian refugees. About two thirds of these refugees, 3.7 million, reside in Turkey. Jordan hosts 675,000 registered Syrian refugees, which is about 12 percent of their total number (UNHCR, 2022b). Most Syrian refugees entered Jordan in 2013 when the total number of registered Syrian refugees increased from just over 120,000 in January 2013 to 576,000 and refugee numbers continued to increase as more people fled Syria up until 2016. Since then, the increases in the number of registered refugees in Jordan is predominantly due to births.

At the end of 2018, Jordan hosted an estimated 1.36 million Syrian refugees (with 665,884 registered), representing 15 percent of Jordan's total population. The majority of registered refugees live in host communities, mainly in Amman and in the northern governorates, with around 10 percent living in camps (Hashemite Kingdom of Jordan, 2020). The Government of Jordan has implemented a series of policy reforms and initiatives that have aided the inclusion of the Syrian people in Jordan, and is one of the first countries globally to pilot a development-focused refugee response program (World Bank, 2021b).[4] However, little is known about the evolution of poverty over time among Syrian refugees in Jordan.

As noted by Verme et al. (2015), Syrian refugees faced three main challenges upon arrival: (i) a limited supply of basic services in the host economy due to institutional capacity constraints; (ii) limited access to the labor market and other economic opportunities due to skill mismatches between the Jordanian labor demand and Syrians' qualifications; and (iii) risk of human capital depletion due to low access to education and lack of use of skills.

The risk of poverty was found to be high and gendered among Syrian refugees in 2013. Hanmer et al. (2020) found that in 2013 half of the female principal applicants[5] for nonnuclear households – a household comprising a family group not composed solely of parents and their children - lived below the poverty line, compared to only one-fifth of male principal applicants for nonnuclear households. Nonnuclear households represented about 92% of all female-headed households and

[4]These reforms and initiatives include granting work permits for Syrian refugees in sectors such as agriculture, construction, retail trade and manufacturing; providing free access to schools and waiving the documentation required for enrollment; and including refugees as part of national plans on COVID-19.

[5]The principal applicant is the person who receives assistance from UNHCR for the family and is self-selected or selected by the family.

31% of all Syrian households. For widows, poverty risk was 10 percentage points higher than for couples with children. However, new households formed as a result of displacement, for example households of siblings or unaccompanied children, were by far the most likely to be poor than other types of households. For unaccompanied children poverty risk was 20 percentage points higher than for couples with children after receiving cash assistance. At the same time, there were differences in poverty risk among female-PA households — single women and single women caregivers were more likely to fall into poverty and more likely to depend on cash assistance than female-PA households who were married and/or had other male adults in the household.

To extend education to Syrian children, the government of Jordan quickly implemented a two-shift system, hired more teachers, and built schools. Although enrollment rates among Syrian refugees remain lower than those of Jordanians, their enrollment rates have increased over time. In 2014, fewer than 45 percent of refugee children were in school (Barbelet et al., 2018). As of 2018, nearly 70 percent of Syrian refugee children between the ages of 6-15 were enrolled, with rates slightly higher for boys than for girls. School enrollment rates are lower than before the war in Syria; primary school enrollment was universal, and secondary school enrollment rates were around 75 percent for both girls and boys. In contrast, 97 percent of Jordanian children are in school (UNICEF, 2014). Hagen-Zanker et al. (2017) found that distance, transportation costs, and harassment are among the reasons preventing access to schooling for refugee children of all ages. Safety concerns, particularly risks of gender-based violence also limited access, especially for girls. For secondary school-age children, these factors combined with the need to generate income for their households – particularly for boys –exacerbated their likelihood of dropping out (UNHCR, 2018b).

Finding employment was particularly challenging, especially in the early years of displacement. Initially, on arrival, Syrians could apply for work permits in specific sectors subject to quotas. However, the majority of Syrians undertook informal sector jobs that often were subject to exploitation. They usually worked long hours, had very low wages, and were subject to arrest by the Ministry of Labour patrols (Bellamy et al., 2017). By 2015, one in two refugees wanted to leave Jordan because they saw no future (Norwegian Refugee Council, 2016). As of 2020, the unemployment rate for Syrian refugees was around 17 percent (ILO, 2021), compared to 13 percent in Jordan nationally, with women's unemployment at 23 percent (World Bank, 2021c). An assessment of individuals and enterprises that received support from the ILO and humanitarian organizations found that COVID-19 disproportionately affected Syrian refugee workers, with one third losing their job following the 2020 pandemic outbreak, compared to 17 percent of Jordanian workers (ILO, 2019).

The Government of Jordan has been acutely aware of the problems facing Syria refugees and has taken important steps to respond to the impact of the Syrian crisis. Several Response Plans sought to integrate policies on livelihoods and education and move from a short-term humanitarian approach to a development perspective (Government of Jordan, 2017). In 2014, the Jordan Response Platform for the Syria Crisis (JRPSC) was established to coordinate and monitor the preparation and implementation of Response Plans in various matters including social protection, food security and livelihoods, water and sanitation, health, education, among others (Government of Jordan, 2015). Humanitarian assistance supported by the United Nations, including the UNHCR and the World Food Program, and by international donors such as USAID, the International Rescue Committee, and CARE, is also a crucial element of support to refugees in Jordan. A significant component of this assistance is cash-based programming. By 2018, Jordan was the UNHCR's second largest cash operation worldwide (after Lebanon) (UNHCR, 2019c), providing US$98 million in assistance to over 435,000 refugees (Brown et al., 2019).

In 2016, Jordan became the first country in the Arab region to facilitate Syrian refugees' access to the labor market (ILO, 2019). The Jordan Compact sought to advance long-term solutions through promoting access to education and to employment opportunities for Syrian refugees – international donors pledged $700 million in the form of grants, $1.9 million in concessional loans from the Multilateral Development Banks at concessionary rates, trade benefits with the European Union, and investments in the Jordanian economy (OECD, 2016b). The payment of grants and loans was tied to specific targets for refugees that included the provision of an additional 50,000 school places in 2017, the issue of 200,000 work permits by 2020 (UNICEF, 2016), and employment quotas in particular sectors (Barbelet et al., 2018; Kelberer, 2017). In 2021, the Jordanian government issued a record 62,000 work permits to Syrians, increasing opportunities for employment (UNHCR, 2022a). Work permits have increased wages and employment stability for Syrian refugees. In 2020, 25 percent of Syrian refugees with a work permit had a written contract with their employers and were covered by social security, compared to 9 percent of those without a permit (Stave et al., 2021).

Although the number of Syrians with work permits has increased, they still mainly find jobs in the informal sector since there is reportedly a mismatch between their skills profiles and the sectors in which they are allowed to work (Barbelet et al., 2018). Interview and survey data of Syrian refugees further revealed that a majority of workers, including those with higher skill levels, felt unsafe in their jobs due to frequent harassment, low bargaining power, harsh labor conditions, and little agency to change their situation due to fear of the authorities (IRC and Airbel Impact Lab, 2016). In this context, women are more disadvantaged than men. Given the direct and indirect costs of acquiring work permits (such as the direct cost of the transaction, the costs of providing all required documents at different institutions, transportation costs to complete those transactions, time and resources forgone that would have been used to work or care for children), it is often only male family members that are in a position to apply for work permits.

The limited availability of employment opportunities for women is strongly associated with social norms that bind women to specific roles in society (Kelberer, 2017). Data from the World Development Indicators (WDI) show that in 2019 only about 13 percent of women over 15 years old participated in the Jordanian labor market. Refugee women's options in the labor market are limited; for most part they are limited to the informal economy, mainly carrying activities such as tailoring, cleaning houses, and cooking for neighbors and family (UN Women, 2018). Changing gender roles in the labor market can cause tensions and have been linked to increased rates of domestic violence in refugee households in Jordan (Hagen-Zanker et al., 2017; Culcasi, 2019).

The most common reason for low levels of female labor force participation is family responsibilities such as childcare and housework, reported by 44 percent of Syrian refugee women and Jordanian women of working age (Stave et al., 2021). This is deeply tied to engrained norms surrounding gender roles. A recent analysis of gender norms in Jordan found that women and girls hold more equitable gender role attitudes than their male counterparts and that there are no significant differences between Syrians and Jordanians (Krafft et al., 2018). Restrictive norms perpetuated by men continue to severely restrict women's access to the formal labor market. Fortunately, the Jordanian Government has worked in recent years to address this problem. For example, in the amendment made to the Jordanian labor law in 2019, the provision of childcare services was made mandatory in every business that has 15 or more employees, male or female, with children under 5 (World Bank, 2021a). However, implementation is not consistent and low-income households report less access to childcare services (Weldali, 2022). Home-based businesses can also provide important access to economic opportunities for women, albeit within a limited number of sectors, and the government of Jordan's 2017 amendments of the regulations governing the licensing of home-based business is particularly important for female entrepreneurs in both host and refugee communities (Ait Ali Slimane, 2020; Turner, 2019).

## 2.3 Data and Descriptive Statistics

We use two rounds of household-level data collected by the UNHCR in 2013-14 and in 2017-18.[6] These data come from two sources, namely, the Profile Global Registration System (ProGres) and Jordan Home Visits (JD-HV). ProGres records demographic information of each individual refugee and their household at the time of registration with the UNHCR, including the relationship of each household member with the principal applicant. In this paper, we refer to the self-identified principal applicant as the household head – the term used in the poverty research literature. Self-identification is commonly used as a means of identifying the household head in household surveys of income, consumption and expenditure, which are used to estimate poverty rates (Hanmer et al., 2020).

Home visits began at the same time as the UNHCR cash assistance program was launched. The program aimed to assist vulnerable Syrian households to meet their basic needs. During the home visits, the UNHCR determined whether a household met the criteria to qualify for assistance, considering economic poverty and the presence of members in the household requiring protection.

[6] For simplicity, from now on we refer to the first wave as the 2013 wave, and to the second wave as the 2018 wave.

Since 2013, these visits have been conducted for every registered household by the UNHCR outside of the camps (UNHCR, 2014).

At registration, each household is assigned a unique registration number that serves to cross-reference the initial registration data with the Home Visits data. Home Visits data are collected for about one-third of the registered households. In both waves, the selected households to be interviewed are those that the UNHCR considers to be most vulnerable in terms of education, specific needs and the use of livelihood coping strategies (UNHCR, 2018c), thus, they are a nonrandom sample of all registered Syrian households in Jordan. During Home Visits, data on household expenditures, aid, income, schooling, and other demographics are collected. Each wave of data is a snapshot of a different nonrandom sample of registered refugees because the pool of registered households changed over time. However, because of data confidentiality, we could not link records across years. For this reason, we treat each wave as an independent sample of cross-sectional data.

We compare 204,941 individuals in 54,900 cases (households) in the 2013 wave, with 195,930 individuals in 43,292 cases in 2018. We focus on comparisons between different types of households; however, we use individual characteristics to determine the household categories. A case is formed when a Principal Applicant (PA) registers at a UNHCR office seeking asylum. Initially the self-identified PA reports all members of the household and their demographic characteristics, which are recorded in ProGres. Subsequently, the UNHCR carries out individual interviews with each family member and records their data in ProGres. The PA is also the person who receives assistance from the UNHCR although he/she may not be the only recipient of humanitarian assistance.[7]

Although income data are available, we use expenditure data from the JD-HV. The existing literature leans toward computing poverty estimates based on expenditure rather than income because: (i) income do not capture the possibility that there are people who can live off their savings or have assets to meet their immediate needs; (ii) consumption is smoother and less variable than

[7]The monthly per capita cash assistance Syrian refugees receive is roughly 60-65% the Survival Minimum Expenditure Basket (SMEB) and is intended to cover rent, water and sanitation costs. Assistance levels are mainly determined by family size. Since 2014, PAs often make the withdraws of their allocated cash assistance using iris scan enabled ATM machines (UNHCR, 2018c).

income, especially in agricultural economies; and (iii) income reporting appears to be more misleading, especially for those with more resources who tend to underreport income more (Carletto et al., 2022; Cutler et al., 1991; Deaton and Zaidi, 2002; Meyer and Sullivan, 2003, 2011, 2012).

Aggregations of expenditure categories were used to compute total expenditure; however, the main challenge was constructing comparable expenditure estimates across waves given a change in the number of spending categories included in the questionnaire. In the 2013 wave, there were nine expenditure categories, namely, rent, utilities, food, water, health treatment, education, transportation, basic households needs and others. In addition to these categories, the 2018 home visit survey records expenditure data on infant needs, hygiene items, debt payment, and telecommunications. This difference in expenditure categories could inflate reported expenditure in the second wave. This is because survey respondents people tend to reveal more accurate information when asked item by item, than when asked for an expenditure or income aggregate. In part, this may be due to an intention to withhold information to obtain the monetary assistance, and in part, it has to do with the recall error that this type of questionnaires elicit.

To aggregate expenditure, we restrict the number of categories to those in both waves which represent a sizable portion of a household expenditure; that is, rent, utilities, food, water, and transportation. These five categories average 97.5 percent and 77 percent of household expenditure in the first and the second waves, respectively. While spending on healthcare, education, and household items were also recorded in both waves, the amounts were not large and were excluded because healthcare and education expenditures may reflect a vulnerability (for example, caused by illness or disability in the case of health care), or a luxury for those who can access the service, and because these data were often missing. We do not include humanitarian cash assistance in the computation of total expenditure because survey respondents would not count cash assistance as part of their spending but rather as part of their income.

Our data enable insights into the change in refugee household's composition five years after their original displacement. The share of PAs (which we classify as household heads) who are women rose from 27 to 38 percent between 2013 and 2018. Interestingly, however, the characteristics of Syrian households observed in the data seem to be similar in both waves. For example, PAs in 2017-18 are about four years older on average compared to PAs in 2013-14. This is partly because most Syrian refugees arrived and were registered by 2013. By December 2017, registrations increased by only 14 percent relative to December 2013, which suggests that we are observing a similar pool of individuals in both waves; according to the UNHCR most of the population increase in 2017-18 is accounted for by babies born to Syrian refugee women living in Jordan.[8]

Table 2.1: Characteristics of Male and Female Principal Applicant (PA) Households

| | 2013-14 Wave | | | 2017-18 Wave | | |
|---|---|---|---|---|---|---|
| Variable | Male PA | Female PA | All PA | Male PA | Female PA | All PA |
| All PAs, % | 72.6 | 27.4 | | 62.0 | 38.0 | |
| Age of PA | 38.0 | 40.5 | 38.7 | 41.9 | 44.1 | 42.7 |
| Elderly PA (%) | 3.9 | 6.5 | 4.6 | 6.5 | 10.2 | 7.9 |
| Able-bodied male adults (%) | 38.0 | 8.4 | 29.9 | 31.4 | 13.3 | 24.5 |
| Married PAs below 18 (#) | 10.0 | 11.0 | 22.0 | 0.0 | 2.0 | 2.0 |
| Marital Status of PA | | | | | | |
| Married with spouse in the hh | 75.3 | 9.9 | 57.4 | 88.0 | 18.4 | 61.6 |
| Married without spouse in the hh | 8.8 | 54.3 | 21.2 | 6.0 | 40.3 | 19.0 |
| Single or engaged | 14.6 | 7.7 | 12.7 | 4.9 | 7.9 | 6.0 |
| Divorced or separated | 0.6 | 5.4 | 1.9 | 0.7 | 8.6 | 3.7 |
| Widowed | 0.8 | 22.7 | 6.8 | 0.5 | 24.8 | 9.7 |
| Education of PA | | | | | | |
| Less than 6 years | 16.6 | 28.5 | 19.8 | 18.8 | 33.1 | 24.2 |
| 6-11 years | 61.8 | 54.2 | 59.7 | 61.8 | 52.2 | 58.1 |
| More than 12 years | 18.5 | 14.9 | 17.5 | 19.5 | 14.7 | 17.6 |
| Family Type | | | | | | |
| Couple residing together with children | 65.7 | 8.8 | 50.1 | 73.5 | 13.9 | 50.9 |
| Couple residing together without children | 6.0 | 0.8 | 4.6 | 6.3 | 1.7 | 4.6 |
| Single caregivers | 4.6 | 47.8 | 16.5 | 5.0 | 36.5 | 16.9 |
| Single -person | 16.6 | 20.3 | 17.6 | 7.3 | 21.3 | 12.6 |
| Unaccompanied children | 0.7 | 1.1 | 0.8 | 0.1 | 0.2 | 0.1 |
| Non-nuclear and other hhs | 6.5 | 21.2 | 10.5 | 7.7 | 26.5 | 14.9 |

*Source:* Own calculations based on ProGres and JD-HV database. Note: PA = principal applicant. PA is determined at the time of registration. This table contains data on 42,505 observations in 2013 and 40,483 observations in 2018. To make both waves comparable, we excluded observations in wave one that did not receive the home visit questionnaire or that reported having no expenditure. For this reason, this table contains figures that slightly differ from Table 1 in Hanmer et al. (2019), which had 54,408 observations.

Hanmer et al. (2020) reported striking differences between the characteristics of male- and

[8]We do not have the information to be able to merge the datasets from both waves based on case identification numbers.

female-PA households in terms of education, marital status, and family types.[9] Table 2.1 confirms the persistence of these differences in the 2018 wave. In both waves, about 80 percent of PAs are married, however most male PAs are accompanied by their spouse (75 percent in 2013 and 88 percent in 2018). In contrast only 10 percent female PAs were married and accompanied by their spouse in 2013, although by 2018 this share had increased to 18 percent. Part of the explanation for this shift may be associated with the spouses being able to reunite years after the crisis. Among female PAs the share of divorced or separated women increased by about 4 percentage points and widows by about 2 percentage points in 2018 compared to 2013.

Most principal applicants have completed primary education – only 20 percent in 2013 and 24 percent in 2018 have less than 6 years of education. Taking a gender lens to human capital achievements, the share of female PAs with more than 12 years of education is about the same in both waves (15 percent), while the share of female PAs who had not completed primary school rose by about 4 percentage points. In contrast the percentage of male PAs with more than 12 years education increased slightly by about 1 percent and the percentage of male PAs who had not completed primary school rose by about 2 percentage points. So, on average male PAs have more years of schooling than female PAs and this gap widened between 2013 and 2018.

Household composition of refugee households can change over time as family splitting is a common response to conflict (Brück and Schindler, 2009; Ibañez, 2009). However, between 2013 and 2018 on average the composition of Syrian refugee households has not changed. In 2013 and 2018, one in two households are couples with children and one in six households has a single caregiver, the vast majority of whom are women. However, there are differences in the distribution of different types of female households in 2018 compared with 2013. There are fewer female single caregivers (49 percent in 2013 compared to 37 percent in 2018) and more female-PA couples with children and female-PA non-nuclear and other household types. Thus, some households that had female single care givers in 2013 may have been joined by their husbands and others absorbed into the households of relatives by 2018. The other notable change over time is in the distribution of male-PA households across family types. Most male-PA households are in couples with children

[9]Following their methodology, family types are defined based on the reported relationship with the principal applicant, age, and disability status. For example, single caregivers are individuals, married or unmarried, who did not report a husband/wife with them at the moment of settlement, but did report elderly, children or disable household members.

(66 and 72 percent in 2013 and 2018 respectively). However, the share of single male PAs falls from one in six male-PA households in 2013 to one in 14 in 2018 (17 to 7 percent). In contrast, among female principal applicants single women account for one in five principal applicants in both 2013 and 2018.

Table 2.2: Household Composition by Household Head

| | 2013-14 Wave | | | 2017-18 Wave | | |
|---|---|---|---|---|---|---|
| Variable | Male PA | Female PA | All PA | Male PA | Female PA | All PA |
| Household size | 4.2 | 3.6 | 4.0 | 4.7 | 3.5 | 4.2 |
| Children (%) | 52.6 | 59.2 | 54.2 | 52.7 | 53.1 | 52.9 |
| Boys (%) | 27.3 | 30.5 | 28.1 | 26.9 | 27.1 | 27.0 |
| Girls (%) | 25.3 | 28.8 | 26.1 | 25.8 | 26.0 | 25.9 |
| Adults (%) | 47.4 | 40.8 | 45.8 | 47.3 | 46.9 | 47.1 |
| Men (%) | 26.6 | 7.8 | 22.0 | 25.7 | 12.8 | 21.6 |
| Women (%) | 20.8 | 33.0 | 23.8 | 21.6 | 34.0 | 25.6 |
| Dependency ratios | | | | | | |
| Children | 1.09 | 1.76 | 1.28 | 1.19 | 1.43 | 1.28 |
| Children and elderly | 1.12 | 1.82 | 1.31 | 1.23 | 1.53 | 1.34 |
| Children, elderly and disabled | 1.14 | 1.84 | 1.34 | 1.27 | 1.58 | 1.38 |

*Source:* Own calculations based on ProGress and JD-HV database. Note: PA = principal applicant. PA is determined at the time of registration. Dependency ratios are defined as the proportion of the groups in the rows to adults.

On average, family size has remained the same over time for female PAs but has increased slightly — by about 0.6 to 4.7 -– in male-PA households. In other words, about half of male-PA households have an additional family member in 2017-18. Households with a female PA have fewer male adults than households with male PAs in both waves. In the average household of four people, in male-PA households, adult males account for one of the household members, in contrast most female-PA households have no male adults. In 2013, only 8 percent of female-PA households had a male adult member, increasing to 13 percent in 2018. This reflects that the path to female headship is often separation or widowhood, which are common after displacement.

Dependency ratios — the proportion of children and people not of working age or unable to work to working age adults — show mixed trends over time by household type, rising among male PA, and falling for female-PA households, the latter possibly because their children reached working age. However, the dependency ratios were larger for female-PA households throughout. These trends suggest that, compared to male-PA households, female-PA households are under greater pressure to provide for their families, although the pressure for male-PA households has increased over time.

## 2.4 Empirical Strategy

The objectives of this paper are: (i) to understand whether, in the context of displacement, the type of expenditure measure leads to different predictions about which types of households are vulnerable (ii) to assess the role that household and PA characteristics play in the level and distribution of expenditure over time.

To address the *first objective*, we compare per capita expenditure with measures that consider an adjustment for economies of scale. Per capita expenditure measures assume that there are no economies of scale in household consumption, that is, the per capita cost of reaching a specific measure of welfare, does not fall as household size increases (Lanjouw et al., 2004). Research in various settings has shown that poverty measurement is sensitive to this assumption. For example, Brown et al. (2019) indicate that, in Sub-Saharan Africa, a small adjustment for economies of scale can reverse conclusions about which households fall below the poverty line. Judged by traditional poverty measures, female-headed households have on average lower poverty rates than male-headed households. However, once consumption is adjusted for economies of scale, female-headed households tend to fare significantly worse in most of Sub-Saharan Africa.

Building on Brown et al. (2019), we adjust for economies of scale by rescaling expenditure in the following manner:

$$y_i^e(\theta) = \frac{y_i}{n_i^\theta} \quad \forall \theta \in (0,1] \tag{2.1}$$

where $y_i$ is total expenditure in household i, $n_i$ is the number of family members, and $\theta$ represents economies of scale in consumption. When $\theta$ equals 1, $y_i^e(\theta)$ is expenditure per capita.

After computing per capita and economies-of-scale-adjusted expenditure, we compare spending levels across different groups and over time to identify whether these measures lead to different conclusions. In this exercise, we compare unconditional distribution plots and run linear regressions that use each expenditure measure as dependent variable. In particular, we estimate the following linear model:

$$\ln y_i^e(\theta) = W_t'\beta_1 + H_i'\beta_2 + (W_t \times H_i)'\beta_3 + I_i'\alpha + \varepsilon_{it} \tag{2.2}$$

where $y_i^e(\theta)$ is expenditure adjusted for economies of scale (or per capita when $\theta = 1$), $W_t$ takes the value of 1 for households observed in the second wave (2017-18) and 0 in households observed in the first wave (2013-14); $H_i$ is set of demographic characteristics of the household, including a set of family categories that disentangle the effect by household headship; and $I_i$ is a set of characteristics of the PA.

In addition to using expenditure as the dependent variable, we do an additional test using a model in which we evaluate how the probability of being below the 40th percentile of the expenditure distribution changes according to the variables in model 1 (see Table B2). We chose to use this relative cutoff instead of a poverty line because we compute expenditure per capita based on a subset of expenditure categories, thus, using the national poverty line would overestimate headcounts. Nevertheless, this approach allows us to determine whether, for example, female-PA households are more likely to be below the bottom tail of the expenditure distribution in each wave and whether this likelihood has changed over time.

We use quantile regressions to evaluate how the expenditure distribution has changed over time and what role demographic variables play in the location of a household in that distribution (second objective). As is well known, the linear regression approach estimates the effects on the mean. However, in this case it is also important to know the distribution of those effects. For example, if we found that expenditure gender gaps have widened over time, one also would like to understand whether that effect is more prominent among households on the 10th quantile of the expenditure distribution or among those on the 90th quantile, conditional on demographic characteristics. For that reason, we estimate conditional quantile regressions as follows:

$$Q^{\tau}(y_i^e(\theta)|H_i, I_i) = H_i^{'}\mu^{\tau} + I_i^{\prime}\gamma^{\tau} + \varepsilon_i^{\tau} \tag{2.3}$$

, where $Q^{\tau}(y_i^e(\theta)|H_i, I_i)$ denotes the value at the $\tau$th percentile of $y_i^e(\theta)$, conditional on household ($H_i$) and PA characteristics ($I_i$). Notice that from this regression method we obtain different $\mu^{\tau}$ and $\gamma^{\tau}$ for each value of $\tau$. Thus, the effect that demographic characteristics have on expenditure may vary over the conditional distribution of expenditure.

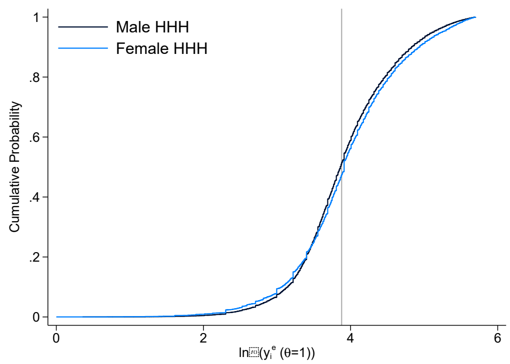
## 2.5 Results

### 2.5.1 Does the Type of Expenditure Measure Matters?

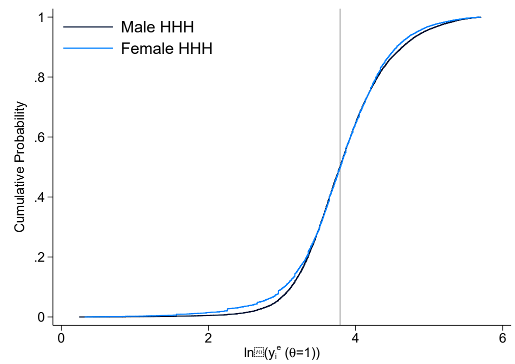
Figure 2.1 shows different expenditure distributions, which vary somewhat depending on the wave of observation and assumptions about economies of scale ($\theta$). Panel A and B show the distribution of expenditure per capita with no economies of scale for the first and the second wave of data, respectively. The per capita expenditure of the household at the median in 2013-14 is JOD$44.2—marked by the grey vertical line, while that of 2017-18 is JOD$44.2. In panel C and D economies of scale are assumed (using $\theta$=0.5). Expenditure levels in those cases are JOD$96.6 and JOD$89.3 for the first and second waves, respectively. If we assume no economies of scale, there are only minor differences between the expenditure distributions of female- and male-PA households, with the former's expenditure cumulating slightly faster in the 2017-18 wave. For example, in Panels A and B, the light blue cumulative distribution curve (female-PA households) is just above the curve for male-PA households below the 50th percentile. Above the 50th percentile in 2013-14 the cumulative distribution curve for female PAs lies below that of male PAs, indicating female PAs are spending slightly more. This gender expenditure gap, although statistically significant (see Table A1), is quite small. By 2017-18 the difference between curves increases, especially below the 50th percentile. This significant difference suggests that the poverty gap between female and male-headed households slightly widened over time negatively affecting female-headed households.

