The Effects of Agglomeration on Customer Traffic & Commercial Real Estate Values: Evidence from Grocery Store Openings

Franklin Qian[*], Qianyang Zhang[†] & Xiang Zhang[‡]

January 4, 2023

Abstract

We investigate the driving forces and effects of agglomeration in the local non-tradable service sector on the value of commercial real estate (CRE). Using a sample of 413 grocery store openings in the U.S. in 2018 and 2019, we find openings of grocery stores lead to significant growth in foot traffic to their opening locations and a 31.6 percent increase in foot traffic to nearby businesses within 0.1 miles. The spillovers of demand are strongest between new grocery stores and existing grocery stores and businesses in wholesale and retail. Landlords capitalized on the benefits of agglomeration and increase the rents in newly-signed leases increase by 23.8% in a half-mile radius from the real grocery store openings in the first two years after the openings.

[*]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 revenue and demand spillover do grocery stores generate in neighboring businesses? How do these spillover effects capitalize on rents for tenants of commercial real estate (CRE) properties and affect the performance and market values of CRE assets? There has been a lack of empirical research evidence on the externalities of grocery anchors. This paper fills this gap by studying the externalities of anchoring grocery stores in nearby retail businesses.

We collect a comprehensive sample of grocery stores that opened across the U.S. in 2018 and 2019. We develop a novel 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, (ii) rents negotiated for nearby commercial tenants.

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 the 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 existing neighborhood-level demographics and business composition in the vicinity. Next, we estimate a propensity score model to predict suitable counterfactual opening locations for each real 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 examine the heterogeneity in the demand spillovers by the categories of businesses nearby. We find the grocery store openings generate additional customer traffic to existing grocery stores within 0.1 miles and increase their foot traffic by 60 percent 6–10 months after the opening. We find further evidence that the synergies in sharing common demand between the grocery stores and other nearby businesses are strongest in the category of wholesale and retail. On average, other grocery stores within 0.1 miles from a grocery store opening experience a 60 percent increase in foot traffic 6–10 months after the opening, while wholesale and retail stores (excluding grocery stores) in the radius of 0.1 miles from a grocery store experience a 43 percent increase in foot traffic 6–10 months after the opening. In comparison, such synergies appear to be the weakest between grocery stores and medical, welfare, and healthcare businesses. In addition, we find that businesses located in the same real estate properties gain more in terms of foot traffic than those that are not. Wholesale and retail businesses in the same property as the grocery store opening see an increase in foot traffic by

68 percent 6–10 months after the opening, whereas wholesale and retail businesses outside the property of the grocery store opening only see an increase of foot traffic of 36 percent. In addition, we find that positive spillovers are driven by the large presence of national chains in our sample, such as Publix and Whole Foods Market. In contrast, the effects of opening a dollar store on nearby businesses' foot traffic is negligible.

Our preliminary rent capitalization analysis provides evidence that landlords of commercial real estate (CRE) properties capitalize on the demand spillovers generated by the grocery store openings. We find that rents on newly signed leases increase by 23.8% in a half-mile radius from real grocery store openings in the first two years after openings. This is equivalent to an increase of $\$4 - \$5$ in rents per square foot. In the future, we will also investigate how grocery store openings affect other terms of the leases for surrounding CRE properties. In addition, we plan to examine the impacts on the business composition. Outcome variables of interest include the number, size, and types of business entrants and exits, etc. These estimates will allow us to better understand how grocery store openings change the nearby business dynamics.

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)). Relative little is known about the strength and impacts of agglomeration in the local non-tradable service sector. Within the non-tradable service sector, a number of papers examine the consequences of dis-economies of agglomeration. For example, they find bankruptcies of firms in non-tradable service sector have significant spillover effects on nearby tradable firms, causing decline in employment and business closures (Shoag and Veuger (2018), Bernstein et

[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).

al. (2019), Benmelech et al. (2019)). 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.[2] We contribute to this literature by moving beyond the labor market outcomes and focus on the local commercial real estate (CRE) market. We provide direct evidence on how demand spillovers due to agglomeration capitalize into the performance and value of CRE assets.[3]

Prior work has shown that the benefits of agglomeration economies such as an increase in 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.[4] Within the non-tradable service sector, the literature suggests that increased productivity due to agglomeration could arise due to shared customer traffic, which reduces consumer search costs and benefit geographically proximate stores (see, for example, Pashigian and Gould (1998), Gould et al. (2005), Bernstein et al. (2019), and Benmelech et al. (2019)). However, direct evidence on customer traffic channels is rare due to a lack of microdata measuring consumer traffic and consumption. We provide micro-level estimates on how anchoring grocery stores that attract customers leads to demand spillovers on nearby stores and how these effects change as a function of distance to other stores.

