Identifying Agglomeration Spillovers: Evidence from Grocery Store Openings*

Franklin Qian†, Qianyang Zhang‡, & Xiang Zhang§

March 2024

Abstract

We estimate the strengths of agglomeration spillovers in the local non-tradable service sector using 413 grocery store openings in the U.S. in 2018–2019. We combine deep learning tools with propensity score estimation to find counterfactual opening sites and compare business outcomes surrounding actual and counterfactual sites. We find openings of grocery stores lead to significant growth in foot traffic to their opening locations and a 39 percent increase in foot traffic to businesses within 0.1 miles. The spillovers of demand are strongest between new grocery stores and businesses in wholesale and retail and hospitality services. We also find that grocery store openings lead to a 6.9 percentage point higher growth in the number of businesses within 0.1 miles of the openings 0–3 years later.

*We would like to thank Brent Ambrose, David Brasington (discussant), Donald Davis, Jonathan Dingel, Jeffery Fisher, Jessie Handbury, Matthias Hoelzlein (discussant), Erik James (discussant), Yuhei Miyauchi (discussant), Michael Pollmann, Steve Redding, Lindsay Relihan (discussant), Jacob Sagi, Philip Ushchev (discussant), David Weinstein, Abdullah Yavas, and Tingyu Zhou (discussant) for helpful comments and suggestions. Yang Chen and Shizhen Fan provided excellent research assistance. We also wish to thank the participants of the Online Spatial and Urban Seminar, UEA, AREUEA-ASSA, AREUEA national conference, RERI conference, CICF, the Eastern Finance Association annual meeting, and the NBER Summer Institute (Real Estate). We acknowledge financial support from the Real Estate Research Institute (RERI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors. They do not necessarily reflect the views of RERI.

†UNC Kenan-Flagler Business School, Department of Finance. Email: franklin_qian@kenan-flagler.unc.edu.

‡Columbia University, Department of Economics. Email: qz2344@columbia.edu.

§Princeton University, Department of Economics. Email: xiangzhang@princeton.edu.

1 Introduction

Businesses in the local non-tradable service industry sector (for example, grocery stores, restaurants and bars, pharmacies, etc.) often cluster together to reduce consumer search costs and benefit from shared demand generated by the economies of agglomeration (Wolinsky (1983)). In the US, most shopping centers are anchored by at least one national name brand or department store expected to be the biggest draw of foot traffic. With the ongoing retail apocalypse since around 2010, strip malls and shopping centers anchored by grocery stores have been more favored by investors, as visitors are still flocking to them. (Fung (2020)). How much demand and revenue spillover do grocery stores generate in neighboring businesses? How do the openings of grocery store anchors change the surrounding business dynamics – what types of new businesses grow fastest in number and size? There has been a dearth of empirical research evidence on the externalities of grocery anchors. This paper fills this gap by studying the externalities of anchoring grocery stores on nearby retail businesses.

We collect a comprehensive sample of grocery stores that opened across the U.S. in 2018 and 2019. We develop an identification strategy in which we define suitable alternative locations for the openings of grocery stores in their vicinity. By comparing businesses surrounding real openings and those surrounding alternative opening locations, we estimate the causal spillover effects of openings of grocery stores on (i) the demand for nearby incumbent businesses measured by foot traffic of visits and business revenue, and (ii) the growth of businesses in terms of number and size.

We follow recent developments in the literature on spatial causal inference to find suitable alternative locations for openings. We borrow tools from Convolutional Neural Networks (CNN) to identify alternative locations for grocery store openings in nearby neighborhoods similar to the actual opening locations from the firms' perspective. Finding candidate locations for grocery stores in a continuous space is a natural setting where CNNs can be useful. We first discretize space into grids and then estimate a CNN model to predict the likelihood of a location having a grocery store using various information about the existing business composition and neighborhood-level demographics in the vicinity. The CNN model helps filter a continuous space with numerous possible counterfactual locations down to a donor pool of high-quality potential counterfactual sites. Next, we estimate a propensity score model with flexible inputs to predict suitable counterfactual opening locations for each actual opening that can be used as a control group. We then compare the outcomes of businesses near the actual and matched counterfactual opening locations to estimate the causal effects of grocery store openings on nearby businesses.

We first show that after a grocery store opening, the monthly foot traffic at the location of the opening increases by 333 percent 6–10 months after the opening, lending credence to the precision of the opening months of these stores. Next, we analyze the demand spillovers measured by foot traffic to the businesses surrounding the grocery store openings. We find that these spillovers are concentrated within 0.1 miles from the openings, with an average increase in foot traffic by 39 percent 6–10 months after the opening, relative to foot traffic to businesses surrounding the counterfactual sites. These local demand spillovers decrease sharply and dissipate to a statistically indistinguishable zero after 0.1 miles. We further find these spillovers are less localized and extend to businesses within 0.2 miles of openings in below-median population density areas.

We further examine the heterogeneity in the demand spillovers by the categories of businesses nearby. We find evidence the spillovers on nearby businesses are the strongest for wholesale and retail as well as hospitality services. On average, wholesale and retail stores (excluding grocery stores) within 0.1 miles of a grocery store experience a 40 percent increase in foot traffic 6–10 months after the opening, while businesses in accommodation, eating, and drinking experience a 38 percent increase in foot traffic. In comparison, the spillovers are weakest for services in medical, welfare, and healthcare and existing grocery stores within 0.1 miles with economically small increases in foot traffic that are statistically insignificant. Further, businesses located in the same real estate properties gain significantly more in terms of foot traffic than those that are not. For instance, wholesale and retail businesses in the same property as the grocery store opening have an increase of 70 percent in foot traffic 6–10 months after the opening, whereas wholesale and retail businesses outside the property of the grocery store opening only have an increase of 30 percent.

In addition, we examine the heterogeneous demand spillovers generated by the openings of different types of grocery stores. We find that positive spillovers are most driven by the national chains (e.g. Whole Foods Market and Trader Joe's) with a 52 percent increase in foot traffic to nearby businesses 6–10 months later. In contrast, the effects of opening a dollar store on nearby businesses' foot traffic are negligible. We also find the demand spillovers generated by grocery stores in higher-income neighborhoods are in general larger, potentially due to more correlated demand with nearby businesses. A higher share of grocery store openings in higher-income neighborhoods belongs to national chains, thus giving them the strongest spillovers.

Interestingly, we do not find evidence of any negative competitive effects of grocery store openings on nearby existing grocers. However, that masks significant heterogeneity in the effects across different pairings of grocery store entry and existing grocers. We find suggestive evidence that openings of big-box retailers and discounters reduce the foot traffic to existing national grocery store chains nearby and vice versa. But neither seems to harm other types of grocers.

Next, we provide evidence of how these grocery store openings change the dynamics in the surrounding business environment. We focus on two measures which are the net growth in the number of businesses and the establishment-level employment, respectively. We further decompose the net growth into growth due to openings and closures of businesses (extensive margin) and expansion and contraction of continuing businesses (intensive margin). First, we find the number of businesses within 0.1 miles of the openings grows faster by 7 percentage points 0-3 years later, relative to the control group. In contrast, businesses within 0.2-0.4 miles experience a 3 percentage points slower growth in the number of businesses and 2 percentage points lower employment growth. The slower business growth further away is potentially due to grocery anchors and the agglomeration economies it introduced drawing customers away from its immediate vicinity. The positive effects on business growth rates are concentrated within 0.1 miles of the grocery store openings and peak 1 year after the openings, before declining in subsequent years. Decomposition of the overall effects on growth rates reveals that grocery anchors not only speed up new business entries but also slow down the closures of existing businesses. The positive effects on the growth rate in employment are driven by the extensive margin due to business entries and closures.

Further examination of treatment effects at the 4-digit NAICS code level suggests that a few nearby industries have the largest synergies with grocery anchor openings in having the fastest growth in number and size. They include supermarkets and grocery stores, furniture stores, consumer goods rentals, drinking places, and banks. They also belong to business categories that gain popularity in customer traffic. On the other hand, "independent artists, writers, performers", "advertising, public relations, and related services", and some healthcare services also grow the fastest in number and size but do not necessarily gain a lot of customer traffic. Overall, our results highlight that grocery store openings also bring substantial business growth at the extensive margin, in addition to the demand spillovers they generate on existing businesses.

Our paper is closely related to a large body of literature on the economics of agglomeration, a concentration of economic activities in a certain geographic sphere. Most existing papers that study spillovers on surrounding businesses focus on the productive performance of firms in manufacturing and tradable services.[1] For example, these papers find that agglomeration in the manufacturing industries could increase the total factor productivity and output of the local plants (Henderson, 2003; Ellison et al., 2010; Greenstone et al., 2010).[2]

[1]See, for example, Greenstone et al. (2010), Ellison and Glaeser (1997), Glaeser et al. (1992), Henderson et al. (1995), and Rosenthal and Strange (2003).

[2]A related strand of literature goes beyond productive outcomes of firms, and shows that the openings and closings of businesses can generate substantial spillovers on the environment nearby. Examples include Currie et al. (2015) on health and housing value consequences of toxic plant openings and closings, Qian and Tan (2021) about the effects of firm entry on incumbent residents' outcomes and welfare, Rosenthal and Urrego (2021) about the crime deterring effects that result from the spatial concentration of retail activities, Gupta et al. (2020) and Diao et al. (2017) about the spillover effects of mass transit entry on housing prices.

Within the non-tradable service sector, a large extant literature has studied how shopping center owners internalize the externalities of anchor tenants by offering anchor tenants significant discounts in rent and differential rental rates for non-anchor tenants based on spatial proximity and sales incentives (Benjamin et al., 1992; Gatzlaff et al., 1994; Pashigian and Gould, 1998; Wheaton, 2000; Konishi and Sandfort, 2003; Gould et al., 2005; Liu and Liu, 2013). Several papers examine the consequences of the dis-economies of agglomeration. For example, they find bankruptcies of firms and closure of national chains in the non-tradable service sector have significant negative externalities on nearby tradable firms, leading to decline in employment, foot traffic, and business closures (Shoag and Veuger, 2018; Bernstein et al., 2019; Benmelech et al., 2019; Knight, 2022). A body of literature finds mixed spillover effects on the openings of big-box retail stores on local labor markets, particularly in the tradable service sector.[3] Despite the existence of substantial recent literature studying agglomeration economies and externalities,[4] causal estimates of positive agglomeration spillovers in the non-tradable service sector are still rare and their demand mechanisms are relatively under-studied. We contribute to this literature by providing a new identification strategy to estimate the causal effects of business openings and quantify the strength of agglomeration spillovers. We present micro evidence on how anchoring grocery stores that attract customers generates demand and revenue spillovers on nearby stores and how these effects change as a function of distance to the opening stores.

Prior work has shown that the benefits of agglomeration economies such as an increase in the productivity of firms can arise due to a variety of reasons, such as (i) reduced transportation cost of goods (ii) knowledge spillovers (iii) labor market pooling that allows for greater specialization (iv) better firm-worker match through the reduction in search friction.[5] Within the non-tradable service sector, the literature suggests that increased productivity due to ag-

[3]See, for example, Basker (2005), Neumark et al. (2008), Jia (2008), Merriman et al. (2012), Arcidiacono et al. (2016), Ellickson and Grieco (2013), Bertrand and Kramarz (2002) and Sadun (2015).

[4]Other examples in this literature include Agrawal and Cockburn (2003), Zhou and Clapp (2015), Liu et al. (2018), Rosenthal and Strange (2020), Liu et al. (2020), Kuiper et al. (2021).

[5]For recent surveys, see Duranton and Puga (2004), Glaeser and Gottlieb (2009), and Moretti (2010).

glomeration could arise due to shared customer traffic, which reduces consumer search costs and benefits geographically proximate stores (Pashigian and Gould, 1998; Gould et al., 2005; Bernstein et al., 2019; Benmelech et al., 2019). In addition, literature has found that the store format matters for the strength of agglomeration spillovers that can be generated. For example, Leung and Li (2021) finds that big-box retailers affect consumer shopping behavior and welfare to a greater extent than other competing retailers. Our paper is also related to recent literature on the welfare consequence of trip-chaining (Miyauchi et al., 2021a; Oh and Seo, 2022; Relihan, 2022) that highlight trip-chaining as one of the key mechanisms of agglomeration. In our paper, we study carefully how these spillovers vary across different types of grocery stores and surrounding business categories, as well as how grocery anchors change the surrounding business dynamics. Our results suggest that the complementarity of businesses is one of the key mechanisms driving agglomeration.

Our paper contributes to the growing literature on causal inference in a spatial setting. For example, Currie et al. (2015) and Ellickson and Grieco (2013) parametrically estimate the treatment effects in different distance bins from the treatment location. Diamond and McQuade (2019) develop a non-parametric method to estimate the spillovers. Butts (2021) develops a method to estimate the spatial spillover effects semi-parametrically. Traditional methods for spatial treatment effects usually compare outcomes in an inner ring with those in an outer ring around the treatment site. However, many observable and unobservable characteristics may correlate with the distance from a business to a grocery store opening, rendering such an identification strategy invalid. We differ from previous papers that rely extensively on ring analysis methods that compare outcomes in the inner ring with the outer ring. Following Pollmann (2020), we build a CNN model to find counterfactual locations that are similar to the locations of actual grocery store openings. We make the following four-fold contributions relative to Pollmann (2020)'s methodology. First, we include a richer set of characteristics as inputs to the CNN model. Second, we create heatmaps that represent the desired output of our CNN model for each opening location. The purpose of these heatmaps is to help us tell how good our CNN model is at predicting a suitable opening location.[6] Third, we demonstrate that we are able to find counterfactual locations for grocery store openings across a broad set of geographies using our GAN-based CNN model. Fourth, we importantly rely on the sharp timing of the grocery store entries as a shock to strengthen the credibility of our identification strategy, whereas Pollmann (2020) only uses cross-sectional variation. Seeing sharp changes in foot traffic around the precise timing of these shocks makes our results more convincing.

The remainder of the paper is organized as follows. Section 2 discusses the data used for the analysis. Section 3 presents the empirical approach and reduced-form effects. Section 4 concludes.

2 Data

2.1 Sample of Grocery Store Openings

We combine data from multiple sources to compile a comprehensive list of grocery store openings in the United States in 2018 and 2019. From Chain Store Guides, we obtain the names and addresses of grocery stores that opened in 2018 and 2019. From SafeGraph's Core Places data, we supplement our sample with newly opened points of interest (POI) classified as grocery stores.[7] Finally, we add a sample of openings from Compstak, which provides commercial real estate lease comps. We identify a grocery store opening from newly signed leases for grocery stores in 2018 and 2019. In total, we have 413 openings in grocery stores. Figure 1 shows the geographic distribution of our sample of grocery store openings. In Table A2, we show the share of openings by category of grocery stores. We divide grocery stores into four categories: (1) chains that account for 33.2% of the sample (2) big-box retailers and

[6] We name this method of creating heatmaps for labels that CNN produces as "heatmap labeling" hereafter.

[7] Grocery stores are identified by NAICS codes 445110 (Supermarkets and Other Grocery (except Convenience) Stores), 445120 (Convenience Stores), 452319 (General Merchandise Stores, including Warehouse Clubs and Supercenters).

discounters (43.8% of the sample) (3) dollar stores (21.8% of the sample) (4) convenience stores (including independent stores) (1.2% of the sample). Appendix table A3 lists the number of openings by grocery store chains in the sample.

We match the addresses of grocery store openings with addresses of commercial POIs from Safegraph to obtain the foot traffic measure of the monthly number of visits to each store. By identifying structural breaks in monthly visits, we impute the opening month for each grocery store.[8] Appendix C provides details on how we find and verify the opening month for each grocery store.

2.2 Construction of Outcome Variables

2.2.1 Visits to Business Establishments from GPS Tracking

We obtain the monthly number of visits to commercial POIs from Safegraph. Safegraph covers the vast majority of POIs in the U.S. and collects foot traffic to these POIs through GPS tracking of apps on cell phone devices. In addition, Safegraph provides POI characteristics such as their names, addresses, and industry classifications. We classify commercial POIs into five categories: 1) Wholesale and Retail; 2) Accommodation, Eating, and Drinking; 3) Medical, Welfare, and Healthcare; 4) Finance, Real Estate, Communication, and Professional Services; 5) Other Services, similar to Miyauchi et al. (2021b).[9]

2.2.2 Consumer Spending at Business Establishments

We obtain the dataset of consumers' spending behavior at specific POIs from SafeGraph. The Spend dataset aggregates anonymized debit and credit card transaction data to individual places in the U.S. at a monthly time interval. The dataset dates back to January 2019. The

[8] Whenever information is available, we manually validate the imputed opening month of a store by looking up online news articles announcing the opening.

[9] Wholesale and retail stores are identified by 2-digit NAICS codes 42, 44, 45; accommodations, eating, and drinking places are identified by 2-digit NAICS code 72; finance, real estate, communication, and professional services are identified by 2-digit NAICS codes 52, 53, 54, 55, 56; medical, welfare and healthcare stores are identified by 2-digit NAICS code 62; and other services include 2-digit NAICS codes 51, 61, 71, 81, 92.

Safegraph Spend data records spending for over 10 million customers at over 1.1 million POIs in the U.S. It covers a total of 5454 brands in the dataset. From the Spend data, we can observe the number of total spending, customers, and transactions at a POI on a monthly basis. In addition, the dataset also allows us to observe spending in-person vs. online, spending by customer demographics such as income at the establishments.

2.2.3 Sales and Employment of Establishments

We obtain annual sales and employment in each business location from Data Axle (formerly Infogroup). The database covers 226 million verified U.S. businesses dating back to 1997, providing the establishment-level names, street addresses, longitudes, and latitudes, and sale volumes and employee counts. The annual snapshots of businesses also allow us to track the entries and exits of businesses over the years. We use the number and employee counts of businesses that are located within 0.5 miles of the grocery store openings and their matched counterfactual sites from 2014 to 2021 in our analysis of business dynamics.

3 Conceptual framework

In this section, we propose a conceptual framework to guide our empirical exercises.

3.1 Model setup

There are K composite of service goods, and a numeraire good which represents all non-service consumption. Each composite of goods is sold by a unique category of business. In our paper, we follow Miyauchi et al. (2021a) to categorize surrounding service businesses into 6 types: (1) wholesale and retail (excluding grocery); (2) grocery stores; (3) accommodations, eating, and drinking; (4) finance, real estate, communication, and professional services; (5) medical, welfare, and healthcare services; (6) Other services.

Each individual i chooses the number of visits to each composite of goods, denoted as

X_{ik}, and a numeraire consumption good x_{i0}, and maximizes the following Cobb-Douglas utility function,

$$\max_{\{x_{i,kj}\}} \quad x_{i0}^{\theta_0} \prod_{k=1}^{K} X_{ik}^{\theta_k}$$

where $\theta_k, k = \{0, 1, \cdots, K\}$ is the taste parameter for each type of composite goods and $\sum_{k=0}^{K} \theta_k = 1$. Each composite good X_{ik} is assumed to be a CES aggregator over one's visits to each place of interest (POI) j.

$$X_{ik} = \left[\sum_{j=1}^{J_{ik}} x_{i,kj}^{\frac{\sigma_k - 1}{\sigma_k}} \right]^{\frac{\sigma_k}{\sigma_k - 1}}, \sigma_k > 1$$

where J_{ik} is the number of type-k POIs in the choice set of individual i, and σ_k is the elasticity of substitution within business type k. We assume that $\sigma_k > 1$ for all k. This assumption implies that within the same category, each POI acts as a gross substitute for the others.

For each individual i, she chooses the number of visits to type-k POI, denoted as $x_{i,kj}$, In what follows, we abstract away from individual-specific choice and remove the individual-specific subscript i.

We assume that the monetary cost of service offered by type-k POI j is h_{kj}. This could be interpreted as the spending at type-k POI j within a single visit. We assume that the spending h_{kj} is constant before and after a grocery store entry.[10] On top of this monetary cost of service, consumers need to pay a monetary cost corresponding to the travel time spent, denoted as γt_{kj}, similar to Su (2022). In this equation, γ is the opportunity cost of the individual's time, and t_{kj} is the time associated with traveling to and from the type-k POI j. As a result, the total price of purchasing service at POI j in category k is,

$$p_{kj} = h_{kj} + \gamma t_{kj}$$

[10] This precludes the scenario wherein a given POI upgrades its service quality to justify an elevated price or implements a higher price markup.

We normalize the price of the numeraire good to be 1. The resident is subject to the budget constraint,

$$x_0 + \sum_{k=1}^{K} \sum_{j}^{J_k} x_{kj}(h_{kj} + \gamma t_{kj}) \leq I$$

where I is the individual's realizable income, if she were not to have any visits to surrounding businesses and consume only the numeraire good x_0.

The Dixit-Stiglitz price index for visits at type-k business is,

$$P_k = \left[\sum_{j}^{J_k} (h_{kj} + \gamma t_{kj})^{1-\sigma_k} \right]^{1/(1-\sigma_k)} \tag{1}$$

The price index reflects the unit cost of consuming a bundle of goods X_k, aggregating the costs of accessing each POI within category k (Dixit and Stiglitz, 1977). First, it is important to understand that this price index rises with the cost of services at each POI (h_{kj}) and the transportation cost (γt_{kj}). Consequently, individuals with access to pricier POIs or enduring longer commutes to surrounding businesses will encounter a higher price index. Additionally, when $\sigma_k > 1$, the access to more POIs leads to a lower price index, reflecting the "love of variety" effect described by Krugman (1979). This principle suggests that variety in consumption choices can reduce the overall cost faced by consumers. Finally, the parameter σ_k determines the sensitivity of the overall price index to the number of accessible POIs (J_k) and the associated costs. On the one hand, a larger σ_k indicates that consumers are more sensitive to the price associated with visiting each POI. As a result, she visits the POIs with lower service prices or travel costs disproportionately more. On the other hand, the "love of variety" force is weaker when σ_k is large. Given the cost associated with the entry POI, an increase in variety only weakly lowers the price index when the elasticity of substitution is large.

The optimal number of visits to store j in category k could be expressed as,

$$x_{kj} = \left(\frac{p_{kj}}{P_k} \right)^{-\sigma_k} \times \frac{I\theta_k}{P_k} = (h_{kj} + \gamma t_{kj})^{-\sigma_k} \times P_k^{\sigma_k - 1} \times I\theta_k \tag{2}$$

From this equation, it is clear that individuals visit POIs with lower total price more. The elasticity of the number of visits to the price is governed by the elasticity of substitution σ_k.

3.2 Model Predictions

From equation (2), we could see that,

$$\ln(x_{kj}) = -\sigma_k \ln(p_{kj}) + (\sigma_k - 1)\ln(P_k) + \ln I\theta_k$$

and consequently,

$$\Delta \ln(x_{kj}) = -\sigma_k \Delta \ln(p_{kj}) + (\sigma_k - 1)\Delta \ln(P_k)$$

In this paper, we focus on the impact trip bundling on the number of visits. Trip bundling can reduce the traveling time associated with visiting a type-k POI j, t_{kj}. Because of this reduced travel time, the total price of visiting that place, p_{kj}, also goes down. When the total price is lower, we expect an increase in the number of visits to POI j. The magnitude of this increase in visits depends on the elasticity of substitution within each category, represented as σ_k. This parameter describes how willing people are to switch to different POIs if there's a change in the price. The higher this elasticity, the more likely people are to change their visiting habits in response to the reduced costs brought about by trip bundling.

Additional assumption I: We assume that stores in proximity to a newly opened grocery store experience a more significant decrease in travel time.

Empirical prediction I: In regressions where we control for a opening case by business category by calendar year fixed effects, a POI closer to the new grocery store will experience a larger increase in the number of visits following the grocery store's opening, compared to a POI in the same business category that is further away from the grocery store entry.

The intuition is, conditional on the opening case by business category by calendar year fixed effects, we have directly controlled for both $(\sigma_k - 1)\Delta \ln P_k$ and $\ln(I\theta_k)$. Consequently, any changes in the number of visits are attributed solely to the term $\Delta \ln(p_{kj})$. On average, across POIs within the same business category, the discrepancy in the change in the number of visits is exclusively linked to the differential changes of the log unit price, represented by $\Delta \ln(p_{kj})$. Assuming the service price at location j remains constant, the differences in the change in x_{kj} is further only due to the differential changes in travel time. With trip bundling, stores in proximity to a newly opened grocery store experience a more significant decrease in travel time (additional assumption I). Consequently, these POIs will witness a more substantial increase in the number of visits.

Empirical prediction II: In regressions where we control for an opening case by business category by calendar time fixed effects, POIs in business categories with a larger σ_k and more likely to form a trip bundle with grocery stores will have a larger increase in the number of visits.

Similarly, after controlling for the opening case by business category by calendar time fixed effects, the change in the number of visits is governed by $\sigma_k \Delta \ln(p_{kj})$. Consequently, POIs within business categories that either (1) possess a higher σ_k, or (2) are more likely to be included in a trip chain with the grocery store opened will benefit from a decrease in travel time, and experience a higher increase in the number of visits.

