DOCUMENTOS DE TRABAJO SOBRE
ECONOMÍA REGIONAL Y URBANA

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

By:
Lukas Delgado-Prieto
Andrea Otero-Cortés
Andrés Calderón

No. 326
Abril, 2024

Centro de Estudios Económicos Regionales (CEER) - Cartagena

The Impact of Hard Discount Stores on Local Labor Markets: Evidence from Colombia[1]

Lukas Delgado-Prieto[2]
ludelgad@eco.uc3m.es

Andrea Otero-Cortés[3]
aoteroco@banrep.gov.co

Andrés Calderón[4]
andresy.calderon@urosario.edu.co

The series **Documentos de Trabajo Sobre Economía Regional y Urbana** is a publication of Banco de la República in Cartagena. The opinions contained in this document are the sole responsibility of the authors and do not commit Banco de la República or its Board of Directors.

Abstract

Hard discount stores (HDS) have changed the dynamics of the traditional retail sector by selling a basket of products at very low prices. This business model has gained significant market share in many countries, but little is known about its impact on the labor market. To fill this gap in the literature, in this paper we study the impact of the entry of hard discounters on local labor markets in Colombia. Making use of the staggered geographic expansion of major discount chains throughout the country as part of our empirical strategy and using information from different sources, such as administrative records on social security and household survey data, we analyze the impact of these stores on labor formality and tax collection. Our results show that the arrival of HDS in a municipality increases local formal employment, especially in retail, manufacturing and agriculture. This suggests that there are significant spillover effects from retail to other industries, as most of the goods sold by these stores are locally produced. As for the informal sector, increased competition between formal and informal businesses has no statistical effect on informal employment. However, there seems to be a decline in labor income of informal retailers, suggesting that the margin of adjustment is not through lower employment but via lower earnings.

Keywords: Hard discount stores, competition, local labor markets, informality.
JEL Codes: E24, J46, O17.

[1] We are thankful to Luis Eduardo Arango, Jhorland Ayala, Jaime Bonet, Leonardo Bonilla, Juan Esteban Carranza, Jorge Flórez, Luis Armando Galvis, Miguel Talamas, and Gabriel Ulyssea, as well as participants at 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 Julían 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.
[2] Corresponding author: Department of Economics, Universidad Carlos III de Madrid, Spain. Email: ludelgad@eco.uc3m.es.
[3] Researcher, Banco de la República, Colombia.
[4] Inter-American Development Bank.

El impacto de las tiendas de descuento en los mercados laborales locales: Evidencia de Colombia

Lukas Delgado-Prieto
ludelgad@eco.uc3m.es

Andrea Otero-Cortés
aoteroco@banrep.gov.co

Andrés Calderón
andresy.calderon@urosario.edu.co

La serie **Documentos de Trabajo Sobre Economía Regional y Urbana** es una publicación del Banco de la República en Cartagena. Las opiniones contenidas en este documento son de exclusiva responsabilidad de los autores y no comprometen al Banco de la República ni a su Junta Directiva.

Resumen

Las tiendas de descuento duro (HDS por sus siglas en inglés) han cambiado las dinámicas del sector minorista tradicional al vender una canasta de productos a muy bajo precio. Este modelo de negocio ha ganado una importante cuota del mercado en muchos países, pero poco se conoce sobre su impacto en el mercado laboral. Para llenar este vacío en la literatura, en este trabajo estudiamos el impacto de la entrada de las tiendas de descuento duro en los mercados laborales locales de Colombia. Haciendo uso de la expansión geográfica escalonada de las principales cadenas de descuento por todo el país como parte de nuestra estrategia empírica y usando información de distintas fuentes, como registros administrativos sobre seguridad social y la encuesta de hogares, analizamos el impacto de estas tiendas sobre la formalidad laboral y recaudo de impuestos. Nuestros resultados muestran que la llegada de las HDS a un municipio aumenta el empleo formal local, sobre todo en los sectores del comercio minorista, la industria manufacturera y la agricultura. Esto sugiere que existen importantes efectos de derrame del comercio minorista a otros sectores económicos, ya que la mayoría de los bienes que venden estas tiendas son productos locales. En cuanto al sector informal, el aumento de la competencia entre comerciantes formales e informales no tiene efectos estadísticos sobre el empleo informal. No obstante, parece haber una disminución en los ingresos laborales de los minoristas informales, lo que sugeriría que el margen de ajuste no se da a través de menor empleo sino vía menores ingresos.

Palabras clave: Tiendas de descuento duro, competencia, mercados laborales locales, informalidad.

Clasificación JEL: E24, J46, O17.

1 Introduction

Hard Discount Stores (HDS) have become prominent in the retail business worldwide. With sales of over 442 billion dollars in 2022 (Euromonitor International, 2023b), they have gained significant market shares in many countries, for example, accounting for around 26% of the food and grocery market in Germany (Euromonitor International, 2023a; MarketLine Industry Profiles, 2023).[1] HDS continue to expand within and between countries based on key market strategies: a limited assortment of products, a high share of own labels offered at lower prices (which helps to maintain a high quality/price ratio), and efficient operations (Jurgens, 2014; Sachon, 2010). The rapid spread of HDS raises questions about how these stores impact retail businesses, consumer behavior, and the labor market. Regarding the latter, the impact is particularly unclear, as these stores can increase labor demand through direct hires, but they also increase competition within the retail sector, potentially making incumbent firms exit, which can lead to job destruction. Additionally, there can be spillovers to other industries upstream of the supply chain (de Paula and Scheinkman, 2010).

The retail sector in Colombia has experienced a dramatic change since 2009 with the introduction of the first hard discount chain. According to market data, hard discounter sales grew by 235% between 2017 and 2022, surpassing the market share of hypermarkets in 2020 and supermarkets in 2021 (Euromonitor International, 2023c). Their rapid success led them to open stores in many cities (along with other competing hard discount chains), such that in 2022, there were more than 4,000 stores in the country (Euromonitor International, 2023c). Moreover, according to administrative records, the three leading hard discount chains hired more than 26,000 formal workers as of 2019. From the beginning, HDS targeted the size and concept of hard discounters in Europe, opening stores in residential neighborhoods and city centers while offering a low-price, limited assortment of mostly own-brand goods.

In this paper, we study the staggered expansion of the leading hard discount chains in Colombia. Unlike the United States or Europe, the labor market in Colombia presents a particular structure, as more than half of the labor force works informally. Specifically, the retail sector in which these stores operate is mainly informal (55% of the workforce does not pay the mandatory contributions to social security, and 83% of the businesses are not legally registered). Still, hard discount chains are formal firms that must comply with labor and tax regulations, which means they only hire formal workers, must comply with minimum wage law, and for tax purposes, their suppliers must be formal firms.

To estimate the impact of HDS on local labor markets, we assemble a unique data set combining administrative records on formal employment, administrative records on subsidized social protection beneficiaries, labor force survey data, direct tax collection at the municipality level, and information on each store's lo-

[1] Hard discount chains have experienced substantial growth in Europe and other countries since 2008. Aldi and Lidl are the most prominent chains, as they are part of the top ten retailers worldwide and the largest hard discount chains (Statista, 2023).

cation and opening year for the three leading hard discount chains in Colombia, from 2010 to 2019 at the municipal level. Still, even with the extensive data collection, there is a challenge in identifying the effect of these stores on local labor markets, as there is an endogenous decision in where these stores decide to open. Therefore, we exploit the rapid staggered rollout of HDS across the country, which we show is unrelated to local labor market trends, and use an event study design, following Callaway and Sant'Anna (2021), with municipalities not yet treated as a control group. Hence, our identification strategy does not assume that the location of stores is exogenous; instead, we claim that the *timing* of opening the first hard discount store is unrelated to local employment or wage trends. Therefore, comparing cohorts of municipalities where hard discount chains opened a store earlier to cohorts of municipalities that opened later helps us 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 several outcomes, like employment, wages, working hours, and taxes, before opening the first store.

With this research design, we capture the dynamic treatment effects of opening HDS on the formal and informal labor markets at the local level.[2] We find three main results in this paper. The first one is that after opening the first hard discount store in a municipality, there is an increase in local formal employment of 1.7 percentage points (pp) using administrative records and of 2.9 pp using survey data, which is driven mainly by retail, manufacturing, and agricultural employment.[3] This suggests that HDS increase the demand on local industries to produce more inputs for the goods they sell in their stores, as reported by the largest hard discounter (La República, 2022). The effect takes a few years to materialize because HDS in Colombia are small (they have, on average, ten workers per store). Thus, it takes time for HDS chains to open up enough stores in each municipality to affect the local labor market. Second, we find estimates for informal retail employment that are small and sometimes positive, yet we do not rule out that as hard discounters open more stores, they can cause closures of informal neighborhood shops as we do not have the most appropriate data to test this hypothesis. We also find that informal retailers' labor income has a negative trend in the last treatment years we analyze, in line with findings from a similar shock in the Mexican retail sector (Talamas Marcos, 2024). Third, using a proxy of tax collection, we find a positive impact on the local economy as the share of taxes over total public revenues increases by 10.1 pp on average. We find that the increase in overall taxes is driven by taxes from industry and commerce, consistent with the story that HDS have spillovers on other sectors.

Related Literature. This paper contributes to two 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

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

[3] The estimates are robust to the inclusion of controls like distance to the capital city or baseline local economic structure.

in the world. The entry of Walmart into new counties in the US negatively impacts local labor markets by reducing retail employment and earnings in the mid-to-long run, as they exploit their monopsony power and ability to affect other local retail stores that might not compete with their 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 (similarly to our empirical strategy), they find negative effects on retail employment and positive employment spillovers in other sectors: Chava et al. (2023) on transportation and warehousing, whereas Cunningham (2023) on tradeable services.[4] Our paper diverts from the previous work as our context is different: we are studying the entry of hard discount chains in a setting with a high prevalence of informal employment and informal competing businesses (neighborhood shops) and traditional supermarkets. In our context, a binding minimum wage also prevents earnings from falling, and hard discount chains fundamentally differ from the concept of Walmart or Amazon FCs. Thus, to the best of our knowledge, this is the first paper that estimates the impact of HDS in developing countries across the formal and informal labor markets. In a developed economy, 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 but also due to positive spillover effects on other retail sectors.

Another strand of the literature that we contribute focuses on the effect of increased competition in the retail sector. To start, there is evidence of increased consumer surplus when hard discounters enter the town as their average prices are lower than traditional supermarkets, and they induce price reductions in competing stores (Atkin et al., 2018; Busso and Galiani, 2019; Hausman and Leibtag, 2007). Although we are not directly testing this result, given the context of Colombia, HDS entry can affect consumption choices as consumers substitute shopping in neighborhood shops (that are not necessarily cheaper than supermarkets) for hard discounters, probably increasing their consumer surplus. But as neighborhood shops may face decreased profits, we also test whether there are changes in informal employment and earnings that could capture neighborhood shops' closure. Lastly, we also aim to measure the spillover effects of HDS entry on employment in other industries, such as manufacturing and agriculture, which are related to the literature on upstream supply chain effects as hard discounters have incentives to source the goods they sell from local formal suppliers (de Paula and Scheinkman, 2010; Gerard et al., 2023; Rios and Setharam, 2018).