Figure 2.1: Unconditional Distribution of Expenditure, by Type of Expenditure Measure, Wave, and Gender of the PA.

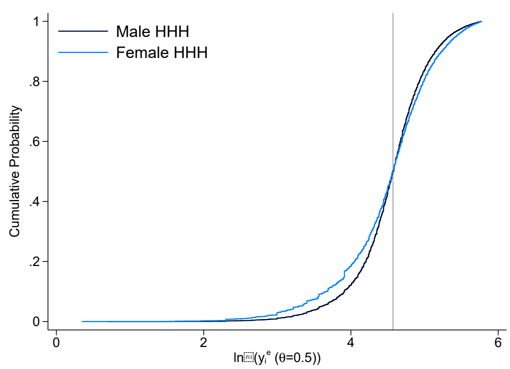
*Panel A.* Per capita expenditure, 2013-14



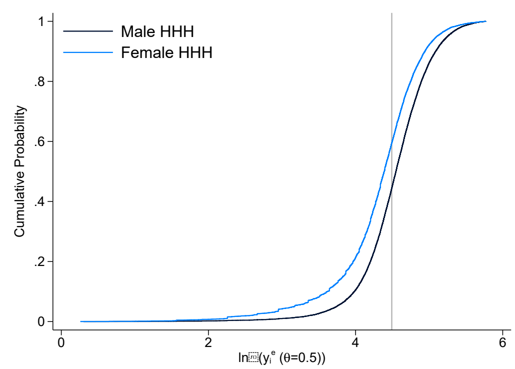

*Panel B.* Per capita expenditure, 2017-18




*Panel C.* Expenditure adjusted for economies of scale, 2013-14




*Panel D.* Expenditure adjusted for economies of scale, 2017-18

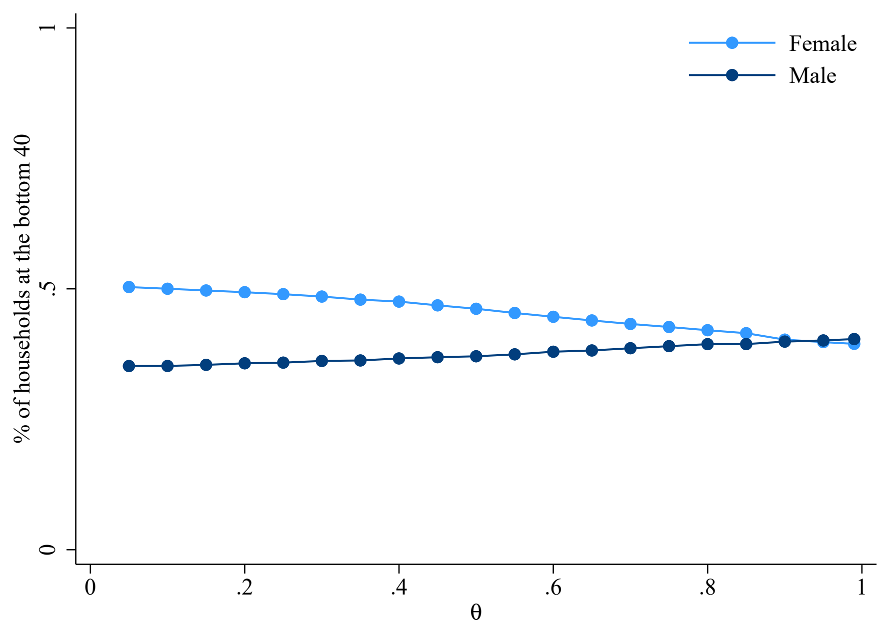



*Notes:* Own calculations based on ProGress and JD-HV database. The graph presents the distribution of per capita expenditure (when $\theta = 1$), and expenditure adjusted by economies of scale (when $\theta = 0.5$). The light blue line shows the distribution of spending by female-PA households while the dark blue line shows that of male-PA households. The vertical gray line locates the 50th percentile of the expenditure distribution, equal to JOD$48.3 in Panel A, JOD$44.2 in Panel B, JOD$96.6 in Panel C, and JOD$89.3 in panel D. Kolmogoroc-Smirnov tests of the differences between the male and female headed household expenditure distributions are shown in Table A1.

Differences between female and male-PA households and trends over time emerge when adjusting for economies of scale (panels C and D). Both graphs show that female-PA households are slightly worse off, below the 50th percentile and that by 2018, female PA households have lower per capita expenditures across the entire expenditure distribution. The largest differences between these two distributions in 2017-18 is 0.156, which is significant according to a Kolmogorov-Smirnov test (see table A1).

Building on Hanmer et al. (2020), Figure 2.2 presents a sensitivity analysis – showing how the share of female and male PA households in the bottom 40 percent of the expenditure distribution changes when $\theta$ changes. Using the per capita measure ($\theta$=1), slightly more male-PA households are at the bottom of the expenditure distribution, but as $\theta$ becomes smaller (economies of scale become larger), the share of female-PA households at the bottom rises and then surpasses the share of male-PA households (at $\theta$= 0.925). That only small adjustments in $\theta$ show differences in poverty between male and female PA households highlights the relative importance of assumed economies of scale in this setting – in our analysis we follow Brown et al. (2019) and use the square root scale (or $\theta$ =0.5) for the analysis of poverty.

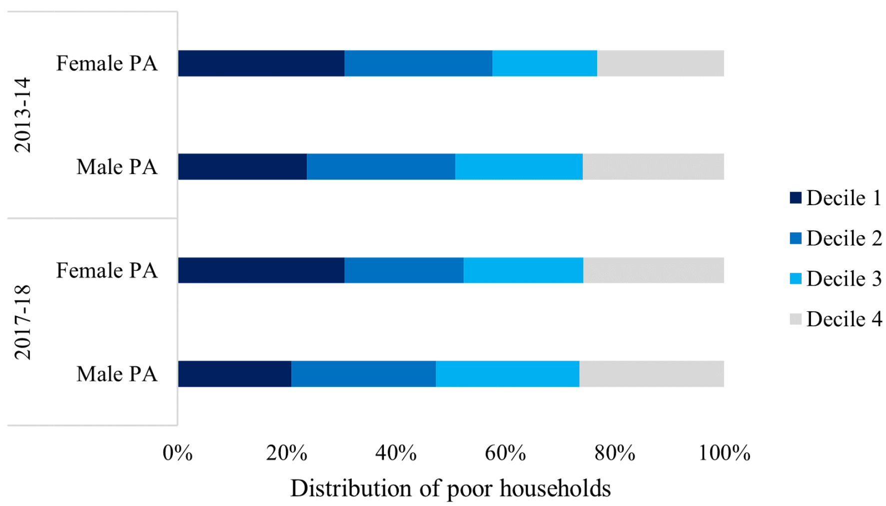
Figure 2.2: Percentage of Households at the Bottom 40 of the Expenditure Distribution, by Gender of the Household Head.



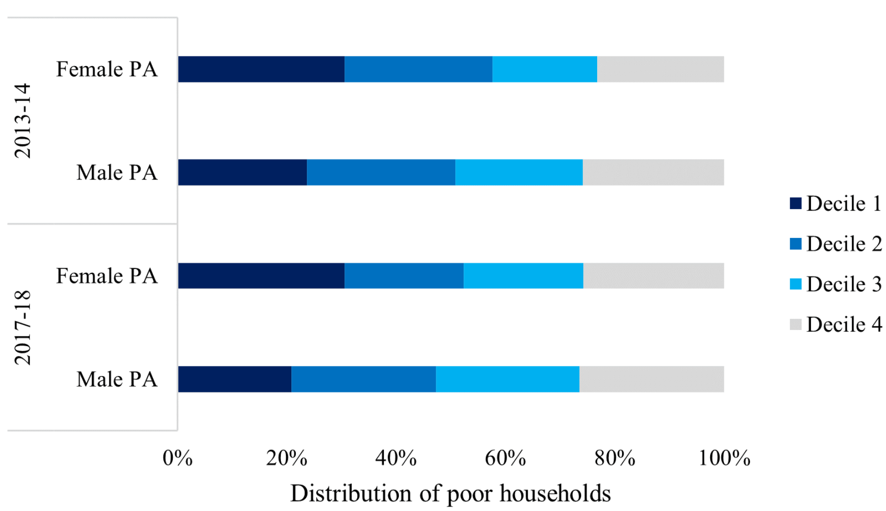

Note: Own calculations based on ProGres and JD-HV database. The graph shows the percentage of households below the percentile 40th by gender of the PA.

Figure 2.3: Intensity of Poverty, by Wave and Gender of the Household Head.

*Panel A.* Distribution of households below the 40th percentile using the unadjusted expenditure ($\theta = 1$)




*Panel B.* Distribution of households below the 40th percentile using expenditure adjusted by economies of scale ($\theta = 0.5$)




*Notes:* Own calculations based on ProGres and JD-HV database. The horizontal axis indicates the percentage of households in each decile of the bottom 40, for each wave. Panel A shows the percentage of households located on each decile of the unadjusted expenditure distribution, while Panel B displays the the same distribution using a measure of expenditure that accounts for economies of scale. .

Besides being more prone to poverty, households with female PAs are more likely than those with male PAs to be in lower deciles of the expenditure distribution. In Figure 2.3, Panels A and B show the percentage of households below the 40th percentile (deciles 1-4). Female-PA households in both waves are slightly over-represented in the bottom decile of the expenditure distribution: 1 in 4 poor female-PA households are in the first decile, the equivalent figure is 1 in 5 for their male counterparts. Panel B of Figure 2.3 shows that when adjustment is made for economies of scale, the gap of how intensely female-PA households experience poverty is about the same.

Table 2.3 shows that, using the unadjusted poverty line, couples with children and non-nuclear and other households with children are more likely to be in the bottom 40 percent of the expenditure distribution than other household types. However, this likelihood fell over time, especially for couples with children – in 2013-14, 54 percent of couples with children were in the bottom 40 percent, compared to 48 percent in 2018. On the other hand, unaccompanied children (although numbering very few – only 0.14 percent of all registered households by 2018) stand out as an increasingly vulnerable group, with the share of households below the 40th percentile going from 34 to 46 percent over the two waves.

Adjusting for economies of scale, single caregivers and single person household are most likely to be in the bottom 40 percent. The share of female single caregivers in the bottom 40 percent rose from 38 to 48 percent between the two time periods. This is in contrast to couples with children, where the share falls by around 5 percentage points. The share of non- nuclear households with children at the bottom of the distribution also decreased.

Table 2.3 also compares households with and without children over time. Using the unadjusted expenditure measure, households with children in both waves are significantly more prone to be at the bottom of the distribution (50.6% and 47.2% vs 13.4% and 19.4% in 2013-14 and 2017-18, respectively). However, this picture is highly sensitive to assumptions about economies of scale. In 2013-14 households with children are more likely to be below the poverty line than households without children (40 vs 28%). In contrast by 2017-18 the share of households without children below the poverty line is 42 percent compared with 38 percent of households with children.

Table 2.3: Share of Refugees Below 40th Percentile, by Demographic Characteristic

| | Unadjusted expenditure $\theta = 1$ | | Expenditure adjusted for economies of scale $\theta = 0.5$ | |
|---|---|---|---|---|
| | 2013-14 | 2017-18 | 2013-14 | 2017-18 |
| Male HHH | 41.0 | 39.8 | 35.7 | 32.6 |
| Female HHH | 37.8 | 40.3 | 38.0 | 49.0 |
| | | | | |
| Couples residing together with children | 54.1 | 47.9 | 39.9 | 34.8 |
| Couples residing together without children | 14.7 | 15.9 | 25.6 | 31.8 |
| Single caregivers | 40.0 | 42.3 | 37.9 | 47.9 |
| Single person households | 10.3 | 19.9 | 29.1 | 51.2 |
| Unaccompanied children | 34.5 | 46.4 | 57.5 | 73.2 |
| Non-nuclear and other households with children | 46.7 | 45.9 | 36.3 | 35.3 |
| Non-nuclear and other households without children | 21.0 | 20.3 | 26.4 | 31.6 |
| | | | | |
| No children in the hh | 13.4 | 19.4 | 28.1 | 42.1 |
| Children in the hh | 50.6 | 47.2 | 39.5 | 37.7 |
| | | | | |
| Dependency ratio ¡1 | 18.1 | 21.7 | 27.0 | 32.3 |
| Dependency ratio ¿=1 | 54.3 | 50.9 | 42.0 | 42.7 |

*Source:* Own calculations based on ProGress and JD-HV database. This table shows the percentage of households below the 40th percentile of the expenditure distribution, for each wave and each demographic group.

Turning to differences between households with higher and lower dependency ratios, households with higher dependency ratios (¿1) are less likely to be in the bottom 40 percent of the distribution by 2017-18, compared with 2013-14, using the unadjusted expenditure measure (54 percent compared to 51 percent), while using the measure adjusted for economies of scale yields no significant difference between the two waves.

Up to this point, we have analyzed whether there are differences in the distribution of unconditional spending using an unadjusted and an adjusted-for-economies-of-scale expenditure measure. We have found that the level of spending is indeed affected by the economies of scale adjustment, making female PA households more likely than male PA households be vulnerable to poverty when correcting for economies of scale. We now assess whether spending levels, conditional on demographic characteristics, are also sensitive to economies of scale adjustment. With this in mind, we first run a linear regression using the expenditure measure that adjusts for economies of scale as a dependent variable (assuming $\theta$=0.5). Then, we implement a sensitivity analysis of the size of the coefficient associated with the gender of the household PA using different $\theta$s to understand the extent to which the conditional estimates change under these adjustments.

Table 2.4: Correlates of Log Expenditure Per Capita Adjusted by Economies of Scale using OLS, by Wave and Gender of the Household Head

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Variables | Female | Male | Pooled | Pooled |
| Female PA | | | 0.010 | -0.011 |
| | | | (0.008) | (0.023) |
| Wave 2017-18 (base: 2013-14) | -0.212*** | 0.002 | -0.003 | 0.001 |
| | (0.008) | (0.004) | (0.004) | (0.004) |
| Female x Wave 2017-18 | | | -0.203*** | -0.213*** |
| | | | (0.009) | (0.009) |
| Characteristics of the PA | | | | |
| Age | 0.004*** | 0.002*** | 0.002*** | 0.003*** |
| | (0.001) | (0.000) | (0.000) | (0.000) |
| Marital Status (base: married) | | | | |
| Single | -0.148*** | -0.121*** | -0.130*** | -0.130*** |
| | (0.020) | (0.014) | (0.012) | (0.012) |
| Divorced | -0.017 | -0.051 | -0.027* | -0.021 |
| | (0.016) | (0.033) | (0.014) | (0.014) |
| Widowed/Widower | -0.082*** | -0.120*** | -0.082*** | -0.076*** |
| | (0.010) | (0.034) | (0.009) | (0.010) |
| Education (base: less than 6 years) | | | | |
| 6-11 years | 0.136*** | 0.134*** | 0.131*** | 0.136*** |
| | (0.010) | (0.006) | (0.005) | (0.006) |
| More than 12 years | 0.213*** | 0.234*** | 0.224*** | 0.235*** |
| | (0.013) | (0.007) | (0.006) | (0.007) |

| | | | | |
|---|---|---|---|---|
| Female PA | | | | |
| x 6-11 years | | | | -0.011 |
| | | | | (0.011) |
| x More than 12 years | | | | -0.034** |
| | | | | (0.014) |
| Characteristics of the household | | | | |
| Family size | -0.015*** | -0.043*** | -0.035*** | -0.044*** |
| | (0.002) | (0.001) | (0.001) | (0.001) |
| Able bodied male adults, % | 0.152*** | 0.164*** | 0.236*** | 0.172*** |
| | (0.027) | (0.016) | (0.011) | (0.016) |
| Number of children below 5 | -0.076*** | -0.035*** | -0.044*** | -0.031*** |
| | (0.005) | (0.003) | (0.002) | (0.003) |
| Number of elderly in the hh | -0.181*** | -0.085*** | -0.106*** | -0.094*** |
| | (0.017) | (0.011) | (0.009) | (0.011) |
| Family type (base: couples residing together with children) | | | | |
| Couples residing together without children | -0.173*** | 0.000 | -0.011 | -0.000 |
| | (0.049) | (0.011) | (0.011) | (0.011) |
| Single caregivers | -0.058*** | 0.054*** | 0.007 | 0.050*** |
| | (0.012) | (0.012) | (0.007) | (0.012) |
| Single person households | -0.130*** | -0.011 | -0.063*** | -0.008 |
| | (0.020) | (0.015) | (0.011) | (0.014) |
| Unaccompanied children | -0.252*** | -0.235*** | -0.241*** | -0.216*** |
| | (0.078) | (0.048) | (0.042) | (0.047) |
| Non-nuclear and other households | -0.004 | -0.012 | 0.059*** | -0.011 |
| with children | (0.012) | (0.013) | (0.008) | (0.012) |
| Non-nuclear and other households | 0.060*** | 0.013 | 0.049*** | 0.008 |
| without children | (0.018) | (0.012) | (0.010) | (0.012) |
| Female PA | | | | |
| x family size | | | | 0.030*** |
| | | | | (0.003) |
| x able bodied male adults, % | | | | -0.010 |

| | | | | |
|---|---|---|---|---|
| | | | | (0.031) |
| x number of children below 5 | | | | -0.050*** |
| | | | | (0.005) |
| x number of elderly in the hh | | | | -0.065*** |
| | | | | (0.018) |
| x couples without children | | | | -0.172*** |
| | | | | (0.050) |
| x single caregivers | | | | -0.108*** |
| | | | | (0.017) |
| x single person households | | | | -0.114*** |
| | | | | (0.023) |
| x unaccompanied children | | | | -0.072 |
| | | | | (0.089) |
| x non-nuclear and other households | | | | 0.010 |
| with children | | | | (0.017) |
| x non-nuclear and other households | | | | 0.059*** |
| without children | | | | (0.021) |
| Constant | 4.420*** | 4.504*** | 4.460*** | 4.481*** |
| | (0.029) | (0.015) | (0.013) | (0.014) |
| Observations | 26,857 | 56,745 | 83,602 | 83,602 |
| R-squared | 0.075 | 0.095 | 0.094 | 0.100 |

*Note:* Robust standard errors in parentheses. *** p¡0.01, ** p¡0.05, * p¡0.1. The dependent variable is the log of household expenditure per cápita adjusted by economies of scale, where $\theta$=5.

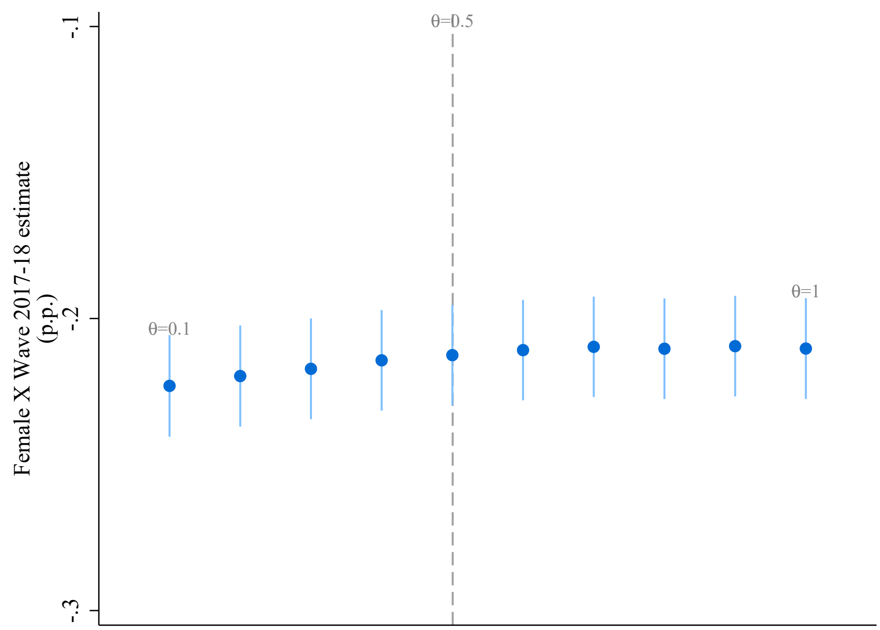
Table 2.4 shows correlates of log expenditure with individual characteristics of the PA, children, the elderly, family type, marital status, and household composition. We discuss results of our preferred model, which controls for household characteristics and interactions of those characteristics with the gender of the PA (See Table 2.4, Column 4). Model (4) shows that the expenditure of female PAs in 2017-18 is 21.3 percent lower than the expenditure of male PA households in 2013-14, suggesting that gender-related differences in poverty are large. This effect is similar to the effect in Column (1), which suggests that the expenditure gap is changing due to the deteriorating position of female PA households over time. With respect to the characteristics of the PAs, we found significant effects for age, marital status, and education for all models. The results also show that being older and more educated is associated with increased spending for both female-PA and male-PA households in both time periods. Female headed households are particularly advantaged when the PA is more educated, as their level of spending is significantly higher. Being single or a widow/er reduces expenditure by 13.0 and 7.6 percent, respectively, compared to married PA households. This result was as expected since these types of households tend to have smaller numbers of able-bodied male adults who are most likely to be able to get work.

With respect to household characteristics, model (4) shows that each additional household member is associated with a decrease in per capita spending of 4.4 percent. Interacting family size with female headship shows the decrease in expenditure associated with larger families is highest for male PA households. This result can be possibly related to the increase in size of male-headed households along with the proportional decrease in males able to participate in labor markets in those households over time. As seen in Table 2.2, while the size of female-headed households remained almost constant, that of male-headed households increased by 0.5. That is, 1 out of every 2 households acquired an additional member. In addition, these additional members may not be labor market participants, as suggested by Table 2.1, which shows that there is no increase in able-bodied males. Altogether, a larger increase in dependents among male-headed households might be causing these differential effects on per capita expenditure.

The presence of able-bodied males increases spending and thus, reduces the risk of poverty. This effect may be explained by the greater capacity of men to enter the labor market given the social norms of Syrians and Jordanians. Although the effect is independent of the gender of the PA, it is worth noting that female-headed households are strikingly less likely to have household members who are able-bodied male adults (see Table 2.1), which makes them even more vulnerable to poverty. A related finding is that the presence of children and elderly in the household increase poverty risk, especially for female-PA households. Table A2 show similar results for a dichotomous variable that takes the value of 1 for households below the 40th percentile and 0 otherwise.

Analyzing how family configurations affect spending, we find that unaccompanied children are the most vulnerable: They spend 21.6 percent less than the average household formed by a couple with children. The table also shows that female-headed households with a family arrangement other than a couple with children have sizeable and significantly lower levels of spending, compared to their male-headed counterparts. Table 2.4 also shows that a greater number of dependents, whether children under 5 or elderly, can have more adverse effects on female-headed households, decreasing their expenditure by 5-6.5 percent, compared with their male counterparts. In summary, these results show that female-headed households tend to be more vulnerable to poverty than those headed by men.

Figure 2.4: Sensitivity of the OLS regression models using different expenditure measures.




Note: Own calculatios based on ProGress and JD-HV database. This table shows, in the vertical axis, the estimated coefficients and confidence intervals (at 95%) associated to the interaction Female×Wave 2017-18, following the model presented in Table 4 and Equation 1. The x axis shows the values of $\theta$ used to compute the dependent variable in each case. The vertical gray line points at the estimate shown in Table 2.4, where $\theta = 0.5$.

To analyze the sensitivity of these results to different economies of scale adjustments, linear regressions are run following Equation (1) using different measures of expenditure as the dependent variable. Each measure varies according to the $\theta$s shown in the horizontal axis of Figure 4. The figure shows on the vertical axis the size of the coefficient estimate associated with the interaction between the gender of the PA and the data wave. The vertical dotted gray line shows where the coefficient from the regression shown in Table 2.4 is located.

Contrary to the findings of the unconditional distributions, we find that the size of the estimates does not vary in the presence of different economies of scale adjustments. If anything, a higher $\theta$ (i.e., a measure of spending closer to unadjusted per capita spending), makes the differences between men and women over time appear smaller. However, it is worth noting that the differences are so subtle that they are non-significant across models. This suggests that once the demographic characteristics of the household are controlled for, the choice of expenditure measure used is of little relevance for statistical inference on the average household. This result is consistent with the fact that displacement situations affect all types of households, many of which are in new family arrangements. Once we control for the different types of family arrangements, economies of scale do not capture differences in expenditure better than the headcount measure.

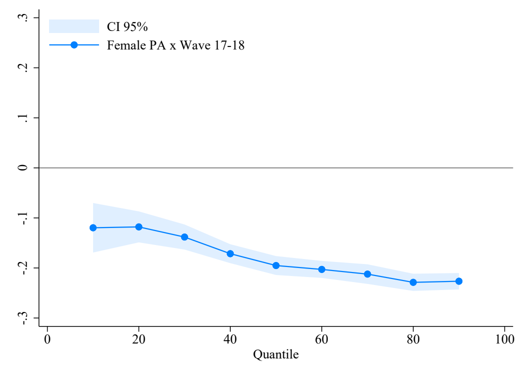
### 2.5.2 How Does the Gender Gap in Spending Evolves Over Time and Across the Expenditure Distribution?

To answer this question, we run quantile-conditional regressions following the model described by Equation 2.3. We run two models per wave, independently shown in Tables B3 and B4. These regressions allow us to identify how gender effects vary at different points in the conditional income distribution. With this in mind, Figure 4 shows how the coefficient associated with the principal applicant's gender changes for each expenditure level. Recall that in the last section we had found that, on average, female-headed households had lower levels of spending compared to those headed by men, especially in the 2017-18 wave.

