Retail Expansion in Developing Countries: Evidence from Hard Discount Stores*

Lukas Delgado-Prieto[†] Andrea Otero-Cortés[‡] Andrés Calderón[§]

June 19, 2025

Abstract

Hard discount stores have reshaped the retail sector by offering low-cost products. While this business model has gained a significant market share in many countries, how it impacts the labor market remains unclear. To address this, we analyze their rapid expansion in Colombia using administrative and survey data. Our findings show that post discounters' entry, local formal employment increases by 10% on average, with strong spillovers from retail to manufacturing. Consistent with this, we find an increase in local tax revenues from manufacturing and commerce activities. These results suggest that hard discount stores can foster formalization in developing countries.

Keywords: Hard discount stores, competition, local labor markets, informality.

JEL Codes: E24, J46, O17.

---

*We are thankful to Luis Eduardo Arango, Jhorland Ayala, Jaime Bonet, Leonardo Bonilla, Juan Esteban Carranza, Jorge Flórez, Luis Armando Galvis, Sergio Ocampo, Miguel Talamas, Jan Stuhler and Gabriel Ulyssea, as well as participants at the 49th Symposium of the Spanish Economic Association, CEDE Seminar, UC3M Applied Reading Group, LACEA-LAMES 2023, Red Investigadores de Economía and Banco de la República seminar for helpful comments and discussions. We thank María Camila Gómez and Julián Gallego Villa for providing excellent research assistance. The views and opinions expressed here are those of the authors and do not necessarily reflect the views of Banco de la República or its Board of Directors. Any errors are exclusively the responsibility of the authors. This paper was previously circulated as: "The Impact of Hard Discount Stores on Local Labor Markets: Evidence from Colombia".

[†]Corresponding author: Department of Economics, University of Oslo, Norway. Email: laprieto@econ.uio.no.

[‡]Researcher, Banco de la República, Colombia

[§]Research fellow, Inter-American Development Bank, Washington, D.C., United States of America

1 Introduction

Hard Discount Stores (HDS) have become prominent in the global retail market, achieving sales that exceed 442 billion dollars in 2022 (Euromonitor International, 2023b,a; MarketLine Industry Profiles, 2023). Leading European hard discount chains Aldi and Lidl, for example, rank among the top ten global retailers (Statista, 2023). HDS continue to expand within and between countries based on key strategies like a limited product assortment, a high share of low-priced, own-labels that ensure a high quality/price ratio, and efficient logistics and operations (Jurgens, 2014; Sachon, 2010). This rapid growth raises questions about their impact on retail businesses, consumer behavior, and the labor market. The latter is particularly unclear, as these stores can boost employment through direct hires. Still, they also increase competition within the retail sector, potentially causing job losses among incumbent formal and informal firms. Furthermore, there can be employment spillovers onto other sectors, for instance, due to upstream supply chain linkages (de Paula and Scheinkman, 2010), agglomeration forces (Evensen et al., 2023), and/or via demand multiplier effects (Gerard et al., 2024).

In this paper, we examine the staggered expansion of hard discount chains in Colombia. Unlike the United States or Europe, Colombia's labor market has a distinct feature, with over half of the workforce employed informally and the majority of businesses lacking legal registration. Hard discount chains, however, are formal firms that must comply with labor and tax regulations, meaning they hire workers formally, adhere to minimum wage laws, and ensure their suppliers operate within the formal sector. Since the entry of the first hard discounter in 2009, the retail sector has undergone a major transformation. Through sizable investments, the leading chains have established more than 4,000 stores nationwide as of 2022 (Euromonitor International, 2023c). Being more efficient than supermarkets, with higher sales per square meter and lower operating costs, their rapid expansion has made them overtake hypermarkets in market share by 2020 and supermarkets by 2021, with a major sales growth of 235% between 2017 and 2022 (Euromonitor International, 2023c). At the same time, social security records indicate that the top three hard discount chains employed over 26,000 formal workers in 2019. Following the model of European hard discounters, these stores have expanded into residential neighborhoods and city centers, offering a limited assortment of predominantly own-brand products at low prices, acting as a bridge between formal and informal retail or modern and traditional retail (Bachas et al., 2024; Lagakos, 2016). Still, hard discounters offer a more limited product range, which means they are not perfect substitutes for supermarkets and can lead consumers to shop at multiple establishments.[1]

To assess the impact of HDS on local labor markets, we assemble a unique dataset combining administrative records of social security contributions, which allows us to capture the universe of formal workers over time, subsidized social protection beneficiaries, labor force survey data, tax revenues, and information on each store's location and opening year for the three leading hard discount chains in Colombia from 2010 to 2019 at the municipal level.[2] Nonetheless, identifying the causal effect of HDS entry on local labor markets is challenging due to the endogenous nature of store location decisions. To address this issue, we leverage the rapid and staggered roll-out of HDS across different municipalities in the country, which was based on logistical and market size factors, to identify the causal effect of HDS entry on local economies. Using an event study research design, we quantify the dynamic treatment effects of HDS on formal and informal labor markets at the local level, based on the estimator of Callaway and Sant'Anna (2021).

Our identification strategy does not assume that store opening decisions are exogenous. In fact, we show there are differential trends between treated and never-treated municipalities before the arrival of the first discounter. Instead, we argue that the *timing* of the first hard discount store opening, excluding the ones opening in capital cities, is unrelated to local unobserved trends. Therefore, by comparing cohorts of intermediate-sized urban municipalities where hard discount chains opened earlier to those where they opened later, we identify the effect of HDS on local labor markets. To start, we provide suggestive evidence that the *timing* assumption holds in our scenario, as the treated and not-yet-treated groups exhibit similar trends in a number of outcomes,

[1] For instance, Florez-Acosta and Herrera-Araujo (2020) documents French households commonly visit multiple supermarkets per week even if they offer similar products, suggesting the entry of HDS may accelerate multi-stop shopping behavior among consumers.

[2] Formal workers are defined as those who contribute to social security, and informal workers are those who do not comply with social security regulations.

such as employment, wages, working hours, and taxes several years before the entry of HDS. A threat to our identification is the presence of time-varying shocks that are correlated with the *timing* of the arrival of hard discounters, such as spillover shocks from the main capital cities, which are potentially more pronounced in the early-treated municipalities or additional firm entry at the same time as the hard discount chains. We address such potential concerns by controlling for distance in our baseline specifications and showing the lack of firm entry post-treatment.

We find three main results. First, the entry of a hard discount store in a municipality leads to a 10% increase in the formal employment-to-population ratio using both administrative and survey data. The increase is robust to allowing for violations in pre-trends and remains statistically significant when randomizing treatment timing (Rambachan and Roth, 2023; Roth and Sant'Anna, 2023) This result is primarily driven by increased formal employment in manufacturing, construction, and retail. We discuss plausible explanations for these inter-sectoral spillovers from retail. On one hand, HDS may stimulate the local demand for intermediate inputs benefiting other complementary industries, as indicated by the largest hard discounter (La República, 2022). In this context, discounters have incentives to source some goods locally, as transportation costs in Colombia are one of the highest in the world (Londoño-Kent, 2009). On the other hand, hard discount chains are large formal firms with access to top-notch technology and cheaper formal credit. They can afford to pay higher wages to attract both formal and informal workers due to their higher productivity (Ulyssea, 2020; Lagakos, 2016). Therefore, HDS entry may stimulate the local economy via increasing earnings of local workers, leading to multiplier effects (Gerard et al., 2024).

Second, we find a positive impact on local tax revenues, as the ratio of taxes-to-baseline public revenues increases by an average of 7.5% after the entry of discount stores. This increase is primarily driven by manufacturing and commerce taxes, supporting the idea that HDS have positive spillovers onto other sectors. Third, while our estimates for informal retail employment are small and sometimes positive, we cannot rule out the possibility that HDS might lead to lower entry of informal neighborhood shops due to increased competition, as we lack precise data to test this hypothesis. In this aspect, Talamas Marcos (2025) shows that an additional convenience store entry in a neighborhood adversely affects the creation of neighborhood shops in Mexico. It also finds that incumbent neighborhood shops survive by leveraging their comparative advantage in supplying fresh products, offering cheaper prices due to their low fixed costs coming from avoiding tax compliance and operating within the owners' houses (Ramos-Menchelli and Sverdlin-Lisker, 2023), and also reducing costs by shrinking inventories.[3] Consistent with this, we also observe an imprecise negative effect on self-reported labor income of informal retailers in the later years of treatment. Taken together, our findings suggest that HDS can promote local economic formalization (through the labor market and consumption) in developing countries with .

This paper contributes to different strands of the literature. Most existing studies on the labor market effects of expanding retail chains focus on the expansion of Walmart. Walmart's entry to new US counties negatively impacts local labor markets by reducing retail employment and earnings in the mid-to-long run, as the company exploits its monopsony power and have adverse effects on other local retailers (Basker, 2005; Neumark et al., 2008; Wiltshire, 2023; Haltiwanger et al., 2010; Dube et al., 2007). Related literature explores the impact of expanding e-commerce fulfillment centers (FCs), like Amazon, in the US (Chava et al., 2023; Cunningham, 2024). Similar to our empirical strategy, Chava et al. (2023) exploits the staggered roll-out of FCs, using areas not yet treated as control, finding negative effects on retail employment but positive employment spillovers to other sectors: in transportation and warehousing, whereas Cunningham (2024) in tradable services. Their results on wages move in opposite directions. Similarly, Greenstone et al. (2010) studies the agglomeration spillovers of the arrival of "Million Dollar Plants" in the US by comparing counties where they enter relative to counties that narrowly lose them.

Our paper diverges from previous work as our context is different: we are studying the entry of hard discount chains in a setting characterized by high informality, where informal competing businesses (neighborhood shops) coexist with formal supermarkets. Furthermore, hard discount chains fundamentally differ from the concept of Walmart or Amazon FCs. To the best of our

[3] In a different market, Macchiavello and Morjaria (2021) find that competition negatively affects several market outcomes when formal employment contracts are not enforceable in Rwanda's coffee mill industry. They find that new mill openings increase farmers' temptation to default on previous relational contracts, worsening farmers' welfare and indirectly reducing mills' profits.

knowledge, this is the first paper to estimate the impact of HDS in developing countries across the formal and informal labor markets.[4] Our work is most similar to Talamas Marcos (2025), which analyzes the impact of chain-run convenience store expansion in Mexico. Yet our focus is not on the responses from neighborhood informal shops. Instead, we center on the broader implications for formal employment and local tax revenues. Finally, we aim to identify spillover effects of HDS entry on employment in other industries, such as manufacturing and construction, contributing to the literature on supply chain effects, as hard discounters have incentives to source their products from local formal suppliers (de Paula and Scheinkman, 2010; Gerard et al., 2023; Rios and Setharam, 2018).

This paper is structured as follows. The next section provides the institutional context of the retail sector in Colombia. Section 3 describes our primary data sources on local labor markets and how we measure the arrival of HDS in the municipalities. Section 4 discusses our identification strategy and its assumptions. Section 5 presents our results on employment, taxes, and labor income and working hours. We conclude in section 7.

2 Labor Market Context and Retail Sector

In Colombia, informal jobs are the prevalent type of employment. However, in recent years, there has been a noticeable upward trend in formal employment rates, both in capital cities and other urban areas (see Appendix Figure C.1). There are several factors contributing to this phenomenon. One key factor is the relatively lower labor costs, since the end of 2012, for employers to hire workers formally (Fernández and Villar, 2017; Morales and Medina, 2017; Kugler et al., 2017).[5]

[4] In developed economies, Cho et al. (2015) studies the impact of large discount stores in Korea, finding positive impacts on local retail employment driven by the large discounters and by the positive spillover effects on other retail sectors. Whereas Evensen et al. (2023) examines how the expansion of discounters affects incumbent local grocery stores in Norway. This paper identifies two opposing effects on sales and consumer traffic: a positive effect driven by store complementarity and a negative effect from fiercer competition. The agglomeration effect dominates when new discounters are located near existing retailers, while the competitive effect prevails when the distance between new and established stores increases.

[5] There are other events that, in turn, might have attenuated the increase in formal employment after 2015, like the arrival of Venezuelan immigrants (Delgado-Prieto, 2024a).

In this context, we want to study the impact of the arrival of HDS on local formal employment, as they only hire workers formally and demand inputs only from legally registered suppliers, and show how this contributes to the national observed increase in formal employment.[6]

2.1 Retail Sector

The grocery retail market in Colombia is a sizeable 40 billion-dollar market (Euromonitor International, 2023c). It represents around 13% of South America's retail market, and at the national level, it accounts for more than half of the total retail sales (MarketLine, 2023). Three actors have historically played a significant role in this market: small neighborhood shops (mainly informal), supermarkets, and hypermarkets. According to market data from Euromonitor International (2023c), in 2017, neighborhood shops accounted for 52% of the retail grocery sales, while large retailers were responsible for around 23% of the sector's income.

Neighborhood shops are mostly informal businesses that offer a limited supply of essential goods, primarily in residential areas.[7] They tend to create a close relationship with their customers, even offering informal credit to them (Talamas Marcos, 2025).[8] Neighborhood shop prices are not necessarily lower than supermarket prices, yet as they sell smaller quantities of essential items, the ticket for daily grocery shopping would fit into lower-income households' daily budgets. In contrast, traditional supermarkets are large formal retailers with infrastructures similar to those in the US, where customers can find goods ranging from fresh products to home appliances. In this context, the arrival of HDS increases competition for the market of certain products with supermarkets and neighborhood shops (Sánchez Duarte, 2017). We provide further descriptive

[6]One potential concern is that the 2012 Colombian tax reform can confound the impacts from the arrival of HDS. However, this reform was enacted at the national level and affected all workers earning up to 10 times the minimum wage and working in firms with at least two employees. Thus, it is unlikely that there is a correlation between the first arrival of a hard discount chain to a municipality and the treatment intensity of the 2012 reform.

[7]Neighborhood shops are not to be confused with convenience stores such as *Oxxo* or *7-Eleven*.

[8]Neighborhood shops are prevalent in the country. In 2019, there were approximately 270,000 stores of such category in the main 100 municipalities of Colombia (Fenaltiendas, 2019), with median weekly sales of around $350 USD and an average ticket per customer of $1.50 USD. They target low-income households that usually only buy everyday groceries. For instance, instead of purchasing a one-kilogram bag of rice in the supermarket, they buy one cup of rice in the local neighborhood shop.

statistics regarding profits, employment and formality of neighborhood shops using data from the Micro Businesses Survey in Appendix A.

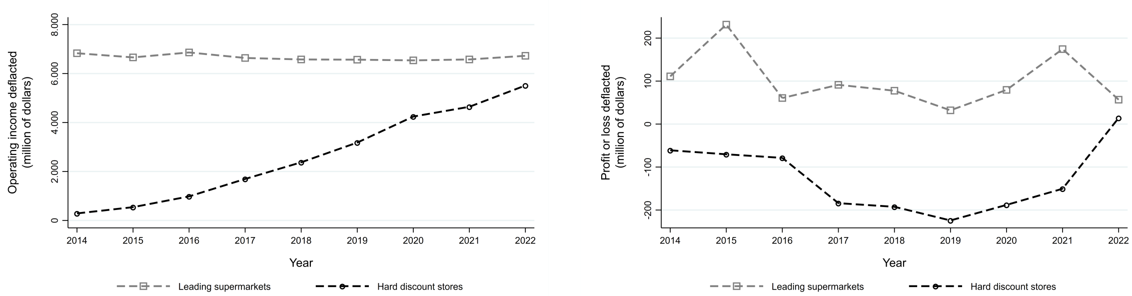
The supermarkets and hypermarkets segment, an important source of formal employment, has been largely dominated by three firms: *Grupo Éxito* (with the *Éxito* and *Carulla* brands), *Supertiendas & Droguerías Olímpica* (*Olímpica* and *Sao*), and *Cencosud S.A.* (*Jumbo* and *Metro*). This format was introduced in Colombia in the late 1940s by José Carulla Vidal (Silva Guerra, 2011), inspired by Mexico's growing supermarket chains like *Sumesa*. During the 1950s, supermarkets spread to major Colombian cities with new owners and chains (Grupo Exito, 2015). According to Euromonitor International (2023c), the sector has a market size of about 4 billion dollars, with around 2,200 outlets. Average store size is 750 square meters, with annual sales of 2 million dollars (US$2,450 per square meter).

By the late 1990s, hypermarkets entered the market, with clear socioeconomic segmentation among customers. Some chains offered lower prices targeting low-to-mid-income consumers, though not under the hard-discount format but through lower-quality products. Colombia has about 210 hypermarkets, mostly in large cities, also generating 4 billion dollars in sales (Euromonitor International, 2023c). The average hypermarket has 4,700 square meters of selling space and sells around 19 million dollars annually (US$4,020 per square meter).

The first hard discount chain, *D1*, opened in Colombia in 2009, inspired by the German Aldi model. The second, *Ara* (owned by a Portuguese company) followed in 2012, and the third, *Justo & Bueno*, in 2016. HDS stores are smaller than traditional supermarkets, typically 250 to 300 square meters, and reduce operating costs through strategies like streamlined distribution with a limited product range, minimal advertising, displaying goods in shipping boxes, and lean staffing. Though HDS have lower annual sales than supermarkets, they generate around US$3,472 per square meter, 41% more than traditional supermarkets (Euromonitor International, 2023c).

Appendix Figure B.1 shows HDS growth from 2010 to 2019, both overall and by chain. Over the decade, they expanded rapidly across Colombia. The three main chains opened nearly 3,000 stores and employed around 26,000 workers across 408 of Colombia's 1,103 municipalities, with the remainder being often rural municipalities in the Amazonía and Orinoquía regions. Figure 1 compares operating income and profits between supermarkets and hard discounters. Supermarkets have historically led in both metrics, but the gap in operating income has narrowed significantly. Today, the leading hard discounter surpasses the top supermarket in operating income. In terms of profits, the two main hard discounters (the third exited in 2021) only became profitable in 2022, after years of heavy investment. In contrast, the three major supermarkets only began seeing a decline in profits that same year.

Figure 1: **Operating income and profits of the leading supermarkets and hard discounters**



(a) Operating income **(b) Profits**

Note: We aggregate for the three leading supermarkets and three leading hard discounters in the country, operating income and profits at real prices (CPI 2018=100). We use the exchange rate of December 2018, 1 USD=3,250 COP. Source: Supersociedades (Operating income and Profits), Banco de la República (Exchange rate), and DANE (CPI series).

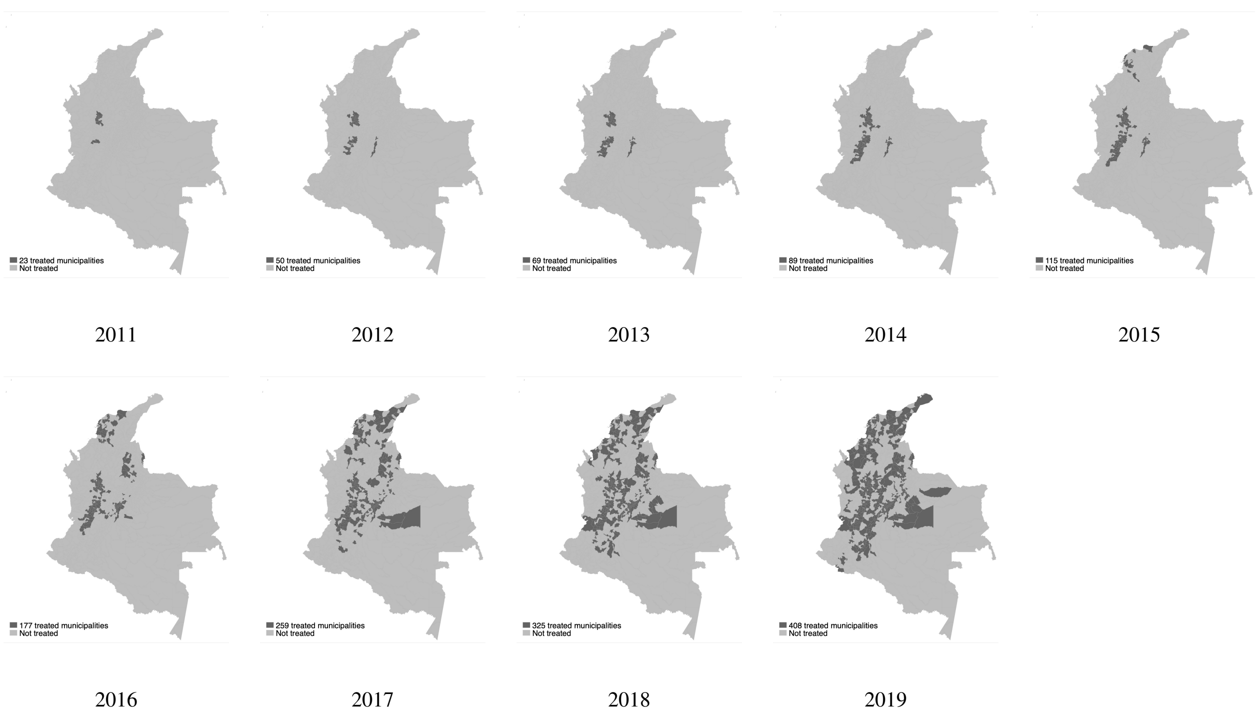
The expansion of hard discount chains not only increased retail options for consumers but also brought substantial investment shocks to the municipalities where they opened. For example, in a 2018 interview, Ara's manager reported that opening 500 stores involved over 400 million euros, about 0.9 million USD per store (Portafolio, 2018; Morante, 2018). Similarly, in 2020, D1's manager stated the firm had invested 123 million USD to open 800 stores (roughly 154,000 USD per store, González Bell (2020)). As a benchmark, Ara's per store investments equal 1.9 times the median annual local tax revenue in our sample of municipalities, while for D1 is 36%. In terms of investment expenditures, they represent 19% and 4%, respectively.[9] Many municipalities host

[9]Median annual tax revenue in our sample (2010–2019) is 425,674 USD, and median investment expenditure is

multiple hard discount chains, greatly increasing the total investment received.

