The Impact of Hard Discount Stores on Local Labor Markets: Evidence from Colombia*

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

October 27, 2024

Abstract

Hard discount stores have reshaped the retail sector by selling low-cost products. While this business model has gained market shares in many countries, how it affects the labor market is unclear. To fill this gap, we study the impact of discount stores on local labor markets in Colombia, where these stores had a rapid, staggered geographic expansion. Our results show that discount stores boost local formal employment, especially in retail, manufacturing, and agriculture, suggesting significant spillover effects from retail to other sectors. Consistent with this finding, we also document increases in local tax revenues from manufacturing and commerce activities.

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, Miguel Talamas, Jan Stuhler and Gabriel Ulyssea, as well as participants at Seminario CEDE (Universidad de los Andes), the UC3M Applied Reading Group, LACEA-LAMES 2023, Red Investigadores de Economía and Banco de la República seminar for the 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 or official policy of Banco de la República or its Board of Directors. Lukas Delgado-Prieto acknowledges the financial support from the Ministry of Science and Innovation in Spain through research grant PDI2019-108144GB-I00. Any errors are exclusively the responsibility of the authors.

†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). They have gained significant market shares in many countries, such as 26% of the food and grocery market in Germany (Euromonitor International, 2023a; MarketLine Industry Profiles, 2023). Notable European chains like Aldi and Lidl lead this category, ranking among the top ten global retailers and being the largest hard discount chains (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, but they also increase competition within the retail sector, potentially causing job losses among incumbent firms. Furthermore, there can be employment spillovers onto other sectors, for instance, due to upstream supply chain effects (de Paula and Scheinkman, 2010).

In this paper, we examine the staggered expansion of the leading hard discount chains in Colombia. Unlike the labor markets in the United States or Europe, Colombia's labor market has a particular structure, with over half of the workforce employed informally. In the retail sector, where these stores operate, employment is predominantly informal (55% of the workforce does not contribute to social security, and 83% of the businesses lack legal registration). However, hard discount chains are formal firms that must comply with labor and tax regulations, which means they must hire only formal workers, comply with minimum wage laws, and ensure their suppliers are formal firms.

The retail sector in Colombia has undergone a major transformation since the launch of the first hard discount chain in 2009. Through sizeable investments, the leading chains have established more than 4,000 stores nationwide as of 2022 (Euromonitor International, 2023c). Their rapid expansion translated into a major sales growth of 235% between 2017 and 2022, overtaking hypermarkets in market share by 2020 and supermarkets by 2021 (Euromonitor International, 2023c). Furthermore, social security records indicate that the top three hard discount chains employed over 26,000 formal workers by 2019. Following the model of European hard discounters, these stores have opened in residential neighborhoods and city centers, offering a limited assortment of predominantly own-brand products at low prices.

To assess the impact of HDS on local labor markets, we assemble a unique data set combining administrative records of formal employment and subsidized social protection beneficiaries, labor force survey data, municipal tax collection, 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. 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, we leverage the rapid and staggered rollout of HDS across the country in an event study design, using municipalities not yet affected by HDS entry as a control group.

Our identification strategy does not assume that the store locations are exogenous. Instead, we argue that the *timing* of the first hard discount store opening, excluding the largest capital cities, is unrelated to local employment or wage trends. Therefore, by comparing cohorts of intermediate-sized municipalities where hard discount chains opened earlier to those where they opened later, we can identify the effect of HDS on local labor markets. We provide suggestive evidence that the *timing* assumption holds in our scenario, as the treated and control groups exhibit similar trends in various outcomes, such as employment, wages, working hours, and taxes before the first store opens.

Using this research design, we capture the dynamic treatment effects of opening HDS on formal and informal labor markets at the local level, based on the method of Callaway and Sant'Anna (2021).[1] We find three main results in this paper. First, the opening of a hard discount store in a municipality leads to a 1.7 percentage points (pp) increase in local formal employment using administrative records and a 2.9 pp increase using survey data. This growth is primarily driven by employment in retail, manufacturing, and agricultural sectors.[2] One plausible explanation for these inter-sectoral spillovers is that HDS boost the local demand for local industries to produce more inputs for the goods sold in their stores, as indicated by the largest hard discounter (La República, 2022). However, this effect takes several years to materialize because HDS are relatively small in Colombia, averaging ten workers per store, and thus require time and investments to reach a scale that impacts the local labor market.

Second, using a proxy of tax collection, we find a positive impact on the local economy, as the share of taxes in total public revenues increases by an average of 7.5%. This increase is primarily driven by the industry and commerce tax, supporting the idea that HDS have positive spillovers on 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 the closure of informal neighborhood shops as they open more stores, given the lack of precise data to test this hypothesis. We also observe a negative trend in the labor income of informal retailers in the later years of treatment, which aligns with findings from a similar shock in the Mexican retail sector (Talamas Marcos, 2024). Taken together, these findings suggest that a relevant channel to foster the formalization of local economies in developing countries is local private investments.

Related Literature. 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, the largest retailer

[1] Formal labor markets refer to all workers who contribute to the social security system.

[2] The estimates are robust to controls such as distance to the capital city or baseline local economic structure. As we explain later, the sample of municipalities is not the same across data sources, causing the estimates to differ.

in the world. The entry of Walmart into 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 affects other local retail stores that struggle to compete with its lower prices (Basker, 2005; Neumark et al., 2008; Wiltshire, 2021; 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, 2023). Leveraging the staggered roll-out of FCs with areas not yet treated as control (similar to our empirical strategy), these studies find negative effects on retail employment but positive employment spillovers in other sectors: Chava et al. (2023) in transportation and warehousing, whereas Cunningham (2023) in tradable services.[3] Similarly, Greenstone et al. (2010) studies the agglomeration spillovers of the arrival of "Million Dollar Plants" in the US by comparing counties where they finally open 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 that sell goods locally within 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 knowledge, this is the first paper estimating the impact of HDS in developing countries across both the formal and informal labor markets.[4]

Another strand of the literature we contribute to focuses on the effect of increased competition in the retail sector. There is evidence of increased consumer surplus when hard discounters enter towns, as their average prices are lower than those of traditional supermarkets, inducing price reductions in competing stores (Atkin et al., 2018; Busso and Galiani, 2019; Hausman and Leibtag, 2007). Although we do not directly test this result, given the context of Colombia, HDS entry might affect consumption choices as consumers substitute shopping in neighborhood shops (not necessarily cheaper than supermarkets) for hard discounters, potentially increasing their consumer surplus. However, as neighborhood shops may experience decreased profits, we also test for changes in informal employment and earnings to capture potential neighborhood shops' closure. Finally, we aim to identify spillover effects of HDS entry on employment in other industries, such as manufacturing and agriculture, contributing to the literature on upstream supply chain effects, as hard discounters have incentives to source their products from local formal suppliers (de Paula and

[3] Chava et al. (2023) documents that the opening of FCs between 2010 and 2016 decreases the labor income of retail workers and sales of brick-and-mortar retail stores. Cunningham (2023) studies the expansion of Amazon FCs between 2010 and 2021 to find a positive effect on overall employment and wages.

[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.

Scheinkman, 2010; Gerard et al., 2023; Rios and Setharam, 2018).

To analyze the impact of hard discounters' entry on incumbent retailers, we draw on several facts to shed light on the underlying mechanisms. First, discount chains are more efficient than traditional supermarkets, with higher sales per square meter and lower operating costs, allowing them to rapidly gain market share. However, 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. 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 multistop shopping behavior among consumers. Second, our findings show that informal employment in the retail sector does not significantly decline after the expansion of hard discount chains, suggesting neighborhood shops continue to operate despite increased competition. The margin of adjustment may occur then through earnings, complementing evidence from Mexico (Talamas Marcos, 2024), where an additional convenience store entry in a neighborhood adversely affects the creation of neighborhood shops. Still, existing neighborhood shops survive by leveraging their comparative advantage in supplying fresh products and reducing costs by shrinking inventories.[5] Neighborhood shops often have low fixed costs by avoiding tax registration and operating within the owners' houses (Ramos-Menchelli and Sverdlin-Lisker, 2023), providing more flexibility in response to increased competition.

This paper is structured as follows. The next section provides the institutional context and information on 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, first on employment, then taxes, and finally on labor income and working hours. We conclude in section 7.

2 Institutional Context

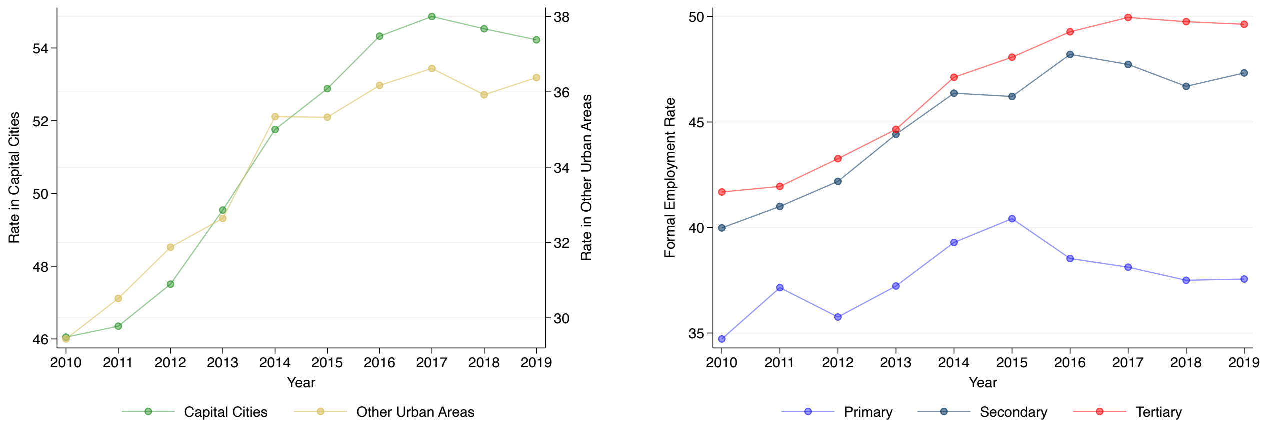
In Colombia, informal jobs are the prevalent type of employment. However, in recent years, there has been a noticeable upward trend in formal employment rates, both in capital cities and other urban areas (see Figure 1). There are several factors contributing to this phenomenon. One key factor is the 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).[6] In this context, we want to study the impact of the arrival of HDS on

[5] 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.

[6] 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, 2024).

local formal employment, as they only hire workers formally and demand inputs only from legally registered suppliers, and determine if this contributes to the national observed increase in formal employment.[7]

Figure 1: Formal employment rates



(a) By geographic location

(b) By industry

Note: This figure shows the formality rate by geographic location and by industry groups using 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.

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 local shops (mainly informal), supermarkets, and hypermarkets. According to market data from Euromonitor International (2023c), in 2017, small local shops accounted for 52% of the retail grocery sales, while large retailers were responsible for around 23% of the sector's income.

Small neighborhood shops are mostly informal businesses that offer a limited supply of essential goods, primarily in residential areas.[8] They tend to create a close relationship with their customers, even offering informal credit to them (Talamas Marcos, 2024).[9] 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

[7]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.

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

[9]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.

would fit into lower-income households' daily budgets. On the other end, traditional supermarkets are large formal firms with infrastructures similar to those of large retailers in the US, where customers can find goods ranging from fresh produce 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 a deeper analysis of the latter using data from the Micro Businesses Survey in Appendix A.

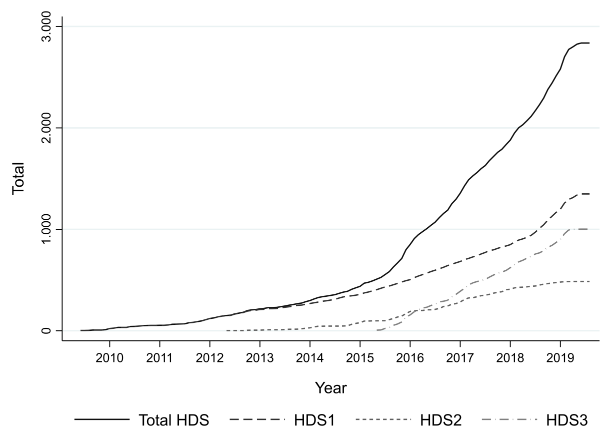
The supermarkets and hypermarkets segment, a relevant source of formal employment, has been mainly dominated by three firms: *Grupo Éxito* (with their brands *Éxito* and *Carulla*), *Supertiendas & Droguerías Olímpica* (with their brands *Olímpica* & *Sao*), and *Cencosud S.A.*, (with the *Jumbo* & *Metro* brands). This format was introduced in Colombia at the end of the 1940s by José Carulla Vidal (Silva Guerra, 2011), after descovering that supermarket stores such as those of the company *Sumesa* were booming in Mexico. During the 1950s, the concept of supermarkets expanded in Colombia, appearing with new chains and new owners in the main cities of the country (Grupo Exito, 2015). According to market data from Euromonitor International (2023c), Colombian supermarkets have an aggregate market size of around 4 billion dollars, with approximately 2,200 outlets. The average selling space is around 750 square meters per store, and the typical store sells about 2 million dollars annually (US$2,450 per square meter).