Studying how the entry of hard discounters affects incumbent retailers, we use the following facts that provide insights into the possible mechanisms. First, hard discount chains seem more efficient than super-

[4] 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.

markets, as they have larger sales per square meter and face lower operation costs, thus increasing their market share in the past few years. However, since HDS are not perfect substitutes for supermarkets due to their limited assortment inventory, consumers may shop in multiple establishments. For instance, Florez-Acosta and Herrera-Araujo (2020) documents that it is common for French households to visit multiple supermarkets per week even if they offer similar products, so the entry of HDS may likely accelerate the multistop shopping behavior among consumers. Second, we find that informal employment in the retail sector does not significantly decline after the expansion of hard discount chains. This suggests that neighborhood shops manage to continue to operate despite increased competition, but the margin of adjustment may occur through earnings. These findings complement the evidence from convenience stores in Mexico (Talamas Marcos, 2024), where an additional convenience store entry in a neighborhood adversely affects the creation of neighborhood shops. Yet, existing neighborhood shops survive as they leverage their comparative advantage by mainly supplying fresh products bought daily and reducing costs by shrinking their inventories.[5] Furthermore, neighborhood shops tend to have low fixed costs, as they do not register to pay taxes, and most operate within the owners' houses (Ramos-Menchelli and Sverdlin-Lisker, 2023), so they have more space to adjust in response to the increased competition.

This paper is divided as follows. The next section provides the institutional context and information on the retail sector in Colombia. Section 3 describes our primary sources of data on local labor markets and the information on how we measure the arrival of HDS in the municipalities. Next, we discuss our identification strategy and its assumptions in Section 4. 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

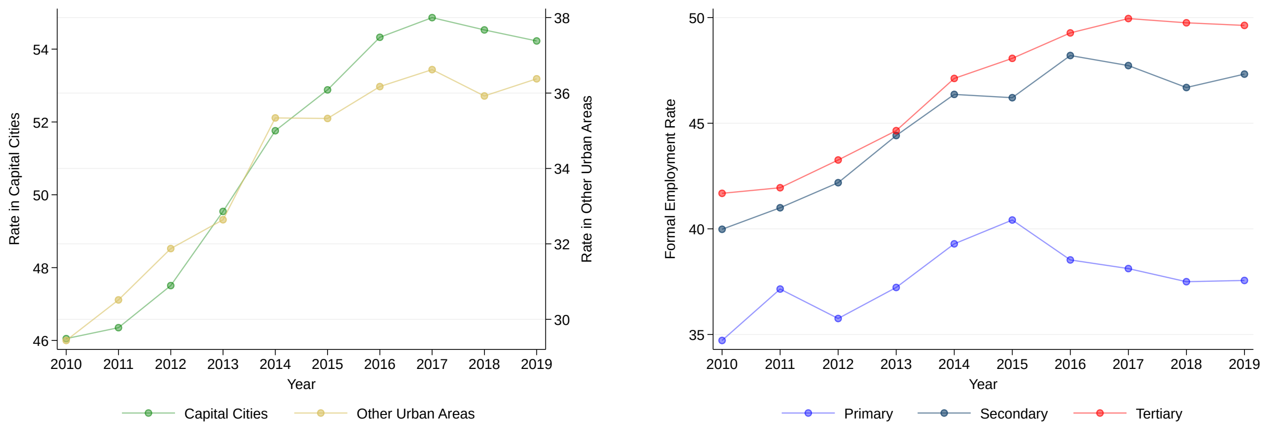
In Colombia, informal labor arrangements are the prevalent type of employment. However, in recent years, there has been a noticeable upward trend in formal employment rates in Colombia, both in capital cities and other urban areas (see Figure 1). There are several factors contributing to this phenomenon. 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 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

[5] In a different market, Macchiavello and Morjaria (2021) finds that competition negatively affects several market outcomes when formal employment contracts are not enforceable in the coffee mill industry in Rwanda. They find that new mill openings increase farmers' temptation to default on their previous relational contract, which causes farmers to be worse off and indirectly reduces 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, 2022).

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. 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). By 2018, these companies were responsible for almost half of the supermarkets' and three-quarters of the hypermarkets' sales.

Supermarkets have had a long history and tradition in the world's retail market. The most common supermarket format, which mainly operates under a self-service concept, offers from meat and fresh products to cleaning supplies. This format was introduced in Colombia at the end of the 1940s by José Carulla Vidal

[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 there is a correlation between the first arrival of a hard discount chain to a municipality and the treatment intensity of the 2012 reform.

(Silva Guerra, 2011). Mr. Carulla had a handful of convenience stores across Bogotá when he discovered the supermarket stores were booming in Mexico, such as the company *Sumesa*. 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).

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.

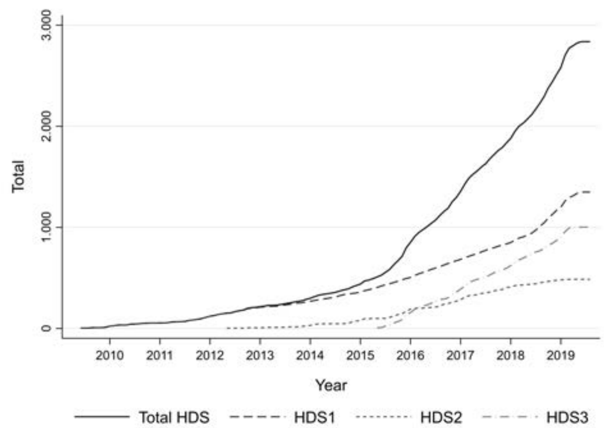
According to market data from Euromonitor International (2023c), Colombian supermarkets have an aggregate market size of around 4 billion dollars, with approximately 2,200 outlets. The average selling space is around 750 square meters per store, and the typical store sells about 2 million dollars annually (US$2,450 per square meter). There are also about 210 hypermarkets in Colombia, mainly in the largest urban centers, and their sales also sum up to 4 billion dollars. 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% and by 86% for hypermarkets (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 the operating income and profits of supermarkets and hard discounters. Historically, supermarkets dominated the retail market regarding operating income, with significant profits. However, since the entry of hard discount chains, the operating income of the new competitors has increased substantially.[8] On aggregate, the main three supermarkets have started to show a decreasing profit in 2022, while the main two hard discounters (as the third chain exits the market in 2021) have started to gain a larger profit.

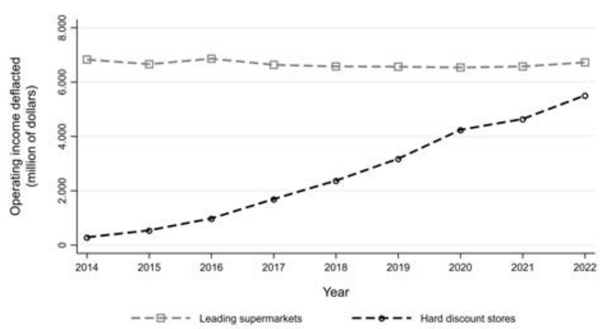
[8]Even the leading hard discounter in the market already has a larger operating income than the leading supermarket in the country.

Figure 2: Hard discount stores

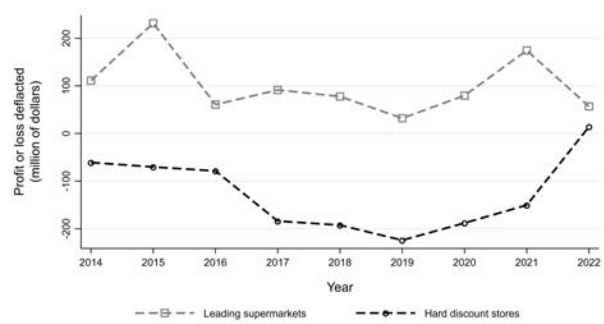


Note: This figure shows the total number of hard discount stores from the three main chains operating in Colombia between 2010 and 2019. Source: Authors' calculations using public location data obtained from the hard discounters' websites.

Figure 3: Operating income and profits of the leading supermarkets and hard discounters



(a) Operating income



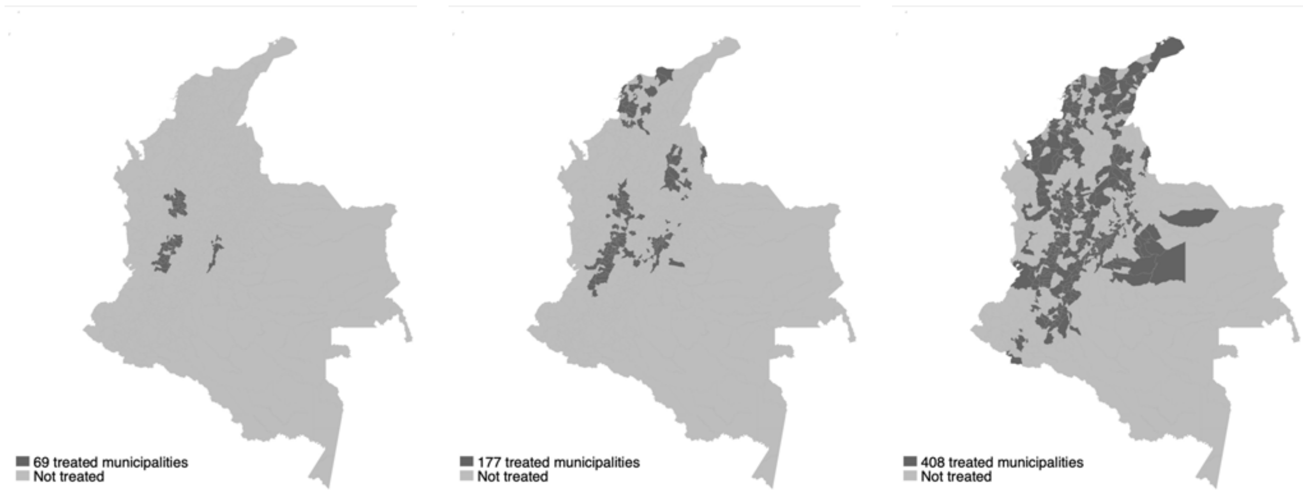
(b) Profits

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).

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 that starts 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.

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 B.1 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 0.53), with population size being the most relevant one.

2.2 Neighborhood Shops

Hard discount chains operate in an industry where traditional supermarkets and neighborhood shops are prevalent. These shops are small, mostly informal businesses that offer a limited supply of essential goods, mainly in residential places.[9] They tend to create a close relationship with their customers, even offering

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

informal credit to them (Talamas Marcos, 2024).[10] Neighborhood shop prices are not necessarily lower than supermarket prices, yet as they sell smaller quantities of essential items, the ticket for daily grocery shopping would fit into lower-income households' daily budgets. On the other end, traditional supermarkets are large formal firms with infrastructures similar to those of large retailers in the US, where customers may find goods from fresh produce to home appliances. They target middle- and high-income households based on their location and the assortment and brands they sell. In 2017, neighborhood shops had 52% of the share of the market, while traditional supermarkets represented 23%. Hence, in this context, the arrival of HDS increases competition for the market of certain products with supermarkets and neighborhood shops (Sánchez Duarte, 2017).

We now analyze extensive information on neighborhood shops using a micro businesses survey (EMICRON, by its acronym in Spanish).[11] These small local grocers represent a crucial channel for the grocery retail industry and a relevant source of informal employment. Appendix Table B.2 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)[12]. 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.[13] 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. Appendix Table

[10] 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.

[11] 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 trade 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.

