Effects of e-commerce on local employment and firms' dynamics

Camila Steffens*

Working Paper
January/2023

Abstract

E-commerce has been growing in the last decades, but little is known about its effects on labor markets. In this paper, I provide the first evidence for the effects of e-commerce on local employment and firms in a developing country. I employ a Difference-in-Differences design, exploiting state-border discontinuities of shipping costs and the expansion of a major e-commerce player to the Northeast region of Brazil. I show that each percentage point increase in e-commerce exposure decreases local employment by 6%. These findings highlight the importance of studying the effects of e-commerce on workers' mobility and wages.

Keywords: E-commerce, Retail, Local labor markets, Employment.

*Universidad Carlos III de Madrid (csteffen@eco.uc3m.es). Acknowledgements: I acknowledge the excellent feedback and comments from my advisor Jan Stuhler, and from Alan Crawford, Luigi Minale, Jesús Fernández-Huertas Moraga, Daniel Rees, Mateus Souza, and participants at the Applied Economics Reading Group at Universidad Carlos III de Madrid. All errors are my own.

1 Introduction

E-commerce participation has been increasing in the 21st century, and especially during the last decade, due to the strong transition to a digital economy.[1] As an important technological change in the retail sector, e-commerce is expected to reduce search costs (as a consequence, prices), and increase the varieties of goods available in the market (Hortaçsu and Syverson, 2015; Goldmanis et al., 2010). For firms, e-commerce is assumed to lower trade costs and raise market access through technological innovations and changes in supply chains (UNCTAD, 2021).

However, this technological change might also represent more intense competition for local retailers. For instance, e-commerce is expected to increase the concentration of firms in more populated areas that have better logistics networks (Hortaçsu and Syverson, 2015). There is also evidence that e-commerce might shift market share from smaller to larger businesses (Goldmanis et al., 2010). This raises the question of how labor markets might be affected by e-commerce expansion. While there is a growing literature studying the effects of e-commerce on labor outcomes (Chava et al., 2022; Bauer, 2021; Chun et al., 2020), findings are still divergent. Moreover, the current research is restricted to developed economies. My paper is the first to study the labor market effects of e-commerce in a developing country, where the retail sector contributes to a substantial fraction of employment.

I study the effects of e-commerce on employment and the number of firms in Brazilian local labor markets. Brazil offers an ideal setting to study the impacts of e-commerce. The country's vast territory, served mostly by road transportation, and characterized by low infrastructure development, generates regional heterogeneity in the access to online markets due to shipping costs. In this setting, online sales often require high investments in logistics and technology, which triggers the concentration of the online market in few firms. Due to investments of the main players, e-commerce has been taking off in the country, with an annual growth of 20% since 2011.

I leverage the expansion of e-commerce to the Northeast region of Brazil in 2011 as a quasi-random shock to local labor markets. Until 2010, the main e-commerce players were concentrated in the Southeast (i.e., Sao Paulo), from where the purchases to the entire

[1] For instance, the share of e-commerce on global retail trade rose from 10.4% in 2017 to more than 14% in 2019 (UNCTAD, 2021).

country were delivered. In order to increase its participation in the Northeast, the major e-commerce company opened a distribution center (DC) in the state of Pernambuco. This implied a substantial reduction of shipping costs in that state, which are usually paid by consumers and added to the final price of the goods (E-bit, 2015). I leverage this increased exposure to e-commerce to assess the effects on local labor markets in that state from 2012 to 2016.[2] Further, the availability of matched employers-employees administrative data allows to assess the effects of e-commerce on the universe of exposed firms and workers.

In my main empirical strategy, I compare formal employment and the number of firms in local labor markets within the state that hosted the distribution center to local labor markets in the other states in the Northeast using a Difference-in-Differences (DiD) design. I leverage state-border discontinuities of the pricing policy of Brazil's national shipping company, Correios, which had the largest share of the delivery market in the period of analysis (Correios, 2020). This policy imposes a fixed shipping cost to all the municipalities within a given state, while costs can more than double for deliveries that cross state borders, even for short travel distances. Due to this discontinuity in the shipping costs, I find that local labor markets in the state hosting the DC were strongly exposed to e-commerce. Measuring the variation in e-commerce share at the state level with data from the Consumption Surveys of 2008/2009 and 2017/2018, I provide evidence that the opening of the DC increased the share of e-commerce in total sales of electronics and home appliances by more than 3 percentage points in the hosting states.

Using the census data of formal establishments in Brazil from 1997 to 2016, I found that e-commerce reduced employment by 20% and the number of firms by 12% in the most exposed retail sectors (electronics, home appliances, furniture). The longitudinal characteristic of the data allowed the decomposition of this effect between inflow and outflow of firms. The negative impact of e-commerce is explained by a reduction in the entrance of new firms in the short run (specially 2013 and 2014), while there is no evidence of effects on firms' exit.

Based on a triple-differences that leverages the national share of e-commerce in total sales across many retail sectors, I found that 1 p.p increase in e-commerce share in the

[2]Many events happened after 2016, such as the integration of physical stores of the main players to e-commerce, allowing for stores pick-up; opening of distribution centers in other states; entrance of Amazon in 2019; and the COVID pandemic in 2020. For this reason, I restrict the analysis in this paper to the period up to 2016.

sector decreased employment by 1.9% up to 2016. Moreover, additional evidence shows that retail categories more directly competing with e-commerce had the largest reduction in employment, while sectors such as shipping, hospitality, beauty services, warehouse, real estate, and informatics services had grown.

The statistical significance of such effects is robust to inference methods that take into account cluster-level correlations at the state level using permutation of residuals across the temporal dimension (Chernozhukov, Wüthrich, and Zhu, 2021) and the comparison of the effects at the cross-section dimension (Hagemann, 2020).

I argue that the opening of a DC, combined with the discontinuities in delivery costs, captures an increase in the potential exposure to e-commerce and would not impact other factors that are also related to employment in retail sectors. Indeed, I show that the evolution of the outcomes were similar in the treated and untreated groups before the treatment. Moreover, the results are similar when controlling for urban population and GDP per capita in 2010, growth in GDP per capita, and log distance to the state capital. In addition, the effects are robust to a Synthetic Difference-in-Differences at the capital level.

I highlight that the findings in this paper can be seen as a lower bound of overall e-commerce effects on local labor markets, since the control group might also have been affected by e-commerce to some extent. This paper contributes to the literature on local labor market responses to different shocks by providing the first evidence of the effects of e-commerce exposure on retail employment in a developing country. The literature has mostly focused on supply-side shocks through immigration, such as Dustmann, Schönberg, and Stuhler (2017); on international trade liberalization by leveraging the exposure to Chinese imports (Kovak, 2013; Dix-Carneiro and Kovak, 2019; Kim and Vogel, 2021); and on local employment responses to recessions (Blanchard and Katz, 1992; Yagan, 2019). The growing literature in e-commerce is restricted to developed countries, finding smaller effects than those in this paper, in the order of 3% in the U.S (Chava et al., 2022) and Korea (Chun et al., 2020). In contrast, Bauer (2021) found that the reduced exposure to e-commerce (due to the introduction of taxation on online goods) might have decreased local retailer employment in the U.S.

The larger effects found in this paper might arise for several reasons. First, the importance of retail in local economies in developing countries. For instance, the retail sector employed around 35% of formal workers in the Northeast region of Brazil. In addition, I observe the population of firms, instead of a sub-sample of large retailers as in Chava et al. (2022). I also take into account relative exposure of different retail categories, and estimate the effects on the sectors that were more directly competing with e-commerce. Differently from Bauer (2021), I study the effects of incipient e-commerce growth, which can be understood as increased competition to local retailers. Furthermore, in my setting, the e-commerce shock is not only through a reduction in shipping time as in the U.S., but also a decrease in effective prices for consumers due to the transference of shipping costs.

My results on local labor markets also contribute to the literature that relates internet spread, and specifically e-commerce, with growth in regional inequality. Forman, Goldfarb, and Greenstein (2012) suggest that the spread of the internet exacerbated regional wage inequality among US counties from 1995 to 2000, since there was no evidence of economic improvement in peripheral areas. Agrawal (2021) also found that internet penetration (through e-commerce) changed the spatial pattern of tax revenues among municipalities in the US, benefiting few municipalities where the online vendors were located. For the majority of towns, on the contrary, higher internet penetration resulted in lower local sales taxes. More recently, Couture et al. (2021) found that e-commerce expansion had no significant effects on the local production and income of rural households in China. However, their results suggest that e-commerce reduced the cost of living of richer rural households that had access to online shopping. My findings are in line with this literature, adding further evidence for e-commerce's effects on local employment.

The remainder of this paper is organized as follows. The next section describes the institutional background of e-commerce exposure in Brazil and the data. Section 3 presents the empirical strategy. The results are discussed in Section 4, and section 5 concludes the paper.

2 Background and Data

Online retail shopping emerged in Brazil mostly in the 2010s, following investments from the main local players. Since then, the prevalence of e-commerce has been steadily increasing, such that 29.4% of the population made at least one online purchase in 2019 (J.P. Morgan, 2019; ITA, 2021). Although its share in total retail was around of 3%

by 2017, there is significant heterogeneity among sectors. Online shopping represented 7% of retail sales by 2017 when excluding sectors in which it is usually limited (such as food and beverage stores, supermarkets, fuel stations and construction stores). In addition, some categories of goods were strongly exposed to e-commerce. For instance, e-commerce represented only 2.7% of sales from electronic stores by 2012, and this share reached almost 15% in 2017 (Table 1).[3]

One of the main challenges for e-commerce in Brazil is the transportation logistics. Brazil's vast territory is served mostly by road transportation, and is characterized by low infrastructure development, which increases shipping cost. For this reason, such costs are usually transferred to the consumers and added to the final price of the good.[4] During the 2010s, the shipping markets was dominated by a public company (*Correios*), which was responsible for more than 80% of deliveries in 2013, and almost 60% by 2017 (Correios, 2020).

The prices of Correios are legislated by the Ministry of Communications. With the aim to connect distant regions with affordable shipping costs, the price policy consisted of a fixed shipping cost to all municipalities that are within a same state, despite the distance. Across state borders, the shipping cost is discontinuous. To illustrate, a delivery from São Paulo to the Northeast of Brazil costs three times more than deliveries within a state.[5] While the shipping costs are fixed within states, the delivery time varies, ranging from one day to up to ten business days depending on the distance. In addition, same-day deliveries (or 24-hour deliveries) are only available to capitals and metropolitan areas (Correios, 2021).

Until 2010, e-commerce companies were concentrated in the Southeast (specifically, in the state of Sao Paulo), from where all the purchases were delivered. In 2011, a major e-commerce player (B2W) opened a Distribution Center (DC) in the state of Pernambuco, in the Northeast region of Brazil. This company was the main e-commerce player selling products directly to consumers (i.e., not restricted to marketplace), with around 30% of

[3]Electronic consists of informatics and communication technologies. Home appliances are in a separate category.

[4]For instance, around 57% of online purchases were made with paid shipping in 2015 (E-bit, 2015).

[5]I obtained the prices per package weight and within/between states for each Brazilian state from 2011 to 2020 directly from Correios through the Brazilian Information Access Law.

e-commerce market share in Brazil.[6]

Figure 1 shows the location and the distance to the nearest DCs in Brazil in 2011. With the expansion to the Northeast, e-commerce exposure was expected to increase in this region due to reduced shipping costs and delivery time. However, due to the discontinuity in the shipping costs from Correios, such exposure would be larger in the state that hosts the DC. Local labor markets within Pernambuco would face an even large decrease in the cost compared to other states in the region. Within the state, more distant local labor markets might be potentially less exposed only by differences in shipping time.

I highlight that, from 2012 to 2016, there were not significant changes in the location of DCs in the Northeast of Brazil.[7] Therefore, local labor markets in Pernambuco were potentially more exposed to e-commerce during this period. In 2017, new distribution centers were opened in the region, and the major e-commerce players integrated online purchases with physical stores, allowing store pick-ups.[8]

2.1 Data

The main source for local labor market outcomes in this paper is the establishment data from the Relação Anual de Informações Sociais (RAIS) from 1997 to 2016. RAIS consists of a yearly census of all public and private firms that employ at least one formal worker in Brazil. The data contain information about the quantity of workers employed as of December; whether the firm is public or private; the municipality where the firm is located; and a five-digit classification of economic activity (CNAE). There are almost 1.8 million firms in the Northeast region in the period of analysis.

The economic activities were aggregated into 43 main sectors, which are listed in Appendix Table A.1. This paper primarily focuses on retail activities which are directly competing with e-commerce. Therefore, the following categories are considered as retail

[6]Another important e-commerce player is the marketplace *Mercado Livre*, with a market share from 30% to 40%. However, since *Mercado Livre* is only for third-party sales, we might not expect a concentration in a specific region. According to Mercado Livre (2021), there are currently 15,480 sellers enrolled in the marketplace, and, until 2017, the deliveries were made by each individual seller located in many cities across Brazil, using the services of Correios. Amazon only fully entered the Brazilian market in 2019. Another important player is Magazine Luiza, which had less than 10% of market share in the period of analysis, and was only delivering from Sao Paulo by 2013/2014.