Figure 2 shows that hard discount store expansion began in Colombia's central region, then moved to the Caribbean and southern areas in a staggered pattern, as chains focused on different regional markets to grow aggressively. According to market data, by 2017 (when discounters already had over 1,700 outlets) hard discounters' sales grew by 235% between that year and 2022, compared to 29% in the aggregate grocery retail sector. The number of stores rose by 133%, and selling space by 141%. The market share increased from 5.2% to 13%, surpassing hypermarkets in 2020 and supermarkets in 2021. By the end of 2022, there were more than 4,000 outlets (Euromonitor International, 2023c), with large-scale investment plans to increase the number of HDS in the following years (La República, 2023), highlighting how policy-relevant it is to study their impacts on local labor markets.

Figure 2: **Geographic location of Hard Discount Stores**



Note: This figure shows the geographic expansion of Colombia's three main hard discount chains using stores' stock between 2011 and 2019. Source: Authors' calculations using public location data from hard discounters' websites.

4,150,126 USD, based on the Colombian municipalities panel (Acevedo and Bornacelly Olivella, 2014). We confirm the chains' investment figures using financial data from *Supersociedades*. Net cash flow from investing activities per new store was about 1.3 million USD for Ara, 47,000 USD for D1, and 90,000 USD for Justo & Bueno.

In line with the geographical pattern, we show that hard discount chains decide store location decisions based on fixed municipality characteristics, such as potential market size and proximity to logistical centers. Appendix Table C.2 presents cross-sectional correlations between the municipality's probability of receiving HDS on its population (serving as a proxy for market size), prevalence of rurality (a proxy for economic development), driving time to Bogotá and Medellín (where the largest logistical centers are based) and the formal employment share in the service industry (a proxy for sectoral composition). We estimate the probability that a HDS enters a new municipality by chains and for all chains. The analysis reveals that the observed variables that we consider play a significant role in determining store openings (R^2 of 0.538), with population size being one of the main factors and the share of formal tertiary employment showing the weakest correlation.

A potential concern for our *timing* assumption is that hard discount chains might decide when to open their stores based on the dynamics of local employment or tax collection. This would violate the assumption of parallel trends on which the differences-in-differences (DiD) relies. We test this empirically by running year-specific regressions whose dependent variable is the probability of a store opening in a given municipality, and we add as covariates both fixed and dynamic characteristics, such as the annual growth in formal employment, wages, and tax revenues with respect to the previous year (all of which come from administrative data).

Appendix Table C.3 compares this set of covariates between municipalities that received their first store in a given year to those where discount chains never opened, and Appendix Table C.4 to those where HDS would arrive later. All static characteristics, but the share of tertiary formal employment, are significantly correlated with the likelihood of opening in both exercises, while municipal trends in formal employment and wages, as well as tax collection, do not seem to play a key role. Therefore, the evidence suggests that discounters based their decisions on fixed characteristics such as potential market size and logistic convenience, and not necessarily on the dynamics of formality or tax collection of municipalities, supporting our identifying assumption.

3 Data

For our treatment exposure, we measure the year of the arrival of a hard discount chain into a municipality. We rely on web-scraped information on the universe of HDS for the three leading chains in 2019. This allows us to identify their location, and with data on business registration from the Chambers of Commerce, we identify the date of entry to a given municipality. Colombian legislation requires that all firms register their establishments, including stores and distribution centers. We use the date of registration of a store as a proxy for its opening date and match the web-scraped spatial data with the Chambers of Commerce data using the store's name and parent company. Our final data set comprises 2,847 stores with their municipality and proxy of the opening date. We then compute the year in which the first HDS opened in a municipality and end up with 414 observations (372 excluding department capital cities and those where HDS arrived before 2011 or after 2019). Appendix B explains the matching process in greater detail.

This paper uses multiple data sources to capture both formal and informal outcomes. First, we use the employer-employee matched administrative records on social security contributions (PILA, by its acronym in Spanish) from 2010 to 2018. PILA contains the universe of mandatory contributions to social security that are made on behalf of each formal worker in Colombia. This data source allows us to identify formal employment by industry level and per municipality.

For the main analysis, we exclude capital cities, where the relative market shock from store openings is small and timing may be more influenced by labor market trends, violating our identification assumption. This restriction removes 83% of observations from the PILA, or about 7 million per month, as most formal jobs are concentrated in capital cities. We also exclude municipalities without HDS by 2019, as their labor markets (typically small or rural) likely follow different dynamics. The final estimation sample focuses on intermediate urban municipalities, with 1.6 million observations on average per month across 372 of the country's 1,120 municipalities, covering 38% of the population. To address PILA's limitations in identifying municipality and industry for some firms in certain years, we use the panel component of the data to extract the information from other years.[10] Still, we note that industry-level results may be noisier due to misclassification by firms in all years.

As the main source of informality, we use a monthly cross-sectional household survey that covers approximately 240,000 households per year (GEIH, by its acronym in Spanish). GEIH is Colombia's labor force survey and has extensive sample coverage across the country, though not in all the municipalities where hard discounters are. Thus, the estimation sample of municipalities using GEIH reduces to 191 from the 372 municipalities we observe in PILA. However, GEIH allow us to characterize the informal and formal labor outcomes as we use the survey questions on workers contribution to the social security system to categorize informal and formal workers. For the analysis, we aggregate this information at the municipality level using department survey weights after restricting to individuals in urban areas between 18 and 64 years and, for the wage analysis, we further consider only workers with positive labor income.[11]

The GEIH survey is not representative at the municipal level, even though the municipalities included are among the largest surveyed outside capital cities. However, since our analysis aggregates similar municipalities into treatment cohorts (ranging from 5 to 35 municipalities, see Table 2), our identifying variation occurs at the cohort not municipal—level, mitigating concerns about statistical noise from limited representativeness. We also aggregate monthly surveys into annual data per cohorts, further reducing noise. Crucially, our GEIH-based estimates for formal employment closely match those from administrative data, supporting the reliability of the survey.

We also use the census of beneficiaries of subsidized social protection (SISBEN, by its acronym

[10] We have two limitations on identifying municipality and industry classification for a fraction of our sample. First, we have missing information on municipality location for some observations from 2017 to 2019. We use two strategies to impute this information: computing the mode of the previous municipalities in which the same worker appears registered or, in case this information is also missing, using the 2020 worker location information. After this procedure, we exclude 1.8% of the workers in PILA from our sample due to wrongful municipality code. Second, industry classification in PILA is self-reported by the firm using a 4-digit International Standard Industrial Classification (ISIC) code. Up to 2013, the ISIC revision three was the standard, and after that classification was updated to revision four. From 2013 to 2019, some firms reported their industry code using the old revision three ISIC list instead of the current one. To cope with this issue, we computed the mode of the 4-digit ISIC self-reported code for each firm in PILA and search for its classification under the ISIC rev. 4 list and if the code reported by the firm does not appear under the revision 4 list, we use the revision 3 version. Approximately 1% of workers per month from our estimation sample are not classified in any industry due to wrongful ISIC coding.

[11] Because the municipality information of the GEIH survey respondents is not publicly available, we obtained it through specialized DANE data centers.

in Spanish) to approximate the stock of non-formal workers at the municipal level. Although this source contains the universe of low-income individuals in the country, including those who can be out of the labor force or unemployed, it also captures informal workers who receive social protection subsidies. Therefore, SISBEN is a proxy we use to measure the size of informality at the local level.

To capture broader tax effects on the local economy, we use data from *Operaciones Efectivas de Caja*, collected by *Departamento Nacional de Planeación* (DNP, by its acronym in Spanish), which contains detailed information on revenue by each type of taxes collected at the municipal level, such as industry and commerce tax, property tax, among others.

3.1 Descriptive Statistics

Administrative Records (PILA). Appendix Figure C.2 panel (a) shows that formal employment in the services sector grew steadily from 2010 to 2019, driven by favorable economic conditions and labor reforms. Nonetheless, informality remains widespread (about half of Colombia's workforce lacks access to the pension system) with even higher rates outside capital cities, as confirmed by GEIH survey data.

Against the persistent informality, the rise of hard discount chains represents a notable push to formal employment. From employing fewer than 500 formal workers in 2012, these chains grew to nearly 26,000 employees by 2019 (see Appendix Figure C.2 panel (b)), highlighting their rapid expansion. Notably, 86% of these workers are full-time employees.[12] As shown in Appendix Figure C.2 panel (c), the average number of workers per store increased until 2016 and then declined through 2019, averaging 10 workers per store.

To better understand who works in these chains, we analyze individual employment histories using the PILA. Between 2010 and 2019, 57,963 people worked at one of the three leading HDS. Of these, 16% were first-time formal workers (appearing in the PILA for the first time in HDS)

[12]Colombian legislation made part-time formal employment relatively costlier than full-time, although regulatory changes in 2014 allowed weekly contracts that promoted part-time employment (de la Parra et al., 2024).

while 31% had previous formal employment elsewhere. First-timers are younger (average age 24), 55% are women, and 47% start on part-time contracts. In contrast, job switchers are typically older (average age 30), 60% are men, and over 75% previously worked in retail, wholesale, or services. These patterns suggest HDS are an entry point into formal employment, particularly for younger workers.

Finally, Appendix Table C.1 presents descriptive statistics for the municipalities in our estimation sample. In 2011, treated municipalities had an average of 6,781 formal workers, a figure that declined as HDS expanded into smaller municipalities over time. While over 90% of formal jobs in these areas lie outside commerce, hotels, and restaurants, the share in that sector grew from 5.1% in 2011 to 10% in 2018. Moreover, the shares of self-employed and minimum-wage workers remained similar, reflecting similar labor market structures in these municipalities.

Labor Force Survey (GEIH). Table 1 shows the mean and standard deviation of several labor market outcomes by year and treatment status in our estimation sample.[13] The average of total employment by treatment cohorts, weighted by the employed population in 2010, shows that hard discount chains prioritized large municipalities for their initial openings (the employed population in the typical early-treated municipality was almost twice as in the typical not-treated-yet municipality in 2011). However, both numbers decreased over time, suggesting that later on, they opened in smaller municipalities.

Despite the difference in employment between treated and not yet treated municipalities, there are no significant gaps between the two groups in most outcomes during the first years of the expansion of HDS. Employment and inactivity rates were similar in 2013, as wages and working hours, even when disaggregated by informality status.

The most considerable differences come from the informality rate and the industry composition by municipalities. For instance, formal employment in 2013 represented, on average, 47% of the total employment in treated municipalities, while the share was 37% in the not yet treated group (nearly a nine pp gap). At the same time, the typical municipality with HDS presence by 2013

[13] Appendix Table C.5 further shows descriptive statistics on wages and working hours for the formal and informal sectors.

was less dependent on retail and more dependent on the primary and secondary sectors: retail workers represented 16% of 2010 total employment in the treated group, compared to 19% in the not yet treated, and the respective shares for the primary and secondary sectors are 34% and 30%. Conversely, the rest of the commerce and the services sectors had a similar weight in the local economies of treated and not yet treated municipalities.

Table 1: **Employment statistics for the estimation sample using survey data**

	Treated				**Not yet treated**			
	2011	**2013**	**2016**	**2018**	**2011**	**2013**	**2016**	**2018**
Employment rate	68.6	70.4	72.3	70.8	69.7	71.1	68.7	68.1
	(1.9)	(6.4)	(5.1)	(6.4)	(7.0)	(6.0)	(7.3)	(9.3)
Unemployment rate	13.1	11.6	10.2	11.3	12.0	10.4	12.5	11.1
	(2.6)	(3.8)	(2.9)	(4.6)	(4.4)	(4.1)	(5.0)	(5.6)
Inactivity rate	21.0	20.7	19.5	20.2	21.0	20.7	21.6	23.7
	(2.9)	(5.4)	(5.0)	(6.1)	(6.6)	(5.6)	(6.0)	(7.9)
Employment share: Retail	19.2	16.2	18.8	19.5	18.1	19.4	21.1	17.4
	(3.3)	(4.2)	(5.0)	(6.6)	(6.6)	(6.4)	(8.4)	(9.4)
Employment share: CHR without retail	13.5	13.8	15.9	15.6	12.2	14.7	17.8	14.4
	(3.5)	(4.5)	(4.6)	(5.2)	(4.3)	(5.7)	(7.6)	(6.8)
Employment share: Primary and secondary	29.3	34.3	33.6	34.3	32.2	30.3	30.9	38.2
	(2.7)	(10.5)	(11.7)	(13.9)	(10.5)	(11.4)	(13.9)	(21.2)
Employment share: Services without commerce	39.1	43.8	48.9	49.4	42.5	46.7	46.5	45.1
	(2.4)	(10.1)	(10.5)	(12.2)	(8.9)	(10.3)	(13.8)	(18.5)
Informal employment share	56.3	61.0	65.1	70.0	70.1	73.4	78.8	80.0
	(11.6)	(15.6)	(17.6)	(21.1)	(17.1)	(20.5)	(22.2)	(36.5)
Formal employment share	44.8	47.1	52.0	48.6	34.8	37.7	37.3	33.6
	(12.5)	(13.6)	(16.1)	(17.9)	(14.5)	(13.7)	(15.0)	(13.2)
Municipalities	5	28	85	156	186	163	106	35
Average 2010 Employed Population	35,923	21,407	26,963	21,058	18,750	18,821	12,974	10,917

Note: This table reports the mean of selected labor market indicators using the municipal panel of the GEIH by year and treatment status. The employment, unemployment, and inactivity rates are constructed by dividing the number of employed, unemployed, and inactive individuals in a municipality over the 2010 municipal working-age population. For shares (by informality status or economic activity), we divided the number of workers in the sector by the total municipal employment in 2010. We use this fixed aggregate to weigh the mean and standard deviation. Standard deviations are reported in parenthesis.

3.2 Pay-premiums of Hard Discounters

To better understand how hard discount chains differ from other formal retail firms in their wage-setting, we estimate the canonical model of Abowd et al. (1999) (hereafter AKM) to isolate firm-specific contributions to formal wages. With the output from this estimation, we answer how much HDS pay, on average, their workers relative to other firms once worker characteristics are netted out. This is relevant as recent evidence from developing countries (Bassier, 2023; Delgado-Prieto, 2024b; Gerard et al., 2021; Pérez Pérez and Nuño-Ledesma, 2024) suggest the role of firm pay policies in explaining wage inequality may be larger than in developed countries. We estimate our AKM model using the standard log-linear specification:

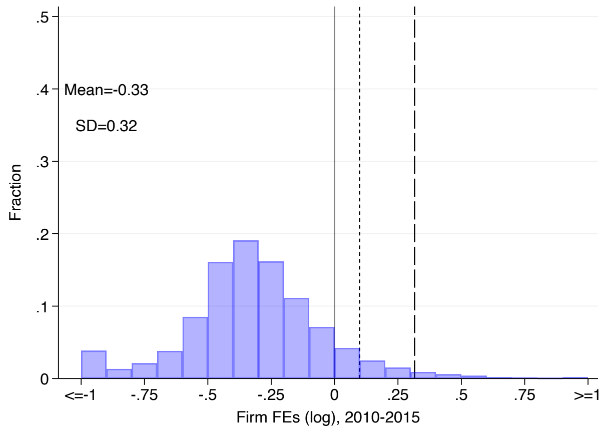
$$log(wages_{it}) = \alpha_i + \psi_{j(i,t)} + X_{it}'\xi + \varepsilon_{it} \tag{1}$$

Here, $log(wages_{it})$ are log monthly wages. They are a function of an additive linear combination of unobserved worker Fixed Effects (FEs) α_i and unobserved firm FEs ψ_j. To identify the latter component, we restrict to the largest set of workers and firms connected by workers' mobility, eliminating around 3% of the initial sample of workers aged 20 to 60. The $j(i,t)$ refers to the firm j of worker i in period t, and the vector X_{it} are time-varying controls, which are age squared and its cubic (after a normalization in age 40) and year FEs. We split the estimation period in two: from 2010 to 2015, when the first HDS opened and employed mainly skilled workers, and from 2015 to 2019, when they grew extensively across the country and employed more intensively blue-collar workers. In this exercise, all the firm FEs are relative to the largest retail store in the country, so on average they are negative.

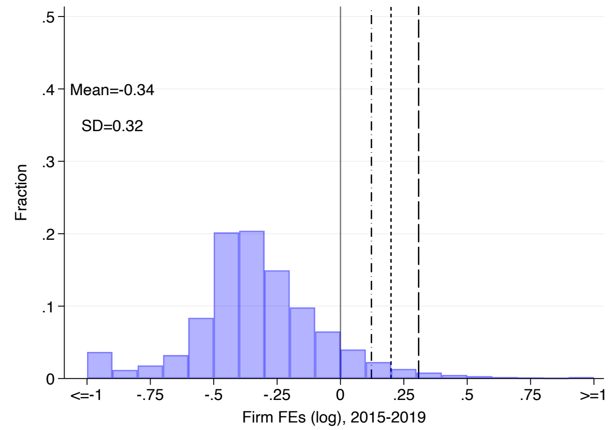
Figure 3 shows that the dashed lines, representing the three hard discount chains, pay consistently higher premiums to all their workers than the country's largest retail store. They are located at the higher end of the distribution of pay premiums, indicating that, once worker characteristics are netted out, these firms contribute positively to all their workers' wages. In both estimating periods, we find this positive contribution of hard discount chains. This in line with evidence from Alfaro-Ureña et al. (2021) that shows multinationals have a positive wage premium on Costa Rican workers. To read more intuitively the figure, when a worker moves from the reference firm at zero to one of the discount chains, it would experience an average positive gain in log wages approximately equal to the value of the dashed line. Lastly, note that the firm's pay premiums are measured at the national level for all workers, including managers and blue-collar workers.[14]

[14] We do not use the leave-out method proposed by Kline et al. (2020) for the estimation as it yields an unbiased variance and covariance moments of the wage decomposition, not the vector of estimated level parameters shown in Figure 3.

Figure 3: **Distribution of Firm FEs**



(a) 2010–2015



(b) 2015–2019

Note: The dashed lines are the HDS chains (from 2010 to 2015, there are only two), while the retail firm used as reference is located at the zero line. For confidentiality reasons, we do not disclose which line belongs to which HDS. For the estimation sample, we eliminate workers with non-positive wages, with less than 30 employment days per month, restrict employees between 20 and 60 years, and leave the highest wage job for workers with more than one contribution to the social security system. Moreover, we eliminate workers and firms that do not belong to the largest connected set of firms and workers and workers that appear only once in the estimation sample. We transform the nominal wages to real terms using the monthly CPI from DANE (with the base year 2018). Source: PILA August 2010–August 2019.