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

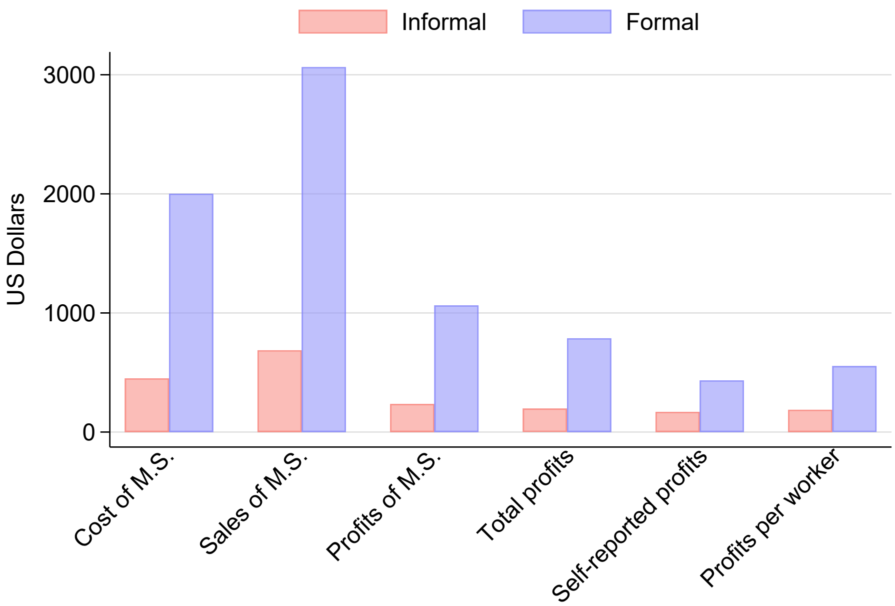
[13] Three taxes are particularly important for businesses in Colombia: the VAT, the income tax, and the industry and commerce tax. The first two are national-level, whereas the last one is paid to the municipality.

B.2 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 5. 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.

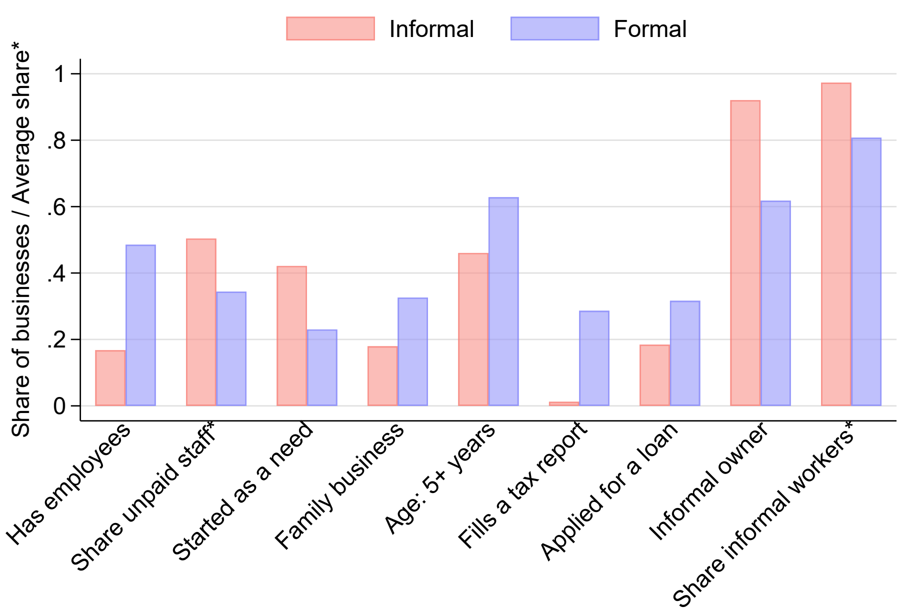
Panel (b) of Figure 5 shows the differences between formal and informal stores in other business characteristics. Formal shops have a larger staff than informal ones and rely less on unpaid workers. They are also less likely to start as needed and more likely to survive in the market.[14] Regarding employment 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).

[14]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).

Figure 5: 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 trade and repair of motor vehicles and motorcycles". The mean and standard deviation are weighed using the survey's sample weights.

3 Data

This paper uses multiple data sources to capture both formal and informal employment. First, we use monthly employer-employee matched administrative records from mandatory social security contributions (PILA, by its acronym in Spanish) from 2010 to 2019. PILA contains the universe of 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.

Nevertheless, we have two limitations on identifying municipality and industry classification for some workers. On the one hand, we do not identify the municipality where workers are located for a subset of workers, especially from 2017 to 2019. We use two different strategies to impute this information. First, we compute the mode of the previous municipalities in which the same worker appears registered and impute it to the missing location. Second, we use the 2020 worker location information to assign that location to the municipality of workers with a wrong municipality code.[15] On the other hand, in Colombia, up to 2013, the standard was to self-report in PILA the industry classification using the International Standard Industrial Classification (ISIC) revision three, but in 2013, the ISIC revision four was adopted, but its implementation in reporting was not mandatory. Therefore, from 2013 to 2019, some firms reported their industry code using the previous version of the ISIC list instead of the current one. To cope with this issue, we computed the mode of the ISIC code for each firm in PILA and took it as their official ISIC code. We do this to identify formal employment in the retail sector.[16]

For the final estimation sample, we exclude capital cities of departments 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 total formal employment, which accounts for approximately 7 million observations in PILA 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 a 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.

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 social protection (SISBEN), which allows us to approximate the stock of low-income informal workers at the municipality level using the same sample

[15] After this procedure, we find that approximately 1.8% of the workers in PILA have reported a wrong municipality code that we are not able to fix. Thus, we do not include them in our estimation sample as we need to know the total employment at the municipality level.

[16] Approximately 1% of workers per month from our estimation sample are not classified in any industry due to wrongful ISIC coding.

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 are beneficiaries of 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.[17] 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.[18]

The data from the GEIH survey has the limitation that it is not representative at the municipality level.[19] Still, as our empirical analysis aggregates similar municipalities into treatment cohorts (that range in size from 5 to 35 municipalities depending on the cohort, see Table 3), 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 the fact 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.[20] From the GEIH survey, we use similar outcomes and fix the denominator again with the census of 2005 to have comparable results. We calculate average wages and working hours within each sector and take the logarithm transformation.

To capture aggregate 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.

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

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

[19]Even though these municipalities are the largest ones surveyed in GEIH, besides capital cities.

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

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

3.1 Descriptive Statistics

Administrative Records (PILA). Figure 6 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 we describe later using the GEIH survey.

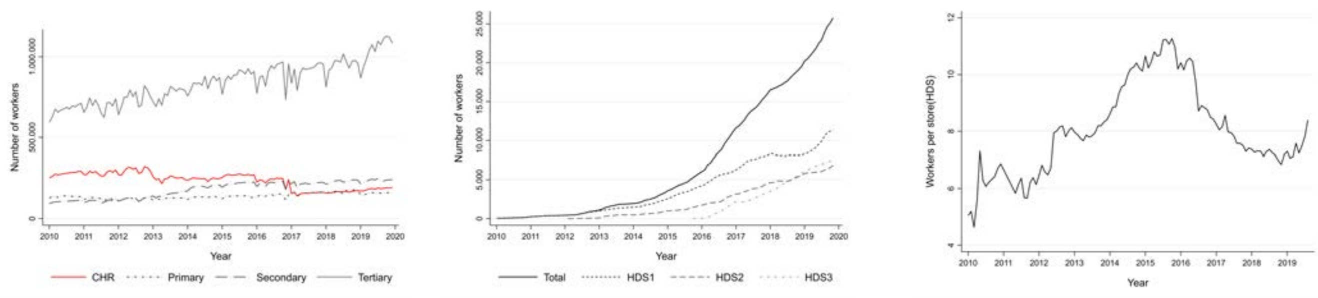
Regarding formal employment in hard discount chains, they went from employing less than 500 formal workers in 2012 to almost 26,000 workers in 2019, showing the rapid expansion of these chains (see Figure 6 panel (b)). Among all these workers, 86% of them are hired as full-time employees (i.e., with a formal work contract for 30 days of the month).[21] Lastly, Figure 6 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 in PILA. First, we find that from 2010 to 2019, 57,963 individuals have worked at any point at one of the three leading hard discount chains. Around 16% of them are first-time formal workers, meaning they appear for the first time in PILA as employees of HDS, and 31% of them used to work at another formal firm and switched to HDS. Among first-time formal workers, 55% are women, the average age is 24, and 47% have part-time work contracts in the first quarter they are hired. On the other hand, those who switch from other firms tend to be older, 30 years old on average, 60% are men, and more than 75% of them come from jobs in the wholesale and retail industry and

[21] Given the nature of the legislation in Colombia, full-time employment is more prevalent in the formal sector than part-time employment, as it was relatively more costly to hire a formal part-time worker than a formal full-time worker. 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., 2023).

services sectors. Overall, we document that hard discount chains play a role in hiring young first-time formal workers.

Figure 6: 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. CHR refers to employment in commerce, hotels, and restaurants; primary refers to employment in agriculture and mining; secondary refers to employment in manufacturing and construction; and tertiary refers to employment in services. Panel B shows the evolution of full-time formal employment by the hard discount chain in the municipalities included in the estimation sample. Panel C reports the average full-time formal employment per store in the municipalities included in the estimation sample. Source: Authors' calculations based on PILA.

Table 1 shows descriptive statistics for the treated municipalities in our estimation sample. First, we observe that in 2011, the average employment was 3,856. This number increased over time as hard discount chains opened in more municipalities. In our time frame, more than 90% of the formal employment in the treated municipalities does not belong to the trade sector. Still, the share of workers in the trade sector increased over time, going from 4.2% in 2011 to 8.3% in 2018. The treated municipalities also have a higher share of workers that are employees relative to the not-yet-treated municipalities. Regarding the minimum wage, both groups have similar shares, ranging from 47% and 57% of the workers earning the monthly minimum wage.[22]

[22]By definition, workers who appear in PILA are bound to earn at least the monthly minimum wage.

Table 1: Employment outcomes for the estimation sample with PILA

	Treated				Not yet treated			
	2011	**2013**	**2016**	**2018**	**2011**	**2013**	**2016**	**2018**
Male workers (%)	62.3	61.6	60	61.4	63	62.1	62	66
Average total employment	3,856	5,572	4,610	6,783	1,318	2,114	2,554	2,145
	(6,139)	(8,396)	(6,411)	(6,079)	(3,436)	(4,100)	(5,011)	(4,090)
CHR sector (%)	4.2	5.7	7.7	8.3	5.6	7	7.7	8.2
Non-CHR sector (%)	95.8	94.3	92.3	91.7	94.4	93	92.3	91.8
Employees (%)	64	73.1	75.7	83.7	63	70.6	74.2	87.5
Independent workers (%)	18.4	18.4	19.4	12.8	20.1	22.5	21.6	9.3
Min wage workers (%)	57.2	54	47	53	53	55	48	54
Average earnings (USD)	301	324	334	353	319	337	311	335
	(278)	(332)	(391)	(376)	(355)	(361)	(268)	(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. Average total employment is computed for all the municipalities treated in that year and only includes full-time employees. The "CHR sector" represents the proportion of formal full-time employees working in commerce, hotels, and restaurants. "No CHR" represents the proportion of formal full-time employees 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-employment on total formal workers. "Min wage workers" is the share of formal full-time workers earning the minimum wage on total formal workers. PILA is bounded at 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 in parentheses.