[7]B2W further expanded in the South and Southeast regions in this period.

[8]Magazine Luiza integrated its more than 1,300 physical stores across Brazil to e-commerce in 2016, and opened a DC in Ceará in 2017, increasing its market share in online sales. B2w was integrated to its controlled company by 2018, a major department stores' chain with more than 1,700 stores in Brazil (B2W, 2021).

for the purpose of this paper: (1) informatics and communication goods (defined, in general, as *electronics*); (2) home appliances; (3) furniture; (4) bookstore and newspaper; (5) cosmetics and pharmacy; (6) clothing; (7) shoes; (8) fabrics; (9) other goods (optics, jewelry, etc.).[9]

I also obtained the share of e-commerce on the total sales of retail sectors at the national level based on Annual Retail Survey (IBGE, n.d.). Table 1 summarizes e-commerce exposure in these sectors in 2012, 2015 and 2017. Finally, information about population and GDP are from the Brazilian Institute of Geography and Statistics (IBGE, 2018).

3 Empirical Strategy

In this paper, I leverage the expansion of a major e-commerce player to the Northeast as a quasi-random shock to local labor markets, which are defined at the microregion level according to the Brazilian Institute of Geography and Statistics (IBGE). In addition, I estimate aggregate effects at the state and respective capital hosting the DC using a Synthetic-DID design. The outcomes consist in the number of formal workers and firms aggregated by sector at the microregion, capital or state levels.

Due to the discontinuity on the price policy of *Correios* across states, I assume that local labor markets in the state hosting a Distribution Center (DC) would be strongly exposed to e-commerce. Using data from the Brazilian Consumption Surveys of 2008/2009 and 2017/2018, I provide evidence that the opening of a B2W's DC in a state is related to an increased exposure to e-commerce by 3.7 percentage points (Table 2). This difference remains significant after excluding Sao Paulo, by 3 percentage points. When comparing only capitals, the share of e-commerce was 7 percentage points larger in those hosting a DC.

3.1 Effects of e-commerce in local labor markets

In order to measure the effects of the e-commerce on local labor markets, I compare the evolution of formal employment and the number of firms in the microregions in the state that hosted the DC (Pernambuco) with microregions in the other eight states in the Northeast region of Brazil based on a Difference-in-Differences (DiD) design. In total, there are 18 treated microregions and 169 non-treated microregions. Figure 2 shows the

[9] I categorized activities that were not integrated to e-commerce apart from retail, such as supermarkets, food and beverage sellers, fuel stations, vehicles and vehicle reparations stores.

treatment and control groups and their location in Brazil.

Although the opening of DCs can be considered as exogenous to non-host microregions located in the same state, the exposure of local markets to e-commerce might be non-random (Borusyak and Hull, 2020). DCs are usually located in the center of the main markets in Brazil. The expansion to the Northeast was motivated by the perceived regional economic growth and the increase of consumer's purchase power, specially after the rise of public policies that highly benefited the region. Furthermore, municipalities closer to the DC' headquarter are expected to benefit more from infrastructure and market access available in the center of the regional economic axis. Thus, they should also be more exposed to e-commerce than the municipalities further from the headquarters, even in the absence of the opening of DC. For this reason, I include a measure of centrality based on the log distance of each local labor market to the state capital as a covariate. I also control for GDP per-capita and urban population in 2010, and GDP growth from 2005 to 2010, so previous to the treatment. Further, by restricting the comparison group to the entire Northeast region, I assure that the treated and non-treated states are more similar in socioeconomic characteristics.

As a main strategy, I estimate the effects of e-commerce on employment and number of firms in electronics, home appliances, and furniture, which I define collectively as *household goods*. These were the main products supplied by the major e-commerce firm and concentrated the largest share of e-commerce sales during the period of analysis (Table 1).[10] In addition, I estimate the effects on overall retail (shown in Table 1) and in shipping and warehouse. Using longitudinal data from 2006 to 2016, the main regression specification is as follows:[11]

$$y_{i,t} = \mu_i + \tau_t + \sum_{t \neq 2010} \gamma_t (D_i \cdot \tau_t) + \beta_t X_i \cdot \tau_t + u_{i,t} \qquad (1)$$

where $y_{i,t}$ is log of number of workers or firms in microregion i at year t; D_i is a dummy variable equal to 1 for microregions in Pernambuco, zero otherwise; μ_i and τ_t

[10] The consumption of electronics represented more than 42% of total e-commerce value in 2019. E-commerce participation in other traditional markets, such as clothing, health and beauty, was not substantial in Brazil (J.P. Morgan, 2019).

[11] After 2016, there have been many events that are expected to impact e-commerce, but require different empirical strategies for the proper identification of their causal effects, such as the integration of e-commerce with the physical stores, the expansion of Mercado Livre and Amazon, and, more recently, the COVID-19 pandemic. I will analyze some of these events in future research.

are microregion and year fixed effects; and X_i is i's log distance to the capital, GDP per capita and urban population in 2010, and growth in GDP per capita from 2006 to 2010. The absence of effects in the pre-treatment years give supporting evidence to the assumption that employment and number of firms were having a similar evolution in the treated state compared to the non-treated states in the absence of the treatment.

In an additional analysis, I estimate the effects of e-commerce on all the sectors shown in Table 1 with a continuous treatment variable based on the level of exposure to e-commerce at the national level in 2017. Using a triple-differences approach, I interact the measure of exposure with the indicator of treatment, as follows:

$$y_{i,s,t} = \mu_i + \tau_t + \eta_s + \sum_{t \neq 2010} \gamma_t (D_i \cdot E_s \cdot \tau_t) + \sum_{t \neq 2010} \lambda_t (D_i \cdot \tau_t) + \alpha (E_s \cdot D_i) + \sum_{t \neq 2010} \delta_t (E_s \cdot \tau_t) + \beta_t X_i \cdot \tau_t + u_{i,s,t} \quad (2)$$

where η_s is sector s fixed effects, and E_s represents the 2017 national level exposure to e-commerce of sector s (Table 1). The large share of other goods/department stores (9.5%) is due to the fact that B2W is registered in these economic activities.

3.2 Effects of e-commerce in the hosting capital

The main identification strategy in the paper is based on the discontinuity of the shipping cost of Correios across states. However, over time, the main e-commerce companies started to invest in own logistics services. B2W acquired two logistic/transportation companies in 2013 and 2014 in a process of vertical integration of shipping, with two main objectives: (I) reduce shipping time; and (II) increase revenues from shipping. These policies contributed to the reduction in the share of sales with free shipping to 15% by the first quarter of 2017 among the top 10 companies in Brazil.

The vertical integration of logistics by B2W allowed the implementation of a shipping policy, named "Shipping Menu," which offered express delivery (48 hours) for purchases in the capitals hosting a distribution center. Further, from 2015 on, it offered fast shipping (3 to 10 days) and economic shipping (6 to 20 days) for all Brazilian cities. Following these policies, B2W's distribution process was responsible for 59% of the deliveries of the company by the end of 2015.

The fact that express delivey was only offered to capitals hosting a DC motivates the focus on the capitals in the Northeast. Even after the reduced participation of Correios, the capital of Pernambuco, Recife, was strongly exposed to e-commerce through faster and cheaper shipping services compared to the other capitals in the Northeast. In order to study the effects of e-commerce in the hosting capital, I rely on a Synthetic Difference-in-Differences (SDID) approach from Arkhangelsky et al. (2021). Instead of requiring an exact fit of pre-treatment outcomes, the SDID matches on pre-treatment trends in order to reweight the comparison units. The synthetic control group follows a similar trend in the pre-treatment period as the treated unit, which allows for differences in the level of the outcome. In addition, the SDID also reweights the pre-treatment years by giving larger weight on "periods that are on average similar to the target (treated) periods" (Arkhangelsky et al., 2021).

For this analysis, I expand the donor pool by including four states in the Center-West region of Brazil. The South and Southeast are not included because they were also treated in the period of analysis. The North is excluded because this region has very different socioeconomic characteristics, small population density, and smaller e-commerce penetration due to geographic characteristics. The donor pool is composed of twelve capitals. The post-treatment period is from 2011 to 2016, while the pre-treatment period is from 2001 to 2010.

I estimate average treatment effects on log of workers and variation in the number of firms using the *Stata* command *sdid*. The standard errors are based on permutations of the treatment across the control units. I also control for urban population in 2010 and GDP growth from 2005 to 2010. Figure 16 shows the weights given for each state capital when the outcome is the log of employment. The results are discussed in the next section.

4 Results

4.1 Effects of e-commerce in local labor markets

The estimated effects of e-commerce on employment in household goods stores (i.e., electronics, home appliances and furniture) are shown in Figures 3. While the coefficients are very close to zero previously to the treatment, we can observe a decreasing trend in employment from 2012 onward. The effects ranges from 9% reduction in 2013 to 16% in 2015 and 20% in 2016. The figure also show that the results are robust to controlling or not for urban population and GDP per capita in 2010, growth in GDP per capita, and log distance to the state capital.

When considering retailers that were not as exposed to e-commerce or directly competing with B2W, in general, the effects on employment are not indistinguishable from zero (Figure 4). In fact, employment increased by 6% in 2011, but this effect disappears in the following years. As an alternative approach, I estimated the effects of e-commerce on all the retail sectors listed in Table 1 using a triple DID. For that, I interacted the treatment indicator with a continuous variable capturing the exposure to e-commerce in 2017 at the national level.

The effects shown in Figure 5 are stronger when not considering the sector in which B2W is registered (department stores/other goods). Although not significant at 10%, the net effect when considering B2W's sector is in the order of 1.2% decrease in employment for each 1 percentage point increase in the exposure to e-commerce. Otherwise, the findings suggest that 1 percentage points increase in the share of e-commerce decreased employment in local retail by 1.9% in 2016. The largest exposure is 15% in electronic goods by 2017, which represents an increment of 12 percentage points after 2012. Based on a back-of-envelope calculation, this implies a reduction in employment by 22.8% in local electronic stores until 2016.

4.1.1 Firms' dynamics

The negative effects of e-commerce on employment are also observed in the number of firms (Figure 6). The cumulative impact on the number of firms in household good stores ranges from 7% in 2014 to 12% decrease by 2016. The longitudinal aspect of the data allows me to disentangle this effect between inflow and outflow of firms. For that, I measured the number of firms entering and exiting the market in each year. The results shown in Figures 7 and 8 provide evidence that the effects on firms is explained by the reduced entrance of new firms into the market in the short run, specially in 2013 and 2014, by 10%. In contrast, there is no evidence of larger exit of incumbent firms in the treated local labor markets as a response to e-commerce.

Similar to the findings for employment, there is no evidence of reduction on the number of firms in retailers not directly competing to e-commerce (Figure 9). Moreover, the point estimates of the continuous treatment according to the exposure of the sector to e-commerce are similar in magnitude, although less precise compared to the average effects on the most exposed sector (Figure 10).

4.1.2 Elasticity

In order to obtain a measure of the elasticity of employment and number of firms to e-commerce exposure, I instrument the share of e-commerce in total sales with an indicator for the opening of the Distribution Center of B2W in the state of Pernambuco. The exposure to e-commerce is measured as the share of online purchases in the total amount of purchases in household goods at the state level. For that, I use two waves of the Brazilian Consumption Survey. The pre-treatment data is from 2008/2009, and the post-treatment is from the 2017/2018 survey. Based on a 2x2 DiD for the years 2009 and 2017, I estimate the following equations:

First Stage

$$EC_{i,c,t} = \delta_1(D_{i,c} \cdot \tau_{2017}) + \delta_2(X_i \cdot \tau_{2017}) + \mu_i + \tau_t + \varepsilon_{i,c,t} \tag{3}$$

Second Stage

$$y_{i,c,t} = \gamma \hat{EC}_{i,c,t} + \beta(X_i \cdot \tau_{2017}) + \mu_i + \tau_t + \epsilon_{i,c,t} \tag{4}$$

where $EC_{i,c,t}$ is the e-commerce share in state c and period t; μ_i and τ_t are microregion and year fixed effects; τ_{2017} is an indicator for the post-treatment year (i.e., equal to one if the year is 2017, and equal to zero if 2009); $D_{i,c}$ is equal to one for microregions i in the treated state, and zero otherwise; and X_i are predetermined covariates at the microregion level. The outcome $y_{i,c,t}$ is either log of employment or the number of firms (normalized by firms in 2010) in microregion i at year t.

The parameter δ_1 captures the effect of the opening of the DC in the exposure to e-commerce in microregions of the treated state. The parameter γ captures the elasticity of employment or the number of firms to the variation of the share of e-commerce between 2009 and 2017. The estimates are shown in Table 3. The first-stage estimate suggests that the opening of the DC is related to a 2.09 percentage points increase in e-commerce between 2009 and 2017. The F-statistic of 38.36 suggests that the opening of the DC is not a weak instrument for e-commerce exposure. The second stage estimates show that employment and the number of firms in the most exposed sectors decreased by 6.6% and 5.8%, respectively, for each 1 percentage point increase in e-commerce exposure. The fact that the OLS estimates are just around one third of the effects from the IV highlights the positive bias related to the exposure to e-commerce. Regions with stronger e-commerce growth are also those with larger economic dynamism, which attenuates the negative effect of e-commerce on local employment.