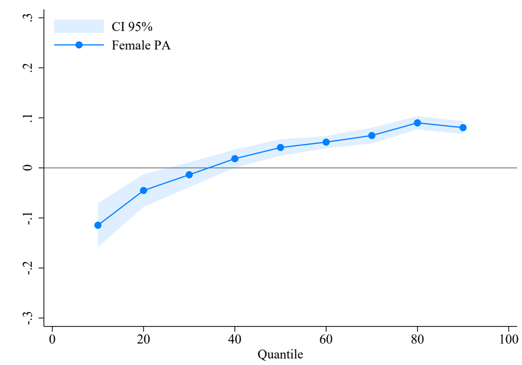
Figure 4 shows that the coefficients associated with the gender of the PA and the interaction between the gender of the PA and the wave of data change depending on the quantile. In particular, we found that the distribution of expenditure has shifted over time making female-headed households in all quantiles worse off than male-headed households, in 2017-18 compared to 2013-14. This finding is supported by two facts shown in the graphs: (i) Panel B shows that female-headed households in the first two quantiles have a lower level of spending compared to male-headed households, with a difference of up to 10 percentage points. At the same time, it shows that in the top quantiles, female-headed households' expenditure is higher than male-headed households' by up to 10 percentage points difference; (ii) Panel C shows that households observed in 17-18 had similar or relatively lower expenditure levels than those observed in 13-14, for all quantiles except for the first quantile. For the latter, expenditure levels improved in 17-18 relative to 13-14; (iii) Panel A displays that the coefficients associated to the interaction term are negative for all quantiles. Taking together the coefficients of Panel A and Panel B to check how the expenditure distribution changed for female-headed households, we observe that for the lower quantiles the expenditure gap is around 20 percentage points disadvantaging women. For the upper quantiles, although women have positive coefficients in Panel B, these are smaller in size than the negative coefficients shown in Panel A. Thus, we see that in the upper quantiles female-headed households also have lower levels of spending relative to male-headed households. This result emphasizes that the risk of poverty has increased steadily for female PA households over time across the whole expenditure distribution.

Figure 2.5: Gender Gap Between PAs (Female – Male). Estimated Coefficients from Log Expenditure Quantile Regressions.

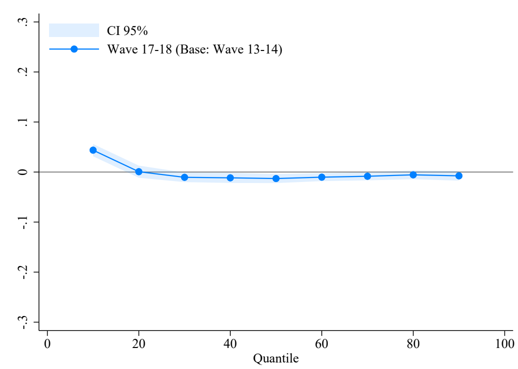
*Panel A*. Coefficients associated with Female x Wave 17-18




*Panel B*. Coefficients associated with Female PA




*Panel C*. Coefficients associated with Wave 17-18




*Notes:* This figure shows estimated coefficients from a quantile regression model that includes as explanatory variables the gender of the PA, the wave of data, an interaction between these variables, and control variables. The y axismeasures the coefficient estimates from the regression, i.e., log changes in expenditure or percent changes (i.e., -0.1 is a 10 percent reduction) due to changes on the explanatory variables. Confidence intervals at 95% are displayed by the shaded area. Complete regression results available in Table A.4.

Next, as a robustness check, we decompose the results to understand the components of the gap between female and male principal applicants. Figure 5 shows Recentered Influence Function (RIF) regressions for both waves by decile, following the methodology by Firpo et al. (2009, 2018).

The RIF methodology generates unconditional quantile estimates, while quantile regressions return conditional quantile estimates. Using this methodology, we decompose the expenditure gap between female-PA and male-PA households into the component attributable to the differences in the observable characteristics of the households (endowment or composition effect), and the unexplained components due to differences in the returns to those characteristics (structural effect).

In summary, we find similar results to those found in the quantile regression models. In 2013, female-headed households were the most disadvantaged, especially among the poorest. Demographic characteristics of these households (age, education, family size, etc) seem to be explaining most of the gap. Above the median, the most disadvantaged households are those headed by males. The scenario changed in 2018. In this wave, female-headed households are the vulnerable across the entire spending distribution. The gap remains largely unexplained. In other words, we find a gap between households with a female PA and those with a male PA even when taking into account demographic factors. The unexplained part of the gap can be associated with factors leading female PAs to experience lower economic standing that are unobserved in the data we have, and which point to the need of more data to further explore this component.

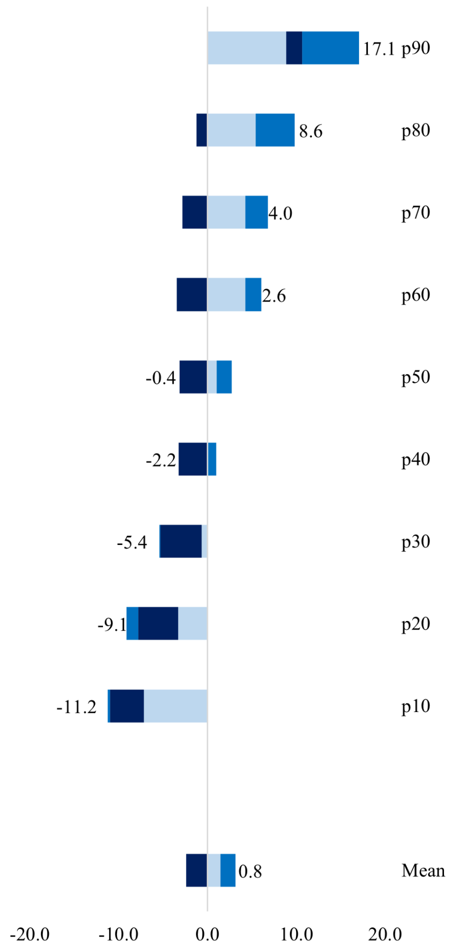
## 2.6 Final Remarks

Our analysis shows how using a gender lens that differentiates between types of households and considers economies of scale can enrich the understanding of poverty in situations of forced displacement. In the case of Syrian refugees, specific household structures, particularly those that result from disruptions caused by displacement, are more prone to experiencing poverty.

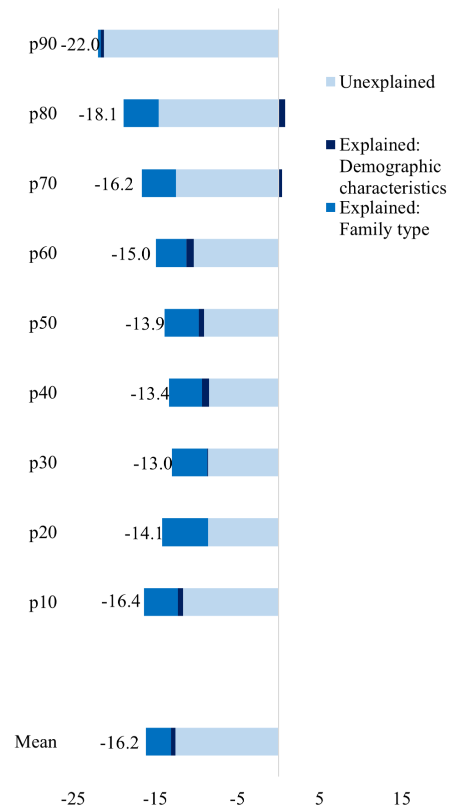
However, while we have described how patterns of poverty have changed over time, lack of data on critical variables, most importantly the amount of social assistance received by households and labor force participation, employment and earnings data, limit our analysis. While some of the changes in poverty over time experienced by male- and female-PA households are consistent with some of the changes in the circumstances in Syrian refugees, specifically their improved access to labor markets and better access to services, it is not possible to understand the extent to which these and potentially other factors have made a difference to their risk of poverty.

Figure 2.6: Gender Gap Between PAs (Female – Male). Estimated Coefficients from Log Expenditure Quantile Regressions.

*Panel A.* Coefficients associated to Female x Wave 17-18




*Panel B.* Coefficients associated to Female PA




*Notes:* This figure shows estimated coefficients from a quantile regression model that includes as explanatory variables the gender of the PA, the wave of data, an interaction between these variables, and control variables. The y axismeasures the coefficient estimates from the regression, i.e., log changes in expenditure or percent changes (i.e., -0.1 is a 10 percent reduction) due to changes on the explanatory variables. Confidence intervals at 95% are displayed by the shaded area. Complete regression results available in Table A.4.

Overall, our results show that the number of female-PA households have increased substantially, from from 26 to 38 percent of all Syrian refugee households between 2013 and 2018. There are differences between the characteristics of male- and female-PA households in terms of education, marital status, and family types. While male PAs are likely to be married, the vast majority of female PAs are either single, divorced or widowed, or married but separated from their spouses. However, by 2018 the share of married female PAs living with their spouses has increased from 10 percent in 2013 to 18 percent, suggesting that some men who had stayed behind in Syria left to rejoin their families in Jordan during this time period.

We use a relative poverty line and examine changes in the composition of the types of households that lie in the bottom 40 percent of the distribution of per capita expenditure between 2013-14 and 2017-18. We show that assessment of the evolution of poverty over time is highly sensitive to assumptions made about economies of scale. Assuming a moderate amount of economies of scale, we find that female-PA households have fared worse than male-PA households between 2013-14 and 2017-18, and are disproportionately represented in the bottom 40 percent of the expenditure distribution. Furthermore, female-PA households have lower per capita expenditures than male-PA households across the whole consumption distribution. Female-PA households have on average 22 percent lower expenditure per capita and are more likely to be in the bottom 40 percent of the distribution. By 2017-18, a 16-percentage point gap has emerged between male and female PA households compared with a 2-percentage point gap in 2013-14.

Particular household types emerge as being especially vulnerable to increased poverty in 2017-18; single caregivers – mostly women – and single persons, both male and female, face increasingly high poverty risks in 2017-18 relative to 2013-14. Unaccompanied children remain at high poverty risk, but thankfully their numbers have decreased substantially between 2013-14 and 2017-8. Further analysis of risk factors is consistent with women's care burden driving some of their increased risk of poverty over time. The number of children below 5 and the number of elderly increased poverty risk more for female-PA households than male-PA households. In contrast for male-PA households, larger household size is most associated with higher poverty risk.

Factors that reduce poverty risk include post-primary education and, having more adult male household members, the latter we take as a proxy for labor market access. While households with these characteristics are associated with lower poverty risk for both male- and female-PA households, the effect is largest for female-PA households.

Displacement is long-term for many Syrian refugees. The challenges and opportunities faced by the displaced in protracted situations are different from those in the phase of emergency. In principal, results from our analysis can be used to inform the design of better targeted social assistance programs that recognize poverty risk of smaller households and aim to design special measures to support those who are caring for young children, the elderly, or other dependents. Our results underline the importance of continued access to secondary education for both boys and girls. It is also important to build capacity for both refugee women and men to access economic opportunities, which can eventually replace social assistance. It is important to note, however, that not everyone will be able to work. Social protections will continue to be needed by those who are elderly, people with disabilities that limit the work they can do, people in poor health, and other vulnerable groups.

Program interventions should identify occupations and sectors where refugees could work given their skills and they should also include components providing support services, particularly for women, such as flexible working arrangement to address specific constraints related to domestic responsibilities.

The environment for job creation is particularly challenging in Jordan. Even before the impact of the COVID crisis, Jordan experienced only modest levels of productivity and low productivity growth. With the impact of COVID the unemployment rate increased to nearly 25%, and women and youth, whose unemployment rates were already the highest, were hardest hit by the crisis. Creating meaningful economic opportunities for refugees and Jordanians alike will require tackling underlying economic constraints that lead to persistently high levels of unemployment and informality. According the World Bank's Job Diagnostic for Jordan (Winkler, Hernan, & Alvarez, 2019) the low level and quality of job creation in the private sector is mostly explained by firm dynamics: limited entry and growth of firms not driven by the most productive firms, while firms exiting are not always the least productive, all resulting in a productive structure dominated by small/micro, low-productivity firms employing half of private sector workers, including two thirds of informal workers. The Government of Jordan launched a National Employment Plan in Response to the Crisis, but structural reforms are also needed to help the private sector create more jobs. (World Bank, 2021a).

However, providing access to economic opportunities is not a guarantee that gender gaps in access to economic opportunities for either refugees or Jordanians will be reduced if men have full control of gains in access to economic opportunities, as determined by patriarchal norms. Economic empowerment programs for refugee women, in particular, should build in guidelines for their protection and should engage refugees in promoting more gender-equitable relationships (Heilman & Barker, 2018).

# Chapter 3

# The Effect of Testing Teachers on Student Performance in Colombia

## 3.1 Introduction

Due to information asymmetries, employers use many strategies to identify the best workers within a set of applicants. Although education and experience may be good predictors of productivity and unobserved ability, these characteristics are imperfect indicators. In this context, pre-employment tests offer a powerful tool to improve screening processes (Goldhaber and Brewer, 2000). However, testing applicants may have unintended consequences, such as deterring the most qualified workers from applying. This may happen because skilled workers could easily find jobs elsewhere at a lower cost. Therefore, there are two forces that pre-employment tests, as a labor market entry barrier, induce. On the one hand, they screen out the least productive applicants who do not meet the requirements to perform well on their job. On the other hand, pre-employment tests may also discourage high-ability workers with better outside options in the market and for whom applying to such jobs requires more effort, given the pre-screening process. Which of the two forces is stronger remains a debate in the literature.

Although research on the effectiveness of pre-employment tests has not been conclusive, many governments have implemented screening processes in an attempt to guarantee that high-quality services are provided to their citizens. This is the case for Colombia's educational sector, which currently has in place a pre-employment test for job applicants who aspire to teach in primary and secondary public schools. This hiring system was put in place through a national hiring system reform, approved in 2002 and implemented in 2005, which conditioned teachers' recruitment and salary increases on a standardized national test. Candidates were evaluated on verbal and numerical reasoning as well as on subject-area knowledge and pedagogy theory. These tests were first implemented in two separate cycles in 2005. But only the results of the teacher tests in the second cycle of 2005 were used to create the list of eligible candidates to be hired in public schools. If schools needed to hire new teachers, the municipal Government where the school was located was required to draw the new teacher from this list according to the teachers' ranking, location, and area of expertise. To be part of the list of eligible candidates and, therefore, be eligible to be employed in public schools, candidates had to score 60 points out of 100 on the pre-employment test. The test was only used as a pre-screening mechanism for new candidates. Thus, teachers already working in public schools in 2005 or earlier were not affected by this reform and were not required to take the test to be employed or promoted. In other words, starting in 2006, schools had a combination of teachers hired through the old and new processes.

Using student-level data on students' test scores and variation in the share of new hires at the school level, I evaluate whether implementing a national standardized test for candidates applying for teaching positions in public schools jobs impacted high-school students' achievement. One year after implementing the reform, children in schools that hired new teachers obtained similar scores to those in schools that did not hire teachers. This result is not surprising given that the effects of new hires are expected to be noticeable in the medium to long run. When I evaluate the effects eight years after the implementation of this new hiring system, no effect was found in any of the student tests analyzed: mathematics, language, biology, philosophy, physics, and chemistry. These results are robust to both difference-in-differences and event study estimations. When I evaluate the possibility of heterogeneous effects by student gender, I found modest declines in girls' test scores in math, philosophy and physics. This result may be explained by the relatively lower recruitment of female teachers in the new system compared to the old system.

This paper contributes to the literature on candidate selection in two ways. First, to the best of my knowledge, this is the first paper that evaluates the individual-level medium-term effects of a nationwide reform in teacher recruitment that introduces pre-employment tests. Second, because standardized tests in Colombia include subjects other than the traditional math and language tests, this is also the first paper that evaluates the effects of pre-employment tests for other areas of knowledge. Within the related literature on this topic, the work of Brutti and Sanchez (2017), who evaluated this reform in Colombia using school-level estimates, stands out. They find modest positive effects of the new regulation on student achievement. However, individual-level estimates may yield more precise results. Other Latin American countries have implemented similar reforms that have also been evaluated. Araujo et al. (2019) evaluates the effects of a similar reform in Ecuador one year after its implementation using individual-level data and information about teacher candidates. They find no evidence that a more competitive recruitment process is reflected in better student performance. Their findings support the effects reported by Cruz-Aguayo et al. (2017), who analyzed the same reform in Ecuador. Estrada (2015) studies the Mexican case, in which teachers could be hired discretionally by the union, or competitively through a test. He finds that the test had no power to predict teacher quality.

The broader literature on pre-employment tests also finds contradictory results regarding the effectiveness of pre-employment tests on productivity. On one hand, Schmidt and Hunter (1998) reviewed the literature on the effect of different types of selection processes and found that cognitive ability tests are excellent predictors of labor productivity in all jobs in all contexts. Additionally, they argue that the studies that did not find such a relationship lacked of statistical power. On the other hand, studies such as Rudner (1992), are more conservative and argue that pre-employment tests effectively select the best applicants only under certain conditions. One of the reasons to believe that pre-employment tests work is that administering this type of testing is costly for applicants in terms of effort. Some employers may even ask applicants to pay a fee for being tested. This additional cost is intended to screen out workers with low skills (Guasch and Weiss, 1981).[1]

The areas evaluated in any pre-employment test are key to determining whether a test is effective in selecting the best candidates. But designing these areas is a challenge, especially in the education sector since the job requires a combination of pedagogical and subject-area knowledge without a clear understanding of the relative importance of those skills. Traditionally, primary and secondary school teachers have a background in pedagogy (Ballou, 1996), and as a consequence, tests focus on the assessment of pedagogical knowledge as opposed to subject-area knowledge. As Angrist and Guryan (2004) find, screening teachers using tests favors those who went to educational programs as opposed to other academic programs. Moreover, they demonstrate that better teachers' scores are not reflected in better student outcomes. This evidence may support the theoretical work by Guasch and Weiss (1981), who find that the more able workers may underestimate their probability of passing the tests (perhaps because they are tested on skills for which they do not have educational background). As a result, they are discouraged from applying, leading to firms' lower profits. Other studies have estimated the effects of teacher licensing on students achievement and have found diverse results (Lankford et al. (2002), Goldhaber and Brewer (2000), Darling-Hammond (2000), Hanushek and Pace (1995), Strauss and Sawyer (1986)). These papers usually use variations in licensing laws at the state level to understand the selection of teachers affects student outcomes, as opposed to national uniform policies.

[1]There is also a concern that tests may discourage minority applicants. However, Autor and Scarborough (2008) show that although minorities performed worse on the test, there has been no impact on minority hiring.

The remainder of the paper proceeds as follows. First, I explain the hiring process of teachers in Colombia before and after implementing the new policy. Then, I describe the data, present descriptive statistics, and explain the identification strategy. Finally, I analyze the results and provide some insights into the policy implications of these results.

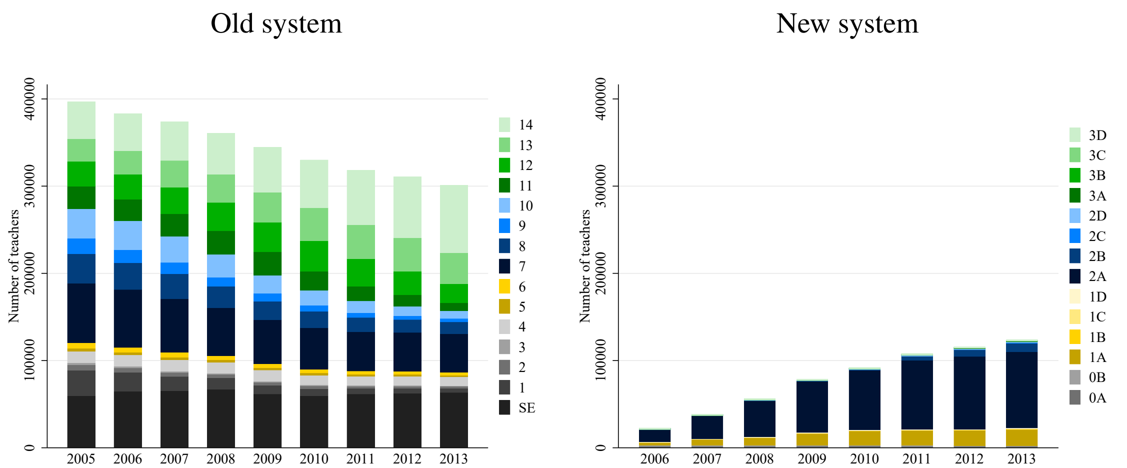
## 3.2 The Teacher Hiring Reform

Investment in education has been one of the cornerstones of the Colombian National Government to promote economic development. As a result, multiple policies have been implemented in past decades to increase the coverage and quality of education. In the 1991 constitution, for example, the Government determined that education should be compulsory for children aged between five and fifteen years, should include at least one year of preschool and nine years of basic education, and should be provided for free at public schools (Government of Colombia, 1991). New funding arrangements between the central government, municipal governments, and public schools have also aimed to increase the efficiency of the education system. Currently, the Ministry of Education transfers funds to Departments[2] and Municipalities depending on the size of the student body and poverty rates. These subnational governments administer these funds, and oversee the selection and promotion processes of teachers. School principals have no power over these processes (OECD, 2016a).

[2]Departments are the highest level of government below the central government. There are 32 Departments in total in Colombia. Each department is further subdivided into municipalities. There are 1119 municipalities in Colombia.

Regarding teacher selection, in 2002 the Government introduced a reform that changed the public school system of selection and promotion of teachers by introducing a pre-employment test to be used as a screening device. Currently, two decrees regulate the teaching staff: Decree 2277 of 1979 (old decree) and Decree 1278 of 2002 (new decree).

Figure 3.1: Number of Primary and Secondary Teachers in Old and New Systems




*Notes:* Own calculations using data from ICFES. The graph on the left shows the number of teachers hired by year and by ladder in the old system. The old system has 14 steps, where 1 is the lowest latter and 14 is the highest one. The highest in the ladder, the more education or experienced the teacher is. The graph on the right depicts the same information for teachers ruled by the new system. As opposed to the old system, the new system has three well defined lanes (1,2,3), where a teacher can be placed based on his education. Within each lane, there are four steps (A, B, C, and D). Temporary positions are displaced as "SE" under the old system and as "0A" and "0B" under the new system.

Teachers that were hired in a public school before 2005 are governed by the old decree. They entered a lane based on their educational level, and could automatically increase a step in the teacher ladder by simply accruing 3-4 years of experience. In total, the old system had 14 ladders, as displayed in Table C1. Under this system, years of experience were the main factor driving career advancement and salary increases.[3] Due to this mechanism, Garcia et al. (2014) report that, by 2005, a large part of teachers in the old public system were concentrated in the highest ladders. Figure 3.1 provides support for this by showing that 67% of teachers in public schools were mainly in ladders six and above.

[3]Every year, the National Government determines teachers' wages for all ladders.

Those candidates who became teachers in public schools on or after 2006 are regulated by the new decree. This set of new teachers had to score 60 out of 100 on the standardized national test to be eligible to be hired, and 80 out of 100 to qualify for a promotion.[4] This system is still in place, but the score requirements and testing modules have changed over time. Figure 3.1 shows that, under the new system, most of the new teachers hired since 2005 hold a bachelor's degree, i.e., they entered lane 2A. However, qualifying for a promotion has been significantly more difficult. Even in 2013, eight years after the implementation of the first test, most teachers are still in lane 2A. This is partly because the test has proven to be difficult for most applicants. For example, when the test was first implemented, only 34% of preschool candidates, 37% of primary school candidates, and 57% of secondary school candidates obtained a score of 60 or higher out of 100 on their first attempt. This, combined with the fact that teachers are allowed to take the test once a year (rarely twice), meant that promotions were rare. In 2015, teachers' low possibilities of promotions unleashed protests that stopped school activities for weeks.[5]

The OECD (2016a) estimated that the transition between these two decrees would be completed by 2030. Two factors affect the length of the transition. First, only a small proportion of teachers regulated by the old decree have chosen to join the new system. Teachers who were already teaching in public schools by 2005 were not required to take the test at any point in their work-life. By 2014, only 28% of the teachers were governed under the new system, 53% were governed under the old system, and 19% were temporary positions that were not regulated (OECD, 2016a). Second, incumbent teachers may have incentives to continue working after retirement age because delaying retirement increases their pensions. The fact that old and new teachers are governed under two different sets of rules has generated a distressing situation in the public-sector teaching staff because the 2002 reform has tougher teacher evaluation processes that teachers consider unfair (Cifuentes, 2013).

The pre-employment test required new teachers to be evaluated on four knowledge areas, each

[4] As mentioned earlier, teachers hired on or before 2005 were not held to such standards.

[5] See https://www.elespectador.com/educacion/se-levanta-paro-de-maestros-tras-acuerdo-entre-gobierno-y-fecode-article-558965/

scored from 0 to 100: numerical reasoning (30 questions), verbal reasoning (30 questions), subject-area knowledge (40 questions) and pedagogical theory (40 questions) (National Civil Service Commission, 2010). The first evaluation took place in January 2005. During its implementation, there were multiple protests by the teachers' union, which strongly opposed the new policy. Indeed, protests often take place around the time the tests are administered. The teachers' union argues that the evaluations are poorly designed and do not improve their teaching skills because instead of obtaining feedback, they only receive a score (Cuevas, 2010). Moreover, the test requires theoretical knowledge of pedagogy (A sample question of this section is displayed in table C2). Therefore, every candidate needs to be proficient in subject-area knowledge and pedagogy theory. Professionals in education and professionals in subject-matter careers (for example, a person with a B.A. in mathematics) have the potential to score favorably on subject-area tests. But only those who study education degrees are competent to obtain favorable results on pedagogy tests. Thus, the pedagogy component, which accounts for more than 28% of the questions in the test, favor teachers with a degree in education.

The standardized test takes place whenever the National Government deems it necessary, often once a year. Candidates voluntarily register and pay a fee to take the test. The fee was COP 34,500 in 2016 (11.3 USD or 2.7% of the monthly salary of teachers at the lowest ladder). At the time of registration, candidates must reveal the level of education for which they are applying to become teachers (preschool, primary or secondary) and the area of knowledge that they wish to specialize in (math, science, physical education, etc.). The subject area portion of the test is tailored to what the candidate reported at registration. All tests are collected and graded by ICFES (*Instituto Colombiano para la Evaluación de la Educación*), the institution that is also responsible for designing standardized tests for students. Teachers with scores equal to or above 60 points out of 100 are placed on the list of eligible candidates in the municipalities or jurisdictions of their choice. They can remain on the list for up two years. Those public schools that need new hires must inform the subnational governments. The latter uses the list of eligible candidates and selects the new hires based on their ranking within the list. Selected candidates are on probation during the first year (Ministry of Education, 2002).

## 3.3 Data and Descriptive Statistics

### 3.3.1 Teachers' Data

To identify which schools hired teachers in the 2005-2013 period, I use the annual Census of Schools. This dataset, provided by the Department of Statistics, contains annual detailed information on the number of staff and students in private and public schools. The dataset reports the number of teachers working in each school, by gender. This paper uses the data only for the public schools, given that the hiring reform affected only candidates in the public system. For public schools, the database also provides information on the number of teachers by type of hiring system (old vs. new system) and ladder. Note that the information in this database is aggregated at the school level, thus, it is not possible to observe to which grades certain teachers are assigned.

Figure 3.1 shows the number of teachers hired under each system. In 2005 there were close to 400,000 teachers in the public system, 67% of whom were classified at scale 6 or higher. These teachers, who were hired before 2006, were not affected by the new reform. This explains why, in 2013, teachers hired under the old system still represent about 71% of the teaching staff. Given the automatic promotion of teachers in the old system, based on experience, the proportion of teachers in the higher ladders (in green) is 47% much greater than the share of teachers in lower ladders (yellow and blue).