Labor Force Survey (GEIH). Table 2 shows the mean and standard deviation of several labor market outcomes by year and treatment status in our estimation sample.[23] 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 percentage points 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

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

the primary and secondary sectors are 34% and 30%. Conversely, the rest of the commerce and the services sectors had a similar weight in the local economies of treated and not yet treated municipalities.

Table 2: Employment outcomes for the estimation sample with 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

Notes: 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.

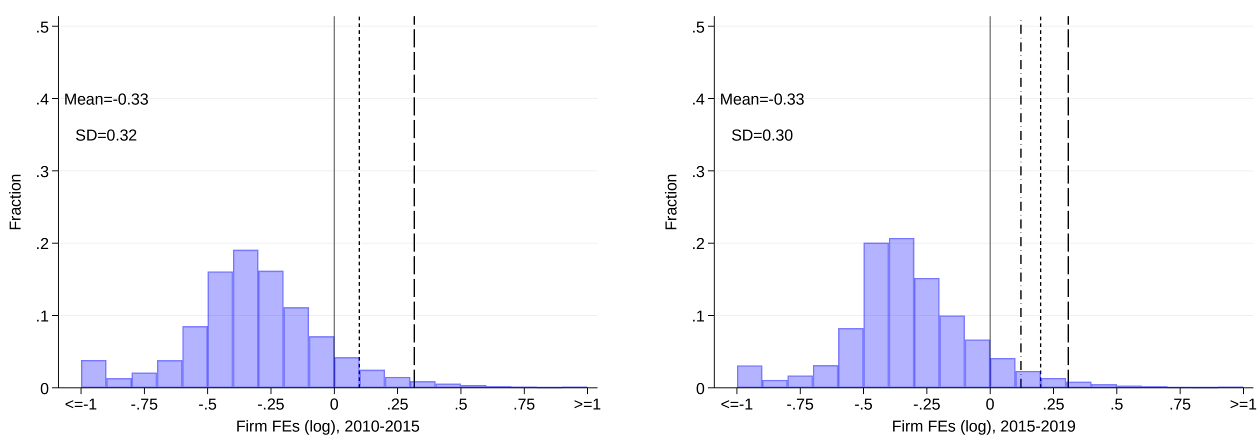
3.2 Pay-premiums of Discount Firms

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 7 shows how the dashed lines, which represent the three hard discount chains, pay, on average, consistently higher premiums to all their workers than the country's largest retail store. In particular, 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 relative to the largest retail store. In both estimation periods, we find this positive contribution of HDS chains. The intuition and assumption of this framework are that if a worker moves from the reference firm to an HDS, it would have an average positive gain in their wages equal to the value of the dashed line depending on the HDS 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.[24]

Figure 7: 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. We do not disclose which line belongs to which HDS for confidentiality reasons. 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.

[24]We do not use the leave-out method proposed by Kline et al. (2020) for the estimation as it yields variance and covariance moments of the wage decomposition, not the vector of estimated parameters we show in Figure 7.

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 how Chava et al. (2023) and Cunningham (2023) study the local impacts of Amazon FCs. Table 3 presents the staggered arrival of hard discount chains in our sample of municipalities (we observe 372 in PILA and 191 in GEIH). Although the number of municipalities increases with time (showing the large expansion of these chains throughout the decade), around 40% of them had a HDS by the end of 2016.

Table 3: 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 results. 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.[25]

The intuition of the C&S estimator is that it breaks down all the sets of possible comparisons into the two periods and two groups (2x2) framework to estimate multiple ATT(g, t) for all 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

[25] 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.

opens in the municipality, so if the treatment started in 2015, then $g = 2015$.[26] More concretely, they are estimated as follows:

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

Here, we use as a control the not yet treated group ($G_l > t$) for a cleaner comparison with treated municipalities, estimate the differences relative to the baseline period ($g - 1$) using ordinary least squares without covariates, and test 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)[27]:

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

Next, as a comparison for the C&S coefficients in the event study figures, we also estimate the traditional TWFE regression that takes the following form:[28]

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

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.[29]

Identification Assumption. The main assumption required in this setup is the standard parallel trends 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 argue that HDS openings are exogenous to local economic trends. Instead, we argue that unobserved trends do not determine the *timing* of HDS openings. Because the roll-out of HDS across the country is rapid, it is less likely to be determined by the local economic trends among the early and late-

[26] 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 $t - g > 5$.

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

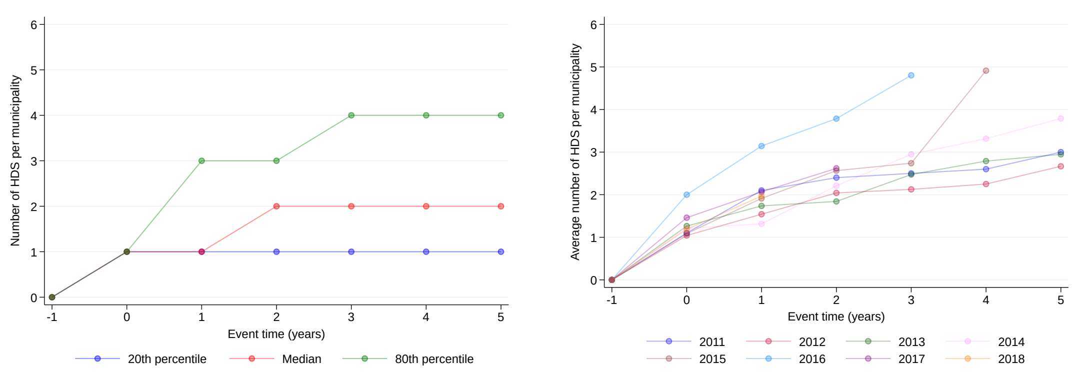
[28] To compare the pre-treatment coefficients of TWFE and C&S accurately, we quantify long gaps when using C&S.

[29] 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).

treated cohorts. In particular, the *timing* of openings is more related to the population levels (as in larger cities, they tend to open first) or the regional location of the municipality (see Appendix Table B.1). In these cases, our specification already controls for constant observed and unobserved characteristics. Furthermore, we show there are no differential employment or wage trends before the arrival of these stores across different sectors and industries. This suggestive evidence supports our *timing* assumption for identifying the effects of HDS openings in local labor markets. One potential concern of this assumption is that HDS might open first in places that are more connected to the capital cities. Hence, the municipalities that open first are more likely to be affected by the spillovers from shocks in the capital cities than the municipalities that open later on. To control for this potential concern, we use the distance to the capital city as a control in the estimation and show that results do not change significantly. We assume that there are no anticipatory effects as a response to store openings. Yet, as the shock is relatively unexpected and its effects take years to materialize, this is less of a 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, then how to interpret the overall ATT is less clear. Figure 8 presents evidence that cohorts receive a similar number of stores over time. In all cohorts, half of the municipalities have two HDS after three years of the arrival of the chains, whereas the top 20% that receive the largest number of stores have four, and those that receive the least number of stores have one. Moreover, if we differentiate by cohort, the average number is similar in all of them. The only one that is higher is the cohort of 2016, the year of the arrival of *Justo & Bueno*. Still, the difference is around one store more.

Figure 8: 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 B.1, 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 B.1).

Apart from the cross-sectional differences in pre-treatment characteristics, the labor markets of never-treated municipalities exhibit different trends than those of treated areas. Panel B of Appendix Table C.1 shows that when we employ the never-treated group as a control, significant pre-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 in all these outcomes. Moreover, in Panel C, we narrowed down 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.[30] 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

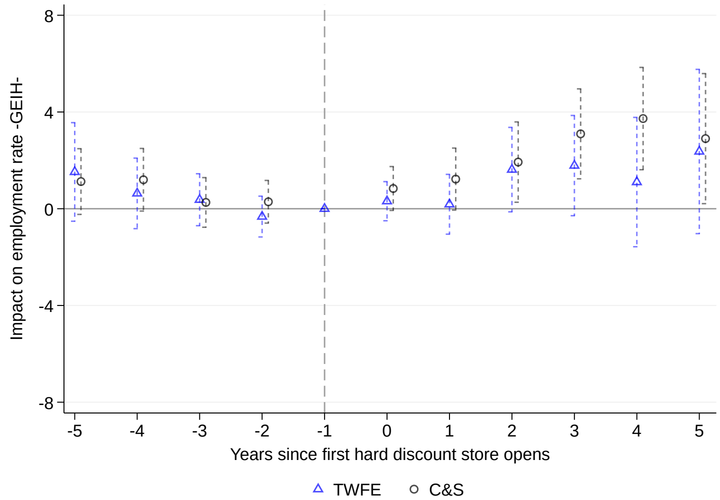
To start, we evaluate the overall effect of HDS on standard labor market outcomes. Figure 9 shows the dynamic treatment effects on the employment rate in the local labor markets. Importantly, treated and not yet treated cohorts of municipalities do not present differing trends on this outcome before the treatment, suggesting that the arrival of HDS is unrelated to local employment rates among these groups.

If a store opening is equivalent to a labor demand shock, then employment in a given local labor market should be positively affected. Consistent with this hypothesis, the estimates in the post-treatment periods are positive and increasing over time; most of them become significant in the latest periods. Note that the TWFE estimates are consistently smaller than the C&S ones, suggesting that heterogeneous treatment

[30] 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 B.1–Column 7.

effects are non-negligible in this outcome. After six years, HDS had an average positive impact on the local employment rate of around 2.3 pp (see Figure 9). Furthermore, in Appendix Table B.4, we find a decrease in the unemployment rate that is not significant, though it is large, and a significant decrease in the inactivity rate following the opening of a store.

Figure 9: 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 behavior of the labor market could mask the response from specific types of employment that HDS demand more, namely, formal employment. Therefore, the subsequent analysis is to quantify the evolution of formal employment in the local labor markets following the opening of stores. For this analysis, we show results using mainly the administrative data, and to support this, we also use the survey data, but 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).[31]

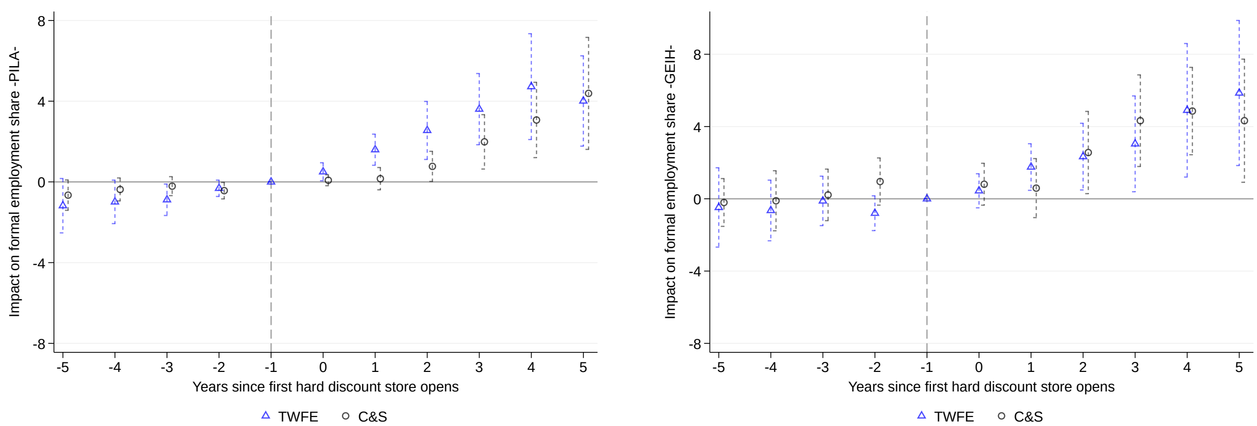
Figure 10 shows that formal employment follows a positive trend that takes a few years to materialize, both using the administrative and the survey data. Hence, the increase in total employment mostly comes from the increase in formal employment. Averaging all six post-treatment periods, there is an increase in formal employment of around 1.7 pp using administrative records and 2.9 pp using survey data, which is robust to the inclusion of controls like distance to the capital city or baseline local economic structure. Hence,

[31]The coefficients can also differ because the employment information of PILA is slightly underreported (firms with several establishments across the country tend to report the information for their workers in the municipality where the headquarters is located). And because the GEIH survey is not representative of all the municipalities we consider in the analysis.

if the *timing* of opening an HDS is unrelated to the local trends in formal employment, this represents a causal change in the type of employment in the affected municipalities.[32]

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.[33] Hence, an increase of 2.9 pp is equivalent to a 10.3% increase in this rate. 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.