Our paper also contributes to the growing literature on causal inference in a spatial setting. Traditional methods for spatial treatment effects usually compare businesses in an inner ring around the opening locations of grocery stores with businesses further away in outer rings. 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. Following Pollmann (2020), we build a CNN model to find counterfactual

[2]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).

[3]Our paper is closely related to Rosenthal and Urrego (2021) who show retail spatial concentration lower neighborhood crime rates and capitalize into higher retail rents.

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

locations that are similar to the locations of real grocery store openings. Relative to Pollmann (2020), we make the following four-fold contributions. First, we include a much 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 to tell how good our CNN model is at predicting a suitable opening location.[5] Third, we demonstrate that we are able to find counterfactual locations for grocery store opening across a broad set of geographies using our GAN-based CNN model. Fourth, we further rely on the sharp timing of the grocery store entries as a shock to strengthen the credibility of our identification strategy, while Pollmann (2020) only uses cross-sectional variation. Seeing sharp changes in foot traffic around the precise timing of these shocks make 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.[6] 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

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

[6]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).

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 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 monthly number of visits to each store. By identifying structural breaks in monthly visits, we impute the opening month for each grocery store.[7] 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. (2021).[8]

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

[8]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.

2.2.2 Commercial Real Estate Lease Comps

We obtain data on the rents, lease characteristics and tenant information at CRE properties from a comprehensive database of lease comps for commercial real estate properties provided by Costar Group. Costar covers more than 3.5 million leases in industrial, office, and retail property types in the US. The starting time at which CoStar began collecting data varies in different markets, so some markets have longer time series of commercial rent data than others. For office and industrial properties, most major markets have data back to at least 2000. Smaller markets and retail properties typically begin in 2006. As our sample of grocery store openings are from 2018 and 2019, we use lease comps that were signed between 2017 and 2021 in our analysis. Given our focus on estimating the causal effects of opening a new grocery store on nearby affected businesses, we identify leasing comps within 0.5 miles of the grocery store openings in our sample and their matched counterfactual sites. In a future version, we will complement our sample with leasing comps from CompStak.

3 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 business, 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 business are already high and the options for grocery shopping are not too many, so that they are more likely to attract a grocery store opening nearby.

To overcome this identification challenge, ideally we would like to compare 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 prior to 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 adapt 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, and Appendix E provides more details.

3.1 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,[9] (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 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 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.

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

3.1.1 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 to 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 openings 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 of 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.

3.1.2 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 adapt 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 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,[10] 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 contain 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 distinguish different 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 simulta-

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

neously 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 the 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.

3.1.3 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 candidates 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:

- 18 Census Block Group level demographic characteristics,
- the number of business establishments belonging to 5 different categories,
- 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 socres. 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 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 opening 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.

3.2 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 surrounding geography of the grocery store along with its counterfactual opening location. Trader Joe's opened on July 18, 2019, and was the first Trader Joe's in the La Habra Area.[11] It replaced Vons, which closed on October 12, 2018.[12] 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). The location is also surrounded by a high level of density of nearby businesses.

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 could have chosen the counterfactual site, which has a similar desirability to open its business.

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

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

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 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 from economies of agglomeration. This may not be 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, 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:

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

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 to remove potential composition biases of businesses across geographic locations.[13] We can run a separate regression comparing rings that are equidistant from the real and

[13]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 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, $\beta_{0,n}$, to zero. Standard errors are clustered at the real or CF opening site level.

4 Empirical Results

4.1 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. [In a future version, we will also perform a balance test for additional covariates not used in training our model.]