4 Empirical Approach

After a grocery store opening, the changes to the foot traffic to nearby businesses could be attributed to different channels: (1) the grocery store as an anchor not only brings new foot traffic to itself, but also draws customer traffic to nearby businesses through increased salience of those businesses, reduced search cost, and trip chaining, etc. (2) There exist some area-wide general equilibrium (GE) effects independent of the grocery store entry. For example, the entire area surrounding the opening could be trending up in business densities and customer demand due to local economic conditions. In the second channel, the increase in foot traffic to nearby businesses would have occurred in the absence of the grocery store entry, as the entire area is on an upward growth trajectory.

To identify the treatment effects of grocery store openings on foot traffic to surrounding businesses, a simple strategy would be to compare foot traffic to businesses close to the opening with that to businesses further away via a spatial difference-in-difference design. However, the key confounding issue is businesses close to the opening could be unobservably different from those further away, and their distance to the opening is correlated to their likelihood of attracting a grocery store nearby. For example, businesses that are closer to the opening could be in locations where the pre-existing densities of businesses are already high and the options for grocery shopping are not too many, so they are more likely to attract a grocery store opening nearby.

To overcome this identification challenge, ideally, we would like to compare the outcomes of businesses around a grocery store opening with those of businesses around an alternative location for the grocery store opening, which is almost identical in its surrounding environment before the opening. Therefore, we need to identify alternative desirable locations for the openings of grocery stores in our sample. When a grocery store chooses a location to open, its desirability is likely influenced by many demographic characteristics of its surrounding neighborhoods where its customers reside and also by the business environment, such as the types and densities of existing businesses in the vicinity. For example, locations that have a nearby customer base with strong purchasing power, are close to a decent amount of businesses, and have easy traffic access and parking availability are more likely to attract grocery store openings. Given the high dimensionality of the set of potential covariates that explain the desirability of a location, we need a parsimonious model to determine the desirability of a location.

Our solution is to adopt a Generative Adversarial Networks (GAN)–based Convolutional Neural Networks (CNN) model, which is commonly used in machine learning applications such as image recognition, to a spatial setting. The model predicts a pool of alternative desirable opening locations in the vicinity of each real opening location. Next, we estimate propensity scores for the likelihood of opening a grocery store for both real and potential locations selected by the model. We then match each real opening with an alternative opening location that most closely resembles the real one. To estimate the treatment effects, we would compare the businesses surrounding real openings with those surrounding the alternative opening locations. We summarize the key details of each step below.

4.1 Predict Counterfactual Locations for Grocery Store Openings

We present essential details of our predictive model for finding the appropriate counterfactual locations. The core of our predictive model is a GAN-based CNN following Pollmann (2020). Our innovations relative to his model are (i) introducing a much richer set of input characteristics that could matter for the desirability of a location for the opening of a grocery store,[11] (ii) adopting heatmap labeling as a technique to smooth the labels and enhance model performance, and (iii) demonstrating that we can use the GAN-based CNN to find counterfactual sites for establishment entries across a broad set of cities and geographies.

Our model construction has three main steps. The first step prepares and transforms the raw data. Our goal is to prepare a spatial data set with economically meaningful input characteristics in an accessible format to train a CNN model. The second step is to find a pool of potential candidates for counterfactual sites. To do that, we build a GAN-based CNN model to find a pool of candidate sites similar to real grocery store opening sites in nearby demographic characteristics and the existing business environment. The third step is to determine the best counterfactual sites for grocery store openings. Given the high dimensionality of the features that affect a grocery store's location decision, we use Principal

[11] The features Pollmann (2020) uses are the location of surrounding grocery stores, restaurants, and a combination of all other kinds of business.

Component Analysis (PCA) to select the essential features. We then use the selected features to estimate a propensity score model for all real opening sites and potential candidate sites from step two. Finally, we match each real opening to one counterfactual site in its vicinity with the closest propensity score from the pool of candidates. These three steps allow us to find a valid control group to draw causal inferences about the impacts of grocery store openings.

4.1.1 Sample Preparation for Training CNN Model

CNN is a type of deep learning model that excels at processing data that has a grid pattern and performing classification. We start with constructing input data for the CNN model. Next, we train the CNN model with 80% of the data and test its predictive accuracy with the remaining 20% of the data. When we obtain a satisfactory result without over- or under-fitting, we use the entire data to train the final version of the CNN model.

Grocery stores take into account a variety of factors when choosing a location to open a new store. We summarize five major factors that grocery store owners consider in their business location strategy (Ghosh, 2022; Krishna, 2021; Waters, 2021): 1) neighborhood demographics and characteristics; 2) accessibility, visibility, and traffic of the location; 3) zoning regulations; 4) competition and neighbors; 5) location costs. For each of the five factors, we collect data from different sources. We present the variables used as input features to the CNN model in Appendix Figure B1.

The typical input data to the CNN model has a grid pattern. We start by discretizing the continuous geographic space surrounding each real opening into a 10×10 grid of cells, each cell has a width of 0.025 miles. The number and size of the cells are determined by a trade-off between the model's predictive accuracy and computational efficiency. A large cell size reduces the predictive accuracy of the model, whereas a small cell size greatly increases the computational burden but has little improvement in accuracy. We position each real opening at the centroid of the bottom left corner of the 10×10 grid. We further assume that the directions of the grid are defined by the four cardinal directions, i.e., north, south, east, and west. We later relax this assumption in the following step of data augmentation.

In deep learning, the limited size of training data constrains the model performance for classification. In our case, we only have 413 grocery store openings that can be used for training, which falls short of the large training samples that traditional deep learning algorithms require. We employ data augmentation techniques to generate additional training data. The key to data augmentation is to generate new samples by transforming the original data. We use three types of data augmentation techniques: translation, rotation, and mirror transformation. A translation shifts a grid vertically or horizontally. Previously, we placed the real opening in the bottom left corner of a 10×10 grid of cells for convenience. Translation allows a real opening to be located in any cell of a grid. A rotation randomly rotates the grid up to 360 degrees clockwise. Rotation relaxes the assumption that the orientation of the axes of the grid aligns with the four cardinal directions. A mirror transformation flips the left and right sides of the grid. Suppose you have an existing opening on the east coast of the US, the mirror transformation creates an opening on the west coast, with all other conditions being identical. We show visually how each transformation works in Appendix Figure B2b. These three data augmentation techniques enable us to greatly expand the size of our training data.

We train the CNN model with the original real openings and the set of opening locations created through data augmentation. When the CNN model completes the training, the next task is to predict counterfactual sites for each of the real openings. To search for counterfactual sites, we again create 10×10 grids of cells at random locations within a 5-mile radius from each real opening. The distance of 5 miles allows us to reasonably control for local unobservable factors that could influence the location choice of each opening. Then we calculate the average value of the chosen input features in each grid cell. In this way, for each real opening, we obtain a three-dimensional matrix of size 24 (input features) $\times 10 \times 10$ as the input data to the CNN model. The model then predicts the probability of having a grocery store opening in each grid cell. Appendix Figure B2a illustrates how we construct these input matrices used by our CNN model.

4.1.2 GAN-based CNN Model to Screen Potential Desirable Opening Locations

Next, we give an overview of the architecture of our CNN model. To improve the predictive accuracy of our model, we adopt the idea of Generative Adversarial Networks (GAN). In addition, we create heatmaps for the labels that our CNN model produces. This technique enables us to tell how good our CNN model is at predicting suitable grocery store opening sites and to improve the model's performance.

CNN has been widely used for image classification (Krizhevsky et al. (2017)). Its core idea is to use a "kernel" (a small matrix) to extract features from the input data through convolution,[12] and to perform the task of feature selection and classification, as illustrated in Appendix Figure B3a.

Our CNN model performs classification in two stages. First, it needs to determine whether the given 10×10 grid of cells contains a real opening that is missing (hence it is an ideal counterfactual site). Second, if the answer is yes to the first, CNN should predict the precise location of this ideal counterfactual site on the 10×10 grid. The first stage is a two-class classification, with one class that has the real openings missing and the other class that does not. The second stage is a classification of 100 classes. CNN needs to predict one correct cell out of the 100 cells.

The main difficulty lies in the first stage, in which CNN tells whether an area lacks a real opening, i.e., how CNN distinguishes between an area with a real opening and a highly similar counterfactual site. To deal with this problem, we adopt the idea of Generative Adversarial Networks (GAN) to improve the accuracy of the CNN model. GAN's core idea is to train a pair of mutually competing networks simultaneously (Goodfellow et al., 2020), a discriminator, and a generator. The discriminator dedicates to distinguishing different

[12] Convolution is a matrix operation that adds each element of the input matrix to its neighbors, using the "kernel" as the weight.

types of samples, and the generator works on generating counterfeit samples to confuse the former. The two networks compete against each other. The generator gradually generates highly similar samples, and the discriminator works hard to improve its ability to identify the disguise made by the generator. Therefore, GAN has more excellent capabilities to discriminate between different samples, especially between highly similar samples, than traditional networks.

However, GAN is highly computationally intensive; training two neural networks simultaneously is complicated and time-consuming. To ease the computational burden, we adopt the philosophy behind GAN and simplify its structure. Specifically, we artificially forge samples as a new data type and then send them to the discriminator. By implementing this method, we are playing the role of the generator ourselves. Specifically, we have three input types, as demonstrated in Appendix Figure B3b. The first type of input samples (hereinafter referred to as type I inputs) are the artificial counterfactual sites. They are made by purposefully deleting the real openings from the grids, so they have surroundings similar to real openings but are missing an opening in a specific location. Type II inputs are the original grids where the real openings are preserved. Type II inputs play the role of a generator in GAN to deceive the CNN. Type III inputs feed CNN with random grids with no counterfactual sites. They are made of grids sampled from arbitrary locations that do not necessarily contain any real openings. When the CNN model is well-trained, it should be able to identify no counterfactual sites in type II or type III inputs. Furthermore, it should be able to identify counterfactual sites from the type I inputs.

To further enhance the accuracy of the GAN-based CNN model, we create heatmap format labels for CNN (Goodfellow et al., 2020). The purpose of creating heatmap labels is to provide information on the extent of the error made by CNN. Intuitively, we are guiding CNN to learn from its mistakes by telling CNN how wrong it is. In contrast, traditional labels only distinguish right from wrong. A comparison of the traditional labeling method and the heatmap labeling method is shown in Appendix Figure B4.

Our GAN-based CNN model eventually achieves 99.7%, 98.8%, and 93.8% accuracy in distinguishing the three types of input (type I, type II, and type III). Further, conditioning on type I inputs, the accuracy in correctly identifying the type I inputs and giving the exact cell of the missing real opening is 90.2%. This accuracy rises to 96.6% if we relax the criteria and count all the predicted cells within the error of one cell width as correct.

4.1.3 Propensity Score Estimation and Matching to Predict Counterfactual Opening Locations

We use the trained CNN model to search for counterfactual sites in areas within 5 miles of the real openings. After searching those areas, CNN gives us millions of candidates for counterfactual sites. We must filter these candidates based on specific criteria and match the real openings with the most suitable counterfactual sites.

We adopt a two-stage propensity score estimation and matching process to screen the candidate counterfactual sites and match the real openings with them. For the pool of real openings and candidate counterfactual sites predicted by the CNN model, we estimate the propensity score of having a grocery store opening for each site using the same 24 local characteristics as input features to our CNN model. They are listed in Appendix Figure B1:

I. 18 Census Block Group level demographic characteristics

II. the number of business establishments belonging to 5 different categories

III. the number of existing grocery stores before opening

For each of these 24 characteristics, we calculate its average value for 10 concentric rings surrounding the real openings and candidate counterfactual sites. Each ring has a bandwidth of 0.025 miles. Therefore, we have 240 features in total to estimate the propensity score model.

We implement Principal Component Analysis (PCA) to reduce the dimensionality of the input features before we estimate the propensity scores. We choose the 30 most informative components from the 240 features using PCA. We then estimate a propensity score model with these 30 features and match each real opening to the top 10 candidate CF sites with the closest propensity scores. After the first round of matching, we further restrict CF sites to be at least 0.2 miles apart from each other and randomly drop one CF site if the distance between two CF sites is smaller than this threshold. Next, we do a second round of propensity score estimation for the pool of real openings and candidate CF sites remaining from round one. This time, we match each real opening with one CF site with the closest propensity score. These CF sites chosen in the second round form the control group for our empirical analysis.

Combining CNN and propensity matching to estimate counterfactual sites allows us to take advantage of both methods. A major challenge for us in finding potential counterfactual locations is the huge class imbalance between 1s (real openings) and 0s (potential counterfactual locations) in geographic space. We address the class imbalance in two steps. First, we over-sample real openings through data augmentation methods. Second, we under-sample potential counterfactual locations by restricting the donor pool to be 0.025 miles grid squares within 5 miles from real openings and then train a CNN model to search for credible potential counterfactual locations within this range. This enables us to greatly reduce the size of the pool of potential counterfactual locations. However, predictions (e.g., activation scores) from the CNN model depend on some model-tuning parameters, which are less relevant for predicting counterfactual locations of grocery openings. We thus perform an additional step of propensity score estimation and matching to supplement the CNN model and find the best counterfactual locations. In this last step, we can include only the most important covariates with flexible input formats (e.g., measures in concentric rings) to estimate the propensity score model and thus obtain counterfactual locations of the highest quality.

4.2 Main Specification for Treatment Effects of Grocery Store Openings on Surrounding Businesses

Figure B5 shows a map for the opening of Trader Joe's at 2101 W Imperial Hwy Ste A, La Habra, California. Panel (a) shows the grocery store's surrounding geography and its counterfactual opening location. Trader Joe's opened on July 18, 2019, and was the first Trader Joe's in the La Habra Area.[13] It replaced Vons, which closed on October 12, 2018.[14] The same location was vacant for nine months. Trader Joe's locates in a strip mall anchored by another tenant, CVS, as shown in Figure B5 Panel (b). A high level of density of nearby businesses also surrounds the location.

From the perspective of our predictive algorithm, the matched counterfactual opening location has a comparable level of business densities and variety nearby and similar demographic characteristics as the real opening location. The counterfactual site is located extremely close to a few strip malls near the intersection of Westminster Boulevard and Beach Boulevard, which currently house businesses like Walgreens, Chase Bank, restaurants, and coffee shops, as well as family medicine clinics, as shown in Figure B5 Panel (c). Moreover, the street connectivity for both sites is similar since both are accessible by state highways. Both sites are located near the intersections of major roads, which is a beneficial factor for the location of the grocery store. Both sites are within 500 meters of Beach Boulevard, a major road in the area that is part of California State Route 39.

We use a Trader Joe's opening to illustrate our identification strategy. Trader Joe's could have chosen the counterfactual site, which has a similar desirability to open its business. For idiosyncratic reasons to the owner of a particular grocery store, the real site is chosen over the counterfactual site. In other words, these idiosyncratic factors that matter for the choice of a particular grocery contribute to the quasi-random variation we leverage for identification. However, such idiosyncrasies could not be correlated with any remaining

[13] https://www.ocregister.com/2019/07/18/la-habra-opens-its-first-trader-joes-after-years-petitioning-for-a-store/

[14] https://www.yelp.com/biz/vons-la-habra?sort_by=date_desc

systematic unobservables that matter for grocery stores' location choices across our sample, that are not already accounted for in our predictive model. In absence of the grocery store entry, businesses surrounding both the real and counterfactual sites would evolve on similar trajectories. After the grocery store opens, it plays the role of an anchor for shopping malls; they draw customer traffic not only to themselves but also to nearby stores. Relative to businesses around the counterfactual site, we expect foot traffic and sales to businesses around the real opening to increase due to economies of agglomeration, rather than reasons related to local economic conditions alone. Based on this intuition, our identification strategy is to compare the outcomes of businesses that are at identical distances from the real and counterfactual opening sites, respectively.

A caveat of this strategy is that we can only capture the effects of grocery store openings on businesses at a certain distance from them due to externalities generated by economies of agglomeration. This may not be the total effect of the opening, which could also include some area-wide common GE effects. However, we difference out those effects by comparing with businesses that would have the same amount of exposure to the grocery store opening had it opened at the counterfactual site as a control group. Our empirical strategy thus allows us to identify the treatment effects due to local agglomeration and spillovers, without taking extra strong structural assumptions on the common GE effects.

To execute our identification strategy, for each grocery store opening, we define a case by pairing it with its matched counterfactual (CF) opening site. We divide the businesses surrounding the real and CF opening sites into successive concentric rings, respectively. We then compare outcomes of businesses located in the rings that are equally distant from the real and CF sites. The implicit assumption is that not only real and CF sites are directly comparable in their surrounding characteristics conducive to grocery store openings, but the businesses that are equidistant from the opening sites are also comparable, hence those near the CF sites would serve as a valid control group. This assumption is reasonable given that the effects of the openings of grocery stores are highly localized, as we will show later. We hence only examine businesses that are mostly within 0.1 miles from the real and CF sites, which ensures the surrounding environment would not differ very much.

To estimate the effects of the grocery store openings on the outcomes of neighboring businesses, we use the following event study design as the baseline specification of our analysis:

$$Y_{ijnt} = \sum_{\tau \in [-s_1, s_2]} \beta_{\tau,n} \text{Treat}_{ijn} \times D_{\tau,ijt} + \alpha_{ij} + \phi_{int} + \varepsilon_{ijnt}, \forall n \in 1, ..., N \tag{3}$$

Here, Y_{ijnt} is the outcome of the business of interest j associated with the case of grocery i's opening in time period t. For example, it could be the log of the monthly number of visits to it. The dummy variable Treat_{ijn} takes the value of one if business j associated with the opening case i is in the $n-$th ring next to the real opening and takes the value of zero if it is in the $n-$th ring next to the CF site matched. The variable $D_{\tau,ijt}$ denotes a dummy equal to one if the outcome Y_{ijnt} is observed in τ time periods relative to the opening of the grocery store i, where τ goes from s_1 periods before the opening to s_2 periods after the opening. α_{ij} denotes the opening case by business pair fixed effect that controls for time-invariant characteristics for each business j associated with opening case i. ϕ_{int} denotes the opening case by ring by time-period fixed effect (e.g., case by ring by calendar (year, month) fixed effect for examining the effects on monthly foot traffic). It allows for a flexible time trend for each local area that contains a pair of real and CF sites. Adding this term ensures that we only compare businesses near the real and CF sites within the same local area defined by a (case, ring) to remove potential composition biases of businesses across geographic locations.[15] We can run a separate regression comparing rings that are equidistant from the real and CF sites for successive rings $n \in 1, ..., N$. Our coefficient of interest, which quantifies the average effects of an opening on nearby businesses in each ring τ time periods relative to the opening, is denoted by $\beta_{\tau,n}$. We normalize the coefficient of the month prior to the opening, $\beta_{-1,n}$, to zero. Standard errors are clustered at the real or

[15]Note the term for Treat_{ijn} alone is collinear with the fixed effect α_{ij}, hence omitted in the regression. The term for $D_{\tau,ijt}$ alone is collinear with the local time trend ϕ_{int} and therefore omitted as well.

CF site level.

5 Empirical Results

5.1 Validation of Predictive Model and Propensity Scores for Treatment

We start our analysis by validating our estimated propensity score model to search for CF sites for grocery store openings. Figure 2 plots the distribution of the estimated propensity score of the real openings and the CF sites matched. We see that the two distributions overlap a lot with each other, suggesting a good matching quality. We also perform a balance test for key neighborhood-level demographic characteristics. Table 1 reports the demographic characteristics of the Census Block Groups that contain the openings of the grocery stores and the CF sites matched. All covariates in the treatment and control groups are perfectly balanced. Furthermore, Figure 3 shows that holding a fixed distance from a real opening or a counterfactual site, the business composition surrounding the real openings and CF sites is very similar. These tests demonstrate our model's ability to locate CF sites that highly resemble the real opening sites in their surrounding environment.

5.2 Validation of Opening Dates using Foot Traffic Measure

Next, we validate the opening dates of the grocery stores in our sample and make sure all the openings indeed happened. We examine whether there is a sharp increase in the foot traffic visiting the location of the opening before and after the month of each grocery store opening. We use an event study of the following specification.

$$Y_{it} = \alpha_i + \gamma_t + \sum_{\tau=-4}^{10} \mu_\tau D_{\tau,it} + \varepsilon_{it}. \tag{4}$$

Here, the outcome variable Y_{it} is the monthly number of visits at the Safegraph POI matched to a grocery store opening i. α_i denotes the fixed effects of the opening of the individual grocery store, and γ_t denotes the fixed effects of the calendar year by month. The variable $D_{\tau,it}$ denotes the set of relative event time dummies, equal to one if Y_{it} is observed in calendar month t is τ months relative to its opening. Our coefficient of interest, which quantifies the effects of the grocery store on foot traffic at the opening location, is denoted by μ_τ.

Figure 4 plots estimated effects, along with 95% confidence intervals. We can see that up to four months prior to the opening of the grocery store, there are no significant pre-trends in foot traffic at the opening location. In six to ten months after the opening, the average foot traffic increased by 588 visits, or 333 percent, relative to the pre-opening period. The sharp increase in foot traffic after openings and the lack of pre-trends prior to openings validate our method of using structure breaks in foot traffic to detect openings. It is also reassuring that there are generally a few months of vacancy between tenancies at the locations of the openings, so the treatment effects on surrounding businesses would not be confounded by the previous tenant at the opening location.

5.3 Treatment Effects on Foot Traffic to Surrounding Businesses

Now, we are ready to examine how foot traffic to nearby businesses is influenced by grocery store openings. As anchors of shopping malls, grocery stores can attract customers not only to themselves but also to nearby stores. Therefore, we expect foot traffic to nearby businesses would increase due to economies of agglomeration. In particular, foot traffic to nearby businesses could increase through increased visibility, reduced search cost, and trip chaining, leading to higher sales and profits as well. To measure such spillover effects, we implement the regression specification in equation (3) using the log of the monthly number of visits as an outcome. For the area surrounding each real opening and its CF site, respectively, we use a distance band of 0.025 miles to define concentric rings in which surrounding businesses are located. Our regression sample consists of a balanced panel of all businesses within 0.2

miles[16] of each real opening site and from each matched CF site, observed from 4 months before the openings to 10 months after the openings. We cluster standard errors at the real or CF site level. Figure 5 plots the average of the estimated coefficients, $\beta_{\tau,n}$, for each concentric ring in the first 10 months after the openings. The figure thus summarizes average treatment effects in percent terms on nearby businesses' foot traffic counts post grocery store openings. These spillover effects are highly localized in that they decay to be statistically not different from zero for businesses located further than 0.1 miles from the actual and CF sites. Within 0.1 miles, businesses surrounding the real openings have a significant increase in foot traffic, relative to those surrounding the CF sites. Due to the localized nature of spillovers, subsequently, we focus on our analysis on comparing outcomes of businesses within 0.1 miles.

We adapt equation (3) to define a single ring made of all businesses within 0.1 miles from either the real or CF opening sites. Figure 6a plots the estimated coefficients β_τ for the effects on log monthly foot traffic for $\tau \in \{-4, 10\}$ months relative to openings. We confirm that there are no differential trends in foot traffic to businesses in the treatment and control groups prior to the opening of the grocery store. We find that openings in grocery stores have economically significant and statistically significant positive effects on surrounding businesses. Average foot traffic in the treatment group increases by 39.1 percent 6–10 months after opening. As a robustness check, we also estimate the treatment effects using inverse propensity score weighting. For the businesses in the treatment group, the weights are inversely proportional to the estimated propensity scores for the opening sites from Section 4. For the control group, the weights are inversely proportional to one minus the estimated propensity scores for the CF sites. Figure 6b shows that the inverse propensity weighting estimators yield similar results on foot traffic spillovers. The average foot traffic in the treatment group increases by 37.3 percent 6–10 months after opening.

One concern for identifying the causal spillovers on foot traffic could be some developers

[16] The geographic range of 0.2 miles we use to define our treatment groups is supported by the maximum precision that Safegraph can identify a visit to a POI. Details of Safegraph algorithms that cluster GPS points to a single place can be found in their white paper: https://www.safegraph.com/guides/visit-attribution-white-paper.

and shopping center owners consider these spillover effects when endogenously determining property tenant mix at the same time as introducing the grocer anchor (Jardim, 2016). In particular, Publix has been increasing its ownership of shopping centers where it operates to control the neighboring tenant mix.[17] In this case, nearby foot traffic growth may not be due to the causal effect of the grocer anchors but rather because the co-tenants are endogenously determined. We perform a robustness check where we exclude the 30 openings of Publix from our sample. Figure B6 shows that the main results in Figure 5 and 6 remain robust.