Figure 10: Event study estimates on formal employment



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

Note: 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 analyze the drivers of the increase in formal employment. Figures 11a and 11b show that the primary and secondary sectors drive the growth in formal employment using both the administrative and survey data (see Appendix Tables B.7 and B.5 for more details).[34] The increased demand from hard discount chains to local formal suppliers for their products seems to explain the increase in employment in these industries, particularly in manufacturing, agricultural, and construction (see Appendix Table B.6), which are the main producers of the type of goods these stores sell.[35] This aligns with the findings from

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

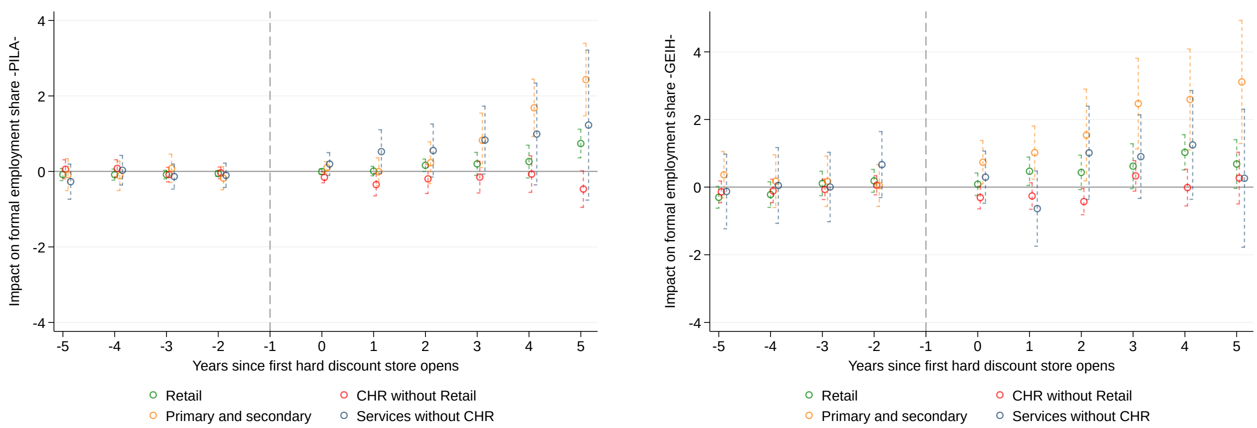
[33] The weights we use are the working-age population in the 2005 census, the same as in the main regression.

[34] In most results, we aggregate the primary and secondary industries due to the small sample size of formal employment in certain urban areas.

[35] We cannot directly estimate that hard discounters buy from the local producers because we lack input-output linkages data. However, the discounters have reported that they have local providers. For instance, according to the largest hard discount chain, around 80% of their goods in 2020 were produced within the country (La República, 2022).

de Paula and Scheinkman (2010) about the business formality chain that occurs when value-added taxation generates incentives for firms to buy from formal firms upstream in the supply chain. Moreover, as consumers decrease the share of the money they spend on groceries after HDS arrive, they can increase their spending on other goods from these industries, similar to an income effect. Lastly, we find a significant increase in employment in the retail sector with the administrative and survey data, in line with a direct increase in the demand for formal retail jobs from HDS.

Figure 11: Event study estimates on formal employment share by industry



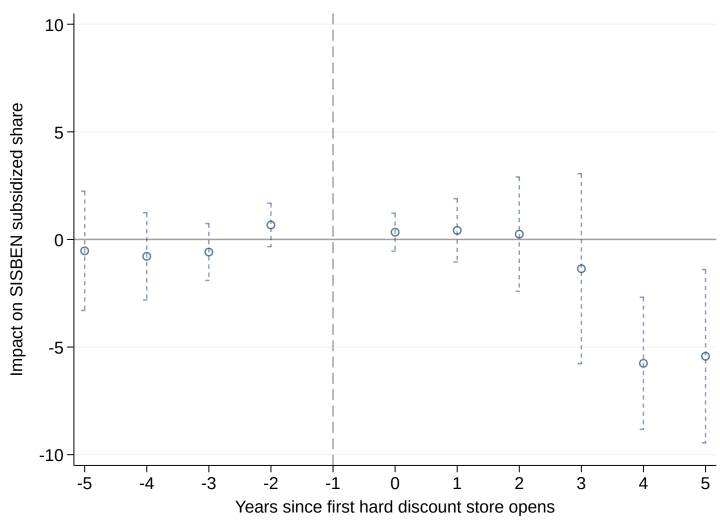
(a) Formal employment (PILA) **(b) Formal employment (GEIH)**

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.

Because the main competitors of HDS are local neighborhood shops that informal workers run, the next outcome we analyze is the aggregate effect on informal employment. In this case, we use the survey data as we only observe the informal sector with GEIH. Appendix Figure C.3 shows that the impact on informal employment is not robust, as the coefficient ranges from negative to positive values when we add further controls like distance to the capital. Hence, we argue that the impact of HDS on informal employment is imprecise. By sectors, the increase in competition between HDS and the informal retail sector does not reduce informal retail employment, and even the point estimate is positive across different specifications (see Appendix Table C.4). Furthermore, the coefficients on informal employment across other sectors vary substantially, but they are again noisy to describe a pattern (see Appendix Table B.8).

A limitation of using the GEIH survey is that we do not observe information for all the municipalities where HDS are located. For that reason, we exploit how Colombia's social security system is designed to get complementary measures of the informal sector. As Colombia grants almost universal access to the health system, all workers can access it by being subsidized or contributing to it. The ones contributing are all formal workers, while those subsidized do not have formal employment. Likely, they are working informally, but they can be not working at all. With this measure of beneficiaries of subsidized social protection, a proxy of informal employment, we show in Figure 12 that in the first four years of the treatment, the effect of HDS openings is close to zero. Still, later on, it has a negative trend similar to the one with the GEIH survey.[36] Still, we cannot differentiate with this data which sector or type of population generates the lagged impact. Overall, the main effect of HDS comes from the substantial change in the workforce composition, as formal employment is increasing more than informal employment in the treated areas.

Figure 12: Event study estimates on subsidized social protection beneficiaries (SISBEN)



Note: The dependent variable is the number of beneficiaries of subsidized social protection 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

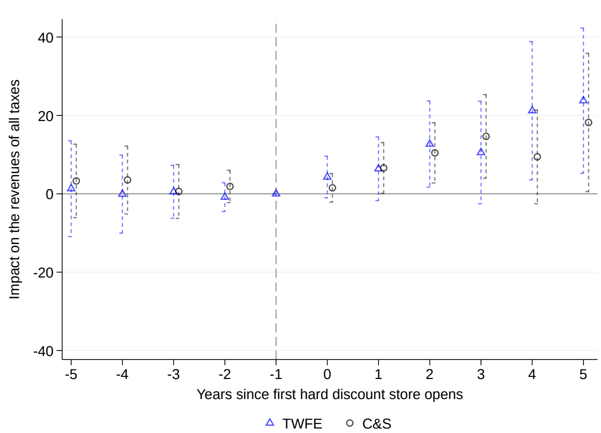
The arrival of hard discount stores can promote economic activity in their municipalities due to direct job creation and spillover effects to other sectors, but also due to increased value of certain areas of the city where the stores are located. We use municipality-level tax revenues as a proxy of economic activity to measure if HDS entry and its significant effects on employment are causing more aggregate effects 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,

[36] Appendix Figure C.4 shows the results are robust to different controls.

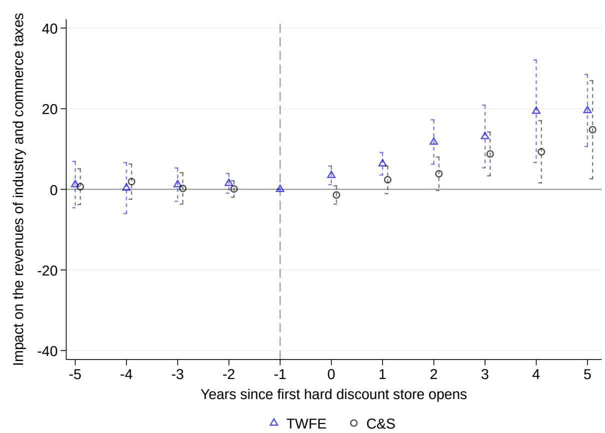
such as property taxes, industry and commerce taxes, vehicle and gasoline taxes, among others.[37] For instance, the industry and commerce taxes are subject to the performance of any industrial, commercial, or service activity registered in the municipality. Thus, in the case of hard discount chains, even though they are present all over the country, they still need to pay industry and commerce taxes based on the local sales and revenues in each of the municipalities where they operate.

Figure 13 shows that before the arrival of stores, the growth of local taxes as a share of total revenues is similar between the municipalities that opened first relative to 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, including all post-treatment periods, and most of this increase is explained by the revenues collected from the industry and commerce taxes (see Appendix Table B.9). Hence, the arrival of stores generates increased tax revenue for the local government, which is explained mainly by the growth of commercial and manufacturing activities. This result goes in line with the evidence presented in the literature on upstream supply chain effects of business formality as HDS purveyors must be formal for them to be able to claim all the tax benefits and discounts provided by the law (de Paula and Scheinkman, 2010; Gerard et al., 2023; Rios and Setharam, 2018).

Figure 13: Event study estimates on local taxes share by type



(a) **Revenues of all taxes**



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

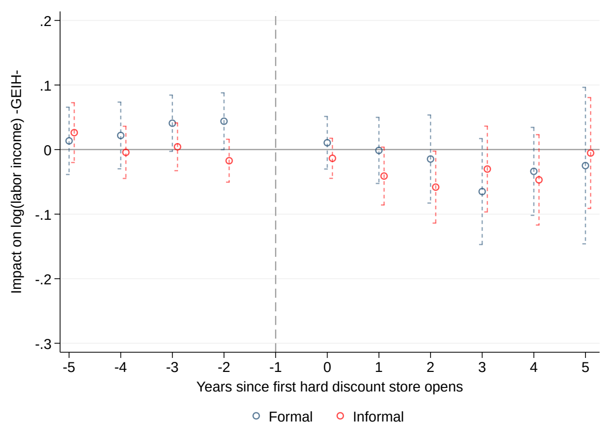
Note: The dependent variable is the revenue by each specific type of local tax over all the revenues (taxes and central government transfers) in a given municipality. We only included taxes collected at the municipality level, such as property taxes, industry and commerce, gasoline taxes, vehicle taxes, 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.