4 Empirical Specification: Cohort Analysis

For our identification strategy, we exploit the staggered rollout of HDS in Colombia to quantify its effects on local labor markets, similarly to the empirical strategy of Chava et al. (2023) and Cunningham (2024) studying the local impacts of Amazon FCs. Table 2 illustrates the staggered introduction of hard discount chains within our sample of municipalities, comprising 372 in PILA and 191 in GEIH. There are no municipalities where the first hard discount store that opens subsequently closes, so all municipalities in our sample remain treated over time. Although the number of municipalities increases with time (showing the large expansion of these chains throughout the decade), around 40% of them had a discount chain by the end of 2016.

Table 2: **Cohorts of Treatment**

	Municipalities (Admin)	Share	Municipalities (Survey)	Share
2011	10	2.7	5	2.6
2012	24	6.5	14	7.3
2013	19	5.1	9	4.7
2014	19	5.1	11	5.8
2015	23	6.2	12	6.3
2016	56	15.1	34	17.8
2017	76	20.4	45	23.6
2018	64	17.2	26	13.6
2019	81	21.8	35	18.3
Total	372	100.0	191	100.0

Note: In this Table, we exclude the municipalities treated in 2009, 2010, and 2020 due to the small number of treated units. Moreover, we dropped 24 capital cities in PILA and 23 in GEIH, plus an additional city in GEIH that did not appear in all survey years.

Using the canonical Two-Way Fixed Effects (TWFE) regression to estimate the Average Treatment Effect on the Treated (ATT) in these settings is common. Yet, recent literature shows that such estimation strategy is potentially biased as treatment effects may be heterogeneous among treated cohorts (Goodman-Bacon, 2021; Sun and Abraham, 2021). For instance, the treatment effects on the earlier cohorts might be distinct from the ones of the later cohorts, but the TWFE regression aggregates these into one single parameter using weights that can be hard to interpret and/or incorrect, which leads to biased coefficients. Therefore, we use the event-study specification of Callaway and Sant'Anna (2021) (C&S, hereafter) as the main estimator since it tackles the main issues regarding the differential timing of treatment and the heterogeneity of labor market effects across cohorts.

The intuition behind the C&S estimator is that it allows us to break down all the sets of possible comparisons into the two-period and two-group (2x2) framework to estimate multiple ATT(g,t) for units treated in the same year (cohort g) measured in period t. In our setup, cohorts refer to the first year that a hard discount store opens in a municipality l. Therefore, if the treatment started in 2015, then $g = 2015$. Due to the small number of treated units at the beginning of the treatment, we restrict the leads before $t - g < -5$ and the lags after $t - g > 5$. More concretely, they are estimated as follows:

$$ATT(g,t) = E(y_{l,t} - y_{l,g-1} \mid G_l = g) - E(y_{l,t} - y_{l,g-1} \mid G_l > t),\ for\ all\ t \geq g. \tag{2}$$

Here, we use as a control the not yet treated group ($G_l > t$) for a cleaner comparison with treated municipalities and estimate the differences relative to the baseline period ($g-1$) using ordinary least squares without covariates, testing its robustness with different controls. Then we aggregate all differences into an overall ATT using weights $w_{g,t}$ that are based on the number of treated units used in the particular ATT(g,t)[15]:

$$ATT_{post} = \sum_{t}^{T} \sum_{g}^{G} 1\{t \geq g\} w_{g,t} ATT(g,t). \tag{3}$$

Next, as a benchmark for the C&S coefficients in the event study figures, we estimate the traditional TWFE regression in certain outcomes.[16] The model takes the following form:

$$y_{lt} = \gamma_l + \gamma_t + \sum_{k=-5, k\neq -1}^{5} \beta_k * \mathbf{1}\{k = t - g\} + \varepsilon_{lt}. \tag{4}$$

Here, the year that the first store opens equals g, and the years are denoted as t, so the relative event time indicators are k. We add time fixed effects (γ_t) and municipality fixed effects (γ_l) to the specification to control for unobserved constant characteristics over time for all units and in all years for each unit, respectively. The parameters of interest are β_k, which come from k event time dummy variables. These dynamic treatment effects measure the effect on y relative to an omitted period, which is when $k = -1$. In this specification, identification arises from two control groups: units not yet treated or units never treated (Schmidheiny and Siegloch, 2023). Again, we select municipalities that have not yet been treated as a control group.[17]

Before discussing the identifying assumptions, it is helpful to describe the functional form of

[15] For the overall ATT, we use eight cohorts $G \in \{2011, ..., 2018\}$ of treatment in the administrative and survey data.

[16] To compare the pre-treatment coefficients of TWFE and C&S accurately, we quantify long gaps when using C&S.

[17] To achieve identification using not yet treated as a control group, we bin or accumulate the leads as $k \leq -5$ and the lags as $k \geq 5$ (Borusyak et al., 2024).

the main outcomes. First, using PILA we quantify average real monthly wages or labor income in each l, after restricting to workers who are working full-time and with 30 days of employment, and apply the standard log transformation. From GEIH, we also quantify average real monthly wages or labor incomes for positive earners, plus average weekly working hours, across the informal and formal sectors in each l and apply the log transformation. For the employment outcomes, we define them in employment-to-population ratios similar to other influential papers, such as Autor et al. (2013) and Dustmann et al. (2017), instead of logs given the considerable variation in the distribution of treated and not-yet-treated outcomes in the baseline year, which can lead to biases in the coefficients and even to the wrong sign (McConnell, 2023). Thus, we define a common denominator across our datasets for comparability, which is the working-age population in each l obtained from the 2005 population census, and then use in the numerator the number of individuals in l either from PILA, SISBEN or GEIH. In this way, we measure the relative growth rates among treated and control municipalities of employment, that avoid the approximation biases of growth rates using large log differences. We also use employment in levels as an outcome to measure the number of formal jobs created and in logs as a robustness check and its results are consistent with the outcome in ratios. For the tax outcomes, we define them in ratios, using the pre-treatment measure of public revenues in the denominator and in the numerator the specific type of revenue that varies over time.

Identification Assumptions. The main assumption required in this setup is the unconditional parallel trends assumption (PTA). It establishes that treated and control units would have evolved similarly in terms of their outcomes in the absence of store openings. As our control group is the municipalities that have not yet been treated, we do not assert that store openings are exogenous to local economic trends. Instead, we argue that unobserved trends do not determine the *timing* of store openings. The rapid roll-out of HDS across the country makes it less likely that local economic trends among the early and late-treated cohorts determine timing. Specifically, the *timing* of openings is mostly related to factors such as population levels (larger cities tend to be prioritized) or the regional location of the municipality (see Appendix Table C.2). Our specification already absorbs any constant observed and unobserved characteristics associated with these factors. As a placebo test, we show there are no differential employment, wage or tax trends before the arrival of these stores across different sectors and industries. This evidence supports our *timing* assumption for identifying the effects of HDS openings in local labor markets.

A potential threat to our identification is a time-varying shock affecting the outcomes that is correlated with the *timing* of entry of discounters, potentially confounding our estimates. For instance, the entry of discount stores might coincide with the entry of other formal firms in the municipality. Appendix Table D.2 shows no significant firm growth in the pre- and post-treatment periods, suggesting that this is not a major threat. Another potential concern is that hard discount chains tend to open first in places closer and better connected to capital cities. Consequently, early treated cohorts could be more affected by spillover trends from capital cities than later treated cohorts. To address this concern, we adjust our empirical strategy to account for baseline controls, resulting in a conditional parallel trends assumption, where treated and not-yet-treated municipalities within similar values of control variables would have evolved in a parallel way in the absence of the treatment. With this weaker identifying assumption, we are comparing treatment cohorts of municipalities with similar baseline distances to the largest cities of the country, Medellín and Bogotá, and with similar development stages (measured with the rurality index) to show that results are robust and do not change significantly, albeit certain outcomes become more imprecise (this is expected as the size of certain cohorts is not large). For this, we use the outcome regression adjustment in the *C&S* framework. Lastly, we assume there are no anticipatory effects in response to store openings. Given that the shock is relatively unexpected and its effects take years to materialize, this is less of a concern.

Another aspect to discuss is the intensive margin of the treatment across municipalities. If cohorts are heterogeneous in the number of stores they have, interpreting the overall ATT becomes less straightforward. Appendix Figure B.2a presents evidence that cohorts receive a similar number of stores over time. In all cohorts, half of the municipalities have two discount stores three years from the arrival of discount chains. The top 20% of municipalities, in terms of the number of stores received, have four, while those that receive the least number of stores have one. When differentiating by cohort, the average number remains similar across cohorts. The only notable exception is the cohort of 2016, which saw the arrival of *Justo & Bueno*. Even in this case, the difference is about one store more on average (see Appendix Figure B.2b).

Lastly, we discuss the choice of using the not-yet-treated municipalities as our main control group instead of the never treated (that is, municipalities that had not received a hard discount chain as of 2020). We argue that the never treated group is not a suitable control given the substantial differences between these municipalities and those where hard discount chains decided to establish stores. As shown in Appendix Table C.2, municipalities that received at least one HDS were more populous, less rural, and closer to the main municipalities of Bogotá and Medellín. Furthermore, we constructed a propensity score using a logit regression that predicted the likelihood of being treated based on these pre-treatment characteristics to find that the distributions are highly skewed, highlighting the stark dissimilarity between the two set of municipalities (see Appendix Figure C.3).

Apart from the cross-sectional differences in pre-treatment characteristics, the labor markets of never-treated municipalities exhibits different trends compared to those of treated areas. Panel B of Appendix Table D.1 shows that when we use the never-treated group as a control, significant pre-treatment trends emerge in employment, unemployment, and inactivity rates. In contrast, comparing treated and not-yet-treated municipalities yields pre-treatment coefficients that are statistically indistinguishable from zero for all these outcomes. Moreover, in Panel C, we restrict the sample of never-treated municipalities to those similar to the treated ones by employing a one-to-one propensity score matching model without replacement. Even with the restricted never-treated group, the pre-trends persist, although they are smaller in magnitude and insignificant. Together, these findings indicate that never treated municipalities differ in unobserved trends to the treated municipalities and thus are not an appropriate control group for our analysis.

5 Effects on Local Employment, Taxes and Labor Income

5.1 Employment

We evaluate the overall effect of HDS on local labor market outcomes. If a store opening induces a positive labor demand shock in the formal sector, then local employment should increase accordingly. Appendix Figure C.4 illustrates the dynamic treatment effects on the local employment rate. Consistent with this hypothesis, Appendix Figure C.4 shows that the post-treatment estimates are positive and grow over time, with most of them becoming significant in the later periods. Averaging the six post-treatment years, HDS openings boosted the local employment rate by approximately 2.3 pp (see Appendix Figure C.4). Additionally, Appendix Table C.6 shows a significant reduction in the inactivity and unemployment rate following store openings.

Importantly, treated and not-yet-treated cohorts of municipalities do not exhibit differing trends before the treatment, suggesting that the arrival of HDS is unrelated to local employment trends among these groups. It is worth noting that the TWFE estimates of the employment rate differ to the C&S ones, suggesting that heterogeneous treatment effects might introduce biases.

The general impact on the labor market masks heterogeneous responses from specific employment types that hard discount chains primarily create, such as formal employment. Therefore, we next analyze the evolution of formal employment in local labor markets following store openings. For this analysis, we primarily use administrative data and complement our findings with survey data. Still, their coefficients are not entirely comparable even if we use the same denominator for the formal employment outcomes in both datasets, since the sample of municipalities in PILA (372) is twice as large as the one in GEIH (191).

The main figure of the paper is Figure 4. It shows that formal employment follows a strongly positive trend after the treatment according to both administrative and survey data, and it takes three years to materialize. Thus, the observed rise in total employment is primarily driven by the increase in formal employment. Averaging all six post-treatment periods, the ratio of formal employment-to-population increases by around 1.7 pp using administrative records and 2.9 pp using survey data. To benchmark the estimates, the weighted mean of the formal employment-to-population ratio before the treatment is 16% in the administrative data and 28.3% in the survey. Hence, the increase in relative terms is equivalent to 10.6% and 10.3%, respectively, which is fairly similar across datasets.[18] These results remain robust when controlling for baseline covariates (such as distance to main capital cities and a proxy for economic development) as well as to violations of parallel pre-trends, and they are also statistically significant under randomization of treatment timing. Hence, as we argue the *timing* of opening a hard discount store is unrelated to the local employment trends, this indicates a positive causal impact on the growth in formal employment in the treated municipalities.

Because the sample of treated and control municipalities is unbalanced over event time, we further decompose these coefficients by cohorts of treatment. For that, we use the properties of the *C&S* estimator that can flexibly aggregate the effects into different components, and aggregate by cohort of treatment rather than by event time. Appendix Table C.7 shows that the effect is concentrated in the municipalities that opened before 2016. This is expected, as these municipalities can be observed for at least three years, yet this distinction is crucial for extrapolating the effects, indicating that not all future-treated municipalities might experience the same effects as the ones initially treated. Altogether, discount stores help to formalize local economies and generate employment growth, partly by reducing the local inactivity and unemployment rate (i.e., the number of individuals not working) as shown before.

These coefficients measure the overall impact from the first opening, thus we also examine the number of stores in each municipality to quantify an intensive margin effect. We use the number of HDS in the municipality as the outcome in our main specification and find that, averaging all post-treatment years after the first opening, there are around 2.4 HDS. From this, we calculate a per-store impact: each additional store increases the formal employment ratio by 0.7 pp (=1.7/2.4) or 4.4% using administrative data and by 1.2 pp (=2.9/2.4) or 4.2% based on survey data. In absolute terms,

[18] Moreover, from the baseline event to the farthest post-treatment event we measured, the weighted mean of the formal employment ratio grew around 13.8 pp in the labor force survey (from 28.3% to 42.1%). So, the entry of HDS would explain around one-fifth of the observed increase in formality.

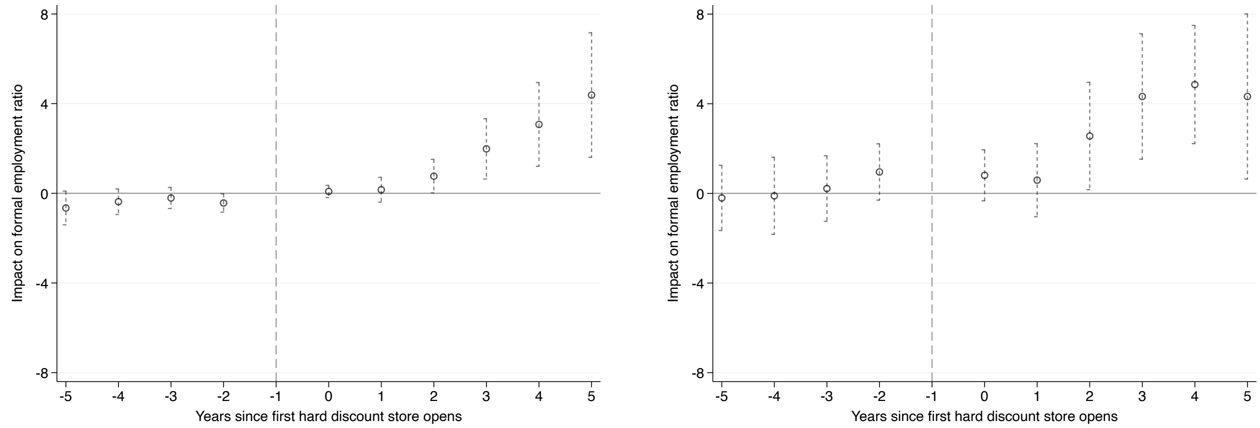
each store generates around 138 direct and indirect formal jobs (95% $CI = [26, 250]$).[19]

The formal employment multipliers we find are large compared to those estimated for the US, where one additional manufacturing job generates about 1.6 jobs in the non-tradable sector (Moretti, 2010). However, given that our sample consists of relatively smaller geographic units averaging fewer than 7,000 formal workers (see Appendix Table C.1), which are much smaller than US metropolitan statistical areas, the multipliers are expected to be higher partly due to a more elastic local labor supply. We explore heterogeneity by dividing municipalities based on a proxy for baseline economic development (whether the labor force survey covers the municipality) and find that the estimated impacts are more pronounced in less developed municipalities without survey coverage (see Appendix Figure C.6).[20] This pattern is consistent with local multipliers being stronger in smaller, less populated areas and with higher levels of informality, that can have a more elastic local labor supply.

[19]For the effect on the number of jobs, we use the total number of full-time formal workers with 30 days of employment in the PILA as an outcome and control for the working-age population of the 2005 census. Then, we divide the resulting coefficient by the post-treatment average number of stores (2.4).

[20]Municipalities covered with the survey are significantly larger in terms of population, have a higher share of formality, and higher labor income.

Figure 4: **Event study estimates on formal employment**



(a) Formal employment (Admin) **(b) Formal employment (Survey)**

Note: We use the *C&S* estimator. The dependent variable in (a) is formal employment using PILA, and in (b) is formal employment using GEIH both over the working-age population according to the 2005 census. Regressions were weighted with the local working-age population in 2005. Observed treated municipalities in PILA are 372, and in GEIH are 191. Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: PILA 2010-2018 in August, and GEIH 2010-2018.

We then examine the sectors driving the formal employment growth. Figures 5a and 5b indicate that the primary and secondary sectors account for most of this growth, according to both administrative and survey data (see Appendix Tables C.8 and C.10 for detailed point-estimates). Focusing on survey data, which provides a more accurate industry classification, Appendix Table C.11 shows that manufacturing employment-to-population ratio rises by 0.9 pp, the agricultural ratio by 0.6 pp, and the construction ratio by 0.4 pp, explaining around 65% of the overall increase (see Appendix Table C.9 for administrative data results). We also find a significant increase in the retail employment-to-population ratio, based on administrative and survey data. This indicates that the direct demand for formal retail jobs from HDS outweighs any potential job losses in incumbent retail firms, such as supermarkets, from the increased competition within the formal retail sector.[21]

[21] Ideally, we would want to separate direct and indirect effects, the former being how many jobs are generated directly by HDS. However, when reporting their payroll on the PILA, discount chains often locate employees in the nearest capital city (that is, instead of reporting payroll in the municipality where the store is located, they often do it in the nearest large urban area). We find that overall, less than 20% of municipalities with hard discount chains report formal employment.

There are plausible explanations for these inter-sectoral spillovers from retail that complement each other. First, hard discount chains have incentives to build local supplier networks for some of their intermediate inputs, as transportation costs in Colombia are one of the highest in the world (Londoño-Kent, 2009), potentially increasing formal employment among these complementary local industries. Unfortunately, we cannot directly link how much hard discounters purchase from local producers because we lack input-output data.[22] Still, the largest hard discount chain reported that around 80% of their goods in 2020 were produced in Colombia (La República, 2022), and according to a discount chain manager, many suppliers have been expanding to meet the increased demand from these stores (Portafolio, 2018). Second, they could create agglomeration effects around HDS that benefit nearby businesses (Evensen et al., 2023). Third, the large-scale investments made by hard discount chains may indirectly support other local sectors, such as construction. Fourth, the additional earnings of labor market entrants and lower expenses on groceries after HDS arrive for consumers, can further stimulate the local economy via demand multiplier effects (Gerard et al., 2024).

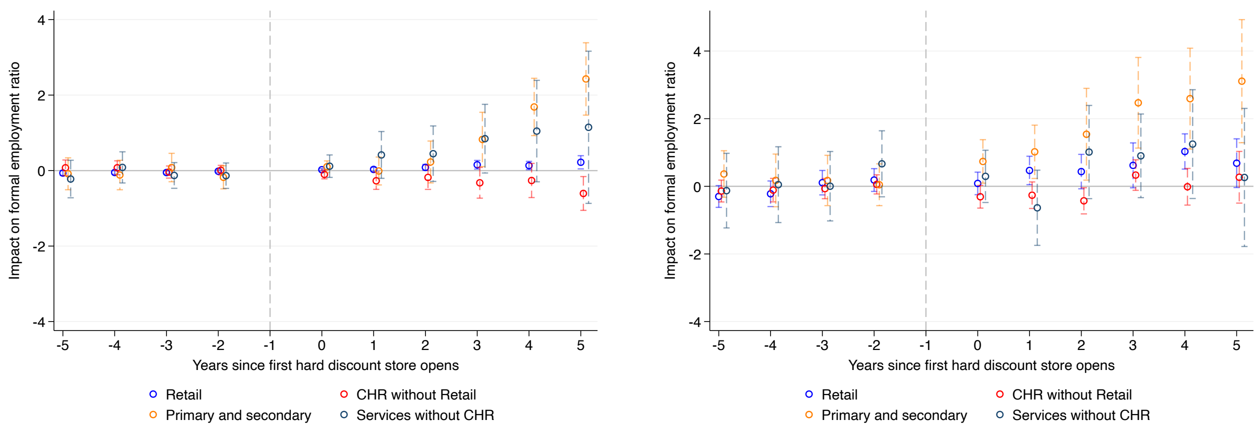
Identifying the main channel driving our results is challenging, as the mechanisms are likely interlinked. However, the sectoral evidence points to upstream supply chain effects and local investment as the most plausible explanations. Two sets of findings support this interpretation. First, if multiplier effects or agglomeration forces were the dominant mechanisms, we would expect to see broader benefits across sectors, including the informal sector, which accounts for a large share of employment in these municipalities. Yet, as shown later, we do not observe such positive effects in the informal sector, at least not in terms of employment or wages. Second, while we find a positive and statistically significant employment effect in construction (consistent with the investment channel) we observe even larger effects in manufacturing, suggesting that supply chain linkages may be more influential in explaining the overall impact of HDS.