5.3.1 Heterogeneous Effects on Foot Traffic by Type of Surrounding Businesses

Next, we examine the heterogeneity in demand spillovers by the categories of the surrounding businesses. We hypothesize that the externalities of the openings of grocery stores for neighboring businesses are stronger for certain business categories. We divide the surrounding businesses into six categories according to their NAICS codes: (1) Accommodation, Eating and Drinking (2) Finance, Real Estate, Communication, and Professional (3) Medical, Welfare, and Healthcare (4) Wholesale and Retail (excluding Grocery Stores) (5) Grocery Stores[18] (6) Other Services. Figure 3 shows the composition of businesses by category in our treatment and control groups.

To perform the heterogeneity analysis, we use the following specification:

$$Y_{ijnht} = \sum_{\tau \in [-s_1, s_2]} \beta_{\tau,n,h} \text{Treat}_{ijn} \times D_{\tau,ijt} + \alpha_{ij} + \phi_{inht} + \varepsilon_{ijnht} \tag{5}$$

Here, Y_{ijnht} is the outcome of the business of interest j in the heterogeneity group h associated with the case of grocery i's opening in time period t. The heterogeneity group refers to the business category here. The dummy variable Treat_{ijn} takes the value of one if business j associated with the opening case i is in the $n-$th ring next to the real opening

[17]Publix owns nearly a third of its 1,167 stores, according to the company's 2017 annual report.

[18]We define grocery stores as the businesses with 4-digit NAICS codes 445110 (Supermarkets and Other Grocery (except Convenience) Stores, e.g., ALDI and Publix Super Markets), 445120 (Convenience Stores, e.g., 7-Eleven), 452210 (Department Stores, e.g., Target), and 452319 (All Other General Merchandise Stores, e.g., Dollar General, Costco, and Walmart).

and takes the value of zero if it is in the $n-$th ring next to the CF site matched. The variable $D_{\tau,ijt}$ denotes a dummy equal to one if the outcome Y_{ijnht} is observed in τ time periods relative to the opening of the grocery store i, where τ goes from s_1 periods before the opening to s_2 periods after the opening. α_{ij} denotes the opening case by business pair fixed effect that controls for time-invariant characteristics for each business j associated with opening case i. ϕ_{inht} denotes the opening case by ring by business category by time-period fixed effect. Our coefficient of interest, which quantifies the average effects of an opening on nearby businesses in each ring in each business category τ time periods relative to the opening, is denoted by $\beta_{\tau,n,h}$. We normalize the coefficient of the month prior to the opening, $\beta_{-1,n,h}$, to zero. Standard errors are clustered at the real or CF opening site level.

Figure 7a summarizes the effects of openings on foot traffic to nearby businesses after opening. For example, the wholesale and retail stores (excluding grocery stores) show an average increase of 40 percent in foot traffic 6–10 months after a grocery store opening. Our results suggest that the synergies between grocery store openings and wholesale & retail stores as well as hospitality services are the strongest. This is expected since these types of businesses most likely share common demand from the same group of customers who visit the shopping center where the grocer anchor is located. Clustering around a grocery store anchor helps attract more customers. However, demand spillovers are smaller in medical, welfare, and healthcare, with a statistically significant difference of negative 31 percent in treatment effect compared to the effect on wholesale and retail stores (excluding grocery stores). These types of businesses may not share the same customer base as nearby grocery stores. People may seek out their designated or preferred dentists or primary care providers close to grocery stores but do not necessarily shop there. Further, the average shopper who does visit the nearby grocery store may not visit the healthcare services frequently. Overall, it leads to weaker synergies between grocery stores and healthcare services. Table A4 presents the treatment effects by industry when we further divide the surrounding businesses into finer industries based on 4-digit NAICS codes.

Interestingly, we find that grocery store openings in our sample on average do not pose a threat to other nearby grocery stores in diverting their existing customers' traffic. On average, they lead to a 4% increase in foot traffic to other existing grocery stores within 0.1 miles, which is economically small and not significant at the 95% level.[19] We further examine how the positive demand spillovers on existing grocers vary by the type of grocery store openings in section 5.3.2 below.

Furthermore, we also examine how spillover effects on foot traffic vary for businesses in the same commercial real estate property as grocery store openings and those not in the same real estate property. We split the sample into two sub-samples: one with all businesses within the same real estate property as the grocery store opening and all businesses within 0.1 miles of the CF site; another sub-sample with all other businesses within 0.1 miles of the grocery store opening, but not in the same real estate property, and all businesses within 0.1 miles of the CF site. We run a separate regression adapted from equation 3 on each sub-sample.[20] Figure 7b shows that wholesale and retail businesses gain the most with an increase of 70 percent in foot traffic if they are located within the same property as the grocery store opening 6–10 months after opening. In comparison, wholesale and retail businesses not in the same property have a much smaller increase in foot traffic of 30 percent. The pattern is broadly similar for other business categories. Businesses within the same property, hence most likely in the same shopping centers or strip malls, generally see a greater increase in foot traffic. It should be noted that the gap in spillovers on hospitality services is smaller, suggesting that spillovers are less localized to those within the same property for restaurants, bars, and hotels.

[19]The differential effect between nearby affected grocery stores and wholesale and retail (excluding grocery stores) is an economically large and significant 36%.

[20]We provide details on how we utilize the structure of Safegraph Plackey to find all POIs in the same property as the grocery openings in Appendix D.

5.3.2 Heterogeneous Effects on Foot Traffic by Type of Grocery Store Openings

We now examine the heterogeneous demand spillovers generated by the openings of different grocery stores. In Figure 8a, we divide the grocery store openings into 3 types: national grocery store chains (e.g. Safeway and Publix), big-box retailers (e.g., Walmart, Target, and Costco) & discounters (e.g., ALDI and LIDL), and Dollar stores (e.g., dollar general, dollar tree). We find that national grocery chains and big-box retailers & discounters mainly drive the positive demand spillovers generated by grocery store openings. In general, 6–10 months after the opening of a national grocery store chain, foot traffic to other businesses within 0.1 miles increases by 52 percent. At the same time, opening a big-box retailer or discounter increases foot traffic to nearby businesses by 31 percent. In contrast, opening a dollar store only increases foot traffic to nearby businesses by 7 percent which is not statistically significant. In Figure 9a, we explore why national grocery store chains have the largest spillovers. One potential explanation is national chains such as Whole Foods Market, Trader Joe's, and Harris Teeter are expanding fast mostly in higher-income areas where the customer base has strong purchasing power that leads to more correlated demand with nearby retailers. We rank the Census Block Groups (CBGs) where the grocery store openings are located in terms of their median household income. Grocery store openings in higher-income neighborhoods indeed have larger spillovers. Figure 9b shows the share of openings by national chains is larger in CBGs in higher-income quartiles. Hence more openings that belong to national chains happen in higher-income neighborhoods, which leads to larger spillovers on nearby businesses.

In Figure 7a, we do not find evidence of negative competitive effects of grocery store openings on nearby existing grocers. However, that could mask significant heterogeneity in the effects across different types of openings and existing grocery stores. For example, policymakers and local communities have been concerned about the potential costs of big-box retailer expansion on local economic development and small businesses and costly local government subsidies when weighted against their economic benefits (Mattera and Purinton, 2004). The literature has found evidence that Walmart's entries damage competitors in the discount store industry (Jia, 2008) and the grocery outlets of larger national chains (Ellickson and Grieco, 2013), decrease local retail and wholesale employment (Basker, 2005), undermine retail wages due to part-time low wage jobs with little healthcare and pension benefits (Mattera and Purinton, 2004).

To examine heterogeneous effects on different types of grocery store entries on existing grocers, we examine the heterogeneous treatment effects by the types of opening grocery stores and surrounding grocery stores. We divide the openings of grocery stores into 3 types: national grocery store chains (e.g. Safeway and Publix), big-box retailers (e.g. Walmart, Target, and Costco), and discounters (e.g. ALDI and LIDL). We divide the surrounding grocery stores into 4 types: national grocery stores, big-box and discounters (which include both big-box retailers and discounters), dollar stores, and convenience and independent stores. We present results in Figure 8b. Consistent with Ellickson and Grieco (2013), we find suggestive evidence that big-box retailers and discounters compete against national grocery store chains, but they do not seem to harm other types of grocers. On the one hand, the opening of a big-box retailer nearby reduces foot traffic to a grocery store belonging to a national chain by an economically large 27% 6–10 months later, though not significant at the 95% level. On the other hand, 6–10 months after the opening of a grocery store under a national chain, the foot traffic to nearby big-box retailers and discounters decreases by 17%.

In Figure 10, we instead classify the opening of grocery stores and surrounding businesses into terciles based on the income bucket of the median customer in 2019.[21] We find that the openings of all types of grocery stores generate positive demand spillovers for surrounding businesses with low- and middle-income customers. In contrast, for surrounding businesses with high-income customers, the openings of grocery stores with middle- and high-income

[21] Safegraph spending data divides the income of customers into 7 buckets: <25k, 25-45k, 45-60k, 60-75k, 75-100k, 100-150k, >150k. For each brand, given the number of customers N_i in the income bucket i in 2019, we define the income bucket of the median customer as $\min\{1 \leq i \leq 7 | \sum_{j=1}^{i} N_j \geq \frac{\sum_{j=1}^{7} N_j}{2}\}$. We then divide all brands into terciles based on the income bucket of the median customer. Note that we treat businesses without brands as individual brands.

customers have a positive impact, while the openings with low-income customers have a negative impact. This may be because the openings of grocery stores that serve lower-income groups leave high-income customers with a negative impression of the quality of the surrounding businesses.

5.3.3 Heterogeneous Effects on Foot Traffic by Population Density of Opening Locations

We further examine the heterogeneous demand spillovers in high versus low population density areas. We divide grocery store openings into high and low-population-density groups according to the population density of the counties where they are located.[22] The results of the difference-in-difference with multiple rings in high-population-density areas and low-population-density areas are presented in Figure 11 separately. Panel (a) suggests that demand spillover effects are clustered within 0.1 miles in high-population-density areas, and Panel (b) suggests that demand spillovers extend to a range of 0.2 miles in low-population-density areas. The spillovers are less localized in low-population density areas as people are more likely to travel further for their shopping trips to consume amenities. In Figure 11, we also present the event study results of grocery store openings on other businesses in affected areas in terms of foot traffic respectively. Panel (c) shows the event study results in high-population-density areas. Treatment groups are all businesses within 0.1 miles from the real openings in high-population-density areas, and control groups are all businesses within 0.1 miles from the corresponding CF sites. Panel (d) shows the event study results in low-population-density areas. Treatment groups are all businesses within 0.2 miles from the real openings in low-population-density areas, and control groups are all businesses within 0.2 miles from the corresponding CF sites. Both openings of grocery stores in high and low-population-density areas increase foot traffic to businesses within 0.1 miles by up to

[22] County-level population density data comes from the 2014-2018 American Community Survey. Openings in counties with a population density above the median are classified into the high-population-density group, and the rest are classified into the low-population-density group.

40% in 6–10 months after openings, relative to the control group. However, the spillovers are less localized and extend to 0.2 miles for low-population-density areas relative to high-population-density areas.

5.4 Treatment Effects on Revenue Spillovers to Surrounding Businesses

In this section, we use data on spending at businesses to estimate the effects of grocery store openings on revenue spillovers to surrounding businesses. So far, we have used foot traffic as a measure of demand spillovers. It has the advantage of measuring the effects on potential demand that businesses can realize. Spending on the other hand allows us to measure the effects on realized demand. Both measures are informative about the degree of spillovers brought by grocer anchors.

We adapt equation (3) to estimate the treatment effects on spending in surrounding businesses. The treatment group consists of all the businesses within the same property as and are 0–0.1 miles from the actual grocery store openings. The control group consists of businesses within 0.1 miles of the counterfactual locations for openings.

Figure 12 plots the estimated coefficients β_τ for the effects on log monthly spending for $\tau \in \{-4, 10\}$ months relative to openings. We confirm that there are no differential trends in spending in businesses in the treatment and control groups prior to the opening of the grocery store. We find that openings in grocery stores have economically significant and statistically significant positive effects on surrounding businesses. Average spending in the treatment group increases by 47.6 percent 6–10 months after opening.

We now examine the heterogeneous demand spillovers generated by the openings of different grocery stores in different surrounding business categories. In 13a, we break down the surrounding businesses into 3 categories, wholesale and retail excluding grocery, grocery, and hospitality. We find that the wholesale and retail stores (excluding grocery stores) show the greatest average increase of 41 percent in spending 6–10 months after a grocery store opening. In Figure 13b, we divide the grocery store openings into 2 types, national grocery store chains and big-box retailers & discounters. We find that the openings of grocery stores belonging to national chains bring the most spending increase of 69% to surrounding businesses.

5.5 Treatment Effects on Surrounding Business Dynamics

In this section, we study how these grocery store openings change the dynamics in the surrounding business environment. In particular, we examine how these grocery store openings lead to openings of new businesses nearby, closure of incumbent businesses, expansions, and contractions of surrounding businesses in terms of their size. Openings and closures of businesses capture the extensive margin of the impact of grocery stores on surrounding business dynamics, while expansions and contractions in size capture the intensive margin of business dynamics.

For the business i in the time period t, we define its outcomes, such as active status[23], its size measured by employment, as E_{it}. Following Ellickson and Grieco (2013), we can calculate the corresponding growth rates as follows:

$$\begin{aligned} g_{it} &= \frac{(E_{it} - E_{it-1})}{X_{it}} \\ X_{it} &= \frac{(E_{it} + E_{it-1})}{2} \end{aligned} \tag{6}$$

Here, X_{it} is the average of outcomes of the business i in periods t and $t-1$. g_{it} is the individual growth rate of the business i from period $t-1$ to t. Note that g_{it} takes the value of 2 if the business i opens and takes the value of -2 if it closes in the period t. We can then aggregate the growth rates at the individual business level to the group level according to the following formulas:

[23] Active status takes the value of 0 before opening and after closing and takes the value of 1 when the business is operational.

$$
\begin{aligned}
JCC_{kt} &= \frac{\sum_{i\in\mathcal{R}_k} X_{it} \max\{g_{it}, 0\} \mathbf{1}[g_{it} < 2]}{\sum_{i\in\mathcal{R}_k} X_{it}} \\
JCE_{kt} &= \frac{\sum_{i\in\mathcal{R}_k} X_{it} * 2 * \mathbf{1}[g_{it} = 2]}{\sum_{i\in\mathcal{R}_k} X_{it}} \\
JDC_{kt} &= \frac{\sum_{i\in\mathcal{R}_k} X_{it} \max\{-g_{it}, 0\} \mathbf{1}[g_{it} > -2]}{\sum_{i\in\mathcal{R}_k} X_{it}} \\
JDE_{kt} &= \frac{\sum_{i\in\mathcal{R}_k} X_{it} * 2 * \mathbf{1}[g_{it} = -2]}{\sum_{i\in\mathcal{R}_k} X_{it}}
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
G_{kt} &= \frac{\sum_{i\in\mathcal{R}_k} X_{it} g_{it}}{\sum_{i\in\mathcal{R}_k} X_{it}} = \frac{\sum_{i\in\mathcal{R}_k} (E_{it} - E_{it-1})}{\sum_{i\in\mathcal{R}_k} X_{it}} \\
&= JCC_{kt} + JCE_{kt} - JDC_{kt} - JDE_{kt}
\end{aligned}
\tag{8}
$$

Here, G_{kt} is the overall growth rate of the corresponding outcome (e.g., employment) in the group k, which can be decomposed into positive contributions from openings (JCE) and expansion of existing businesses (JCC), and negative contributions from closings (JDE) and contraction of existing businesses (JDC). Note that G_{kt} could be interpreted as the growth rate of the number of businesses in the group when E_{it} denotes the active status of businesses. In this case, JCE and JDE can be interpreted as the share of the number of openings and closures relative to the number of existing businesses, respectively.

In our setting, the group k is determined by the n-th ring next to the real or CF site j associated with the case of grocery i's opening. We use a distance band of 0.1 miles to define concentric rings in which surrounding businesses are located. Our regression sample consists of a panel of all businesses within 0.5 miles of each real opening site and from each matched CF site, observed from 3 years before the openings to 3 years after the openings. To estimate the effects of the grocery store openings on surrounding business dynamics, we use the following event study design as the baseline specification of our analysis:

$$
Y_{ijnt} = \sum_{\tau\in[-s_1,s_2]} \beta_{\tau,n} \text{Treat}_{ijn} \times D_{\tau,ijt} + \alpha_{inj} + \phi_{int} + \varepsilon_{ijnt}, \forall n \in 1, ..., N \tag{9}
$$

Here, Y_{ijnt} is the outcome of the n-th ring next to the real or CF site j associated with the case of grocery i's opening in time period t. For example, it could be the growth of the number of businesses in the ring. The dummy variable Treat_{ijn} takes the value of one if site j is a real opening site and takes the value of zero if it is a CF site. The variable $D_{\tau,ijt}$ denotes a dummy equal to one if the outcome Y_{ijnt} is observed in τ time periods relative to the opening of the grocery store i, where τ goes from 3 years before the opening to 3 years after the opening. α_{inj} denotes the opening case by ring by site fixed effect that controls for time-invariant characteristics of the n-th ring next to the real or CF site j associated with opening case i. ϕ_{int} denotes the opening case by ring by calendar year fixed effect. Our coefficient of interest, which quantifies the average effects of an opening on nearby businesses in each ring τ time periods relative to the opening, is denoted by $\beta_{\tau,n}$. We normalize the coefficient of the month prior to the opening, $\beta_{-1,n}$, to zero. Standard errors are clustered at the real or CF opening site level.

Figure 14 plots the average treatment effects of grocery store openings on the growth rate of the number of businesses within 0.5 miles 0-3 years later in Panel (a) and the growth rate of the employment by businesses within 0.5 miles in Panel (b). Businesses grow faster in total number and size within 0.1 miles of the grocery store openings, relative to the control group. In contrast, businesses within 0.2–0.4 miles have slower growth rates. The scale economies introduced by these grocery anchors lead to more business entries and faster growth in business size closer to their locations and potentially draw business growth further away from them.

In Figure 15, we zoom in on the effects of grocery store openings on business growth within 0.1 miles where the positive spillovers are most pronounced, and plot the event study coefficients from estimating equation 9. Panel (a) shows that the growth rate of businesses within 0.1 miles of the openings peaks 1 year later and is 8.3 percentage points higher than those surrounding the CF sites. Panel (c) shows that the growth rate of employment for businesses within 0.1 miles of the openings peaks 1 year later and is 17.2 percentage points higher than those surrounding the CF sites.[24] The results are robust to using inverse propensity score weighting in Panels (b) and (d). The positive effects on business growth rates quickly decline beyond 1 year after the openings. However, the cumulative effects on business growth within the first 3 years after openings are still substantial and positive. To see this, we examine the effects on the arcsinh of the number of businesses and employment within 0.1 miles in Figure B7.[25]. Panel (a) shows that the number of businesses within 0.1 miles of the openings is 44.6 percentage points higher than those surrounding the CF sites 3 years after the opening. Panel (c) shows that the size of employment within 0.1 miles of the openings is 78.9 percentage points higher than those surrounding the CF sites.

Figure 16 further decomposes the effects on growth rates into contributions from openings, closures, expansion, and contraction of incumbent businesses. Panel (a) shows that the new business openings within 0.1 miles of the grocer opening increase by 4 percentage points in 0–3 years after opening, relative to the same range surrounding the CF sites, whereas the closures of existing businesses slow down by 3 percentage points. Panel (b) shows the overall growth in firm size surrounding the grocery store openings is driven by the higher growth of new businesses (6 percentage points higher) and fewer business closures (4 percentage points lower), relative to the control group.

In Figure 17, we examine the heterogeneous effects on business growth by the category of surrounding businesses and the type of grocery store openings.[26] We also separate the effects contributed by openings and closures respectively. Hence the overall effects on growth equal the difference between the two. Panel (a) and (c) highlight that the effects on business growth in number and size are the largest for wholesale and retail. Panel (b) and (d) show

[24] We can also measure the size of businesses using sales revenue from Reference USA, and the effects on growth rate are similar to those on employment. Results are available upon request.

[25] The arcsinh or inverse hyperbolic sine is defined as $\operatorname{arsinh} x = \ln\left(x + \sqrt{x^2+1}\right)$. The function is approximately equivalent to the log when x is large, but its advantage over the log is that it can allow x to be 0.

[26] For heterogeneity by business category, we construct the growth rate by determining a group k in Equations 7 & 8 by a specific business category in a specific ring next to a site. We then adapt Equation 9 to run a separate regression for each category in the rings within 0.1 miles from the sites, which is conceptually equivalent to the specification of Equation 5.

that Dollar stores have the biggest effects on the growth of businesses nearby.

Figure 18 plots treatment effects by 4-digit NAICS code on surrounding business growth and contributions from openings and closures of nearby businesses within 0–0.1 miles from the grocery store openings in the year of opening and the first year after opening. We present the industries with significantly positive effects at the 95% level and a pre-existing share of the industry among all surrounding businesses ranking in the top 2/3. Panel (a) shows the treatment effects on the growth of the number of businesses. Among the top industries that experience the highest growth at the extensive margin shortly after the grocery store openings, "supermarkets and grocery stores", "furniture stores", and "consumer goods rental" belong to categories such as "Wholesale and Retail" that also experience a substantial increase in foot traffic as we saw earlier. However, businesses in "independent artists, writers, performers", "advertising, public relations, and related services", and some healthcare services experience large growth in number, but they did not have a substantial increase in foot traffic. This is likely because even though there are more entries and fewer closures, the average customer does not visit these types of services very often, hence the effect at the extensive margin dominates. Panel (b) shows the treatment effects on the growth of employment, mostly due to growth at the extensive margin as we see in Figure 16. The patterns are consistent with those for growth in number. Industries related to healthcare experience some of the largest growth in employment. In addition, "drinking places", "supermarkets and grocery stores", and "depository credit intermediation" (banks) belong to business categories that experience a large increase in foot traffic. They also experience some of the largest growth in employment.

Overall, results in this section highlight that grocery store openings also bring substantial business growth at the extensive margin, in addition to generating demand spillovers on incumbent businesses. These results help us gain additional insight into which types of businesses have the largest growth at the extensive margin. Some industries had higher growth in number and size, despite not having a substantial increase in demand measured by foot traffic and spending.

6 Conclusion

We investigate the driving forces and effects of agglomeration in the local non-tradable service sector. We use anchoring grocery store openings in the U.S. in 2018 and 2019 to study the impacts of grocery store openings on nearby affected businesses. We borrow tools from deep learning and combine them with propensity score estimation techniques to find credible counterfactual opening sites. We then compare the outcomes of businesses surrounding the actual and these counterfactual sites. We find that the openings of grocery anchors generate substantial positive demand spillovers to nearby businesses. Such demand spillovers are concentrated within 0.1 miles of the openings. On average, a grocery store opening increases the foot traffic to nearby businesses within 0.1 miles by 39 percent 6 to 10 months later. Such strong positive spillovers are strongest between grocery store openings and wholesale and retail stores, with a 40 percent increase in foot traffic 6–10 months later. We also find that grocery store openings encourage local business growth mainly at the external margin. Within 3 years after a grocery store opens, there is a 6.9 percentage point higher growth in the number of businesses in 0.1 miles surrounding the openings.

References

Agrawal, Ajay and Iain Cockburn, "The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems," *International Journal of Industrial Organization*, November 2003, *21* (9), 1227–1253.

Arcidiacono, Peter, Patrick Bayer, Jason R. Blevins, and Paul B. Ellickson, "Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition," *The Review of Economic Studies*, July 2016, *83* (3), 889–931.

Basker, Emek, "Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion," *The Review of Economics and Statistics*, 2005, *87* (1), 174–183.

Benjamin, John D., Glenn W. Boyle, and C. F. Sirmans, "Price discrimination in shopping center leases," *Journal of Urban Economics*, November 1992, *32* (3), 299–317.

Benmelech, Efraim, Nittai Bergman, Anna Milanez, and Vladimir Mukharlyamov, "The Agglomeration of Bankruptcy," *The Review of Financial Studies*, July 2019, *32* (7), 2541–2586.

Bernstein, Shai, Emanuele Colonnelli, Xavier Giroud, and Benjamin Iverson, "Bankruptcy Spillovers," *Journal of Financial Economics*, September 2019, *133* (3), 608–633.

Bertrand, Marianne and Francis Kramarz, "Does Entry Regulation Hinder Job Creation? Evidence from the French Retail Industry*," *The Quarterly Journal of Economics*, November 2002, *117* (4), 1369–1413.

Butts, Kyle, "Difference-in-Differences Estimation with Spatial Spillovers," November 2021. arXiv:2105.03737 [econ].

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker, "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings," *American Economic Review*, February 2015, *105* (2), 678–709.

Diamond, Rebecca and Tim McQuade, "Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Property Development," *Journal of Political Economy*, June 2019, *127* (3), 1063–1117. Publisher: The University of Chicago Press.