4.1.3 Inference

In the main results discussed in this paper, the standard errors are clustered at the microregion level. As the introduction of the distribution center and the price policy are set at the state level, the error term is supposed to be correlated within states. However, inference methods for few clusters, such as the Wild-Cluster Bootstrap, are not valid in settings with very few treated groups (specifically, a single treated state in this paper) (MacKinnon and Webb, 2017).

As an alternative, in Appendix B, I show that my results are robust to inference methods that are consistent with a very small number of clusters. Permuting the residuals of the treated state over the time series based on Chernozhukov, Wüthrich, and Zhu (2021), I find supporting evidence of a break from 2012 to 2016, suggesting that the effects are in fact statistically significant. I also show that the baseline results are robust to a more conservative approach that leverages the differences of the outcomes at the cross-sectional level (Hagemann, 2020).

I also highlight that the baseline standard errors, which are clustered at the microregion level, can be interpreted under the assumption that the state-specific error term is not random. Thus, instead of an inference issue, such shock should be seen as deviations from the parallel trend assumption (Roth et al., 2022). I will provide formal sensitivity analysis to assess whether my findings are robust to potential state-level shocks in a future version of the paper.

4.1.4 Effects of e-commerce in other sectors

So far this paper has been focusing on local retailers, which are more directly competing with e-commerce. However, some sectors are expected to be complementary to online sales, such as shipping and warehouse. In Figure 11, I show that employment increased by 12% in 2012, 15% in 2015 and more than 18% by 2016 in these sectors. The number of firms increased by 9% in 2013 (Figure 12).

Additional evidence show that retail categories more directly competing with e-commerce had the largest reduction in employment, while sectors such as shipping, hospitality, beauty services, warehouse, real estate and informatics services had grown. I estimate heterogeneous effects on each 43 sectors listed in Appendix Table A.1 using a canonical 2x2 DiD for each post-treatment year, where the baseline year is 2010. For that, I interact the treatment indicator D_i and the post-treatment dummy τ_t with an indicator variable for each sector (η_s), as follows:

$$y_{i,s,t} = \mu_i + \tau_t + \eta_s + \sum_s \gamma_{s,t}(D_i \cdot \eta_s \cdot \tau_t) + \beta_t X_i \cdot \tau_t + u_{i,s,t} \tag{5}$$

where η_s is sector s fixed effects; μ_i is the microregion fixed effects; τ_t is a dummy variable for each post-treatment year; and X_i are the predetermined covariates at the microregion level. The dependent variable is the number of workers in year t normalized by the baseline number of workers. Therefore, the parameter $\gamma_{s,t}$ captures the percent change in employment in sector s from 2010 to the post-treatment year t. The regression is weighted by the baseline number of workers, and the standard errors are two-way clustered at microregion and sector.

The distribution of the estimators and the t-statistics for 2015 are represented in Figures 13 and 14 respectively. The figures for 2013, 2014 and 2016 are available in Appendix C. The sector with the largest reduction in employment is *Electronics*, ranging from 5% in 2012, 25% in 2014 to almost 43% decrease in 2015. The negative impact on books, furniture, fabrics and home appliances stores also increased over time, reaching, in 2015, a reduction in employment by 16%, 10.8%, 10.7%, and 8.2% respectively. In contrast, sectors such as shipping/Correios, Warehouse, Real Estate, Informatics and basic services (e.g., hospitality and beauty) had a rise in their level of employment after 2010. By 2015, employment in food preparation and restaurants increased by more than 10%, and by almost 28% in bakery, while beauty and sporting services increased by 38%. The variation of employment in Correios and shipping companies was more than 20%, and warehouses employed almost 100% more in 2015. Finally, employment in informatics consistently increased in the period, going from 8% in 2011, 25% in 2013, to 56% in 2015 in the treated microregions.

One might be concerned about the net effects of e-commerce on local labor markets. Figure 15 provide supporting evidence of a decreasing trend in local employment after 2013, reaching 10% reduction in 2015. In contrast, the number of establishments increased by 5% until 2014, following a decreasing trend afterwards.

I highlight that these findings should not be interpreted as causal effects of e-commerce, since the parallel trends assumption might not hold for all the sectors. However, the main take away is that there might be winners and losers from the e-commerce growth, depending on the degree of substitution and complementary between the local firms and online retail. This calls attention to the importance of measuring distributive effects of e-commerce on local workers. For instance, considering that retail employment requires more skills compared to basic services (beauty, hospitality, sports, shipping, warehouse), there might be a displacement of workers from medium skills to low skills jobs. As a consequence, we might also expect effects on wages. Studying the effects of e-commerce on wages and workers' displacement (e.g., mobility across sectors) is a research in progress.

4.2 Effects of e-commerce in the hosting capital

The previous section has discussed the findings for local labor markets that were strongly exposed to e-commerce after the entrance of a major distribution center in Pernambuco. However, the exposure to e-commerce is expected to be stronger at the capital hosting the DC, due to the access to cheaper and faster shipping. Using a Synthetic Difference-in-Differences (SDID) approach, I estimate the effects of e-commerce on employment and on the number of firms at the capital level. For that, I include twelve capitals from the Northeast and the Center-West in the pool of potential controls.

The map in Figure 16 presents the weights given for each state capital for log of employment. The largest weights are for capitals in the Northeast, especially those nearby the treated capital. For instance, Maceió, in the state of Alagoas, receives a weight of 0.14, while Salvador, in Bahia, receives 0.09. In the Center-West, the largest weight is for Cuiabá, in Mato Grosso, of almost 0.11. The smallest weight is in the order of 0.007 in Campo Grande, Mato Grosso do Sul.

There is supporting evidence that the trajectory of log of employment in household good stores was parallel before the treatment between the synthetic control and the treated capitals (Figure 17). After the treatment, employment reduced more in the treated capital, with an estimated average treatment effect of 12% reduction from 2011 to 2016 (p-value of 0.007). The difference-in-differences point estimate in 2015 is of 21% reduction in employment in the treated capital. Figure 18 shows a similar pattern on the number of firms, with an estimated average treatment effect of 13% reduction from 2011 to 2016 in the treated capital.

When looking at employment and the number of firms in overall retail (Figures 19 and 20), there is also evidence of negative effects after the treatment. For instance, the point-estimate of the DiD in 2015 suggest a decrease in employment by almost 10%. Although the average treatment effect in the entire post-treatment period is negative by 5%, it is less precise (p-value of 0.184). For firms, the average treatment effect suggests a reduction by 6%, but it is not significant at standard levels (p-value of 0.111).

There is also suggestive evidence that e-commerce negatively impacted employment and the number of firms in household good stores in the rest of the state, excluding the hosting capital (Appendix Figure C.7). However, the average treatment effects are smaller compared to the effects at the capital: more than 7% reduction in employment and 8.6% in the number of firms by 2016. The average treatment effects on the treated state using a SDID excluding all the capitals are of similar magnitude, but much less precise (Appendix Figures C.9 and C.10). Overall, the findings in this section point to the conclusion that e-commerce had larger negative effects on retailers in the capital hosting the distribution center. Nevertheless, it is important to highlight that the effects on peripheral labor markets are non-negligible, and might be stronger in later stages of e-commerce in Brazil, specially with the possibility of stores pick-ups.

5 Conclusion

Although e-commerce has been sharply increasing in the last decades, little is known about its effects on local labor markets. The growing literature focus on developed countries, finding effects of around 3% reduction on retail employment (Chava et al., 2022; Chun et al., 2020). This paper is the first to provide evidence about the effects of e-commerce on local labor markets in a developing country. Using census-data of formal establishments in Brazil from 1997 to 2016, I found that e-commerce reduced employment and the number of firms in the most exposed retail sectors by 20% and 12% until 2016 in the treated state in the Northeast.

The larger effects found in this paper compared to previous evidence from developed countries might be explained by the fact that local economies usually rely more on retail in developing countries. In addition, this paper estimates effects on the full population of retail firms, not restricting to larger companies as in Chava et al. (2022). I also took into account relative exposure of different retail categories, and estimated the effects on the sectors that were more directly competing with e-commerce.

I showed that the results are robust to different levels of analysis, specifications, and to inference methods that take into account cluster-level correlations within states. Moreover, additional evidence show that retail categories more directly competing with e-commerce had the largest reduction in employment, while sectors such as shipping, hospitality, beauty services, warehouse, real estate and informatics services had grown.

I highlight that the findings in this paper can be seen as a lower bound of overall e-commerce effects on local labor markets, since the control group might also have been affected by e-commerce to some extent. Effects on wages and workers (such as mobility across sectors) is a research in progress.

References

Agrawal, David R (2021). “The Internet as a tax haven?” *American Economic Journal: Economic Policy.*

Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager (2021). “Synthetic difference-in-differences”. *American Economic Review* 111(12), pp. 4088–4118.

B2W (2021). *The company and investor information.*

Bauer, Anahid (2021). “Effects of e-commerce on local labor markets”. *Job Market Paper.*

Blanchard, Olivier J and Lawrence F Katz (1992). “Regional evolutions”. *Brookings Papers Econ. Activity* 23(1).

Borusyak, Kirill and Peter Hull (2020). “Non-random exposure to exogenous shocks: Theory and applications”. *National Bureau of Economic Research (NBER).*

Chava, Sudheer, Alexander Oettl, Manpreet Singh, and Linghang Zeng (2022). “Creative Destruction? Impact of E-Commerce on the Retail Sector”. *Management Science, accepted.*

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu (2021). “An exact and robust conformal inference method for counterfactual and synthetic controls”. *Journal of the American Statistical Association* 116(536), pp. 1849–1864.

Chun, Hyunbae, Hailey Hayeon Joo, Jisoo Kang, Yoonsoo Lee, et al. (2020). *Diffusion of E-Commerce and Retail Job Apocalypse: Evidence from Credit Card Data on Online Spending.* Tech. rep. Center for Economic Institutions, Institute of Economic Research ...

Correios (2020). *Relatorio Intergrado Correios.*

Correios (2021). *Saiba mais - Nacional.*

Couture, Victor, Benjamin Faber, Yizhen Gu, and Lizhi Liu (2021). “Connecting the countryside via E-Commerce: Evidence from China”. *American Economic Review: Insights* 3(1), pp. 35–50.

Dix-Carneiro, Rafael and Brian K Kovak (2019). “Margins of labor market adjustment to trade”. *Journal of International Economics* 117, pp. 125–142.

Dustmann, Christian, Uta Schönberg, and Jan Stuhler (2017). “Labor supply shocks, native wages, and the adjustment of local employment”. *The Quarterly Journal of Economics* 132(1), pp. 435–483.

E-bit (2015). *Webshopper 32rd Edition.*

Forman, Chris, Avi Goldfarb, and Shane Greenstein (2012). “The Internet and local wages: A puzzle”. *American Economic Review* 102(1), pp. 556–75.

Goldmanis, Maris, Ali Hortaçsu, Chad Syverson, and Önsel Emre (2010). “E-commerce and the Market Structure of Retail Industries”. *The Economic Journal* 120(545), pp. 651–682.

Hagemann, Andreas (2020). “Inference with a single treated cluster”. *arXiv preprint arXiv:2010.04076.*

Hortaçsu, Ali and Chad Syverson (2015). “The ongoing evolution of US retail: A format tug-of-war”. *Journal of Economic Perspectives* 29(4), pp. 89–112.

IBGE (2018). *Municipal GDP (2002-2018).*

IBGE (n.d.). *Pesquisa Anual de Comércio (PAC).*

ITA (2021). *Brazil - Country Commercial Guide: eCommerce.*

J.P. Morgan (2019). *E-commerce Payments Trends: Brazil e-commerce insights.*

Kim, Ryan and Jonathan Vogel (2021). "Trade shocks and labor market adjustment". *American Economic Review: Insights* 3(1), pp. 115–30.

Kovak, Brian K (2013). "Regional effects of trade reform: What is the correct measure of liberalization?" *American Economic Review* 103(5), pp. 1960–76.

MacKinnon, James G and Matthew D Webb (2017). "Wild bootstrap inference for wildly different cluster sizes". *Journal of Applied Econometrics* 32(2), pp. 233–254.

Mercado Livre (2021). *Tudo o que você precisa saber sobre o Mercado Livre.*

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe (2022). "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. Section 5: Relaxing sampling assumptions". *arXiv preprint arXiv:2201.01194.*

UNCTAD (2021). *COVID-19 and E-commerce: A Global Review.* Tech. rep. United Nations Conference on Trade and Development.

Yagan, Danny (2019). "Employment hysteresis from the great recession". *Journal of Political Economy* 127(5), pp. 2505–2558.