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

5.3 Labor Income and Working Hours

Other margins of adjustment, like labor income or working hours, can exist to cope with 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 one hand, Figure 14 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, retail 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 makes average formal wages go down. On the other hand, formal wages using PILA are slightly decreasing, but only in the latest periods (see Appendix Figure B.2).

Figure 14: Event study estimates on formal and informal labor income



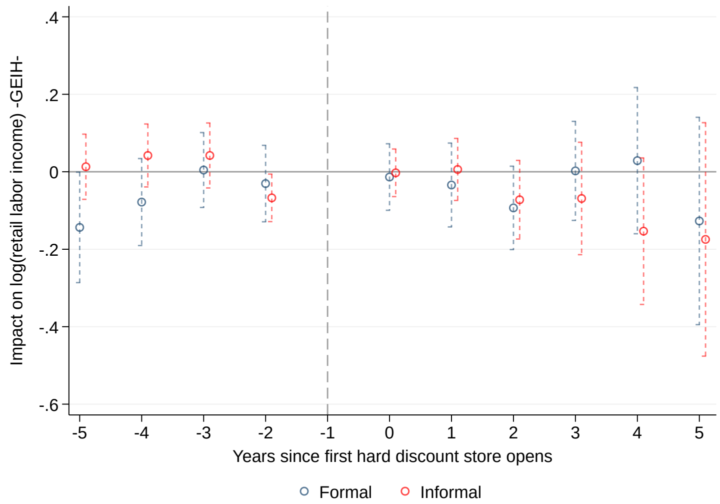
Note: We use the $C\&S$ estimator. The dependent variable is the logarithm of formal and informal labor income in a given municipality. Regressions were weighted with the local working-age population in the 2005 census. Observed treated municipalities are 191. Standard errors are clustered at the municipality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

Next, we measure the impact on the labor income of retail workers by sector. For informal workers, it is an indirect measure of the earnings of neighborhood shop owners, while for formal workers, it refers to regular wages. Figure 15 shows a negative trend in the latest post-treatment periods on the labor income of informal workers, indicating that the increase in competition from HDS could affect them more through the income margin than the employment one.[38] Due to their labor contract, they do not pay mandatory minimum wages to their workers, so they have income space to adjust to demand shocks. Even if the estimates are insignificant and imprecise, they are large ($ATT_{post} = -7.8\%$).[39]

[38] The estimated effect on the labor income of formal retail workers can seem contradictory to the pay premiums of hard discounters in subsection 3.2. However, the sample of analysis is different (for instance, the AKM sample includes capital cities), and the fact that HDS pay a national wage premium does not necessarily imply that this will increase average wages at the sector level in treated municipalities.

[39] However, when we add the distance control, the point estimates get close to zero (see Appendix Figure C.5).

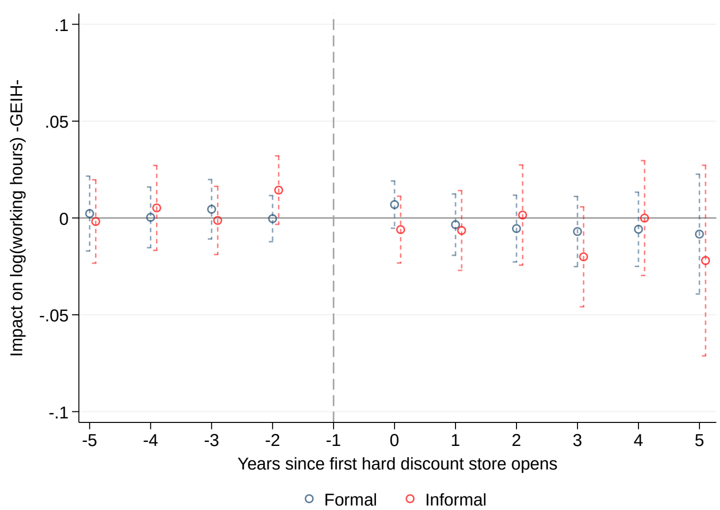
Figure 15: Event study estimates on labor income of retail workers by sector



Note: We use the *C&S* estimator. The dependent variable is the logarithm of formal and informal labor income 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.

Last, we focus on working hours across the informal and formal sectors. For these outcomes, we do not observe significant changes. Thus, the hypothesis that informal neighborhood shops can adjust to the increase in competition via working more hours is not occurring (see Figure 16).

Figure 16: Event study estimates on working hours by sector



Note: We use the *C&S* estimator. The dependent variable is the logarithm of formal and informal 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 to the inclusion of different controls and estimators. First, we use as a control the distance in kilometers from each municipality to the department's capital to adjust for the degree of connectivity to the largest city and have 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 economic structure in the service industries before the arrival of hard discount chains. With this in mind, Appendix Table C.2 and Panel A of Figure C.3 show the estimates of formal employment using these two controls with the survey data and with the outcome regressions estimation method of Sant'Anna and Zhao (2020). Overall, the positive effects remain robust across various specifications. Specifically, the estimate of total formal employment for the last three periods holds statistically significant with the control for the economic structure, yet its magnitude is reduced when we add the distance control. On the other hand, the coefficients for the primary and secondary sectors remain significant when using each control separately. In the last panel, we include both controls jointly, and the standard errors increase substantially. In this case, the coefficients on total formal employment and primary and secondary employment decrease but remain statistically significant at the 10% level.

Then, Appendix Table C.3 shows the results with the administrative data. 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, while the coefficient on commerce, hotels, and restaurants without retail gets more negative and becomes statistically significant.

Regarding informal employment, Table C.4 and Panel B of Figure C.3 show that the estimates are volatile when including controls, ranging from positive to negative values, but all are insignificant. Meanwhile, Figure C.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. Moving to Panel A of Table C.5, we replicate the ATTs from Figure 14 of wages and Figure 16 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 C.5, which shows robustness to the effects on the income of informal retail workers (Figure 15). Lastly, Table C.6 shows how robust the effects on labor market rates are after including controls. The estimates for employment and inactivity rates demonstrate stability across specifications, while the decrease in the unemployment rate becomes significant 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 C.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. To partially address this, we have excluded capital cities from our estimation sample. However, the estimates can still be biased by those municipalities that are more closely linked to capitals. To mitigate this concern, we conducted additional analysis excluding the six largest metropolitan areas.[40] This further exclusion drops 19 municipalities from our main sample. Panel C of Appendix Table C.7 reports that the treatment effects do not vary substantially in magnitude. Another concern is that since we use cohorts that aggregate different municipalities of GEIH that are not uniquely representative, the survey weights can induce biases in our results, so we remove them from the analysis to find a positive impact on formal employment and an insignificant one on informal employment, in line with our previous results (see Appendix Figure C.6).

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 C.7 and C.8 reports the estimates of formal employment with PILA and GEIH using the working-age population of the 2018 census as the denominator. There are no major changes

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

in the direction or statistical significance of the effects.[41] 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 1.2 pp. This reduction is mainly due to the downward adjustment of the dependent variables.[42]

7 Conclusion

In this paper, we study the impact of the expansion of Colombia's leading hard discount chains on local labor markets. Unlike previous studies examining the impact of large retailers like Walmart, we focus on the effects on a developing country where the informal sector is large. Notably, we find that after the introduction of HDS, there have been significant changes in local labor markets. First, HDS entry increases local formal employment, mainly in retail, manufacturing, and agricultural industries. These spillover effects suggest that the entry of HDS chains into a new municipality generates a positive formality chain as they demand inputs and goods from other local formal firms in their supply chain that also hire more formal workers to satisfy the increased demand. We also document that the employment effects are not immediate, as they show up to three years after the first opening of HDS, which is consistent with the time it takes for a hard discount chain to gain more customers and open more stores in treated municipalities. Besides, 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 incumbent informal retail sector, we find suggestive evidence that HDS do not decrease informal retail employment. However, we cannot empirically test if the entry of HDS is causing the exit of local neighborhood shops as we do not have 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 via earnings and not via quantity of employment, as neighborhood shops are small entrepreneurial activities run mostly by their owner. They rarely hire workers (only 16% of informal neighborhood shops have employees besides the owner).

We also find suggestive evidence that the positive impacts on employment also translate into aggregate effects in the local economies, as after the entry of HDS into municipalities, the share of collected taxes over total public revenues increases by 10.1 pp on average, and most of this increase is driven by the revenues

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

[42] The pre-treatment mean of the dependent variables reduces substantially; for instance, the share of formal employment lowers from 16 to 10.4 in PILA and from 28.3 to 18.4 in GEIH.

derived from the industry and commerce taxes and property taxes. Highlighting how relevant the entry of hard discount chains is in their local labor markets, where informal businesses coexist with formal retailers. These findings then have important policy implications for developing countries. Further research is needed to understand better the potential effects on neighborhood shop profits and the long-term implications of the expansion of HDS in developing countries.

References

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

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., Morales, L. F., et al. (2023). *The Labor Market Effects of Facilitating Social Security Contributions Under Part-Time Employment Contracts: Evidence from Colombia*. SSRN.

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. A. (2022). Immigration, wages, and employment under informal labor markets.

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].

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.

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

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.

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.

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.

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.

Sant'Anna, P. H. and Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122.

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 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.

A.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.

A.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.

A.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.[A.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.

[A.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).

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.[A.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 Additional Results

Table B.1: Selection into treatment by hard discount chain

	HDS1		HDS2		HDS3		Any chain	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(2005 Population)	0.207***	0.229***	0.149***	0.141***	0.204***	0.230***	0.261***	0.265***
	(0.017)	(0.014)	(0.033)	(0.037)	(0.029)	(0.022)	(0.020)	(0.017)
Log(Distance to department capital)	-0.012	-0.040**	-0.035	-0.010	-0.016	-0.020	-0.030	-0.039
	(0.020)	(0.018)	(0.029)	(0.024)	(0.024)	(0.019)	(0.021)	(0.024)
2005 Unsatisfied basic needs index	-0.007***	-0.005***	-0.003	-0.002*	-0.005***	-0.006***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2010 Share of retail formal employment	-0.004	-0.002	-0.003	-0.003	0.001	0.001	-0.005**	-0.004*
	(0.003)	(0.002)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)
2010 Share of tertiary formal employment	-0.002***	-0.001	-0.000	-0.000	-0.001	0.000	-0.003**	-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	0.003**	0.002	0.003***	0.001	0.002	0.005**	0.005***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Constant	-1.244***	-1.491***	-0.985***	-1.068***	-1.385***	-1.684***	-1.583***	-1.641***
	(0.192)	(0.144)	(0.260)	(0.362)	(0.269)	(0.222)	(0.238)	(0.182)
Observations	1,025	1,025	1,025	1,025	1,025	1,025	1,025	1,025
R-squared	.416	.488	.229	.411	.329	.465	.472	.534
Dep Var Mean	.265	.265	.157	.157	.221	.221	.374	.374
Dep Var SD	.442	.442	.364	.364	.415	.415	.484	.484
Department FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[A.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.