By the end of the 1990s, hypermarkets appeared in the national retail sector with a precise socioeconomic segmentation among customers of the different supermarkets operating in the country. Certain chains were known to have lower prices that targeted low-to-mid-income clients. However, these lower-price stores did not operate under the hard-discount concept but by selling lower-quality products. There are about 210 hypermarkets in Colombia, mainly in the largest urban centers, and their sales also sum up to 4 billion dollars (Euromonitor International, 2023c). The average selling space per outlet is 4,700 square meters, with the typical store selling about 19 million dollars per year (US$4,020 per sq meter.)

The first hard discount chain (*D1*) opened in Colombia in 2009, following the German model used by Aldi. The second hard discount chain (*Ara*) opened in 2012, and the third one (*Justo & Bueno*) opened in 2016. HDS are smaller in area than traditional supermarkets (HDS have, on average, 250 to 300 square meters), and they reduce their operational costs through different strategies. The most important ones are having efficient distribution chains based on a limited portfolio of goods, low investment in ads, exhibiting products in shipping boxes, and smaller staff. The typical Colombian HDS is also smaller than a supermarket in annual sales. However, they sell around US$3,472 per square meter, which is larger than traditional supermarkets' sales per square meter by 41%. (Euromonitor International, 2023c).

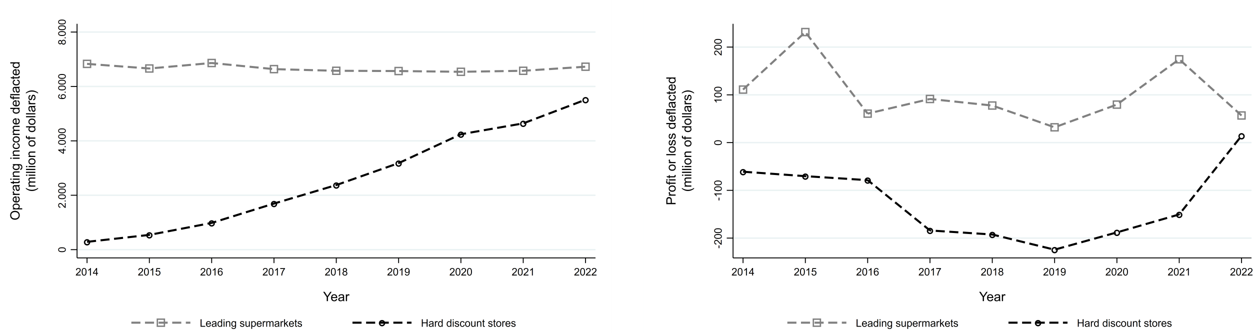
Figure 2 shows the evolution of HDS from 2010 to 2019, both at the aggregate level and by chains. Over ten years, hard discount chains in Colombia rapidly expanded nationwide. Among the three leading chains, they opened almost 3,000 stores with a stock of approximately 26,000 employees in 408 municipalities of Colombia (out of 1,103 municipalities). Figure 3 then compares supermarkets' and hard discounters' operating income and profits. Historically, supermarkets have led the retail market by operating income while making significant profits. However, since the entry of hard discount chains, the gap in operating income between them has decreased substantially over time.[10] Regarding profits, the main two hard discounters (as the third chain exited the market in 2021) did not manage to earn positive profits until 2022, showing how much they have over-invested in the country to expand and gain market shares. In contrast, the three main supermarkets only started to show a decrease in profit in 2022.

Figure 2: Hard discount stores 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.

Figure 3: 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).

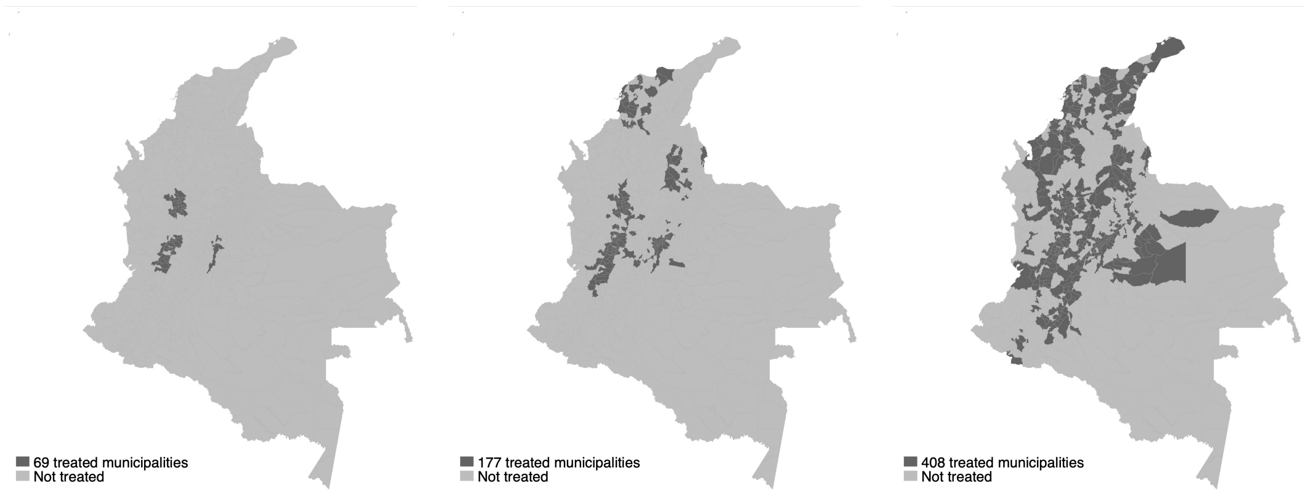
[10] Even the leading hard discounter in the market already has a larger operating income than the leading supermarket in the country.

The expansion of hard discount chains not only increased the number of stores from which retail consumers may buy but also represented a sizeable investment shock for the municipalities where the chains decided to open. For instance, in an interview in 2018, Ara's manager stated that its overall investment for opening 500 stores was more than 400 million euros, an average investment per store of around 0.8 million euros, or approximately 0.9 million USD (Portafolio, 2018; Morante, 2018). Similarly, D1's manager announced in 2020 that it had invested 123 million USD in opening 800 stores (approximately 154,000 dollars per store, González Bell (2020)). These investments are equivalent to 1.9 times the median annual local tax revenue of the municipalities included in our sample for the case of Ara, and 36% for the case of D1. They also represent about 19% of these municipalities' median annual investment expenditures for the case of Ara, while for D1, represent around 4%.[11] Many municipalities have multiple hard discount chains and stores, increasing the investment received by several orders of magnitude.

Figure 4 shows that the expansion began in the country's central region. Then, it expanded to the Caribbean region and the southern part of the country in a staggered fashion, partly because hard discount chains focused on different regions to expand its operations more aggressively. According to market data starting in 2017 (when discounters already had more than 1,700 outlets), hard discounters' sales grew by 235% between 2017 and 2022, compared to 29% in the aggregate grocery retail sector. The number of stores grew by 133% and the 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), and there are large-scale investment plans to increase the number of HDS in the following years in Colombia (La República, 2023), highlighting how policy-relevant it is to study the impact of these stores on local labor market outcomes.

[11] The median annual tax revenue of the municipalities in our sample during the 2010–2019 period is 425,674 USD, while the median annual investment expenditure is 4,150,126 USD, according to the annual panel of Colombian municipalities (Acevedo and Bornacelly Olivella, 2014). We corroborate the chains' investment values using public financial data from *Supersociedades*. The net cash flow from investing activities per open store was computed to be roughly 1.3 million USD for Ara, 47,000 USD for D1, and 90,000 USD for Justo & Bueno.

Figure 4: Location of hard discount stores



(a) 2013 **(b) 2016** **(c) 2019**

Note: This figure shows the geographic expansion of Colombia's three main hard discount chains using stores' stock in 2013, 2016, and 2019. Source: Authors' calculations using public location data from hard discounters' websites.

In addition to the geographical pattern, hard discount chains decide where to locate their stores based on pre-existing municipality characteristics, such as potential market size and industrial composition. 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), poverty level (indicating economic conditions), distance to the department's capital, and the formal employment share in the retail, non-retail commerce (including hotels and restaurants), and services industries. We estimate the probability of HDS entering to 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 0.529), with population size being the most relevant one.

3 Data

This paper uses multiple data sources to capture both formal and informal employment. First, we use the employer-employee matched administrative records on social security contributions (PILA, by its acronym in Spanish) from 2010 to 2019. 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 final estimation sample, we exclude capital cities to avoid confounding the effect with labor market trends of the largest areas that presumably are much less affected by the relatively small shock of HDS openings. Due to this restriction, we dropped 83% of observations from PILA, which accounts for approximately 7 million observations per month, as capital cities concentrate most of the formal jobs in the country. Moreover, we exclude all municipalities that do not have HDS up to 2019, as the labor market in those municipalities evolves differently than those that eventually opened an HDS. Our final estimation sample from PILA has, on average, 1.6 million observations per month and includes 372 out of the 1120 municipalities of the country, where 38% of the total population is located.[12]

As in PILA we only observe the formal sector, we also use administrative records to approximate the size of the informal sector through the census of beneficiaries of subsidized social protection (SISBEN), which allows us to approximate the stock of low-income informal workers at the municipality level using the same sample of municipalities as in PILA. Although this source contains the universe of low-income individuals in the country, including those who can be out of the labor force, it also captures informality as individuals who receive social protection subsidies, by definition, must not be formal workers because they lose immediate access to the subsidies. Thus, SISBEN is a good proxy of the size of informality in a given municipality.

As a second source aiming to capture informality, we use GEIH, a monthly cross-sectional household survey that covers approximately 240,000 households per year. GEIH is the Colombian 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 data is reduced to 191 from the 372 municipalities we observe in PILA. However, GEIH does allow us to characterize the informal labor market as we use the survey questions about workers' contribution to the social security system to capture wages and employment of informal and formal workers.[13] 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 dropping capital cities.[14]

The data from the GEIH survey has the limitation that it is not representative at the municipality level, even though these municipalities are the largest ones surveyed in GEIH, besides capital cities. Still, as our empirical analysis aggregates similar municipalities into treatment cohorts (that range in size from 5 to 35

[12] 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 ISIC (International Standard Industrial Classification) 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.

[13] We only consider employed workers with positive labor income in the analysis.

[14] The geolocation information of the GEIH survey is not publicly available, so we used specialized data centers of DANE to access this information.

municipalities depending on the cohort, see Table 2), our identifying variation is not at the municipality level, but at the cohort level, which alleviates concerns about the statistical noise induced from measurement error. Furthermore, we aggregate monthly to yearly information, which also helps to reduce noise. Importantly, our results on formal employment with GEIH and PILA are similar, which supports that the data we are using from GEIH is accurate.

The primary outcomes we use in PILA are formal employment, defined as the rate of formal employment in area l over the working-age population of l that we obtain from the 2005 Census, and formal wages, defined as the logarithm of average real monthly wages in l.[15] From the GEIH survey, we use similar outcomes and fix the denominator again with the census of 2005 to have comparable results.[16] We calculate average wages and working hours within each sector and take the logarithm transformation.

To capture broader effects on the local economy, we use data from *Operaciones Efectivas de Caja*, collected by *Departamento Nacional de Planeación* (DNP), 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.

Finally, for our identification strategy, we measure the year of the arrival of a hard discount chain into a municipality. We relied 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.

3.1 Descriptive Statistics

Administrative Records (PILA). Figure 5 panel (a) shows that formal employment in the services industry grew steadily from 2010 to 2019, mainly due to favorable economic conditions that fostered the creation of formal jobs. Despite this positive result, around half of the working population in Colombia still works informally and lacks access to the pension system. Informality rates are even higher outside the main capital cities, as shown using the GEIH survey.

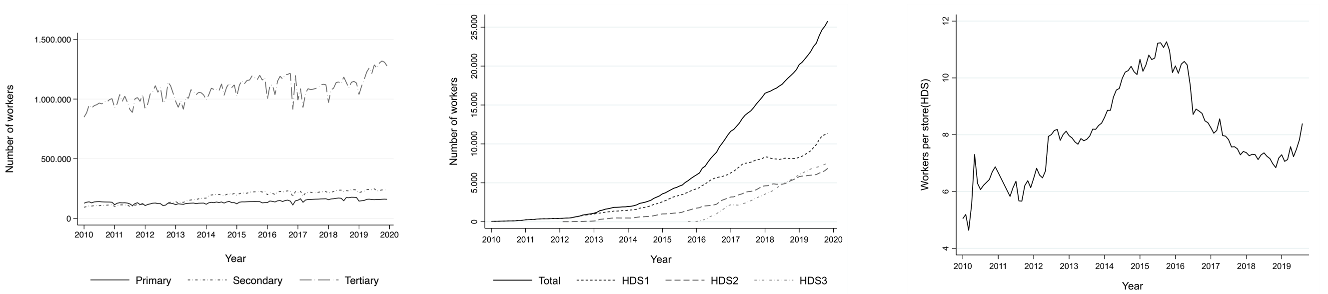
[15] We include only wages from workers who reported working on a full-time schedule

[16] We do not use a logarithm transformation for employment outcomes given their considerable variation among treated and not-yet-treated groups in the baseline year, which can lead to biases in the coefficients (McConnell, 2023).