4.2 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.

$$Y_{it} = \alpha_i + \gamma_t + \sum_{\tau=-4}^{10} \mu_\tau D_{\tau,it} + \varepsilon_{it}. \tag{2}$$

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, 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 break 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.

4.3 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 could 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 (1) using log of 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 panel of all businesses within 0.2 miles 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.[14]

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 real 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 (1) 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

[14] The Safegraph uses a visits attribution algorithm that first cluster GPS points associated with a visit to a single place and then decide the most-likely visit among a set of potential options. Before clustering, Safegraph filter out high horizontal accuracy pings and jumpy GPS pings to ensure the accuracy of the visits attribution algorithm. In the first step of clustering GPS points, Safegraph uses a modified version of the canonical DBSCAN clustering algorithm, which addresses the issue of GPS signal drift. They consider a sequence of pings that are within some 80 meters (0.05 miles) a cluster as long as they never go more than 100 meters (0.067 miles) away from the last ping. In the second step of choosing the most-likely visit among a set of potential options, Safegraph develops a learning-to-rank model that accurately learns how to rank a set of nearby places by comparing the feature vectors for those places.

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 3. 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.

We further examine the heterogeneity in demand spillovers by the categories of the surrounding businesses with a triple difference-in-difference specification. We hypothesize that the externalities of the openings of grocery stores in neighboring stores 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 (6) Other Services. Figure 3 shows the composition of businesses by category in our treatment and control groups. 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 43 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 business most likely share common demand. Clustering around a grocery store anchor helps attract more customers. However, demand spillovers are the smallest in medical, welfare, and healthcare, with a statistically significant difference of negative 35 percent in treatment effect compared to the baseline category. These types of business do not share much common demand with grocery stores, hence the synergies are weak. Interesting, we find openings of grocery stores do not pose a threat to other nearby grocery stores in diverting their existing customers' demand. On the contrary, on average, they lead to substantial positive demand spillovers for other grocery stores in the vicinity.[15]

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 triple difference-in-difference regression adapted from equation 1 on each sub-sample.[16] Figure 7b shows that wholesale and retail businesses gain the most with an increase of 68 percent in foot traffic if they are located within the same property as the grocery store opening 6–10 months after opening. Whereas wholesale and retail businesses not in the same property have much smaller increase in foot traffic of 35 percent. The pattern is broadly similar for other business categories. Businesses within the same property 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 hospitality services.

In addition, we also examine the heterogeneous demand spillovers generated by the openings of different types of grocery stores with a triple difference-in-difference specification. Figure 10 suggests that the positive demand spillovers generated by grocery stores opening are mainly driven by national grocery chains and warehouse club store or discounter 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. Whereas an opening of a big-box retailer or discounter increases foot traffic to nearby businesses by 31 percent. In

[15]The differential effect between nearby affected grocery stores and wholesale and retail excluding grocery stores is an economically large 17% despite being not significant at the 95% level. Hence the positive spillovers on existing grocery stores are at least as large as on other wholesale and retail businesses.

[16]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.

contrast, opening a dollar store only increases foot traffic to nearby businesses by 7 percent.

We further examine the heterogeneous demand spillovers in urban areas versus non-urban areas. We divide grocery store openings into urban and non-urban groups according to the population density of the counties where they are located.[17] The results of the difference-in-difference with multiple rings in urban areas and non-urban areas are presented in Figure 8 separately. Panel (a) suggests that demand spillover effects are clustered within 0.1 miles from the openings of grocery stores, and Panel (b) suggests that demand spillovers extend to a range of 0.2 miles in non-urban areas. In Figure 9, we also present the event study results of grocery store openings on other businesses in affected areas in terms of foot traffic respectively. Panel (a) shows the event study results in urban 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 (b) shows the event study results in non-urban 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. Both openings of grocery stores in urban areas and non-urban areas increase foot traffic to businesses in affected areas by up to 40% in 6–10 months after openings.