Table B.2: 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

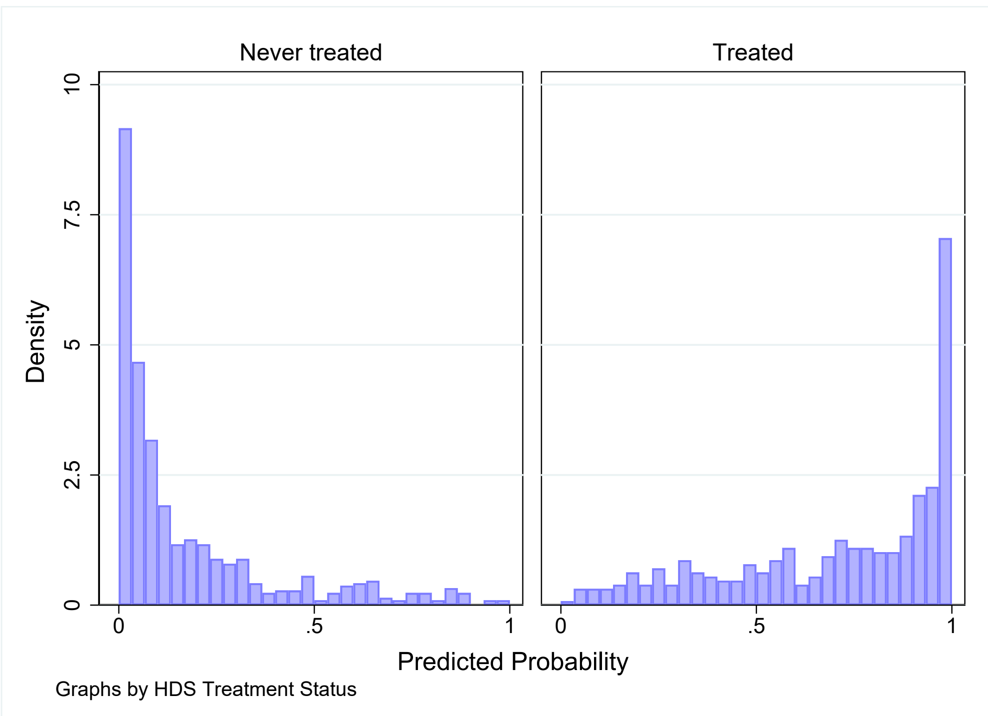
Notes: This Table shows the mean of corner shops by formality status. A business is formal if it has an updated register at a chamber of commerce. We use a USD to COP exchange rate of 3,250 (which approximates the average rate in 2019) and define corner shops as businesses classified under code 45 of the ISIC Revision 4: "Wholesale and retail trade 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.

Table B.3: Labor income and working hours outcomes for estimation sample with GEIH

	Treated				Not treated yet			
	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

Notes: 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 B.1: Distribution of predicted treatment probabilities



Notes: 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 B.1 (without department dummies). We then predict the probability, which is depicted in the x-axis.

Table B.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.

Table B.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 B.6: **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.

Table B.7: **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}	-0.418	-0.077	0.003	-0.075	-0.120
	(0.294)	(0.065)	(0.112)	(0.205)	(0.199)
ATT_{post}	1.742***	0.228**	-0.232	0.877***	0.720
	(0.607)	(0.115)	(0.175)	(0.296)	(0.462)
N	3,348	3,348	3,348	3,348	3,348
Clusters	372	372	372	372	372
Mean pre-treatment	16	1.2	2.9	4.5	11.8

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 B.8: **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.

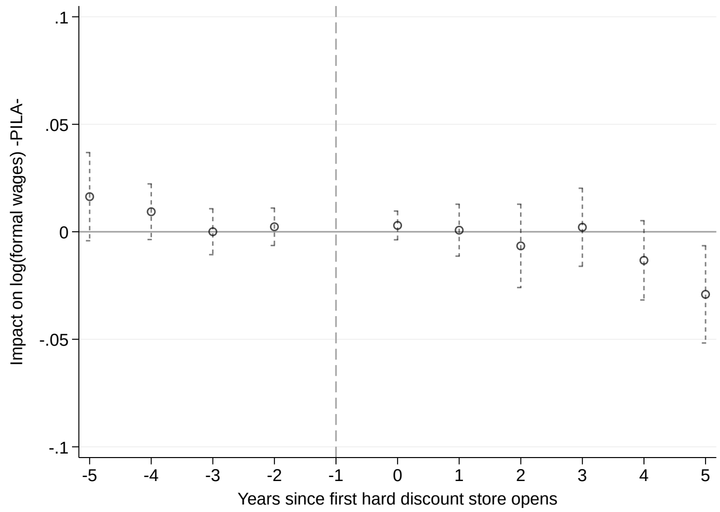
Table B.9: **Average $C\&S$ estimates of tax revenues by type**

	(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	146.8	14.4	43.9	40	18.8	44

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.

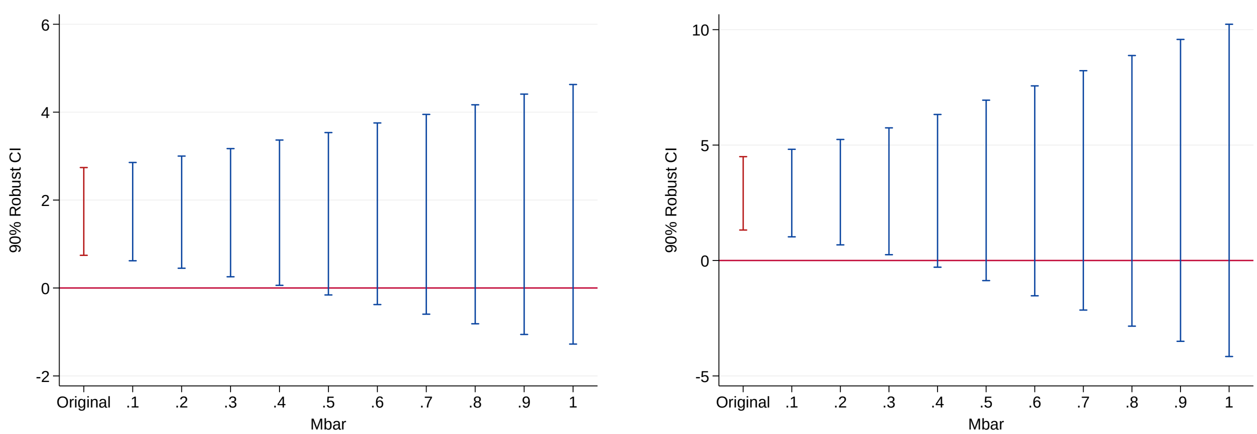
Figure B.2: Event study estimates on formal wages using PILA



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.

C Robustness Checks

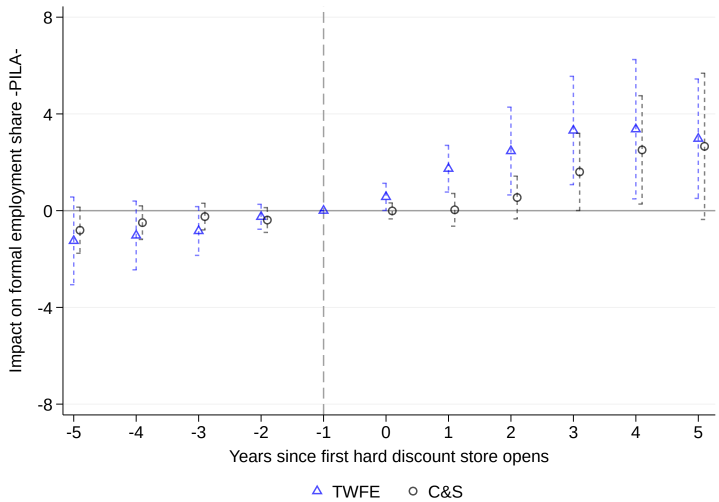
Figure C.1: Sensitivity analysis of ATT_{Post} for formal employment using PILA and GEIH



(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 C.2: Event study estimates on formal employment using PILA with the municipalities observed in GEIH



Note: 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 C.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	1,629	1,712
Clusters	191	191	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
Panel C: Never treated as control group (restricted)			
ATT_{pre}	1.232**	-0.853*	-0.617
	(0.614)	(0.517)	(0.513)
ATT_{post}	0.188	0.446	-0.645
	(0.685)	(0.588)	(0.731)
N	3,020	2,773	2,989

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. 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 B.1–Column 7. The propensity score was estimated using a logit model. Source: GEIH 2010-2018.

Table C.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
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
Panel C: Share of tertiary sector in 2010					
ATT_{pre}	0.046	-0.094	-0.031	0.170	0.001
	(0.786)	(0.181)	(0.184)	(0.393)	(0.554)
ATT_{post}	3.092***	0.565**	-0.056	1.814***	0.769
	(1.113)	(0.230)	(0.234)	(0.612)	(0.688)
$ATT_{post_{k=3,4,5}}$	4.808***	0.779***	0.177	2.561***	1.291
	(1.575)	(0.295)	(0.320)	(0.788)	(1.001)
N	1,719	1,719	1,719	1,719	1,719
Panel D: Log(Distance) and share of tertiary sector					
ATT_{pre}	0.280	-0.148	0.036	0.171	0.221
	(0.903)	(0.183)	(0.195)	(0.411)	(0.633)
ATT_{post}	1.915	0.458*	-0.040	1.152	0.345
	(1.471)	(0.238)	(0.278)	(0.836)	(0.994)
$ATT_{post_{k=3,4,5}}$	3.708*	0.675**	0.107	1.932*	0.994
	(2.121)	(0.293)	(0.335)	(1.109)	(1.497)
N	1,710	1,710	1,710	1,710	1,710

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

Note: This table reports the robustness of the formal employment estimates (see Table B.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.

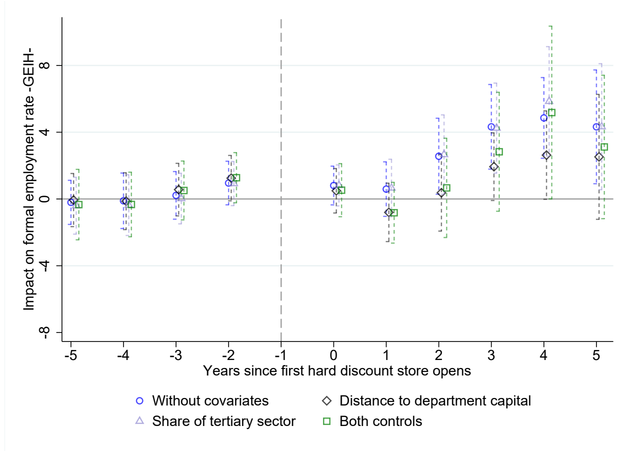
Table C.3: **Average $C\&S$ estimates of formal employment by sector using PILA and including controls**