Figures

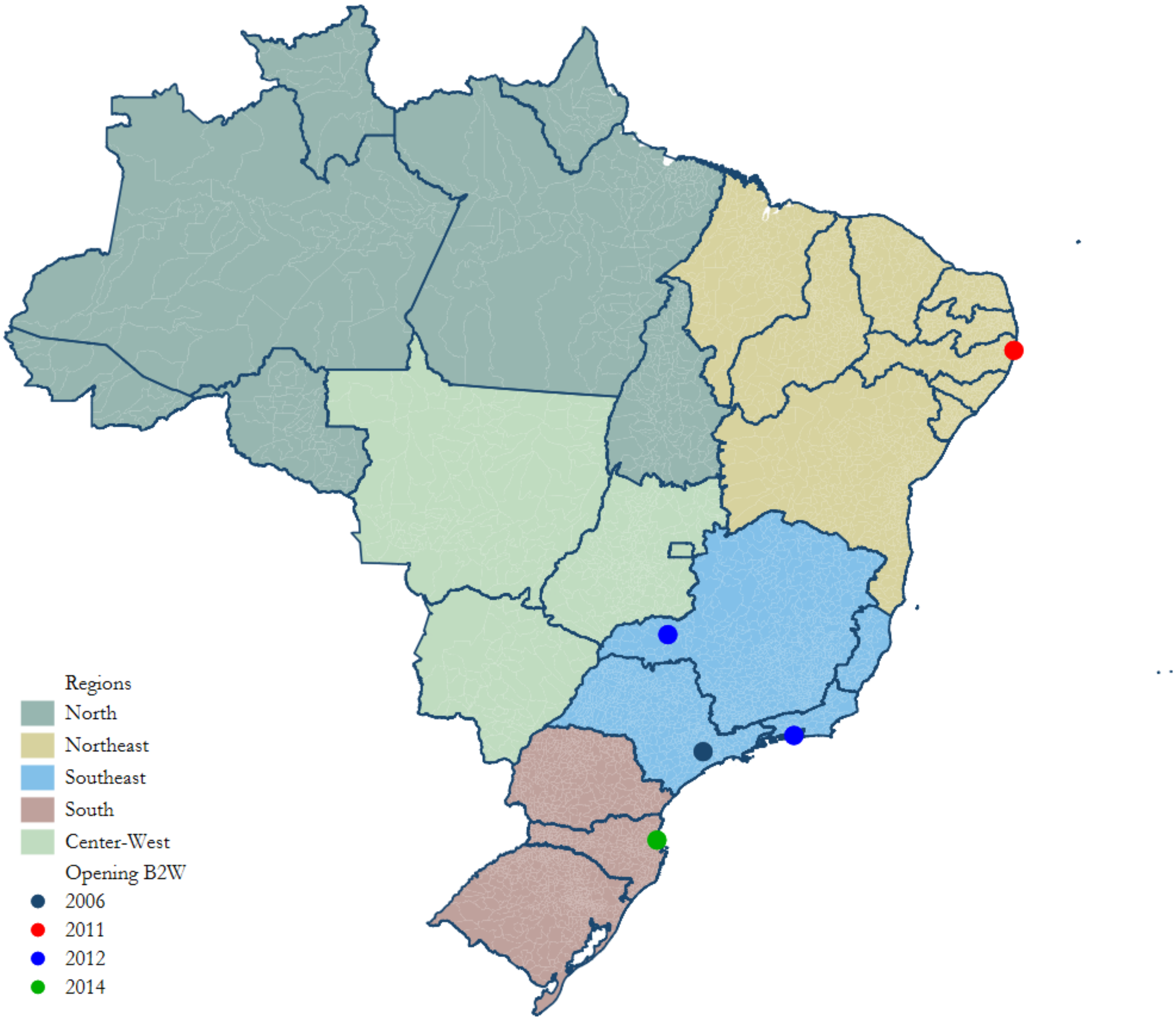

Figure 1: Location of B2W Distribution Centers in Brazil

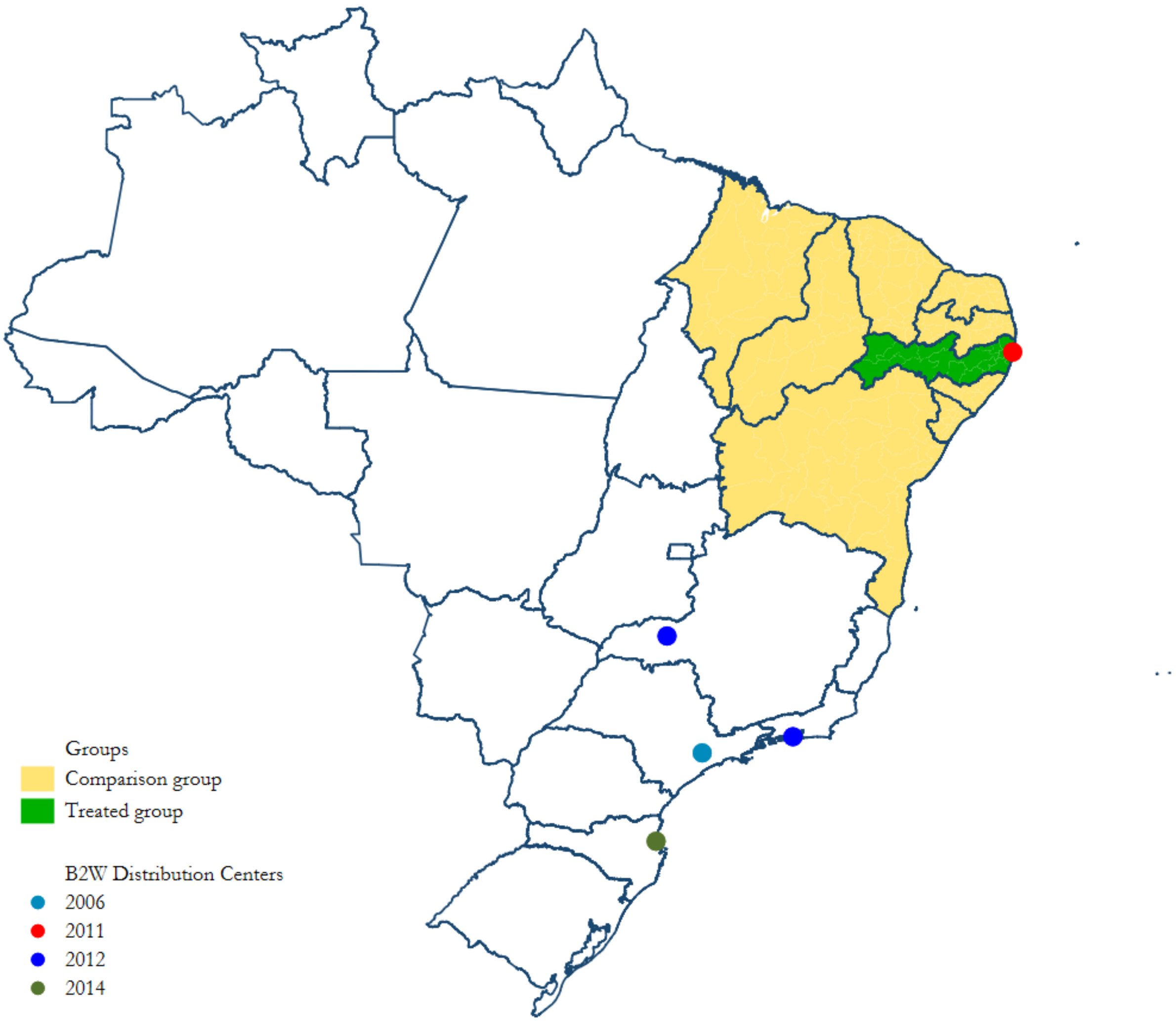

Figure 2: Treated and untreated microregions in the Northeast region of Brazil

Effects on employment in local labor markets

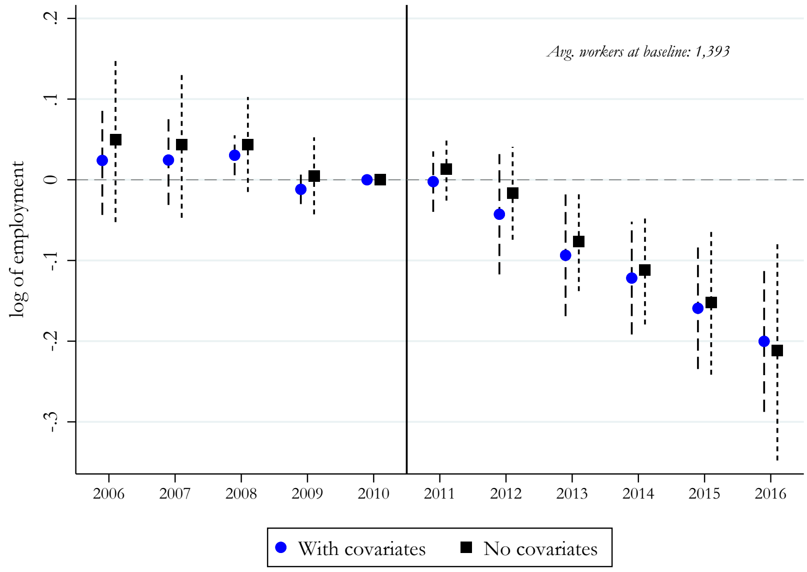

Figure 3: Effects of E-commerce on Employment in Household Goods Stores

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

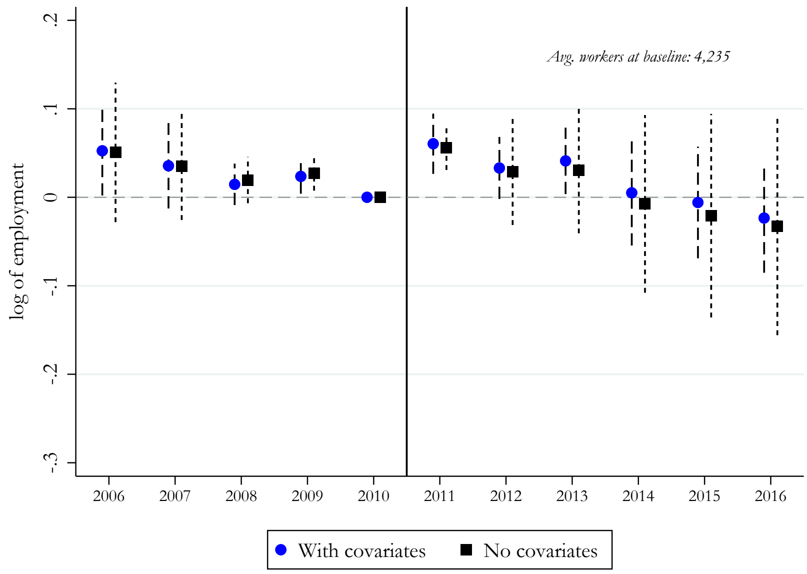

Figure 4: Effects of E-commerce on Employment in Other Retailers

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

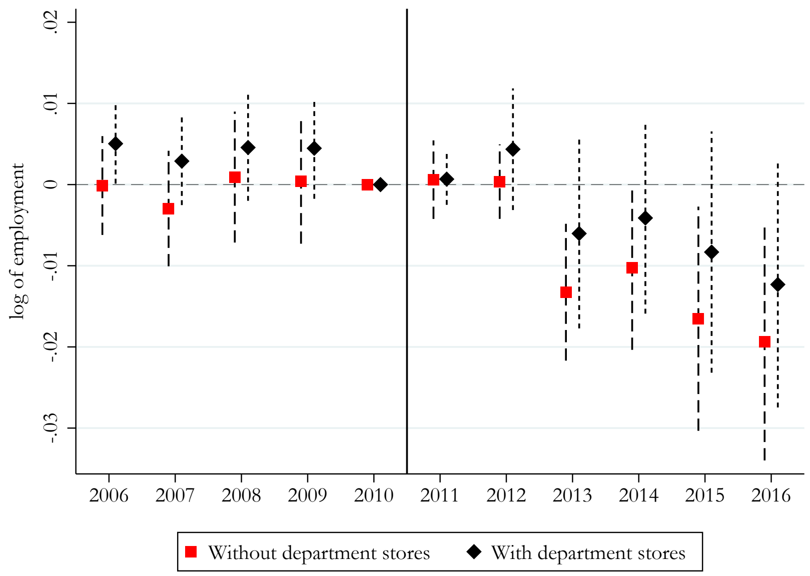

Figure 5: Effects of E-commerce on Employment in Retail by Exposure to E-commerce