Table 3.1: Descriptive Statistics: Teachers Test Scores

| Variables | 2005-1 | 2005-2 | Difference |
|---|---|---|---|
| Preschool | | | |
| Above eligibility threshold (%) | 33.69 | 16.95 | $-16.74^{***}$ |
| Above wage increase threshold (%) | 0.00 | 0.00 | 0.00 |
| Verbal reasoning score | 57.69 | 53.79 | $-3.90^{***}$ |
| Numerical reasoning score | 55.62 | 51.77 | $-3.85^{***}$ |
| Subject-area score | 59.90 | 55.49 | $-4.41^{***}$ |
| Pedagogy theory score | 58.70 | 56.38 | $-2.32^{***}$ |
| Female (%) | 99.10 | 98.18 | $-0.92^{***}$ |
| Disabled (%) | 0.58 | 0.04 | $-0.54^{***}$ |
| Age | 33.15 | 33.28 | 0.13 |
| Primary | | | |
| Above eligibility threshold (%) | 36.73 | 13.40 | $-23.33^{***}$ |
| Above wage increase threshold (%) | 0.01 | 0.00 | $-0.01^{**}$ |
| Verbal reasoning score | 57.96 | 54.03 | $-3.93^{***}$ |
| Numerical reasoning score | 57.20 | 53.64 | $-3.56^{***}$ |
| Subject-area score | 59.95 | 51.52 | $-8.43^{***}$ |
| Pedagogy theory score | 59.38 | 57.39 | $-1.99^{***}$ |
| Female (%) | 80.59 | 76.99 | $-3.60^{***}$ |
| Disabled (%) | 0.48 | 0.02 | $-0.46^{***}$ |
| Age | 31.85 | 32.57 | $0.72^{***}$ |
| Secondary | | | |
| Above eligibility threshold (%) | 56.99 | 40.07 | $-16.92^{***}$ |
| Above wage increase threshold (%) | 0.04 | 0.00 | $-0.04^{***}$ |
| Verbal reasoning score | 61.66 | 58.24 | $-3.42^{***}$ |
| Numerical reasoning score | 62.48 | 59.41 | $-3.07^{***}$ |
| Subject-area score | 59.92 | 55.49 | $-4.43^{***}$ |
| Pedagogy theory score | 60.13 | 57.83 | $-2.30^{***}$ |
| Female (%) | 52.80 | 50.48 | $-2.32^{***}$ |
| Disabled (%) | 0.56 | 0.03 | $-0.53^{***}$ |
| Age | 33.42 | 32.74 | $-0.68^{***}$ |

Source: ICFES. This table describes the data available in the first year of testing applicants. Applicants disclosed at registration the level of education at which they aspired to become teachers (preschool, primary or secondary education). *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Data on applicants' test scores are available only for the first year of the test, and were obtained from ICFES. However, these data are not used in the regression analyses since it was not possible to match applicants to schools because there is no public information that would allow one to identify which schools hired the applicants who passed the test. Nevertheless, these data provide important insights into how teachers performed on the first wave of the test. Table 3.1 presents descriptive statistics from this dataset for two exams that took place in 2005. The first test took place in January of 2005 (2005-1 in the table). Due to candidates' discontent with the test and their poor performance, a second test was administered in December of 2005. The second test (2005-2) was used to create the list of eligible candidates from which new hires were drawn in 2006.

Table 3.1 shows that, in 2006, only 17% and 13% of the candidates who applied to positions in preschool and primary schools, respectively, obtained a score above the eligibility threshold (60 points). Moreover, if these candidates had requested a promotion, none of them would have qualified for a salary increase (none of them obtained a score of 80 points or higher). Performance on the test is a little better for secondary school job candidates. Many more candidates passed the eligibility threshold, but less than 1% qualified for a salary increase. Regarding the demographic characteristics of job candidates, most of them are between 31 and 34 years old. Preschool and primary school job candidates were mostly female (98% and 77%, respectively), while job candidates for secondary school positions were almost equally distributed by gender (50.5% were female).

### 3.3.2 Students' Data

ICFES administers a standardized test to Grade 11 students every six months,[6] called Saber 11. Performance on this test is usually used as a requirement for high school graduation and college admission. Public schools, which use Calendar A, typically take the test during the second half of the year (around October). Students take the test on the day and place assigned by ICFES. The test requires the attendance of students for a full day at the assigned site. This site often does not correspond to their own school. The test lasts 9 hours divided into two sections of 4 hours and 30 minutes each. The test has multiple components, including math, language, biology, philosophy, physics and chemistry.

[6]Colombia has two academic calendars: A and B. The school year in Calendar A starts in February and ends in November. The school year in Calendar B begins in September and ends in May. ICFES offers the Saber 11 test twice a year so that students from both calendar years can take the test in their respective second semesters.

In this paper, I use student-level data from students taking the Saber 11 test between 2005-2013. I use the subsample of students in public schools who are between 14 and 20 years old when they took the first test for the analysis. Each component of the test was standardized by wave, given that the test questions change in each wave of the exam. I also computed a global score, which is the result of standardizing the sum of the scores in all tests.

Table 3.2: Descriptive Statistics of High-School Students in 2005, by School-Treatment Status in 2006

| Variable | Hired in 2006<br>(1) | Did not hire in 2006<br>(2) | Difference<br>(3) | p-value<br>(4) |
|---|---|---|---|---|
| Female | 54.6% | 54.5% | -0.1% | 0.864 |
| Age (years) | 17.2 | 17.3 | 0.1 | 0.000 |
| Household Size | 5.2 | 5.0 | 0.2 | 0.000 |
| Internet in the household | 14.8% | 14.5% | 0.3% | 0.669 |
| Low income household | 78.8% | 80.9% | -2.2% | 0.039 |
| Father with secondary or less | 92.3% | 91.0% | 1.3% | 0.000 |
| Number of students per school | 65.9 | 66.2 | -0.3 | 0.425 |
| | | | | |
| Students in the sample | 85,698 | 182,329 | | |
| Schools in the sample | 1,329 | 4,263 | | |

Note: Descriptive statistics using administrative student-level data from ICFES in 2005. Column 1 shows descriptive statistics for students in schools that hired teachers in 2006, while column 2 summarizes demographics for students in schools that did not hire teachers in 2006. The difference between these two columns is shown in column 3. P-values of these differences, computed with standard errors clustered at the school level, are shown in column 4. Low income households are those households in stratums 1 and 2 (out of 6 possible income stratums).

Table 3.2 shows descriptive statistics of students that took the Saber 11 test in 2005. Column 1 shows descriptive statistics for those students in schools that hired teachers in 2006 and column 2 for those in schools that did not. The table shows statistically significant differences between both groups of children; however, those differences are minor in size. For example, children in schools that hired teachers in the first year after the implementation of the new system are 0.1 years younger, have on average 0.2 more family members, are 2 percentage points less likely to be in a low-income household and are 1.3 percentage points more likely to have a father with secondary education completed or less education. These differences, although significant, are not large. I do not see significant differences in terms of the gender distribution at school, the availability of

Internet on the household or on the number of students per school. For that reason, I conclude that there is no evidence that schools that hired teachers in 2006 were different from schools that did not hire teachers.

## 3.4 Identification Strategy

The main specification exploits school-level variation in the number of teachers hired under the new system in the first year after implementing the policy (2006). This specification compares students in schools that hired new teachers and those in schools that did not hire new teachers using test scores from 2005 and 2006, the last year before and the first year after the implementation of the policy. The objective of this specification is to identify the first-year effects of the new policy, which allows one to clearly identify a group of schools that did not hire new teachers. This group will be the comparison group. The specification of the difference-in-differences model is as follows:

$$Y_{ist} = \beta(Fraction\ NS_{s,2006} \times Y2006_t) + X_i'\gamma + \tau_t + \mu_s + \varepsilon_{ist}, \ \ \forall t = 2005, 2006 \tag{3.1}$$

where $Y_{it}$ is the standardized test score of student i in school s in the year t. Notice that each student only takes the Saber 11 exam once, thus, each i is observed only once. $Fraction\ NS_s$ is the number of teachers reported to be classified in the new system in 2006 as a share of the total number of teachers in that year in school s. $Y2006_t$ takes the value of 1 in year 2006 and 0 in 2005. Finally, $X_i$ is a set of student-level controls that include gender and age, $\tau_t$ is a year-wave fixed effect, and $\mu_s$ is a school-level fixed effect. Standard errors for individual-level regressions are clustered at the school-level. The main effect is measured by $\beta$. It measures how an increase in one percentage point in the share of newly hired teachers changes students' standardized tests scores. In other words, $\beta$ is estimated by comparing the change in test scores of students in schools that hired new teachers, accounting for the intensity of hiring those new teachers.

One concern with this specification is that nothing prevented incumbent teachers from taking the test, and therefore, being ruled by the new system. The voluntary transition of teachers from the old system to the new system could contaminate the treatment group and underestimate the effect of new hires. However, it is unlikely this is the case despite the fact that entry-level salaries in the new system were 2-8% higher. To see why, note that after passing the exam and being hired, teachers entered a probationary period. After this probationary period, they are subject to an annual performance evaluation that assesses teaching and behavioral competencies. The annual performance evaluation has no implications for career advancement. In the worst case scenario, a teacher who switches from the old system to the new system who obtained an unsatisfactory result in the performance evaluation would have to make a personal and professional improvement plan. While this seems to be a minor consequence, the reputational costs of the exam can be very high and could have persuaded many incumbent teachers to remain in the old system. On the other hand, career advancement in the old system required only the accumulation of experience regardless of the quality of the service provided, while career advancement under the new system required taking a standardized exam and passing it with 80 out of 100 points, which rarely happened. That this probability is low is also supported by anectotal evidence, which confirms that incumbent teachers were unlikely to take the test (Cuevas, 2010).

Another concern is that schools that hire teachers in 2006 would have been performing worse or better than schools that did not hire teachers. In particular, one of the requirements of the difference-in-differences model is that the academic performance of schools that hired teachers in the new system could not have been progressing differently from the academic performance of schools that did not hire teachers in 2006. However, I cannot test this hypothesis because I do not have historical data on student performance prior to 2005. Despite not being able to test this hypothesis, Table 3.2 showed that students from both groups did not have notable demographic differences.

A final concern is that the hiring of teachers would have taken place only in schools located in regions of lower educational quality. As shown in Figure C1, municipalities involved in the hiring process vary nationally. Still, those municipalities that hired teachers and those that did not may differ in observable and unobservable characteristics. Table C4 displays descriptive statistics comparing these two groups. As can be seen, municipalities that did not hire teachers in 2006 are richer, larger, have fewer police stations, higher unemployment rates, and higher density. For the specifications I estimate, I use control variables and school fixed effects to account for these differences. All specifications include either school-level or municipality-level fixed effects, using the former whenever possible.

Given that on 2005 no schools had teachers hired through the new system, the first-year results offer the the cleanest effect of the policy. However, it may take some time for teacher recruitment to bear fruit. For that reason, this paper also presents evidence about the results of implementing the employment test eight years after it was implemented. The specification of the 2005-2013 repeated cross-sectional data is similar to the one explained in Equation 3.1. In particular, I estimate the following equation:

$$Y_{ist} = \beta(Fraction\ NS_{st}) + X_{ist}^{'}\gamma + \tau_t + \mu_s + \varepsilon_{ist}, \ \ \forall t \in [2005, 2013] \tag{3.2}$$

where $Fraction\ NS_{st}$ is the number of teachers in the new system at school $s$ and time $t$ divided by the total number of teachers in school $s$ at time $t$. Notice that this ratio takes the value of 0 for all schools when $t = 2005$. This regression also includes as explanatory variables the age and gender of the student. Unfortunately, there are no controls varying over time and across municipalities that can be included due to the lack of such data for the study period. Time and school fixed effects are included in this specification to control for observable and unobservable variables that have changed over time for all schools, and that vary between schools at all points in time. Because there is a longer observation time, these estimators are expected to be more precise about what happened to school performance in the medium run. In this case, $\beta$ shows how student performance changes as teacher hiring intensity changes. The next section presents the results.

## 3.5 Results

### 3.5.1 One-Year After the Implementation of the Policy

Table 3.3 presents the results of comparing the scores of students who took the Saber 11 test in 2005, before the implementation of the policy, and in 2006, after the implementation. The coefficient displayed is the estimated $\beta$ coefficient from a linear regression as shown in Equation 3.1. That is, each regression uses individual-level data, controls for the individual's age and gender, and includes year-wave and school fixed effects.

Table 3.3: Individual-Level Results: One Year After Implementation

| Test | $Fraction\ NS_s$ $\times Y2006_t$ (1) | Robust p-value (2) | Romano-Wolf p-value (3) |
|---|---|---|---|
| Global | 0.031 | [0.379] | $\{0.765\}$ |
| Math | 0.006 | [0.865] | $\{1.000\}$ |
| Language | -0.036 | [0.294] | $\{0.765\}$ |
| Biology | 0.071 | [0.014] | $\{0.078\}$ |
| Philosophy | 0.040 | [0.289] | $\{0.765\}$ |
| Physics | -0.004 | [0.909] | $\{1.000\}$ |
| Chemistry | 0.072 | [0.085] | $\{0.353\}$ |

*Notes:* The results in each row of this table come from independent linear regressions in which the variable of interest is students' standardized test scores in the related subject. The explanatory variable is computed as the number of teachers registered in the new system as a percentage of the size of the teaching staff. This ratio takes the value of zero for those schools that did not hire teachers in 2006 and for all schools in 2005. Estimates are shown in Column 1. Each regression uses school-level and year-wave fixed effects. Column 2 presents p-values computed with robust standard errors, displayed in brackets. Column 3 presents p-values computed with the Romano-Wolf correction in curly brackets. The number of observations in all regressions is 529,584.

The results show that one year after the change in the teacher hiring policy, there are no significant effects on tests scores in either language or mathematics, which are usually the subjects in which the curriculum is quite similar between schools. The results are also generally insignificant for the student tests in other subject areas, for which curricula widely varies across schools. There is only a modest positive result for the biology test, which implies that a 1 percentage point increase in the number of teachers hired under the new system would generate an increase in biology tests scores of 0.00071 standard deviations. This effect not only is quite small, but is only significant at the 5% level. The effects pare less significant after applying the Romano-Wolf correction. In the case of the results of biology tests, the significance with the Romano-Wolf correction drops to the 10% level. The Romano-Wolf correction (asymptotically) calculates the probability of rejecting at least one true null hypotheses in a set of hypothesis under test, also known as the familywise error rate (FWER).

These results are to be expected, for several reasons. First, given that I do not observe student-teacher matches, it is not clear from the data whether the students who took the test were exposed to the new teachers hired under the new system. This means that, even if the school was treated, the student within the school might not have been treated. Second, given that the teachers hired are new to the public sector and most likely are new to the teaching profession, even if they are qualified, it may take time for them to reach their teaching potential. The literature in education asserts that experience plays an essential role in teachers' quality (Elacqua et al., 2017). These findings are also in line with Araujo et al. (2019)'s work one year after the implementation of the Ecuadorian teacher reform.

### 3.5.2 Eight Years After the Policy Changed

Table 3.4: Individual-Level Results: Eight Years After Implementation

| Test | $Fraction\ NS_{st}$ (1) | Robust p-value (2) | Romano-Wolf p-value (3) |
|---|---|---|---|
| Global | 0.022 | [0.049] | {0.078} |
| Math | -0.042 | [0.283] | {0.600} |
| Language | -0.047 | [0.274] | {0.600} |
| Biology | 0.057 | [0.104] | {0.518} |
| Philosophy | -0.004 | [0.934] | {1.000} |
| Physics | -0.025 | [0.523] | {0.899} |
| Chemistry | 0.079 | [0.071] | {0.427} |

*Notes:* The results shown in each row of this table come from independent linear regressions in which the variable of interest is the standardized tests scores of students in the corresponding subject. The explanatory variable shown in this table corresponds to the number of teachers registered in the new system as a percentage of the size of the teaching staff, which takes the value of zero for those schools that did not hire teachers in 2006 and for all schools in 2005. Estimates are shown in Column 1. Each regression uses school-level and year-wave fixed effects. p-values computed with robust standard errors are displayed in brackets. p-values computed with the Romano-Wolf correction are shown in column 3 in curly brackets. The number of observations in each regression is 2,898,252.

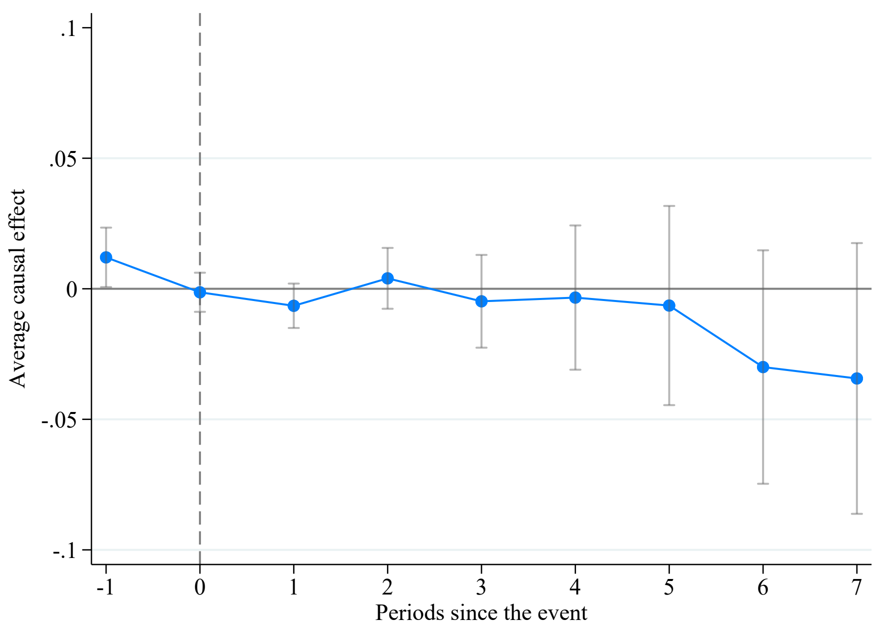
Focusing on longer-term results, Table 3.4 shows evidence that the progressive hiring of more teachers under the new system during the eight years following the policy's implementation has not affected students' test scores. In particular, for the global score, I find that a one percentage point increase in the number of teachers in the new system as a percentage of the total number of teachers generated an average increase in student scores of 0.00022 standard deviations after the reform. The size of this effect makes it economically insignificant, although statistically it is significant at the 95% level of confidence. After applying the Romano-Wolf correction, the significance level drops but passes the 10% level. Concerning math and language scores, I find small average decreases in student achievement, but these effects are not significant. Similarly, I find non-substantial changes in tests scores in biology, philosophy, physics, and chemistry, none of which are statistically significant.

The above specification allows one to evaluate the average effect of hiring intensity on student outcomes after implementing the reform. However, one concern with this specification is that the treatment units were treated at different points in time. That is, schools began hiring teachers from the new system in different years according to their needs. Given this staggered adoption of teachers under the new system, a two-way fixed effects model may generate biases in the estimation, even in the presence of parallel trends. This can happen when treatment effects are dynamic and take some time to be realized. In the case I am studying, this can happen because new teachers, although they may be better selected and therefore could have better potential quality, do not have teaching experience, which is key to being a good teacher. Given that the effect shown in the table is only the average effect over the 8-year period after the hiring of teachers, the effect is less likely to be close to zero if these better-trained teachers turn out to be better than incumbent teachers once they have acquired some experience.

One of the approaches to estimate dynamic effects is to implement an event study using a centered time variable. For each school, this variable is centered at 0 when the school began to hire teachers. That is, the event variable $\tau_t$ is 0 when the school $s$ hired the first teacher of the new system, and can take a maximum value of 7, which corresponds to the maximum number of periods to which a school was exposed to the teachers of the new system. The period dummies resulting from this variable, interacted with hiring intensity, can help determine the effect of new hires on school performance one to eight years after the first hiring.

But implementing event studies in this context also has its challenges. For example, recent papers have identified that in a staggered specification, the treatment group in each period is compared to a control group that changes from period to period. That is, the estimated average treatment effect is a weighted sum of the effects of each treatment group in each period from several difference-in-differences estimators. De Chaisemartin and d'Haultfoeuille (2020) demonstrate that some of the weights in this computation may be negative due to these changes in the comparison group, which in turn, may eben lead to obtaining a negative ATE even when the treatment effect in each treatment period is positive. De Chaisemartin and d'Haultfoeuille (2020) propose a correction to this problem, which I use to estimate dynamic treatment effects, one period before and seven periods after treatment.

Figure 3.2: Chaisemartin and D'haultfeuille Staggered Event Study: Global Test Score

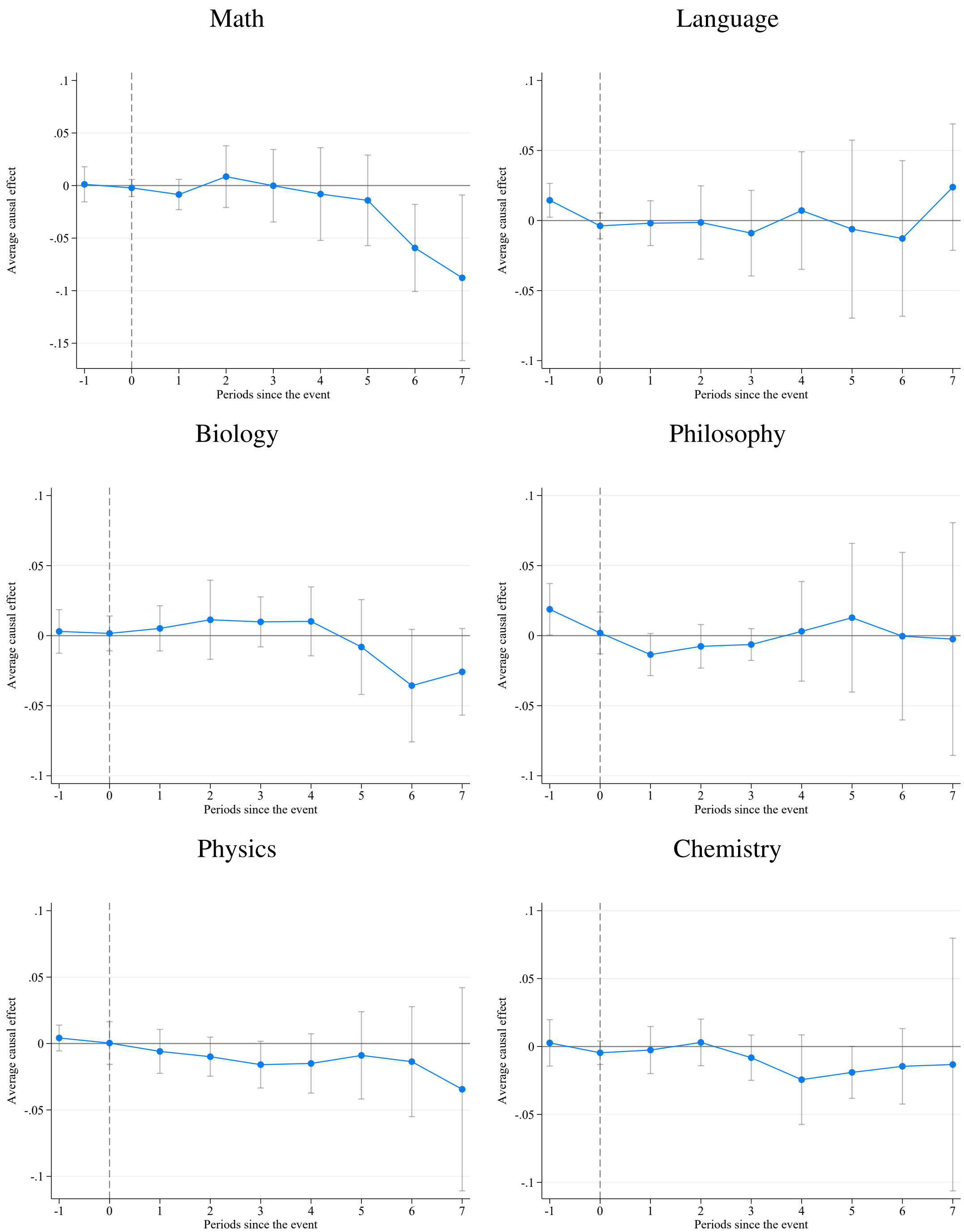



*Notes:* This graph shows the results of the event study estimation using the De Chaisemartin and d'Haultfoeuille (2020) methodology. The x-axis shows the number of periods that occurred from the time when the first hiring of teachers in the new system occurred. Each estimate, in blue, has an associated 95% confidence interval, in gray.

Figure 3.2 shows, for each period, the causal effects using this methodology on the global test, together with its 95% confidence interval. The graph shows the effects one year before hiring the first teacher (period -1), and 7 years after the year in which the first teacher was hired (periods 1-7). Period 0 corresponds to the year in which the first teacher from the new system was hired. No significant effects are found in any of the treatment periods, except for a marginally significant effect in the period *before* hiring. For this period, I find that the treatment group had relatively better scores than the control group. The graph also shows that there is a slightly decreasing trend that starts in years 6 and 7 after the implementation of the policy. However, the estimation of these coefficients is less accurate as the number of periods from time zero increases because there are fewer treated units with observations there.

Figure 3.3 shows the results of this methodology for each subject-area test score. In the case of math, effects are not found until years 6 and 7 after the first teacher of the new system is hired. For year 6, for example, I find that an increase of 1 percentage point in the number of teachers from the new hiring system generates a *decrease* of 0.0005 standard deviations in the math test. While the effect is significant, its size is quite small. No significant effects are found in any other test. These estimates may indicate that an evaluation that includes a longer period of time may be needed to observe the effects of this reform.

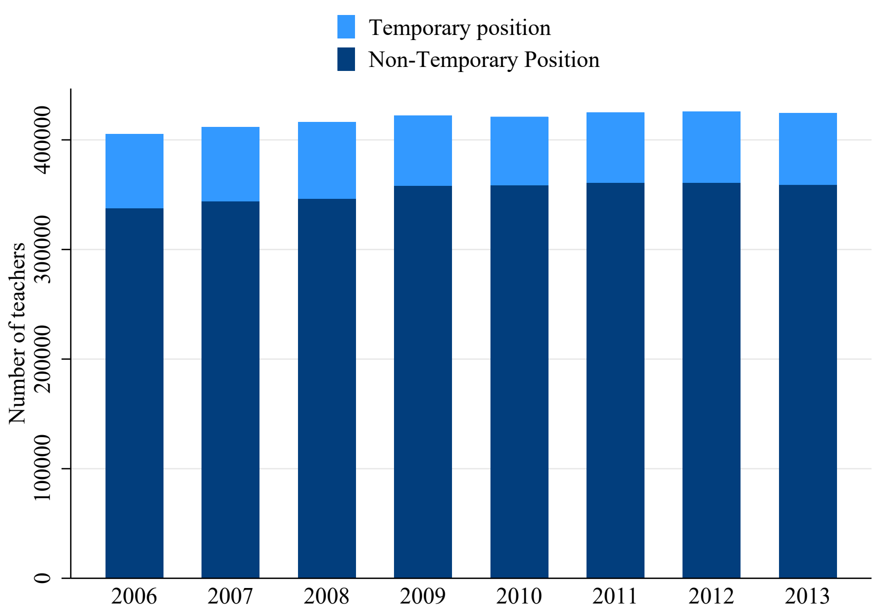
Figure 3.3: Chaisemartin and D'haultfeuille Staggered Event Study: Subject-Area Test Scores




*Notes:* This graph shows the results of the event study estimation using the De Chaisemartin and d'Haultfoeuille (2020) methodology. The x-axis shows the number of periods that occurred from the time when the first hiring of teachers in the new system occurred. Each estimate, in blue, has an associated 95% confidence interval, in gray.