	(1) Total	(2) Retail	(3) CHR without retail	(4) Primary and secondary	(5) Services without CHR
Panel A: Without controls					
ATT_{pre}	-0.418	-0.077	0.003	-0.075	-0.120
	(0.294)	(0.065)	(0.112)	(0.205)	(0.199)
ATT_{post}	1.742***	0.228**	-0.232	0.877***	0.720
	(0.607)	(0.115)	(0.175)	(0.296)	(0.462)
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}	-0.518*	-0.111*	0.032	-0.107	-0.205
	(0.315)	(0.059)	(0.112)	(0.199)	(0.211)
ATT_{post}	1.928***	0.266**	-0.398***	1.091***	0.802*
	(0.656)	(0.107)	(0.140)	(0.230)	(0.461)
N	3,339	3,339	3,339	3,339	3,339
Panel C: Share of tertiary sector in 2010					
ATT_{pre}	-0.482	-0.085	0.010	-0.105	-0.146
	(0.303)	(0.066)	(0.127)	(0.222)	(0.204)
ATT_{post}	1.757***	0.241**	-0.235	0.896***	0.691
	(0.606)	(0.116)	(0.187)	(0.297)	(0.475)
N	3,348	3,348	3,348	3,348	3,348
Panel D: Log(Distance) and share of tertiary sector					
ATT_{pre}	-0.571*	-0.115**	0.034	-0.137	-0.201
	(0.317)	(0.058)	(0.115)	(0.215)	(0.210)
ATT_{post}	1.959***	0.287***	-0.397***	1.138***	0.726
	(0.640)	(0.104)	(0.147)	(0.240)	(0.449)
N	3,339	3,339	3,339	3,339	3,339

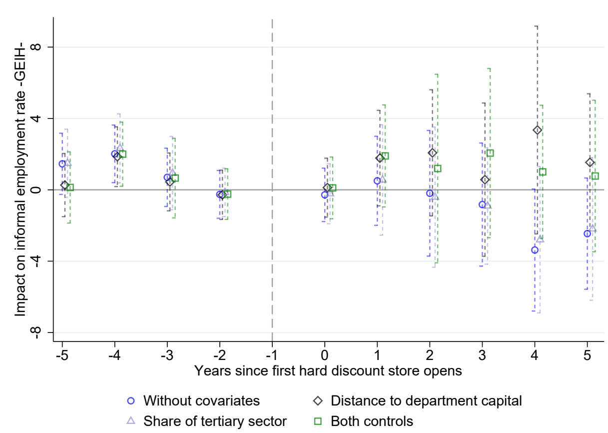
Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table reports the robustness of the formal employment estimates using PILA (Table B.7) 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 C.3: Event study estimates on type of employment using controls



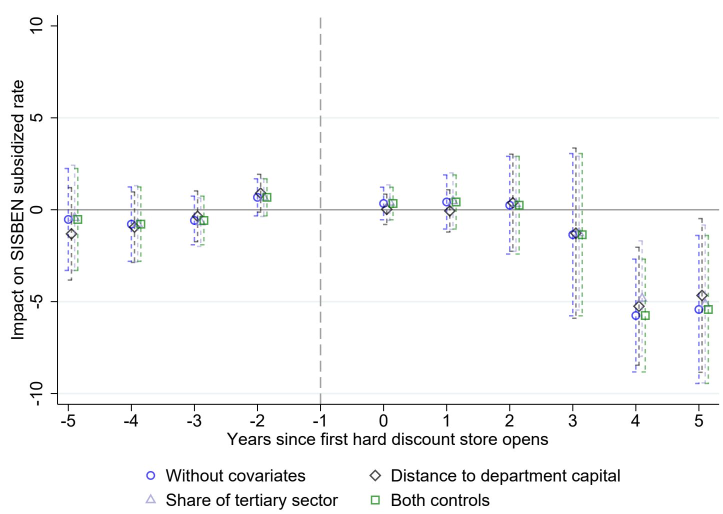
(a) Formal employment



(b) Informal employment

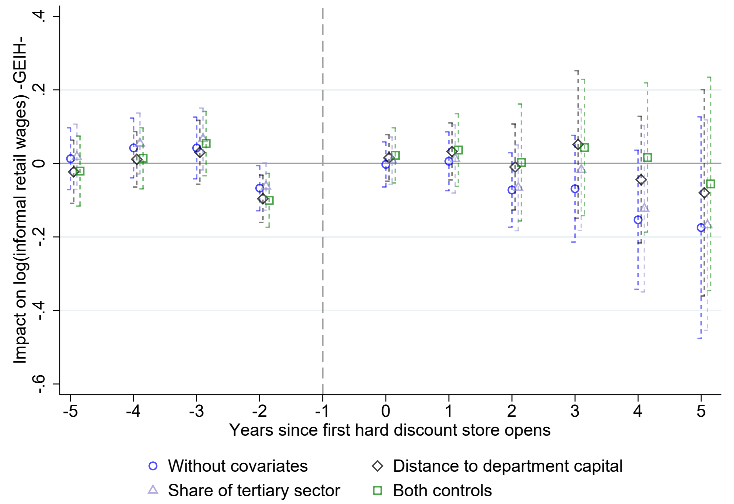
Note: The dependent variable is the type of employment relative to the working-age population in the 2005 census in a given municipality. We control for the logarithm of the driving distance from the municipality to the department capital in kilometers and for the share of tertiary sector employment 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 C.4: Event study estimates on subsidized SISBEN share using controls



Note: The dependent variable is the number of subsidized individuals 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 C.5: 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 C.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
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	-0.018
	(2.106)	(1.541)	(1.153)	(1.532)	(1.221)
N	1,710	1,710	1,710	1,710	1,710
Panel C: Share of tertiary sector in 2010					
ATT_{pre}	1.150	0.146	0.461	-0.750	1.293**
	(0.902)	(0.393)	(0.406)	(0.504)	(0.542)
ATT_{post}	-0.995	0.581	-0.396	-0.845	-0.334
	(1.642)	(0.583)	(0.484)	(1.033)	(0.555)
$ATT_{post_{k=3,4,5}}$	-1.965	0.484	-0.996	-0.412	1.291
	(1.913)	(0.804)	(0.707)	(1.363)	(1.001)
N	1,719	1,719	1,719	1,719	1,719
Panel D: Log(Distance) and share of tertiary sector					
ATT_{pre}	0.638	0.027	0.426	-0.873	1.058**
	(0.811)	(0.419)	(0.412)	(0.597)	(0.485)
ATT_{post}	1.174	1.361	0.446	-0.527	-0.106
	(1.244)	(1.038)	(0.720)	(1.298)	(0.597)
$ATT_{post_{k=3,4,5}}$	1.280	1.711	0.219	-0.263	0.994
	(1.783)	(1.484)	(0.951)	(1.805)	(1.497)
N	1,710	1,710	1,710	1,710	1,710

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

Note: This table reports the robustness of the informal employment estimates (Table B.8) 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 C.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	1,714	1,714	1,714	1,714
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	1,710	1,710	1,710	1,710
Panel C: Share of tertiary sector in 2010				
ATT_{pre}	0.030	0.006	0.002	0.004
	(0.022)	(0.017)	(0.010)	(0.010)
ATT_{post}	-0.014	-0.037	-0.004	-0.011
	(0.042)	(0.030)	(0.009)	(0.013)
N	1,719	1,719	1,719	1,719
Panel D: Log(Distance) and share of tertiary sector				
ATT_{pre}	0.016	-0.019	-0.001	-0.007
	(0.026)	(0.017)	(0.009)	(0.012)
ATT_{post}	-0.026	-0.005	-0.004	0.013
	(0.047)	(0.030)	(0.010)	(0.018)
N	1,710	1,710	1,710	1,710

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

Note: This table reports the robustness of the wages and working hours estimates (Figures 14 and 16) 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 C.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	1,629	1,712
Clusters	191	191	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
Panel C: Share of tertiary sector in 2010			
ATT_{pre}	0.887	-0.446	-0.643
	(0.567)	(0.417)	(0.474)
ATT_{post}	2.430**	-1.121*	-1.960**
	(0.952)	(0.635)	(0.873)
N	1,719	1,674	1,713
Panel D: Log(Distance) and share of tertiary sector			
ATT_{pre}	1.167*	-0.450	-0.902*
	(0.644)	(0.590)	(0.542)
ATT_{post}	2.331***	-1.056	-1.824***
	(0.747)	(0.670)	(0.606)
N	1,710	1,665	1,704

Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table reports the robustness of the labor market rates estimates (Table B.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 C.7: **Average $C\&S$ estimates of formal employment by sector using 2018 working-age population and excluding metropolitan areas (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}	-0.418	-0.077	0.003	-0.075	-0.120
	(0.294)	(0.065)	(0.112)	(0.205)	(0.199)
ATT_{post}	1.742***	0.228**	-0.232	0.877***	0.720
	(0.607)	(0.115)	(0.175)	(0.296)	(0.462)
N	3,348	3,348	3,348	3,348	3,348
Mean pre-treatment	16	1.2	2.9	4.5	11.8
Panel B: 2018 census working-age population					
ATT_{pre}	-0.165	-0.037	-0.002	-0.010	0.013
	(0.202)	(0.045)	(0.073)	(0.135)	(0.130)
ATT_{post}	1.088***	0.140*	-0.123	0.548***	0.438
	(0.376)	(0.077)	(0.125)	(0.202)	(0.296)
N	3,348	3,348	3,348	3,348	3,348
Mean pre-treatment	10.4	.79	1.9	2.9	7.7
Panel C: Excluding municipalities in metropolitan areas					
ATT_{pre}	-0.549*	-0.088	0.057	-0.144	-0.238
	(0.317)	(0.072)	(0.126)	(0.229)	(0.222)
ATT_{post}	1.732**	0.250*	-0.226	0.858***	0.657
	(0.676)	(0.128)	(0.190)	(0.309)	(0.510)
N	3,213	3,213	3,213	3,213	3,213

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

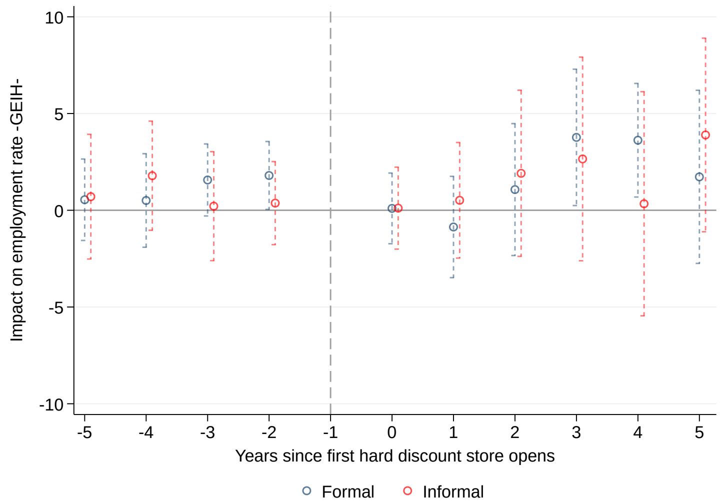
Note: This table reports the robustness of the formal employment estimates using PILA (Table B.7) by substituting the denominator of the dependent variables and excluding metropolitan areas. Panel B uses the 2018 census working-age population (individuals aged 15 years or older) to construct the shares, 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 C.8: **Average $C\&S$ estimates of formal employment by sector using 2018 working-age population (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
Mean pre-treatment	28.3	2.6	2.2	8.4	15.1
Panel B: 2018 census working-age population					
ATT_{pre}	0.368	-0.010	-0.055	0.379*	0.054
	(0.395)	(0.098)	(0.086)	(0.200)	(0.297)
ATT_{post}	1.234**	0.247**	0.002	0.610**	0.374
	(0.495)	(0.105)	(0.103)	(0.246)	(0.308)
$ATT_{post_{k=3,4,5}}$	2.098***	0.380***	0.206	0.899***	0.612
	(0.645)	(0.132)	(0.139)	(0.325)	(0.423)
N	3,770	3,770	3,770	3,770	3,770
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 B.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 C.6: 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.