Diao, Mi, Delon Leonard, and Tien Foo Sing, "Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values," *Regional Science and Urban Economics*, November 2017, *67*, 64–77.

Dixit, Avinash K and Joseph E Stiglitz, "Monopolistic competition and optimum product diversity," *The American economic review*, 1977, *67* (3), 297–308.

Duranton, Gilles and Diego Puga, "Chapter 48 - Micro-Foundations of Urban Agglomeration Economies," in J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4 of *Cities and Geography*, Elsevier, January 2004, pp. 2063–2117.

Ellickson, Paul B. and Paul L. E. Grieco, "Wal-Mart and the geography of grocery retailing," *Journal of Urban Economics*, May 2013, *75*, 1–14.

Ellison, Glenn and Edward L. Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, October 1997, *105* (5), 889–927.

_ , _ , and William R. Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, June 2010, *100* (3), 1195–1213.

Fung, Esther, "Strip Centers Shine as Some Shoppers Sour on Malls," *Wall Street Journal*, January 2020.

Gatzlaff, Dean, Stacy Sirmans, and Barry Diskin, "The Effect of Anchor Tenant Loss on Shopping Center Rents," *Journal of Real Estate Research*, January 1994, *9* (1), 99–110. Publisher: Routledge _eprint: https://doi.org/10.1080/10835547.1994.12090738.

Ghosh, Utpala, "How to choose a perfect location for a grocery store? A 101 guide," Apr 2022.

Glaeser, Edward L. and Joshua D. Gottlieb, "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, December 2009, *47* (4), 983–1028.

_ , Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer, "Growth in Cities," *Journal of Political Economy*, 1992, *100* (6), 1126–1152.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial networks," *Communications of the ACM*, 2020, *63* (11), 139–144.

Gould, Eric D., B. Peter Pashigian, and Canice J. Prendergast, "Contracts, Externalities, and Incentives in Shopping Malls," *The Review of Economics and Statistics*, 2005, *87* (3), 411–422. Publisher: The MIT Press.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Gupta, Arpit, Stijn Van Nieuwerburgh, and Constantine Kontokosta, "Take the Q Train: Value Capture of Public Infrastructure Projects," *Journal of Urban Economics*, 2020, p. 50.

Henderson, J.Vernon, "Marshall's Scale Economies," *Journal of Urban Economics*, January 2003, *53* (1), 1–28.

Henderson, Vernon, Ari Kuncoro, and Matt Turner, "Industrial Development in Cities," *Journal of Political Economy*, October 1995, *103* (5), 1067–1090.

Jardim, Eduardo, "Essays on the Dynamic Decisions of Homeowners and Retail," *Ph.D. Dissertation Essays*, 2016.

Jia, Panle, "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Knight, Samsun, "Retail Demand Interdependence and Chain Store Closures," October 2022.

Konishi, Hideo and Michael T. Sandfort, "Anchor stores," *Journal of Urban Economics*, May 2003, *53* (3), 413–435.

Krishna, Priya, "Why Do American Grocery Stores Still Have an Ethnic Aisle?," Aug 2021.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, 2017, *60* (6), 84–90.

Krugman, Paul R, "Increasing returns, monopolistic competition, and international trade," *Journal of international Economics*, 1979, *9* (4), 469–479.

Kuiper, Niels, Mark Van Duijn, and Arno Van der Vlist, "Retail Externalities and Distance in Shopping Malls," *ERES*, January 2021. Number: eres2021_205 Publisher: European Real Estate Society (ERES).

Leung, Justin and Zhonglin Li, "Big-Box Store Expansion and Consumer Welfare," *SSRN Electronic Journal*, 2021.

Liu, Crocker and Peng Liu, "Is What's Bad for the Goose (Tenant), Bad for the Gander (Landlord)? A Retail Real Estate Perspective," *Journal of Real Estate Research*, January 2013, *35* (3), 249–282. Publisher: Routledge _eprint: https://doi.org/10.1080/10835547.2013.12091369.

Liu, Crocker H., Stuart S. Rosenthal, and William C. Strange, "The vertical city: Rent gradients, spatial structure, and agglomeration economies," *Journal of Urban Economics*, July 2018, *106*, 101–122.

_ , _ , and _ , "Employment density and agglomeration economies in tall buildings," *Regional Science and Urban Economics*, September 2020, *84*, 103555.

Mattera, Philip B. and Anna Purinton, "Shopping for Subsidies: How Wal-Mart Uses Taxpayer Money to Finance Its Never-Ending Growth," *Good Jobs First*, 2004.

Merriman, David, Joseph Persky, Julie Davis, and Ron Baiman, "The Impact of an Urban WalMart Store on Area Businesses: The Chicago Case," *Economic Development Quarterly*, November 2012, *26* (4), 321–333.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen Redding, "Consumption Access and the Spatial Concentration of Economic Activity: Evidence from Smartphone Data," Technical Report w28497, National Bureau of Economic Research, Cambridge, MA February 2021.

_ , _ , and _ , "Consumption Access and the Spatial Concentration of Economic Activity: Evidence from Smartphone Data," Technical Report w28497, National Bureau of Economic Research, Cambridge, MA February 2021.

Moretti, Enrico, "Local Labor Markets," Working Paper 15947, National Bureau of Economic Research April 2010.

Neumark, David, Junfu Zhang, and Stephen Ciccarella, "The Effects of Wal-Mart on Local Labor Markets," *Journal of Urban Economics*, March 2008, *63* (2), 405–430.

Oh, Ryungha and Jaeeun Seo, "OhSeo_2022April.pdf," 2022.

Pashigian, B. Peter and Eric D. Gould, "Internalizing Externalities: The Pricing of Space in Shopping Malls," *The Journal of Law and Economics*, April 1998, *41* (1), 115–142.

Pollmann, Michael, "Causal Inference for Spatial Treatments," *arXiv:2011.00373 [econ, stat]*, October 2020.

Qian, Franklin and Rose Tan, "The Effects of High-skilled Firm Entry on Incumbent Residents," December 2021.

Relihan, Lindsay, "Is Online Retail Killing Coffee Shops? Estimating the Winners and Losers of Online Retail Using Customer Transaction Microdata," March 2022.

Rosenthal, Stuart and Joaquin Urrego, "Eyes on the Street, Spatial Concentration of Retail Activity and Crime," December 2021.

Rosenthal, Stuart S. and William C. Strange, "Geography, Industrial Organization, and Agglomeration," *The Review of Economics and Statistics*, 2003, *85* (2), 377–393.

_ and _ , "How Close Is Close? The Spatial Reach of Agglomeration Economies," *Journal of Economic Perspectives*, August 2020, *34* (3), 27–49.

Sadun, Raffaella, "Does Planning Regulation Protect Independent Retailers?," *The Review of Economics and Statistics*, December 2015, *97* (5), 983–1001.

Shoag, Daniel and Stan Veuger, "Shops and the City: Evidence on Local Externalities and Local Government Policy from Big-Box Bankruptcies," *The Review of Economics and Statistics*, July 2018, *100* (3), 440–453.

Su, Yichen, "Measuring the value of urban consumption amenities: A time-use approach," *Journal of Urban Economics*, 2022, *132*, 103495.

Waters, Shari, "Choosing a Retail Store Location," Mar 2021.

Wheaton, William C., "Percentage Rent in Retail Leasing: The Alignment of Landlord–Tenant Interests," *Real Estate Economics*, 2000, *28* (2), 185–204.

Wolinsky, Asher, "Retail Trade Concentration Due to Consumers' Imperfect Information," *The Bell Journal of Economics*, 1983, *14* (1), 275–282.

Zhou, Tingyu and John M. Clapp, "The location of new anchor stores within metropolitan areas," *Regional Science and Urban Economics*, January 2015, *50*, 87–107.

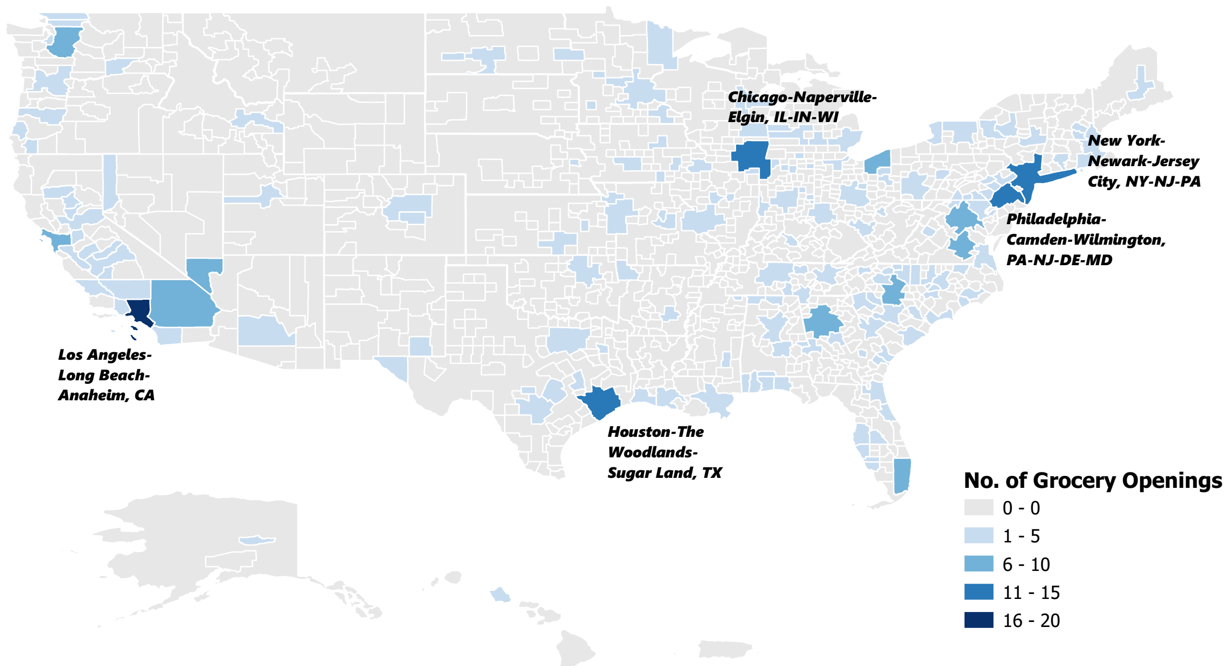


Figure 1: Locations of Grocery Store Openings

Notes: This figure plots the distribution of grocery store openings by CBSA in our sample. There are a total of 413 grocery store openings across the U.S. in 2018 and 2019 in our sample.

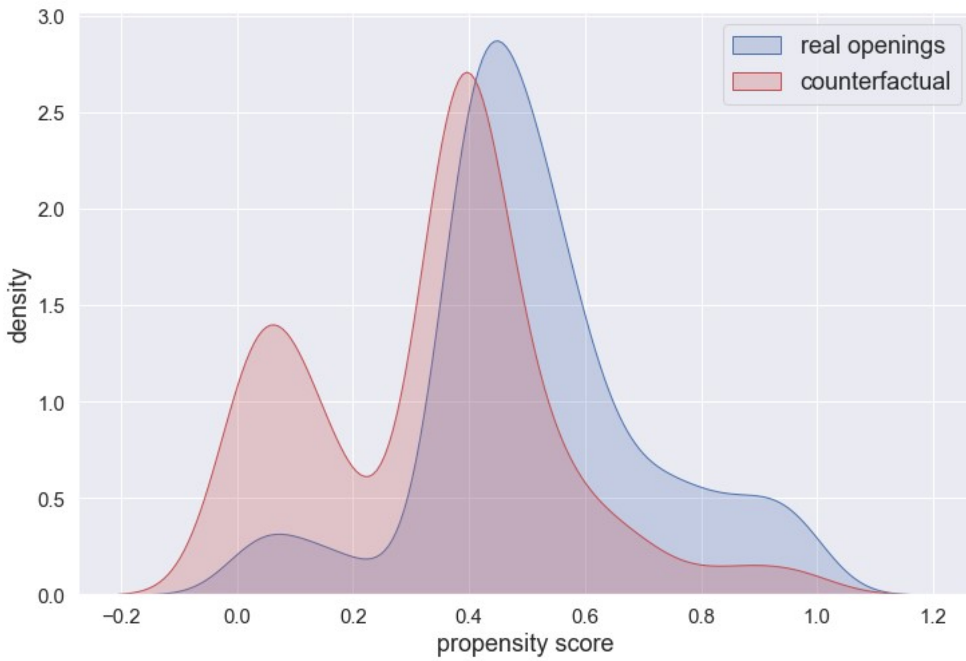
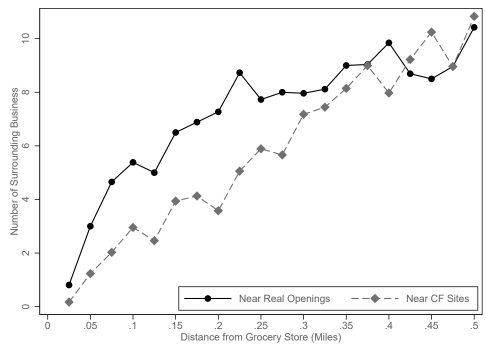
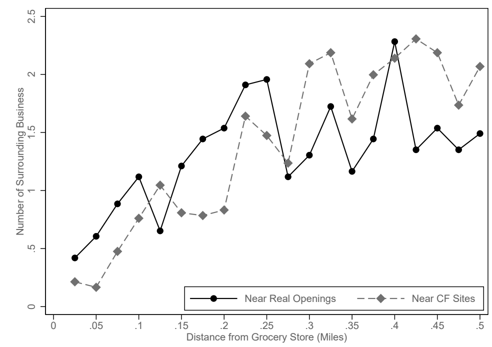


Figure 2: Distribution of Propensity Scores for Real Openings and for Matched CF Sites

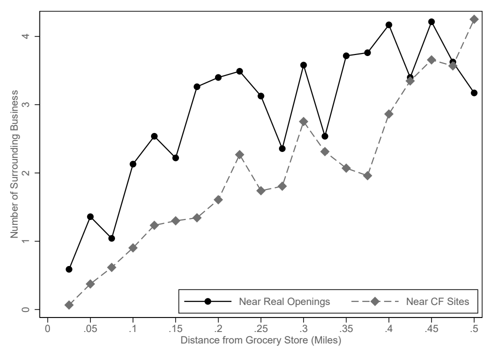
Notes: This figure plots the distribution of estimated propensity scores for real openings and for counterfactual sites matched.



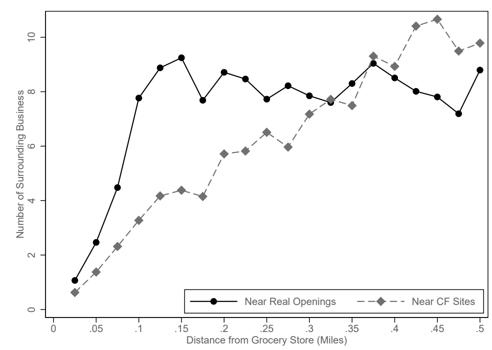
(a) Wholesale and Retail (excluding Grocery)



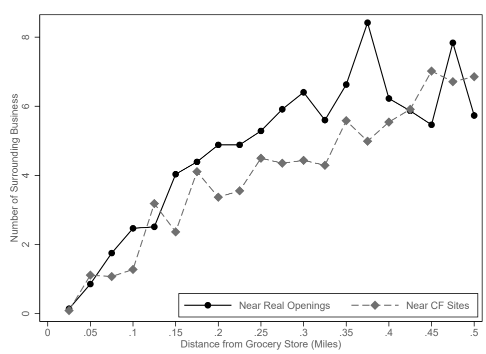
(b) Grocery



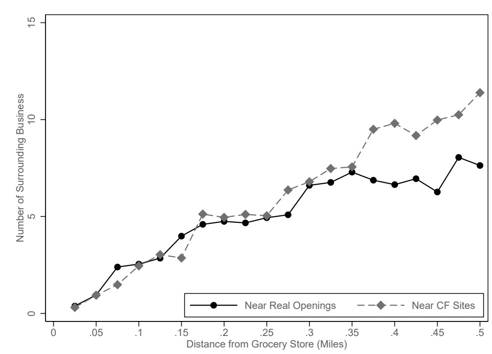
(c) Finance, Real estate, Communication, and Professional



(d) Accommodations, Eating, and Drinking



(e) Medical, Welfare, and Healthcare



(f) Other Services

Figure 3: Number of Businesses Surrounding Real Openings and Counterfactual Sites

Notes: We plot the average number of businesses by category by distance in a half-mile radius from real openings and CF sites, respectively. We follow Miyauchi et al. (2021b) to categorize surrounding businesses into 6 categories: wholesale and retail (excluding grocery) are identified by 2-digit NAICS codes 42, 44, 45 and exclude NAICS code corresponding to grocery stores; grocery stores are identified by 6-digit NAICS codes 445110, 445120, 452210, 452319; accommodations, eating, and drinking are identified by 2-digit NAICS code 72; finance, real estate, communication, and professional are identified by 2-digit NAICS code 52, 53, 54, 55, 56; medical, welfare, and healthcare are identified by 2-digit NAICS code 62; and other services include 2-digit NAICS code 51, 61, 71, 81, 92.

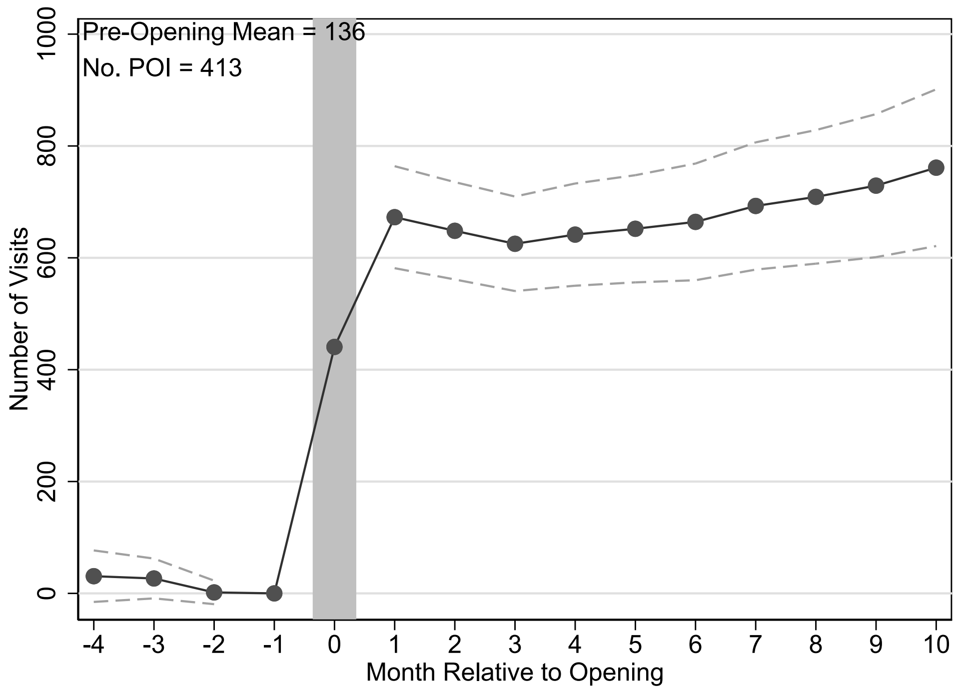


Figure 4: Validation of Opening Dates using Foot Traffic Measure

Notes: Sample consists of all grocery stores opening in our sample. This figure plots the estimated coefficients of μ_τ of equation (4), along with 95% confidence intervals. μ_τ captures the effects of the grocery store opening on the foot traffic at that location. The outcome variable is the monthly number of visits. The average number of monthly visits in the four months prior to the opening is 136. The standard errors are clustered at the grocery store level.

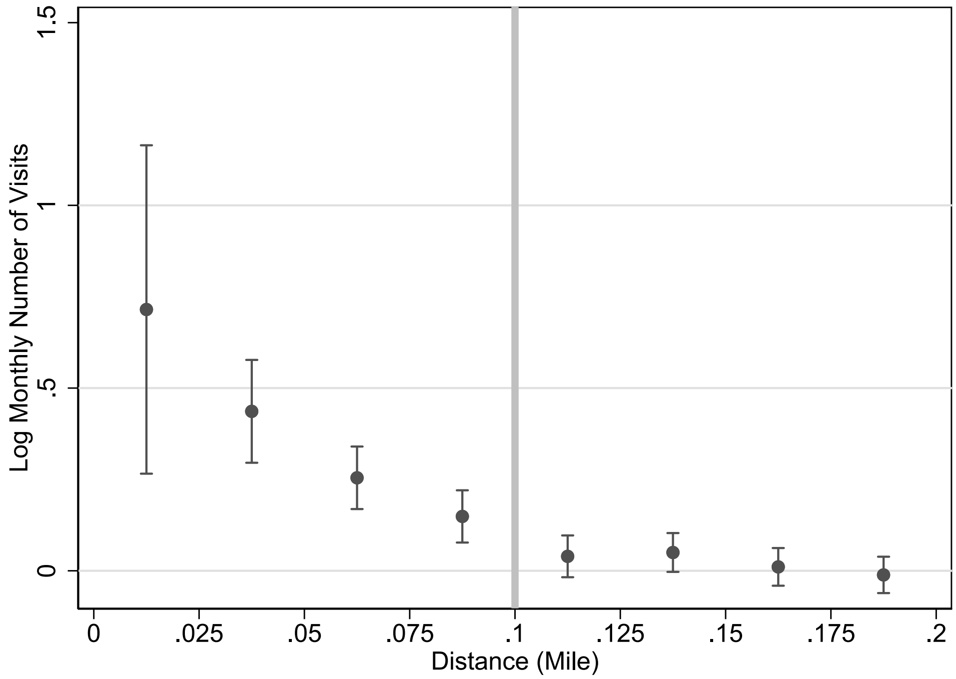
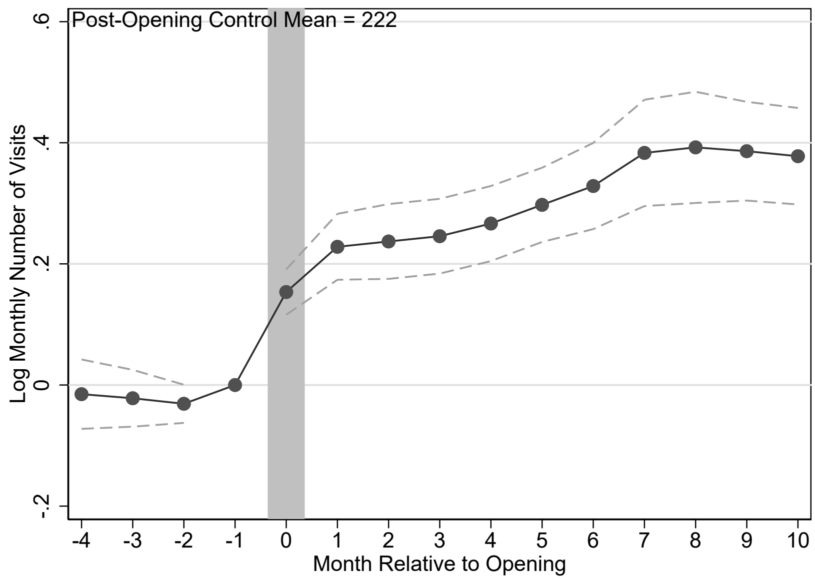
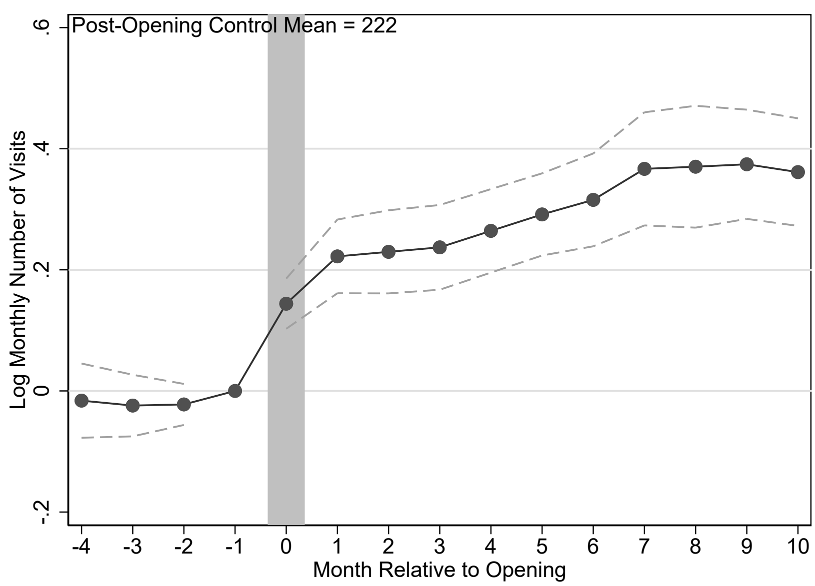


Figure 5: Treatment Effects on Foot Traffic to Surrounding Businesses as a Function of Distance

Notes: This figure plots the average treatment effects of grocery store openings on nearby businesses in each concentric ring in the first 11 months from the openings relative to the last 4 months before the openings by estimating equation (3). 95% confidence intervals are shown along with point estimates. The regression sample consists of a panel of all businesses within 0.2 miles from each of the grocery store opening sites and its matched CF site, observed from 4 months prior to the openings to 10 months after the openings. For the area surrounding each real opening site and CF site, respectively, we use a distance band of 0.025 miles to define concentric rings in which surrounding businesses are located for a distance up to 0.2 miles from each site. Standard errors are clustered at the real or CF site level.