Regarding formal employment in hard discount chains, they went from employing less than 500 formal workers in 2012 to almost 26,000 workers in 2019, illustrating the rapid expansion of these chains (see Figure 5 panel (b)). Among these workers, 86% of them are hired as full-time employees (i.e., with a formal work contract for 30 days of the month).[17] Lastly, Figure 5 panel (c) shows how the average number of employees per store grew until 2016, and then it started to decrease up to 2019. There are, on average, 9.8 employees per store.

Next, we analyze the individual work histories of HDS employees using the PILA. From 2010 to 2019, we identified 57,963 individuals who have worked at one of the three leading hard discount chains at any point. Of these, approximately 16% are first-time formal workers, meaning they appear for the first time in PILA as employees of HDS, while 31% of them worked previously at other formal firms before switching to HDS. Among first-time formal workers, 55% are women, with an average age of 24, and 47% have part-time work contracts during the first quarter of employment. Conversely, those who switch from other firms tend to be older, averaging 30 years old, with 60% being men, and more than 75% of them come from jobs in the wholesale and retail industry and services sectors. Overall, the data suggests that hard discount chains play a role in providing opportunities for young first-time formal workers.

Figure 5: Descriptive statistics on formal employment at HDS using PILA



(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.

Appendix Table C.1 shows descriptive statistics for the treated municipalities in our estimation sample. First, we observe that in 2011, the average formal employment in treated municipalities was 6,781. This number decreased over time as hard discount chains expanded to more municipalities smaller in size. In

[17] Due to Colombian legislation, full-time employment is more prevalent in the formal sector than part-time employment, as it was relatively more costly to hire a part-time worker than a full-time one. However, part-time employment has grown faster than full-time employment since 2014 due to changes in the regulation that allowed for weekly formal labor contracts (de la Parra et al., 2024).

our sample, at least 90% of the formal employment in the treated municipalities does not belong to the commerce, hotels, and restaurant industries. Still, the share of workers in that sector increased over time, going from 5.1% in 2011 to 10% in 2018. Regarding the share of workers who work independently or as employees who earn the monthly minimum wage, both groups have similar shares.

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.[18] 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 very 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 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.

[18] Appendix Table C.3 further shows descriptive statistics on wages and working hours for the formal and informal sectors.

Table 1: Descriptive statistics for the estimation sample using GEIH

	Treated				Not treated yet			
	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

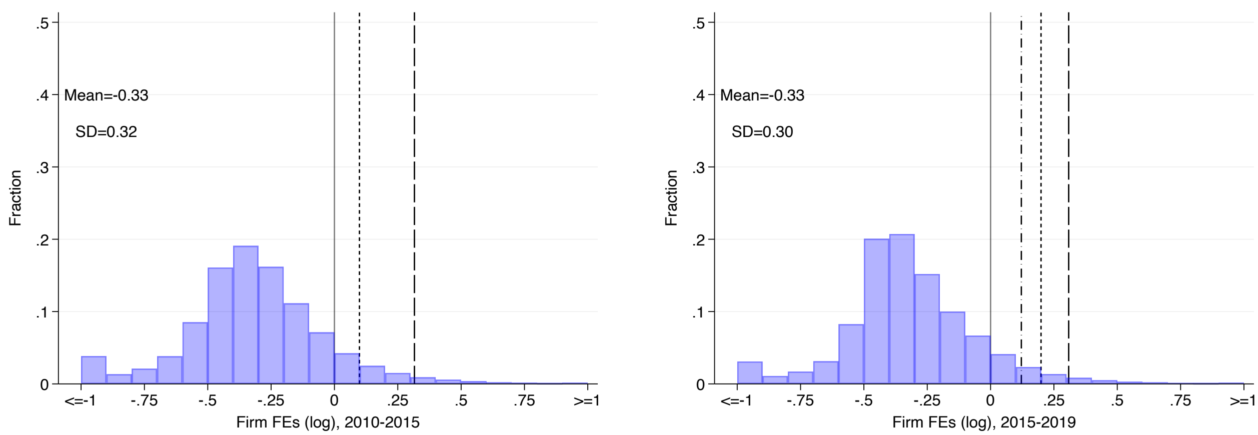
To understand in more detail how hard discount chains operate relative to other formal retail firms, we estimate the canonical model from Abowd et al. (1999) (AKM hereafter) to disentangle the firms' contributions to formal wages. With the output from this estimation, we answer how much HDS pay, on average, their workers relative to other firms once worker characteristics are netted out. We estimate our AKM model using the standard specification:

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

Here, $log(wages_{it})$ are total log 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 capture the latter component, workers must move across different firms. Therefore, we restrict to the largest set of workers and firms connected by workers' mobility. 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) and year FEs. We split the estimation period in two: from 2010 to 2015, when the first HDS opened and employed mainly educated workers, and from 2015 to 2019, when they grew extensively across the country and employed more blue-collar workers. In this exercise, all the firm FEs are relative to the largest retail store in the country.

Figure 6 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. The intuition of this framework is that if a worker moves from the reference firm to a discount chain, it would have an average positive gain in wages equal to the value of the dashed line depending on the chain it moves. If it were the opposite movement, it would decrease their wages equal to the respective dashed line. Lastly, note that the firm's pay premiums are measured at the national level for all workers, including managers and low-skilled workers.[19]

Figure 6: Distribution of Firm FEs



(a) Firm FEs, 2010–2015 **(b) Firm FEs, 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 or 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 methodology of Chava et al. (2023) and Cunningham (2023) studying the local impacts of Amazon FCs. Table 2 illustrates the staggered introduction of hard discount chains

[19] 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 6.

within our sample of municipalities, comprising 372 in PILA and 191 in GEIH.[20] 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: Staggered treatment adoption by year

	Municipalities (PILA)	Share	Municipalities (GEIH)	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 it is potentially incorrect 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 ATT using weights that can be hard to interpret, 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.[21]

The intuition of the C&S estimator is that it breaks down all the sets of possible comparisons into the two-period and two-group (2x2) framework to estimate multiple $\text{ATT}(g, t)$ 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. Therefore, if the treatment started in 2015, then $g = 2015$.[22] More concretely, they are estimated as follows:

[20] There are no instances of municipalities where the first hard discount store that opens subsequently closes, so all municipalities in our sample remain treated over time.

[21] We do not select Sun and Abraham (2021) as the main estimator due to the availability of control groups. Only the never-treated and last-treated cohorts are available as control groups with that estimator. Yet, arguably, it is more challenging to assume that local economic trends in cities that HDS never opened or opened in the latest period would behave similarly to the treated cities without the treatment.

[22] 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 $t - g > 5$.

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

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)[23]:

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

Next, as a benchmark for the C&S coefficients in the event study figures, we estimate the traditional TWFE regression in certain outcomes. The model takes the following form:[24]

$$y_{lt} = \alpha + \sum_{k=-5}^{5} \beta_k \mathbf{1}\{k = t - g\} + \gamma_l + \gamma_t + \epsilon_{lt}. \tag{4}$$

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 unit 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 selected municipalities that have not yet been treated as a control group.[25]

Identification Assumption. The main assumption required in this setup is the standard parallel trend assumption (PTA). It establishes that treated and control units would have evolved similarly in terms of their outcomes in the absence of HDS openings. As our control group is the municipalities that have not yet been treated, we do not assert that HDS openings are exogenous to local economic trends. Instead, we argue that unobserved trends do not determine the *timing* of HDS 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 more 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 controls for constant observed and unobserved characteristics associated with these

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

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

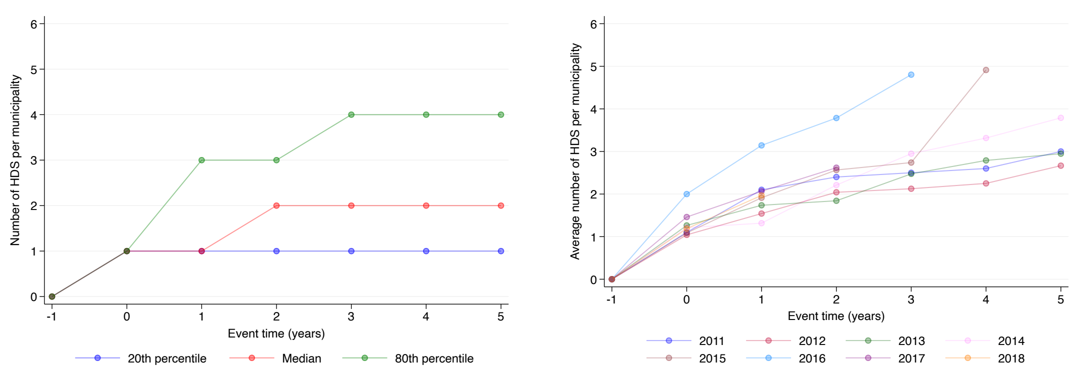
[25]To achieve identification using not yet treated as a control group, we bin or accumulate the leads as $k \leq -5$ and the lags as $k \geq 5$ (Borusyak et al., 2024).

factors.

Moreover, we show there are no differential employment or wage 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 concern with this assumption is that HDS might open first in places better connected to capital cities. Consequently, municipalities that experience early openings might be more affected by the spillovers from time-varying shocks in the capital cities than those with later openings. To address this concern, we include the distance to the capital city as a control in our estimations and show that results do not change significantly. Additionally, we assume that 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 a less significant concern.

Another important aspect to evaluate is the intensity of the treatment across municipalities. If cohorts are heterogeneous in the number of stores they have, interpreting the overall ATT becomes less clear. Figure 7a 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 Figure 7b).

Figure 7: 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.

Lastly, we discuss the choice of using the not-yet-treated municipalities as our main control group instead of the never-treated (i.e., municipalities that, as of 2020, had not received a hard discount chain). 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, closer to the department's capital, less impoverished, and had a higher proportion of their formal employment in the non-retail commerce, hotels, and restaurants sectors compared to the never treated municipalities. 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 groups (see Appendix Figure C.1).

Apart from the cross-sectional differences in pre-treatment characteristics, the labor markets of never-treated municipalities exhibit 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.[26] Even with the restricted never-treated group, the pre-trends persist, although they are smaller in magnitude. Altogether, these findings indicate that never-treated municipalities do not serve as an appropriate control group.

5 Results

5.1 Employment

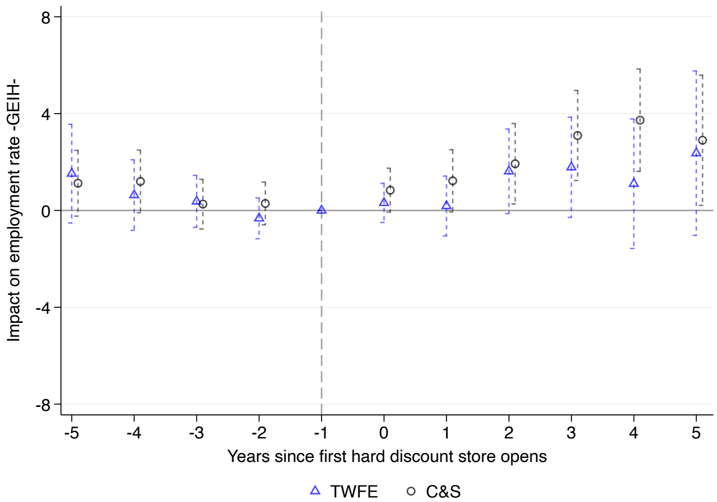
To start, we evaluate the overall effect of HDS on standard labor market outcomes. Figure 8 illustrates the dynamic treatment effects on the employment rate in local labor markets. Importantly, treated and not-yet-treated cohorts of municipalities do not exhibit differing trends before the treatment, suggesting that the arrival of HDS is unrelated to local employment rates among these groups.

If a store opening represents a labor demand shock, then local employment should be positively affected. Consistent with this hypothesis, the post-treatment estimates are positive and grow over time, with most of them becoming significant in the later periods. It is worth noting that the TWFE estimates are consistently smaller than the C&S ones, suggesting that heterogeneous treatment effects might introduce biases in the former. Therefore, we are going to show only C&S estimates hereafter.

[26] The model identifies the closest never-treated municipality for each treated municipality in the GEIH panel using a propensity score constructed with the covariates of Table C.2–Column 7.