One reason why there may be a lack of effects of this hiring system does not necessarily have to do with the fact that evaluating teachers by means of a test is not appropriate. It may be that by implementing this system, unintended side effects were generated. One of these unintended effects has to do with the hiring of teachers in temporary positions, who do not meet the qualifying criteria. Figure 3.4 shows that the percentage of teachers hired temporarily has remained stable since 2006, and represent a significant portion of the total number of teachers hired. That is, there is a possibility that the reason for not finding effects is not related to poor screening, but could potentially be associated with a lower quality of teachers hired in temporary positions.

Figure 3.4: Evolution of Temporary Positions Over Time

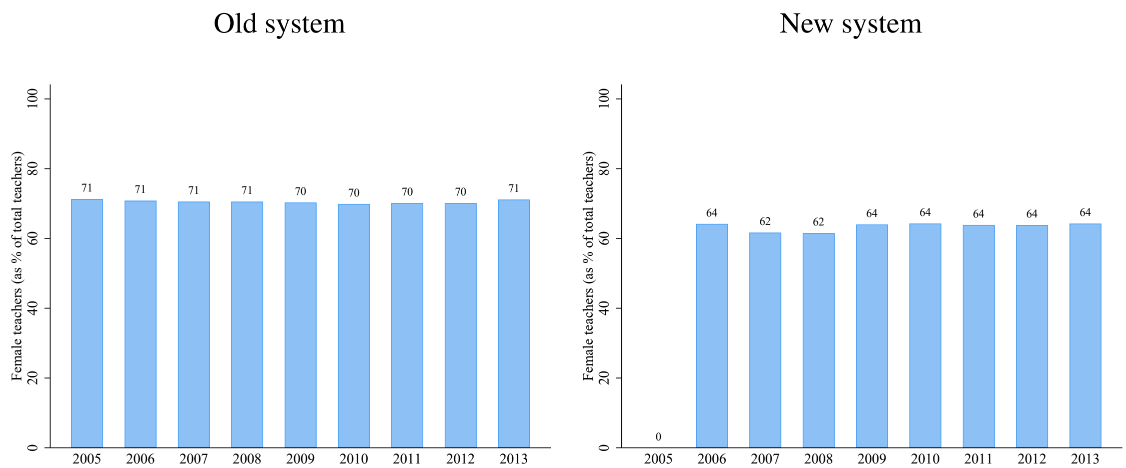



*Notes:* This graph shows the results of number of teachers over time, by type of contractual relationship.

### 3.5.3 Heterogeneous Effect by Gender of the Student

The teaching profession in Colombia has been characterized as being mainly a profession in which women are overrepresented, that is, there are proportionally more women than men in this profession. The Figure 3.5 shows that in 2005, 71% of school teachers were women. However, when comparing the gender distribution in the new system, there is a marked difference. For example, in 2006, the proportion of teachers hired in the new system who are women was 64%, compared to 71% of women in the old system. In other words, there is a relatively greater entry of men into the teaching profession after 2006.

Figure 3.5: Share of Female Teachers Over Time




*Notes:* This graph shows the number of female teachers hired in each hiring system, relative to the total number of teachers in the public sector.

Several recent studies have found that female teachers improve female student outcomes (Gong et al., 2018; Winkelmann, 2016), especially in mathematics (Hwang and Fitzpatrick, 2021). Therefore, the question that arises in the Colombian context is whether the reduction in the percentage of women hired generated adverse effects on female students relative to male students. With this in mind, I estimate the following equation to investigate whether if there are differential effects by gender of teacher hires:

$$Y_{ist} = \delta(Fraction\ NS_{st} \times Female_{ist}) + \mu(Percent\ NS_{st}) + X'_{ist}\gamma + \tau_t + \mu_s + \varepsilon_{ist},\ \ \forall t \in [2005, 2013] \tag{3.3}$$

The difference between equations 3.2 and 3.3 lies in the interaction term $Fraction\ NS_{st} \times$

$Gender_{ist}$. The estimate of $\delta$ will give us an estimate of the differential effect of female students compared to their male counterparts, in schools that had more teacher hires after 2006. The results of this model are shown in Table 3.5.

Table 3.5: Heterogeneous Effects by Gender: Eight Years After Implementation

| Test | $Female \times Fraction\ NS_{st}$ (1) | Robust p-value (2) | Romano-Wolf p-value (3) |
|---|---|---|---|
| Global | -0.032 | [0.000] | $\{0.020\}$ |
| Math | -0.213 | [0.000] | $\{0.020\}$ |
| Language | 0.071 | [0.000] | $\{0.020\}$ |
| Biology | 0.015 | [0.094] | $\{0.196\}$ |
| Philosophy | -0.041 | [0.000] | $\{0.020\}$ |
| Physics | -0.184 | [0.000] | $\{0.020\}$ |
| Chemistry | 0.007 | [0.426] | $\{0.490\}$ |

*Notes:* The results shown in each row of this table come from independent linear regressions in which the variable of interest is the standardized tests scores of students in the corresponding subject. The explanatory variable shown in this table corresponds to an interaction between the gender of the student and the number of teachers registered in the new system as a percentage of the size of the teaching staff. The latter takes the value of zero for those schools that did not hire teachers in 2006 and for all schools in 2005. Estimates are shown in Column 1. Each regression uses school-level and year-wave fixed effects. p-values computed with robust standard errors are displayed in brackets. p-values computed with the Romano-Wolf correction are shown in column 3 in curly brackets. The number of observations in each regression is 2,898,252.

I find small reductions in girls' school performance for mathematics, philosophy and physics tests caused by the increased recruitment of teachers in the new system. In math, for example, the hiring of one percentage point more teachers in the new system generated a reduction in the girls' test score of 0.0021 standard deviations compared to boys. In the case of language, there is evidence of a relative improvement of girls in the test, but the effect size is very small: 0.00071 standard deviations per each increase in one percentage point in the share of new hires. These effects, although significant, are quite moderate and very close to zero in size. One would also like to know if these results are due to different impacts between boys and girls when analyzed separately. Table C6 show separate regressions for boys and girls of being exposed to more newer teachers. The specification controls for the same charateristics as the model used in Table 3.5. The model shows that there is an improvement in the test score of boys, especially in math language and biology. In contrast, girls in schools that hired more teachers are doing worse in math, philosophy, physics and chemistry. Therefore, the differential effect observed in Table 3.5 can be attributed to two mechanisms: boys are improving their performance in certain tests, but this improvement does not outweigh the worsening of tests scores of girls due to the hiring of new teachers. Lower hiring of female teachers may have to do with the fact that they tend to perform worse on tests compared to their male counterparts. To assess whether this is the case, I use the test scores of candidates who took the test in 2005 using a linear regression model as follows:

$$Y_{jmt} = \beta Female + X_j^{'}\gamma + \tau_t + \mu_m + \varepsilon_{jmt} \tag{3.4}$$

where $Y_{jmt}$ is the standardize test score of a component of the test or a dummy taking the value of 1 if the candidate passed the eligibility thresholds, and zero otherwise. Since there were two waves of testing in 2005, wave fixed effects are included ($\tau_t$). Each regression controls for teacher's age, disability status and municipality of residence. $\beta$ is the estimate of how different the test results were for female versus male candidates.

Table 3.6: Correlates of Teacher Test Score Results with Gender in 2005

| Test | $\beta_{female}$ (1) | Robust p-value (2) | Romano-Wolf p-value (3) |
|---|---|---|---|
| *Panel A. Standardized Test Scores* | | | |
| Verbal | 0.020*** | [0.009] | 0.059 |
| Numeracy | -0.400*** | [0.000] | 0.020 |
| Subject area | -0.115*** | [0.000] | 0.020 |
| Pedagogy | -0.084*** | [0.000] | 0.020 |
| Average | -0.213*** | [0.000] | 0.020 |
| *Panel B. Probability of passing the threshold* | | | |
| $Score \geq 60$ | -0.085*** | [0.000] | 0.019 |
| $Score \geq 80$ | -0.000*** | [0.000] | 0.019 |

*Notes:* The results shown in each row of this table come from independent linear regressions using observations from 269,488 candidates. The variables of interest in Panel A are the standardized test scores of teachers in the respective component of the test. The dependent variables of the regressions in Panel B are dummies that take the value of 1 if the teacher met the score criteria, and zero otherwise. Each regression controls for the wave of the exam, and teacher's age, disability status and municipality of residence.

Table 3.6 shows that females performed better than males only in the verbal aptitude test (0.020 standard deviations above). However, in the other three components of the test, females scored significantly lower than males. For example, in the numeracy component, women's scores were 0.4 standard deviations below those of men. When taking the test, each candidate declared the area of knowledge to which he/she wanted to apply (e.g., mathematics, science, biology, etc.). When looking at the test results in the area component, women scored 0.11 standard deviations lower than men. Similarly, in the pedagogical theory component, females scored 0.08 standard deviations lower than males. When I look at the probability of passing the eligibility tests, I find that the probability of women passing the hiring threshold (60) is 8.5 percentage points lower than that of men. While the effect on the probability of passing the career advancement threshold is significant, the size of the effect is negligible.

### 3.5.4 Results in Dropout Rates

The school census also has additional information about student attrition rates during the first year of program implementation. Table C5 presents linear regressions at the school level, which for each grade level regresses dropout rates on the percentage of teachers in the new system who arrived at the school in 2006. Although significant effects are found, these effects are small. In general, we see that there is a small drop in the school dropout rate. This drop appears to be more prevalent among younger students (those in grades 5 and below). Unfortunately, these data are only available for 2005 and 2006, so I am unable to make estimates for a longer period of time.

## 3.6 Final Remarks

In this paper, I evaluate the effect of a reform in the hiring process of teachers in public schools in Colombia that began to be implemented in 2006. The reform required teachers to be hired through a screening process that required them to take a test to evaluate their skills. Only new candidates were required to take the test, while incumbent teachers would continue to be promoted in their profession according to the old promotion system. Using difference-in-differences and event study estimates, I compared students in schools that hired relatively more teachers since 2006 with students in those schools that did not hire, or hired at lower rates.

I find that one year after the implementation of the reform there are no effects. Eight years after the implementation of the reform, no effects are found either. When looking at the dynamic effects of the reform, no effects are found in most periods. However, a negative but not significant trend begins to be seen six and seven years after the first recruitment took place in math test scores. These results contrast with those found by Brutti and Sanchez (2017), who run school-level estimates over the same period of analysis to identify the effects of this reform in Colombia, and found small positive and significant effects of the reform. It is worth mentioning that when running regressions at the school level,[7] I find similar results to those found at the individual level and have not been able to replicate the results of this paper. While the current research paper improves on the estimates provided by them by focusing on individual-level outcomes, it should be noted that there are data restrictions that limit the ability to make better inference in this paper. For example,

[7] Available upon request

the unavailability of data on the match between students and hired teachers does not allow me to quantify the effect of direct exposure to new teachers. Also, it should be noted that the effects I see are those on high school seniors. Not being able to see the results in students in other grades does not allow me to see differential effects between younger students compared to older students.

These limitations aside, the results from this paper are insightful but do not necessarily imply that implementing a recruitment process using screening tests does not work. Pre-employment tests that do not accurately test proxies of productivity may be harmful. One of the complaints of candidates at the beginning of the reform was that the tests were very complex, especially caused by the test in pedagogical theory. This section of the test did not benefit those professionals in specific subject-areas of knowledge (for example a B.A. in Math), while it did benefit those who were professionals in education. However, it is not clear that not knowing pedagogical theory harms teacher quality.

In addition, the difficulty in hiring teachers who passed the exam generated secondary effects. In particular, the high percentage of temporary teachers hired after the reform is noteworthy. About 19% of teachers in 2014 were unregulated temporary positions (OECD, 2016a). These hires did not go through the test screening process and, therefore, it is not possible to know or guarantee the quality of these teachers. In general, more evidence needs to be generated using more detailed data and over a longer time range to improve the results presented in this paper.

# References


**Abadie, Alberto and Javier Gardeazabal**, "The economic costs of conflict: A case study of the Basque Country," *American economic review*, 2003, *93* (1), 113–132.

**Allen, Maha Kattaa Patrick Daru Ait Ali Slimane Fiona**, "New economic opportunities for Jordanians and Syrian refugees," 2020.

**Angrist, Joshua D. and Jonathan Guryan**, "Teacher Testing, Teacher Education, and Teacher Characteristics," *The American Economic Review*, 2004, *94* (2), 241–246.

**_ and Jorn-Steffen Pischke**, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

**Arango, Diana, Lucia Hanmer, Eliana Rubiano, and Julieth Santamaria**, "How does poverty differ among refugees Taking an age and gender lens to the data," *World Bank Publications, forthcoming*, 2018.

**Araujo, P, Maria Daniela et al.**, "Measuring the effect of competitive teacher recruitment on student achievement: Evidence from Ecuador," Technical Report, Bamberg University, Bamberg Economic Research Group 2019.

**Autor, David and David Scarborough**, "Does job testing harm minority workers? Evidence from retail establishments," *The Quarterly Journal of Economics*, 2008, *123* (1), 219–277.

**Bahar, Dany, Ana Ibanez, and Sandra Rozo**, "Give Me Your Tired and Your Poor: Impact of a Large-Scale Amnesty Program for Undocumented Refugees," 2020.

**_ , Prithwiraj Choudhury, and Hillel Rapoport**, "Migrant inventors and the technological advantage of nations," *Research Policy*, 2020, p. 103947.

**_ , T Piccone, and Harold Trinkunas**, "Venezuela: a path out of misery," *Brookings. edu*, 2018.

**Ballou, Dale**, "Do Public Schools Hire the Best Applicants?," *Quarterly Journal of Economics*, 1996, *111* (1), 97–133.

**Banerjee, Abhijit and Esther Duflo**, "The economic lives of the poor," *Journal of Economic Perspectives*, 2007, *21* (1), 141–168.

**Bank, World**, *Hashemite Kingdom of Jordan: Understanding How Gender Norms in MNA Impact Female Employment Outcomes*, World Bank, 2018.

**Barbelet, Veronique, Jessica Hagen-Zanker, and Dina Mansour-Ille**, "The Jordan Compact," *Lessons learnt and implications for future refugee compacts. Briefing Papers*, 2018.

**Bardhan, Pranab and Christopher Udry**, *Development Microeconomics*, Oxford University Press, 1999.

**Basso, Gaetano, Francesco D'Amuri, and Giovanni Peri**, "Immigrants, labor market dynamics and adjustment to shocks in the Euro Area," *IMF Economic Review*, 2019, *67* (3), 528–572.

**Bellamy, C, S Haysom, C Wake, and V Barbelet**, "The lives and livelihoods of Syrian Refugees: a study of refugee perspectives and their institutional environment in Turkey and Jordan," *HPG Commissioned Report. London: ODI*, 2017.

**Bernstein, Shai, Rebecca Diamond, Timothy McQuade, Beatriz Pousada et al.**, "The contribution of high-skilled immigrants to innovation in the United States," Technical Report 2018.

**Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan**, "How Much Should We Trust Differences-in-Differences Estimates?," *The Quarterly Journal of Economics, MIT Press*, 2004, *119* (1).

**Bjorkman, Martina**, "Income Shocks and Gender Gaps in Education: Evidence from Uganda," *Seminar Papers from Stockholm University, Institute for International Economic Studies*, 2006, (744).

**Borjas, George J**, "Self-selection and the earnings of immigrants," Technical Report, National Bureau of Economic Research 1987.

__ , "The economics of immigration," *Journal of economic literature*, 1994, *32* (4), 1667–1717.

__ , "The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market," *The quarterly journal of economics*, 2003, *118* (4), 1335–1374.

**Boustan, Leah Platt, Price V Fishback, and Shawn Kantor**, "The effect of internal migration on local labor markets: American cities during the Great Depression," *Journal of Labor Economics*, 2010, *28* (4), 719–746.

**Brown, Caitlin and Dominique Van de Walle**, "Headship and poverty in Africa," *The World Bank Economic Review*, 2021, *35* (4), 1038–1056.

**Brown, Harry, Nicola Giordano, Charles Maughan, and Alix Wadeson**, "Vulnerability assessment framework: population study 2019.," 2019.

**Brück, Tilman and Kati Schindler**, "The impact of violent conflicts on households: What do we know and what should we know about war widows?," *Oxford Development Studies*, 2009, *37* (3), 289–309.

**Brücker, Herbert and Elke J Jahn**, "Migration and wage-setting: reassessing the labor market effects of migration," *Scandinavian Journal of Economics*, 2011, *113* (2), 286–317.

**Brutti, Zelda and Fabio Sanchez**, "Does Better Teacher Selection Lead to Better Students? Evidence from a Large Scale Reform in Colombia," *Documento CEDE*, 2017, (11).

**Cameron, Maxwell A. and Flavie Major**, "Venezuela's Hugo Chavez: Savior or Threat to Democracy," *Latin American Research Review*, 2001, *36* (3), 255–266.

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

— , "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

**Carletto, Gero, Marco Tiberti, and Alberto Zezza**, "Measure for measure: Comparing survey based estimates of income and consumption for rural households," *The World Bank Research Observer*, 2022, *37* (1), 1–38.

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

**Caruso, German, Christian Gomez Canon, and Valerie Mueller**, "Spillover effects of the Venezuelan crisis: migration impacts in Colombia," *Oxford Economic Papers*, 2019.

**Carvalho, Carlos, Ricardo Masini, and Marcelo C. Medeiros**, "ArCo: An artificial counterfactual approach for high-dimensional panel time-series data," *Journal of Econometrics*, 2018, *207* (2), 352 – 380.

**Ceritoglu, Evren, H. Burcu Gurcihan Yunculer, Huzeyfe Torun, and Semih Tumen**, "The impact of Syrian refugees on natives' labor market outcomes in Turkey: evidence from a quasi-experimental design," *IZA Journal of Labor Policy*, Jun 2017, *6* (1), 5.

**Chaisemartin, Clément De and Xavier d'Haultfoeuille**, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–96.

__ , **Xavier D'Haultfoeuille, and Yannick Guyonvarch**, "DID_MULTIPLEGT: Stata module to estimate sharp Difference-in-Difference designs with multiple groups and periods," 2019.

**Cifuentes, Carolina**, "Estatuto de Profesionalización Docente en Colombia. Análisis de los dos estatutos vigentes. Diferencias en la labor docente," *Universidad National de Colombia*, 2013.

**Cochran, William and Donald Rubin**, "Controlling Bias in Observational Studies," *Sankhya: The Indian Journal of Statistics*, 1973, *Series A* (35), 417–446.

**Cruz-Aguayo, Yyannú, Pablo Ibarrarán, and Norbert Schady**, "Do tests applied to teachers predict their effectiveness?," *Economics Letters*, 2017, *159*, 108–111.

**Cuevas, Maria**, "Los mejores maestros critican la evaluación docente," *El Espectador*, 2010.

**Culcasi, Karen**, ""We are women and men now": Intimate spaces and coping labour for Syrian women refugees in Jordan," *Transactions of the Institute of British Geographers*, 2019, *44* (3), 463–478.

**Cutler, David M, Lawrence F Katz, David Card, and Robert E Hall**, "Macroeconomic performance and the disadvantaged," *Brookings papers on economic activity*, 1991, *1991* (2), 1–74.

**Dal-Bo, Ernesto and Pedro Dal-Bo**, "Workers, Warriors and Criminals: Social Conflict In General Equilibrium," *Journal of the European Economic Association*, 2011, *9* (4), 646–677.

**DANE**, "Rueda de prensa. Resultados Mercado Laboral," https://www.facebook.com/watch/live/?v=3366521323562668&ref=watch_permalink 2019.

**Darling-Hammond, Linda**, "Teacher quality and student achievement," *Education policy analysis archives*, 2000, *8*, 1.

**Deaton, Angus**, "Saving and Income Smoothing in Cote d'Ivoire," *Journal of African Economies*, 1992, *1* (1), 1–24.

__ **and Salman Zaidi**, *Guidelines for constructing consumption aggregates for welfare analysis*, Vol. 135, World Bank Publications, 2002.

**Dekker, Rianne and Godfried Engbersen**, "How social media transform migrant networks and facilitate migration," *Global Networks*, 2014, *14* (4), 401–418.

**Dercon, Stefan**, "Wealth, risk and activity choice: cattle in Western Tanzania," *Journal of Development Economics*, February 1998, *55* (1), 1–42.

**Dinero**, "Maduro aumenta 300% el mínimo en Venezuela," https://www.dinero.com/internacional/articulo/maduro-aumenta-300-salario-minimo-en-venezuela/266002 2019. Accessed: 6.30.2019.

**Drew, Jacqueline M. and Michael E. Drew**, "The Identification of Ponzi Schemes: Can a Picture Tell a Thousand Frauds?," Discussion Papers in Finance 1, Griffith University, Department of Accounting, Finance and Economics August 2010.

**Dube, Oeindrila and Juan Vargas**, "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *Review of Economic Studies*, Forthcoming 2013.

**Duryea, Suzanne, David Lam, and Deborah Levison**, "Effects of economic shocks on children's employment and schooling in Brazil," *Journal of Development Economics*, September 2007, *84* (1), 188–214.

**Dustmann, Christian, Tommaso Frattini, and Ian P. Preston**, "The Effect of Immigration along the Distribution of Wages," *The Review of Economic Studies*, 04 2012, *80* (1), 145–173.

**_ , Uta Schönberg, and Jan Stuhler**, "The impact of immigration: Why do studies reach such different results?," *Journal of Economic Perspectives*, 2016, *30* (4), 31–56.

**El Comercio**, "Salario mínimo en Venezuela: Qué se puede comprar con 18 mil bolívares (20,9 dólares)?," https://elcomercio.pe/mundo/venezuela/salario-minimo-venezuela-comprar-18000-bolivares-pais-nicolas-maduro-nnda-nnlt-notic 2019. Accessed: 6.30.2019.

**Elacqua, Gregory, Diana Hincapie, Emiliana Vegas, Mariana Alfonso, Solange Berstein, Alejandro Puente, Rafael Porto de Almeida, Lucas Azevedo Fonseca, Dillon Clarke et al.**, "Profesión: Profesor en América Latina¿ Por qué se perdió el prestigio docente y cómo recuperarlo?(Resumen ejecutivo)," 2017.

**Escalona, Alejandro**, "Venezuela: cronologia de los ultimos 20 anos," 7 2017.

**Estrada, Ricardo**, "Rules rather than discretion: Teacher hiring and rent extraction," 2015.

**Ferman, Bruno and Cristine Pinto**, "Revisiting the synthetic control estimator," *Working paper*, 2016.

**Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux**, "Unconditional quantile regressions," *Econometrica*, 2009, *77* (3), 953–973.

**Firpo, Sergio P, Nicole M Fortin, and Thomas Lemieux**, "Decomposing wage distributions using recentered influence function regressions," *Econometrics*, 2018, *6* (2), 28.

**Flavin, Marjorie A**, "The Adjustment of Consumption to Changing Expectations about Future Income," *Journal of Political Economy*, October 1981, *89* (5), 974–1009.

**Foged, Mette and Giovanni Peri**, "Immigrants' effect on native workers: New analysis on longitudinal data," *American Economic Journal: Applied Economics*, 2016, *8* (2), 1–34.

**Fonseca, Yuri R, Ricardo M Prates, Marcelo C Medeiros, and Gabriel FR Vasconcelos**, "ArCo: An R package to Estimate Artificial Counterfactuals.," *R Journal*, 2018, *10* (1), 91.

**Freitez, Anitza**, "La emigracion desde Venezuela durante la ultima decada.," *Revista temas de coyuntura*, 2011, (63).

**Friedberg, Rachel M**, "The impact of mass migration on the Israeli labor market," *The Quarterly Journal of Economics*, 2001, *116* (4), 1373–1408.

**Friedman, Milton**, *A Theory of the Consumption Function* number frie57-1. In 'NBER Books.', National Bureau of Economic Research, Inc, Jan-Jun 1957.

**Gallup**, "World Split on Whether Immigrants Take Jobs Away," `https://news.gallup.com/poll/186425/world-split-whether-immigrants-jobs-away.aspx` 2015.

**Garber, Peter**, "Famous first bubbles," *Journal of Economic Perspectives*, 1990, *4*, 35/54.

**Garcia, Sandra et al.**, "Cómo Mejorar la Calidad de la Educación para todos los colombianos," *Fundación Compartir*, 2014.

**Goldhaber, Dan D. and Dominic J. Brewer**, "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement," *Educational Evaluation and Policy Analysis*, 2000, *22* (2), 129–145.

**Gong, Jie, Yi Lu, and Hong Song**, "The effect of teacher gender on students' academic and noncognitive outcomes," *Journal of Labor Economics*, 2018, *36* (3), 743–778.

**Gordon, G**, "Solving the Refugee Employment Problem in Jordan: A Survey of Syrian Refugees," *International Rescue Committee report*, 2017.

**Government of Colombia**, "National Constitution," 1991.

**Government of Jordan**, "Funding Status for Jordan Response Plan. The Jordan Response Plan for the Syria Crisis," 2015.

__ , "The Jordanian Response Plan for the Syria Crisis 2017-2019," 2017.

**Gregoriou, Greg and Francois L'habitant**, "Madoff: A flock of red flags," *Journal of Wealth Management*, 2009, *12* (1), 89–97.

**Guarcello, Lorenzo, Fabrizia Mealli, and Furio Rosati**, "Household vulnerability and child labor: the effect of shocks, credit rationing, and insurance," *Journal of Population Economics*, January 2010, *23* (1), 169–198.

**Guasch, J. Luis and Andrew Weiss**, "Self-Selection in the Labor Market," *The American Economic Review*, 1981, *71* (3), 275–284.

**Hagen-Zanker, J, M Ulrichs, R Holmes, and Z Nimeh**, "Cash transfers for refugees: the economic and social effects of a programme in Jordan," *London: Overseas Development Institute and Aya Consultancy*, 2017.

**Haider, Huma**, "Syrian refugee onward migration from Jordan to Europe," 2016.

**Hanmer, Lucia, Eliana Rubiano, Julieth Santamaria, and Diana J Arango**, "How does poverty differ among refugees? Taking a gender lens to the data on Syrian refugees in Jordan," *Middle East Development Journal*, 2020, *12* (2), 208–242.

**Hanushek, Eric A and Richard R Pace**, "Who chooses to teach (and why)?," *Economics of education review*, 1995, *14* (2), 101–117.

**Hashemite Kingdom of Jordan**, "Jordan Response Plan for the Syria Crisis 2020-22," 2020.

**Heilman, Brian and Gary Barker**, *Masculine norms and violence: Making the connections*, Promundo-US, 2018.

**Hernandez, Igor, Francisco Monaldi et al.**, "Weathering Collapse: An Assessment of the Financial and Operational Situation of the Venezuelan Oil Industry," Technical Report, Center for International Development at Harvard University 2016.