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

Effects on the number of firms in local labor markets

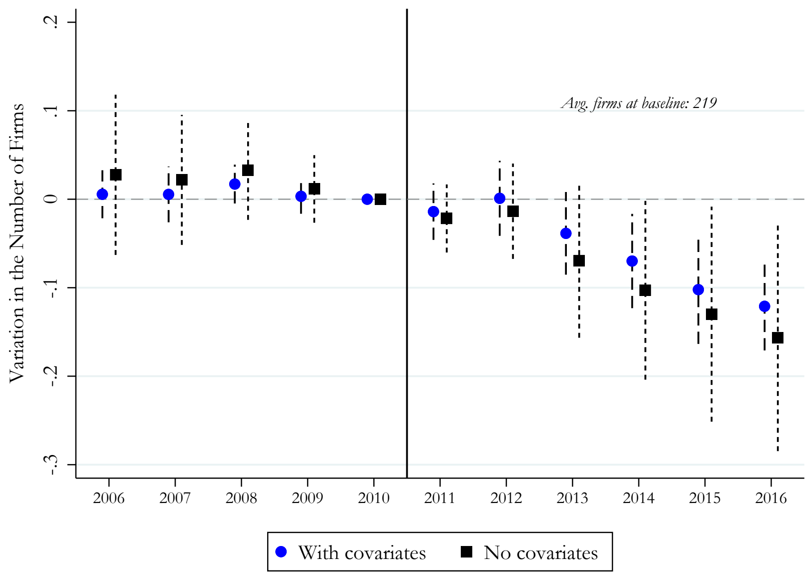

Figure 6: Effects of E-commerce on the Number of Firms in Household Goods Stores

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

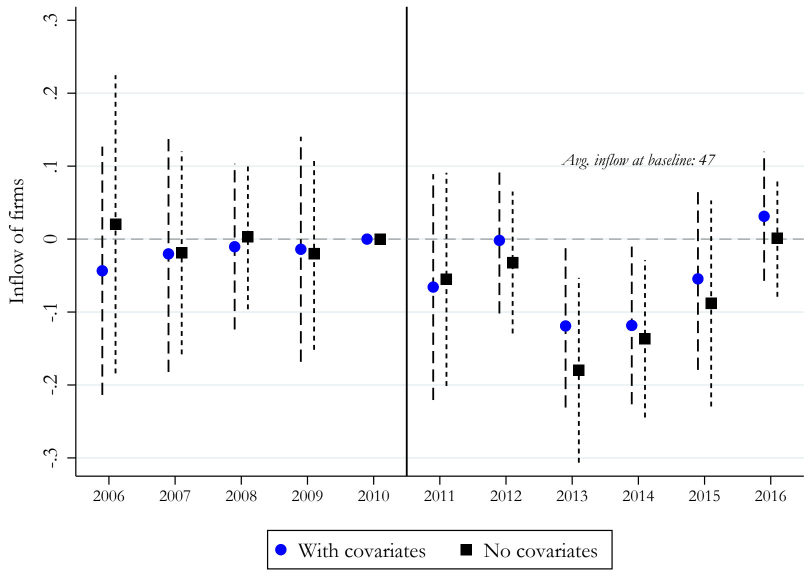

Figure 7: Effects of E-commerce on the Inflow of Firms in Household Goods Stores

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

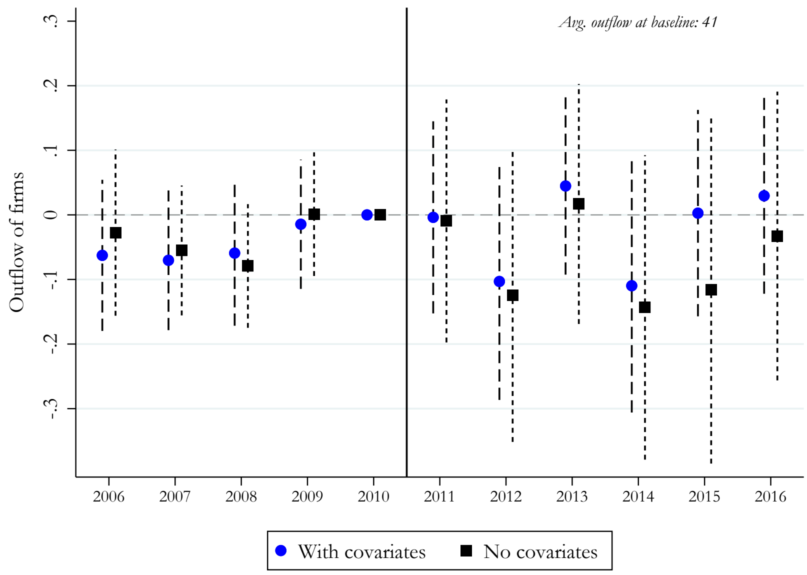

Figure 8: Effects of E-commerce on the Outflow of Firms in Household Goods Stores

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

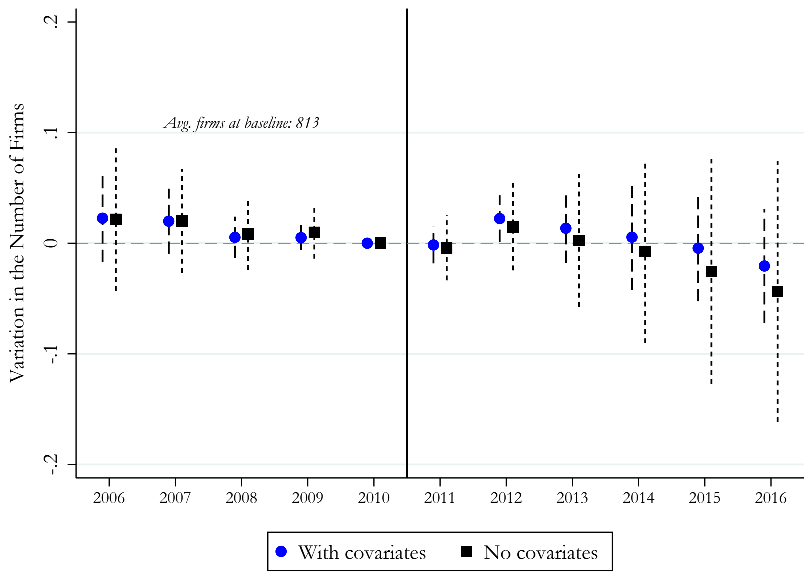

Figure 9: Effects of E-commerce on the Number of Firms in Other Retailers

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

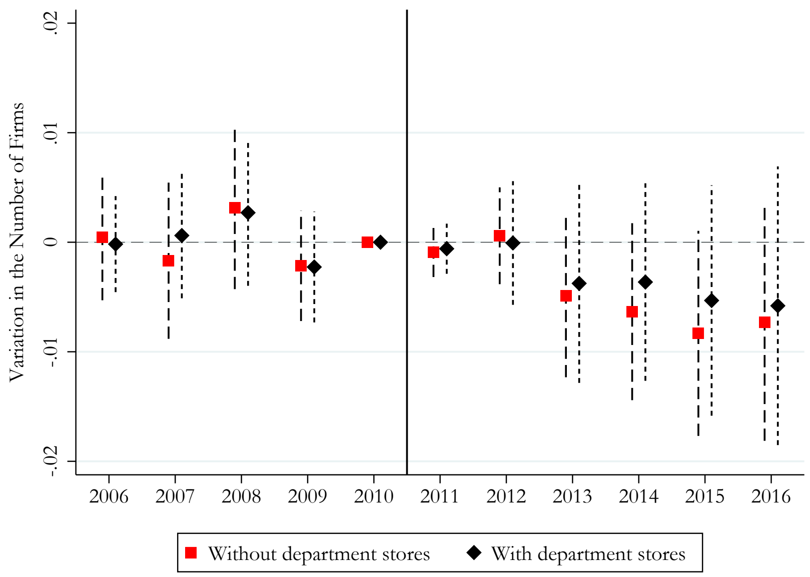

Figure 10: Effects of E-commerce on the Number of Firms by Exposure to E-commerce

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

Effects on other sectors in local labor markets

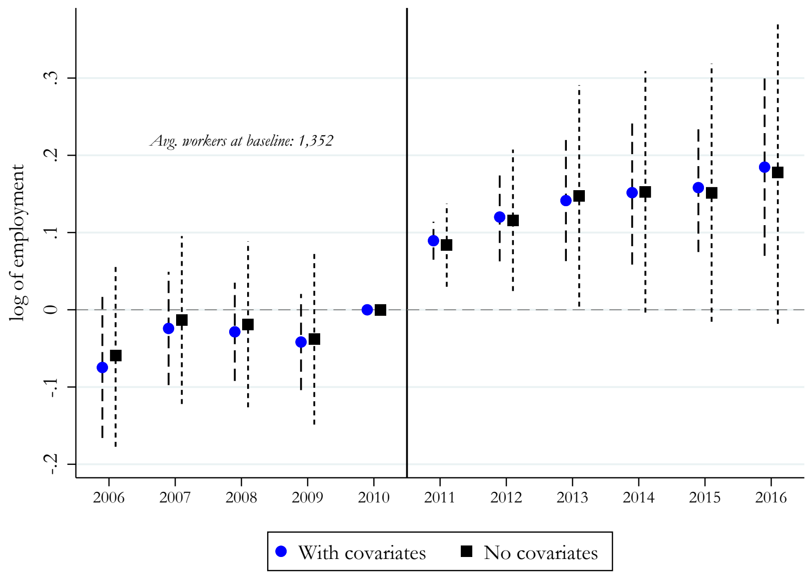

Figure 11: Effects of E-commerce on Employment in Shipping and Warehouse

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

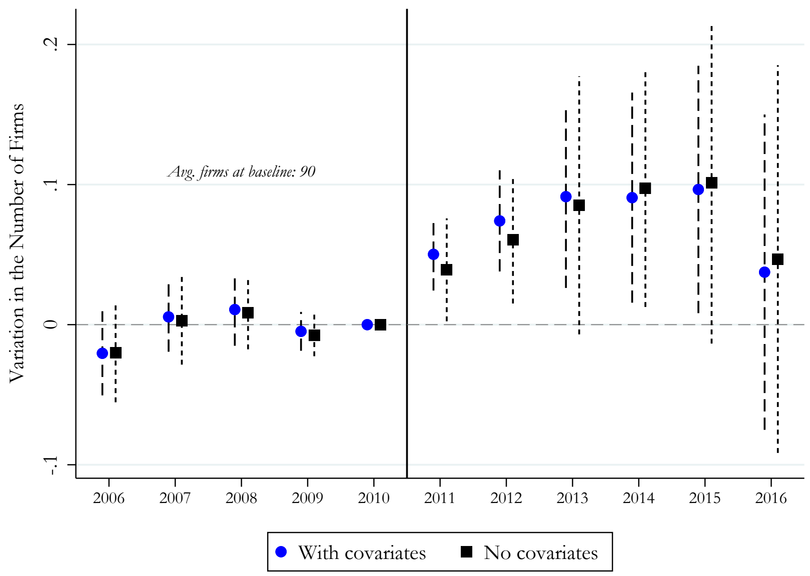

Figure 12: Effects of E-commerce on the Number of Firms in Shipping and Warehouse

Notes: Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

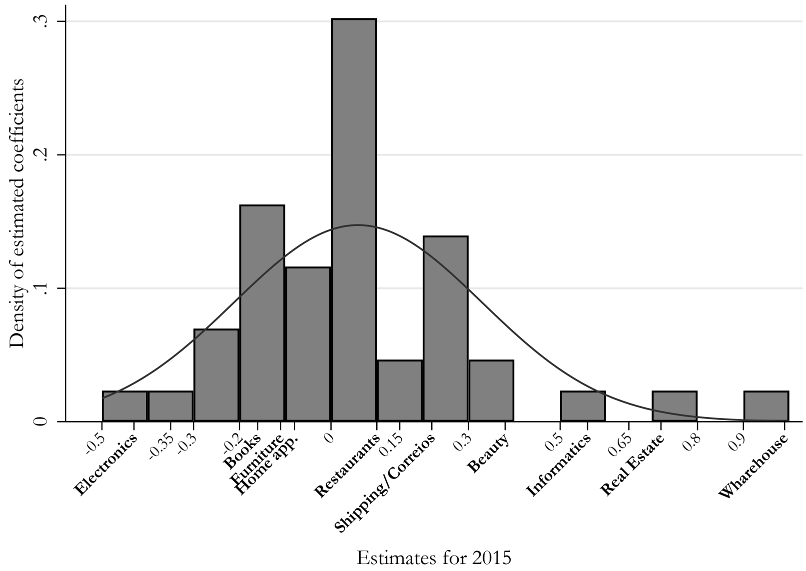

Figure 13: Effects of E-commerce on Employment across Sectors in 2015

The regression is performed for two periods, where 2010 is the pre-treatment and 2015 is the post-treatment year. I interacted an indicator for each sector with the treatment variable. There are 43 sectors. The regression includes year, sector and microregion fixed effects. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

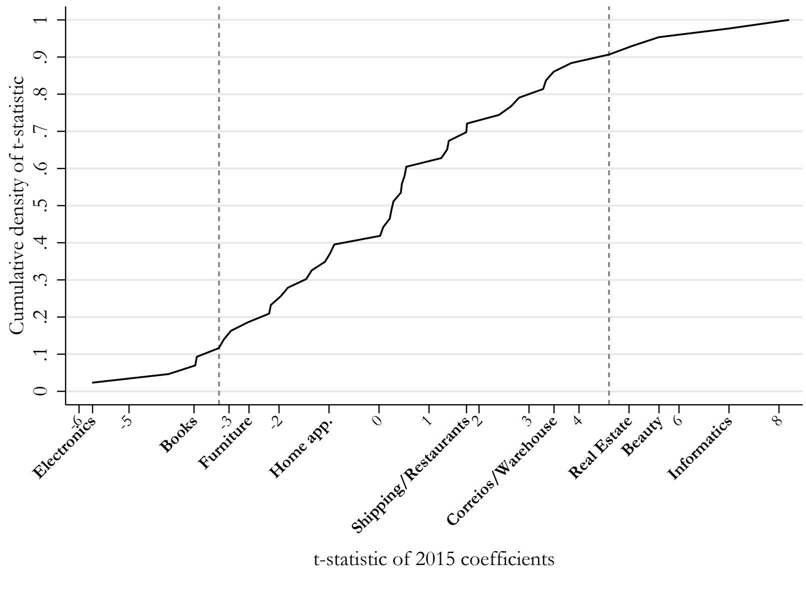

Figure 14: Distribution of Statistic-t for the Effects of E-commerce on Employment across Sectors in 2015