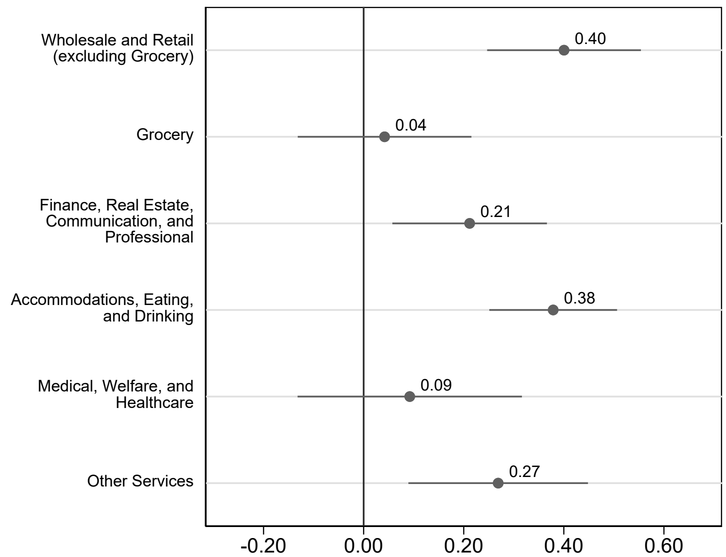
(a) Event Study Results on Surrounding POIs: No Weights



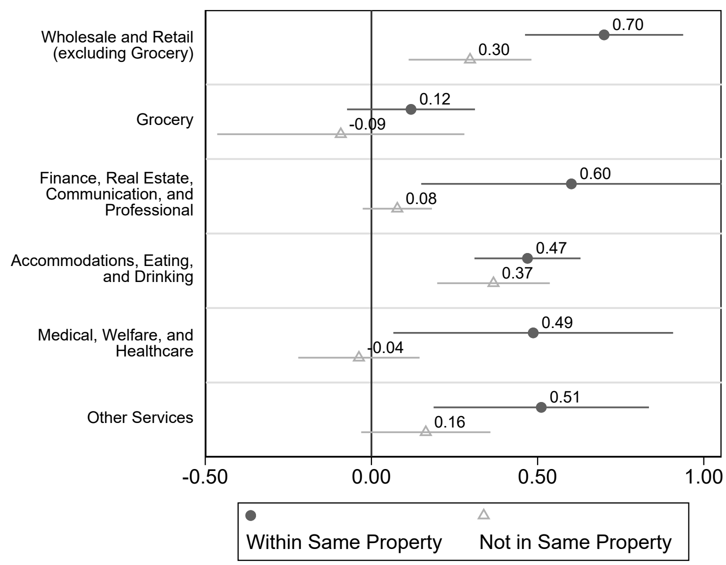
(b) Event Study Results on Surrounding POIs: IPW

Figure 6: Treatment Effects on Foot Traffic to Surrounding Businesses Within 0.1 Miles

Notes: Panel (a) plots the coefficients β_τ^g and corresponding 95% confidence intervals for each treatment group by estimating equation (3). Coefficients β_τ^g summarize the average effects of an opening on nearby businesses in treatment group g in period τ after the opening. The regression sample consists of a panel of all businesses within 0.1 miles from the real grocery store openings and their matched counterfactual locations, observed from 4 months prior to the openings to 10 months after the openings. The treatment group consists of all businesses that are 0–0.1 miles from the real grocery store openings. The control group consists of all businesses that are 0–0.1 miles from the counterfactual locations for openings. Standard errors are clustered at the real or CF site level. Panel (b) plots the coefficients β_τ^g and corresponding 95% confidence intervals for each treatment group by estimating equation (3) and implementing an inverse propensity score weighting.



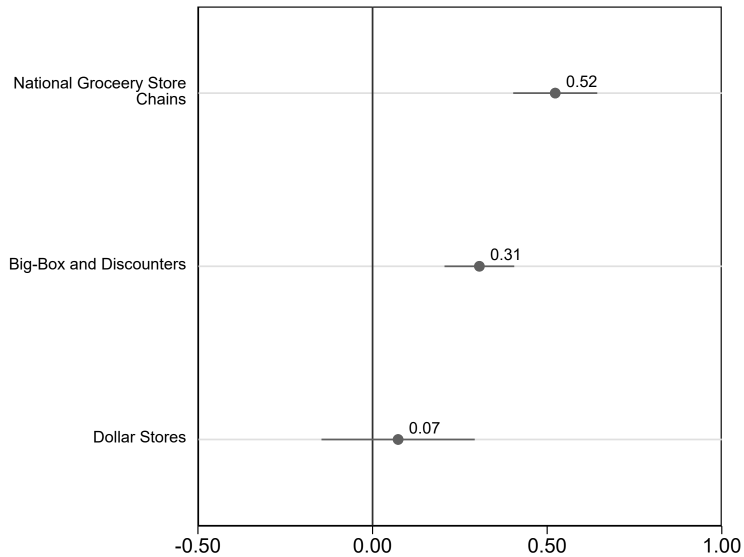
(a) Heterogeneity by Business Category



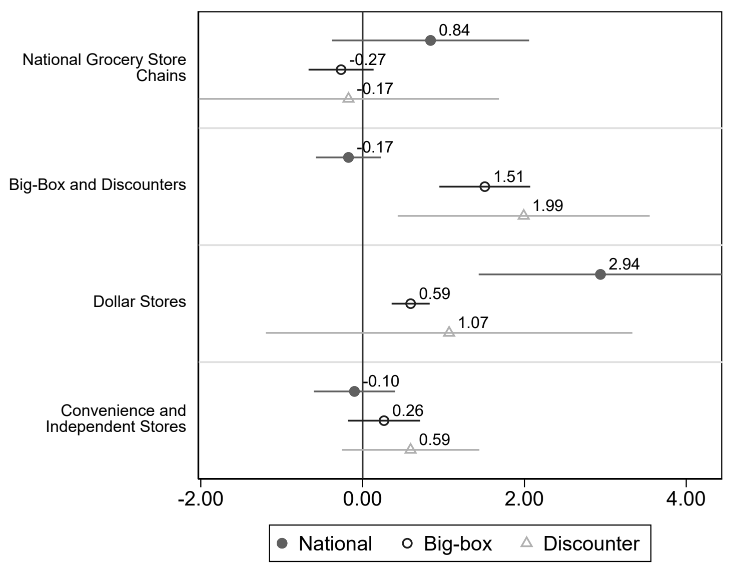
(b) Heterogeneity by Within Same Real Estate Property and Business Category

Figure 7: Heterogeneous Treatment Effects on Foot Traffic: by Type of Surrounding Businesses

Notes: Panel (a) plots the heterogeneous treatment effects on the log of monthly visit counts to nearby businesses within 0–0.1 miles from the grocery store openings 6–10 months after opening relative to 1-4 months before opening. We report the treatment effects on different business categories by row. We include "Grocery Stores" as a separate category. Examples of categories "Finance, Real Estate, Communication, and Professional" include real estate brokerages and banks. Examples of categories "Accommodations, Eating, and Drinking" include restaurants and bars. Examples of categories "Medical, Welfare, and Healthcare" include dentists and primary care providers. Examples of the category "Other services" include religious organizations and public administrations. Standard errors are clustered at the real or CF site level. Panel (b) plots the heterogeneous treatment effects on the five business categories while distinguishing whether the nearby businesses are within the same real estate property as the grocery store openings or not.



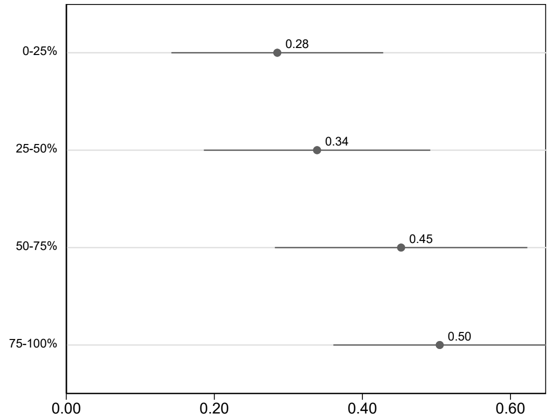
(a) Heterogeneity by Opening Grocery Store Type



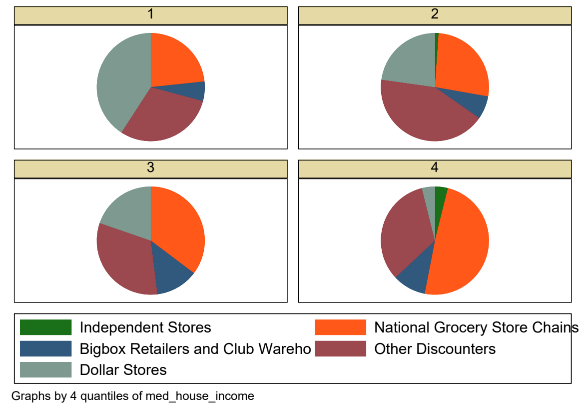
(b) Heterogeneity by Opening Grocery Store Type and Surrounding Grocery Store Type

Figure 8: Heterogeneous Treatment Effects on Foot Traffic: by Type of Grocery Store Openings

Notes: This figure plots the heterogeneous treatment effects by type of grocery store on nearby businesses within 0–0.1 miles from the grocery store openings 6–10 months after opening relative to 1-4 months before opening. Panel (a) shows the heterogeneity by type of grocery store openings. We report the treatment effects of openings of "National Grocery Store Chains", "Big-Box and Discounters", and "Dollar Stores" by row. Panel (b) shows the heterogeneity by opening grocery store type and surrounding grocery store type. The surrounding grocery stores are divided into "National Grocery Store Chains", "Big-Box and Discounters", "Dollar Stores", and "Convenience and Independent Stores", which are labeled on the left of the figure. Standard errors are clustered at the real or CF site level.



(a) Heterogeneity by Median Household Income of the Census Block Groups



(b) Distribution of Opening Grocery Store Types by Median Household Income

Figure 9: Heterogeneous Treatment Effects on Foot Traffic: by Median Household Income of the Census Block Groups

Notes: We rank the Census Block Groups (CBGs) where the grocery store openings are located in terms of their median household income and divide the openings into quartiles. Panel (a) plots the heterogeneous treatment effects by median household income of the CBGs on the businesses within 0–0.1 miles from the grocery store openings 6–10 months after opening relative to 1-4 months before opening. 95% confidence intervals are shown along with point estimates. Standard errors are clustered at the real or CF site level. Panel (b) shows the distribution of types of grocery store openings in each quartile from low-income to high-income.

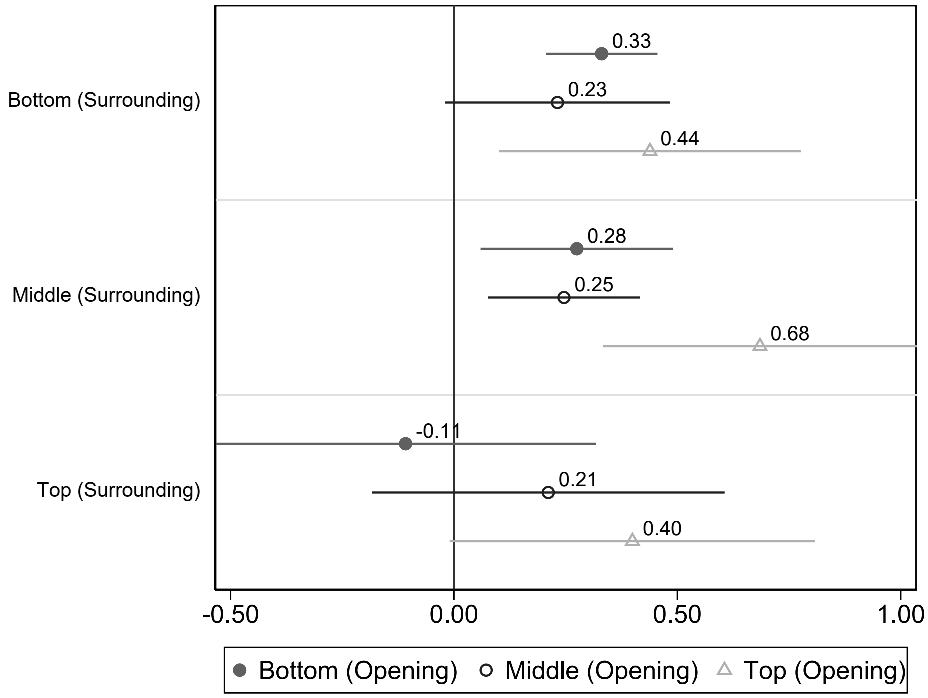
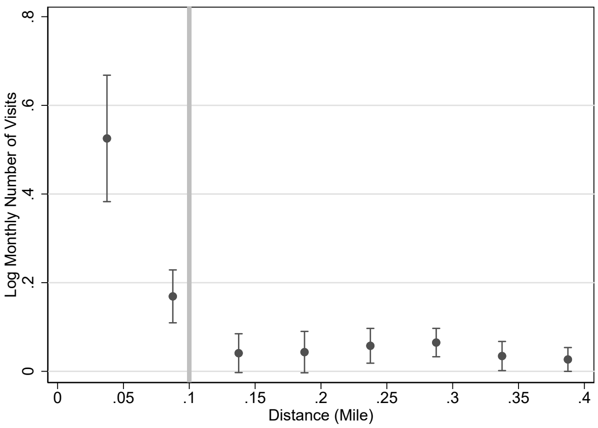
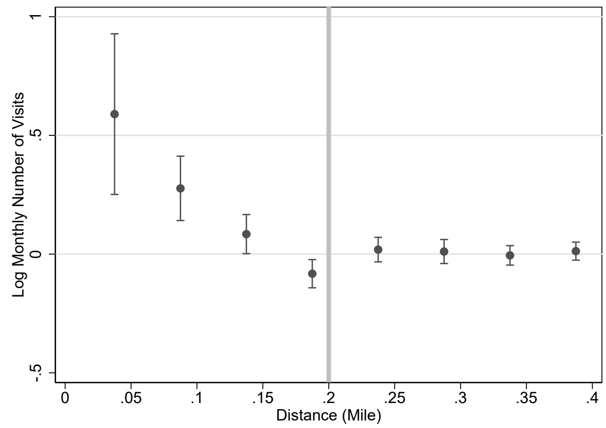


Figure 10: Heterogeneous Treatment Effects on Foot Traffic: by Income Bucket of the Median Customer of Opening Grocery Stores and Surrounding Businesses

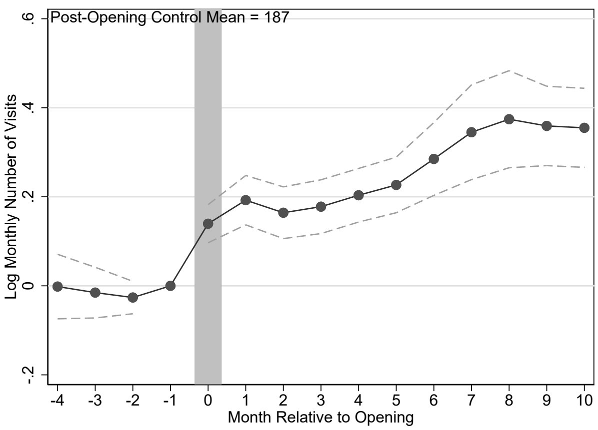
Notes: The figure plots the heterogeneous treatment effects by income bucket of the median customer of opening grocery stores and surrounding businesses within 0–0.1 miles from the grocery store openings 6–10 months after opening relative to 1-4 months before opening. The Safegraph spending data divide the income of customers into 7 buckets: <25k, 25-45k, 45-60k, 60-75k, 75-100k, 100-150k, >150k. For each brand, given the number of customers N_i in the income bucket i in 2019, we define the income bucket of the median customer as $\min\{1 \leq i \leq 7 | \sum_{j=1}^{i} N_j \geq \frac{\sum_{j=1}^{7} N_j}{2}\}$. We then divide all brands into terciles based on the income bucket of the median customer. Note that we treat businesses without brands as individual brands. 95% confidence intervals are shown along with point estimates. Standard errors are clustered at the real or CF site level.



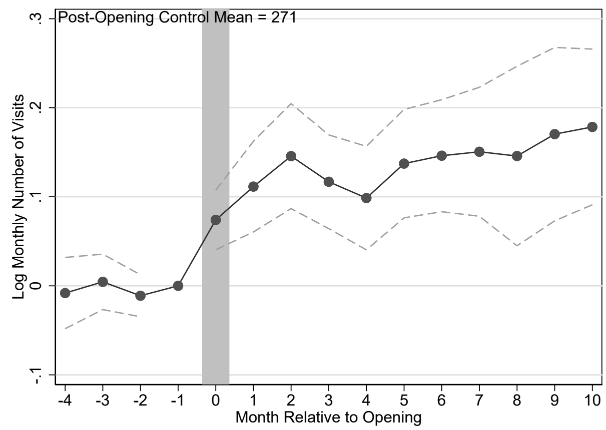
(a) DID Results in High Pop. Density Areas



(b) DID Results in Low Pop. Density Areas



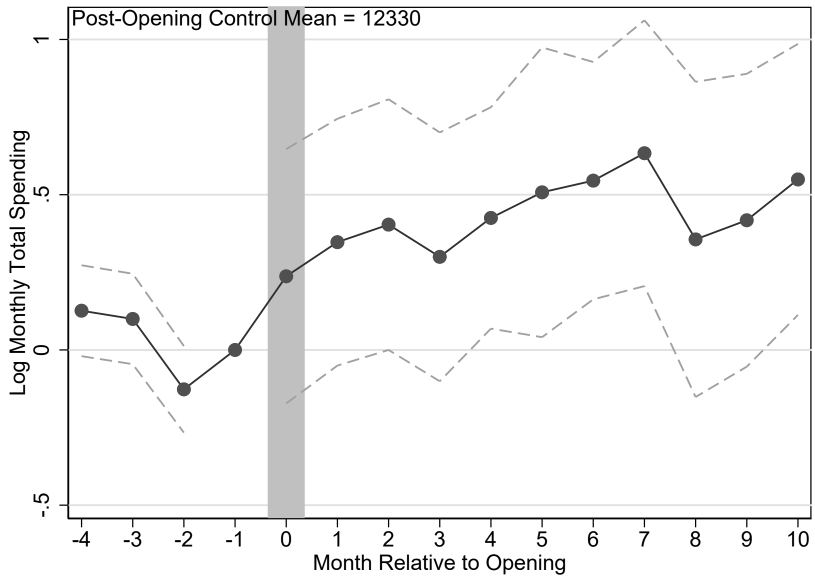
(c) Event Study Results in High Pop. Density Areas



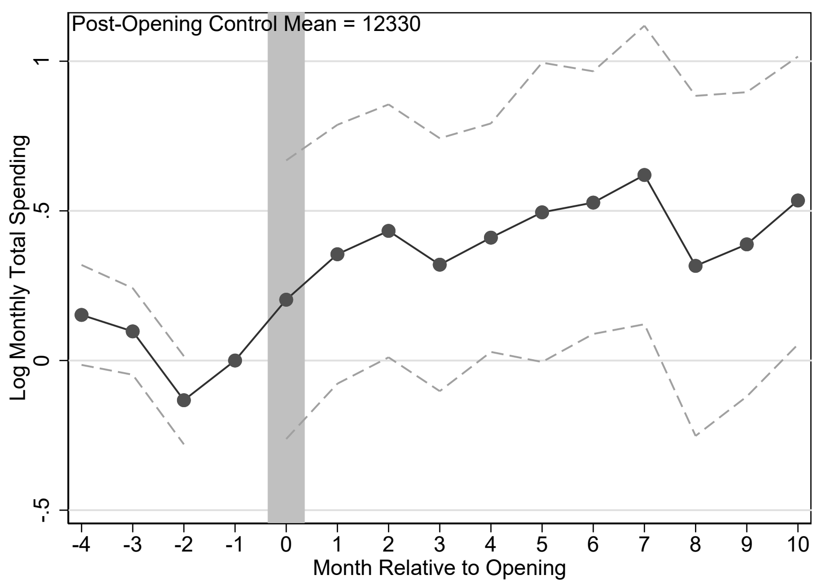
(d) Event Study Results in Low Pop. Density Areas

Figure 11: Treatment Effects on Foot Traffic by Population Density

Notes: Panel (a) shows the treatment effects of grocery store openings in high-population-density areas 0-10 months after opening relative to 1-4 months before opening. Panel (b) presents the treatment effects of grocery openings in low-population-density areas. Panel (c) shows the event study results of grocery store openings in high-population-density areas. Treatment groups are all POIs within 0.1 miles from the real openings in urban areas, and control groups are all POIs within 0.1 miles from the corresponding CF sites. Panel (d) shows the event study results of grocery store openings in low-population-density areas. Treatment groups are all POIs within 0.2 miles from the real openings in non-urban areas, and control groups are all POIs within 0.2 miles from the corresponding CF sites.



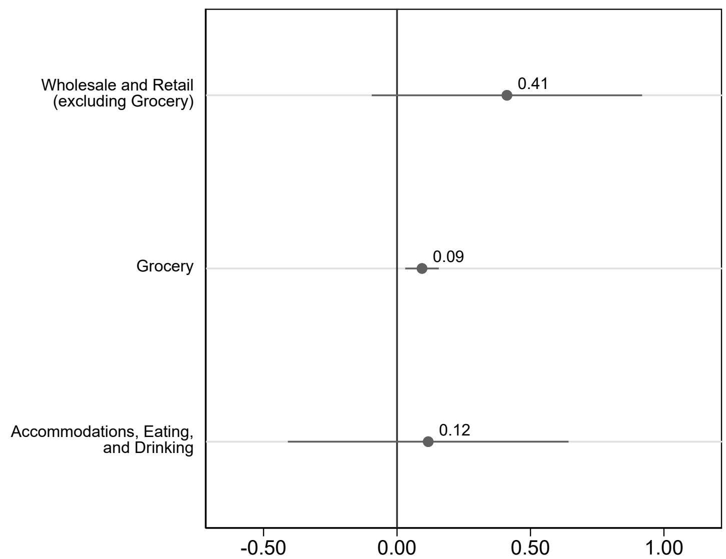
(a) Event Study Results on Surrounding POIs: No Weights



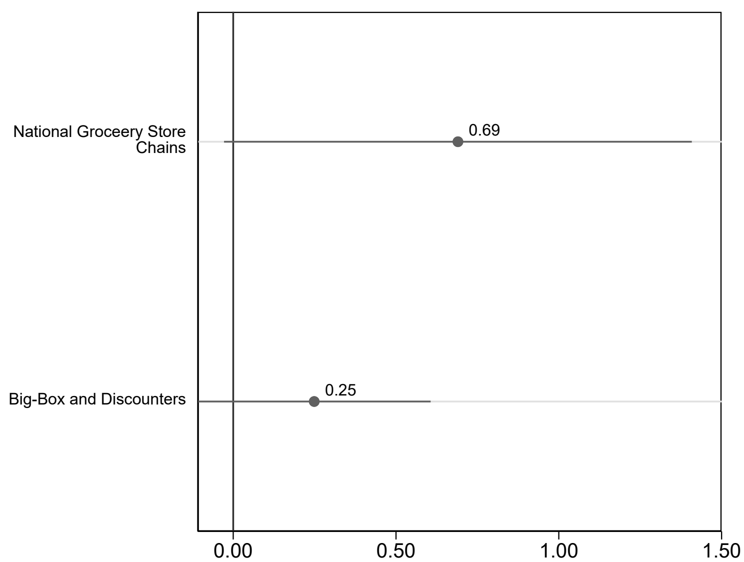
(b) Event Study Results on Surrounding POIs: IPW

Figure 12: Treatment Effects on Spending in Surrounding Businesses within the Same Property

Notes: Panel (a) plots the coefficients β_τ^g and corresponding 95% confidence intervals for each treatment group by estimating equation (3). Coefficients β_τ^g summarize the average effects of an opening on nearby businesses in treatment group g in period τ after the opening. The treatment group consists of all businesses within the same property and within 0.1 miles of the real grocery store openings. The control group consists of all businesses that are 0–0.1 miles from the counterfactual locations for openings. Standard errors are clustered at the real or CF site level. Panel (b) plots the coefficients β_τ^g and corresponding 95% confidence intervals by estimating equation (3) and implementing an inverse propensity score weighting.