**Hidalgo, Daniel, Suresh Naidu, Simeon Nichter, and Neal Richardson**, "Occupational Choices: Economic Determinents of Land Invasions," *Review of Economics and Statistics*, 2010.

**Howard, Greg**, "The migration accelerator: Labor mobility, housing, and demand," 2019.

**Hwang, NaYoung and Brian Fitzpatrick**, "Student–Teacher Gender Matching and Academic Achievement," *AERA Open*, 2021, *7*, 23328584211040058.

**Ibañez, Ana Maria**, "Forced displacement in Colombia: Magnitude and causes," *The Economics of Peace and Security Journal*, 2009, *4* (1).

**ILO**, "ILO Response to Syrian Refugee Crisis in Jordan," 2019.

__ , "Impact of work permits on decent work for Syrians in Jordan," 2021.

**ILO and FAFO**, "Impact of COVID-19 on Syrian Refugees and Host Communities in Jordan," 2022.

**IOM**, "Determinants of Internet Use Among Migrants in South-East Asia: A case study of internal migrants in Thailand and Viet Nam," *Migration Research Series*, 2019, (58).

__ , "World Migration Report 2020," 2020.

**IRC and Airbel Impact Lab**, "New survey reveals extent of hardship and abuse experienced by Syrian children working on streets of Lebanon," 2016.

**Jacoby, Hanan G and Emmanuel Skoufias**, "Risk, Financial Markets, and Human Capital in a Developing Country," *Review of Economic Studies*, July 1997, *64* (3), 311–35.

**Jarvis, Christopher**, "The Rise and Fall of Albania's Pyramid Schemes," *Finance and Development*, 2000, *37* (1).

**Jensen, Robert**, "Agricultural Volatility and Investments in Children," *The American Economic Review*, 2000, *90* (2), 399–404.

**Kelberer, Victoria**, "The work permit initiative for Syrian refugees in Jordan," *Policy*, 2017.

**Krafft, Caroline, Assaad Ragui, and Isabel Pastoor**, "How do gender norms shape education and domestic work outcomes? The case of Syrian refugee adolescents in Jordan," 2021.

**__ , Maia Sieverding, Colette Salemi, and Caitlyn Keo**, "Syrian refugees in Jordan: Demographics, livelihoods, education, and health," in "Economic Research Forum Working Paper Series," Vol. 1184 2018.

**Kuncel, Nathan R. and Sarah A. Hezlett**, "Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions," *Current Directions in Psychological Science*, 2010, *19* (6), 339–345.

**Lanjouw, Jean O, Peter Lanjouw, Branko Milanovic, and Stefano Paternostro**, "Relative price shifts, economies of scale and poverty during economic transition 1," *Economics of Transition*, 2004, *12* (3), 509–536.

**Lankford, Hamilton, Susanna Loeb, and James Wyckoff**, "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis," *Educational Evaluation and Policy Analysis*, 2002, *24* (1), 37–62.

**Larson, Donald F. and Frank Plessmann**, "Do farmers choose to be inefficient? Evidence from Bicol," *Journal of Development Economics*, September 2009, *90* (1), 24–32.

**Lebow, Jeremy, Jonathan Moreno Medina, and Horacio Coral**, "Immigration and Trust: The Case of Venezuelans in Colombia," *Available at SSRN 3660641*, 2020.

**Lopez, Virginia and Jonathan Watts**, "Nicolas Maduro narrowly wins Venezuelan presidential election," 04 2013.

**MacKay, Charles**, *Extraordinary Popular Delusions and the Madness of Crowds*, Farrar, Straus and Giroux, 1841.

**Martin, S and Lisa Singh**, "Data analytics and displacement: Using big data to forecast mass movements of people," *Digital lifeline? ICTs for refugees and displaced persons*, 2018, pp. 185–206.

**Meyer, Bruce D and James X Sullivan**, "Measuring the well-being of the poor using income and consumption," 2003.

— **and** — , "Further results on measuring the well-being of the poor using income and consumption," *Canadian Journal of Economics/Revue canadienne d'économique*, 2011, *44* (1), 52–87.

— **and** — , "Identifying the disadvantaged: Official poverty, consumption poverty, and the new supplemental poverty measure," *Journal of Economic Perspectives*, 2012, *26* (3), 111–36.

**Migración Colombia**, "Todo lo que quiere saber sobre la migración venezolana y no se lo han contado," 2019.

— , "Más de 1 millón 825 mil venezolanos estarían radicados en Colombia," 2020.

— , "Número de venezolanos radicados en Colombia desciende por primera vez en 5 años," 2020.

**Miguel, Edward, Shanker Satyanath, and Ernest Sergenti**, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, *112* (4), pp. 725–753.

**Ministerio de Relaciones Exteriores**, "Resolución 0872," 2019.

**Ministry of Education**, "Decree 2277 of 1979," 1979.

— , "Decree 1278 of 2002," 2002.

**Ministry of Labor**, "Work Permit progress report June," 2018.

**Modigliani, Franco**, *Post-Keynesian Economics*, Rutgers University Press, 1954.

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

**National Civil Service Commission**, "Orientaciones para la presentación de las pruebas de aptitudes, competencias básicas y psicotécnica para la selección de docentes," 2010.

**Norwegian Refugee Council**, "Thousands of refugees return to Syria from Jordan," 2015.

_ , "Drivers of despair: Refugee protection failures in Jordan and Lebanon," 2016.

**OECD**, "Education in Colombia," 2016.

**OECD**, "Support to the Middle East and North Africa (MENA)," 2016.

**Ome, Alejandro**, "Meritocracy for teachers. Evidence from Colombia," *Working paper*, 2012.

**Organization, World Health**, "Mental health and psychosocial considerations during the COVID-19 outbreak, 18 March 2020," Technical Report, World Health Organization 2020.

**Palotti, Joao, Natalia Adler, Alfredo Morales-Guzman, Jeffrey Villaveces, Vedran Sekara, Manuel Garcia Herranz, Musa Al-Asad, and Ingmar Weber**, "Monitoring of the Venezuelan exodus through Facebook's advertising platform," *Plos one*, 2020, *15* (2), e0229175.

**Peri, Giovanni and Chad Sparber**, "Task specialization, immigration, and wages," *American Economic Journal: Applied Economics*, 2009, *1* (3), 135–69.

_ **and Vasil Yasenov**, "The labor market effects of a refugee wave synthetic control method meets the mariel boatlift," *Journal of Human Resources*, 2019, *54* (2), 267–309.

**Puech, Frederic and Patrick Guillaumont**, "Macro-Economic Instability and Crime," Working Papers, CERDI 2006.

**Romero, Simon**, "Crisis at Colombia Border Spills Into Diplomatic Realm," *New York Times*, 2008, *4.*

**Rose, Elaina**, "Consumption Smoothing and Excess Female Mortality in Rural India," *The Review of Economics and statistics*, 1999, *81* (1), 41–49.

**Rosenbaum, Paul and Donald Rubin**, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 1983, *70* (1), 41–55.

**Rosenzweig, Mark R. and Kenneth I. Wolpin**, "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India," *Journal of Political Economy*, 1993, *101* (2), 223–244.

**Rubin, Donald**, "Bias reduction using mahalanobis-metric matching," *Biometrics*, 1980, *36*, 293–298.

**Rudner, Lawrence**, "Pre-employment testing and employee productivity," *Public Personnel Management*, 1992, *21* (2), 133.

**Ruiz, Isabel and Carlos Vargas-Silva**, "The economics of forced migration," *The Journal of Development Studies*, 2013, *49* (6), 772–784.

**Sarsons, Heather**, "Rainfall and conflict," *Yale Economics Department Working Paper*, 2011.

**Schmidt, Frank L. and John E. Hunter**, "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings," *Psychological Bulletin*, 1998, *124* (2), 262–274.

**Slimane, Meriem Ait Ali, Jonna Lundvall, Komal Mohindra, Shereen Al Abbadi, Nato Kurshitashvili, and Ola Hisou**, "Women's Economic Empowerment in Jordan," 2020.

**Stave, Svein Erik, Tewodros Aragie Kebede, Maha Kattaa et al.**, "Impact of work permits on decent work for Syrians in Jordan," 2021.

**Strauss, Robert P. and Elizabeth A. Sawyer**, "Some new evidence on teacher and student competencies," *Economics of Education Review*, 1986, *5* (1), 41 – 48.

**Thomas, Duncan, Kathleen Beegle, Elizabeth Frankenberg, Bondan Sikoki, John Strauss, and Graciela Teruel**, "Education in a crisis," *Journal of Development Economics*, June 2004, *74* (1), 53–85.

**Tiltnes, Åge A, Huafeng Zhang, and Jon Pedersen**, "The living conditions of Syrian refugees in Jordan," *FAFO Report. Amman: Fafo Research Foundation*, 2019.

**Townsend, Robert M.**, "Risk and Insurance in Village India," *Econometrica*, 1994, *62* (3), 539–591.

**Tribín-Uribe, Ana María, Adhvaryu Achyuta, Cesar Anzola, Oscar Ávila-Montealegre, Leonardo Bonilla-Mejía, Juan Carlos Castro-Fernández, Luz Adriana Flórez, Ánderson Grajales-Olarte, Alexander Guarín-López, Franz Hamann-Salcedo et al.**, "Migración desde Venezuela en Colombia:

caracterización del fenómeno y análisis de los efectos macroeconómicos," *Revista Ensayos Sobre Política Económica; No. 97, octubre 2020. Pág.: 1-74*, 2020.

__ **and Brian Knight**, "Immigration and Violent Crime: Evidence from the Colombia-Venezuela Border," *Borradores de Economía; No. 1121*, 2020.

**Tumen, Semih**, "The use of natural experiments in migration research," *IZA World of Labor*, 2015.

**Turner, Lewis**, "Three years of the Jordan compact: The (gendered) challenges of providing work permits for Syrian refugees," 2019.

**UN Women**, "Unpacking Gendered Realities in Displacement: The status of Syrian refugee women in Jordan," 2018.

**UNHCR**, "Living in the shadows: Jordan home visits report 2014," *United Nations High Commissioner for Refugees, Geneva. http://www. unhcr. org/54b685079. html*, 2014.

__ , "The sea route to Europe: The Mediterranean passage in the age of refugees," 2015.

__ , "As Venezuelans flee throughout Latin America, UNHCR issues new protection guidance," 2018.

__ , "Jordan factsheet, January – December 2018," 2018.

__ , "Jordan Multipurpose Cash Assistance," 2018.

__ , "Desperate journeys. Refugees and migrants arriving in Europe and at Europe's borders (January–December 2018)," 2019.

__ , "Figures at Glance," https://www.unhcr.org/figures-at-a-glance.html 2019.

__ , "UNHCR Implementation of the Policy on Cash-based Inverventions: 2016-2019," 2019.

__ , "Jordan issues record number of work permits to Syrian refugees," 2022.

__ , "Total Persons of Concern. UNHCR Operations Data Portal," 2022.

__ , "Total Registered Syrian Refugees. UNHCR Operations Data Portal," 2022.

**UNICEF**, "Jordan country report on out-of-school children," *MENA Regional Office, Jordan: UNICEF*, 2014.

__ , "Summary Overview Document: Leaders' Summit on Refugees," 2016.

**Universidad de Antioquia**, "La carrera de los venezolanos para convalidar sus títulos," *Academia Sociedad*, 2020.

**Universidad del Rosario**, "Guía de derechos y deberes para venezolanos en Colombia," *Consultorio Jurídico y Centro de Conciliación*, 2020.

**Venturini, Alessandra and Claudia Villosio**, "Labour market effects of immigration into Italy: An empirical analysis 1," *International Labour Review*, 2006, *145* (1-2), 91–118.

**Verme, Paolo, Chiara Gigliarano, Christina Wieser, Kerren Hedlund, Marc Petzoldt, and Marco Santacroce**, *The welfare of Syrian refugees: evidence from Jordan and Lebanon*, World Bank Publications, 2015.

**Weldali, Maria**, "Study highlights correlation between childcare services and women's workforce participation," 2022.

**Winkelmann, Timothy J**, *The Impact of Teacher Gender on Elementary Students Academic Performance*, Baker University, 2016.

**Winkler, Hernan and Alvaro Gonzalez**, "Jordan jobs diagnostic," 2019.

**World Bank**, "Migración desde Venezuela a Colombia: Impactos y estrategia de respuesta en el corto y mediano plazo," 2018.

—, "Improving Women Economic Opportunities. Select Entry Points for Policy Dialogue and Operational Interventions," 2021.

—, "Jordan Support to Private Sector Employment and Skills," 2021.

—, "Performance and Learning Review of the Country Partnership Framework for Hashemite Kingdom of Jordan for the Period FY17-FY22," 2021.

**Zeldes, Stephen P**, "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence," *The Quarterly Journal of Economics*, 1989, *104* (2), 275–98.

**Zuckoff, Mitchell**, *Ponzi's Scheme: The True Story of a Financial Legend*, Random House Digital, Inc, 2005.

—, "What Madoff could learn from Ponzi," *CNNMoney.com*, January 13, 2009, *retrieved April 15, 2009.*

# Appendix A

# Additional Tables and Figures for Chapter 1

## A1 Definitions

*Experience*: It is potential experience. That is the age reported by the individual minus the years of schooling.

*Hourly real wages*: They are defined as the labor earnings received last month, adjusted by a monthly price index, and divided by the number of hours that an individual usually works per month. The latter is defined as four (weeks) times the number of hours that a person usually works per week. Wages in the formal sector were also adjusted by the contribution employers make to health insurance and retirement plants of their employees. In Colombia, employers contribute 8.5% of their wages to their employees' health insurance plan and 12% to their retirement plans. I also dropped observations below the 1% and above the 99% of the wage distribution.

*Labor force*: It is defined as the economically active. That is, those people who are 12+ years old and, who are either working or looking for employment[1].

*Informality*: This paper uses two definitions of informality. The fist measure defines an informal workers as the individuals who do not have access to a health insurance through their employer or does not have a

[1] Although the official definition from the National Bureau of Statistics in Colombia takes into account people who are 10+ years old in rural areas, and those who are 12+ years old in urban areas, I dropped those people aged 10-11 years old from the sample.

retirement plan. The main results of this paper use this measure. an alternative measure, used by DANE, defines informality according to the size of the firm. If the worker is employed by a firm of 5 or less employees, he is classified as informal. Self-employed also enter in this category. Only if specified in the footnote, usually on appendix tables, this measure will be used.

# A2 Google Searches Index

Google Trends queries return a Google search index for the location and time range specified. In this paper, I use a query for key terms in Colombia for the past 4 years (2015-2018). The result of this query is a dataset with a unique index per city. The google search index for citi $i$ using this type of query would be:

$$SI_i = \frac{hits_i}{\max_{j \in N}\{hits_j\}} \tag{A.1}$$

, where $SI_i$ is the search intensity measure that one gets from the query, $hits_i$ is the number of hits of searches of key words in geographical area i during the period 2015-2018. The denominator of the equation returns the number of hits of the city where most of the hits took place. Thus, the Google search intensity index returns, per city, a number that ranges from 0 to 100. The city with most of the searches will be assigned a value of 100 and every other city in the country will be assigned a value relative to that value. Cities with very low search are automatically assigned a value of zero. Google Search also removes multiple searches made by the same person within a specific time period. [2]

Let's calculate a composite index that is the summation of the city indexes

$$\sum_{i=1}^{N} SI_i = \frac{1}{\max_{j \in N}\{hits_j\}} \sum_{i=1}^{N} hits_i \tag{A.2}$$

Thus, one can calculate the share of hits in location i in the following manner:

$$\begin{aligned} m_i^2 &= \frac{SI_i}{\sum_{i=1}^{N} SI_i} && \text{(A.3)} \\ &= \frac{\frac{hits_i}{\max_{j \in N}\{hits_j\}}}{\frac{1}{\max_{j \in N}\{hits_j\}} \sum_{i=1}^{N} hits_i} && \text{(A.4)} \\ &= \frac{hits_i}{\sum_{i=1}^{N} hits_i} && \text{(A.5)} \end{aligned}$$

$m_i^2$ is a proxy for the number of migrants in location i as a share of the total number of migrants.

[2] See https://support.google.com/trends/answer/4365533?hl=en

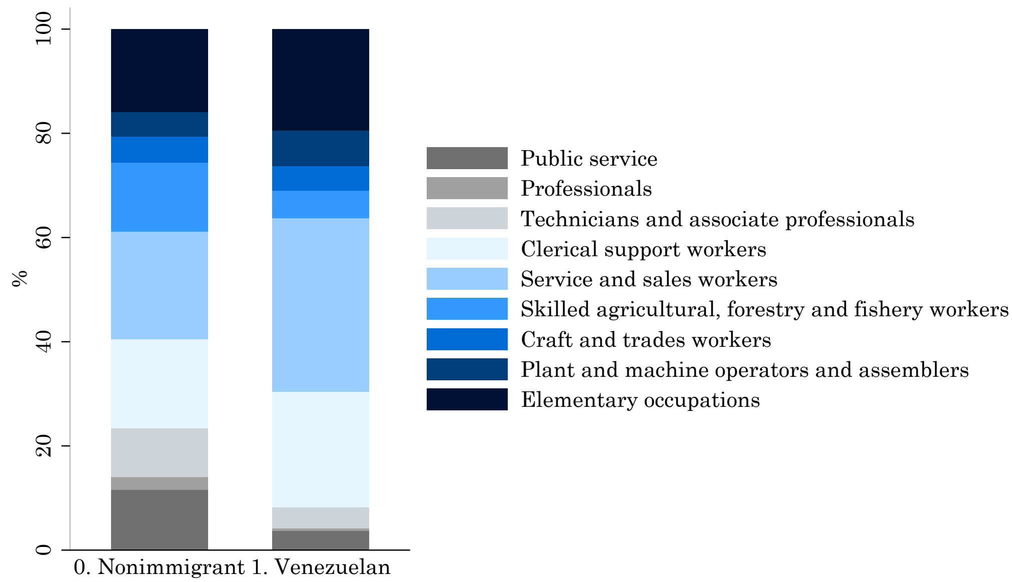
## A3 Figures and Tables

Table A1: Timing of the events

| | | |
|---|---|---|
| August 19, 2015 | • | Maduro orders to close the border crossing in Táchira |
| **First wave** | • | |
| August 13, 2016 | • | Venezuela-Colombia border is reopened after almost one year |
| **Second wave** | • | |
| July 16, 2017 | • | Venezuelan Referendum takes place |
| October 15,2017 | • | Maduro's political party wins 17 out of 22 Governorships |
| **Third wave** | • | |
| May 20, 2018 | • | Presidential elections take place. Maduro is re-elected. |

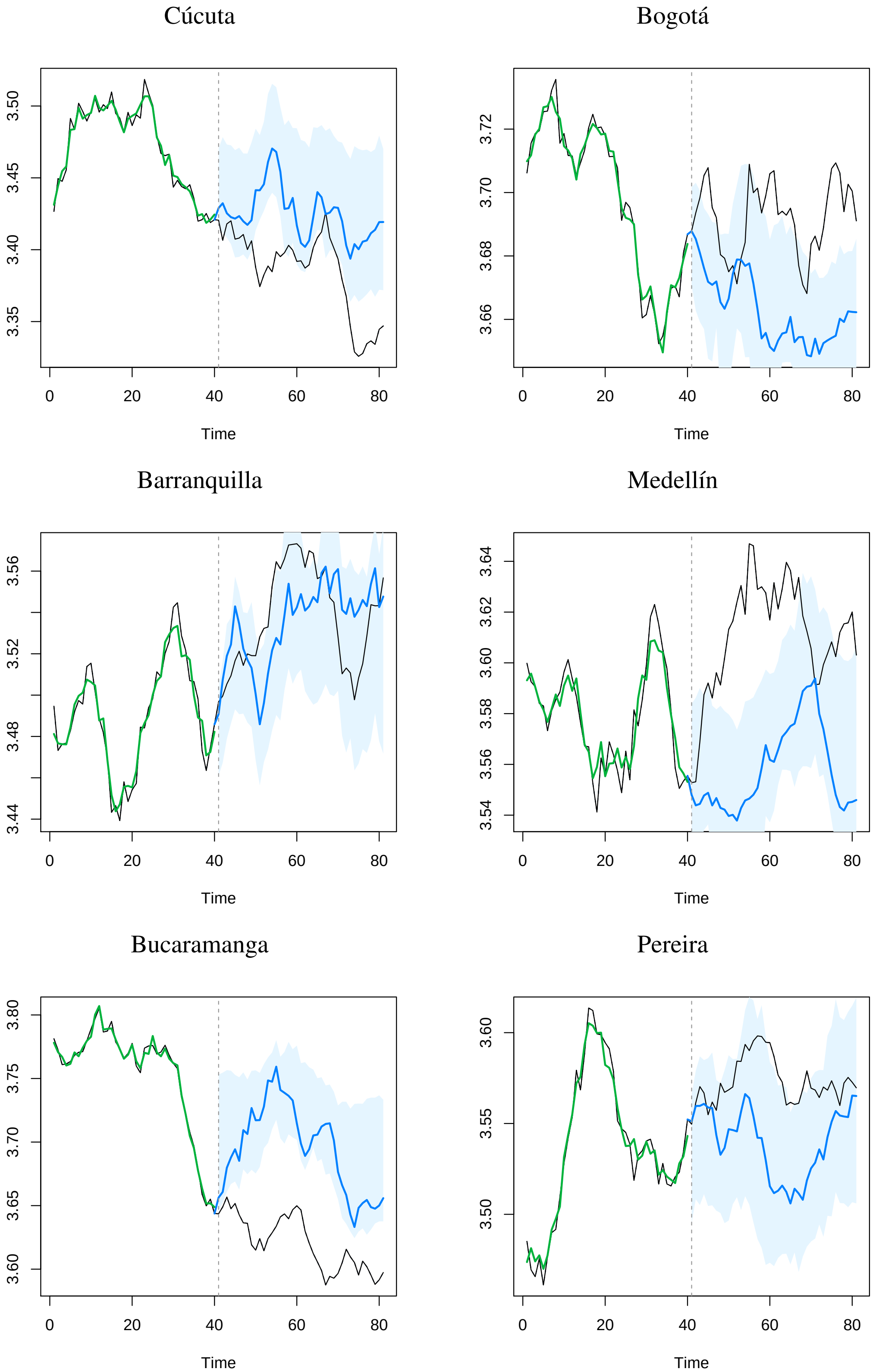
*Notes:* This table summarizes three key events that unfolded the Venezuelan exodus. Accuracy of the data was cross-checked using multiple journals cited on the references section.

Figure A1: Occupation of Nationals and Immigrants in Colombia



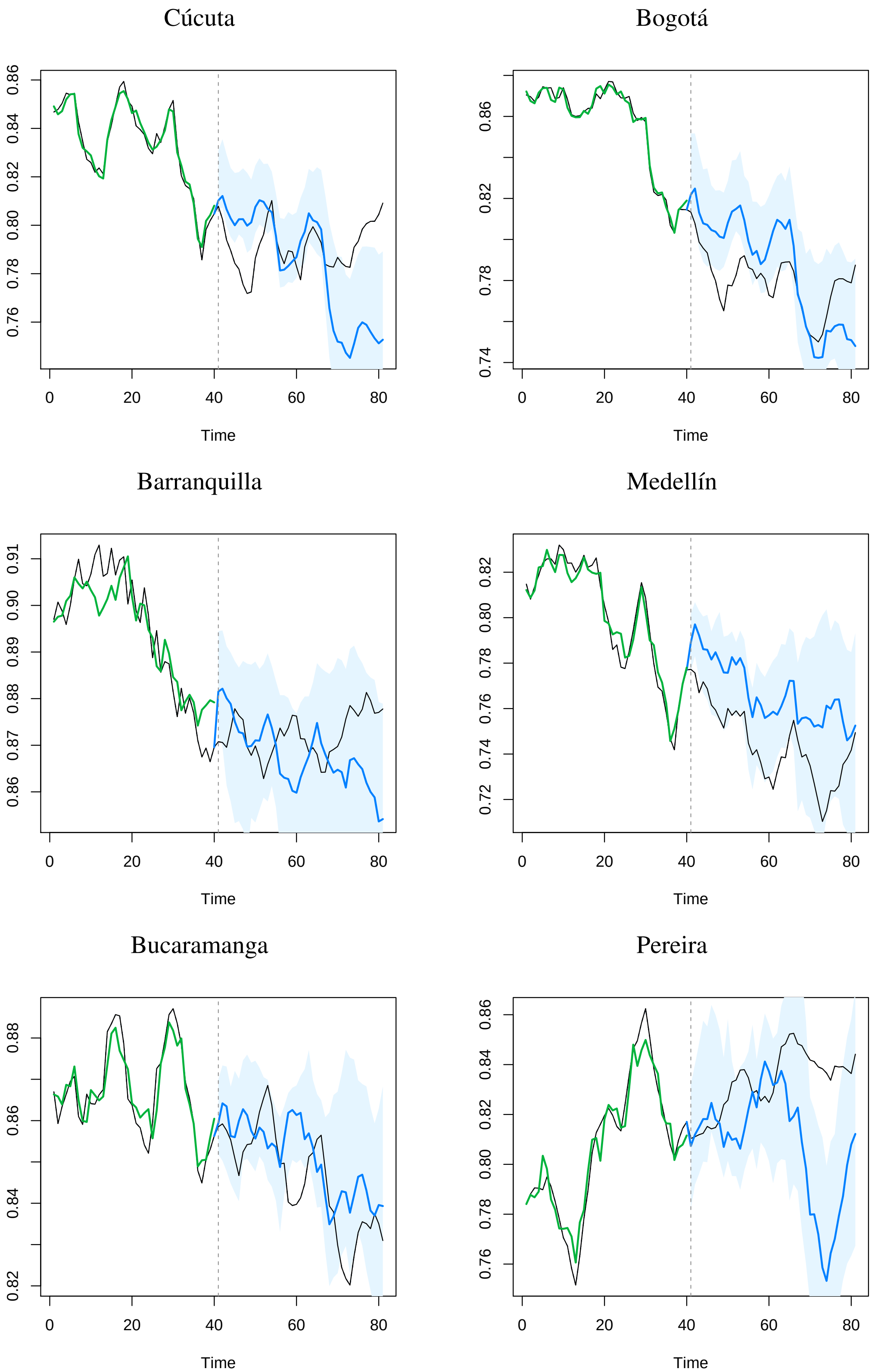

*Notes:* This graph displays the distribution of nonimmigrants, Venezuelan immigrants and Colombian returnees in the GEIH sample. The distribution is computed using all individuals observed after July 2016, i.e., after the re-opening of borders.

Figure A2: ArCo Results for Informal Wages




*Notes:* This set of graphs depict the fitting of the ArCo methodology in 6 regions that had a relatively higher influx of immigrants. Al the graphs depict results in informal wages. The black line presents the observed data. The green line shows how the elastic net fits the observed data in the pre-treatment period. The light blue line depicts the prediction of the model in the post-treatment period along with a 95% confidence interval. The vertical dashed line signals the moment in time when borders re-opened.