Averaging the six post-treatment years, HDS openings boosted the local employment rate by approximately 2.3 pp (see Figure 8). Additionally, Appendix Table C.4 shows a large, though not statistically significant, decrease in the unemployment rate and a significant reduction in the inactivity rate following store openings.

Figure 8: Event study estimates on employment rate



Note: The dependent variable is the yearly employment rate. Regressions were weighted with local employment in 2010. Observed treated municipalities are 191. The panel is not balanced for certain outcomes, so 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. Source: GEIH 2010-2018.

The general impact on the labor market masks heterogeneous responses from specific types of employment that HDS 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 support 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. Their coefficients differ mainly because the sample of municipalities in PILA (372) is twice as large as the one in GEIH (191).[27]

Figure 9 shows that formal employment follows a positive trend, taking a few years to materialize, according to both administrative and survey data. Thus, the rise in total employment is primarily driven by the increase in formal employment. Averaging all six post-treatment periods, formal employment increases by around 1.7 pp using administrative records and 2.9 pp using survey data. These results are robust to the inclusion of controls such as distance to the capital city or baseline local economic structure. Hence, if the *timing* of opening a hard discount store is unrelated to the local employment trends, this indicates a

[27]Difference in coefficients may also arise because PILA slightly underreported employment data (firms with several establishments across the country often report the information for their workers in the municipality where the headquarters is located). And because the GEIH survey is not representative to all the municipalities considered in the analysis.

positive causal impact on formal employment in affected municipalities.[28] Altogether, discount stores help to formalize local economies and generate employment growth by reducing the number of individuals out of the labor force (see Appendix Figure C.2).

To benchmark the estimates, the weighted mean of the formal employment share before the treatment is 28.3% in the sample of treated municipalities in GEIH.[29] Hence, an increase of 2.9 pp is equivalent to a 10.3% increase in this share. Moreover, from the baseline event to the farthest post-treatment event we measured, the weighted mean of the formal employment share grew around 13.8 pp (from 28.3% to 42.1%). So, the entry of HDS would explain around one-fifth of the observed increase in formality in these municipalities.

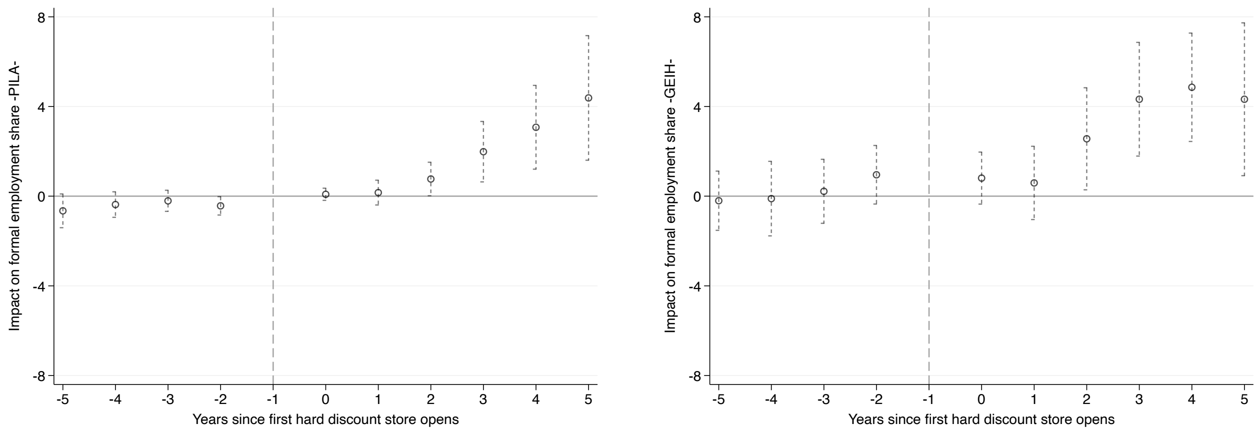
Since these coefficients measure the overall impact from the first opening, we also examine the number of stores in each municipality to quantify an intensive margin effect. We use the number of HDS in the municipality as the outcome in our main specification and find that, averaging all post-treatment 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 share by 0.7 pp (=1.7/2.4) using administrative data and by 1.2 pp (=2.9/2.4) using survey data. In absolute terms, each store generates around 138 direct and indirect formal jobs (95% $CI = [26, 250]$).[30]

[28] In Appendix Figure D.2, we show the estimates of PILA using the estimation sample of GEIH where we find a smaller coefficient (1.2 pp).

[29] The weights we use are the working-age population in the 2005 census, the same as in the main regression.

[30] 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 average number of stores (2.4).

Figure 9: Event study estimates on formal employment



(a) Formal employment (PILA) **(b) Formal employment (GEIH)**

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 in a given municipality. Regressions were weighted with the 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. Source: PILA 2010-2018 in August, and GEIH 2010-2018.

We then investigate the drivers behind the increase in formal employment. Figures 10a and 10b show that the primary and secondary sectors explain the growth in formal employment, according to both administrative and survey data (see Appendix Tables C.6 and C.5 for detailed point-estimates). A plausible explanation for this is that as hard discount chains increase demand for products and inputs from local formal suppliers, they boost formal employment in these industries. Specifically, manufacturing employment rises by 0.9 pp and agricultural employment by 0.6 pp, industries that are key producers of goods sold by these stores (see Table 3). Unfortunately, we cannot directly link how much hard discounters buy from the local producers because we lack input-output data. Still, the discounters have reported that they primarily source from local suppliers.[31] For instance, the largest hard discount chain reported that around 80% of their goods in 2020 were domestically produced (La República, 2022). Moreover, according to a discount chain manager, many local suppliers have been expanding to meet the increased demand from these stores (Portafolio, 2018).

In addition, HDS may generate spillover effects in these industries through other channels, such as income effects, large-scale investments, or agglomeration effects. First, as consumers spend less on groceries after HDS arrive, they can increase their spending on other goods. Second, large-scale investments by hard discount chains can indirectly benefit other local businesses, notably in construction and real estate. Third,

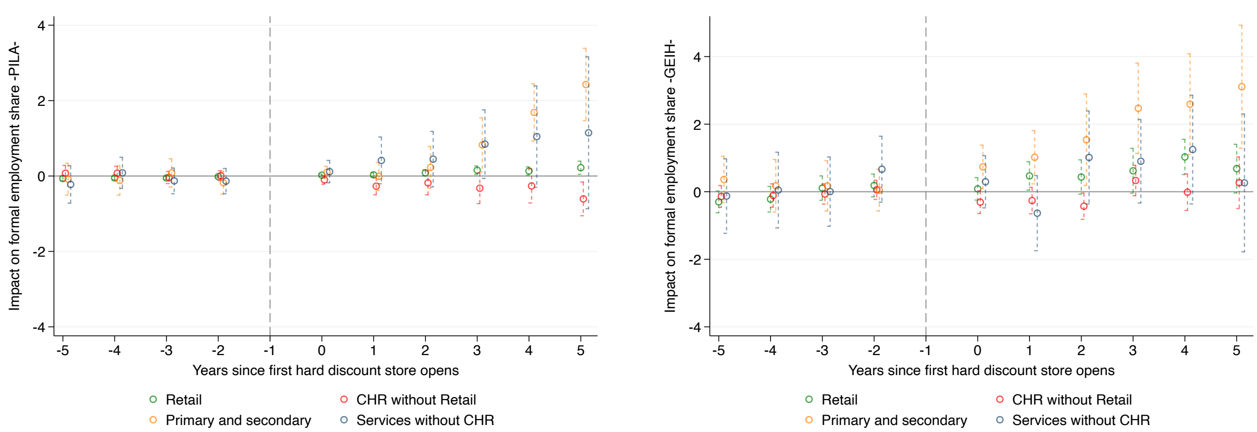
[31] These findings align with those of de Paula and Scheinkman (2010), who discusses the "business formality chain" that occurs when value-added taxation incentivizes firms to purchase more from formal firms upstream in the supply chain.

HDS may create agglomeration effects that benefit nearby businesses (Evensen et al., 2023).

It is challenging to pinpoint which channel explains our findings, yet the results by sector suggest that the upstream supply chain effects are one of the most important. Two sets of results support this conclusion. On the one hand, if the income effects were predominant, there is no apparent reason why it would substantially benefit the formal agriculture and manufacturing sectors. Given the context of these municipalities, where informality plays a crucial role, we should expect the informal sector to benefit from the income effect. However, as shown later, we do not find such positive effects, at least as measured by employment or wages. On the other hand, if the investment channel were the most impactful, we would expect the results to be driven by the construction sector. While Table 3 shows a positive and statistically significant (at the 10% level) employment effect in construction, the effects in agriculture and manufacturing are quantitatively larger. Given the more substantial impacts in agriculture and manufacturing compared to construction or other sectors, the evidence strongly suggests that supply chain factors are the primary drivers behind HDS impacts.

Lastly, we observe a significant increase in employment in the retail sector, according to both 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, due to increased competition within the formal retail sector.

Figure 10: Event study estimates on formal employment share by industry



(a) Formal employment (PILA) **(b) Formal employment (GEIH)**

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 in a given municipality. 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 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. Source: PILA 2010-2018 in August, and GEIH 2010-2018.

Table 3: Average $C\&S$ estimates of formal employment by subsector using GEIH

	(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
Clusters	191	191	191	191
Mean pre-treatment	28.3	2.1	4.5	1.3

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

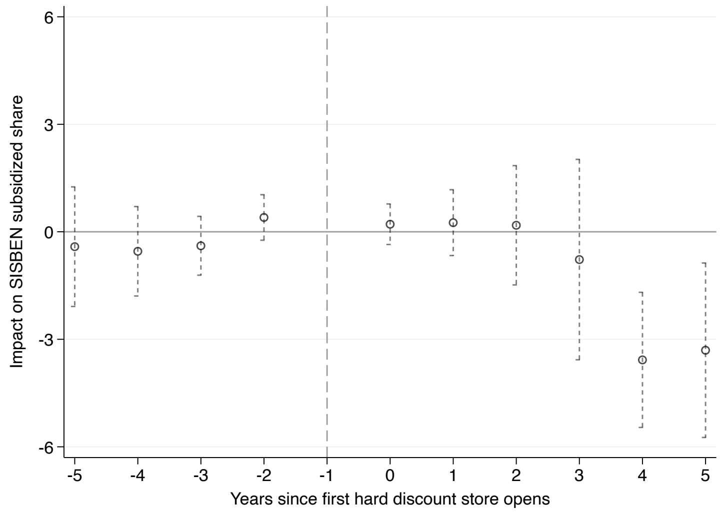
Note: The dependent variable is formal employment by the given sector over the working-age population according to the 2005 census in a given municipality. Regressions were weighted with the local working-age population in the 2005 census. Observed treated municipalities are 191. Standard errors are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

Given that the main competitors of HDS are local neighborhood shops, which are primarily informal, the next outcome we analyze is the impact on informal employment. For this, we rely on survey data from the GEIH, as the informal sector is only observed in the survey. Appendix Figure D.3 shows that the impact on informal employment is not robust, as the coefficient ranges from negative to positive values when we add controls such as distance to the capital. Therefore, the impact of HDS on informal employment is imprecise. Regarding sectoral impacts, the increase in competition between HDS and the informal retail sector does not reduce informal retail employment. In fact, the point estimate is positive across different specifications (see Appendix Table D.4). Furthermore, the coefficients on informal employment across other sectors vary substantially but are too noisy to describe a clear pattern (see Appendix Table C.7).

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 informal sector. In Colombia, almost universal access to the health system is granted, either by subsidies or contributions. Formal workers contribute to the system, while those who are subsidized usually do not have formal employment and are likely working informally or not working at all. Using this measure of beneficiaries of subsidized social protection as a proxy for informal employment, we show in Figure 11 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.[32] Overall, the main impact of HDS is a substantial shift in the workforce composition, with formal employment increasing more than informal employment in the treated areas.

[32] Appendix Figure D.4 shows the results are robust to including different controls.

Figure 11: Event study estimates on subsidized social protection beneficiaries (SISBEN)



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 in a given municipality. 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. Source: Health Ministry 2010-2018.