The regression is performed for two periods, where 2010 is the pre-treatment and 2015 is the post-treatment year. I interacted an indicator for each sector with the treatment variable. There are 43 sectors. The regression includes year, sector and microregion fixed effects. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

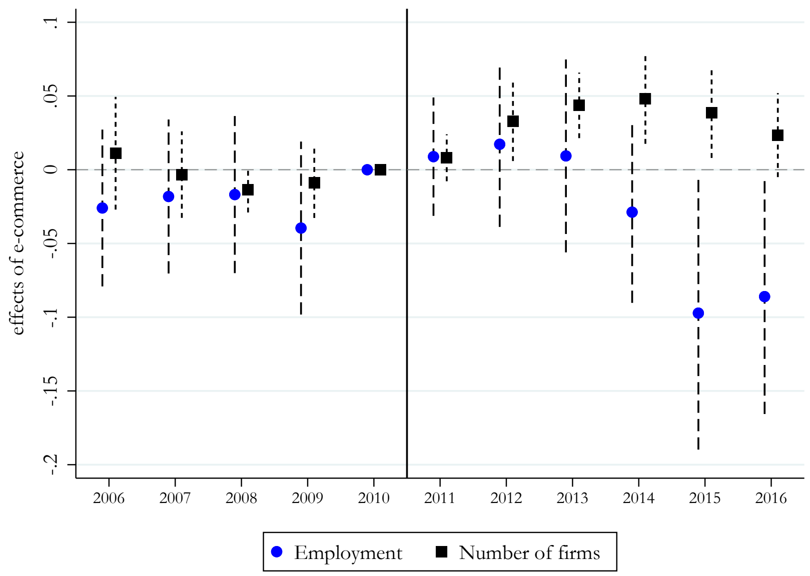

Figure 15: Effects of E-commerce on Employment and the Number of Firms in Local Labor Markets

Notes: Regressions include year, microregion fixed effects and sector fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

Effects of E-commerce on Employment and Number of Firms in the Capital

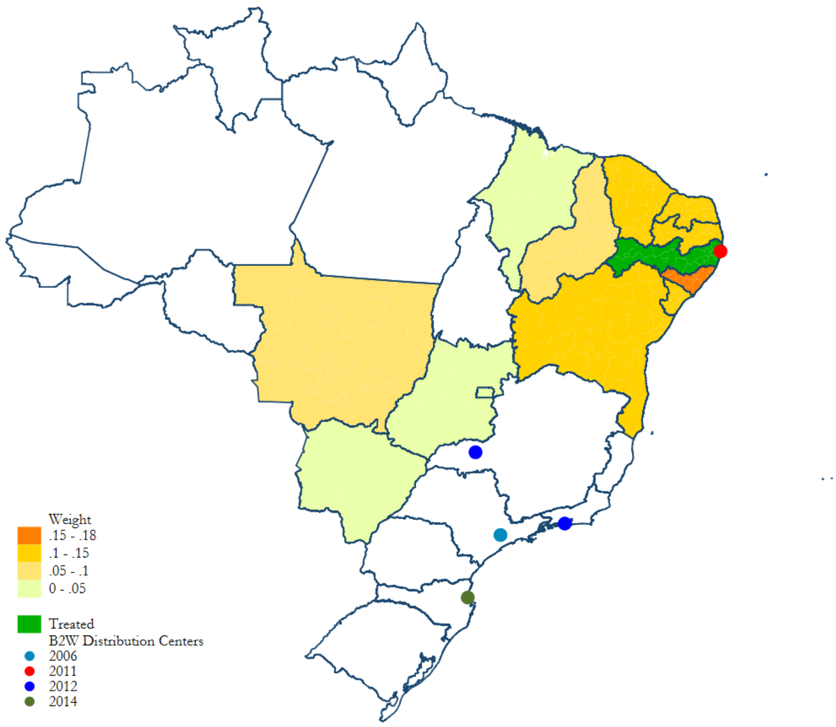

Figure 16: Synthetic Difference-in-Differences: Weights and Control Pool at Capital Level

The donor pool is composed by twelve state capitals in the Northeast and Center-West. 2010 Urban population and GDP per capita considered for the synthetic control's selection, in addition to pre-determined outcome.

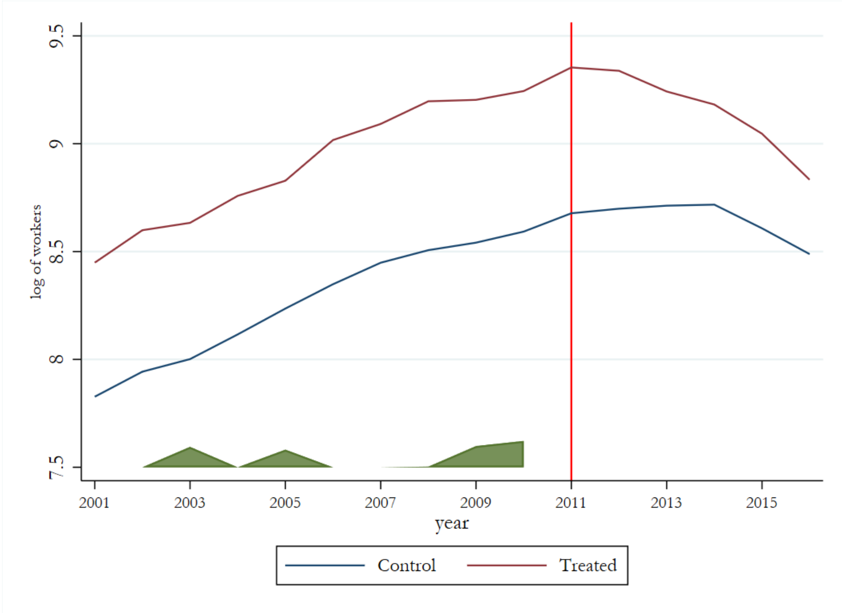

Figure 17: Effects of E-commerce on Employment in Household Goods Stores in the Hosting Capital

The outcome is the log of workers. The synthetic control is composed by twelve capitals, whose weights minimize the trends of pre-determined outcomes up to a constant difference in levels. $\hat{ATT}$ = **-0.12 (0.05), p-value = 0.007.** Inference based on placebo allocation of treatment in the donor pool.

Figure 18: Effects of E-commerce on the Number of Firms in Household Goods Stores in the Hosting Capital

The outcome is the number of firms in year t normalized by the number of firms in the baseline. The synthetic control is composed by twelve capitals, whose weights minimize the trends of pre-determined outcomes up to a constant difference in levels. $\hat{ATT}$ = **-0.13 (0.03), p-value = 0.0001.** Inference based on placebo allocation of treatment in the donor pool.

Figure 19: Effects of E-commerce on Employment in Local Retailers in the Hosting Capital

The outcome is the log of workers. The synthetic control is composed by twelve capitals, whose weights minimize the trends of pre-determined outcomes up to a constant difference in levels. $\hat{ATT}$ = **-0.05 (0.04), p-value = 0.184.** Inference based on placebo allocation of treatment in the donor pool.

Figure 20: Effects of E-commerce on the Number of Firms in Local Retailers in the Hosting Capital

The outcome is the number of firms in year t normalized by the number of firms in the baseline. The synthetic control is composed by twelve capitals, whose weights minimize the trends of pre-determined outcomes up to a constant difference in levels. $\hat{ATT}$ = **-0.06 (0.04), p-value = 0.111.** Inference based on placebo allocation of treatment in the donor pool.

Tables

Table 1: Share of e-commerce in total sales in retail

Sector	2012	2015	2017
Electronics	2.7%	10.5%	14.9%
Home appliances	4.2%	5.9%	6.7%
Furniture	1.4%[a]	2.2%	3.5%
Sports/Books	4.1%	6.0%	6.4%
Pharmacy and cosmetics	0.5%	1.3%	1.4%
Clothing	0.2%	2.0%	4.0%
Shoes	3.9%	·	7.9%
Fabrics	0.03%	0.8%	1.7%
Other goods[b]	4.2%	9.3%	9.5%

Notes: E-commerce share measured as revenues from online sales per total revenues based on the Annual Retail Survey (IBGE, n.d.).
(a) 2013 data; (b) jewelry, department stores, etc.

Table 2: Opening of distribution center and variation in e-commerce exposure

	All state		Capitals	
Δ e-commerce	(1)	(2)	(3)	(4)
DC	3.70	3.01	7.02	6.82
	(1.26)	(1.37)	(1.96)	(2.19)
Constant	3.52	3.52	4.65	4.65
N	27	26	26	25

Notes: The dependent variable is the change in the share of e-commerce from 2008/2009 to 2017/2018 at the state or the capital level. The independent variable is an indicator equal to 1 for the states/capitals with a Distribution Center from B2W in this period (Sao Paulo, Rio de Janeiro, Minas Gerais, Pernambuco, and Santa Catarina), and equal to 0 otherwise. Columns (1) and (3) include Sao Paulo; (2) and (4) exclude Sao Paulo.

Table 3: Effects of E-commerce on Employment and Number of Firms in Household Goods Stores

	Workers	**Firms**
	(1)	(2)
OLS	-0.022	-0.015
	(0.014)	(0.009)
2SLS	-0.066	-0.058
	(0.019)	(0.016)
	E-commerce share	
First Stage	2.09	
	(0.337)	
F-stat	38.36	
N	187	

Notes: Regressions include year and microregion fixed effects, and control for urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

A Data Appendix

Table A.1: Sectors

Sector/Industry	Code CNAE
1. Agriculture	All starting by 0
Manufacture	
2. Extractive, food, textile	All starting by 1
3. Processing	All starting by 2 and 3
Services	
4. Construction	All starting by 4
5. Transportation	Groups 60, 61, 62, 63, 64
6. Road cargo shipping	60267, 63118, 63215, 63401, 64122, 64114
7. Warehouses	63126
8. Travel agencies	63304
9. Telecommunication	64203
10. Finance and Insurance	Groups 65, 66, 67
11. Real estate	Group 70
12. Other rental services	Group 71
13. Computation and research	Groups 72, 73
14. Public services and others	Groups 74, 75
15. Juridic and accounting	74110, 74128, 74144, 74160
16. Publicity, photos	74136, 74209, 74306, 74403, 74918
17. Education	Group 80
18. Health	Group 85
19. Other services (e.g., repair)	All starting by 9 and 527
20. Sports and physical activities	92614, 92622, 93041
21. Beauty (hair saloon)	93025
22. Wholesale	Group 51
23. Hospitality	55115, 55123, 55131, 55190
24. Restaurants	55212, 55220, 55239, 55247, 55298
Retail (sales)	
25. Vehicles	50105, 50415
26. Car repair	50202, 50423, 50300
27. Fuel	50504, 52477
28. Supermarkets	Group 521
29. Bakery	52213
30. Beverages	52248
31. Food stores	52221, 52230, 52299
32. Used goods, fairs, stands	52442, 525, 526
33. Department stores	52159
34. Fabrics	52310
35. Clothing	52329
36. Shoes	52337
37. Pharmacy and cosmetics	52418
38. Books	52469
39. Other goods	52493
Household goods	
40. Furniture	52434
41. Home appliances	52426
42. Electronics	52450

Notes: The classification is based on the National Classification of Economic Activities from 1995 (CNAE 95), which was available for every year from 1995 to 2016.

B Inference Appendix

In this Appendix, I provide evidence that my results are robust to inference methods that are consistent with a very small number of clusters. First, I apply the rearrangement test proposed by Hagemann (2020) to compare the evolution of average employment of the treated state to each of the eight control state. In the sequence, permuting the residuals of the treated state over the time series based on Chernozhukov, Wüthrich, and Zhu (2021), I find supporting evidence of a break from 2012 to 2016, suggesting that the effects are in fact statistically significant.

Table B.1 shows whether we reject (R) or don't reject (D-R) the null hypothesis of no treatment effect for each post-treatment year based on the rearrangement test proposed by Hagemann (2020). As a brief overview, I compare the average effects on the treated state in each post-treatment year with an estimated effect for each control state. The test rejects the null hypothesis when the effect on the treated group is larger than the effects in all the control states even after deflating it by $(1-\omega)$. The weight ω is numerically chosen such as the size of the test is the level of significance α, given the number of control clusters (eight in my case) and the heterogeneity measure ρ. The level of heterogeneity restricts how much more variable is the estimate of the effect in the treated state compare to almost all the estimates in the control states (Hagemann, 2020).