Figure A3: ArCo Results for Informal Employment




*Notes:* This set of graphs depict the fitting of the ArCo methodology in 6 regions that had a relatively higher influx of immigrants. Al the graphs depict results in informal employment. The black line presents the observed data. The green line shows how the elastic net fits the observed data in the pre-treatment period. The light blue line depicts the prediction of the model in the post-treatment period along with a 95% confidence interval. The vertical dashed line signals the moment in time when borders re-opened.

Table A2: Heterogeneous Effects on Wages

| | (1) Formal | (2) Informal | (3) Both |
|---|---|---|---|
| *Panel A. By gender* | | | |
| $M_{Ven,rt} \times Female$ | 0.002* | 0.005*** | 0.007*** |
| | (0.001) | (0.001) | (0.001) |
| Observations | 276,332 | 689,062 | 965,394 |
| R-squared | 0.058 | 0.097 | 0.151 |
| *Panel B. By age category* | | | |
| $M_{Ven,rt} \times 15-24$ | 0.004** | -0.007*** | 0.001 |
| | (0.002) | (0.002) | (0.002) |
| $M_{Ven,rt} \times 25-34$ | -0.004** | -0.003* | 0.003* |
| | (0.002) | (0.002) | (0.002) |
| $M_{Ven,rt} \times 35-44$ | -0.006*** | -0.006*** | -0.001 |
| | (0.002) | (0.002) | (0.001) |
| $M_{Ven,rt} \times 45-54$ | -0.006*** | -0.002 | 0.000 |
| | (0.002) | (0.001) | (0.001) |
| Observations | 276,332 | 689,062 | 965,394 |
| R-squared | 0.058 | 0.097 | 0.151 |
| *Panel C. By occupation* | | | |
| Clerical support workers | 0.005 | -0.016* | -0.016* |
| | (0.010) | (0.010) | (0.009) |
| Observations | 33,676 | 141,766 | 169,272 |
| Service workers | -0.004 | -0.025*** | -0.017*** |
| | (0.004) | (0.004) | (0.004) |
| Observations | 76,659 | 168,195 | 236,751 |
| Elementary occupations | -0.013 | -0.053*** | -0.045*** |
| | (0.009) | (0.010) | (0.010) |
| Observations | 67,019 | 140,107 | 200,540 |

*Notes:* This table summarizes the findings of heterogeneous effects of the migration-induced supply shock on wages. These regressions use the sample of Colombians who have not changed their city of residence in the past year and who have completed secondary or less. Panel A and Panel B show the results of a model of triple differences that tests heterogeneous effects by gender and age category, respectively. Each subpanel from Panel C comes from a different regression that uses the subsample of workers in the corresponding occupation. In this case, each regression uses a migration shock at the region-month-occupation level given that the household survey allows to identify which occupations immigrants enter. Column (1) presents the results in the formal sector, Column (2) in the informal sector, and Column (3) in both. Standard errors, in parentheses, are clustered at the region-month/year level. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Heterogeneous Effects on Employment

| | (1) Formal | (2) Informal | (3) Both |
|---|---|---|---|
| *Panel A. By gender* | | | |
| $M_{Ven,rt} \times Female$ | 0.009*** | 0.000 | 0.001 |
| | (0.001) | (0.001) | (0.000) |
| Observations | 404,347 | 927,731 | 1,223,911 |
| R-squared | 0.160 | 0.054 | 0.038 |
| *Panel B. By age category* | | | |
| $M_{Ven,rt} \times 15-24$ | 0.001 | -0.005*** | -0.004*** |
| | (0.002) | (0.001) | (0.001) |
| $M_{Ven,rt} \times 25-34$ | 0.001 | -0.004*** | -0.002*** |
| | (0.002) | (0.001) | (0.001) |
| $M_{Ven,rt} \times 35-44$ | 0.001 | -0.002*** | -0.001** |
| | (0.002) | (0.001) | (0.001) |
| $M_{Ven,rt} \times 45-54$ | 0.004*** | -0.000 | 0.000 |
| | (0.001) | (0.001) | (0.000) |
| Observations | 380,171 | 820,690 | 1,103,035 |
| R-squared | 0.159 | 0.055 | 0.039 |
| *Panel C. By occupation* | | | |
| Clerical support workers | 0.006 | 0.003 | 0.002 |
| | (0.008) | (0.003) | (0.003) |
| Observations | 59,215 | 196,722 | 234,447 |
| Service workers | 0.007* | 0.004 | 0.004** |
| | (0.004) | (0.003) | (0.002) |
| Observations | 115,555 | 214,779 | 295,385 |
| Elementary occupations | 0.007 | 0.006 | 0.004 |
| | (0.008) | (0.004) | (0.003) |
| Observations | 92,674 | 175,457 | 245,705 |

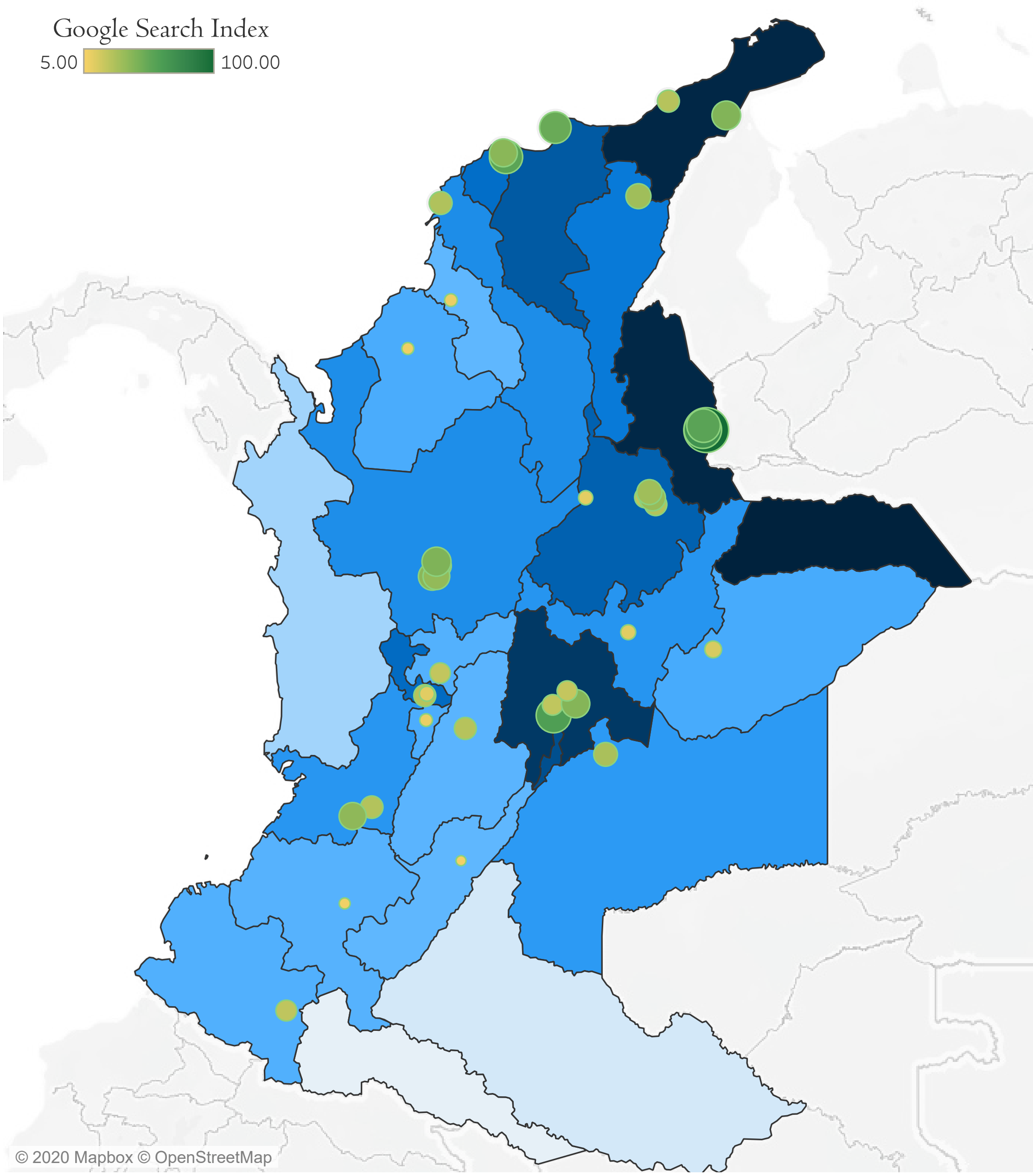
*Notes:* This table summarizes the findings of heterogeneous effects of the migration-induced supply shock on employment. These regressions use the sample of Colombians who have not changed their city of residence in the past year and who have completed secondary or less. Panel A and Panel B show the results of a model of triple differences that tests heterogeneous effects by gender and age category, respectively. Each subpanel from Panel C comes from a different regression that uses the subsample of workers in the corresponding occupation. In this case, each regression uses a migration shock at the region-month-occupation level given that the household survey allows to identify which occupations immigrants enter. Column (1) presents the results in the formal sector, Column (2) in the informal sector, and Column (3) in both. Standard errors, in parentheses, are clustered at the region-month/year level. *** p<0.01, ** p<0.05, * p<0.1

Table A4: Effect of Venezuelans Immigrants on Internal Migration

| | (1) |
|---|---|
| Variable | $M_{Int,mr}$ |
| $M_{Ven,mr}$ | -0.011 |
| | (0.016) |
| | |
| Observations | 1,418 |
| R-squared | 0.771 |

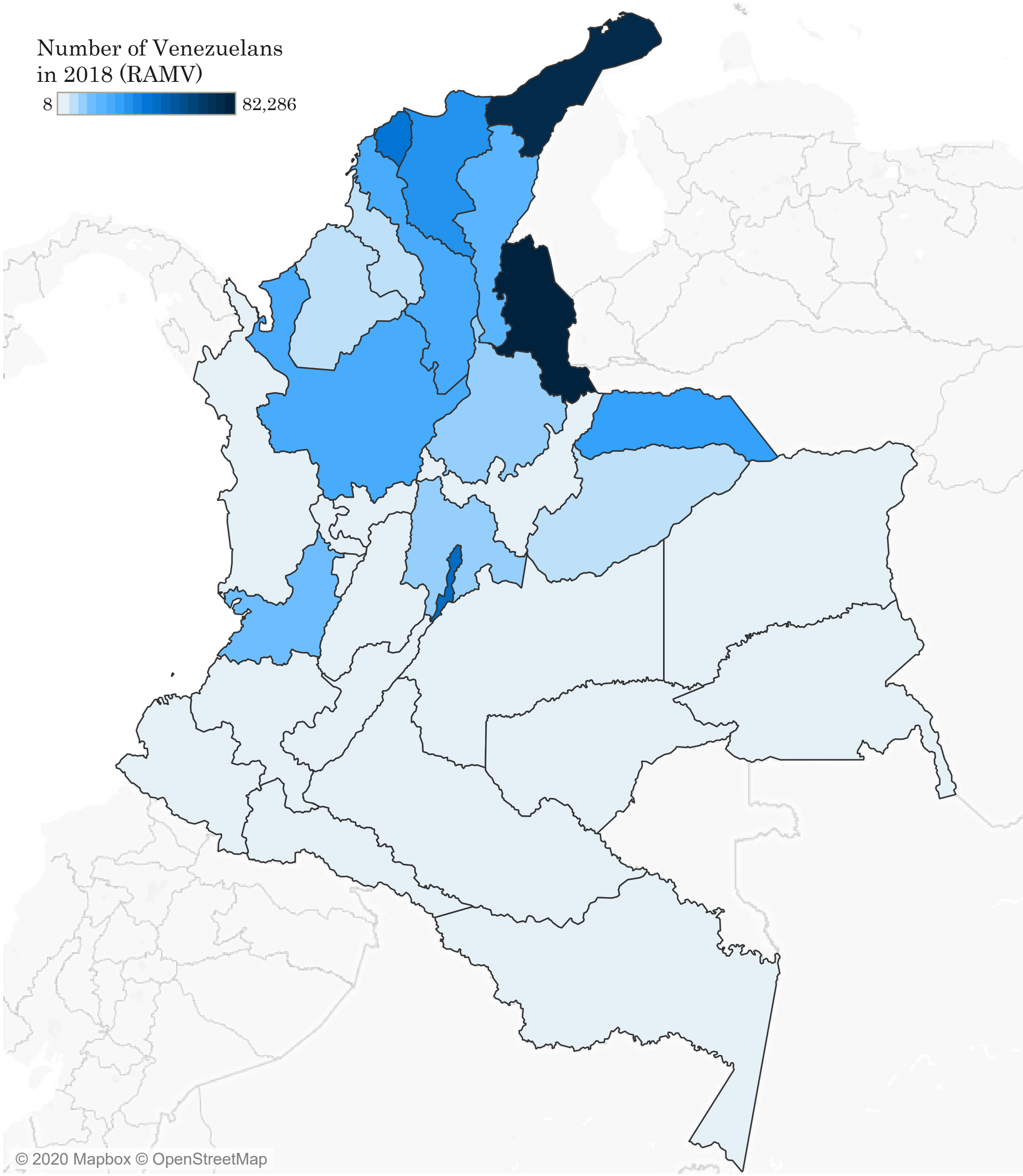
*Notes:* This table regresses net migration of Venezuelan immigrants on the net migration of internal migrants. The results are obtained from a dataset that contains time and regional variation in the outcomes of interest. Robust standard errors are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure A4: Geographical Distribution of the Raw Internet Search Index



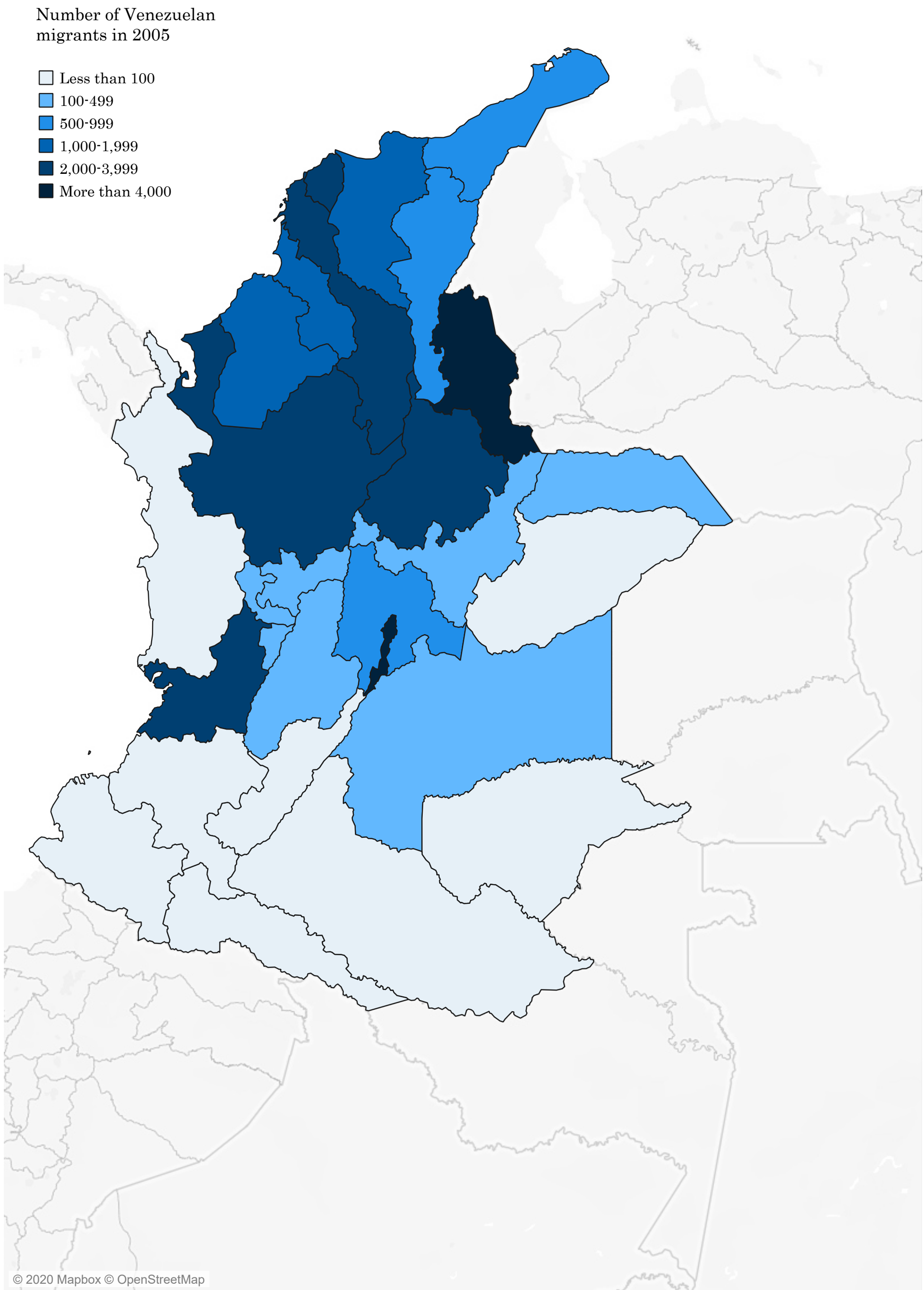

Note: This map depicts the geographical distribution of the number of immigrants according to the Internet search index. Two overlapping distributions are shown. In a scale of blues, the geographical density at the Department level, while in green the geographical density of selected cities. The maximum value of the raw index is 100.

Figure A5: Venezuelan Migrant Counts According to Official Records



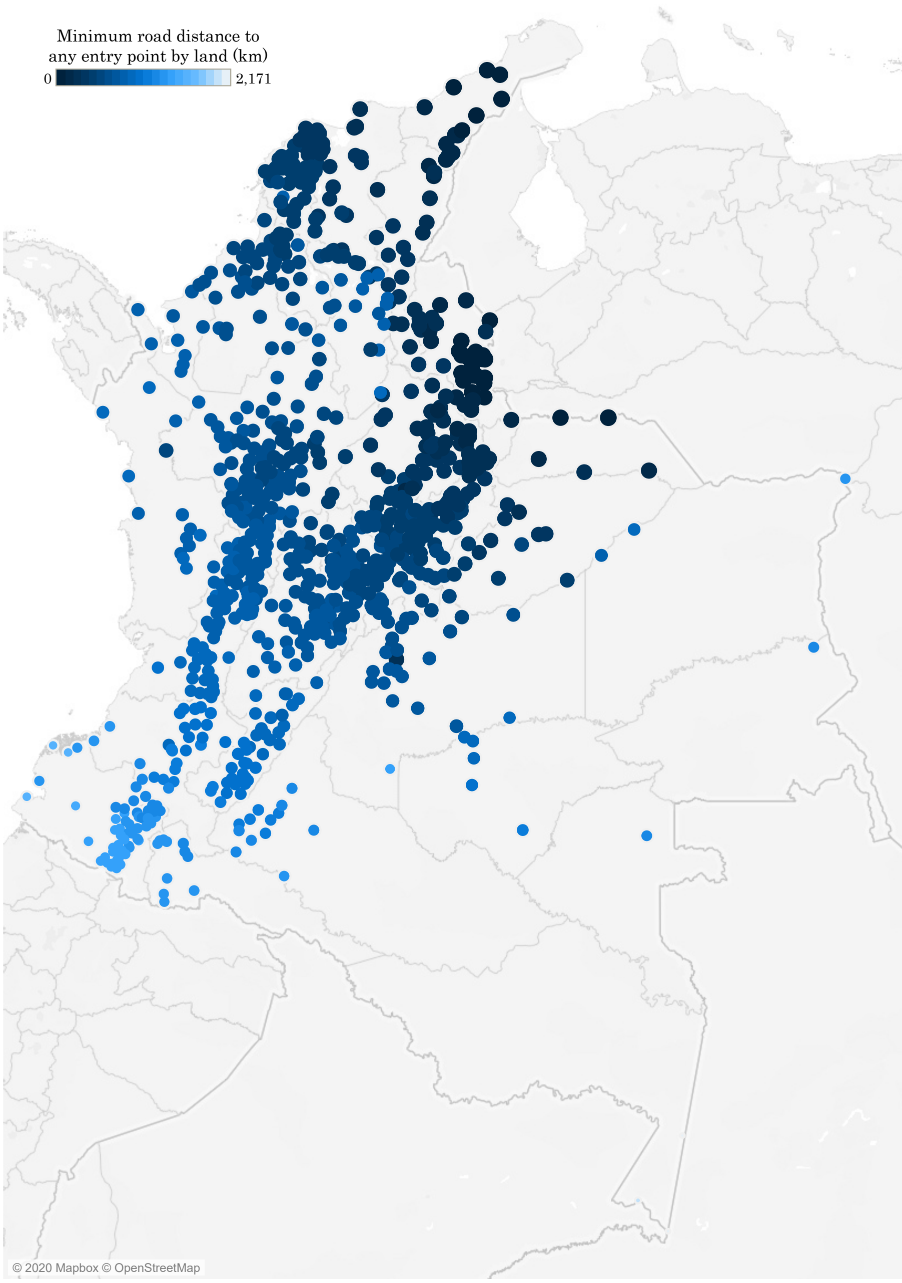

Note: This map depicts the geographical distribution of the number of immigrants by Department using Data from administratiive records (PEP-RAMV). Data was collected from extractions of Tableau dashboards produced by Migración Colombia.

Figure A6: Venezuelan Migrants in 2005




Note: Produced by the author using information from IPUMS international 2005. Number of immigrants in 2005 correspond to the weighted sum of the number of people born in Venezuela surveyed in the 10% Census sample.

Figure A7: Road Distance to the Border




Note: This map depicts the road distance between the closest entry point and the target city. Three official entry points are used: Cucuta, Maicao and Arauca. Notice that the road will not necessarily match linear distance between two points.

# Appendix B

# Additional Tables and Figures for Chapter 2

## B1 Additional Tables

Table B1: Kolmogorov-Smirnov equality-of-expenditure-distributions test

| Type of expenditure distribution | $D^+ =$ $\max_y M(y) - F(y)$ | $D^- =$ $= \min_y M(y) - F(y)$ | $D$ $\max(\mid D^+ \mid, \mid D^- \mid)$ |
|---|---|---|---|
| $\theta = 1$, 2013-14 | 0.040 | -0.023 | 0.040 |
| | $[0.000]$ | $[0.000]$ | $[0.000]$ |
| $\theta = 1$, 2017-18 | 0.007 | -0.030 | 0.030 |
| | $[0.369]$ | $[0.000]$ | $[0.000]$ |
| $\theta = 0.5$, 2013-14 | 0.034 | -0.065 | 0.065 |
| | $[0.000]$ | $[0.000]$ | $[0.000]$ |
| $\theta = 0.5$, 2017-18 | 0.000 | -0.156 | 0.156 |
| | $[0.999]$ | $[0.000]$ | $[0.000]$ |

Note: Own calculations based on ProGress and JD-HV database. The table shows the results of comparing male and female-headed household expenditure distributions. The rows correspond to the type of expenditure measured employed, which is a combination of the wave and the economies of scale used ($\theta$). The first column shows the largest difference between the male and female PA expenditure distributions when the expenditure measure of male PA is below that of female Pas. The second column shows the largest difference between the distributions when the expenditure measure of male PA households is above that of female PA households. The third column shows the largest absolute difference between the differences shown in columns 1 and 2. P-values are shown.

Table B2: Correlates of Expenditure Categories (Above vs Below the 40th Percentile) using OLS, by Wave and Gender of the Household Head

| | Wave 13-14 | | Wave 17-18 | |
|---|---|---|---|---|
| | Female HHH | Male HHH | Female HHH | Male HHH |
| VARIABLES | (1) | (2) | (3) | (4) |
| Characteristics of the PA | | | | |
| Age | -0.013*** | -0.006*** | -0.012*** | -0.010*** |
| | (0.002) | (0.001) | (0.002) | (0.002) |
| Age squared | 0.000*** | 0.000*** | 0.000*** | 0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Marital status (base: married) | | | | |
| Single or engaged | 0.122*** | 0.040*** | -0.008 | 0.041*** |
| | (0.017) | (0.010) | (0.016) | (0.014) |
| Divorced or separated | 0.061*** | 0.012 | -0.022* | 0.027 |
| | (0.017) | (0.024) | (0.013) | (0.027) |
| Widowed/Widower | 0.066*** | 0.017 | 0.007 | 0.077* |
| | (0.011) | (0.024) | (0.010) | (0.040) |
| Education (base: less than 6 years) | | | | |
| 6-11 years | -0.094*** | -0.078*** | -0.029*** | -0.066*** |
| | (0.010) | (0.006) | (0.009) | (0.007) |
| More than 12 years | -0.125*** | -0.135*** | -0.058*** | -0.126*** |
| | (0.013) | (0.008) | (0.012) | (0.008) |
| Characteristics of the household | | | | |
| Family Size | 0.111*** | 0.125*** | 0.113*** | 0.127*** |
| | (0.003) | (0.002) | (0.002) | (0.002) |
| Able bodied male adults, % | -0.118*** | -0.119*** | -0.091*** | -0.063*** |
| | (0.031) | (0.016) | (0.023) | (0.014) |
| Number of children below 5 | 0.028*** | 0.038*** | 0.048*** | 0.028*** |
| | (0.006) | (0.003) | (0.006) | (0.004) |
| Number of elderly in the household | 0.061*** | 0.016 | 0.053*** | -0.006 |
| | (0.021) | (0.013) | (0.015) | (0.013) |
| Family type (base: couples with children) | | | | |
| Couples without children | 0.106** | 0.075*** | 0.181*** | 0.101*** |
| | (0.048) | (0.010) | (0.032) | (0.011) |
| Single caregivers | -0.004 | -0.010 | 0.081*** | 0.050*** |
| | (0.016) | (0.012) | (0.013) | (0.012) |
| Single person households | 0.045** | 0.154*** | 0.205*** | 0.209*** |
| | (0.022) | (0.013) | (0.018) | (0.014) |
| Unaccompanied children | 0.054 | 0.146*** | 0.201* | 0.307*** |
| | (0.047) | (0.030) | (0.110) | (0.083) |
| Non-nuclear and other households | -0.035* | 0.034** | 0.006 | 0.008 |
| with children | (0.018) | (0.016) | (0.014) | (0.017) |
| Non-nuclear and other households | -0.033 | 0.010 | 0.036** | 0.010 |
| without children | (0.022) | (0.013) | (0.017) | (0.014) |
| Constant | 0.330*** | 0.094*** | 0.219*** | 0.043 |
| | (0.044) | (0.028) | (0.041) | (0.032) |
| Observations | 12,271 | 32,310 | 15,384 | 25,141 |
| R-squared | 0.201 | 0.324 | 0.199 | 0.295 |

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is a dummy that takes the value of 1 for households in the bottom 40th of the expenditure per capita distribution and 0 for those households above the 40th percentile household.