5.2 Taxes

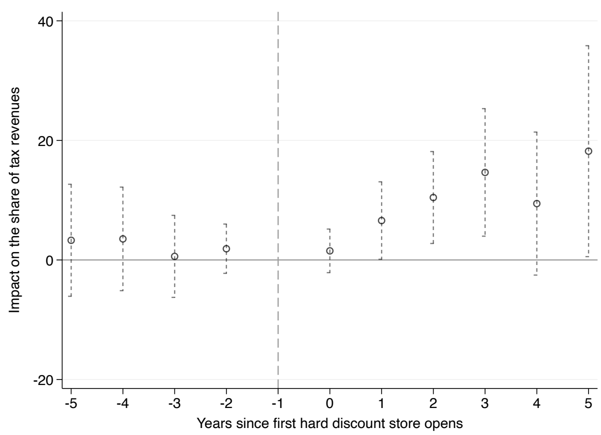
The arrival of hard discount stores can boost economic activity in their municipalities through direct job creation, spillover effects in other sectors, and increased property value in areas where the stores are located. We use municipality-level tax revenues as a proxy for economic activity to determine if HDS entry and their significant effects on employment lead to broader impacts on local economies. Although the main taxes in Colombia are collected nationally, such as the income tax, social security contributions, and VAT, other significant taxes in Colombia are collected at the municipality level. These include property taxes, industry and commerce taxes, and vehicle and gasoline taxes.[33] Industry and commerce taxes, in particular, reflect the performance of industrial, commercial, or service activity registered in the municipality. Therefore, even though hard discount 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 12 shows that prior to the arrival of stores, the growth of local taxes as a share of total revenues was similar between the municipalities that opened early and those that opened later. Then, with the arrival of discount stores, the growth rate of all local tax revenue increases by 10.1 pp on average, or 7.5% relative to the pre-treatment mean, across all post-treatment periods. Most of this increase is attributed to revenues collected from the industry and commerce taxes (see Appendix Table C.8). Hence, the arrival of stores

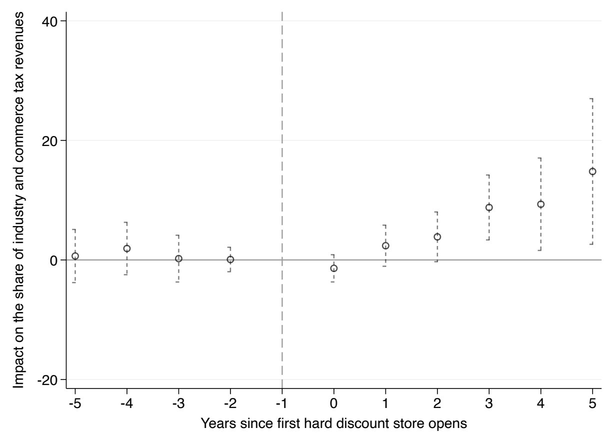
[33]The owners of low-income residential households are not required to pay property tax.

substantially boosts tax revenues for local governments, primarily driven by growth in manufacturing and commercial activities. This result aligns with the literature on upstream supply chain effects of business formality, indicating HDS purveyors must be formal to claim all tax benefits and discounts provided by the law (de Paula and Scheinkman, 2010; Gerard et al., 2023; Rios and Setharam, 2018). Increasing local revenues may increase local public spending, which can stimulate the labor market. This suggests another particular channel where the arrival of HDS can increase local employment.

Figure 12: Event study estimates on local taxes share 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) in a given municipality. 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 are weighted using the municipality's share of revenues in 2010. Observed treated municipalities are 371. Standard errors are clustered at the municipality level. 90% confidence interval. Source: DNP, 2010-2018.

5.3 Labor Income and Working Hours

Other margins of adjustment, like labor income or working hours, may be used to respond to the increase in labor demand or competition from these stores. First, we focus on the impacts on labor income across the formal and informal sectors. On the one hand, Figure 13a shows a slight drop in the first post-treatment years in the labor income of informal workers, while the labor income of formal workers is rather stable. As shown with the AKM model, 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 more likely they hire minimum wage workers, which decreases average formal wages. In the PILA, formal wages slightly decrease after the entry of hard discount chains, but only in the latest post-treatment periods (see Appendix Figure C.3).

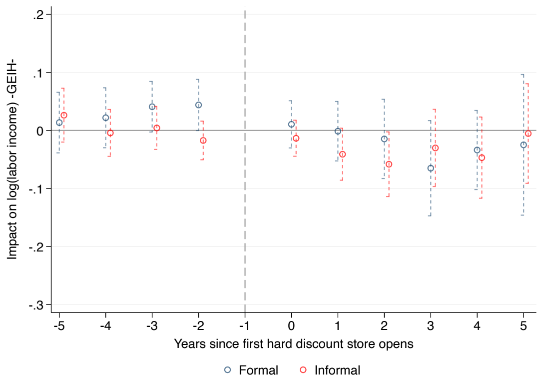
Next, we measure the impact on the labor income of retail workers by sector. For informal workers, this serves as an indirect measure of the earnings of neighborhood shop owners, while for formal workers, it reflects regular wages. Figure 13b shows a negative trend in the labor income of informal workers in the latest post-treatment periods, indicating that the increase in competition from HDS may affect them more through reduced earnings rather than employment.[34] 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 ($ATT_{post} = -7.8\%$).[35]

Last, we focus on the logarithm of average working hours across the informal and formal sectors as it is another margin of adjustment. 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 13c shows that there are no significant changes in these outcomes after the first entry. Thus, we cast aside the hypothesis that informal or formal workers adjust to the arrival of HDS by working more hours.

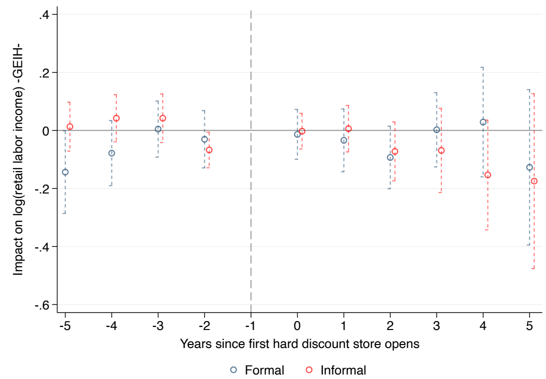
[34] The estimated effect on the labor income of formal retail workers might seem contradictory to the pay premiums of hard discounters in subsection 3.2. However, the sample differs (e.g., the AKM sample includes capital cities), and HDS paying a national wage premium does not necessarily increase average wages at the sector level in treated municipalities.

[35] However, when we add the distance control, the point estimates approach zero (see Appendix Figure D.6).

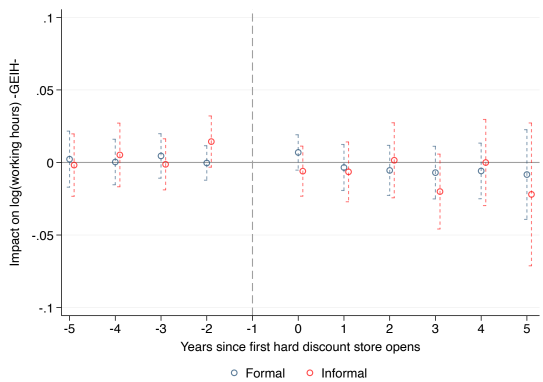
Figure 13: **Event study estimates on labor income and working hours**



(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 in a given municipality. 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. Source: GEIH 2010-2018.

6 Robustness Checks

In this section, we test the robustness of our results by including different controls and using other estimators. First, we use the distance in kilometers from each municipality to the department's capital as a control to account for the degree of connectivity to the largest city, enabling a cleaner comparison of municipalities within cohorts of treatment. Second, we include the share of municipal tertiary employment in 2010 as a control to adjust for the pre-existing economic structure in the service industries before the arrival of hard discount chains. Appendix Table D.2 and Panel A of Figure D.3 show the estimates of formal employment using these two controls with the survey data. Overall, the positive effects remain robust across various specifications. Specifically, the coefficient of total formal employment for the last three periods remains statistically significant when controlling for the economic structure, although its magnitude is reduced when we add the distance control. Regarding the coefficients for the primary and secondary sectors, they remain significant when using each control separately. In the final panel, we include both controls jointly, which results in a substantial increase in standard errors. Despite this, the coefficients for total formal employment and primary and secondary employment decrease but remain statistically significant at the 10% level.

Then, Appendix Table D.3 shows the results with the administrative data are robust to the robustness tests. The impact on total employment is stable across different specifications, ranging from 1.7 pp to 2 pp, and remains statistically significant. Notably, when we use the distance to the capital as a control, the impact on retail, primary, and secondary employment is slightly higher. In contrast, the coefficient on commerce, hotels, and restaurants without retail gets more negative and becomes statistically significant at the 1% level.

Regarding informal employment, Table D.4 and Panel B of Figure D.3 show that the estimates are volatile when including controls, ranging from positive to negative values, but all are insignificant. Meanwhile, Figure D.4 shows that the impacts on the rate of subsidized health over time are more robust, mainly because the sample of municipalities almost doubles the one in the GEIH survey. Regarding taxes, controlling for the distance to the department's capital reduces the estimates when the dependent variable is all tax revenues, while when it is industry and commerce tax revenues, it remains robust, as shown in Figure D.5). Moving to Panel A of Table D.5, we replicate the ATTs from Figure 13 of wages and Figure 13 of working hours. Again, all estimates are insignificant across the formal and informal sectors, and the addition of controls does not notably alter the size or statistical significance of the effects. This is further illustrated in Figure D.6, which shows robustness to the effects on the income of informal retail workers (Figure 13). Lastly, Table D.6 demonstrates the robustness of the effects on labor market rates after including controls. The estimates for employment and inactivity rates remain stable across specifications, while the decrease in the unemployment rate becomes significant at the 5% level only when controlling for distance to the department capital.

To provide additional robustness to our main findings, we relax the assumption of parallel pre-trends and perform sensitivity analysis on the average coefficients following Rambachan and Roth (2023). For instance, 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, according to Rambachan and Roth (2023). 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 D.1). The value does not reach violations close to 1, partly due to the number of periods we analyze (six periods), as they can increase the confidence intervals and decrease the breakdown values (Rambachan and Roth, 2023).

Another potential concern revolves around 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 control of distance. To mitigate this concern, we conducted additional analysis excluding the six largest metropolitan areas.[36] This further exclusion drops 19 municipalities from our main sample. Panel C of Appendix Table D.7 reports that the treatment effects do not vary substantially in magnitude. Another concern is that using cohorts that aggregate different municipalities from GEIH, which are not uniquely representative, might introduce biases due to the use of survey weights. To address this, we exclude them from constructing the local outcomes and still find a positive impact on formal employment and an insignificant one on informal employment, in line with our previous results (see Appendix Figure D.7).

6.1 Immigration Shocks

We also analyze whether the Venezuelan mass migration to Colombia, which started during the expansion of hard discount chains, can impact the results. First, we argue that we exclude capital cities, where most migrants are located, from the analysis and focus on intermediate cities. Still, to address this more directly, we substitute the denominator of the dependent variables from the working-age population of the 2005 census with the one in the 2018 census. If the timing of the arrival of hard discount chains is unrelated to the immigration shock, then the adjusted dependent variables would take into account the population growth, and the coefficient would not change its sign, only its scale.

Panel B of Appendix Tables D.7 and D.8 reports the estimates of formal employment based on the PILA and GEIH using the working-age population of the 2018 census as the denominator. There are no major changes in the direction or statistical significance of the effects.[37] Still, there is a reduction in the magnitude of the estimates that changes the interpretation of the estimates. For instance, the effect in PILA using the ratio of formal employment to 2005 working age is 1.7 pp while the ratio of formal employment to 2018 working age is 1.1 pp. In GEIH, using the ratio of formal employment to 2005 working age yields a coefficient of 2.9 pp, while the ratio of formal employment to 2018 working age population yields a coefficient of 1.8 pp. This reduction is mainly due to the downward adjustment of the dependent variables.[38]

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

[37]Only the retail employment in PILA and total and primary employment in GEIH become significant at the 10% level.

[38]The pre-treatment mean of the dependent variables reduces substantially, as the share of formal employment lowers from

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, mainly in retail, manufacturing, and agricultural industries. This suggests that the entry of hard discount chains into a municipality generates spillover effects from retail to other sectors. A consistent explanation is the increased demand for goods from local formal firms within the discounters' supply chain, thereby hiring 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.

Regarding the informal retail sector, our evidence 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 a negative trend after the entry of hard discounters. We rationalize that the adjustment occurs through earnings rather than employment, as neighborhood shops are typically small entrepreneurial activities run mostly by their owner, with only 16% employing additional workers.

Furthermore, we find suggestive evidence that the positive impacts on employment contribute to aggregate effects in the local economies. Post-HDS entry, the share 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 promoting local private investments that can spur the formalization of local economies on top of other 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.

16 to 10.4 in PILA and from 28.3 to 18.4 in GEIH.

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*.

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.

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

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

Busso, M. and Galiani, S. (2019). The causal effect of competition on prices and quality: Evidence from a field experiment. *American Economic Journal: Applied Economics*, 11(1):33–56.

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. (2023). 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*, Forthcoming.

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

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

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., Naritomi, J., Seibold, A., and Zulian, B. (2023). Two-tier tax systems and firms: 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.

Hausman, J. and Leibtag, E. (2007). Consumer benefits from increased competition in shopping outlets: Measuring the effect of wal-mart. *Journal of Applied Econometrics*, 22(7):1157–1177.

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. Technical report, National Bureau of Economic Research.

La República (2022). Prácticamente el 80% de nuestros productos en Tiendas D1 son de producción nacional. URL: https://www.larepublica.co/empresas/practicamente-80-de-nuestros-productos-en-tiendas-d1-son-de-produccion-nacional-3073792.