I highlight that, with eight control clusters, I can apply the rearrangement test only for $\alpha \geq 0.10$. In column (1) of Table B.1, I show that the null hypothesis is rejected at 10% significance level from 2014 to 2016. The maximum level of heterogeneity that solves for some ω is 0.9, which means that the estimated effect in the treated state can be less precise than the most precise control group, but should have smaller variance than all remaining controls. $\rho = 0.9$ is not imposing homogeneity across clusters, but only restricting their relative variance. In column (3), I show that the null hypotheses are also rejected at 20% level with $\rho = 1.5$, which considers that the effect on the treated is less precise compare to all control states. Given that the effects are estimated using administrative data with the population of formal workers in each microregion, in column (4) I allow for $\rho = 1.1$ at 10% of significance level by removing the uncertainty from the estimated average effect of the control clusters.

Overall, the conclusions from the rearrangement test are in the same direction of my main results. The main restriction of the test is that parallel trends should hold for all the untreated states, instead of only on average. Figure B.1 shows the evolution of the log differences of employment in each year with respect to the baseline (2010) in all the states in the Northeast. When comparing the treated state (blue line) with the average of the controls (black line), there is supporting evidence for the parallel trends assumption, and a clear break in the trend of the treated group after 2011. However, the figure also shows that a unique state (red line) is binding in the rearrangement test, while it is clear that it was following a specific trend over all the period.

Alternatively, I leverage the temporal dimension to perform inference using permutation of residuals as suggested by Chernozhukov, Wüthrich, and Zhu (2021). First, I estimated the average counterfactual employment for the treated group using data from all the states in the Northeast from 1997 to 2016. The residuals are the differences of the observed log employment and the counterfactual of the treated state, and are shown in Figure B.2. While the residuals at the pre-treatment years are more closely distributed around zero, there is a clear break in the time series after the treatment. The test statistic (S-statistic) is the absolute value of the sum of residuals for the post-treatment period (i.e., across 5 years, from 2012 to 2016). The distribution of the S-statistic, shown in Figure B.3, was obtained using a block-permutation of the residuals. The test statistic of interest is the largest one observed in the data, thus the p-value for the null of no average treatment effects is 0.05. When performing inference for each post-treatment year individually, the estimated p-value is 0.065 from 2013 to 2016. I highlight that 0.065 is the smallest possible p-value, giving the number of periods available.

Therefore, both the rearrangement test and the permutation of residuals across the time series provide supporting evidence to the main findings of the paper. I also highlight that the baseline standard errors, which are clustered at the microregion level, can be interpreted under the assumption that the state-specific error term is not random. Thus, instead of an inference issue, such shock should be seen as deviations from the parallel trend assumption (Roth et al., 2022). I will provide formal sensitivity analysis to assess whether my findings are robust to potential state-level shocks in a future version of the paper.

Table B.1: Inference based on the rearrangement test

	Log difference		Decision			
Year	$\Delta Y_1 - \Delta\overline{Y}_0$	$\Delta Y_{0k} - \Delta\overline{Y}_0$	(1)	(2)	(3)	(4)
2012	-0.033	-0.144	D-R	D-R	D-R	D-R
2013	-0.088	-0.123	D-R	D-R	D-R	D-R
2014	-0.128	-0.093	R	D-R	R	R
2015	-0.162	-0.126	R	D-R	R	R
2016	-0.228	-0.124	R	R	R	R
Level of significance (α)			0.10	0.15	0.20	0.10
Heterogeneity (ρ)			0.9	1.3	1.5	1.1
Weight (ω)			0.2208	0.2691	0.088	0.226

Notes: The test decision is R when we reject the null hypothesis at the level of significance α, and $D-R$ otherwise. The level of heterogeneity ρ restricts how much more variable is the estimate of the effect in the treated state compare to almost all the estimates in the control states. The weight ω is numerically chosen such as the size of the test is α, given the number of control clusters and the heterogeneity ρ. I highlight that, with eight control clusters, I can't apply the rearrangement test for $\alpha < 0.10$ (i.e., there are no weights that solve the problem). In column (4) I can increase the level of heterogeneity under 10% of significance level by removing the uncertainty from the estimated average effect of the control clusters.

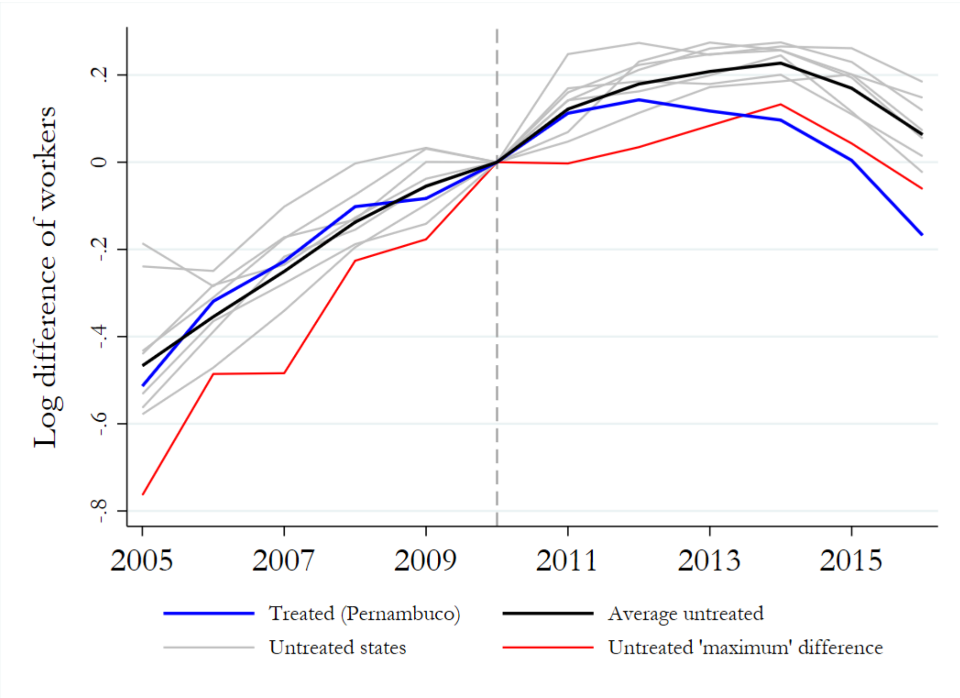

Figure B.1: Evolution of average employment in household goods across states in the Northeast

Notes: Log differences of average employment in each year t to the baseline (2010)

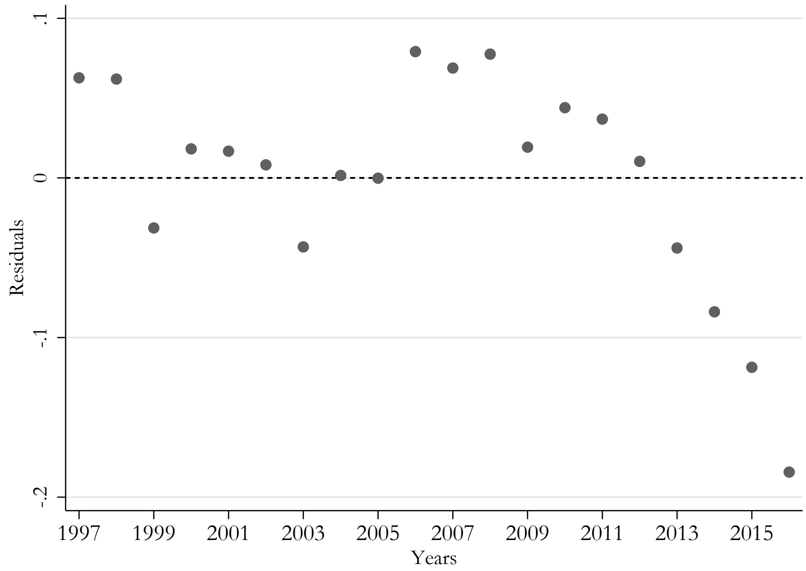

Figure B.2: Distribution of the residuals of the treated state over time

Notes: The residuals are obtained as the difference of average log employment in household goods stores in each year in the treated stated and the estimated counterfactual, based on the approach from Chernozhukov, Wüthrich, and Zhu (2021).

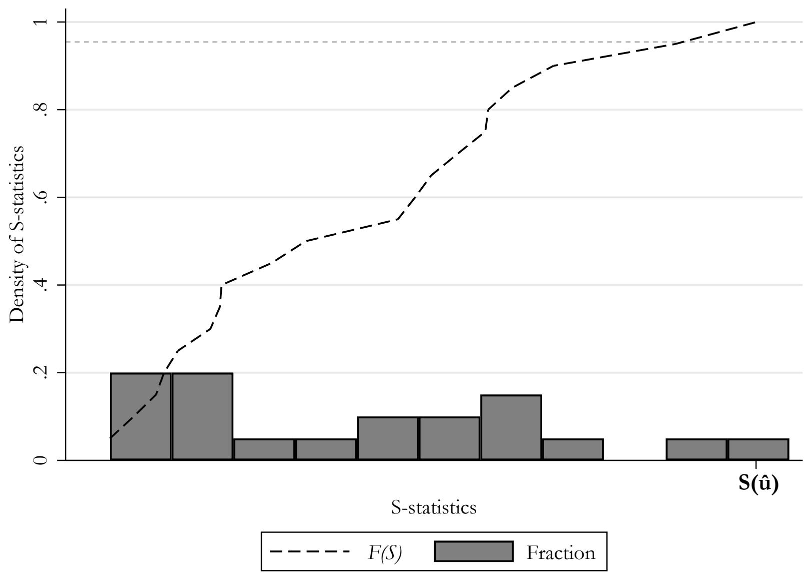

Figure B.3: Distribution of the S-statistics

Notes: The distribution of the S-statistics is obtained from the block-permutation of the residuals from 1997 to 2016. The statistics of interest is the absolute value of the sum of the residuals during the post-treatment period (Chernozhukov, Wüthrich, and Zhu, 2021). The p-value for testing the null of no average treatment effect is 0.05. The p-value for testing null effects on pre-treatment period is 0.82 ($H_0 : \theta_{2009} = \theta_{2010} = \theta_{2011} = 0$. The pointwise p-value is 0.065 (smallest possible value) for all post-treatment years from 2013 to 2016.

C Additional results

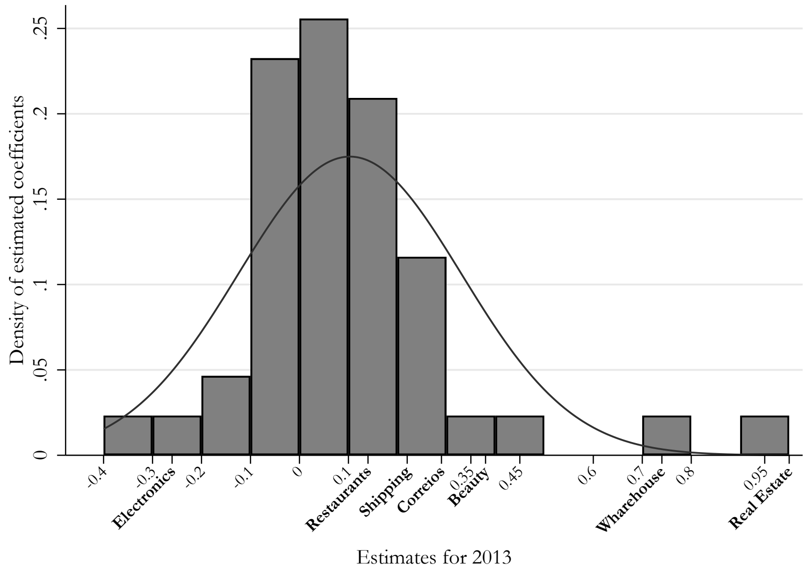

Figure C.1: Effects of E-commerce on Employment across Sectors in 2013

The regression is performed for two periods, where 2010 is the pre-treatment and 2013 is the post-treatment year. I interacted an indicator for each sector with the treatment variable. There are in total 43 sectors. The regression includes year, sector and microregion fixed effects. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

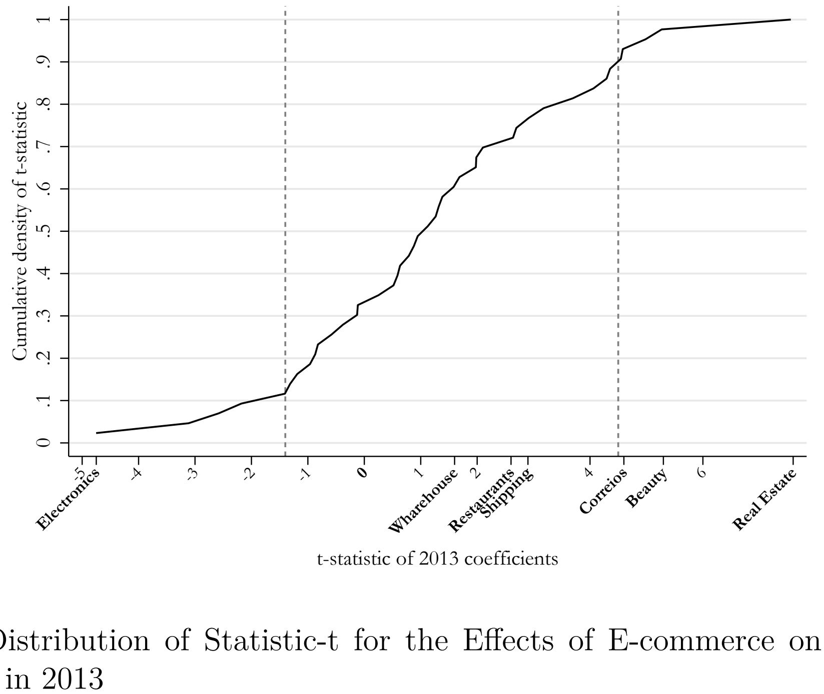

Figure C.2: Distribution of Statistic-t for the Effects of E-commerce on Employment across Sectors in 2013