Additional support for this interpretation comes from Atkin et al. (2018), who find no local

[22]These findings are consistent with de Paula and Scheinkman (2010), who discusses the "business formality chain" that emerges when increased enforcement at the firm's production stage-combined with the way value-added taxes are collected in Brazil-incentivizes formality both upstream and downstream in their supply chain.

employment effects from the entry of foreign retailers in Mexico, despite these firms being considerably larger than HDS. Crucially, those retailers import most of their products from the US or Europe, which limits their integration into local supply chains. In contrast, the positive employment effects we document are likely amplified by discount chains stronger reliance on domestic suppliers.

Figure 5: **Event study estimates on formal employment ratios by industry**



(a) Formal employment (Admin) **(b) Formal employment (Survey)**

Note: We use the *C&S* estimator. The dependent variable in (a) is formal employment in each sector using PILA, and in (b) is formal employment in each sector using GEIH both over the working-age population according to the 2005 census. The CHR refers to commerce, hotels, and restaurants. Primary and secondary are manufacturing, construction, agriculture, and mining. Regressions were weighted with the local working-age population in 2005. Observed treated municipalities in PILA are 372, and in GEIH are 191. Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: PILA 2010-2018 in August, and GEIH 2010-2018.

Given that the main competitors of HDS are local neighborhood shops, which are primarily informal, the next outcome we analyze is informal employment. For this, we rely on survey data from the GEIH, as the informal sector is only observed in the survey. Appendix Table D.5 shows that the impact on informal employment-to-population ratio is not robust, as the coefficient ranges from negative to positive values when we include baseline covariates. Therefore, the impact on informal employment is imprecise. Regarding sectoral impacts, the increase in competition between HDS and the informal retail sector does not affect their employment growth. In fact, the point estimate is positive across different specifications (see Appendix Table D.5). Lastly, the coefficients on informal employment across other sectors vary substantially but are too noisy to describe a clear pattern (see Appendix Table C.12).

A limitation of using the GEIH survey is that it does not cover all the municipalities where HDS operate. To address this, we leverage Colombia's social security system to get complementary measures of the population outside of the formal sector. In Colombia, universal access to the health system is granted, either by subsidies or contributions. Formal workers contribute to the system, while those who are subsidized are either informally employed, unemployed or not in the labor force. Using the beneficiaries of subsidized social protection as a proxy for population outside the formal sector, we show in Appendix Figure C.7 that in the first four years of the treatment, the effect is close to zero. However, in later years, we observe a negative trend consistent with the observed increase in formality. Overall, the main impact of HDS is a substantial shift in the workforce composition, with formality driving the increase in employment.

5.2 Taxes

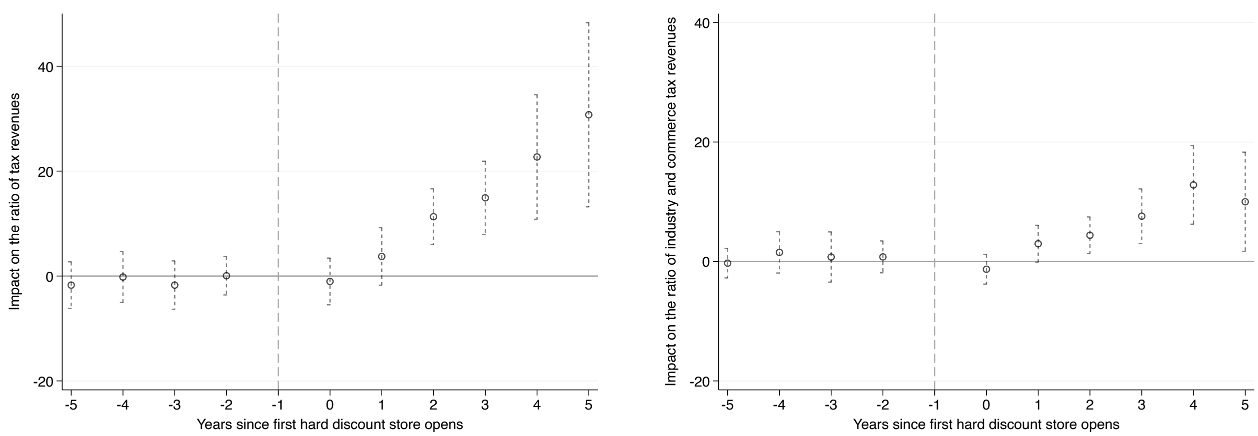
The arrival of hard discount stores boosts local economic activity through direct job creation but also through spillover effects to other sectors or by agglomeration effects in areas where the stores are located. We use municipality-level tax revenues to explore whether HDS entry has broader impacts on local tax revenues. Although the central taxes in Colombia are collected nationally, such as the income tax, social security contributions, and VAT, other significant taxes are collected at the municipality level. These include property, industry, commerce, and vehicle and gasoline taxes.[23] Industry and commerce taxes, in particular, are highly correlated to the performance of industrial, commercial, or service activities registered in the municipality. Therefore, even though HDS chains operate nationwide, they are required to pay industry and commerce taxes based on local sales and revenues in each municipality where they have stores.

Figure 6 shows before the arrival of HDS, the growth of taxes-to-baseline public revenues was

[23] The owners of low-income residential households are not required to pay property tax.

similar between the early and late treated municipalities.[24] After the entry of discount stores, the ratio of local tax revenues increases by 10.1 pp on average, or 7.5% relative to the pre-treatment mean, for all post-treatment periods.[25] Most of this increase is attributed to revenues collected from the industry and commerce taxes (see Appendix Table C.13 for detailed point-estimates). Hence, the arrival of HDS substantially boosts tax revenues for local governments, primarily driven by the growth in manufacturing and commercial activities. Increasing local revenues could trigger local public spending, which can stimulate, in turn, the labor market, indicating another mechanism where HDS could promote local formal employment. These results align with the previous findings by sectors and with the literature on supply chain effects of business formality (de Paula and Scheinkman, 2010; Gerard et al., 2023; Rios and Setharam, 2018), indicating that HDS purveyors are likely to be formal if they want to claim all tax benefits and discounts provided by the law.

Figure 6: **Event study estimates on the ratio of taxes by type**



(a) Revenues of all taxes **(b) Revenues of industry and commerce taxes**

Note: We use the *C&S* estimator. The dependent variable is the revenue by each specific type of local tax over all the revenues (taxes and central government transfers) . We only included taxes collected at the municipality level, such as property taxes, industry and commerce, gasoline taxes, vehicle taxes, and other local taxes, such as the rights to post ads on public streets. Regressions were weighted using municipality's revenues in 2010. Observed treated municipalities are 371. Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: DNP, 2010-2018.

[24] A potential concern is that local policymakers might reduce municipal taxes to attract discounters and other investors. However, tax rates remained largely stable during our study period (2010–2018).

[25] We use pre-shock local tax revenues as regression weights, rather than population, to capture differences in the size and capacity of municipal tax authorities.

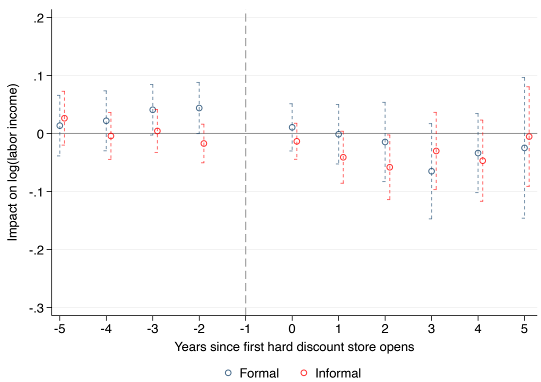
5.3 Labor Income and Working Hours

Other margins of adjustment from employers, such as on labor income or working hours, may occur in response to the increase in labor demand or labor market competition from these stores. One limitation with our measure of these outcomes is compositional changes over time, as we only observe local averages of all workers. To start, we focus on the impacts on labor income across the formal and informal sectors. On the one hand, Figure 7a shows a slight negative effect in the first post-treatment years in the labor income of informal workers, while the impact on the labor income of formal workers is rather stable. Moreover, we find with the administrative data that formal wages experience an effect close to zero, that turns negative in the last post-treatment period (see Appendix Figure C.8). As shown with the AKM model, hard discount firms indeed pay their workers a premium that should increase average formal wages, but as primary and secondary sectors drive the main increase in formal employment, it is likely they hire minimum wage workers, which in turn decreases average formal wages. Another potential explanation is that the local labor supply is infinitely elastic and this mutes any formal wage effect (Moretti, 2010).

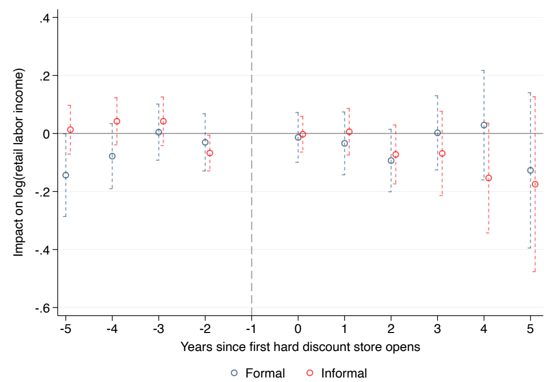
Next, we measure the impact on the labor income of retail workers by sector. For informal workers, this serves as an indirect measure of the earnings of neighborhood shop owners, while for formal workers, it reflects regular wages. Figure 7b shows an imprecise negative growth in the labor income of informal workers in the latest post-treatment periods, indicating that the increase in competition from HDS may affect them mostly through reduced earnings rather than employment losses. Due to the nature of their labor contract, informal shop owners do not pay mandatory minimum wages to their workers, thereby having the flexibility to adjust to demand shocks. While the estimates are insignificant and imprecise, they are notably large ($ATT_{post} = -7.8\%$). For formal retail workers, we do not find an effect, which could contradict the AKM findings, yet the sample differs (e.g., the AKM sample includes capital cities), and HDS paying a national wage premium does not necessarily increase average wages at the sector level in treated municipalities.

Last, we focus on the average weekly working hours across the informal and formal sectors as it is another possible margin of adjustment from employers. For instance, discount stores may have longer opening hours to capture more customers, so other businesses might change their operating hours, possibly affecting the working hours of their workers. Figure 7c shows that there are no significant changes in these outcomes after the first entry. Thus, it is less likely that informal or formal workers adjust to the arrival of HDS by working more hours.

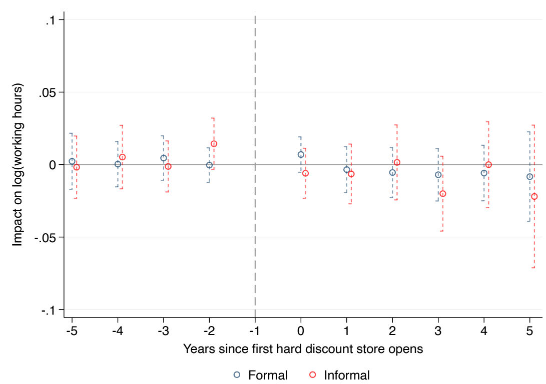
Figure 7: **Event study estimates on labor income and working hours**



(a) Labor income by sector



(b) Retail labor income by sector



(c) Working hours by sector

Note: We use the *C&S* estimator. The dependent variable is the logarithm of formal and informal labor income and working hours. Regressions were weighted with the local working-age population in the 2005 census. Observed treated municipalities are 191. Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage changes. Source: GEIH 2010-2018.

6 Robustness Checks

We provide additional robustness to our main findings. First, we relax the assumption of parallel pre-trends and perform sensitivity analysis on the average coefficients following Rambachan and Roth (2023). A potential concern in this setting is that there may be unobserved time-varying shocks at the municipal level that would have affected differently early treated municipalities relative to later ones, even in the absence of HDS. So, if we assume that these shocks are of similar magnitude before and after the treatment, then we can bound the estimates on relative magnitudes with respect to the maximal violation in pre-trends. In that sense, we can obtain breakdown values from which the coefficient is no longer significant. In this exercise, we find that the average post-treatment coefficient of formal employment is robust up to between 0.3 to 0.4 maximal violations of pre-trends (see Appendix Figure D.1). The value does not reach violations close to 1, partly due to the number of periods we analyze (six periods), which can increase the confidence intervals and decrease the breakdown values (Rambachan and Roth, 2023).

Second, we conduct Fisher randomization tests (FRT), which rely on permuting the timing of treatment, that is, the year HDS enters a municipality. Since our baseline estimator computes p-values using standard errors clustered at the municipal level, we follow the procedure proposed by Roth and Sant'Anna (2023), which is designed for staggered adoption settings where treatment timing is as-good-as-random and is compatible with the C&S estimator. Table D.3 presents the FRT based p-values for the event study estimates of our two main formal employment outcomes: total formal employment (columns 1–3) and formal employment in the primary and secondary sectors (columns 4–6). The treatment effect estimates correspond to panel B of Table D.8, as the Roth and Sant'Anna (2023) method does not support the inclusion of regression weights. For each outcome, we report three p-values: the standard z-score based value used in C&S (relying on clustered standard errors), and two FRT based p-values (one using covariate adjusted standard errors and the other using Neyman standard errors). Across all specifications, the p-values are closely similar, and the statistical significance of the results remains unchanged when we randomize treatment timing.

Third, we test the robustness of our results to including baseline covariates. The models of selection into treatment that we discussed in Section 2 showed that population, driving time to Bogotá and Medellín, and the share of rurality are strongly correlated both with the overall probability of receiving a store both over our sample and in a given year. Our baseline estimates already take into account the population as we weight them using the working-age population in the 2005 census. Appendix Figures D.2 to D.4 report the estimates when we include the other three covariates.[26] While we report figures for the main outcomes (formal employment and taxes), we show in tables the results for the rest (informal employment, wages, and labor market indicators). Figures display the robustness to accounting for each specific covariate, while the Tables compare estimates without covariates with estimates with the full battery of them.

Most of the main results remain robust to the inclusion of covariates. In the case of formal employment based on administrative data, Figure D.2a shows the estimates remain positive and statistically significant, especially in the later years. Regarding the primary and secondary sectors, we still find an upward trend but estimates become more imprecise (see Figure D.2b). Relatedly, Figure D.3b shows that when measuring formal employment in these sectors with survey data, the estimates are robust to controls, suggesting that these stores continue to have a positive impact even when accounting for covariates, and relying on a weaker identifying assumption.

Other labor market outcomes do not show substantial changes when adjusting for covariates, with the exception of formal employment in the survey. Figure D.3a shows that when adding controls the post-treatment estimates reduce their magnitude and become more imprecise. However, they remain positive and with an upward trend over time.[27] Regarding the employment, unemployment, and inactivity rates, Table D.4 shows that the estimated impacts on these outcomes become larger and more significant. Finally, the effects on informal employment (Table D.5), labor income

[26]We opt for incorporating these covariates using the outcome regression DiD estimator based on ordinary least squares instead of other estimators such as the doubly robust. The other estimators are based on propensity-score methods that rely on the common support assumption. This assumption is not likely to hold in our case, as shown in Appendix Figure for all municipalities C.3.

[27]For some periods (second and fourth years after the arrival) the coefficients are less precise but their magnitude does not change, and for others (the third and the fifth years) the magnitudes decrease only when controlling for the distance to Bogotá and Medellín.

and working hours (Table D.6) remain non-significant.

Regarding the main tax outcomes, Figures D.4a and D.4b display how the dynamic impact estimates differ when including one control at a time and when including all of them. Columns 2 and 4 of Table D.7 report the estimated impacts when all covariates are taken into account. Both dependent variables appear robust to including these controls, with the coefficients increasing in magnitude. Altogether, results are robust to controlling for pre-treatment differences in distances to the main cities and rurality index between early- vs late-treated municipalities.

Fourth, another potential concern stems from the concentration of the effects in metropolitan areas, where the expansion of hard discount chains is easier due to their proximity to capital cities. We have excluded capital cities from our estimation sample to partially address this. However, the estimates can still be biased by those municipalities that are more closely linked to capitals in ways not captured by the controls of distance. To mitigate this concern, we conducted additional analysis excluding the six largest metropolitan areas.[28] This further exclusion drops 19 municipalities from our main sample. Panel B of Appendix Table D.11 reports that the treatment effects do not vary substantially in magnitude.

Fifth, an additional confounding factor may be migration, both internal and from Venezuela. This is unlikely to drive our results, as population movements would need to be systematically correlated with the timing of discounter's entry. Nonetheless, we assess whether the Venezuelan exodus to Colombia affects our findings. Note that we already exclude capital cities from the analysis (where most migrants settle and where internal migration is most concentrated) and focus instead on intermediate cities. Still, we directly account for the immigration shock, using data from the 2018 census. We define this as the ratio of employed Venezuelan migrants in a municipality to the total employed population aged 18 to 64. Panel C of Table D.11 presents the estimates, which remain similar in magnitude and statistical significance to those in the specification without controls (Panel A).[29]

[28]The classification of metropolitan areas is performed by DANE.

[29]Besides, internal migration patterns remained stable during our study period. According to Acosta and Gu (2024), five-year migration flows from 2012 to 2019, based on household surveys, consistently show Bogotá as the primary destination. This pattern holds across different definitions (one-year vs. five-year flows) and does not vary significantly

6.1 Regression Weights and Log Transformation

Our preferred estimation use population weights to account for potential market size and to focus on the effect on the average individual, rather than on the average region. It also uses outcomes in ratios to omit potential concerns due to differences in the pre-treatment outcome distributions between treated and not-yet-treated municipalities. In this subsection, we show that our main results remain largely unchanged when we opt for unweighted models or use log transformation of the outcomes. Panel B of Table D.8 in the Appendix reports the estimated impacts on formal employment using administrative data with an unweighted specification. Compared to the weighted estimates in Panel A, the unweighed for total formal employment remain positive and statistically significant, with similar results for employment in the primary and secondary sectors. On the other hand, the estimated impacts on retail employment, as well as in hotels and restaurants, gain statistical significance. Finally, the inconclusive impacts on service employment continue.

Regarding the log transformation, Table D.9 compares the estimates on total formal employment using controls when the dependent variable is the ratio (Column 1) to when it is in logs (Column 2). Figure D.5a exhibits the dynamic effects. Despite the coefficients becoming more imprecise, the effect is still positive and growing with time: on average, formal employment grows 4.9% after a municipality receives a HDS, and around 10% after the third year. Robustness is stronger for the results about taxes, as shown in Table D.10 and Figure D.5b. Columns 2 and 4 of the table display the estimated impacts when using log of total and industry and commerce tax revenues, respectively. For the case of total tax revenues, coefficients are positive and significant: tax collection in the municipality grows after a HDS arrives by 10.3% on average. The effects also remain positive and with an upward trend for the case industry and commerce tax revenues, but they lose statistical significance.

over time. If our findings were driven by workers relocating in anticipation of higher labor demand, we would expect shifts in internal migration flows, which we do not observe.

7 Conclusion

In this paper, we examine the impact of the expansion of Colombia's leading hard discount chains on local labor markets. Unlike previous studies focusing on large retailers like Walmart or Amazon FCs, our research targets the effects in a developing country with a significant informal sector. Our findings reveal that following the introduction of HDS, there are notable shifts in local labor markets. Specifically, HDS entry increases local formal employment by around 10%, driven by primary and secondary industries, on top of retail. This suggests that the entry of hard discount chains into a municipality generates strong spillover effects from retail to other sectors. A plausible explanation is the increased demand for some of the goods local formal firms produce within the discounters' supply chain, thereby affected firms hire more formal workers to satisfy the increased demand. Importantly, the employment effects are not immediate, as they show up to three years after the first opening of HDS, aligning with the time and investments required for a hard discount chain to attract more customers and expand within the treated municipalities. Additionally, we show that hard discount chains pay consistently higher premiums to their workers than most firms in the Colombian formal sector, including the largest retail firm in the country, although this does not translate into an increase in local formal wages.