La República (2023). Tiendas D1, Tiendas Ara e Ísimo planean la apertura de 830 puntos durante este año. URL: https://www.larepublica.co/empresas/tiendas-d1-tiendas-ara-e-isimo-planean-la-apertura-de-830-puntos-durante-este-ano-3619732.

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.

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.

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*.

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.

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.

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. (2024). Surviving competition: Neighborhood shops vs. convenience chains. *The Review of Economic Studies*, Forthcoming.

Wiltshire, J. C. (2021). 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 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

[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.

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).

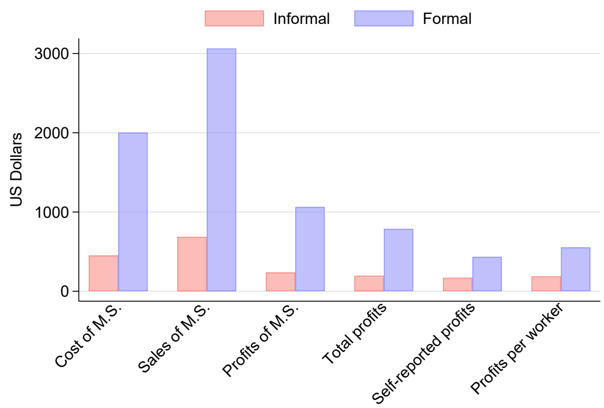
[A.4] Stores with an updated register at the Chamber of Commerce are also more likely to be classified as formal using other definitions: about 30% of them report income, VAT, or commerce tax (compared to only 1.2% of the informal stores).

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

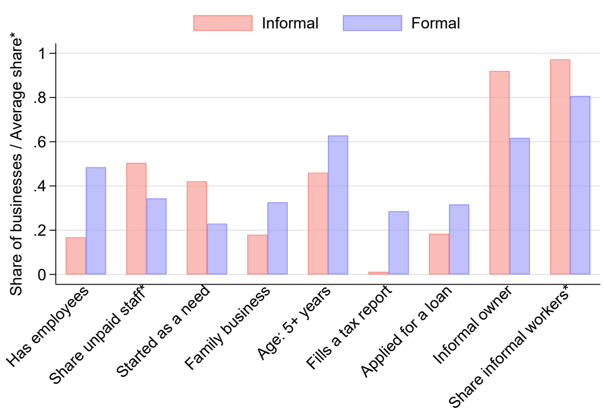
	Total sample mean (S.D.)	Formal mean	Informal mean
Panel A: Basic business characteristics			
Has employees	0.222	0.485	0.168
	(0.415)	(0.500)	(0.373)
Total business staff	1.320	1.832	1.215
	(0.744)	(1.210)	(0.549)
Share of unpaid staff in personnel	0.444	0.344	0.504
	(0.486)	(0.456)	(0.493)
Owner started 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 $q = 3$ 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.

[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).

[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.

C Additional Results

Table C.1: Descriptive statistics for the estimation sample using PILA

	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%
Total employment	67,819	244,286	841,385	1,122,119	776,792	814,881	467,375	106,767
Average total 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: Selection into treatment by hard discount chain

	HDS1		HDS2		HDS3		Any chain	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(2005 Population)	0.217***	0.237***	0.154***	0.148***	0.209***	0.233***	0.275***	0.276***
	(0.017)	(0.013)	(0.034)	(0.036)	(0.028)	(0.022)	(0.019)	(0.016)
Log(Distance to department capital)	-0.006	-0.037*	-0.033	-0.008	-0.013	-0.019	-0.021	-0.034
	(0.022)	(0.019)	(0.028)	(0.023)	(0.023)	(0.019)	(0.022)	(0.024)
2005 Unsatisfied basic needs index	-0.008***	-0.005***	-0.003	-0.003**	-0.005***	-0.006***	-0.007***	-0.007***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2010 Share of retail formal employment	-0.003	-0.001	-0.002	-0.002	0.001	0.001	-0.004	-0.003
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
2010 Share of tertiary formal employment	-0.002**	-0.001	-0.000	0.000	-0.001	0.000	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2010 Share of non-retail CHR formal employment	0.003	0.002*	0.001	0.002**	0.001	0.002	0.004*	0.004***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Constant	-1.343***	-1.558***	-1.028***	-1.143***	-1.431***	-1.702***	-1.722***	-1.748***
	(0.203)	(0.144)	(0.262)	(0.340)	(0.261)	(0.221)	(0.230)	(0.177)
Observations	1,025	1,025	1,025	1,025	1,025	1,025	1,025	1,025
R-squared	.411	.485	.227	.407	.328	.465	.464	.529
Dep Var Mean	.265	.265	.157	.157	.221	.221	.374	.374
Dep Var SD	.442	.442	.364	.364	.415	.415	.484	.484
Department F.E	NO	YES	NO	YES	NO	YES	NO	YES

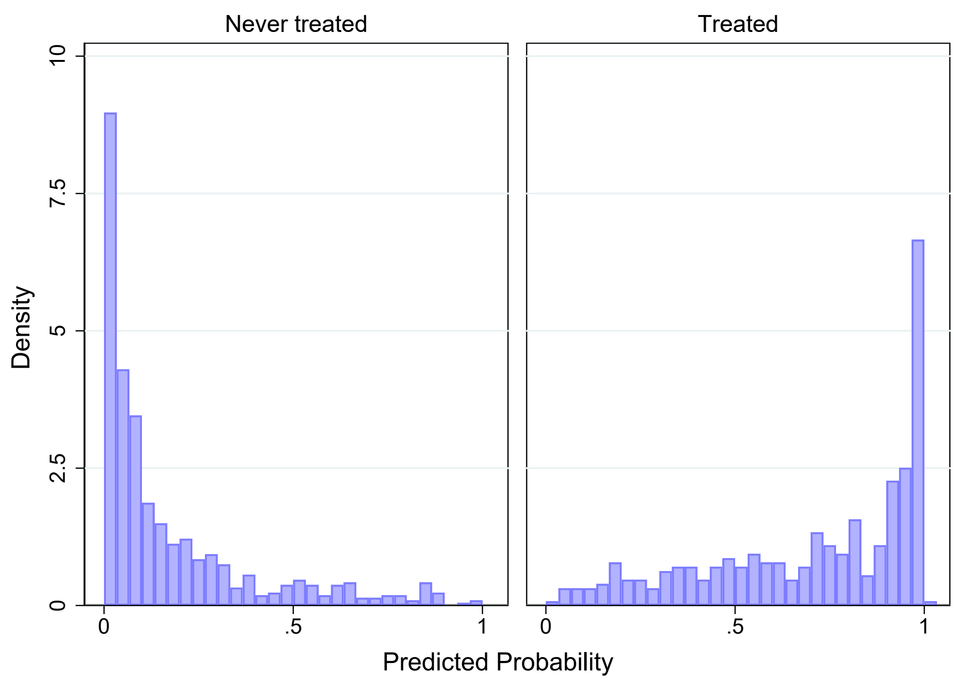
Note: This table presents the estimation results of four cross-sectional linear probability models. In these models, the dependent variable is binary, taking the value of 1 if a municipality received a Hard Discount Store from a specific chain before 2020 and 0 if the municipality was never treated. In columns 7 and 8, we employ a dummy variable to indicate whether a municipality received HDS from any chain before 2020. Columns 2, 4, 6, and 8 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: Labor income and working hours outcomes for estimation sample using GEIH

	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.1: Distribution of predicted treatment probabilities by HDS 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, which is depicted in the x-axis.

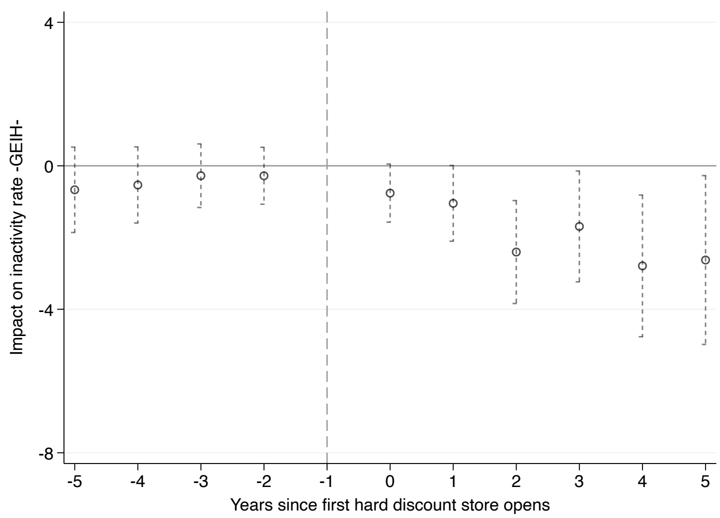
Table C.4: **Average $C\&S$ estimates of labor market rates**

	(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)
N	1,719	1,629	1,712
Clusters	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. Source: GEIH 2010-2018 in August.

Figure C.2: Event study estimates on 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. Source: GEIH 2010-2018 in August.

Table C.5: **Average $C\&S$ estimates of formal employment by sector using GEIH**

	(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
Clusters	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 by the given sector relative to the working-age population according to the 2005 census in a given municipality. 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. 90% confidence interval. Source: GEIH 2010-2018.

Table C.6: **Average $C\&S$ estimates of formal employment by sector using PILA**

	(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)
N	3,348	3,348	3,348	3,348	3,348
Clusters	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 of a given municipality. 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. Source: PILA 2010-2018 in August.

Table C.7: **Average $C\&S$ estimates of informal employment by sector**

	(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
Clusters	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 by the given sector over the 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. Regressions were weighted with local employment in 2010. Observed treated municipalities are 191. Standard errors are clustered at the municipality level. Source: GEIH 2010-2018.

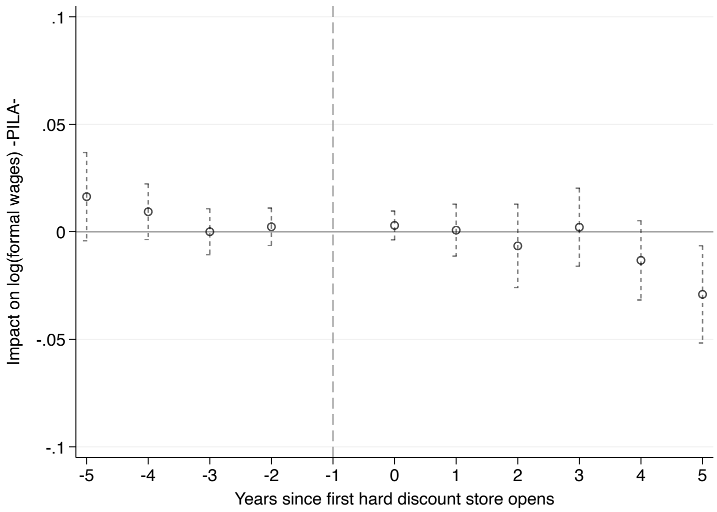
Table C.8: **Average $C\&S$ estimates of tax revenues by type**

	(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)
N	3,339	3,339	3,339	3,339	3,339	3,339
Clusters	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 specific type of taxes or revenues over all the revenues in 2010 in a given municipality. Regressions were weighted with all the revenues in 2010. Observed treated municipalities are 371. Standard errors are clustered at the municipality level. Source: DNP, 2010-2018.

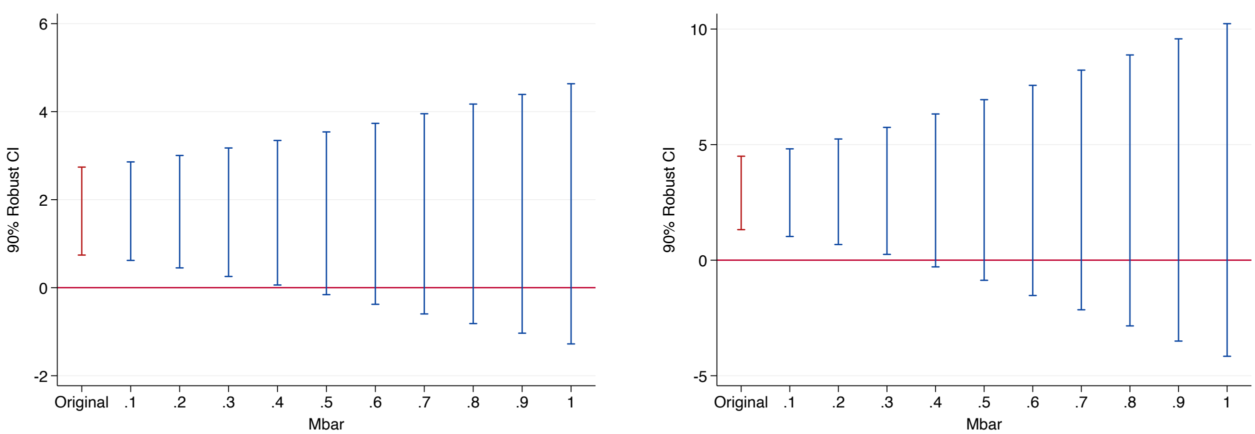
Figure C.3: Event study estimates on formal wages using PILA



Note: We use the $C\&S$ estimator. The dependent variable is the logarithm of average wages in a given municipality. 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. Source: PILA 2010-2018 in August.