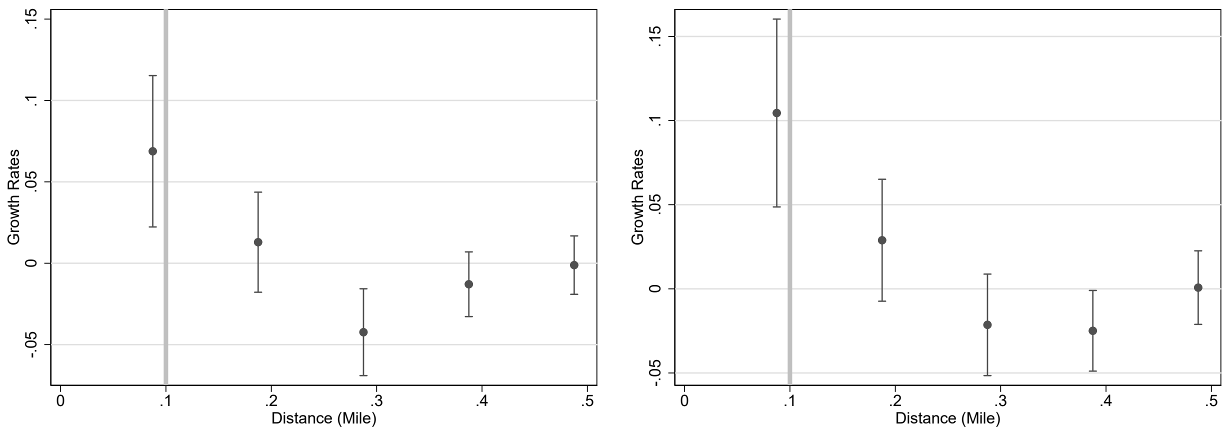
(a) Heterogeneity by Surrounding Business Category



(b) Heterogeneity by Type of Grocery Store Openings

Figure 13: Heterogeneous Treatment Effects on Spending: by Surrounding Business Category and by Type of Grocery Store Openings

Notes: Panel (a) plots the heterogeneous treatment effects by surrounding business category and Panel (b) plots the heterogeneous treatment effects by type of grocery store openings on spending in nearby businesses within the same property as the grocery store openings 6–10 months after opening relative to 1–4 months before opening. The treatment group consists of all businesses within the same property as and are 0–0.1 miles from the real grocery store openings. The control group consists of all businesses that are 0–0.1 miles from the counterfactual locations for openings. 95% confidence intervals are shown along with point estimates. The surrounding businesses are divided into "Wholesale and Retail (excluding Grocery)", "Grocery", and "Accommodations, Eating, and Drinking" by row. We report the treatment effects of openings of "National Grocery Store Chains" and "Big-Box and Discounters" by row. Standard errors are clustered at the real or CF site level.

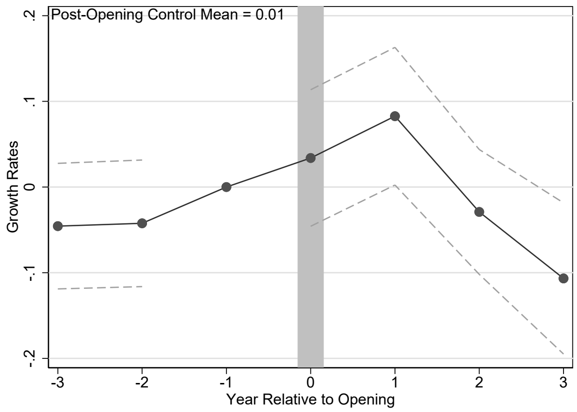


(a) Growth of Numbers of Businesses

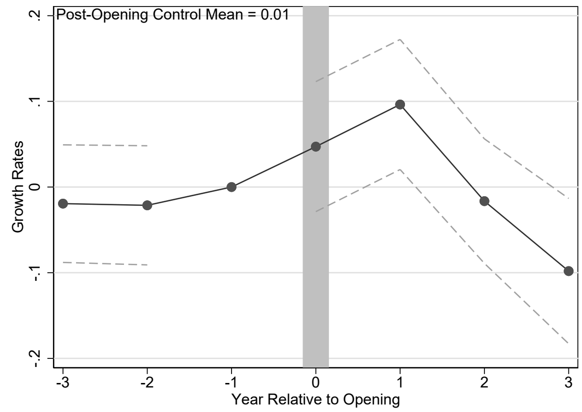
(b) Growth of Employment

Figure 14: Treatment Effects on Surrounding Business Dynamics as a Function of Distance

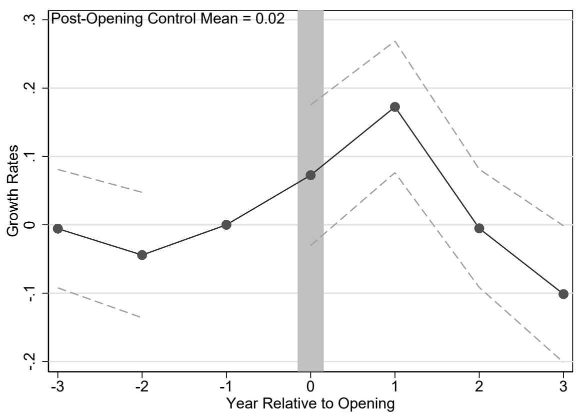
Notes: This figure plots the average treatment effects of grocery store openings on nearby businesses in each concentric ring 0-3 years after opening relative to 1-3 years before opening by estimating equation (9). 95% confidence intervals are shown along with point estimates. The regression sample consists of a panel of all businesses within 0.5 miles from each of the grocery store opening sites and its matched CF site, observed from 3 years prior to the openings to 3 years after the openings. For the area surrounding each real opening site and CF site, respectively, we use a distance band of 0.1 miles to define concentric rings in which surrounding businesses are located for a distance up to 0.5 miles from each site. Panel (a) shows the treatment effects on the growth of the number of businesses. Panel (b) shows the treatment effects on the growth of employment. Standard errors are clustered at the real or CF site level.



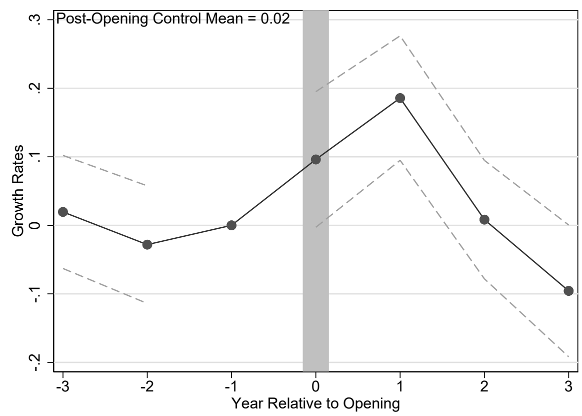
(a) Growth of Numbers of Businesses: No Weights



(b) Growth of Numbers of Businesses: IPW



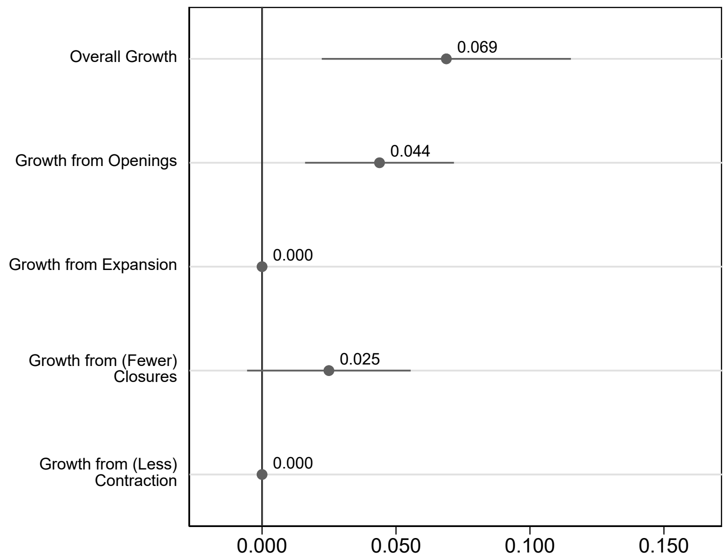
(c) Growth of Employment: No Weights



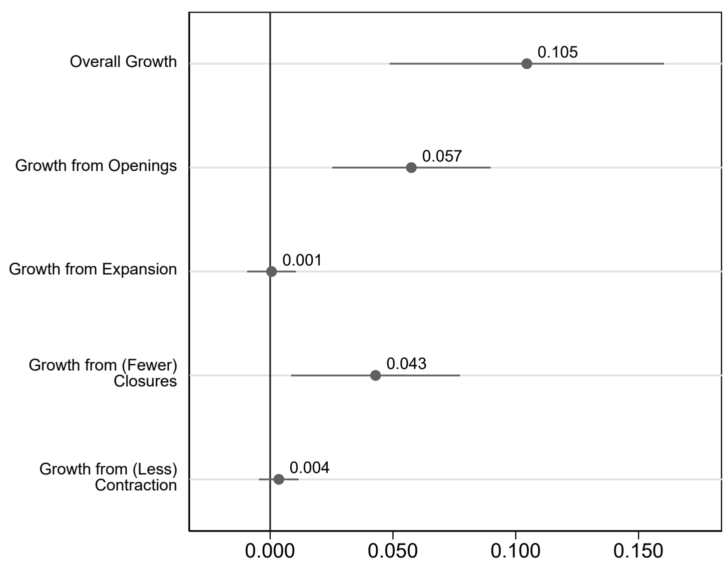
(d) Growth of Employment: IPW

Figure 15: Treatment Effects on Surrounding Business Dynamics Within 0.1 Miles

Notes: This figure plots the coefficients $\beta_{\tau,1}$ and corresponding 95% confidence intervals by estimating equation (9). Coefficients $\beta_{\tau,1}$ summarize the average effects of an opening on nearby businesses within 0.1 miles in period τ after the opening. The regression sample consists of a panel of all businesses within 0.1 miles from the real grocery store openings and their matched counterfactual locations, observed from 3 years prior to the openings to 3 years after the openings. The treatment group consists of all businesses that are 0–0.1 miles from the real grocery store openings. The control group consists of all businesses that are 0–0.1 miles from the counterfactual locations for openings. Standard errors are clustered at the real or CF site level. Panel (a) & (c) plot the results of numbers of businesses, employment, and sales, respectively. Panel (b) & (d) additionally implement an inverse propensity score weighting.



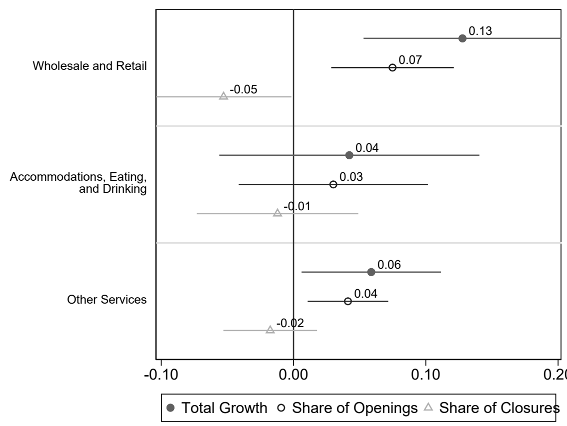
(a) Growth of Numbers of Businesses



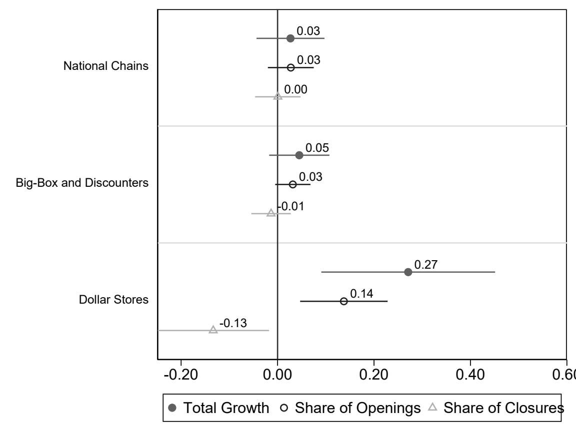
(b) Growth of Employment

Figure 16: Decomposition of Treatment Effects on Surrounding Business Dynamics

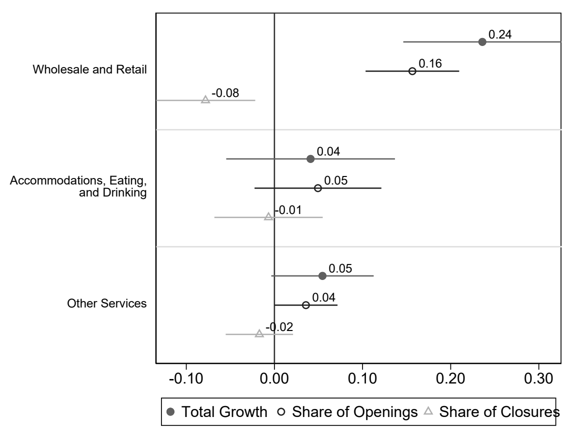
Notes: This figure plots the treatment effects on surrounding business growth decomposed by openings, expansions, closures, and contractions of nearby businesses within 0–0.1 miles from the grocery store openings 0–3 years after opening relative to 1-3 years before opening. 95% confidence intervals are shown along with point estimates. Panel (a) shows the treatment effects on the growth of the number of businesses. Panel (b) shows the treatment effects on the growth of employment. Standard errors are clustered at the real or CF site level.



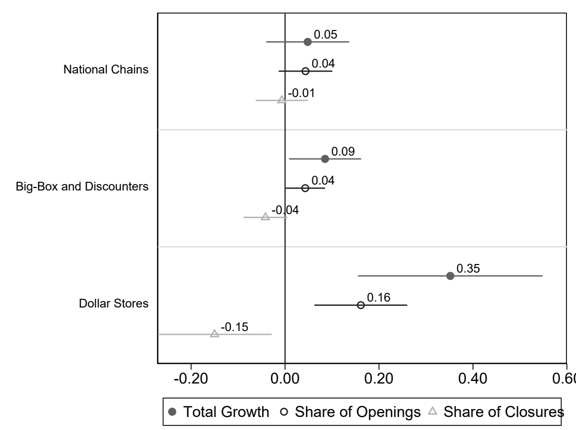
(a) Growth of Numbers of Businesses: by Surrounding Business Category



(b) Growth of Numbers of Businesses: by Opening Grocery Store Type



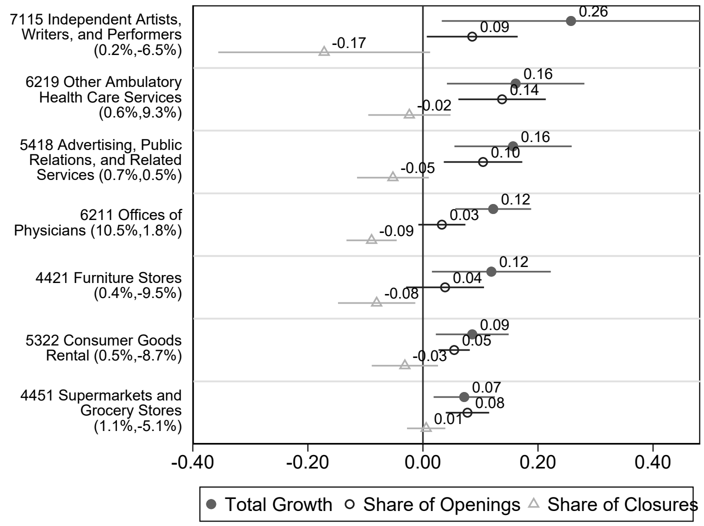
(c) Growth of Employment: by Surrounding Business Category



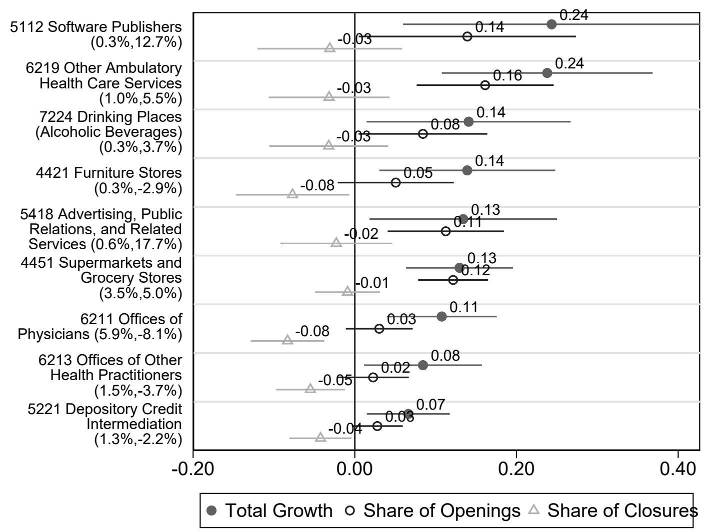
(d) Growth of Employment: by Opening Grocery Store Type

Figure 17: Heterogeneity Treatment Effects on Business Dynamics by Surrounding Business Category and by Type of Grocery Store Openings

Notes: This figure plots the treatment effects by surrounding business category and by opening grocery store type on surrounding business growth and contributions from openings and closures of nearby businesses within 0–0.1 miles from the grocery store openings 0-3 years after opening relative to 1-3 years before opening. 95% confidence intervals are shown along with point estimates. Panel (a) & (b) shows the treatment effects on the growth of the number of businesses. Panel (c) & (d) shows the treatment effects on the growth of employment. For heterogeneity by business category, we construct the growth rate by determining a group k in Equations 7 & 8 by a specific business category in a specific ring next to a site. We then adapt Equation 9 to run a separate regression for each category in the rings within 0.1 miles from the sites. Standard errors are clustered at the real or CF site level.



(a) Growth of Numbers of Businesses



(b) Growth of Employment

Figure 18: Heterogeneity by 4-Digit NAICS Code Industry: Business Dynamics

Notes: This figure plots the treatment effects by 4-digit NAICS code on surrounding business growth and contributions from openings and closures of nearby businesses within 0–0.1 miles from the grocery store openings 0-1 years after opening relative to 1-3 years before opening. 95% confidence intervals are shown along with point estimates. Industries are arranged in rows from top to bottom according to the magnitude of the treatment effect. The numbers in brackets below the industry name represent, in order, the share of the industry among all surrounding businesses before the opening and the average growth rate of the industry around the CF sites after the opening. We only present here the industries with significantly positive effects and pre-existing shares ranking in the top 2/3. Panel (a) shows the treatment effects on the growth of the number of businesses. Panel (b) shows the treatment effects on the growth of employment. Standard errors are clustered at the real or CF site level.

Variable	Treatment (Real Openings)	Control (CF Sites)	Difference	T-statistics
Population	1512.21	1514.79	2.57	0.05
	(799.45)	(761.98)		
Share under 16	0.19	0.19	-0.00	-0.04
	(0.08)	(0.08)		
Share 16-64	0.65	0.65	-0.00	-0.35
	(0.11)	(0.08)		
Share College	0.47	0.47	0.00	0.17
	(0.14)	(0.13)		
Share White	0.71	0.70	0.01	0.63
	(0.26)	(0.27)		
Median Household Income	63,586.11	64,112.34	-525.93	-0.20
	(28,116.01)	(39,094.86)		
Share Unemployed	0.07	0.07	-0.00	-0.75
	(0.06)	(0.06)		

Table 1: Summary Statistics

Notes: We perform a balance test for key neighborhood-level demographic characteristics between our treatment group of real openings of grocery stores and the control group of matched CF sites. Table 1 reports the demographic characteristics of the Census Block Groups that contain the grocery store openings and the matched CF sites.

Appendix For Online Publication

A Additional Tables

count	413
mean	35.566
std	139.770
min	1.165
25%	6.907
50%	15.063
75%	33.527
max	2407.211

Table A1: Descriptive Statistics for Closest Distance between Real Openings (unit: mile)

Notes: This table A1 shows the descriptive statistics for the closest distance between the real openings. We summarize the distance between each real opening and the closest real opening to it.

Grocery Store Type	No.	%
National Grocery Store Chains	137	33.2
Big-Box and Discounters	181	43.8
Dollar Stores	90	21.8
Independent Stores	5	1.2
Total	413	100.0

Table A2: Grocery Store Openings by Type

Notes: This table A2 shows the distribution of grocery stores by type in our sample. We categorize grocery store openings in our sample into 4 categories: national grocery store chains, big-box retailers and discounters, dollar stores, and convenience stores and independent stores

Grocery Store Chain	No.	%
National Grocery Store Chains		
99 Ranch Market	1	0.2
Albertsons	1	0.2
Albertsons Market	1	0.2
C and R Market	1	0.2
Central Market	1	0.2
County Market	1	0.2
Earth Fare	1	0.2
El Ahorro Supermarket	1	0.2
El Rancho Supermercado	2	0.5
El Super	1	0.2
Fareway Stores	2	0.5
Food Lion	1	0.2
Foodland	1	0.2
Foodland Hawaii	1	0.2
Fresh Thyme	2	0.5
GetGo	1	0.2
Giant Food	1	0.2
Giant Food Stores	3	0.7
H-E-B	4	1.0
Harps Food Store	2	0.5
Harris Teeter	1	0.2
Hy-Vee	1	0.2
King Soopers	1	0.2
La Michoacana Meat Market	1	0.2
Lunds & Byerlys	1	0.2
Martin's Foods	2	0.5
Natural Grocers	7	1.7
Net Cost Market	1	0.2
Pete's Market	1	0.2
Price Less Foods	1	0.2
Publix Super Markets	30	7.3
Raley's	1	0.2
Rouses Markets	4	1.0
Schnucks	1	0.2
Seafood City	2	0.5

Continued on next page

Table A3 – continued from previous page

Grocery Store Chain	No.	%
ShopRite	2	0.5
Smart & Final	3	0.7
Sprouts Farmers Market	12	2.9
Super One Foods	1	0.2
Tom Thumb Food & Pharmacy	1	0.2
Trader Joe's	11	2.7
Wegmans Food Markets	1	0.2
Weis Markets	1	0.2
Whole Foods Market	13	3.1
WinCo Foods	1	0.2
Woodman's Market	1	0.2
Big-Box and Discounters		
ALDI	102	24.7
Cash Wise	2	0.5
Costco	11	2.7
Grocery Outlet	23	5.6
Kroger	2	0.5
Lidl	15	3.6
Meijer	8	1.9
Sam's Club	2	0.5
Save-A-Lot	1	0.2
Target	11	2.7
Walmart	4	1.0
Dollar Stores		
Dollar General	80	19.4
Dollar Tree	5	1.2
Family Dollar Stores	5	1.2
Convenience Stores and Independent Stores		
ampm	6	1.5
Indenpendent Stores	5	1.2
Total	413	100.0

Table A3: Grocery Store Openings by Chain

NAICS	Industry	Category	Share	Coefficient
5223	Activities Related to Credit Intermediation	Finance, Real Estate, Communication, and Professional	0.002	0.365
6231	Nursing Care Facilities (Skilled Nursing Facilities)	Medical, Welfare, and Healthcare	0.003	0.242
6116	Other Schools and Instruction	Other Services	0.008	0.204
5191	Other Information Services	Other Services	0.002	0.203
8111	Automotive Repair and Maintenance	Other Services	0.027	0.189
5419	Other Professional, Scientific, and Technical Services	Finance, Real Estate, Communication, and Professional	0.006	0.170
5311	Lessors of Real Estate	Finance, Real Estate, Communication, and Professional	0.017	0.154
4531	Florists	Wholesale and Retail (excluding Grocery)	0.005	0.128
4461	Health and Personal Care Stores	Wholesale and Retail (excluding Grocery)	0.038	0.103
7225	Restaurants and Other Eating Places	Accomodations, Eating, and Drinking	0.221	0.100
6111	Elementary and Secondary Schools	Other Services	0.016	0.095
7139	Other Amusement and Recreation Industries	Other Services	0.035	0.093
4453	Beer, Wine, and Liquor Stores	Wholesale and Retail (excluding Grocery)	0.007	0.084
5221	Depository Credit Intermediation	Finance, Real Estate, Communication, and Professional	0.017	0.077
4539	Other Miscellaneous Store Retailers	Wholesale and Retail (excluding Grocery)	0.022	0.075
4471	Gasoline Stations	Wholesale and Retail (excluding Grocery)	0.023	0.066
6211	Offices of Physicians	Medical, Welfare, and Healthcare	0.035	0.066
4451	Grocery Stores	Grocery	0.028	0.060

Table A4: Heterogeneity by 4-Digit NAICS Code Industry: Foot Traffic

Notes: This table A4 shows the heterogeneous treatment effect coefficients on the log of monthly visit counts to nearby businesses within 0–0.5 miles from the grocery store openings 6–10 months after opening. We report the treatment effects on different 4-digit NAICS code industries by row. We also report the business category that each industry belongs to and the share of the industry among all surrounding businesses before the opening. We only present here the industries with significantly positive effects at 95% confidence level.