Regarding the informal retail sector, the findings suggests that discount stores do not decrease informal retail employment. However, we cannot empirically verify whether the entry of HDS leads to the closure of some local neighborhood shops due to lack of data on such outcomes. On the other hand, labor earnings in the informal retail sector exhibit an insignificant negative trend after the entry of hard discounters. We rationalize that the adjustment may occur through earnings rather than employment, as neighborhood shops are typically small entrepreneurial activities run mostly by their owner, with only 16% employing additional workers.

Lastly, we find suggestive evidence that the positive impacts on employment contribute to aggregate tax effects in the local economies. Post-HDS entry, the ratio of collected taxes over total public revenues increases by an average of 7.5%, driven by the revenues from the industry and commerce taxes. This highlights how relevant the entry of hard discount chains in a local labor market is. Collectively, these findings have important policy implications for developing countries, such as the promotion of specific type of businesses that may spur the formalization of local economies on top of other public policies. Further research is needed to better understand the potential effects on neighborhood shop profits and the long-term implications of the expansion of hard discount chains in developing countries.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acevedo, K. and Bornacelly Olivella, I. D. (2014). Panel municipal del cede. *Documentos CEDE*.

Acosta, K. and Gu, H. (2024). Reexamining a classical question: Does development push migration? *The Annals of Regional Science*, 77:2091–2115.

Alfaro-Ureña, A., Manelici, I., and Vasquez, J. P. (2021). The effects of multinationals on workers: Evidence from costa rican microdata. Technical report, Working paper.

Atkin, D., Faber, B., and Gonzalez-Navarro, M. (2018). Retail globalization and household welfare: Evidence from mexico. *Journal of Political Economy*, 126(1):1–73.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–2168.

Bachas, P., Gadenne, L., and Jensen, A. (2024). Informality, consumption taxes, and redistribution. *Review of Economic Studies*, 91:2604–2634.

Basker, E. (2005). Job creation or destruction? labor market effects of wal-mart expansion. *Review of Economics and Statistics*, 87(1):174–183.

Bassier, I. (2023). Firms and inequality when unemployment is high. *Journal of Development Economics*, 161:103029.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. *The Review of Economic Studies*, 91(6):3253–3285.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Chava, S., Oettl, A., Singh, M., and Zeng, L. (2023). Creative destruction? impact of e-commerce on the retail sector. *Management Science*, 0(0):1–20.

Cho, J., Chun, H., and Lee, Y. (2015). How does the entry of large discount stores increase retail employment? evidence from korea. *Journal of Comparative Economics*, 43:559–574.

Cunningham, E. R. (2024). Local labor market effects of amazon. *Unpublished manuscript*.

de la Parra, B. S., Otero-Cortés, A., and Morales, L. F. (2024). The labor market effects of facilitating social security contributions under part-time employment contracts: Evidence from colombia. *Labour Economics*, 91.

de Paula, A. and Scheinkman, J. (2010). Value-added taxes, chain effects, and informality. *American Economic Journal: Macroeconomics*, 2(4):195–221.

Delgado-Prieto, L. (2024a). Immigration, wages, and employment under informal labor markets. *Journal of Population Economics*, 37(2):55.

Delgado-Prieto, L. (2024b). Worker responses to immigration across firms: Evidence from colombia.

Dube, A., Lester, T. W., and Eidlin, B. (2007). Firm entry and wages: Impact of wal-mart growth on earnings throughout the retail sector.

Dustmann, C., Schönberg, U., and Stuhler, J. (2017). Labor supply shocks, native wages, and the adjustment of local employment. *The Quarterly Journal of Economics*, 132(1):435–483.

Euromonitor International (2023a). Discounters in germany [industry report].

Euromonitor International (2023b). Discounters retail sales by region [industry report].

Euromonitor International (2023c). Retail in colombia [industry report].

Evensen, C. B., Steen, F., and Ulsaker, S. A. (2023). Co-Location, Good, Bad, or Both: How do New Entries of Discount Variety Stores Affect Local Grocery Businesses? *Journal of the European Economic Association*, 22(4):1798–1843.

Fenaltiendas (2019). En defensa del pequeño emprendedor: los efectos de un eventual impuesto a las bebidas no alcohólicas en las tiendas de barrio.

Fernández, C. and Villar, L. (2017). The impact of lowering the payroll tax on informality in colombia. *Economía*, 18(1):125–155.

Florez-Acosta, J. and Herrera-Araujo, D. (2020). Multiproduct retailing and consumer shopping behavior: The role of shopping costs. *International Journal of Industrial Organization*, 68.

Gerard, F., Lagos, L., Severnini, E., and Card, D. (2021). Assortative matching or exclusionary hiring? the impact of employment and pay policies on racial wage differences in brazil. *American Economic Review*, 111(10):3418–3457.

Gerard, F., Naritomi, J., Seibold, A., and Zulian, B. (2023). Two-tier tax systems and firms: Evidence from brazil.

Gerard, F., Naritomi, J., and Silva, J. (2024). Cash transfers and the local economy: Evidence from brazil.

González Bell, J. (2020). Las movidas que hicieron los discounters en 2019 y los planes de expansión para 2020.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of political economy*, 118(3):536–598.

Grupo Exito (2015). En su cumpleaños 110, carulla anuncia que culminará el 2015 con 100 almacenes.

Haltiwanger, J., Jarmin, R., and Krizan, J. (2010). Mom-and-pop meet big-box: Complements or substitutes? *Journal of Urban Economics*, 67(1):116–134.

Jurgens, U. (2014). German grocery discounters: Dynamics and regional impact. the case of schleswig-holstein (germany). *Quastiones Geographicae*, 33(4):17–26.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

Kugler, A., Kugler, M., and Prada, L. O. H. (2017). Do payroll tax breaks stimulate formality? evidence from colombia's reform. *National Bureau of Economic Research*, (w23308).

La República (2022). Prácticamente el 80% de nuestros productos en Tiendas D1 son de producción nacional. URL.

La República (2023). Tiendas D1, Tiendas Ara e Ísimo planean la apertura de 830 puntos durante este año. URL.

Lagakos, D. (2016). Explaining cross-country productivity differences in retail trade. *Journal of Political Economy*, 124(2):579–620.

Londoño-Kent, P. (2009). Freight transport for development toolkit: Road freight. *Washington DC: The World Bank*.

Macchiavello, R. and Morjaria, A. (2021). Competition and relational contracts in the rwanda coffee chain. *Quarterly Journal of Economics*, 136(2):1089–1143.

MarketLine (2023). Food & grocery retail in colombia [industry report].

MarketLine Industry Profiles (2023). Food grocery retail in europe.

McConnell, B. (2023). What's logs got to do with it: On the perils of log dependent variables and difference-in-differences. *arXiv preprint arXiv:2308.00167*.

Morales, L. F. and Medina, C. (2017). Assessing the effect of payroll taxes on formal employment: The case of the 2012 tax reform in colombia. *Economía*, 18(1):75–124.

Morante, A. (2018). Petroleras fueron claves en más ganancias de empresas.

Moretti, E. (2010). Local multipliers. *American Economic Review*, 100(2):373–377.

Neumark, D., Zhang, J., and Ciccarella, S. (2008). The effects of wal-mart on local labor markets. *Journal of Urban Economics*, 63(2):405–430.

Pérez Pérez, J. and Nuño-Ledesma, J. G. (2024). Workers, workplaces, sorting, and wage dispersion in mexico. *Economía LACEA Journal*, 23(1):311–332.

Portafolio (2018). Ara da los pasos para duplicar sus tiendas en el país.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5):2555–2591.

Ramos-Menchelli, D. and Sverdlin-Lisker, D. (2023). Fragmented markets and the proliferation of small firms: Evidence from mom-and-pop shops in mexico.

Rios, J. and Setharam, I. (2018). Propagating formality via value-added tax networks: Evidence from india.

Roth, J. and Sant'Anna, P. H. C. (2023). Efficient estimation for staggered rollout designs. *Journal of Political Economy Microeconomics*, 1(4):669–709.

Sachon, M. (2010). The hard discount model in retailing.

Sánchez Duarte, L. A. (2017). El impacto de los establecimientos hard discount o tiendas de descuento en el sector comercial de colombia.

Schmidheiny, K. and Siegloch, S. (2023). On event study designs and distributed-lag models: Equivalence, generalization and practical implications. *Journal of Applied Econometrics*.

Silva Guerra, H. (2011). Comportamiento de las superficies de retail en colombia. *Pensamiento y Gestión*, 30:3–20.

Statista (2023). Leading retailers worldwide in 2021. Technical report.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Talamas Marcos, M. Á. (2025). Surviving competition: Neighborhood shops vs. convenience chains. *The Review of Economic Studies*, 92:553–585.

Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics*, 12:525–546.

Wiltshire, J. C. (2023). Walmart supercenters and monopsony power: How a large, low-wage employer impacts local labor markets. *Job market paper*.

Online Appendix

A Neighborhood Shops Characteristics

In this part of the appendix, we analyze extensive information on neighborhood shops exploiting a micro businesses survey that contains informal and formal firms (EMICRON, by its acronym in Spanish).[A.1] These small local grocers represent a crucial channel for the grocery retail industry and a relevant source of informal employment. Table A.1 shows that 88% of the surveyed businesses are run by its owner, who is also the only employee, and for those that have more than one employee, 44% of them are unpaid (usually relatives)[A.2]. They also have lower survival rates than the rest of the micro-businesses in the survey, as almost one-third of them have been operating for at least ten years compared to 48% of the businesses in other sectors.

Neighborhood shops are highly informal. Only 17% of them have an updated register in the local chamber of commerce, and less than 6% fill out any tax report.[A.3] Regarding employment informality, almost 9 out of 10 business owners do not contribute to the social security system. For businesses with employees, the informality rate is also around 91%. This is considerably higher than Colombia's overall employment informality rate in 2019, which was 50%.

The average neighborhood store spends around US$717 monthly on merchandise and sells goods for around US$1,000 monthly. Other monthly expenses (such as utilities, rent, and transportation) only account for US$100. For the stores with paid employees, the average monthly wage is approximately US$200, which is 30% below the mandatory monthly minimum wage for 2019. This results in an average monthly profit of around US$380. When self-reporting their average monthly profits, the typical response is around US$215.

There are important differences when doing the descriptive analysis by formality status. Table A.1

[A.1] The survey was conducted in 2019, which only includes businesses with up to 10 employees. We have data at the two-digit industry level, so we identify neighborhood shops as those businesses under code 45 of ISIC Revision 4: "Wholesale and retail commerce and repair of motor vehicles and motorcycles". Throughout the analysis, we use an exchange rate of 3,250 Colombian pesos per US dollar to convert the financial variables. This is an approximate of the average USD/COP rate in 2019.

[A.2] In all the results, we use the survey's sampling weights to compute the reported mean and standard deviation.

[A.3] Three taxes are particularly important for businesses in Colombia: the VAT, the income tax, and the industry and commerce tax. The first two are national-level, whereas the last one is paid to the municipality.

reports the mean and standard error of key indicators for two subgroups: stores with an updated register at the local chamber of commerce (that we classify as formal) and stores without (classified as informal). We plot some of these results in Figure A.1. Panel (a) shows that formal stores have, on average, merchandise sales that are 3.5 times higher than informal stores. They also have higher costs for merchandise sold: 6.5 million pesos for the typical formal store (US$2,000), compared to 1.5 million pesos for the typical informal one (US$450). Thus, formal stores are, on average, more profitable than informal stores: profits for merchandise sold per worker are US$187 in the typical informal store during a month and US$553 on average for formal shops.

Panel (b) of Figure A.1 shows the differences between formal and informal stores in other business characteristics. Formal shops have a larger staff than informal ones and rely less on unpaid workers. They are also less likely to start as needed and more likely to survive in the market.[A.4] Regarding employment informality, almost 4 in 10 formal business owners contribute to the social security system (compared to 1 in 10 for informal businesses), and the average share of informal paid workers is 80% for formal stores (versus 97% in informal stores).

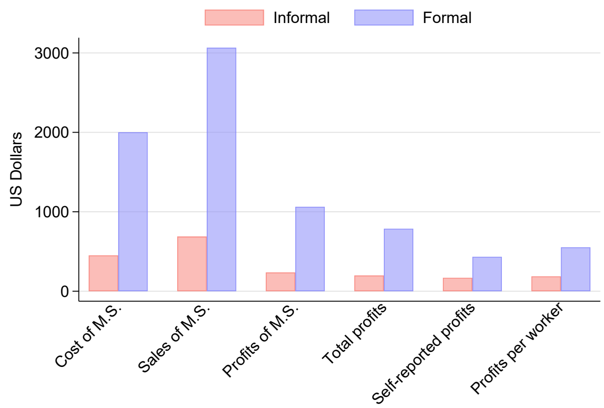
[A.4] Stores with an updated register at the Chamber of Commerce are also more likely to be classified as formal using other definitions: about 30% of them report income, VAT, or commerce tax (compared to only 1.2% of the informal stores).

Table A.1: **Summary statistics of corner shops by formality status**

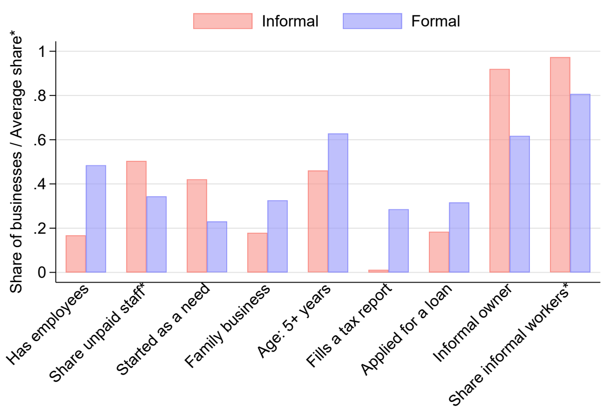
	Total sample mean (S.D.)	Formal mean	Informal mean
Panel A: Basic business characteristics			
Has employees	0.222	0.485	0.168
	(0.415)	(0.500)	(0.373)
Total business staff	1.320	1.832	1.215
	(0.744)	(1.210)	(0.549)
Share of unpaid staff in personnel	0.444	0.344	0.504
	(0.486)	(0.456)	(0.493)
Owner started bussiness alone	0.736	0.595	0.765
	(0.441)	(0.491)	(0.424)
Family business	0.204	0.326	0.179
	(0.403)	(0.469)	(0.384)
Business started as a need	0.389	0.230	0.422
	(0.488)	(0.421)	(0.494)
Business age: less than a year	0.164	0.070	0.183
	(0.370)	(0.255)	(0.387)
Business age: more than 1 year and less than 5 years	0.347	0.302	0.356
	(0.476)	(0.459)	(0.479)
Business age: more than 5 years	0.489	0.628	0.461
	(0.500)	(0.483)	(0.498)
Business located in household dwelling	0.332	0.306	0.338
	(0.471)	(0.461)	(0.473)
Panel B: Informality of the business and the owner			
Business reports income, VAT, or commerce tax	0.059	0.286	0.012
	(0.235)	(0.452)	(0.109)
Business applied for a loan	0.207	0.316	0.184
	(0.405)	(0.465)	(0.388)
Owner does not contribute to health or pension	0.869	0.618	0.921
	(0.337)	(0.486)	(0.270)
Panel C: Costs, sales, and profits (in USD)			
Cost of merchandise sold during last month	717.293	2,001.617	451.123
	(2,212.441)	(4,379.250)	(1,233.752)
Merchandise sales during last month	1,095.262	3,065.198	687.003
	(2,918.777)	(5,787.930)	(1,539.870)
Profits for merchandise sold during last month	377.970	1,063.581	235.880
	(1,090.237)	(2,046.299)	(670.479)
Total profits during last month	297.626	787.681	197.162
	(963.807)	(1,799.183)	(629.619)
Self-reported average monthly profits	214.648	434.129	169.653
	(486.733)	(745.249)	(399.647)
Average merchandise profits by employee	250.576	553.774	187.739
	(602.826)	(924.643)	(488.391)
Panel D: Personnel characteristics			
Share of women in personnel	0.541	0.560	0.529
	(0.466)	(0.446)	(0.477)
Average employee tenure in months	56.791	61.062	54.256
	(78.647)	(80.189)	(77.606)
Share of fully-informal employees	0.912	0.808	0.973
	(0.272)	(0.376)	(0.154)
Average wage (for paid employees) in USD	199.282	227.973	164.504
	(104.991)	(88.772)	(112.363)
Maximum obs (unweighted)	22,675	3,994	18,681
Maximum weighted obs	1,408,925	239,698	1,169,227

Note: This Table shows the mean of corner shops by formality status. A business is formal if it has an updated register at a chamber of commerce. We use a USD to COP exchange rate of 3,250 (which approximates the average rate in 2019) and define corner shops as businesses classified under code 45 of the ISIC Revision 4: "Wholesale and retail commerce and repair of motor vehicles and motorcycles". The mean and standard deviation are weighed using the survey's sample weights. Standard deviations are in parenthesis. Source: 2019 Microbusinesses Survey, DANE.

Figure A.1: **Neighborhood shops characteristics by formality status**



(a) Average monthly costs of merchandise sold (M.S), sales, and profits



(b) Employment, business origin, and other definitions of informality

Note: This figure reports the mean of selected characteristics for corner shops by formality status, using the 2019 Microbusinesses Survey. A business is formal if it has an updated register at a chamber of commerce. We use a USD to the COP exchange rate of 3,250 (which approximates the average rate in 2019) and define corner shops as businesses classified under code 45 of the ISIC Revision 4: "Wholesale and retail commerce and repair of motor vehicles and motorcycles". The mean and standard deviation are weighed using the survey's sample weights.

B Hard discount entry data

To obtain the year of the entrance of hard discount chains to Colombian municipalities, we built a dataset of active HDS in October 2020, containing their location and a proxy of the opening date. The location variable was obtained via web scraping, whereas the opening date comes from the store's register date in the

Chambers of Commerce. We then matched the store's location with the date using the store's name. This section describes the process of constructing this match in further detail.

B.1 Web scraping of HDS location

We web-scraped the websites of the three largest chains in October 2020, specifically the sites on the location of the active stores. The sites typically contain the name of the store, its address (with the name of the municipality and the department), and its opening time. We collected information on the first two variables for 2,938 HDS. Importantly, we added the chamber of commerce associated with the municipality for matching purposes.

B.2 Store date of register

We collected data on the universe of establishments that the three large chains had registered in the chambers of commerce by October 2020. In this dataset, establishments refer to stores, distribution centers, or stockrooms, either active or nonactive. Each table (one per company) contains the name of the establishment, the chamber of commerce where it was registered, the date of registration, and the status (active or nonactive). This dataset comprises 3,449 observations.

B.3 Match process

We matched the web-scraped stores with the chambers of commerce dataset on the establishment date of the register using exact, fuzzy, and manual matching.[B.1]

1. We consider two variables when executing the exact matching: the name of the store and the chamber of commerce. That means a web-scraped store must match an establishment that shares its name but is registered in its municipality's corresponding chamber of commerce. Around half of the stores (1,453) are matched using this method.

2. We executed two rounds of fuzzy matching, using the *Jaccard* index as string distance measure with

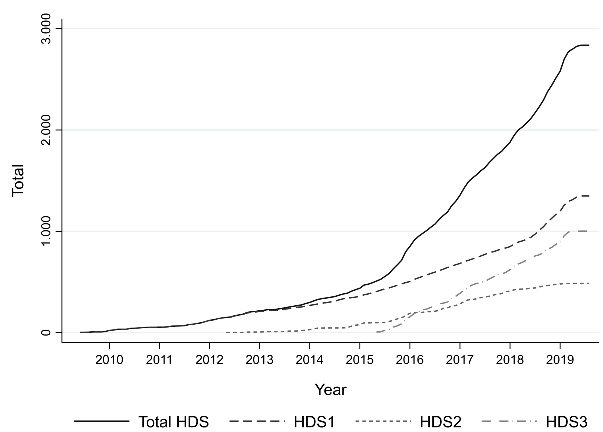
[B.1] Regardless of the matching algorithm, we always matched datasets of the same chain (i.e., web-scraped *Justo y Bueno* stores are always matched with *Justo y Bueno* registered establishments).