Traditional views are concerned about the competitive effects of big-box retailers on nearby businesses. To test this idea, we examine the heterogeneous treatment effects by type of opening grocery stores and surrounding grocery stores. We divide the openings of grocery stores into 3 types: national grocery store chains (examples include Safeway and Publix), big-box retailers (examples include Walmart, Target, and Costco), and discounters (examples include 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 11. Our findings suggest big-box retailers and discounters compete against national

[17] 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 urban group, and the rest are classified into the non-urban group.

grocery store chains. On one hand, the opening of a big-box retailer nearby reduces foot traffic to a grocery store belonging to a national chain by 27% 6–10 months later. 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%.

4.4 Capitalization of Demand Spillovers on CRE Rents

Do commercial real estate landlords capitalize on the demand spillovers generated by the openings of the grocery store and increase the rents they are able to charge their tenant? In this section, we compare the terms of the new leases signed by tenants close to the openings of the grocery store and those close to the CF sites, before and after openings, respectively, using the following specification.

$$Y_{ijt} = \eta_{it} + \beta_0 \mathrm{Treat}_{ij} + \beta_1 \mathrm{Post}_{ijt} + \beta_2 \mathrm{Treat}_{ij} \times \mathrm{Post}_{ijt} + X_{ijt}\theta + \varepsilon_{ijt} \tag{3}$$

Here, Y_{ijt} is the log rent per square-foot of a newly signed lease j associated with the case of grocery store i's opening in calendar year-quarter t. η_{it} is a fixed effect interacting the opening case and the calendar year-quarter in which the lease is signed. The variable Treat_{ij} denotes the dummy for the treatment group, equal to one if the business associated with the signing of lease j is located within 0.5 miles from a real grocery store opening, and equal to zero if it is located within 0.5 miles from a CF site. $Post_{ijt}$ denotes a dummy equal to one if lease j is signed after the opening of the grocery store i. X_{ijt} is a set of lease-level controls, including the lease term, type of lease (net vs. gross), purpose of space use (retail, office, industrial, etc.), and a dummy for whether the rent is asking rent or effective rent. Our coefficient of interest, capturing the average treatment effect of a grocery store opening on the rents negotiated for new leases for commercial CRE tenants nearby, is denoted by β_2. Our regression sample consists of new leases signed within 0.5 miles from each real opening and from each CF site, during the 4 quarters before the openings to 8 quarters after the openings.

Table 2 column (2) shows the estimated treatment effects on rent capitalization after controlling for characteristics of the leases. On average, grocery store openings increase rents in newly negotiated leasing contracts by 23.8% if they are located within 0.5 miles from the openings, and the effect is significant at 90% level. Applying the inverse propensity score weighting in column (4) yields a similar estimate of 29.7% increase in rents, the effect is significant at 95% level. On average, the rent per square foot for new leases signed before the real openings in the control group surrounding the CF sites is $17. Hence, the estimated treatment effects imply an increase of $4-$5, which is economically significant. We focus on comparing the rents as the outcome for now, and in a future version, we will also compare other terms of the leases. We also plan to investigate how the openings of grocery stores change the business composition and the dynamics of the entry of new tenants and the exits of existing tenants in nearby CRE properties.

5 Conclusion

We investigate the driving forces and effects of agglomeration in the local non-tradable service sector on the value of commercial real estate. 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 find that openings at grocery stores lead to significant growth in foot traffic to its opening locations. The grocery anchors generate substantial positive demand spillovers to nearby businesses by increasing the foot traffic to nearby businesses within 0.1 miles by 39 percent on average 6 to 10 months later. Such demand spillovers decay to zero sharply as a function of distance. The synergies between a new grocery store opening and existing grocery stores within 0.1 miles are positive and substantial, leading to a 60 percent increase in foot traffic 6–10 months later. More broadly, such strong positive spillovers are strongest between grocery store openings and wholesale and retail store, with a 46 percent increase in foot traffic 6–10 months later. Preliminary analysis suggests that CRE values capitalize on the benefits of agglomeration economies caused by grocery anchors. Rents for CRE properties within 0.5 miles of an opening increase by 23.8 percent after the opening.

References

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.

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.

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.

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*, August 2005, *87* (3), 411–422.

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.

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.

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

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.

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.

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.

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.

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.

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.