The regression is performed for two periods, where 2010 is the pre-treatment and 2013 is the post-treatment year. I interacted an indicator for each sector with the treatment variable. There are in total 43 sectors. The regression includes year, sector and microregion fixed effects. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

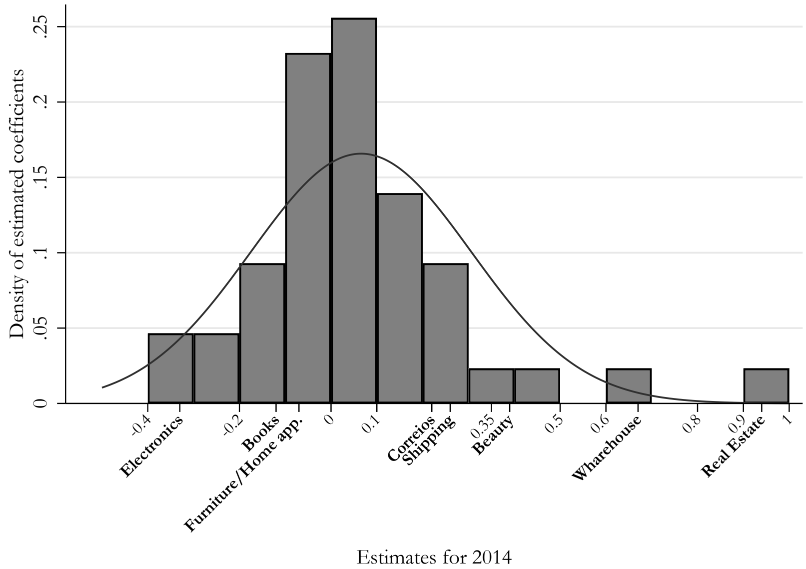

Figure C.3: Effects of E-commerce on Employment across Sectors in 2014

The regression is performed for two periods, where 2010 is the pre-treatment and 2014 is the post-treatment year. I interacted an indicator for each sector with the treatment variable. There are in total 43 sectors. The regression includes year, sector and microregion fixed effects. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

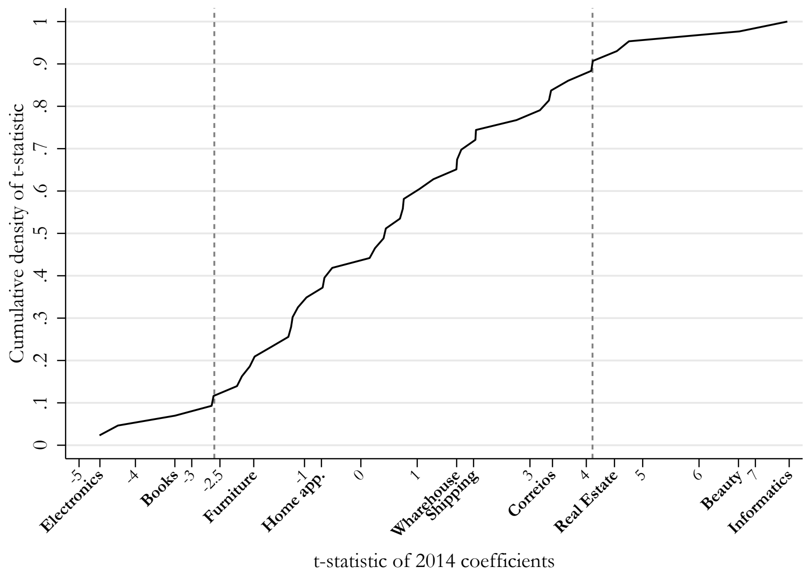

Figure C.4: Distribution of Statistic-t for the Effects of E-commerce on Employment across Sectors in 2014

The regression is performed for two periods, where 2010 is the pre-treatment and 2014 is the post-treatment year. I interacted an indicator for each sector with the treatment variable. There are in total 43 sectors. The regression includes year, sector and microregion fixed effects. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

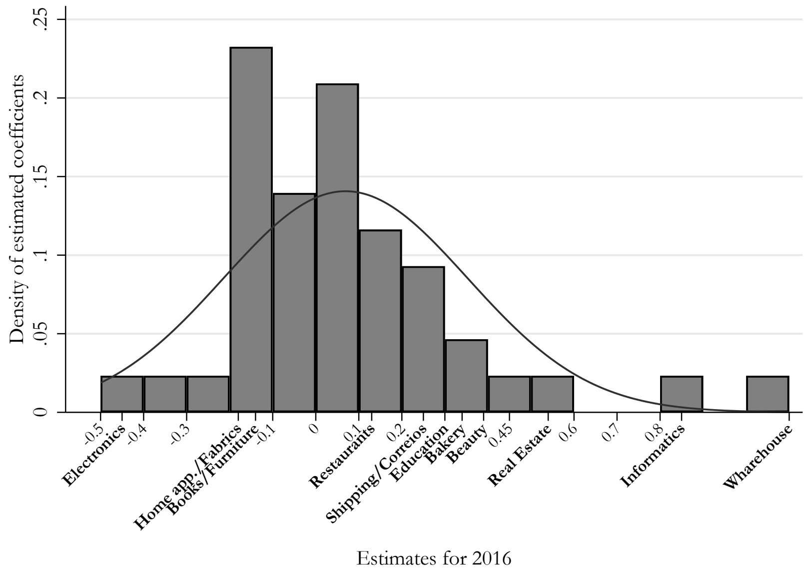

Figure C.5: Effects of E-commerce on Employment across Sectors in 2016

The regression is performed for two periods, where 2010 is the pre-treatment and 2016 is the post-treatment year. I interacted an indicator for each sector with the treatment variable. There are in total 43 sectors. The regression includes year, sector and microregion fixed effects. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

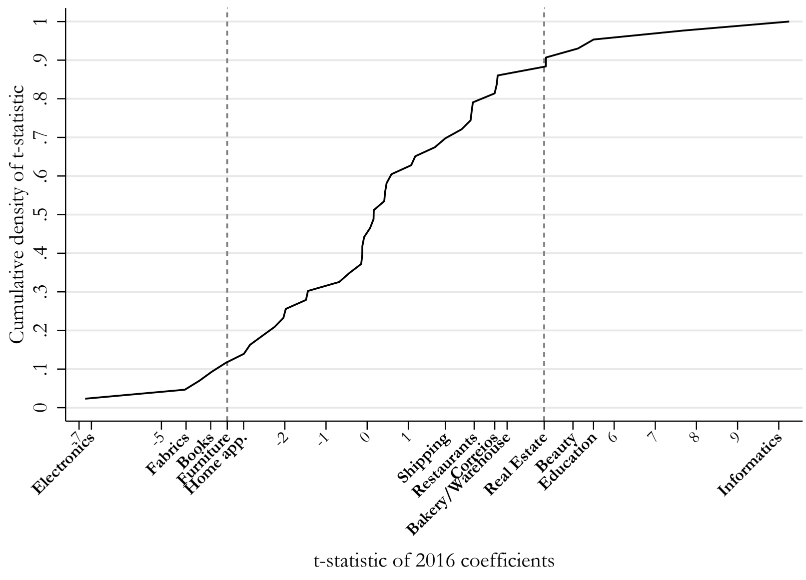

Figure C.6: Distribution of Statistic-t for the Effects of E-commerce on Employment across Sectors in 2016

The regression is performed for two periods, where 2010 is the pre-treatment and 2016 is the post-treatment year. I interacted an indicator for each sector with the treatment variable. There are in total 43 sectors. The regression includes year, sector and microregion fixed effects. Cluster-robust standard errors at the microregion and sector levels. Regressions weighted by number of workers in 2010.

C.0.1 Effects of E-commerce on Employment and Number of Firms in the Rest of the State

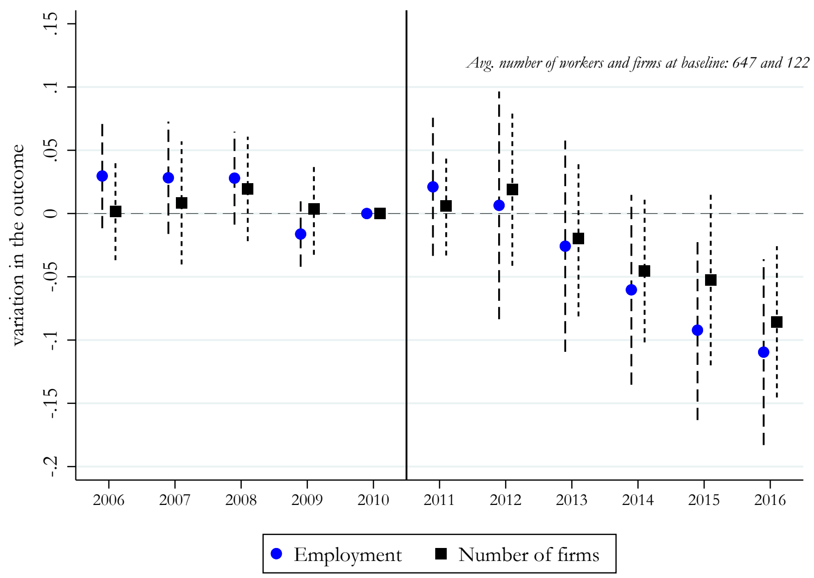

Figure C.7: Effects of E-commerce on Employment and the Number of Firms in Household Goods Stores

Notes: Effects at the microregion level, excluding the microregion hosting the DC. Regressions include year and microregion fixed effects. The covariates are urban population and GDP per capita in 2010, log of distance to the capital, and GDP growth from 2006 to 2010 interacted with years. Cluster-robust standard errors at the microregion level. Regressions weighted by number of workers in 2010.

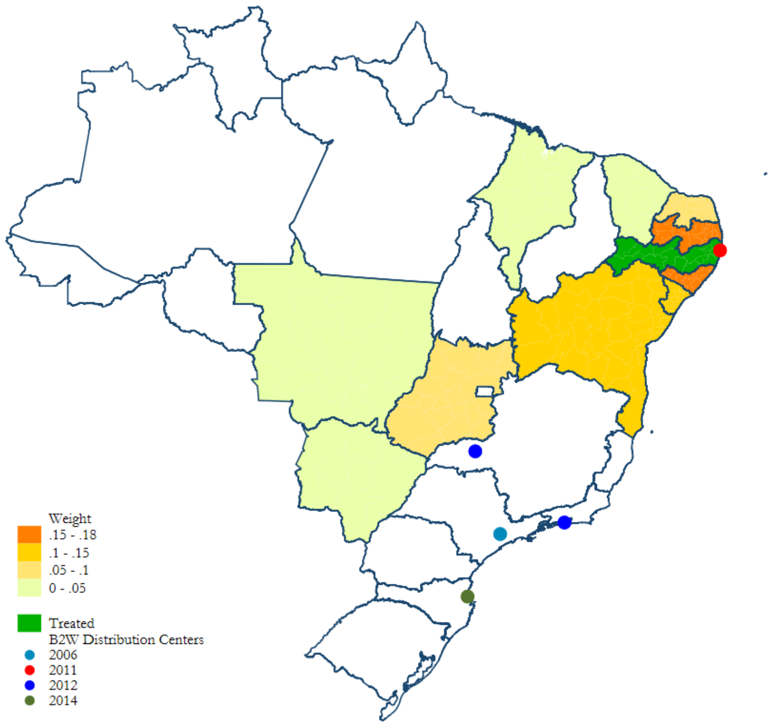

Figure C.8: Synthetic Difference-in-Differences: Weights and Control Pool in the Rest of the State

The donor pool is composed by eleven states in the Northeast and Center-West, and excludes the state capitals to measure the outcomes. 2010 Urban population and GDP per capita considered for the synthetic control's selection, in addition to pre-determined outcome.

Figure C.9: Effects of E-commerce on Employment in Household Goods Stores in the Rest of the State

The outcome is the log of workers. The synthetic control is composed by eleven states, whose weights minimize the trends of pre-determined outcomes up to a constant difference in levels. $\hat{ATT}$ = **-0.09 (0.09), p-value = 0.330.** Inference based on placebo allocation of treatment in the donor pool.

Figure C.10: Effects of E-commerce on the Number of Firms in Household Goods Stores in the Rest of the State

The outcome is the number of firms in year t normalized by the number of firms in the baseline. The synthetic control is composed by eleven states, whose weights minimize the trends of pre-determined outcomes up to a constant difference in levels. $\hat{ATT}$ = **-0.08 (0.07), p-value = 0.308.** Inference based on placebo allocation of treatment in the donor pool.