$q = 3$ as the size of the q-gram.[B.2] For the first round, we discarded the matches with an index higher than 0.8 (43 stores). After manually revising the matches and guaranteeing the coincidence of the chamber of commerce, we discarded 431 stores. At the end of the first round, 2,484 stores (around 85% of the active stores) are matched.

3. We repeat the fuzzy matching using the sample of unmatched web-scraped stores and unmatched registered establishments. In this second round, we did not discard matches based on the Jaccard Index or the coincidence of the chamber of commerce. After a manual revision, we additionally matched 259 stores. By the end of this round, 2,743 stores are matched, representing 93%

4. Finally, the manual matching comprises changing the name of unmatched registered stores to coincide with that of the web-scraped stores. Using this method, 105 additional stores are matched. After all the steps, we have a final dataset with 2,847 stores matched.

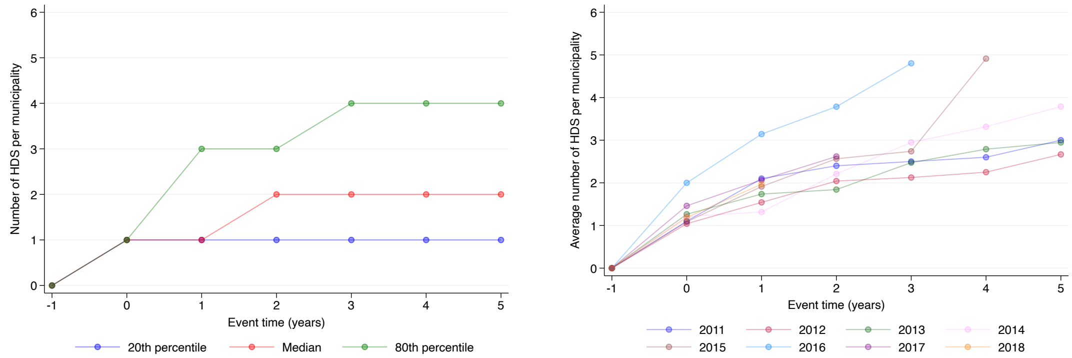
Figure B.1: **Number of HDS over time**



Note: This figure shows the total number of hard discount stores from the three main chains operating in Colombia between 2010 and 2019. Source: Authors' calculations using public location data obtained from the hard discounters' websites.

[B.2] The Jaccard index is defined as $1 - |X \cap Y|/|X \cup Y|$, where X and Y represent the set of q-grams of size $q = 3$ (subsequences of 3 consecutive characters) in the two strings that are being compared.

Figure B.2: **Number of HDS by event time per municipality**

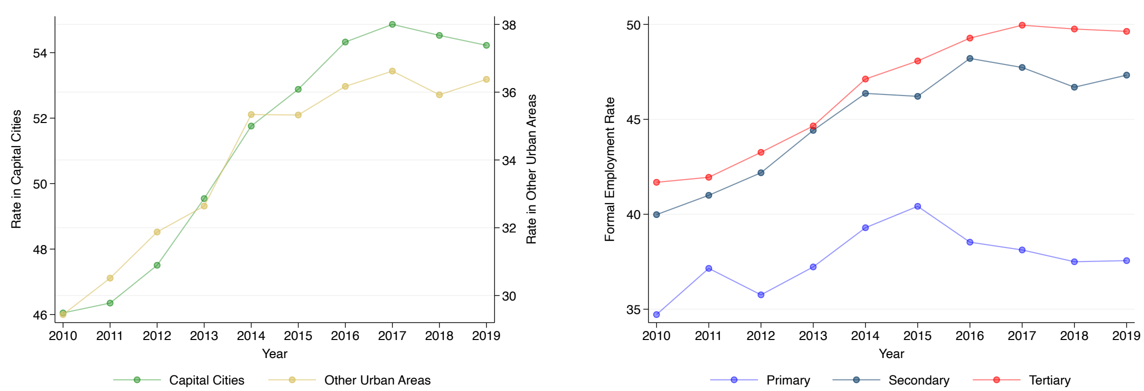


(a) Statistics of all cohorts **(b) Average by cohort**

Note: We restrict the event time after five years. Source: Authors' calculations using public location data from hard discounters' websites.

C Supplementary Results

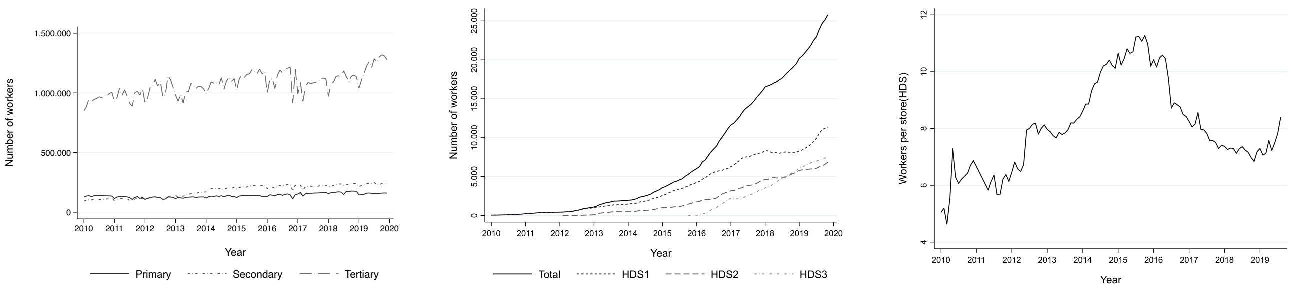
Figure C.1: **National formal employment rates**



(a) By geographic location **(b) By industry**

Note: This figure shows the formality rate by geographic location and by industry groups based on survey data. A worker is considered formal if they contribute to the social security system. We restrict the sample to workers between the ages of 18 and 64 located in urban areas. Source: GEIH 2010-2019.

Figure C.2: **Descriptive statistics on formal employment at HDS using administrative data**



(a) Total FE **(b) FE in HDS** **(c) Average full-time FE per HDS**

Note: Panel A shows the evolution of total full-time formal employment (FE) in the 372 municipalities included in our sample by industry. Primary refers to employment in agriculture and mining; secondary refers to employment in manufacturing and construction; and tertiary refers to employment in services. Panel B shows the evolution of full-time formal employment by the hard discount chain in the municipalities included in the estimation sample. Panel C reports the average full-time formal employment per store in the municipalities included in the estimation sample. Source: Authors' calculations based on PILA.

Table C.1: **Descriptive statistics for the estimation sample using administrative data**

	Treated				**Not yet treated**			
	2011	**2013**	**2016**	**2018**	**2011**	**2013**	**2016**	**2018**
Proportion male workers	62.3%	61.6%	60%	61.4%	62.9%	62.1%	62.0%	66.4%
Average formal employment	6,781	4,609	5,572	3,856	2,146	2,554	2,115	1,318
	(6,079)	(6,412)	(8,397)	(6,139)	(4,090)	(5,011)	(4,101)	(3,437)
CHR workers (%)	5.1%	6.6%	9.3%	10.0%	6.5%	8.5%	9.3%	10.1%
Non-CHR workers (%)	94.8%	93.3%	90.6%	89.9%	93.4%	91.4%	90.6%	89.8 %
Employees (%)	64.0%	73.0%	75.7%	83.7%	63.9%	70.6%	74.2%	87.5%
Independent workers (%)	18.4%	18.4%	19.4%	12.7%	20.1%	22.5%	21.6%	9.4%
Min wage workers (%)	57.2%	53.6%	46.9%	53.4%	53.5%	54.8%	47.6%	54.4%
Average earnings	301.6	323.8	334.7	353.7	319.8	337.0	311.2	334.6
	(279)	(332)	(391)	(377)	(356)	(362)	(269)	(277)
Municipalities	10	53	151	291	362	319	221	81

Note: This table reports the mean of selected labor market indicators using administrative records from PILA by year and treatment status. A municipality is considered treated when the first hard discount store opens in the local market. The average total employment is computed for all the municipalities treated that year and only includes full-time employees. "CHR workers" represent the proportion of formal full-time workers working in commerce, hotels, and restaurants. "Non-CHR workers" represent the proportion of formal full-time workers working in other industries outside commerce, hotels, and restaurants. "Employees" is the share of dependent workers on total formal workers. "Independent workers" is the share of formal self-employed on total formal workers. "Min wage workers" is the share of formal full-time workers earning the minimum wage on total formal workers. In PILA no workers are earning less than the the minimum wage as, by definition, formal full-time workers cannot earn less than the monthly minimum wage. "Average earnings" is the average reported labor income of full-time formal workers. Standard deviations are in parentheses.

Table C.2: **Treatment selection by hard discount chain**

	HDS1		HDS2		HDS3		Any chain	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total population (in logs)	0.174***	0.193***	0.107***	0.102***	0.164***	0.176***	0.187***	0.200***
	(0.013)	(0.011)	(0.023)	(0.025)	(0.020)	(0.016)	(0.018)	(0.014)
Driving time to Bogota in hours	-0.014***	-0.006	-0.010	-0.007	-0.017***	-0.013**	-0.010***	-0.005
	(0.005)	(0.009)	(0.007)	(0.007)	(0.004)	(0.006)	(0.003)	(0.008)
Driving time to Medellin in hours	-0.027***	-0.030***	0.000	0.000	0.004	-0.009	-0.019***	-0.029***
	(0.005)	(0.007)	(0.012)	(0.009)	(0.005)	(0.008)	(0.003)	(0.006)
Share of rural population	-0.005***	-0.005***	-0.005***	-0.004***	-0.005***	-0.006***	-0.007***	-0.006***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Share of formal tertiary employment	-0.001**	-0.000	0.000	0.001	-0.001	0.000	-0.002**	-0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Municipalities	1,038	1,037	1,038	1,037	1,038	1,037	1,038	1,037
R-squared	.472	.516	.29	.454	.398	.502	.494	.538
Dep Var Mean	.282	.282	.168	.167	.237	.236	.387	.387
Dep Var SD	.45	.45	.374	.373	.425	.425	.487	.487
Department F.E	NO	YES	NO	YES	NO	YES	NO	YES

Note: This table presents the estimation results of four cross-sectional linear probability models whose dependent variables indicate whether a municipality had received a store from a particular hard-discount chain by 2019 (in Columns 1 to 6), or from any chain in Columns 7 and 8. The municipal population and the share of the rural population were measured in 2005, while the share of formal tertiary employment was measured in 2010 and the driving times in 2024. Even columns also incorporate department dummies. Standard errors, clustered at the department level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: **Treatment selection by year (excluding not yet treated)**

	(1) 2011	(2) 2012	(3) 2013	(4) 2014	(5) 2015	(6) 2016	(7) 2017	(8) 2018
Static characteristics:								
Total population (in logs)	0.050**	0.087***	0.064**	0.081***	0.094***	0.142***	0.173***	0.142***
	(0.024)	(0.022)	(0.027)	(0.028)	(0.028)	(0.017)	(0.022)	(0.025)
Driving time to Bogota in hours	-0.004	-0.008**	-0.008*	-0.008**	-0.006**	-0.015***	-0.019***	-0.014***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)
Driving time to Medellin in hours	-0.008***	-0.012***	-0.006**	-0.007**	-0.003	-0.002	-0.003	-0.007*
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)	(0.004)	(0.004)
Share of rural population	-0.002**	-0.003***	-0.002**	-0.002**	-0.003***	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of formal tertiary employment	0.000	0.000	0.000	0.000	0.001*	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Dynamic characteristics:								
Formal employment growth	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Formal wages growth	-0.000	-0.000	0.000	0.000	0.001	0.001	0.001*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Income tax growth	-0.000	-0.000*	-0.000	-0.000	0.000*	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
# of total municipalities	643	649	655	656	662	697	718	701
# of muni. treated that year	11	27	19	20	26	62	82	65
R-squared	.208	.348	.189	.248	.26	.449	.403	.258

Note: This table presents the estimation results of yearly cross-sectional linear probability models whose dependent variables indicate whether a municipality received its first hard-discount store in a given year, with never-treated municipalities as control group. The municipal population and the share of the rural population were measured in 2005, while the share of formal tertiary employment was measured in 2010 and the driving times in 2024. Dynamic characteristics are calculated as annual percentage changes with respect to the previous year. Standard errors, clustered at the department level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: **Treatment selection by year (excluding never treated)**

	(1) 2011	(2) 2012	(3) 2013	(4) 2014	(5) 2015	(6) 2016	(7) 2017	(8) 2018
Static characteristics:								
Total population (in logs)	0.010	0.024	-0.014	0.029	0.030	0.101***	0.143***	0.120**
	(0.012)	(0.017)	(0.012)	(0.018)	(0.019)	(0.034)	(0.036)	(0.049)
Driving time to Bogota in hours	0.002*	-0.001	-0.002	-0.001	0.006	-0.011	-0.028**	-0.041***
	(0.001)	(0.003)	(0.004)	(0.003)	(0.006)	(0.014)	(0.011)	(0.007)
Driving time to Medellin in hours	-0.014***	-0.026***	-0.014**	-0.020**	0.000	-0.001	0.004	-0.009
	(0.004)	(0.005)	(0.006)	(0.009)	(0.005)	(0.015)	(0.013)	(0.006)
Share of rural population	-0.000*	-0.001*	-0.001	-0.000	-0.001	-0.004***	-0.004***	-0.004*
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Share of formal tertiary employment	-0.000	-0.000	-0.000	-0.000	0.001	-0.000	0.002	0.003**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Dynamic characteristics:								
Formal employment growth	0.000	0.000	-0.001	-0.001	-0.001	0.000	0.001	-0.002
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Formal wages growth	0.001	-0.000	0.001	0.002	0.005*	0.001	0.001	-0.008
	(0.001)	(0.000)	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)
Income tax growth	0.000	-0.000	-0.000	0.000	0.000	-0.000	-0.000	0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)
# of total municipalities	389	378	352	333	313	287	225	143
# of muni. treated that year	11	27	19	20	26	62	82	65
R-squared	.0936	.144	.0702	.0952	.068	.136	.148	.159

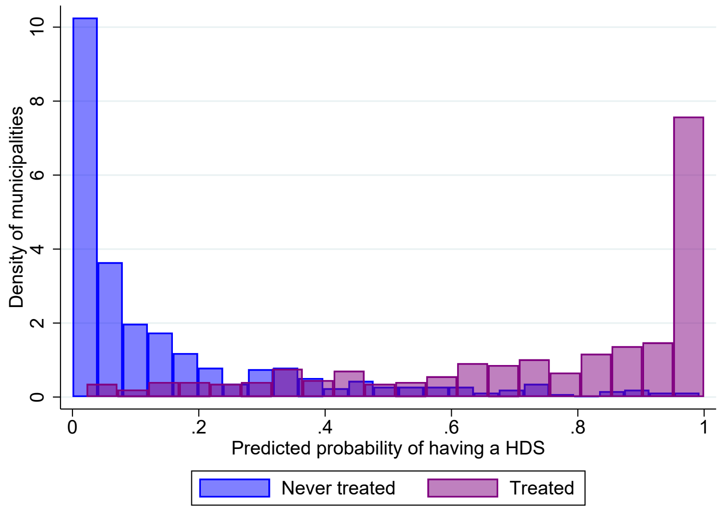
Note: This table presents the estimation results of yearly cross-sectional linear probability models whose dependent variables indicate whether a municipality received its first hard-discount store in a given year, with later-treated municipalities as control group. The municipal population and the share of the rural population were measured in 2005, while the share of formal tertiary employment was measured in 2010 and the driving times in 2024. Dynamic characteristics are calculated as annual percentage changes with respect to the previous year. Standard errors, clustered at the department level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.5: **Labor income and working hours outcomes for estimation sample using survey data**

	Treated				**Not yet treated**			
	2011	**2013**	**2016**	**2018**	**2011**	**2013**	**2016**	**2018**
Wages (USD)	274.7	300.3	305.4	300.4	276.4	294.3	274.6	282.7
	(110.9)	(60.6)	(85.8)	(77.1)	(69.5)	(95.9)	(61.4)	(70.5)
Wages: Informal sector (USD)	191.2	222.8	219.9	209.0	206.6	213.8	194.4	194.6
	(61.2)	(67.9)	(48.1)	(42.7)	(54.1)	(60.2)	(47.7)	(39.9)
Wages: Formal sector (USD)	372.4	413.3	417.7	441.7	434.6	453.1	443.9	479.8
	(114.2)	(74.6)	(106.8)	(133.5)	(110.2)	(129.4)	(114.3)	(107.9)
Working hours	47.1	45.3	46.1	45.7	47.0	46.4	45.9	43.8
	(1.9)	(3.7)	(2.5)	(3.0)	(3.2)	(3.0)	(3.0)	(3.7)
Working hours: Informal sector	44.6	43.8	44.2	44.1	46.2	45.5	45.5	42.2
	(3.0)	(4.4)	(3.7)	(3.8)	(3.6)	(3.9)	(3.9)	(3.6)
Working hours: Formal sector	51.7	50.2	49.3	48.4	49.4	49.2	47.7	46.3
	(1.4)	(2.6)	(2.2)	(3.1)	(4.3)	(3.6)	(4.7)	(5.0)
Municipalities	5	28	85	156	186	163	106	35
Average 2010 Employed Population	35,923	21,407	26,963	21,058	18,750	18,821	12,974	10,917

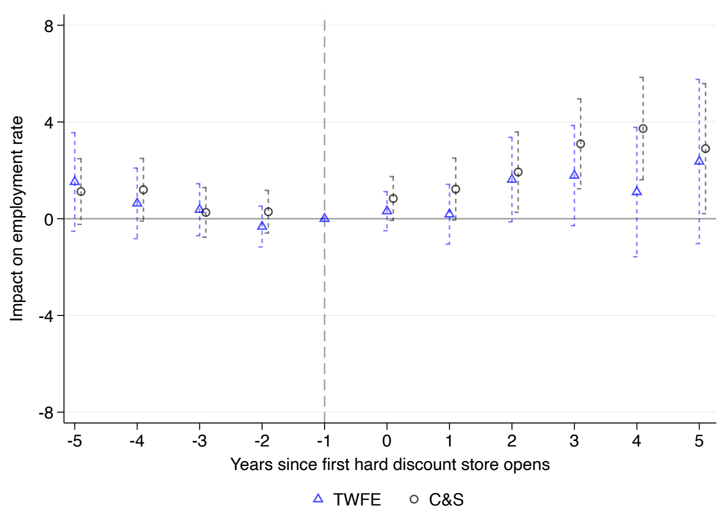
Note: This Table shows the mean values of labor income and working hours indicators using the municipal panel of the GEIH, categorized by year and treatment status. The descriptive statistics are weighted by total employment in each municipality in 2010. Standard deviations are in parentheses.

Figure C.3: **Distribution of propensity scores by treatment status**



Note: This figure plots the distribution of the predicted probabilities of receiving HDS before 2020 by treatment status. We estimate a logit model where the independent variables match those in Table C.2 (without department dummies). We then predict the probability at the municipality level, which is depicted in the x-axis.

Figure C.4: **Event study estimates on the employment rate**



Note: The dependent variable is the yearly employment rate. Regressions were weighted with local employment in 2010. Observed treated municipalities are 191. When the panel is not balanced in certain outcomes, we use only the observations with a balanced pair (observed in the pre- and post-treatment period). Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: GEIH 2010-2018.