Data Source	Description
Panel A. Census Block Group Level Demographics	
ACS	Population
ACS	Share of Female Population
ACS	Share of population under 16
ACS	Share of working-age Population (16-65)
ACS	Share of population over 65
ACS	Share White
ACS	Share Black
ACS	Share Asian
ACS	Share Hispanic
ACS	Share education attainment: less than high school
ACS	Share education attainment: high school
ACS	Share education attainment: college
ACS	Share education attainment: college or more
Panel B. Census Block Group Level Commuting Patterns	
ACS	Share Commuting by Car
ACS	Share Commuting by Public Transportation
ACS	Share Working at Home
ACS	Share Travel Time $<=$ 15 Minutes
ACS	Share Travel Time $15-30$ Minutes
ACS	Share Travel Time $30-45$ Minutes
ACS	Share Travel Time $45-60$ Minutes
ACS	Share Travel Time $60-90$ Minutes
Panel C. Census Block Group Level Income and Employment Information	
ACS	Median Household Income
ACS	Share of population in the labor force
ACS	Unemployment rate
Panel D. Census Block Group Level Housing Information	
ACS	Housing Vacancy rate
ACS	Share of house occupied by owners
ACS	Share of house occupied by renters
ACS	Median rent
ACS	Median housing value
Panel E. Business Density Information	
Safegraph	Number of surrounding "Grocery stores"
Safegraph	Number of surrounding "Wholesale and Retail (excluding grocery stores)" businesses
Safegraph	Number of surrounding "Finance, real estate, communication, and professional" business
Safegraph	Number of surrounding "Accommodation, eating, and drinking" businesses
Safegraph	Number of surrounding "Medical, welfare, and healthcare" businesses
Safegraph	Number of other surrounding businesses

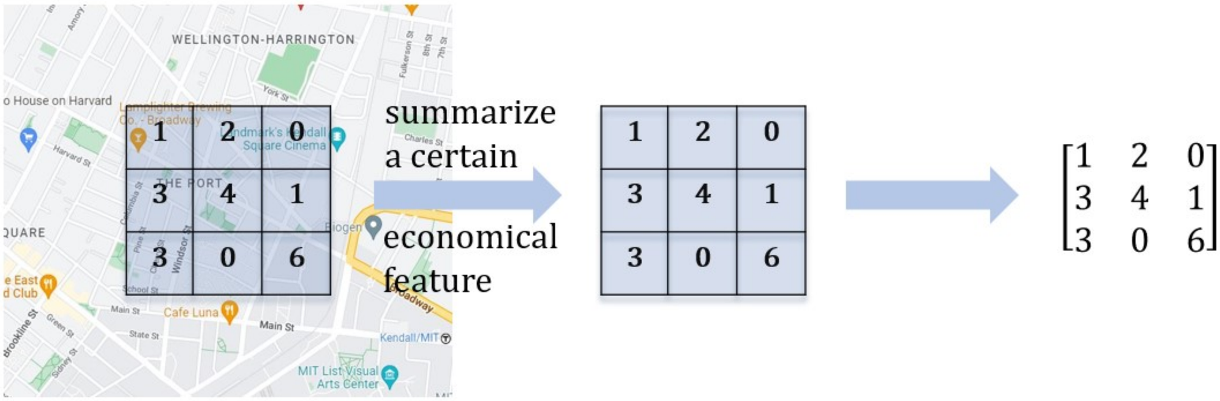
Table A5: List of Input Variables to CNN

B Additional Figures

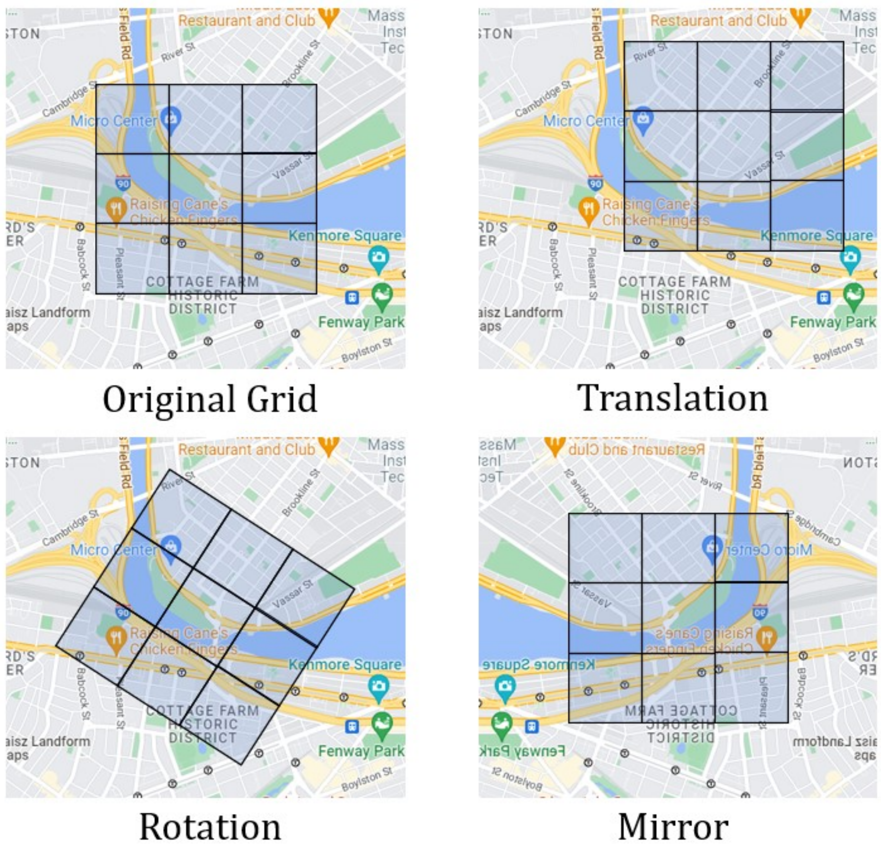
Shopkeeper-concerning Elements	Sub-elements	Variable Description	Source
Neighborhood Demographics and Characteristics	the number of people living in the area	population	Census American Community Survey (ACS) data
		working population	not included
		employment/residential population	not included
		the closest distance to CBDs	not included
		land supply elasticity	not included
		employment density	not included
	the number of potential customers and their purchasing power	age profile	Census American Community Survey (ACS) data
		race profile	Census American Community Survey (ACS) data
		education profile	Census American Community Survey (ACS) data
		household income profile	Census American Community Survey (ACS) data
		unemployment	Census American Community Survey (ACS) data
		industry densities	Dun & Bradstreet
	the neighborhood appearance	median household value	Census American Community Survey (ACS) data
		crime rate	not included
Accessibility, Visibility, and Traffic	accessibility of suppliers	(not important/ have correlations with other features)	not included
	the average foot and vehicle traffic	transportation means ratio: e.g., cars, public transportation, etc.	Census American Community Survey (ACS) data
		travelling time profile	Census American Community Survey (ACS) data
		foot traffic of surroundings (grocery stores etc.)	Census American Community Survey (ACS) data
	availability of ample parking space	parking lot data	not included
Zoning Regulations	avoid residential-only areas	share residential units	not included
Competition and Neighbors	near complementary businesses	surrounding tradable businesses	SafeGraph
		surrounding "Wholesale and Retail"	SafeGraph
		surrounding "Accommodations"	SafeGraph
		surrounding "Eating, Drinking",	SafeGraph
		surrounding "Medical and Health Care"	SafeGraph
	stay away from discounters	surrounding competitors	SafeGraph
Location Costs	the cost of leasing a shop (rents)	commercial rent or alternatives with high correlation	not included

Figure B1: Predictive Model for Counterfactual Sites: List of Input Features

Notes: We specifically focus on five big categories of shopkeepers' concern elements: "Neighbourhood Demographics and Characteristics", "accessibility, Accessibility, Visibility, and Traffic", "Zoning Regulations", "Competition and Neighbors", "Location Costs". And we decide on each characteristic in detail accordingly. The ACS-sourced data is block group-level data.



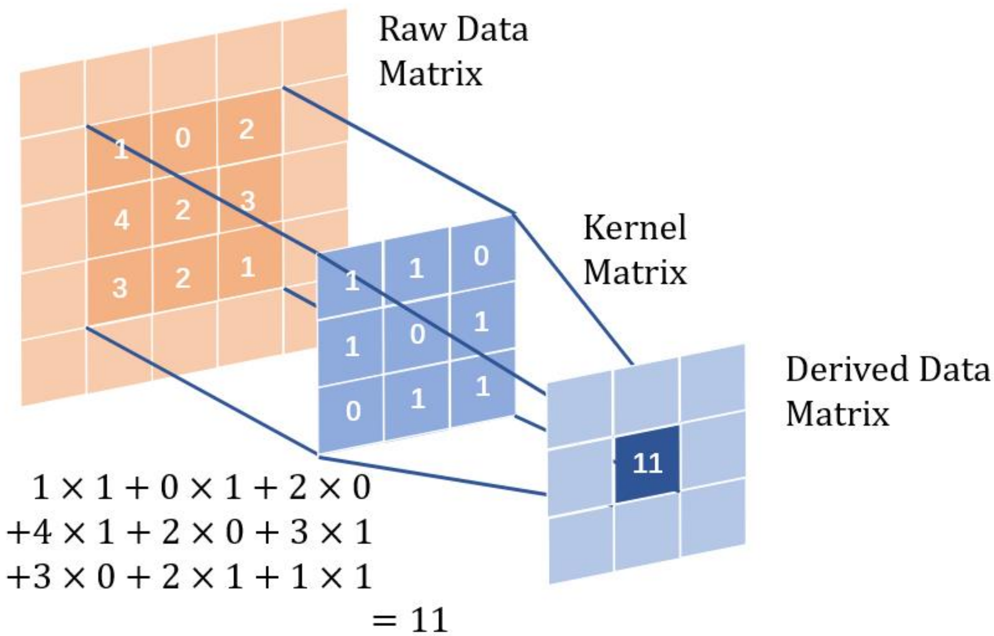
(a) CNN Input Matrix



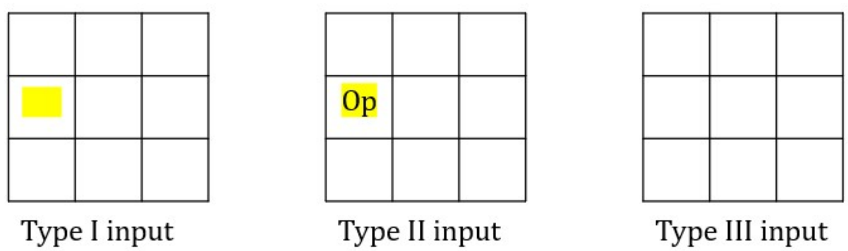
(b) Data Augmentation Methods

Figure B2: CNN Input Matrix and Data Augmentation Methods

Notes: (a). We discretize the geographical space into grids, calculate economical features within each cell of the grid and finally generate the matrices acceptable by CNN. (b). We augment the data using three kinds of transformations, the translation transformation, the rotation transformation, and the mirror transformation. The translation transformation moves the grids vertically or horizontally. The rotation transformation is mimicking a similar opening sample with direction differences. For example, an opening happened with a different river course. As for mirror transformation, we conduct a left-right interchange.



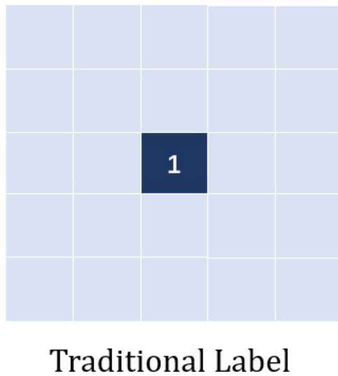
(a) How a CNN kernel extracts features



(b) the Three kinds of Input Sample

Figure B3: How a CNN kernel extracts features and the Three types of Input Sample

Notes: (a) CNN uses the kernel matrix to extract features from the input raw data matrix. Starting from the (0, 0) position of the raw data matrix, the elements of the kernel matrix and the corresponding raw data elements are multiplied and then summed. This sum is the extracted feature and will be recorded in the corresponding position of the derived data matrix. Then, the same feature extraction calculation will be performed on the next corresponding matrix of the raw data matrix (moving one grid to the left or right or up or down). After scanning the entire raw data matrix, a complete derived data matrix is generated. This derived data matrix will be used as a new "raw data matrix" to enter the next stage of feature extraction together with the new kernel matrix. The specific values of the kernel matrix are decided by CNN during its training procedure to meet the optimization conditions. (b) We incorporate the philosophy of GAN into CNN by dividing the samples into three types. The first type consists of at least one real opening, but we gauge out the real opening to mimic an ideal counterfactual location. The second type also consists of a real opening. The third type is randomly picked on the map, not necessarily consisting of any real openings. ("Op" means real opening)



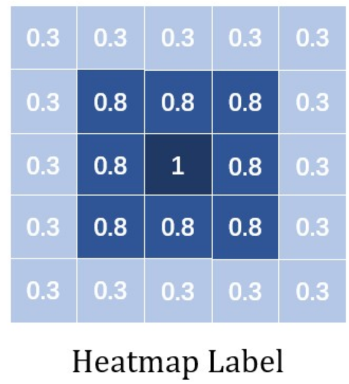
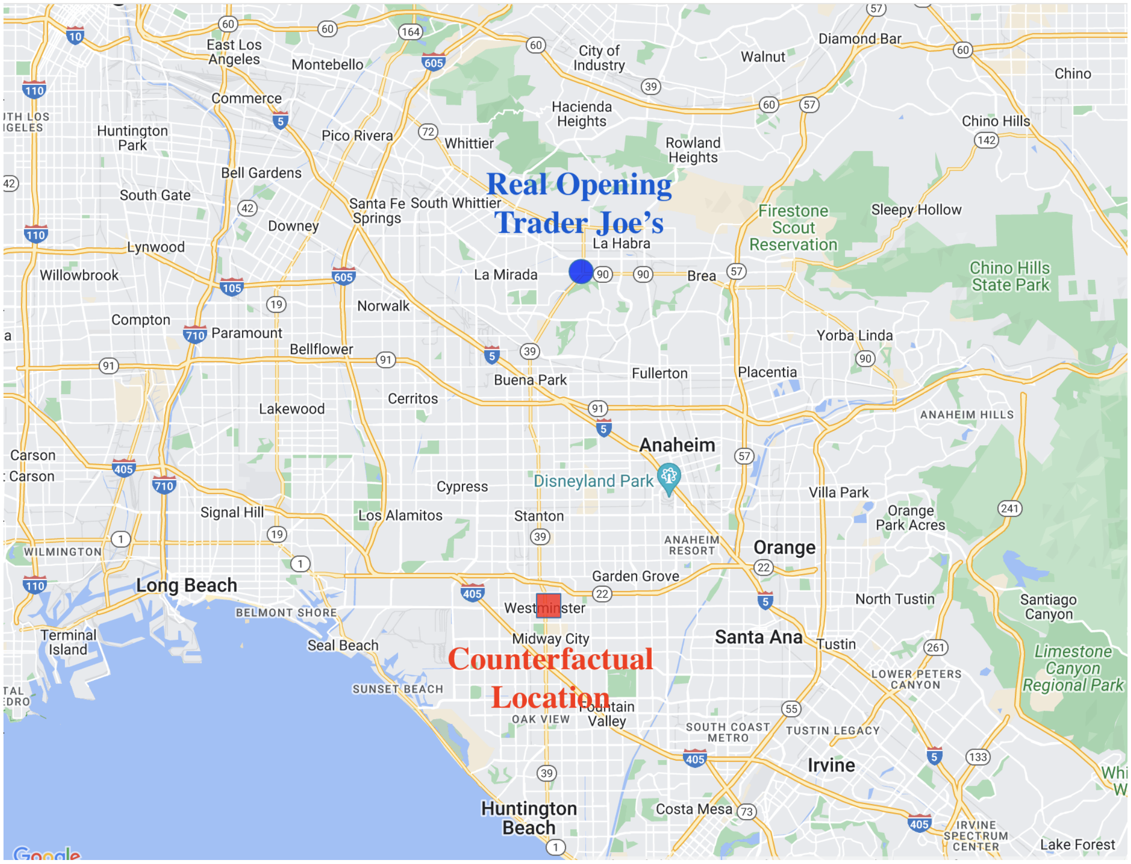
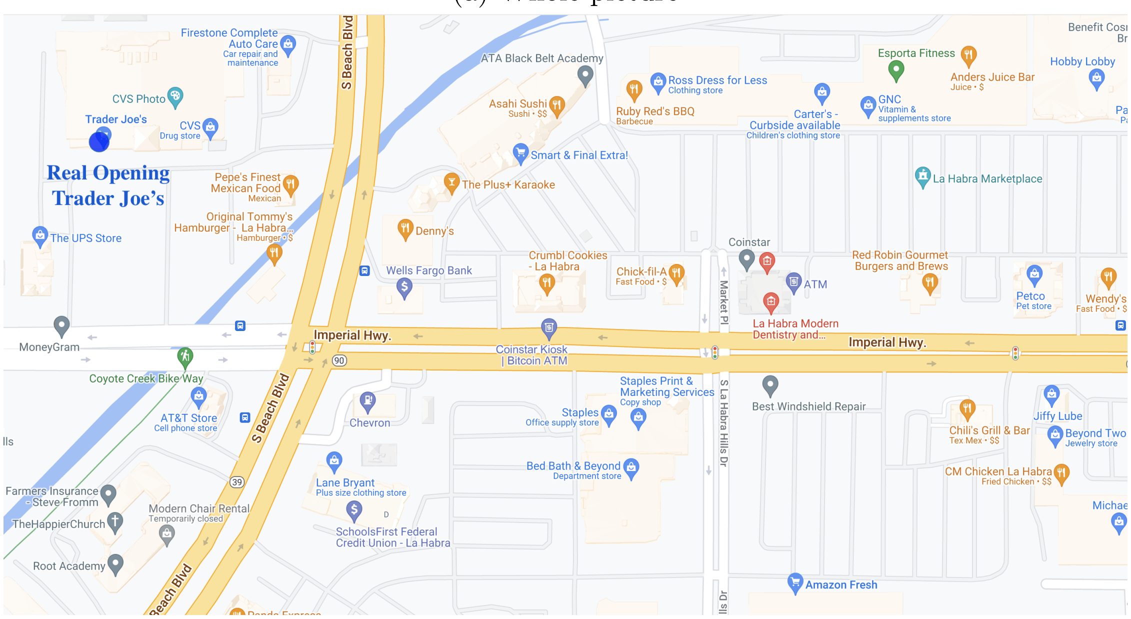


Figure B4: A Traditional label and a Heatmap Label

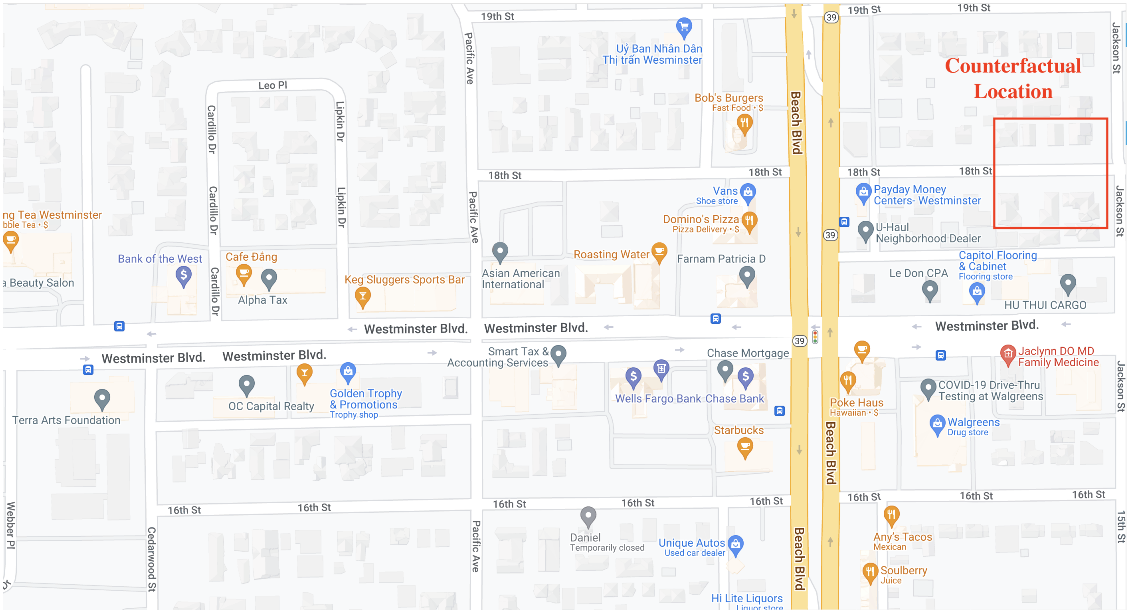
Notes: The traditional label only gives a mark at the correct position, and eventually compresses the matrix into a single character, while the Heatmap label assigns a certain mark value to the entire matrix to guide CNN to reflect and learn, and finally optimizes the overall loss function based on Heatmap regression. We assign 1 to the specific location of real openings and assign decreasing numbers as labels to the locations surrounding them, the number decay function can vary, here we merely present one possible assignment of the label, in our model, we adopt Gaussian's function: $f(x, y) = e^{-A(x^2+y^2)}$. In which A denotes the amplitude, and we assign $A = 1$.



(a) Whole picture



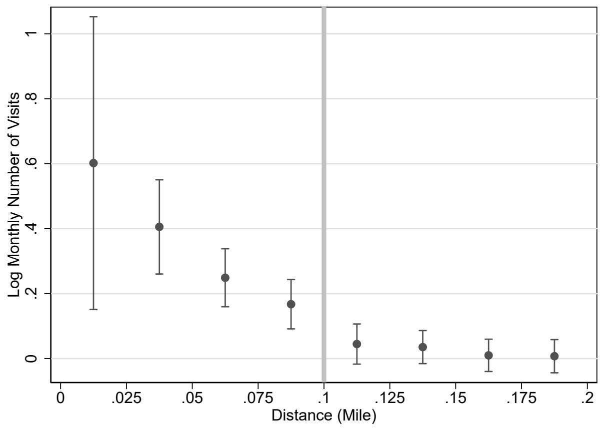
(b) Real opening



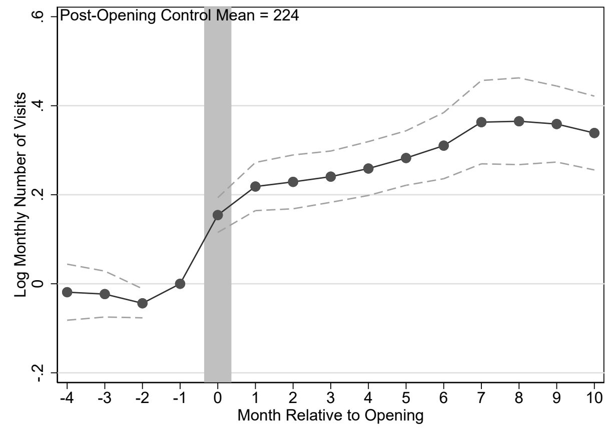
(c) Counterfactual site

Figure B5: Real Opening and Counterfactual Location of Case 333

Notes: This figure shows the relative position of the real opening and the matched counterfactual site of case 333. Panel (a) plots the whole picture, and panels (b) and (c) plot the real opening and counterfactual site separately. The real opening is a Trader Joe's located at latitude 33.91851, longitude -117.969530. The counterfactual site is a 0.025 mile $\times$ 0.025-mile square rather than a single point, with a centroid at latitude 33.76020, longitude -117.988047. The surrounding area of the counterfactual site has comparable features to the real opening, including business density and accessibility to major roads. The real opening operates in a strip mall surrounded by many other businesses near the state highway Beach Boulevard. The counterfactual site is near strip malls along Westminster Blvd and the state highway intersection, making it a possible candidate site. This case demonstrates that the CNN model performs well in giving CF sites comparable to real openings.