Table B3: Results from Quantile Regressions - Quantiles 10 to 50

| VARIABLES | Quantile 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| Female PA x Wave 17-18 | -0.120*** | -0.118*** | -0.138*** | -0.171*** | -0.195*** |
| | (0.015) | (0.008) | (0.008) | (0.007) | (0.007) |
| Wave 17-18 (base: Wave 13-14) | 0.044*** | 0.001 | -0.011** | -0.012*** | -0.013*** |
| | (0.007) | (0.007) | (0.005) | (0.003) | (0.004) |
| Characteristics of the PA | | | | | |
| Female PA | -0.115*** | -0.045*** | -0.014 | 0.018** | 0.041*** |
| | (0.012) | (0.012) | (0.009) | (0.008) | (0.008) |
| Age | 0.001** | 0.002*** | 0.002*** | 0.002*** | 0.002*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Marital status (base: married) | | | | | |
| Single or engaged | -0.232*** | -0.192*** | -0.160*** | -0.163*** | -0.161*** |
| | (0.020) | (0.012) | (0.013) | (0.018) | (0.019) |
| Divorced or separated | -0.088** | -0.069*** | -0.047*** | -0.045*** | -0.039*** |
| | (0.041) | (0.018) | (0.017) | (0.013) | (0.011) |
| Widowed/Widower | -0.100*** | -0.126*** | -0.110*** | -0.103*** | -0.100*** |
| | (0.023) | (0.015) | (0.009) | (0.007) | (0.011) |
| Education (base: less than 6 years) | | | | | |
| 6-11 years | 0.241*** | 0.186*** | 0.151*** | 0.127*** | 0.115*** |
| | (0.014) | (0.010) | (0.007) | (0.006) | (0.005) |
| More than 12 years | 0.343*** | 0.267*** | 0.228*** | 0.201*** | 0.187*** |
| | (0.013) | (0.010) | (0.007) | (0.006) | (0.006) |
| Characteristics of the household | | | | | |
| Family Size | 0.069*** | 0.056*** | 0.048*** | 0.042*** | 0.037*** |
| | (0.003) | (0.002) | (0.001) | (0.001) | (0.001) |
| Able bodied male adults, % | 0.221*** | 0.228*** | 0.254*** | 0.254*** | 0.246*** |
| | (0.025) | (0.012) | (0.015) | (0.010) | (0.012) |
| Number of children below 5 | -0.068*** | -0.044*** | -0.039*** | -0.036*** | -0.034*** |
| | (0.006) | (0.003) | (0.002) | (0.002) | (0.002) |
| Number of elderly in the household | -0.211*** | -0.178*** | -0.153*** | -0.122*** | -0.101*** |
| | (0.025) | (0.011) | (0.008) | (0.009) | (0.012) |
| Family type (base: couples with children) | | | | | |
| Couples without children | -0.256*** | -0.174*** | -0.130*** | -0.102*** | -0.080*** |
| | (0.033) | (0.022) | (0.015) | (0.013) | (0.012) |
| Single caregivers | -0.081*** | -0.044*** | -0.013 | -0.007 | -0.004 |
| | (0.013) | (0.011) | (0.008) | (0.008) | (0.009) |
| Single person households | -0.524*** | -0.459*** | -0.420*** | -0.362*** | -0.320*** |
| | (0.020) | (0.015) | (0.015) | (0.013) | (0.013) |
| Unaccompanied children | -0.722*** | -0.597*** | -0.596*** | -0.526*** | -0.450*** |
| | (0.165) | (0.054) | (0.069) | (0.070) | (0.043) |
| Non-nuclear and other households | 0.054*** | 0.052*** | 0.062*** | 0.068*** | 0.061*** |
| with children | (0.012) | (0.012) | (0.008) | (0.008) | (0.005) |
| Non-nuclear and other households | -0.052** | -0.028* | -0.014 | 0.001 | 0.015* |
| without children | (0.021) | (0.015) | (0.011) | (0.008) | (0.008) |
| Constant | 4.005*** | 4.308*** | 4.505*** | 4.655*** | 4.775*** |
| | (0.030) | (0.017) | (0.018) | (0.016) | (0.014) |
| Observations | 83,606 | 83,606 | 83,606 | 83,606 | 83,606 |

Note:Coefficients come from quantile regressions. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table B4: Results from Quantile Regressions - Quantiles 60 to 90

| | Quantile | | | |
|---|---|---|---|---|
| VARIABLES | 60 | 70 | 80 | 90 |
| Female PA x Wave 17-18 | -0.203*** | -0.212*** | -0.229*** | -0.227*** |
| | (0.006) | (0.006) | (0.007) | (0.010) |
| Wave 17-18 (base: Wave 13-14) | -0.010** | -0.008* | -0.006 | -0.008 |
| | (0.004) | (0.004) | (0.004) | (0.005) |
| Characteristics of the PA | | | | |
| Female PA | 0.051*** | 0.065*** | 0.090*** | 0.080*** |
| | (0.007) | (0.008) | (0.007) | (0.009) |
| Age | 0.002*** | 0.003*** | 0.003*** | 0.003*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Marital status (base: married) | | | | |
| Single or engaged | -0.149*** | -0.145*** | -0.118*** | -0.076*** |
| | (0.017) | (0.017) | (0.015) | (0.012) |
| Divorced or separated | -0.046*** | -0.045*** | -0.044*** | -0.025 |
| | (0.012) | (0.013) | (0.012) | (0.016) |
| Widowed/Widower | -0.094*** | -0.082*** | -0.059*** | -0.042*** |
| | (0.008) | (0.008) | (0.008) | (0.009) |
| Education (base: less than 6 years) | | | | |
| 6-11 years | 0.108*** | 0.098*** | 0.099*** | 0.094*** |
| | (0.004) | (0.004) | (0.004) | (0.005) |
| More than 12 years | 0.179*** | 0.177*** | 0.180*** | 0.182*** |
| | (0.006) | (0.005) | (0.004) | (0.007) |
| Characteristics of the household | | | | |
| Family Size | 0.032*** | 0.026*** | 0.019*** | 0.011*** |
| | (0.001) | (0.001) | (0.001) | (0.002) |
| Able-bodied male adults, % | 0.221*** | 0.199*** | 0.162*** | 0.107*** |
| | (0.010) | (0.009) | (0.010) | (0.012) |
| Number of children below 5 | -0.032*** | -0.029*** | -0.027*** | -0.023*** |
| | (0.002) | (0.002) | (0.002) | (0.003) |
| Number of elderly in the household | -0.096*** | -0.078*** | -0.050*** | -0.034*** |
| | (0.010) | (0.010) | (0.010) | (0.010) |
| Family type (base: couples with children) | | | | |
| Couples without children | -0.061*** | -0.056*** | -0.028** | -0.023* |
| | (0.011) | (0.010) | (0.011) | (0.012) |
| Single caregivers | -0.000 | -0.004 | -0.014** | -0.006 |
| | (0.007) | (0.008) | (0.006) | (0.009) |
| Single person households | -0.261*** | -0.190*** | -0.127*** | -0.050*** |
| | (0.012) | (0.012) | (0.010) | (0.011) |
| Unaccompanied children | -0.332*** | -0.279*** | -0.134** | -0.068 |
| | (0.046) | (0.054) | (0.059) | (0.063) |
| Non-nuclear and other households with children | 0.058*** | 0.061*** | 0.045*** | 0.048*** |
| | (0.006) | (0.007) | (0.008) | (0.007) |
| Non-nuclear and other households without children | 0.031*** | 0.029*** | 0.029*** | 0.029*** |
| | (0.007) | (0.008) | (0.007) | (0.007) |
| Constant | 4.882*** | 4.998*** | 5.125*** | 5.318*** |
| | (0.012) | (0.014) | (0.013) | (0.015) |
| Observations | 83,606 | 83,606 | 83,606 | 83,606 |

Note:Coefficients come from quantile regressions. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table B5: RIF Results, Wave 2013-14

| Component | Mean | Percentile | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
| Male PA | 105.2 | 50.1 | 66.5 | 77.7 | 87.2 | 96.5 | 106.9 | 120.3 | 138.2 | 170.7 |
| Female PA | 106.0 | 38.9 | 57.4 | 72.3 | 85.0 | 96.1 | 109.5 | 124.3 | 146.8 | 187.7 |
| Difference | -0.8 | 11.2 | 9.1 | 5.4 | 2.2 | 0.4 | -2.6 | -4.0 | -8.6 | -17.1 |
| Explained | 0.6 | 4.1 | 5.8 | 4.8 | 2.3 | 1.4 | 1.7 | 0.3 | -3.2 | -8.2 |
| Unexplained | -1.5 | 7.2 | 3.3 | 0.6 | -0.1 | -1.1 | -4.3 | -4.3 | -5.4 | -8.9 |
| Explained | | | | | | | | | | |
| Age of PA | -1.1 | -1.1 | -1.0 | -1.2 | -1.1 | -1.0 | -1.1 | -1.3 | -1.2 | -2.1 |
| Marital Status of PA | 2.0 | 1.7 | 1.7 | 1.8 | 1.8 | 2.0 | 2.3 | 2.7 | 2.4 | 4.2 |
| Education of PA | 2.1 | 2.0 | 2.6 | 2.3 | 2.0 | 1.8 | 1.9 | 2.1 | 2.0 | 3.1 |
| Family size | -2.7 | 0.6 | 0.1 | -0.7 | -1.4 | -2.2 | -3.6 | -4.9 | -6.1 | -8.2 |
| Able bodied male adults, % | 2.7 | 0.6 | 1.7 | 3.2 | 2.9 | 3.6 | 5.0 | 5.1 | 4.5 | 1.0 |
| Number of children below 5 | -1.0 | -0.6 | -1.1 | -1.2 | -1.4 | -1.3 | -1.3 | -1.4 | -0.9 | -0.3 |
| Number of elderly in the household | 0.4 | 0.6 | 0.6 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.5 | 0.5 |
| Family type | -1.7 | 0.3 | 1.3 | 0.1 | -0.9 | -1.7 | -1.8 | -2.5 | -4.4 | -6.4 |
| Unexplained | | | | | | | | | | |
| Age of PA | 0.7 | -9.8 | -7.8 | -7.8 | -5.1 | -2.3 | -2.1 | 0.5 | 11.4 | 7.6 |
| Marital Status of PA | 1.3 | 1.1 | 1.8 | 1.8 | 1.7 | 1.7 | 1.6 | 1.8 | 1.2 | 0.7 |
| Education of PA | -1.7 | 2.0 | -5.1 | -4.9 | -3.9 | -2.5 | -2.9 | -3.0 | 0.5 | -4.3 |
| Family size | -3.8 | -15.0 | -14.0 | -9.6 | -8.8 | -6.2 | -1.0 | 3.8 | 6.0 | 17.3 |
| Able bodied male adults, % | -0.3 | 1.2 | 0.4 | -1.8 | -0.6 | -1.4 | -3.2 | -2.4 | -1.3 | 4.4 |
| Number of children below 5 | 1.7 | 1.1 | 2.0 | 1.6 | 2.0 | 1.1 | 1.5 | 1.8 | 1.0 | 1.4 |
| Number of elderly in the household | 0.5 | 0.8 | 0.8 | 0.7 | 0.5 | 0.4 | 0.5 | 0.5 | 0.5 | 0.9 |
| Family type | -0.5 | -0.4 | 0.5 | 0.1 | -1.1 | -1.6 | -1.8 | -1.6 | -1.3 | -3.7 |
| Constant | 0.6 | 26.2 | 24.5 | 20.6 | 15.2 | 9.7 | 3.3 | -5.7 | -23.4 | -33.2 |

Table B6: RIF Results, Wave 2017-18

| Component | Mean | Percentile | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
| Male PA | 102.4 | 53.6 | 66.9 | 76.6 | 85.2 | 94.5 | 104.8 | 116.7 | 133.5 | 162.1 |
| Female PA | 86.3 | 37.2 | 52.8 | 63.6 | 71.9 | 80.6 | 89.9 | 100.5 | 115.5 | 140.1 |
| Difference | 16.2 | 16.4 | 14.1 | 13.0 | 13.4 | 13.9 | 15.0 | 16.2 | 18.1 | 22.0 |
| Explained | 3.6 | 4.8 | 5.5 | 4.5 | 4.9 | 4.8 | 4.6 | 3.7 | 3.5 | 0.7 |
| Unexplained | 12.5 | 11.6 | 8.6 | 8.5 | 8.5 | 9.0 | 10.3 | 12.5 | 14.6 | 21.3 |
| Explained | | | | | | | | | | |
| Age of PA | -0.8 | 0.0 | -0.1 | -0.3 | -0.4 | -0.5 | -0.7 | -0.9 | -1.2 | -1.5 |
| Marital Status of PA | 1.0 | -0.6 | 0.0 | 0.5 | 0.9 | 1.0 | 1.5 | 1.7 | 1.6 | 2.0 |
| Education of PA | 1.8 | 0.6 | 0.7 | 0.8 | 1.0 | 1.0 | 1.4 | 1.7 | 2.5 | 3.7 |
| Family size | -2.7 | 3.4 | 1.3 | 0.3 | -0.3 | -1.5 | -2.9 | -4.7 | -7.3 | -9.3 |
| Able bodied male adults, % | 2.9 | 0.2 | 0.7 | 1.4 | 2.1 | 2.8 | 3.6 | 3.6 | 4.8 | 5.5 |
| Number of children below 5 | -2.2 | -3.5 | -3.2 | -3.1 | -2.8 | -2.5 | -2.5 | -2.4 | -1.8 | -0.7 |
| Number of elderly in the household | 0.6 | 0.6 | 0.5 | 0.5 | 0.4 | 0.5 | 0.5 | 0.5 | 0.6 | 0.6 |
| Family type | 3.1 | 4.1 | 5.6 | 4.3 | 4.0 | 4.1 | 3.7 | 4.2 | 4.3 | 0.3 |
| Unexplained | | | | | | | | | | |
| Age of PA | -16.3 | -4.2 | -1.6 | -5.3 | -7.9 | -9.5 | -14.3 | -16.7 | -25.8 | -30.4 |
| Marital Status of PA | -0.5 | -0.6 | -0.4 | -0.3 | -0.3 | -0.2 | -0.3 | -0.3 | -0.5 | -0.4 |
| Education of PA | 0.6 | 4.8 | 4.7 | 4.7 | 3.2 | 3.8 | 2.6 | 1.4 | -2.0 | -5.0 |
| Family size | -9.2 | -16.6 | -14.2 | -15.1 | -18.0 | -16.8 | -14.7 | -11.7 | -2.6 | 3.5 |
| Able bodied male adults, % | 2.2 | 2.0 | 1.3 | 1.0 | 0.6 | 0.2 | 0.0 | 1.7 | 1.9 | 8.6 |
| Number of children below 5 | 1.4 | 2.7 | 2.8 | 3.1 | 2.6 | 2.1 | 1.6 | 1.3 | 0.0 | -0.6 |
| Number of elderly in the household | 0.3 | 0.6 | 0.3 | 0.3 | 0.1 | 0.2 | 0.3 | 0.0 | 0.2 | 0.3 |
| Family type | 3.3 | 1.0 | 2.1 | 1.0 | 1.4 | 1.9 | 2.5 | 4.3 | 6.5 | 8.0 |
| Constant | 30.6 | 21.8 | 13.6 | 19.1 | 26.8 | 27.5 | 32.7 | 32.5 | 36.8 | 37.2 |

# Appendix C

# Additional Tables and Figures for Chapter 3

## C1 Additional Tables

Table C1: Teacher Ladder Requirements Before and After the Decree of 2002

| | *Old system* | | *New system* | |
|---|---|---|---|---|
| Ladder | *Education* | *Experience* | *Education* | *Experience* |
| L1 | High School | | | |
| L2 | High School | 2 yrs in L1 | | |
| L3 | High School | 3 yrs in L2 | Technical training on education | |
| L4 | High School | 3 yrs in L3 | – | 3 yrs in L3 |
| L5 | High School<br>Technical training on education | 3 yrs in L4 | – | 3 yrs in L4 |
| L6 | High School<br>Technical training on education<br>Bachelor's degree in other areas | 3 yrs in L5<br>3 yrs in L5 | – | 3 yrs in L5 |
| L7 | High School<br>Technical training on education<br>Bachelor's degree in other areas<br>Bachelor's degree in education | 3 yrs in L6<br>3 yrs in L6<br>3 yrs in L6 | Bachelor's degree | |
| L8 | High School<br>Technical training on education<br>Bachelor's degree in other areas<br>Bachelor's degree in education | 3 yrs in L7<br>3 yrs in L7<br>3 yrs in L7<br>3 yrs in L7 | – | 3 yrs in L7 |
| L9 | Technical training on education<br>Bachelor's degree in other areas<br>Bachelor's degree in education | 3 yrs in L8<br>4 yrs in L8<br>3 yrs in L8 | – | 3 yrs in L8 |
| L10 | Technical training on education<br>Bachelor's degree in other areas<br>Bachelor's degree in education | 4 yrs in L9<br>3 yrs in L9<br>3 yrs in L9 | – | 3 yrs in L9 |
| L11 | Technical training on education<br>Bachelor's degree in other areas<br>Bachelor's degree in education | 4 yrs in L10<br>3 yrs in L10<br>3 yrs in L10 | Master's or PhD | |
| L12 | Bachelor's degree in other areas<br>Bachelor's degree in education | 4 yrs in L11<br>4 yrs in L11 | – | 3 yrs in L11 |
| L13 | Bachelor's degree in education | 3 yrs in L12 | – | 3 yrs in L12 |
| L14 | Bachelor's degree in education | 2 yrs in L13 | – | 3 yrs in L13 |

*Source:* Government of Colombia. *Notes:* This is a simplified version of how I understand the decrees. ”-” means that the requirement of that Lis the same as in the previous ladder. The first 2 ladders of the Decree of 2002 are empty because the norm defines the classification in 12 ladders. However, the national government annually publishes the salary allocation for 14 ladders. The first two levels are used to classify those teachers who do not meet the requirements and yet are hired.

Table C2: Sample Question of the Pedagogy Section of the Test

| |
|---|
| Teachers and managers in a school are changing their teaching approach to respond to the different learning paces of students. They see that the principles of the New School go in line with these interests. They identify that they should define criteria for selecting their resources and methodologies, as for example: |
| 1. Implement experiential learning in accordance with the interests of students.<br>2. Define procedures for the continuous evaluation of theoretical contents.<br>3. Implement activities for the activation of students' previous ideas.<br>4. Use exercises that involve the development of demonstrative processes |

Table C3: Determinants of New Hires at the Municipality Level

| VARIABLES | (1)<br>% new hirings | (2)<br>% new hirings | (3)<br>% new hirings |
|---|---|---|---|
| Log municipality income | -0.106*** | -0.083*** | -0.112 |
| | (0.030) | (0.029) | (0.202) |
| Log population | 1.024*** | 0.793*** | 1.574 |
| | (0.284) | (0.277) | (1.742) |
| Distance to the capital | -0.000 | -0.000 | 0.001 |
| | (0.000) | (0.000) | (0.001) |
| Unemployment rate | 0.010*** | 0.008*** | -0.014*** |
| | (0.001) | (0.001) | (0.002) |
| Students' score | 0.008*** | 0.002 | 0.001 |
| | (0.002) | (0.002) | (0.004) |
| Constant | -1.705*** | -1.131*** | -2.441 |
| | (0.375) | (0.369) | (1.799) |
| Time FE | | ✓ | ✓ |
| Municipality FE | | | ✓ |
| Observations | 1,631 | 1,631 | 1,631 |
| R-squared | 0.052 | 0.105 | 0.777 |

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table C4: Descriptive Statistics of Municipalities by Treatment Status in 2006

| Variables | Hired | Did not hire | Difference |
|---|---|---|---|
| Monthly income per capita | 441.51 | 508.89 | $-67.38^{***}$ |
| Population | 58323.79 | 22500.83 | $35822.96^{*}$ |
| Police stations per capita | 0.06 | 0.07 | $-0.01^{*}$ |
| Unemployment rate | 11.38 | 10.45 | $0.93^{***}$ |
| Density | 201.45 | 103.37 | $98.08^{**}$ |

Source: DNP. Income per capita is measured in thousand colombian pesos. Density is measured as number of people per $Km^2$

*** p<0.01, ** p<0.05, * p<0.1

Table C5: School Level Effects on Students' Dropout Rates, by Grade

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| VARIABLES | Total | Grade 1 | Grade 2 | Grade 3 | Grade 4 | Grade 5 |
| New hirings | -0.005*** | -0.007*** | -0.005*** | -0.003 | -0.005** | -0.004* |
| | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Constant | 9.174*** | 13.420*** | 9.253*** | 10.278*** | 8.262*** | 7.811*** |
| | (0.305) | (0.464) | (0.447) | (0.522) | (0.483) | (0.506) |
| Observations | 95,670 | 93,850 | 93,496 | 93,828 | 90,588 | 88,111 |
| R-squared | 0.036 | 0.033 | 0.018 | 0.014 | 0.012 | 0.009 |
| Number of Ins_code | 15,742 | 15,675 | 15,667 | 15,621 | 15,375 | 15,117 |

| | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|
| VARIABLES | Grade 6 | Grade 7 | Grade 8 | Grade 9 | Grade 10 | Grade 11 |
| New hirings | -0.017*** | -0.015*** | -0.010** | -0.022*** | -0.006 | -0.004 |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.005) | (0.004) |
| Constant | 6.319*** | 3.170*** | 3.386*** | 1.507** | 1.304** | 1.572*** |
| | (0.688) | (0.586) | (0.631) | (0.641) | (0.653) | (0.557) |
| Observations | 39,268 | 38,176 | 36,959 | 35,448 | 27,125 | 25,684 |
| R-squared | 0.015 | 0.008 | 0.006 | 0.005 | 0.001 | 0.002 |

Clustered standard errors at the school level in parentheses. Regressions include school FE, year FE and a time trend. *** p<0.01, ** p<0.05, * p<0.1

## C2 Additional figures

Figure C1: New Hirings After the Policy Implementation

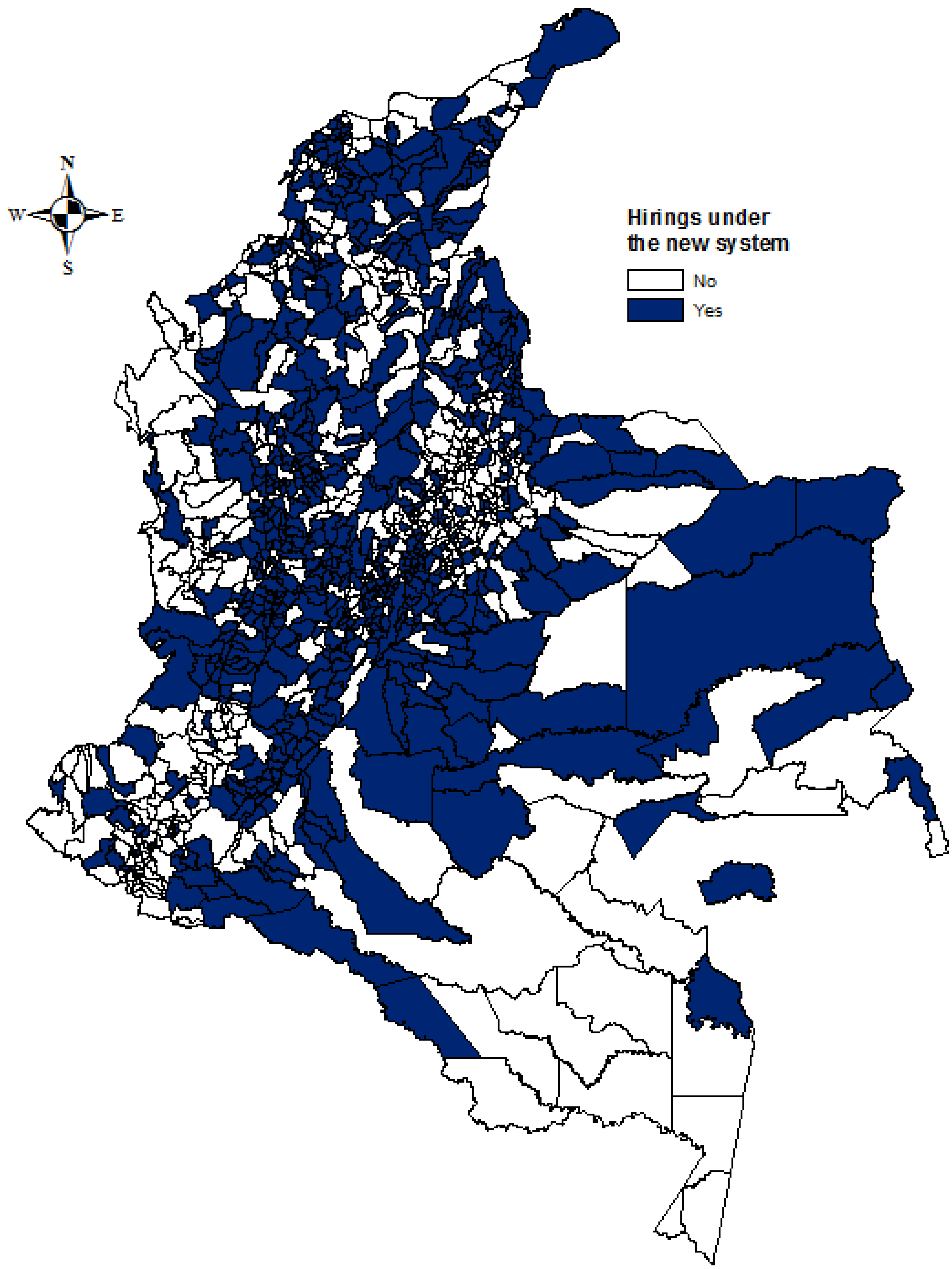


Table C6: Heterogeneous Effects by Gender: Eight Years After Implementation

| | Male students | | | Female students | | |
|---|---|---|---|---|---|---|
| | Fraction | Robust | Romano-Wolf | Fraction | Robust | Romano-Wolf |
| Test | $NS_{st}$ | p-value | p-value | $NS_{st}$ | p-value | p-value |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Global | 0.059 | [0.000] | $\{0.019\}$ | -0.008 | [0.474] | $\{0.490\}$ |
| Math | 0.024 | [0.078] | $\{0.255\}$ | -0.027 | [0.025] | $\{0.078\}$ |
| Language | 0.063 | [0.000] | $\{0.019\}$ | 0.025 | [0.021] | $\{0.078\}$ |
| Biology | 0.055 | [0.000] | $\{0.019\}$ | 0.017 | [0.107] | $\{0.216\}$ |
| Philosophy | 0.011 | [0.367] | $\{0.647\}$ | -0.031 | [0.006] | $\{0.019\}$ |
| Physics | 0.018 | [0.127] | $\{0.314\}$ | -0.050 | [0.000] | $\{0.019\}$ |
| Chemistry | -0.001 | [0.960] | $\{0.961\}$ | -0.043 | [0.000] | $\{0.019\}$ |

*Notes:* The results shown in columns (1)-(3) and columns (4)-(6) of this table come from independent linear regressions in which the variable of interest is the standardized tests scores of students in the corresponding subject. The explanatory variable shown in this table corresponds to the number of teachers registered in the new system as a percentage of the size of the teaching staff. This variable latter takes the value of zero for those schools that did not hire teachers in 2006 and for all schools in 2005. Estimates for male students are shown in Column 1. Estimates for female students are shown in Column 4. Each regression uses school-level and year-wave fixed effects. p-values computed with robust standard errors are displayed in brackets. p-values computed with the Romano-Wolf correction are shown in column 3 in curly brackets. The number of observations in in the regressions of male students is 1,302,516 and of female students is 1,595,736.