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

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

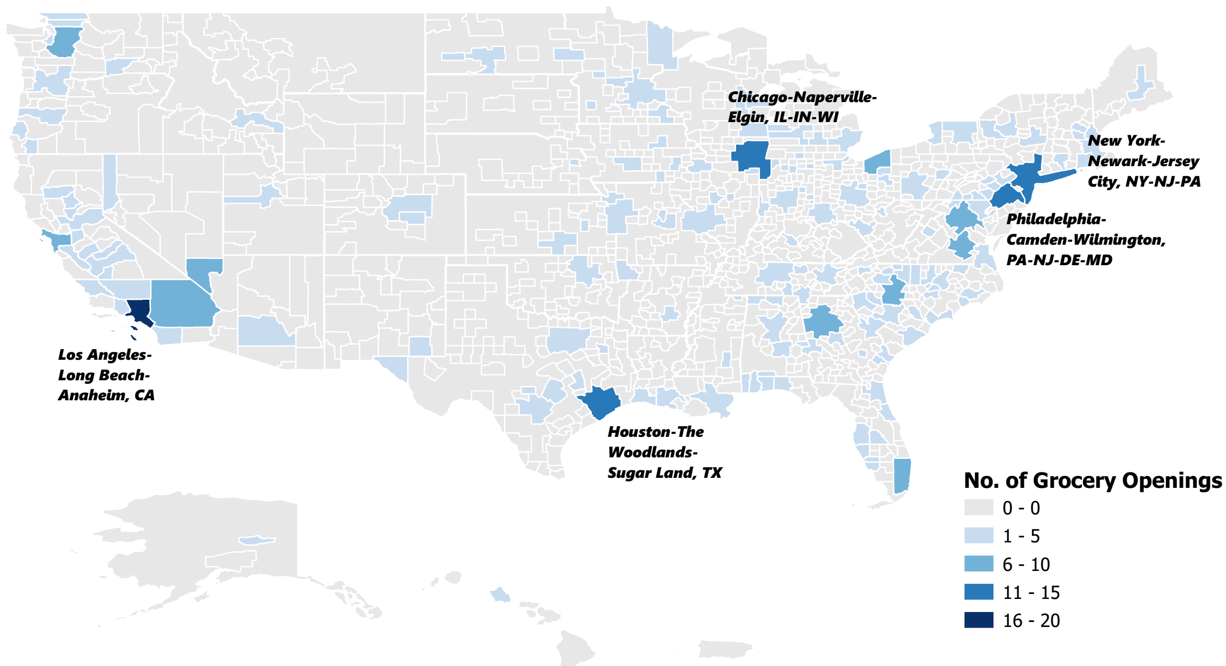


Figure 1: Locations of Grocery Store Openings

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.