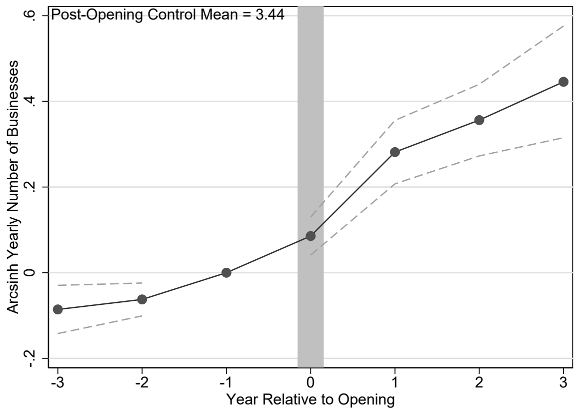
(a) Ring Analysis



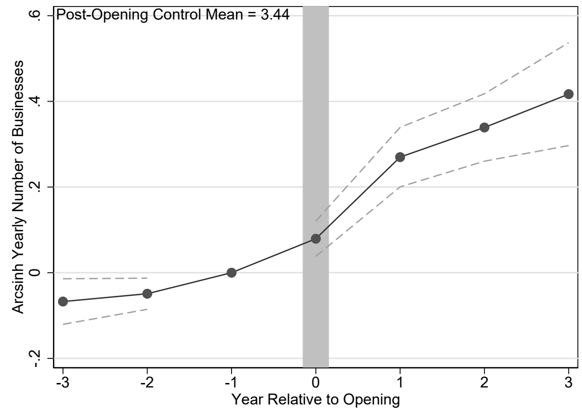
(b) Event-study

Figure B6: Robustness: Excluding Publix Openings

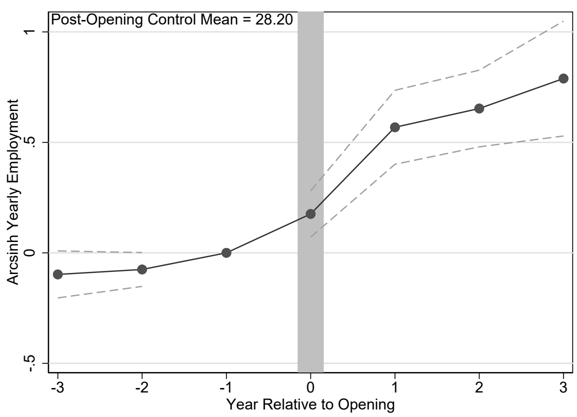
Notes: This figure shows the robustness of the main results after removing the grocery store openings of Publix. Panel (a) corresponds to Figure 5 and plots the average treatment effects of grocery store openings on nearby businesses in each concentric ring in the first 10 months after the openings by estimating equation (3). 95% confidence intervals are shown along with point estimates. The regression sample consists of a panel of all businesses within 0.2 miles from each of the grocery store opening sites and its matched CF site, observed from 4 months prior to the openings to 10 months after the openings. For the area surrounding each real opening site and CF site, respectively, we use a distance band of 0.025 miles to define concentric rings in which surrounding businesses are located for a distance up to 0.2 miles from each site. Panel (b) corresponds to Figure 6a and plots the coefficients $\beta_{\tau,n}$ and corresponding 95% confidence intervals for each treatment group by estimating equation (3). Coefficients $\beta_{\tau,n}$ summarize the average effects of an opening on nearby businesses in the n-th ring in period τ after the opening. The regression sample consists of a panel of all businesses within 0.1 miles from the real grocery store openings and their matched counterfactual locations, observed from 4 months prior to the openings to 10 months after the openings. The treatment group consists of all businesses that are 0–0.1 miles from the real grocery store openings. The control group consists of all businesses that are 0–0.1 miles from the counterfactual locations for openings. Standard errors are clustered at the real or CF site level.



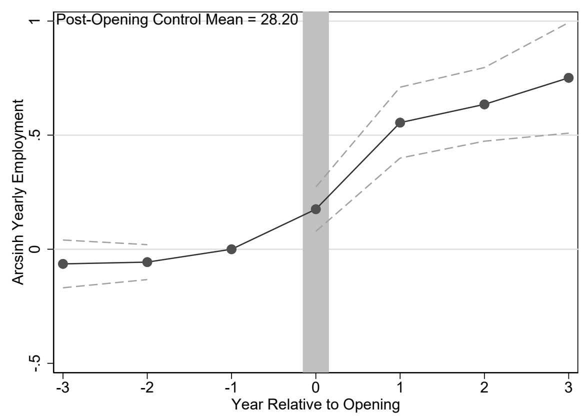
(a) Numbers of Businesses: No Weights



(b) Numbers of Businesses: IPW



(c) Employment: No Weights



(d) Employment: IPW

Figure B7: Treatment Effects on Surrounding Business Dynamics Within 0.1 Miles

Notes: This figure shows the cumulative treatment effects of grocery store openings on the number of surrounding businesses and their employment. This figure plots the coefficients $\beta_{\tau,1}$ and corresponding 95% confidence intervals by estimating equation (9). Coefficients $\beta_{\tau,1}$ summarize the average effects of an opening on nearby businesses within 0.1 miles in period τ after the opening. The regression sample consists of a panel of all businesses within 0.1 miles from the real grocery store openings and their matched counterfactual locations, observed from 3 years prior to the openings to 3 years after the openings. The treatment group consists of all businesses that are 0–0.1 miles from the real grocery store openings. The control group consists of all businesses that are 0–0.1 miles from the counterfactual locations for openings. Standard errors are clustered at the real or CF site level. Panel (a) & (c) plot the results of numbers of businesses, employment, and sales, respectively. Panel (b) & (d) additionally implement an inverse propensity score weighting.

C Determination of Opening Month

We obtain a list of opening grocery stores from Safegraph and Chains Store Guide. However, opening months for grocery stores provided by those two data sources are often inaccurate. In section C.1, we discuss how we impute and correct them. In section C.2, we assess the quality of our imputation.

C.1 Methodology

In order to determine the opening month of a POI, we follow a two-step procedure.

1. In step one, we run POI-level regressions with two structural breaks to determine a candidate opening month

2. In step two, we manually evaluate the quality of the candidate opening month. It is possible that the imputation from the first step is of low quality for a number of reasons. In these cases, we manually correct for/remove the candidate's opening month.

Step 1: Determining candidate opening month

We impute candidate opening month using the monthly number of visits series of each POI, restricted between January 2018 and December 2021[27]. Specifically, we estimate a POI-specific OLS regression with two structural breaks and search for the location of the breaks that maximizes the R-squared of the following regression:

$$N_{it} = \omega_i + \tau_i t + \gamma_{i1}\mathbb{1}\{t \geq t_i^{1*}\} + \rho_{i1} t \times \mathbb{1}\{t \geq t_i^{1*}\} + \gamma_{i2}\mathbb{1}\{t \geq t_i^{2*}\} + \rho_{i2} t \times \mathbb{1}\{t \geq t_i^{2*}\} + \varepsilon_{it}$$

where N_{it} is the monthly number of visits for POI i in month t; τ_i is a POI-specific linear time trend before the first structural break; ρ_{i1} is a POI-specific change in linear time trend after the first structural break; ρ_{i2} is a POI-specific change in linear time trend after the second

[27]Time series length can vary by POI. We use all the observations available. In order to ensure precision in our imputation, we only impute the opening month for POIs with at least 20 months of foot traffic data.

structural break; γ_{i1} and γ_{i2} are the size of the POI-specific structural breaks, respectively; t_i^{1*} and t_i^{2*} are the months of the first and the second structural break, respectively. We estimate the regressions with two structural breaks since the time series of foot traffic spans January 2018 to December 2021, during which the COVID shock could affect foot traffic. Having obtained the imputed months of structural breaks, we label t_i^{1*} as the candidate opening month for grocery stores.

Step 2: Manual Validation and Correction

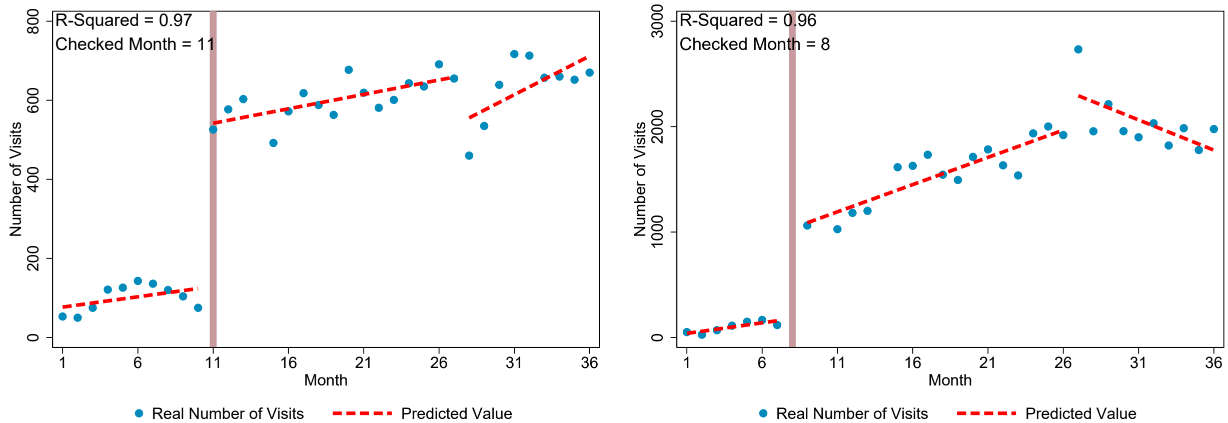
It is possible that the imputation from the first step is of low quality for a number of reasons. In this section, we illustrate what a high-quality candidate's opening month is, and how we manually correct the candidate's opening month when the imputation quality is low.

Keep stores with high-quality imputation

Figure C8 plots two examples of high-quality imputation from the first stage. In these figures, the red bar represents our candidate opening months, and the red dashed line represents the imputed foot traffic. It can be seen from the figures that the foot traffic before the candidate's opening month is close to zero, while the number of visits increases significantly after the candidate's opening month. Meanwhile, the predicted foot traffic aligns well with the real foot traffic data, with an R-squared larger than 0.95.

Drop stores with low-quality imputation

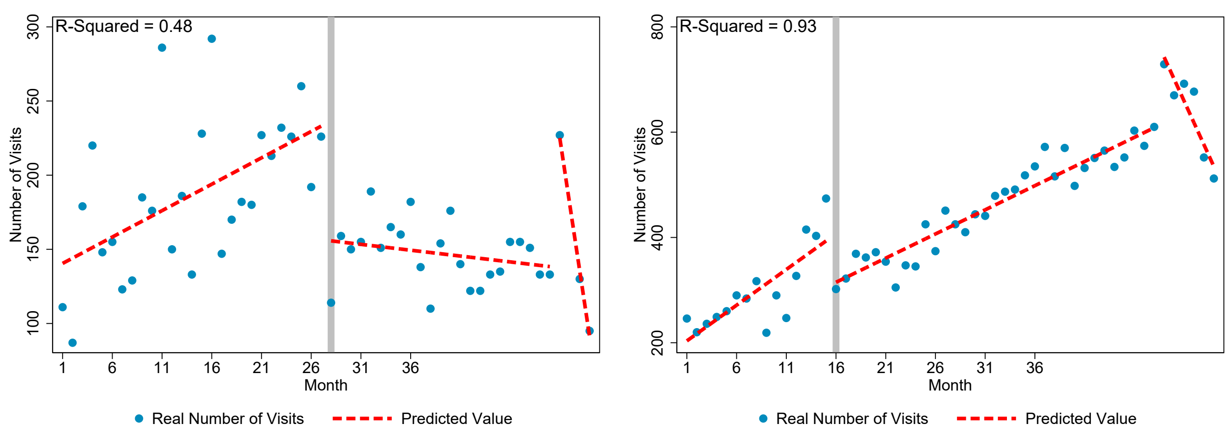
It is possible that there are no clear structural breaks in the foot traffic time series. Figure C9 plots two examples of low-quality imputation from the first stage. In Figure C9(a), the foot traffic is approximately a flat line over time. On the other hand, the foot traffic in Figure C9(b) is almost linear over time. There is no clear structural break in the foot traffic data in either of these cases, even the R-squared could be sometimes extremely high as shown in Figure C9(b). We cannot identify the opening month from foot traffic data when we



(a) Example 1 (b) Example 2

Figure C8: Examples of High-quality Imputed Opening Month

encounter such examples. Because we cannot accurately identify the opening month of such new grocery stores, we have excluded them from our sample.



(a) Example 1 (b) Example 2

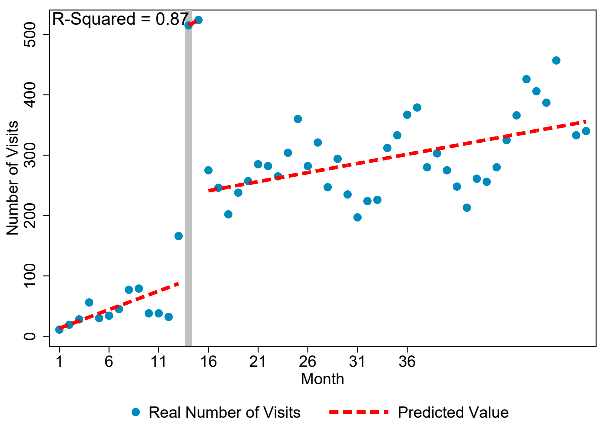
Figure C9: Examples of Low-quality Imputed Opening Month

Correcting for opening month

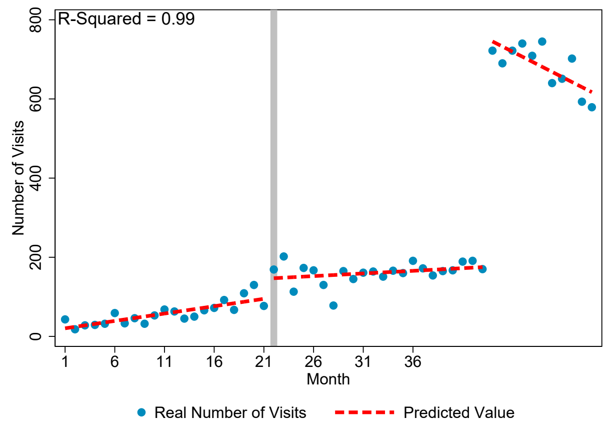
There are times when the structural break algorithm does not capture the most plausible opening month. We adjust the candidate opening months in two situations in our practice.

Case 1. In this case, the structural break algorithm performs well in identifying the first full month of store openings. Since most grocery stores do not open at the beginning of every month, the increase in the number of visits in the opening month can be limited[28]. The algorithm may fail to capture the opening month t in these situations but will assign the month $t+1$ as the opening month instead.

Figure C10(a) provides one of such examples. The candidate opening month from the first step is month 14, whereas we can see a sharp increase in the number of visits in month 13. As a result, we have adjusted the grocery store opening month to 13 instead of 14.



(a) Example 1



(b) Example 2

Figure C10: Examples where we adjusted imputed opening month

Case 2. In this case, the structural break algorithm fails in identifying the store opening months because there are fewer than two structural breaks in data. Figure C10(b) provides one such example. The candidate opening month from the first step is month 22, whereas we can see there is no clear structural break at that time. Instead, we see a sharp increase in the number of visits in month 44. As a result, we have adjusted the grocery store opening month to 44 instead of 22.

[28]For instance, if a newly opened grocery store attracts 30 customers per day, we could expect to see 900 visits within a month if the grocery store openings on the 1st day of the month. When, however, the grocery store opens on the 20th of the month, there will only be 300 visits in the first month.

C.2 Assessing imputation quality

The structural break algorithm is a powerful tool for identifying structural breaks, and we spend a considerable amount of time validating those imputed opening months. We manually searched the Internet for the opening months of grocery stores in our sample in order to further assess the quality of our imputation. We finally successfully found the opening month for 294 newly opened grocery stores in our sample on the Internet.

Table C1 reports the quality of our imputation. When an imputed opening month falls within two months of the opening month we find on the Internet, it is considered to be "accurate.". Among all 294 grocery stores that we can find opening month online, 90.47% of the imputed opening month is accurate.

Note that there is a subset of grocery stores where we directly use the output of structural break estimation as the opening month, and there is another subset of grocery stores that we manually corrected for the opening month. The first subset of grocery stores is referred to as a "high-quality sample" and the second set as a "correction sample". There are 222 grocery stores in the "high-quality sample", and 72 stores in the "correction sample". We then assess the quality by subsample.

We do not see any significant differences in imputation accuracy between these two samples. Among the 222 grocery stores in the "high-quality sample", the accuracy rate is 91.89%, while the accuracy for the 72 stores in the "correction sample" is 86.11%. All these results suggest that our imputation does a good job of guessing the real opening month.

Sample	No. grocery stores	Number accurate	Share accurate
Full sample	294	266	90.47%
High-quality sample	222	204	91.89%
Correction sample	72	62	86.11%

Table C1: Accuracy of Imputed Opening

D Use Safegraph Placekeys to Identify POIs in the Same Property

The Structure of Safegraph Placekey

We utilize the structure of the Placekey for each POI in the Safegraph data to determine whether the POI is in the same property as the grocery store openings. The Placekey of a POI has the format: Address Encoding-POI Encoding@the centroid of the hexagon built on Uber's H3 grid system. Thus, if two POIs have the same Address Encoding and the same centroid of the H3 grid, then we consider these two POIs to have the same address. In addition, Safegraph also provides the parent Placekey for some POIs. If a place is encompassed by a larger place (e.g. mall, airport), then the parent Placekey of the place lists the Placekey of the parent place; otherwise, the parent Placekey of the place is null. Therefore, if we combine the Placekey for the POI itself and the Placekey for the parent of the POI, we are able to identify all POIs in the same property.

Methodology

We identify the POIs in the same property using the following steps.

- **Step 1:** We find all POIs that have the same structure of Address Encoding@the centroid of the hexagon as the grocery opening.
- **Step 2:** We find all corresponding parent Placekeys of these POIs with the same-address Placekey structure if they exist.
- **Step 3:** We find all Placekeys that share the same parent Placekeys found in Step 2, and combine the Placekeys of which the parent Placekeys are null from Step 1. These two parts constitute all POIs in the same property as the grocery opening.

We provide here an example of the POIs at the Columbus Circle shopping mall in Manhattan. There is one Whole Foods Market in the mall. We show how we use our methodology to identify all POIs in the same property as Whole Foods Market.

	A	B	C	E	F	H	I	J	K	L	M	N
6610	22c-222@627-s4r-2rk	225-22v@627-s4r-2tv	Mandarin Oriental Hotels	Mandarin Oriental	Traveler Accommodatio	721110	40.769133	-73.983089	80 Columbus Cir	New York	NY	10023
6611	22b-226@627-s4r-2rk		GamEffective		Wired and Wireless Tele	517312	40.769818	-73.98327	33 W 60th St Ste 1103	New York	NY	10023
6612	**225-22y@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Professional Physical Therapy**		**Offices of Other Health**	**621340**	**40.768532**	**-73.983192**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6613	**225-22f@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Coach**	**Coach**	**Jewelry, Luggage, and L**	**448320**	**40.76787**	**-73.982771**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6614	**225-227@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Amazon Books**	**Amazon Books**	**Book Stores and News I**	**451211**	**40.768551**	**-73.983185**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6615	**25p-223@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**J Crew**	**J.Crew**	**Clothing Stores**	**448140**	**40.768546**	**-73.9832**	**10 Columbus Cir Ste 206B**	**New York**	**NY**	**10019**
6616	23f-222@627-s4r-2tv	225-22v@627-s4r-2tv	Alice Tully Hall		Promoters of Performin	711310	40.76866	-73.982518	1941 Broadway	New York	NY	10023
6617	**225-22p@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**lululemon athletica**	**lululemon athletic**	**Clothing Stores**	**448190**	**40.768422**	**-73.983144**	**10 Columbus Cir Ste 113**	**New York**	**NY**	**10019**
6618	**225-232@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Jo Malone London**	**Jo Malone Londo**	**Health and Personal Ca**	**446120**	**40.768203**	**-73.983042**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6619	**225-22c@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Michael Kors**	**Michael Kors**	**Clothing Stores**	**448140**	**40.768311**	**-73.982835**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6620	**225-22w@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**The Appel Room**				**40.768582**	**-73.98305**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6621	**225-22q@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**First Republic Bank**	**First Republic Ban**	**Depository Credit Interr**	**522110**	**40.768558**	**-73.983201**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6622	**235-222@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Eileen Fisher**	**Eileen Fisher**	**Clothing Stores**	**448120**	**40.768394**	**-73.983159**	**10 Columbus Cir Ste 205**	**New York**	**NY**	**10019**
6623	**225-222@627-s4r-2tv**	**227-229@627-s4r-2tv**	**Juice Press**	**Juice Press**	**Restaurants and Other I**	**722515**	**40.768487**	**-73.982703**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6624	**24x-222@627-s4r-2tv**	**227-229@627-s4r-2tv**	**WFM Coffee Bar**		**Restaurants and Other I**	**722515**	**40.768561**	**-73.983203**	**10 Columbus Cir Ste SC101**	**New York**	**NY**	**10019**
6625	**253-222@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Sephora**	**Sephora**	**Health and Personal Ca**	**446120**	**40.768537**	**-73.983003**	**10 Columbus Cir Ste 201**	**New York**	**NY**	**10019**
6626	**249-222@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Dizzy's Club**		**Other Miscellaneous St**	**453920**	**40.768549**	**-73.98306**	**10 Columbus Cir Fl 5**	**New York**	**NY**	**10019**
6627	**225-233@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Pop Bag USA**		**Jewelry, Luggage, and L**	**448320**	**40.768528**	**-73.983206**	**10 Columbus Cir # 2**	**New York**	**NY**	**10019**
6628	**225-22g@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Tumi**	**Tumi**	**Jewelry, Luggage, and L**	**448320**	**40.768567**	**-73.983175**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6629	**26v-222@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Therabody**		**Health and Personal Ca**	**446120**	**40.768567**	**-73.983193**	**10 Columbus Cir Lbby 02K**	**New York**	**NY**	**10019**
6630	**225-22d@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Stuart Weitzman**	**Stuart Weitzman**	**Shoe Stores**	**448210**	**40.768557**	**-73.983188**	**10 Columbus Cir Spc 101B**	**New York**	**NY**	**10019**
6631	**225-22z@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**WeBop**		**Other Amusement and**	**713990**	**40.768662**	**-73.982818**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6632	**225-22v@627-s4r-2tv**		**Shops At Columbus Circle The**		**Lessors of Real Estate**	**531120**	**40.768141**	**-73.983136**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6633	**227-227@627-s4r-2tv**	**227-229@627-s4r-2tv**	**Per Se**		**Restaurants and Other I**	**722511**	**40.768218**	**-73.98292**	**10 Columbus Cir Ste 4**	**New York**	**NY**	**10019**
6634	**252-222@627-s4r-2tv**	**227-229@627-s4r-2tv**	**Genji Sushi Bars**		**Restaurants and Other I**	**722513**	**40.768555**	**-73.983166**	**10 Columbus Cir Ste 101**	**New York**	**NY**	**10019**
6635	**224-222@627-s4r-2tv**	**225-22v@627-s4r-2tv**	**Condos at 25 Columbus Cir**		**Lessors of Real Estate**	**531110**	**40.767991**	**-73.983164**	**25 Columbus Cir**	**New York**	**NY**	**10019**
6636	**225-23c@627-s4r-2tv**	**225-22v@627-s4r-2tv**	**Condos at 10 Columbus Cir**		**Lessors of Real Estate**	**531110**	**40.768263**	**-73.983105**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6637	**225-224@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Whole Foods Market**	**Whole Foods Mar**	**Grocery Stores**	**445110**	**40.768539**	**-73.983178**	**10 Columbus Cir**	**New York**	**NY**	**10019**
6638	**22r-222@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Solstice Sunglasses**	**Solstice Sunglasse**	**Other Miscellaneous St**	**453998**	**40.768303**	**-73.983068**	**10 Columbus Cir Ste 306**	**New York**	**NY**	**10019**
6639	**23g-222@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Williams-Sonoma**	**Williams-Sonoma**	**Home Furnishings Store**	**442299**	**40.76855**	**-73.982941**	**10 Columbus Cir Ste 114**	**New York**	**NY**	**10019**
6640	**24y-222@627-s4r-2tv**	**zzw-226@627-s4r-4d9**	**Jackrabbit**	**JackRabbit**	**Shoe Stores**	**448210**	**40.768859**	**-73.982527**	**10 Columbus Cir Ste 210**	**New York**	**NY**	**10019**

Figure D11: Example of POIs in the Same Property as Grocery Store Openings

Notes: This table shows an example of the POIs that can be identified as located in the same property as a Whole Foods Market at Columbus Circle in Manhattan. The first column is the Placekey for the POI itself, and the second column is the Placekey for the parent.

By manual checks, we identify that all POIs in bold text are located in the Columbus Circle Mall. We proceed with finding the same POIs in the property using the following steps:

- First, we find all POIs that have the same POI Placekey format of 225-xxx@627-s4r-2tv as the Whole Foods Market.
- Second, we find the corresponding parent POI Placekeys (zzw-226@627-s4r-4d9 and 227-229@627-s4r-2tv and 225-22v@627-s4r-2tv).
- Third, we select all POIs with the parent POI Placekeys in step 2, and all POIs without a parent POI from step 1. These are all the POIs that we identify as in the same property as the Whole Foods Market in Columbus Circle.