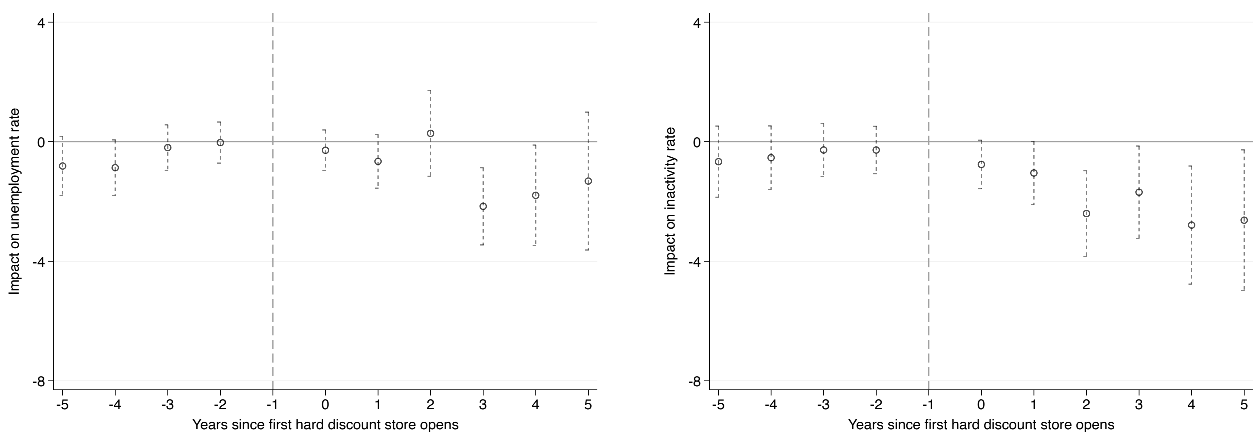
Table C.6: **Average *C&S* estimates of labor market rates**

	(1) Employment rate	(2) Unemployment rate	(3) Inactivity rate
ATT_{pre}	0.718	-0.477	-0.439
	(0.536)	(0.398)	(0.443)
ATT_{post}	2.286***	-0.990	-1.886***
	(0.849)	(0.628)	(0.728)
$ATT_{post_{k=3,4,5}}$	3.242***	-1.758**	-2.369**
	(1.191)	(0.884)	(1.058)
N	1,719	1,674	1,713
Municipalities	191	191	191
Mean pre-treatment	70.3	11.3	20.9

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is the yearly labor market rate. Regressions were weighted with local employment in 2010. Observed treated municipalities are 191. Since the panel is not balanced for certain outcomes, we use only observations with pair balanced (that is, observed during pre- and post-treatment period). Standard errors are clustered at the municipality level. The coefficients represent percentage points changes. Source: GEIH 2010-2018 in August.

Figure C.5: **Event study estimates of labor market rates**



(a) Unemployment Rate

(b) Inactivity Rate

Note: We use the C&S estimator. The dependent variable is the yearly labor market rate. Regressions were weighted with local employment in 2010. Observed treated municipalities are 191. Since the panel is not balanced for certain outcomes, we use only observations with pair balanced (that is, observed during pre- and post-treatment period). Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: GEIH 2010-2018 in August.

Table C.7: **Average formal employment effects by cohorts of treatment**

	Admin	Survey
$g = 2011$	2.083**	2.853
	(0.864)	(2.173)
$g = 2012$	1.455*	1.705
	(0.768)	(1.300)
$g = 2013$	3.283	3.459
	(3.097)	(2.593)
$g = 2014$	1.417*	4.964***
	(0.834)	(1.098)
$g = 2015$	0.872	4.854*
	(1.234)	(2.735)
$g = 2016$	-0.111	0.408
	(0.485)	(1.485)
$g = 2017$	0.177	0.196
	(0.456)	(1.512)
$g = 2018$	-0.163	-0.401
	(0.316)	(2.412)

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: We use the *C&S* estimator. The dependent variable in column (1) is formal employment using PILA, and in column (2) is formal employment using GEIH both over the working-age population according to the 2005 census. Regressions were weighted with the local working-age population in 2005. Observed treated municipalities in PILA are 372, and in GEIH are 191. Standard errors are clustered at the municipality level. The coefficients represent percentage points changes. Source: PILA 2010-2018 in August, and GEIH 2010-2018.

Table C.8: **Average *C&S* estimates of formal employment ratios by sector using administrative data**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
ATT_{pre}	-0.417	-0.045*	0.030	-0.074	-0.100
	(0.294)	(0.025)	(0.090)	(0.205)	(0.211)
ATT_{post}	1.742***	0.108**	-0.290*	0.877***	0.670
	(0.608)	(0.044)	(0.158)	(0.296)	(0.474)
$ATT_{post_{k=3,4,5}}$	3.147***	0.170***	-0.397*	1.647***	1.013
	(1.049)	(0.066)	(0.228)	(0.426)	(0.756)
N	3,348	3,348	3,348	3,348	3,348
Municipalities	372	372	372	372	372
Mean pre-treatment	16	.5	1.7	4.5	12.5

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is formal employment over the working-age population according to the 2005 census. Regressions were weighted with the local working-age population in the 2005 census. The CHR refers to commerce, hotels, and restaurants. Primary and secondary are manufacturing, construction, agriculture, and mining. Observed treated municipalities are 372. Standard errors are clustered at the municipality level. The coefficients represent percentage points changes. Source: PILA 2010-2018 in August.

Table C.9: **Average *C&S* estimates of formal employment ratios by subsector using administrative data**

	(1) Total	(2) Agriculture	(3) Manufacturing	(4) Construction
ATT_{pre}	-0.417	0.037	-0.210**	0.076
	(0.294)	(0.112)	(0.097)	(0.056)
ATT_{post}	1.742***	-0.005	0.611***	0.175**
	(0.608)	(0.160)	(0.167)	(0.082)
$ATT_{post_{k=3,4,5}}$	3.147***	0.028	1.118***	0.338***
	(1.049)	(0.226)	(0.268)	(0.122)
N	3,348	3,348	3,348	3,348
Municipalities	372	372	372	372
Mean pre-treatment	16	1.7	1.8	0.8

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is formal employment in each sector over the working-age population according to the 2005 census. Regressions were weighted with the local working-age population in the 2005 census. Observed treated municipalities are 191. Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: PILA 2010-2018.

Table C.10: **Average *C&S* estimates of formal employment ratios by sector using survey data**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
ATT_{pre}	0.215	-0.057	-0.066	0.190	0.148
	(0.738)	(0.182)	(0.158)	(0.365)	(0.538)
ATT_{post}	2.910***	0.554**	-0.069	1.911***	0.514
	(0.967)	(0.220)	(0.207)	(0.535)	(0.524)
$ATT_{post_{k=3,4,5}}$	4.501***	0.778***	0.196	2.724***	0.804
	(1.318)	(0.286)	(0.293)	(0.719)	(0.768)
N	1,719	1,719	1,719	1,719	1,719
Municipalities	191	191	191	191	191
Mean pre-treatment	28.3	2.6	2.2	8.4	15.1

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is formal employment in each sector relative to the working-age population according to the 2005 census. The CHR refers to commerce, hotels and restaurants. Primary and secondary are manufacturing, construction, agriculture, and mining. Regressions were weighted with the working-age population of the 2005 census. Observed treated municipalities are 191. Standard errors are clustered at the municipality level. The coefficients represent percentage points changes. Source: GEIH 2010-2018.

Table C.11: **Average *C&S* estimates of formal employment ratios by subsector using survey data**

	(1) Total	(2) Agriculture	(3) Manufacturing	(4) Construction
ATT_{pre}	0.215	-0.114	0.112	0.080
	(0.738)	(0.156)	(0.222)	(0.184)
ATT_{post}	2.910***	0.578*	0.920***	0.373*
	(0.967)	(0.348)	(0.288)	(0.194)
$ATT_{post_{k=3,4,5}}$	4.501***	0.766*	1.272***	0.581*
	(1.318)	(0.432)	(0.391)	(0.311)
N	1,719	1,719	1,719	1,719
Municipalities	191	191	191	191
Mean pre-treatment	28.3	2.1	4.5	1.3

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is formal employment in each sector over the working-age population according to the 2005 census. Regressions were weighted with the local working-age population in the 2005 census. Observed treated municipalities are 191. Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: GEIH 2010-2018.

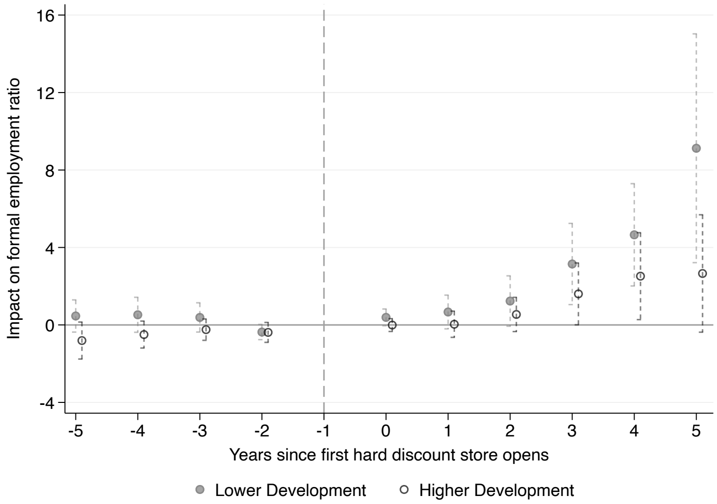
Table C.12: **Average *C&S* estimates of informal employment ratios by sector using survey data**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
ATT_{pre}	0.980	0.085	0.413	-0.718	1.199***
	(0.782)	(0.357)	(0.318)	(0.473)	(0.417)
ATT_{post}	-1.105	0.281	-0.223	-0.806	-0.357
	(1.467)	(0.573)	(0.429)	(0.856)	(0.496)
$ATT_{post_{k=3,4,5}}$	-2.219	0.017	-0.674	-0.441	0.804
	(1.729)	(0.756)	(0.569)	(1.054)	(0.768)
N	1,719	1,719	1,719	1,719	1,719
Municipalities	191	191	191	191	191
Mean pre-treatment	45.3	9.7	7.6	12.6	15.5

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

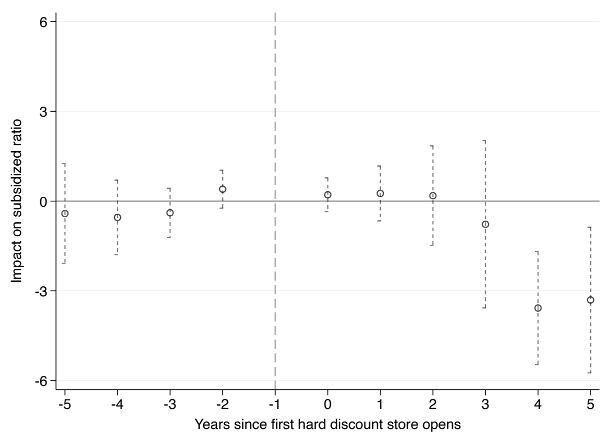
Note: The dependent variable is informal employment in each sector over the working-age population according to the 2005 census. The CHR refers to commerce, hotels, and restaurants. Primary and secondary are manufacturing, construction, agriculture, and mining. Regressions were weighted with local employment in 2010. Observed treated municipalities are 191. Standard errors are clustered at the municipality level. The coefficients represent percentage points changes. Source: GEIH 2010-2018.

Figure C.6: **Heterogeneous event study estimates on formal employment ratios**



Note: We use the *C&S* estimator. The outcome is the formal employment ratio. We distinguish municipalities by baseline development: 191 municipalities with survey coverage are classified as higher development, while the remaining 181 municipalities without survey coverage are classified as lower development. Regressions are weighted using the local working-age population from the 2005 census. Standard errors are clustered at the municipality level. 90% confidence intervals. Source: PILA 2010–2018 in August.

Figure C.7: **Event study estimates on subsidized social protection beneficiaries**



Note: The dependent variable is the number of beneficiaries of subsidized social protection times the share of individuals from 15 to 59 in the 2018 census over the working-age population in the 2005 census. Regressions were weighted with the local working-age population in the 2005 census. Observed treated municipalities are 372. Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: Health Ministry 2010-2018.

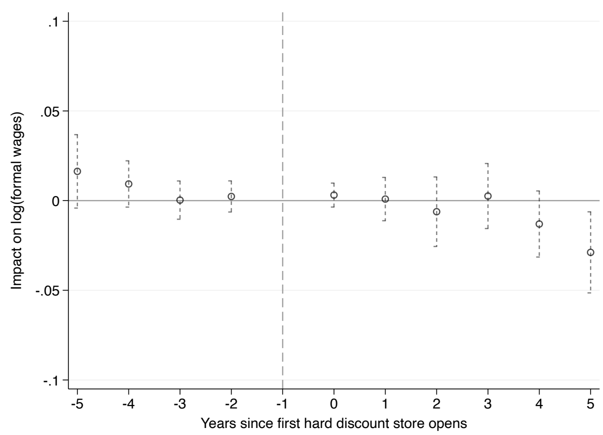
Table C.13: **Average *C&S* estimates of tax revenue ratios by type**

	(1) All	(2) Non taxes	(3) Property	(4) Industry and commerce	(5) Gasoline	(6) Other taxes
ATT_{pre}	2.321	1.696	-2.196*	0.718	0.122	3.678*
	(4.003)	(1.470)	(1.232)	(1.927)	(0.311)	(2.158)
ATT_{post}	10.138**	2.265	3.599*	6.291**	0.059	0.189
	(4.740)	(1.735)	(2.025)	(2.939)	(0.509)	(1.738)
$ATT_{post_{k=3,4,5}}$	14.091*	3.886	5.822*	10.967**	0.270	-2.968
	(7.374)	(2.781)	(3.161)	(4.741)	(0.718)	(2.582)
N	3,339	3,339	3,339	3,339	3,339	3,339
Municipalities	371	371	371	371	371	371
Mean pre-treatment	134.3	14.2	42.2	44.3	13.8	34

Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is the type of tax or revenue over all the revenues in 2010. Regressions were weighted using municipality's revenues in 2010. Observed treated municipalities are 371. Standard errors are clustered at the municipality level. The coefficients represent percentage points changes. Source: DNP, 2010-2018.

Figure C.8: **Event study estimates on formal wages using administrative data**

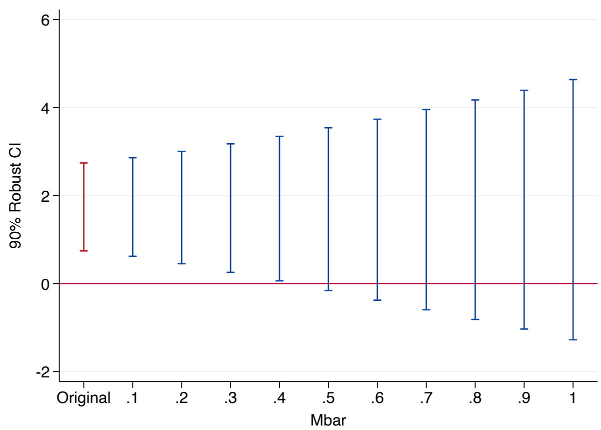


Note: We use the *C&S* estimator. The dependent variable is the logarithm of average wages . Regressions were weighted with the local working-age population in the 2005 census. Observed treated municipalities are 372. Standard errors are clustered at the municipality level. 90% confidence interval. The coefficients represent percentage changes. Source: PILA 2010-2018 in August.

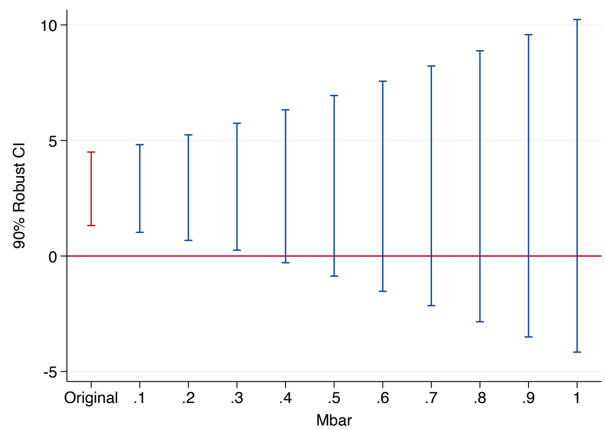
D Robustness Checks

D.1 Figures

Figure D.1: **Sensitivity analysis of ATT_{Post} for formal employment**



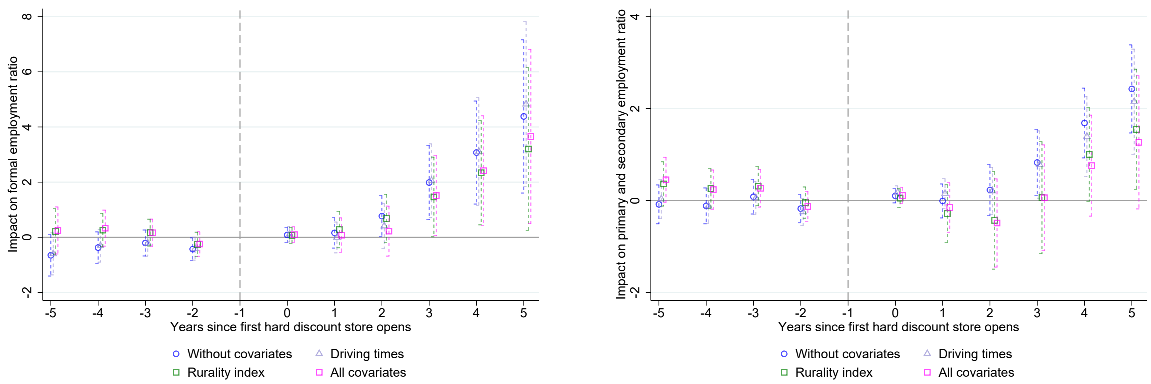
(a) Formal employment (Admin)



(b) Formal employment (Survey)

Note: The coefficient is the average of all post-treatment periods. The $Mbar$ refers to how robust the coefficient is to the maximal violation in pre-trends. For instance, $Mbar = 1$ assumes the maximal pre-treatment violation while $Mbar = 0.5$ assumes half the maximal violation. Standard errors are clustered at the municipality level. 90% robust confidence intervals with conditional-least favorable option.

Figure D.2: **Event study estimates for administrative data formal employment ratios including controls**

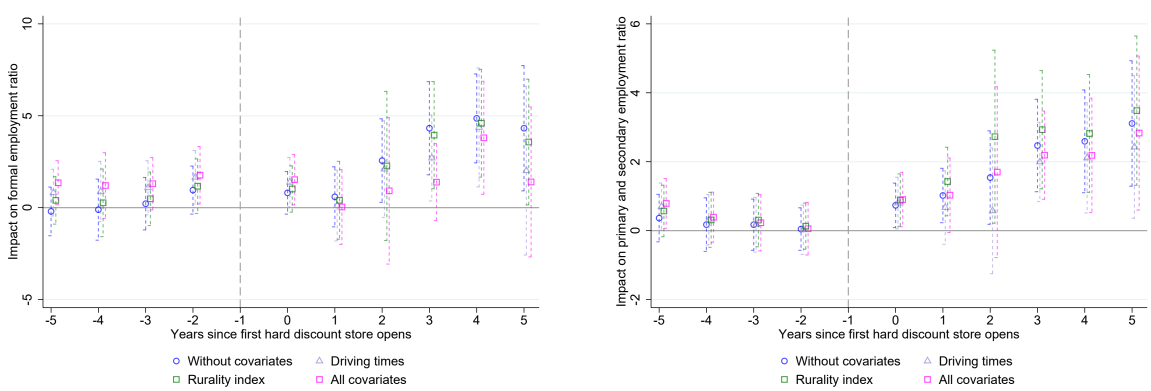


(a) **Total Employment**

(b) **Primary & secondary**

Note: The dependent variable is the formal employment in a given sector using administrative data over the working-age population in the 2005 census. We control for the share of rural population measured in 2005 and the driving times to Bogotá and Medellín using the outcome regression DiD estimator based on ordinary least squares. Regressions were weighted with the local working-age population in the 2005 census, and standard errors were clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: PILA 2010-2018.

Figure D.3: **Event study estimates for survey data formal employment ratios including controls**

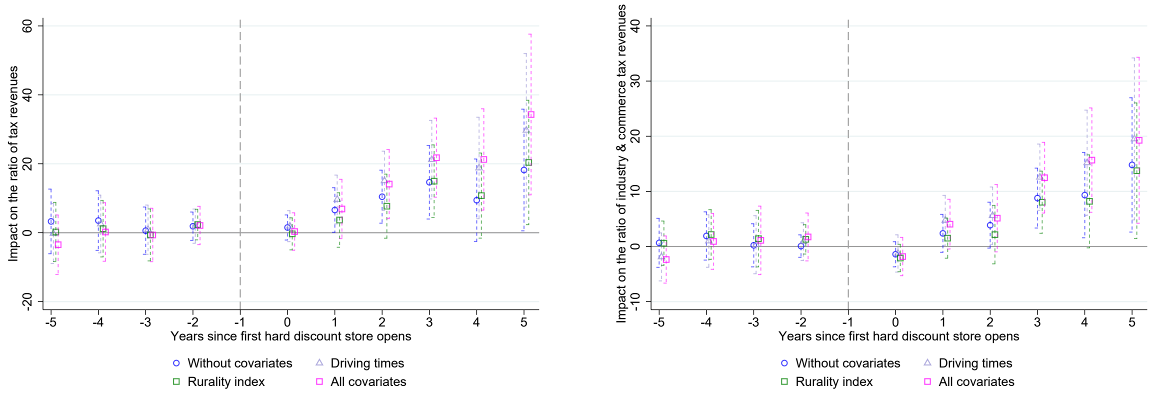


(a) **Total Employment**

(b) **Primary & secondary**

Note: The dependent variable is the formal employment in a given sector using survey data over the working-age population in the 2005 census. We control for the share of rural population measured in 2005 and the driving times to Bogotá and Medellín using the outcome regression DiD estimator based on ordinary least squares. Regressions were weighted with the local working-age population in the 2005 census, and standard errors were clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: GEIH 2010-2018.