D Robustness Checks

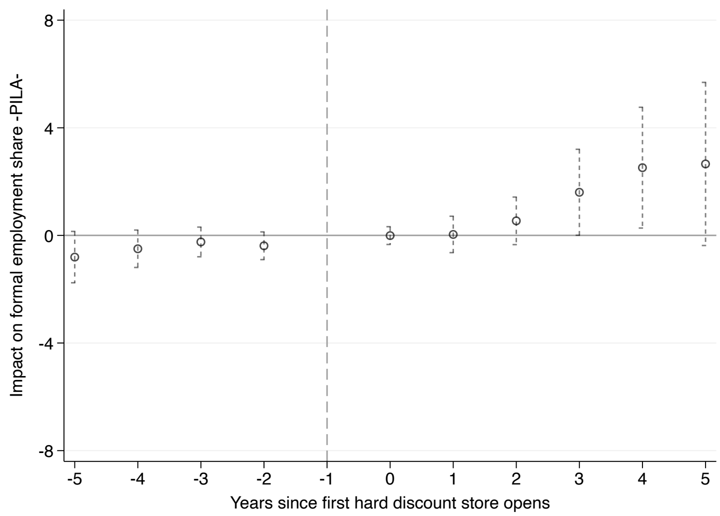
Figure D.1: Sensitivity analysis of ATT_{Post} for formal employment using PILA and GEIH



(a) Formal employment (PILA) **(b) Formal employment (GEIH)**

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 on formal employment using PILA with the municipalities observed in GEIH



Note: We use the *C&S* estimator. The dependent variable is the population of formal employment over the working-age population according to the 2005 census in a given municipality. Observed treated municipalities are 191. Standard errors are clustered at the municipality level. 90% confidence interval. Source: PILA 2010-2018 in August.

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

	(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)
N	1,719	1,674	1,713
Clusters	191	189	191
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)
N	3,770	3,442	3,728
Clusters	377	372	377
Panel C: Never treated as control group (restricted)			
ATT_{pre}	1.168*	-0.860*	-0.540
	(0.608)	(0.510)	(0.499)
ATT_{post}	0.043	0.441	-0.498
	(0.688)	(0.592)	(0.726)
N	3,080	2,825	3,048
Clusters	308	303	308
Mean pre-treatment	70.3	11.3	20.9

Note: The dependent variable is the yearly labor market rate. Regressions were weighted with local employment in 2010. The sample of never-treated municipalities in Panel C is restricted to those selected by a one-to-one propensity score matching using the model from Table C.2–Column 7. The propensity score was estimated using a logit model. Source: GEIH 2010-2018.

Table D.2: **Average $C\&S$ estimates of formal employment by sector using controls**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: Without controls					
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
Clusters	191	191	191	191	191
Panel B: Log(Distance to department capital in km)					
ATT_{pre}	0.403	-0.128	-0.000	0.161	0.371
	(0.813)	(0.178)	(0.175)	(0.411)	(0.557)
ATT_{post}	1.192	0.394	-0.209	1.197*	-0.189
	(0.954)	(0.272)	(0.272)	(0.699)	(0.833)
$ATT_{post_{k=3,4,5}}$	2.363*	0.614*	-0.142	1.910**	-0.018
	(1.350)	(0.328)	(0.332)	(0.939)	(1.221)
N	1,710	1,710	1,710	1,710	1,710
Clusters	190	190	190	190	190
Panel C: Share of tertiary sector in 2010					
ATT_{pre}	0.109	-0.067	-0.043	0.189	0.030
	(0.795)	(0.193)	(0.185)	(0.416)	(0.545)
ATT_{post}	3.136***	0.560**	-0.057	1.848***	0.784
	(1.190)	(0.249)	(0.241)	(0.630)	(0.736)
$ATT_{post_{k=3,4,5}}$	4.861***	0.762**	0.195	2.575***	1.330
	(1.623)	(0.314)	(0.327)	(0.806)	(1.055)
N	1,719	1,719	1,719	1,719	1,719
Clusters	191	191	191	191	191
Panel D: Log(Distance) and share of tertiary sector					
ATT_{pre}	0.322	-0.145	0.028	0.172	0.268
	(0.908)	(0.183)	(0.194)	(0.451)	(0.618)
ATT_{post}	1.839	0.421*	-0.030	1.265	0.183
	(1.515)	(0.245)	(0.277)	(0.854)	(1.106)
$ATT_{post_{k=3,4,5}}$	3.571*	0.646**	0.127	2.076*	0.723
	(2.127)	(0.295)	(0.325)	(1.061)	(1.608)
N	1,710	1,710	1,710	1,710	1,710
Clusters	190	190	190	190	190

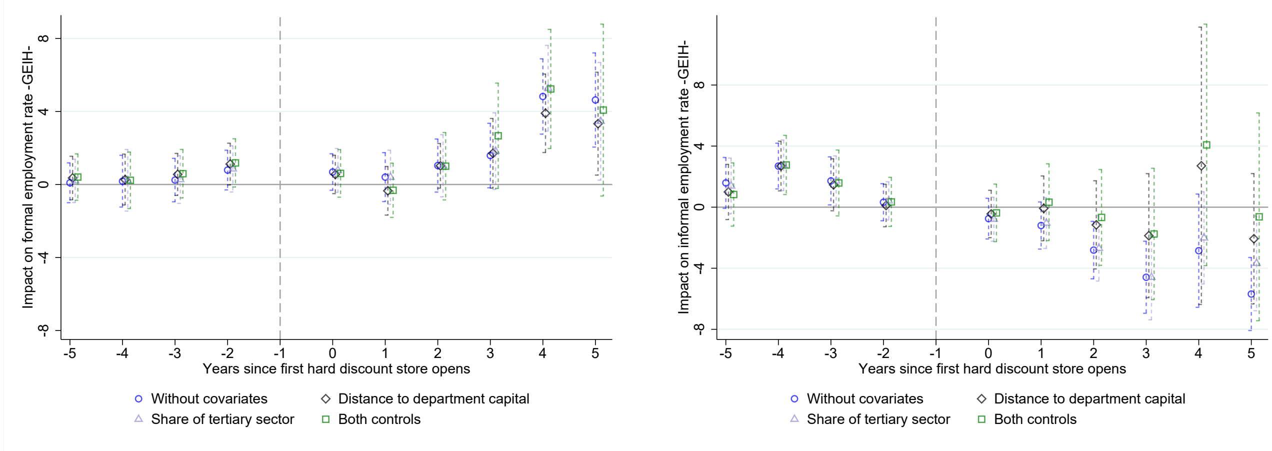
Note: This table reports the robustness of the formal employment estimates (see Table C.5) through the introduction of controls. The dependent variable is formal employment by the given sector relative to the working-age population according to the 2005 census in a given municipality. The CHR refers to commerce, hotels and restaurants. Primary and secondary are manufacturing, construction, agriculture, and mining. Panel B controls for the logarithm of the driving distance from the municipality to the department capital in kilometers, which we retrieved using Google Maps. Panel C controls for the share of tertiary sector employment in total formal municipality employment in 2010, based on PILA. Meanwhile, Panel D incorporates both covariates. 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. Source: GEIH 2010-2018.

Table D.3: **Average $C\&S$ estimates of formal employment by sector using PILA and including controls**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: Without controls					
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)
N	3,348	3,348	3,348	3,348	3,348
Clusters	372	372	372	372	372
Panel B: Log(Distance to department capital in km)					
ATT_{pre}	-0.515	-0.062**	0.078	-0.106	-0.187
	(0.315)	(0.025)	(0.086)	(0.199)	(0.223)
ATT_{post}	1.925***	0.140***	-0.451***	1.088***	0.716
	(0.657)	(0.041)	(0.113)	(0.230)	(0.469)
N	3,339	3,339	3,339	3,339	3,339
Clusters	371	371	371	371	371
Panel C: Share of tertiary sector in 2010					
ATT_{pre}	-0.476	-0.048*	0.040	-0.112	-0.121
	(0.306)	(0.025)	(0.095)	(0.218)	(0.215)
ATT_{post}	1.761***	0.108**	-0.296*	0.897***	0.638
	(0.606)	(0.047)	(0.170)	(0.300)	(0.484)
N	3,348	3,348	3,348	3,348	3,348
Clusters	372	372	372	372	372
Panel D: Log(Distance) and share of tertiary sector					
ATT_{pre}	-0.568*	-0.061**	0.080	-0.138	-0.194
	(0.320)	(0.025)	(0.085)	(0.213)	(0.223)
ATT_{post}	1.971***	0.143***	-0.455***	1.141***	0.645
	(0.641)	(0.040)	(0.118)	(0.242)	(0.456)
N	3,339	3,339	3,339	3,339	3,339
Clusters	371	371	371	371	371

Note: This table reports the robustness of the formal employment estimates using PILA (Table C.6) through the introduction of controls. The dependent variable is formal employment over the working-age population in 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 logarithm of the driving distance from the municipality to the department capital in kilometers, which we retrieved using Google Maps. Panel C controls for the share of tertiary sector employment in total formal municipality employment in 2010, calculated based on PILA data. Meanwhile, Panel D incorporates both covariates. 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. Source: PILA 2010-2018 in August.

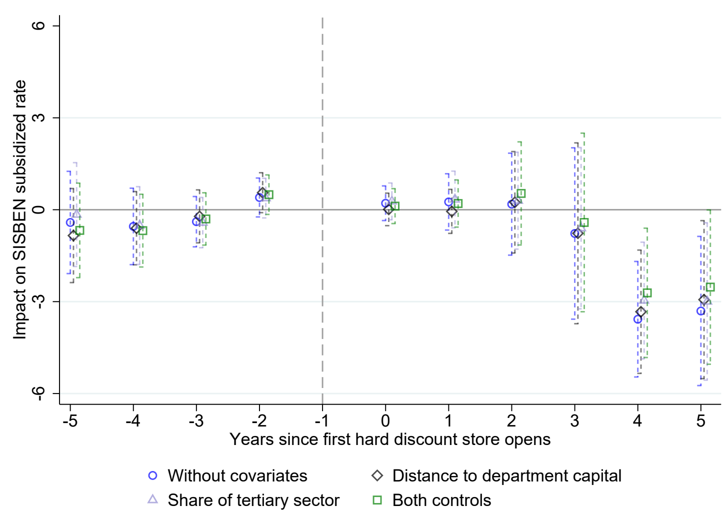
Figure D.3: Event study estimates on type of employment using controls



(a) Formal employment **(b) Informal employment**

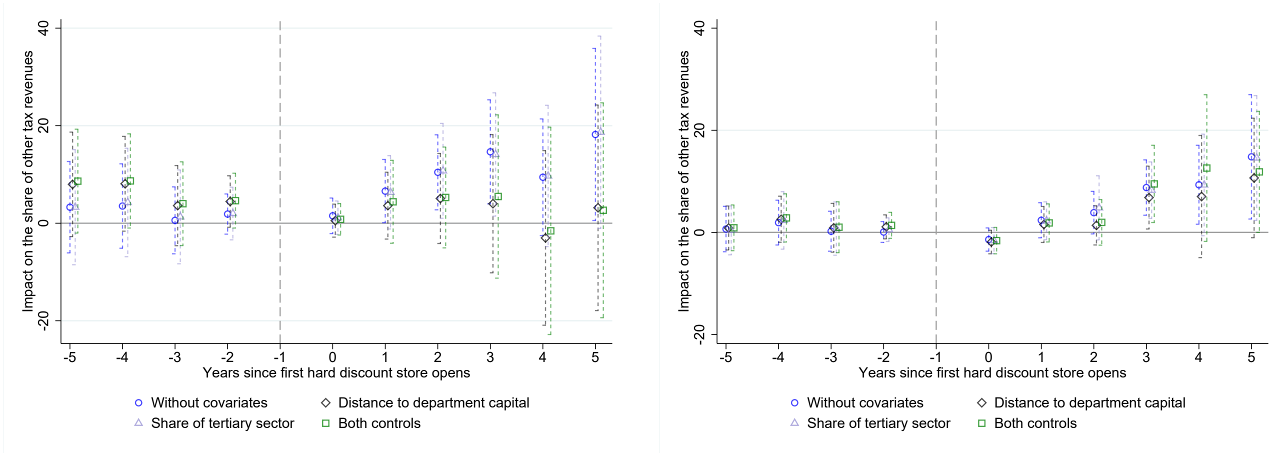
Note: The dependent variable is the type of employment relative to the working-age population in the 2005 census in a given municipality. We control for the logarithm of the driving distance from the municipality to the department capital in kilometers and for the share of tertiary sector employment in total formal municipality employment in 2010. 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. Source: GEIH 2010-2018.