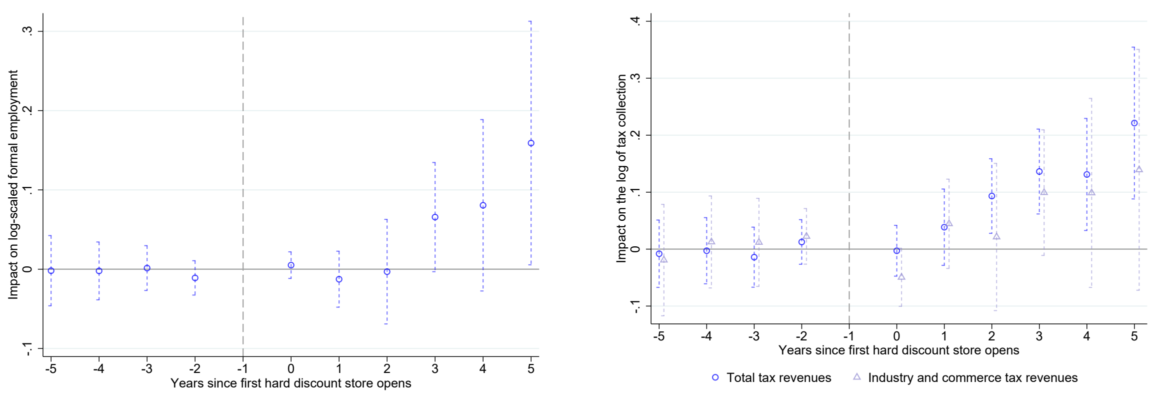
Figure D.4: **Event study estimates for tax outcomes including controls**



(a) **Tax revenues** (b) **Industry and commerce**

Note: The dependent variable in panel (a) is the ratio of taxes over all revenues (including taxes and central government transfers) and the ratio of industry and commerce tax revenues over all revenues in panel (b). We control for the share of rural population measured in 2005 and the driving times to Bogotá and Medellín using the outcome regression DiD estimator based on ordinary least squares. Regressions are weighted using municipality's revenues in 2010, and standard errors were clustered at the municipality level. 90% confidence interval. The coefficients represent percentage points changes. Source: DNP 2010-2018.

Figure D.5: **Event study estimates for formal employment and main taxation outcomes in logs**



(a) **Total formal employment** (b) **Taxation outcomes**

Note: The dependent variable in panel (a) is the natural logarithm of total formal employment using administrative data, while in panel (b) we log-transform the two main tax outcomes: total tax revenues, and industry and commerce tax revenues. The regression in panel (a) is weighted with the working-age population of the 2005 census, and those in panel (b) with the municipality's revenues in 2010. All regressions control for the share of rural population and driving times to Bogotá and Medellín using the outcome regression DiD estimator based on ordinary least squares. Standard errors clustered at the municipal level. 90% confidence interval. Sources: PILA 2010-2018, and DNP 2010-2018.

D.2 Tables

Table D.1: **Average *C&S* estimates of labor market rates using never treated municipalities as control group**

	(1) Employment rate	(2) Unemployment rate	(3) Inactivity rate
Panel A: Not-yet-treated as control group			
ATT_{pre}	0.718	-0.477	-0.439
	(0.536)	(0.398)	(0.443)
ATT_{post}	2.286***	-0.990	-1.886**
	(0.849)	(0.628)	(0.728)
$ATT_{post_{k=3,4,5}}$	3.242***	-1.758**	-2.369**
	(1.191)	(0.884)	(1.058)
Panel B: Never treated as control group			
ATT_{pre}	1.737***	-1.201***	-1.022**
	(0.521)	(0.388)	(0.435)
ATT_{post}	-0.234	0.404	0.003
	(0.534)	(0.470)	(0.492)
$ATT_{post_{k=3,4,5}}$	-0.036	0.409	-0.152
	(0.711)	(0.649)	(0.663)
Panel C: Never treated as control group (restricted)			
ATT_{pre}	1.153	-0.204	-1.163
	(0.919)	(0.751)	(0.730)
ATT_{post}	1.367	0.873	-2.163*
	(0.996)	(0.785)	(1.152)
$ATT_{post_{k=3,4,5}}$	2.824*	0.860	-3.856**
	(1.645)	(1.095)	(1.830)

Note: The dependent variable in each column is the yearly labor market rate. The sample of never-treated municipalities in Panel C is restricted to those selected by a one-to-one logit-based propensity score matching using the dependent variable and covariates in Column 7 of Table C.2. Regressions were weighted with the local employment in 2010, and standard errors are clustered at the municipality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: GEIH 2010-2018.

Table D.2: **Average $C\&S$ estimates of formal firm growth by firm size**

	(1) Firm size ($1-50$ workers)	(2) Firm size (>50 workers)
ATT_{pre}	1.830	0.195
	(1.861)	(0.158)
ATT_{post}	-0.657	-0.256
	(3.490)	(0.269)
N	3,348	3,348
Municipalities	372	372

Note: The dependent variable is the number of formal firms by size over the total number of firms in 2010. Regressions were weighted with local firm size in 2010. Observed treated municipalities are 372. Standard errors are clustered at the municipality level. The coefficients represent percentage points changes. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Source: PILA 2010-2018.

Table D.3: **Fisher Randomization Test p-values for event study estimates of formal employment**

	Total:Clustered	Total:Neyman Fisher	Total:Adjusted Fisher	P&S:Clustered	P&S:Neyman Fisher	P&S:Adjusted Fisher
$ATT_{post_{k=0}}$	0.555	0.566	0.566	0.123	0.121	0.121
$ATT_{post_{k=1}}$	0.530	0.538	0.538	0.947	0.941	0.941
$ATT_{post_{k=2}}$	0.259	0.267	0.267	0.468	0.453	0.453
$ATT_{post_{k=3}}$	0.005	0.004	0.004	0.002	0.000	0.000
$ATT_{post_{k=4}}$	0.001	0.001	0.001	0.000	0.000	0.000
$ATT_{post_{k=5}}$	0.004	0.007	0.007	0.000	0.001	0.001

Note: This table reports robustness of the main formal employment event study estimates to the use of p-values based on randomization tests. The dependent variable is total employment in Columns 1 to 3, and primary & secondary employment in Columns 4 to 6, both according to the administrative data and relative to the working-age population according to the 2005 census. P-values in columns 1 and 4 are those computed using Callaway & Sant'Anna (they rely on standard errors clustered at the municipality level). Those in columns 2, 3, 5, and 6 follow Fisher Randomization Tests (FRT) suggested by Roth and Sant'Anna (2023), whose standard errors are computed using two methods: Neyman in Columns 2 and 5, and covariate-adjusted in Columns 3 and 6. All FRT p-values are based on 5,000 permutations. Source: PILA 2010-2018.

Table D.4: **Average *C&S* estimates of labor market indicators using controls**

	(1) Employment rate	(2) Unemployment rate	(3) Inactivity rate
Panel A: Baseline			
ATT_{pre}	0.718	-0.477	-0.439
	(0.536)	(0.398)	(0.443)
ATT_{post}	2.286***	-0.990	-1.886**
	(0.849)	(0.628)	(0.728)
$\text{ATT}_{post_{k=3,4,5}}$	3.242***	-1.758**	-2.369**
	(1.191)	(0.884)	(1.058)
Panel B: Including Covariates of Distance to Main Cities and Rurality Index			
ATT_{pre}	0.077	-0.181	0.159
	(0.656)	(0.493)	(0.538)
ATT_{post}	3.234***	-1.539**	-2.694***
	(1.077)	(0.762)	(0.867)
$\text{ATT}_{post_{k=3,4,5}}$	4.611***	-2.247**	-3.765***
	(1.334)	(0.927)	(1.104)

Note: This table reports the robustness to the inclusion of covariates of the labor market rates estimates (Table C.6). The dependent variable in each column is the yearly labor market rate. Panel B controls for the share of rural population and the driving times to Bogotá and Medellín. We incorporate these covariates to the model using the outcome regression DiD estimator based on ordinary least squares. Regressions were weighted with the local employment in 2010, and standard errors clustered at the municipality level. The coefficients represent percentage points changes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: GEIH 2010-2018.

Table D.5: **Average *C&S* estimates of informal employment ratios by sector using controls**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: Baseline					
ATT_{pre}	0.980	0.085	0.413	-0.718	1.199***
	(0.782)	(0.357)	(0.318)	(0.473)	(0.417)
ATT_{post}	-1.105	0.281	-0.223	-0.806	-0.357
	(1.467)	(0.573)	(0.429)	(0.856)	(0.496)
$\text{ATT}_{post_{k=3,4,5}}$	-2.219	0.017	-0.674	-0.441	-1.120*
	(1.729)	(0.756)	(0.569)	(1.054)	(0.627)
Panel B: Including Covariates of Distance to Main Cities and Rurality Index					
ATT_{pre}	-0.290	-0.115	-0.094	-1.001*	0.920*
	(1.007)	(0.455)	(0.424)	(0.534)	(0.519)
ATT_{post}	0.503	0.904	-0.290	-0.244	0.133
	(1.558)	(0.676)	(0.691)	(1.191)	(0.602)
$\text{ATT}_{post_{k=3,4,5}}$	-0.193	0.824	-0.776	0.323	-0.563
	(1.974)	(0.890)	(0.901)	(1.353)	(0.749)

Note: This table reports the robustness to the inclusion of covariates of the informal employment estimates (Table C.12). The dependent variable is informal employment by the given sector relative to working-age population according to the 2005 census of a given municipality. The CHR refers to commerce, hotels, and restaurants. Primary and secondary are manufacturing, construction, agriculture, and mining. Panel B controls for the share of rural population and the driving times to Bogotá and Medellín. We incorporate these covariates to the model using the outcome regression DiD estimator based on ordinary least squares. Regressions were weighted with the working-age population of the 2005 census, and standard errors clustered at the municipality level. The coefficients represent percentage points changes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: GEIH 2010-2018.

Table D.6: **Average *C&S* estimates of labor income and working hours using controls**

	(1)	(2)	(3)	(4)	(5)	(6)
	Labor income				Working hours	
	Total formal	Total informal	Retail formal	Retail informal	Total formal	Total informal
Panel A: Baseline						
ATT_{pre}	0.030	0.002	-0.062	0.007	0.002	0.004
	(0.023)	(0.016)	(0.050)	(0.035)	(0.007)	(0.009)
ATT_{post}	-0.022	-0.033	-0.040	-0.078	-0.004	-0.009
	(0.034)	(0.027)	(0.069)	(0.067)	(0.008)	(0.012)
$\text{ATT}_{post_{k=3,4,5}}$	-0.041	-0.027	-0.032	-0.132	-0.007	-0.014
	(0.045)	(0.037)	(0.102)	(0.106)	(0.010)	(0.017)
Panel B: Including Covariates of Distance to Main Cities and Rurality Index						
ATT_{pre}	0.036	-0.025	-0.087	0.002	0.005	0.002
	(0.025)	(0.020)	(0.054)	(0.038)	(0.007)	(0.009)
ATT_{post}	-0.061*	-0.016	-0.058	-0.106	0.013	0.015
	(0.036)	(0.033)	(0.094)	(0.072)	(0.013)	(0.014)
$\text{ATT}_{post_{k=3,4,5}}$	-0.082	0.002	0.001	-0.138	0.007	0.005
	(0.054)	(0.041)	(0.143)	(0.112)	(0.015)	(0.020)

Note: This table reports the robustness to the inclusion of covariates of wages and working hours estimates using survey data (Figures 7 and 7). The dependent variables are the logarithm of sectorial labor incomes in Columns 1 to 4, and of working hours in Column 5 and 6. Panel B controls for the share of rural population and the driving times to Bogotá and Medellín. We incorporate these covariates to the model using the outcome regression DiD estimator based on ordinary least squares. Regressions were weighted with the working-age population of the 2005 census, and standard errors clustered at the municipality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: GEIH 2010-2018.

Table D.7: **Average *C&S* estimates of tax ratios using controls**

	(1)	(2)	(3)	(4)
	Ratio of tax revenues		Ratio of industry & commerce tax revenues	
	Without controls	With controls	Without controls	With controls
ATT_{pre}	2.321	-0.453	0.718	0.354
	(4.003)	(4.244)	(1.927)	(2.778)
ATT_{post}	10.138**	16.447***	6.291**	9.127**
	(4.740)	(5.779)	(2.939)	(3.616)
$\text{ATT}_{post_{k=3,4,5}}$	14.091*	25.776***	10.967**	15.799**
	(7.374)	(9.010)	(4.741)	(5.673)

Note: This table reports the robustness of the two main tax outcomes (total tax revenues and revenues from the industry and commerce tax) to the inclusion of covariates. The dependent variable in Columns 1 and 2 is the ratio of taxes over all revenues (including taxes and central government transfers), and it corresponds to the ratio of industry and commerce tax revenues over all revenues in Columns 3 and 4. While estimates in Columns 1 and 3 do not include covariates (they match those in Table C.13), Columns 2 and 4 control for the share of rural population and the driving times to Bogotá and Medellín. We incorporate these covariates to the model using the outcome regression DiD estimator based on ordinary least squares. Regressions were weighted using municipality's revenues in 2010, and standard errors clustered at the municipality level. The coefficients represent percentage points changes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.Source: DNP 2010-2018.

Table D.8: **Average *C&S* estimates of formal employment ratios by sector excluding weights**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: Weighted (Baseline)					
ATT_{pre}	-0.417	-0.045*	0.030	-0.074	-0.100
	(0.294)	(0.025)	(0.090)	(0.205)	(0.211)
ATT_{post}	1.742***	0.108**	-0.290*	0.877***	0.670
	(0.608)	(0.044)	(0.158)	(0.296)	(0.474)
$ATT_{post_{k=3,4,5}}$	3.147***	0.170**	-0.397*	1.647***	1.013
	(1.049)	(0.066)	(0.228)	(0.426)	(0.756)
Panel B: Unweighted					
ATT_{pre}	-0.099	-0.073***	0.057	0.085	0.074
	(0.299)	(0.019)	(0.078)	(0.160)	(0.212)
ATT_{post}	1.901***	0.115***	-0.359***	0.930***	0.469
	(0.643)	(0.035)	(0.126)	(0.275)	(0.384)
$ATT_{post_{k=3,4,5}}$	3.499***	0.185***	-0.378**	1.743***	0.772
	(1.081)	(0.054)	(0.189)	(0.425)	(0.642)

Note: This table reports the robustness of the formal employment estimates when we do not weight them. Panel A reports the results from the regressions that use administrative data and are weighted with the 2005-census working age population (see Table C.8), while estimates in Panel B do not include weights. The dependent variable is formal employment by the given sector relative to the working-age population according to the 2005 census. The CHR refers to commerce, hotels and restaurants. Primary and secondary are manufacturing, construction, agriculture, and mining. Standard errors clustered at the municipality level. The coefficients represent percentage points changes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: PILA 2010-2018.

Table D.9: **Average *C&S* estimates of formal employment ratio using outcomes in logs**

	(1) Formal employment ratio	(2) Logarithm of formal employment
ATT_{pre}	0.124	-0.003
	(0.329)	(0.018)
ATT_{post}	1.329*	0.049
	(0.678)	(0.038)
$ATT_{post_{k=3,4,5}}$	2.526**	0.102
	(1.178)	(0.062)

Note: This table reports the robustness of the total formal employment estimates to using a log-transformation on the dependent variable, instead of the preferred ratio-interpreted outcome. The dependent variable in Column 1 is formal employment from administrative data relative to the working-age population according to the 2005 census, whereas it corresponds to the natural logarithm of formal employment in Column 2. Both regressions were weighted with the working-age population of the 2005 census and control for the share of rural population and driving times to Bogotá and Medellín using the outcome regression DiD estimator based on ordinary least squares. Standard errors clustered at the municipality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: PILA 2010-2018.

Table D.10: **Average *C&S* estimates of tax ratios using outcomes in logs**

	(1)	(2)	(3)	(4)
	Tax revenues		Industry & commerce tax revenues	
	Ratio	Logarithm	Ratio	Logarithm
ATT_{pre}	-0.453	-0.003	0.354	0.007
	(4.244)	(0.028)	(2.778)	(0.042)
ATT_{post}	16.447***	0.103***	9.127**	0.059
	(5.779)	(0.038)	(3.616)	(0.062)
$\text{ATT}_{post_{k=3,4,5}}$	25.776***	0.163***	15.799***	0.112
	(9.010)	(0.056)	(5.673)	(0.090)

Note: This table reports the robustness of the two main tax outcomes (total tax revenues and revenues from the industry and commerce tax) to using a log-transformation instead of the preferred ratios. The dependent variables in Columns 1 and 3 are the total tax revenues and the industry and commerce tax revenues respectively, both divided by all the revenues (taxes and central government transfers). Columns 2 and 4 have as outcomes the natural logarithm of total and industry and commerce tax revenues, respectively. Both regressions were weighted using municipality's revenues in 2010 and control for the share of rural population and driving times to Bogotá and Medellín using the outcome regression DiD estimator based on ordinary least squares. Standard errors clustered at the municipality level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: DNP 2010-2018.

Table D.11: **Average *C&S* estimates of formal employment ratios excluding metropolitan areas and accounting for the immigration shock**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: Baseline					
ATT$_{pre}$	-0.417	-0.045*	0.030	-0.074	-0.100
	(0.294)	(0.025)	(0.090)	(0.205)	(0.211)
ATT$_{post}$	1.742***	0.108**	-0.290*	0.877***	0.670
	(0.608)	(0.044)	(0.158)	(0.296)	(0.474)
ATT$_{post_{k=3,4,5}}$	3.147***	0.170**	-0.397*	1.647***	1.013
	(1.049)	(0.066)	(0.228)	(0.426)	(0.756)
Panel B: Excluding metro-area municipalities					
ATT$_{pre}$	-0.547*	-0.055**	0.069	-0.144	-0.194
	(0.318)	(0.027)	(0.101)	(0.229)	(0.240)
ATT$_{post}$	1.732**	0.133***	-0.293*	0.857***	0.591
	(0.676)	(0.045)	(0.168)	(0.309)	(0.524)
ATT$_{post_{k=3,4,5}}$	3.098***	0.206***	-0.405	1.636***	0.871
	(1.154)	(0.066)	(0.250)	(0.442)	(0.844)
Panel C: Including the Venezuelan immigration shock					
ATT$_{pre}$	-0.444	-0.048*	0.033	-0.099	-0.140
	(0.291)	(0.025)	(0.091)	(0.204)	(0.212)
ATT$_{post}$	1.753***	0.114***	-0.288*	0.889***	0.664
	(0.609)	(0.042)	(0.160)	(0.303)	(0.490)
ATT$_{post_{k=3,4,5}}$	3.145***	0.175***	-0.393*	1.657***	1.016
	(1.050)	(0.065)	(0.229)	(0.431)	(0.766)

Note: This table reports the robustness of the formal employment estimates to the exclusion of municipalities belonging to a metropolitan area and the inclusion of a covariate on the Venezuelan migration shock. The baseline estimates correspond to those from administrative data (Table C.8). The CHR refers to commerce, hotels, and restaurants. Primary and secondary are manufacturing, construction, agriculture, and mining. Panel B excludes 19 municipalities classified by DANE as part of metropolitan areas, and Panel C controls for the share of migrants using the outcome regression DiD estimator based on ordinary least squares. We construct the share of migrants based on the 2018 census and define it as the ratio of employed migrants from Venezuela in a municipality over the total employed population between 18 and 64 years. Regressions were weighted with the working-age population of the 2005 census, and standard errors clustered at the municipality level. The coefficients represent percentage points changes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: GEIH 2010-2018.