Figure D.4: Event study estimates on subsidized SISBEN share using controls



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 in a given municipality. We control for the logarithm of the driving distance from the municipality to the department capital in kilometers and for the share of tertiary sector employment in total formal municipality employment in 2010. 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. Source: Health Ministry 2010-2018.

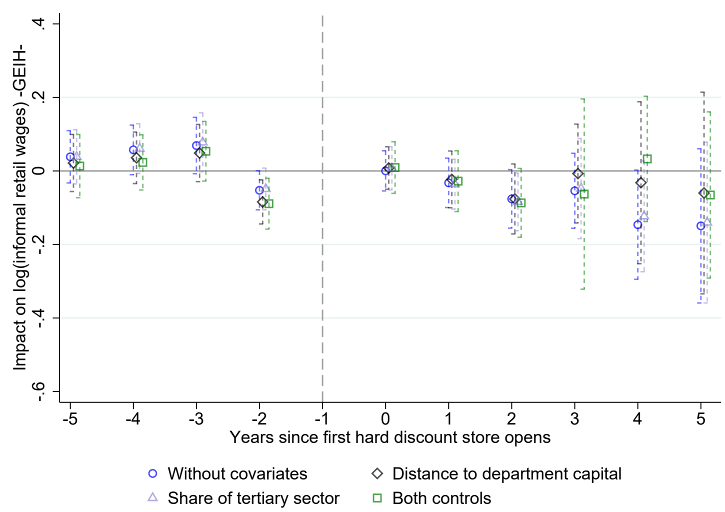
Figure D.5: Event study estimates on local taxes share by type using controls



(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) in a given municipality. 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. We control for the logarithm of the driving distance from the municipality to the department capital in kilometers and for the share of tertiary sector employment in total formal municipality employment in 2010. Regressions are weighted using the municipality's share of revenues in 2010. Observed treated municipalities are 371. Standard errors are clustered at the municipality level. 90% confidence interval. Source: DNP, 2010-2018.

Figure D.6: Event study estimates on informal labor income of retail workers using controls



Note: The dependent variable is the logarithm of informal labor income in a given municipality. We control for the logarithm of the driving distance from the municipality to the department capital in kilometers and for the share of tertiary sector employment in total formal municipality employment in 2010. 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. Source: GEIH 2010-2018.

Table D.4: **Average $C\&S$ estimates of informal employment by sector using controls**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: Without controls					
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
Clusters	191	191	191	191	191
Panel B: Log(Distance to department capital in km)					
ATT_{pre}	0.573	0.026	0.396	-0.883*	1.033**
	(0.741)	(0.404)	(0.366)	(0.498)	(0.460)
ATT_{post}	1.572	1.245	0.950	-0.800	0.178
	(1.306)	(1.019)	(0.844)	(1.171)	(0.561)
$ATT_{post_{k=3,4,5}}$	1.820	1.575	0.845	-0.658	0.058
	(2.106)	(1.541)	(1.153)	(1.532)	(0.852)
N	1,710	1,710	1,710	1,710	1,710
Clusters	190	190	190	190	190
Panel C: Share of tertiary sector in 2010					
ATT_{pre}	1.113	0.154	0.400	-0.732	1.291**
	(0.869)	(0.393)	(0.406)	(0.541)	(0.564)
ATT_{post}	-1.140	0.571	-0.401	-0.907	-0.403
	(1.683)	(0.633)	(0.494)	(1.085)	(0.566)
$ATT_{post_{k=3,4,5}}$	-2.251	0.483	-1.009	-0.566	-1.158
	(1.971)	(0.863)	(0.737)	(1.486)	(0.729)
N	1,719	1,719	1,719	1,719	1,719
Clusters	191	191	191	191	191
Panel D: Log(Distance) and share of tertiary sector					
ATT_{pre}	0.731	0.105	0.394	-0.860	1.092**
	(0.852)	(0.413)	(0.438)	(0.558)	(0.525)
ATT_{post}	1.289	1.438	0.626	-0.732	-0.043
	(1.411)	(1.025)	(0.759)	(1.412)	(0.635)
$ATT_{post_{k=3,4,5}}$	1.331	1.845	0.387	-0.614	0.723
	(1.814)	(1.452)	(0.966)	(2.010)	(1.608)
N	1,710	1,710	1,710	1,710	1,710
Clusters	190	190	190	190	190

Note: This table reports the robustness of the informal employment estimates (Table C.7) through the introduction of controls. 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 introduces control for the natural logarithm of the driving distance from the municipality to the department capital in kilometers, we retrieved using Google Maps. Panel C considers the share of tertiary sector employment in total formal municipality employment in 2010, calculated based on PILA data. Meanwhile, Panel D incorporates both covariates. 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. Source: GEIH 2010-2018.

Table D.5: **Average $C\&S$ estimates of labor income and working hours**

	(1) Formal labor income	(2) Informal labor income	(3) Formal working hours	(4) Informal working hours
Panel A: Without controls				
ATT_{pre}	0.030	0.002	0.002	0.004
	(0.023)	(0.016)	(0.007)	(0.009)
ATT_{post}	-0.022	-0.033	-0.004	-0.009
	(0.034)	(0.027)	(0.008)	(0.012)
N	1,718	1,719	1,718	1,719
Clusters	191	191	191	191
Panel B: Log(Distance to department capital in km)				
ATT_{pre}	0.028	-0.008	-0.004	-0.005
	(0.023)	(0.016)	(0.008)	(0.010)
ATT_{post}	-0.021	-0.007	-0.007	0.008
	(0.038)	(0.028)	(0.012)	(0.015)
N	1,709	1,710	1,709	1,710
Clusters	190	190	190	190
Panel C: Share of tertiary sector in 2010				
ATT_{pre}	0.030	0.006	0.001	0.003
	(0.023)	(0.016)	(0.009)	(0.011)
ATT_{post}	-0.017	-0.036	-0.003	-0.010
	(0.041)	(0.031)	(0.010)	(0.015)
N	1,718	1,719	1,718	1,719
Clusters	191	191	191	191
Panel D: Log(Distance) and share of tertiary sector				
ATT_{pre}	0.026	-0.005	-0.004	-0.004
	(0.025)	(0.017)	(0.009)	(0.012)
ATT_{post}	-0.016	-0.009	-0.007	0.006
	(0.045)	(0.029)	(0.011)	(0.017)
N	1,709	1,710	1,709	1,710
Clusters	190	190	190	190

Note: This table reports the robustness of the wages and working hours estimates (Figures 13 and 13) through the introduction of controls. The dependent variables are the logarithm of formal and informal labor income in a given municipality (columns 1 and 2), and the logarithm of formal and informal working hours (columns 3 and 4). Panel B introduces control for the natural logarithm of the driving distance from the municipality to the department capital in kilometers, we retrieved using Google Maps. Panel C considers the share of tertiary sector employment in total formal municipality employment in 2010, calculated based on PILA data. Meanwhile, Panel D incorporates both covariates. 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. Source: GEIH 2010-2018.

Table D.6: **Average $C\&S$ estimates of labor market rates using controls**

	(1) Employment rate	(2) Unemployment rate	(3) Inactivity rate
Panel A: Without controls			
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)
N	1,719	1,674	1,713
Clusters	191	189	191
Panel B: Log(Distance to department capital in km)			
ATT_{pre}	1.128*	-0.535	-0.810
	(0.619)	(0.528)	(0.586)
ATT_{post}	2.163***	-1.164**	-1.568***
	(0.692)	(0.579)	(0.561)
N	1,710	1,665	1,704
Clusters	190	188	190
Panel C: Share of tertiary sector in 2010			
ATT_{pre}	0.833	-0.430	-0.599
	(0.541)	(0.411)	(0.456)
ATT_{post}	2.300**	-1.071*	-1.881**
	(0.935)	(0.614)	(0.883)
N	1,719	1,674	1,713
Clusters	191	189	191
Panel D: Log(Distance) and share of tertiary sector			
ATT_{pre}	1.182*	-0.494	-0.889
	(0.617)	(0.553)	(0.542)
ATT_{post}	2.272***	-1.089*	-1.752***
	(0.752)	(0.657)	(0.608)
N	1,710	1,665	1,704
Clusters	190	188	190

Note: This table reports the robustness of the labor market rates estimates (Table C.4) through the introduction of controls. The dependent variable is the yearly labor market rate. Panel B introduces control for the natural logarithm of the driving distance from the municipality to the department capital in kilometers, which we retrieved using Google Maps. Panel C considers the share of tertiary sector employment in total formal municipality employment in 2010, calculated based on PILA data. Meanwhile, Panel D incorporates both covariates. 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. Source: GEIH 2010-2018 in August.

Table D.7: **Average $C\&S$ estimates of formal employment by sector using 2018 working-age population and excluding metropolitan areas in PILA**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: 2005 census working-age population					
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)
N	3,348	3,348	3,348	3,348	3,348
Clusters	372	372	372	372	372
Mean pre-treatment	16	.549	1.75	4.54	12.5
Panel B: 2018 census working-age population					
ATT_{pre}	-0.164	-0.037	-0.012	-0.009	0.025
	(0.202)	(0.045)	(0.057)	(0.135)	(0.139)
ATT_{post}	1.087***	0.141*	-0.102	0.548***	0.417
	(0.376)	(0.077)	(0.110)	(0.202)	(0.306)
N	3,348	3,348	3,348	3,348	3,348
Clusters	372	372	372	372	372
Mean pre-treatment	10.4	.799	1.5	2.97	8.15
Panel C: Excluding municipalities in metropolitan areas					
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)
N	3,213	3,213	3,213	3,213	3,213
Clusters	357	357	357	357	357

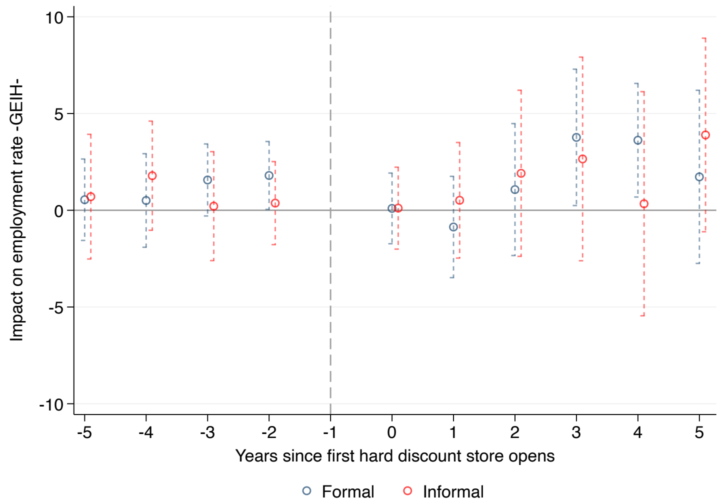
Note: This table reports the robustness of the formal employment estimates using PILA (Table C.6) by substituting the denominator of the dependent variables and excluding metropolitan areas. Panel B uses the 2018 census working-age population (individuals aged 15 years or older) to construct the shares, while Panel C excludes 19 municipalities classified by DANE as part of metropolitan areas. 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 the 2005 census in Panels A and C and with the local working-age population in the 2018 census in Panel B. Standard errors are clustered at the municipality level. Source: PILA 2010-2018 in August.

Table D.8: **Average $C\&S$ estimates of formal employment by sector using 2018 working-age population in GEIH**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: 2005 census working-age population					
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
Clusters	191	191	191	191	191
Mean pre-treatment	28.3	2.64	2.22	8.37	15.1
Panel B: 2018 census working-age population					
ATT_{pre}	0.312	-0.015	-0.021	0.150	0.197
	(0.483)	(0.120)	(0.106)	(0.238)	(0.361)
ATT_{post}	1.793***	0.341**	-0.068	1.280***	0.239
	(0.637)	(0.149)	(0.140)	(0.364)	(0.354)
$ATT_{post_{k=3,4,5}}$	2.800***	0.490**	0.104	1.822***	0.385
	(0.842)	(0.194)	(0.196)	(0.479)	(0.510)
N	1,719	1,719	1,719	1,719	1,719
Clusters	191	191	191	191	191
Mean pre-treatment	18.4	1.72	1.44	5.44	9.81

Note: This table reports the robustness of the formal employment estimates (see Table C.5) by substituting the denominator of dependent variables with the 2018 census working-age population (individuals aged 15 years old or older) in Panel B. 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 in Panel A, and with the local working-age population in the 2018 census in Panel B. Standard errors are clustered at the municipality level. Source: GEIH 2010-2018.

Figure D.7: Event study estimates on formal and informal employment using GEIH without survey weights



Note: The dependent variable is formal employment and informal employment using GEIH without survey weights both over the employed population in 2010. Observed treated municipalities are 191. Standard errors are clustered at the municipality level. 90% confidence interval. Source: GEIH 2010-2018.