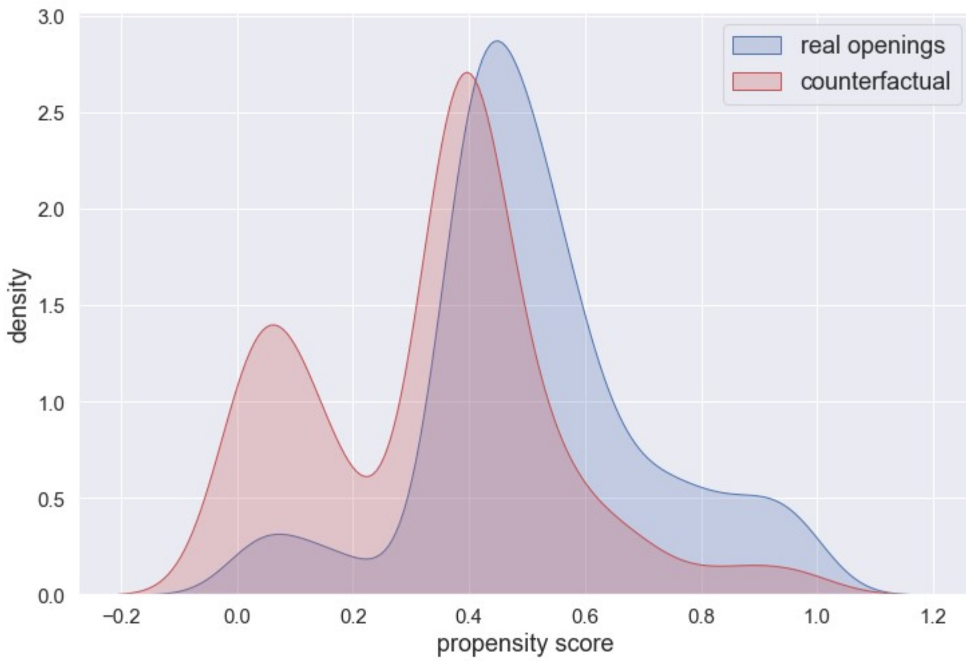
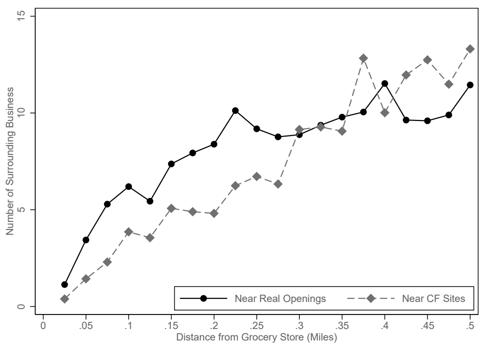
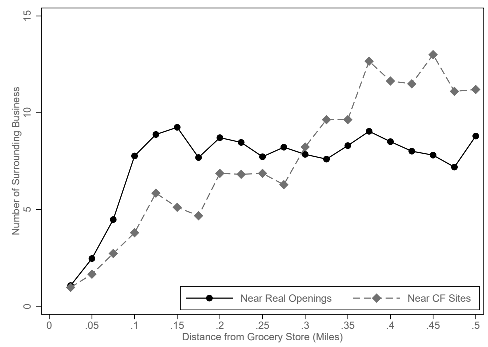


Figure 2: Distribution of Propensity Scores for Real Openings and for Matched CF Sites

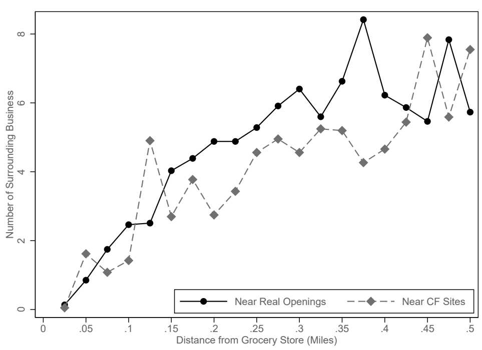
Notes: This figure plots the distribution of estimated propensity scores for real openings and for counterfactual sites matched.



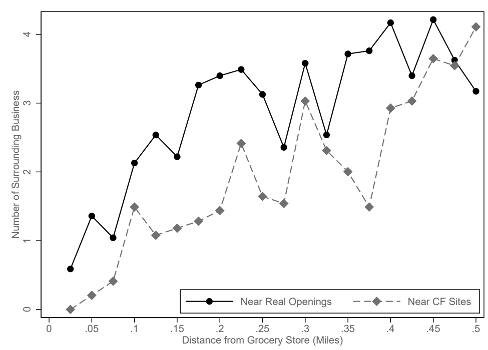
(a) Wholesale and Retail



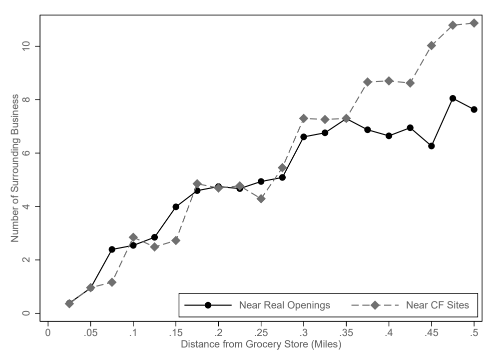
(b) Accommodations, Eating, and Drinking



(c) Medical, Welfare, and Healthcare



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



(e) 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. (2021) to categorize surrounding businesses into 5 categories: wholesale and retail are identified by 2-digit NAICS codes 42, 44, 45; 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.

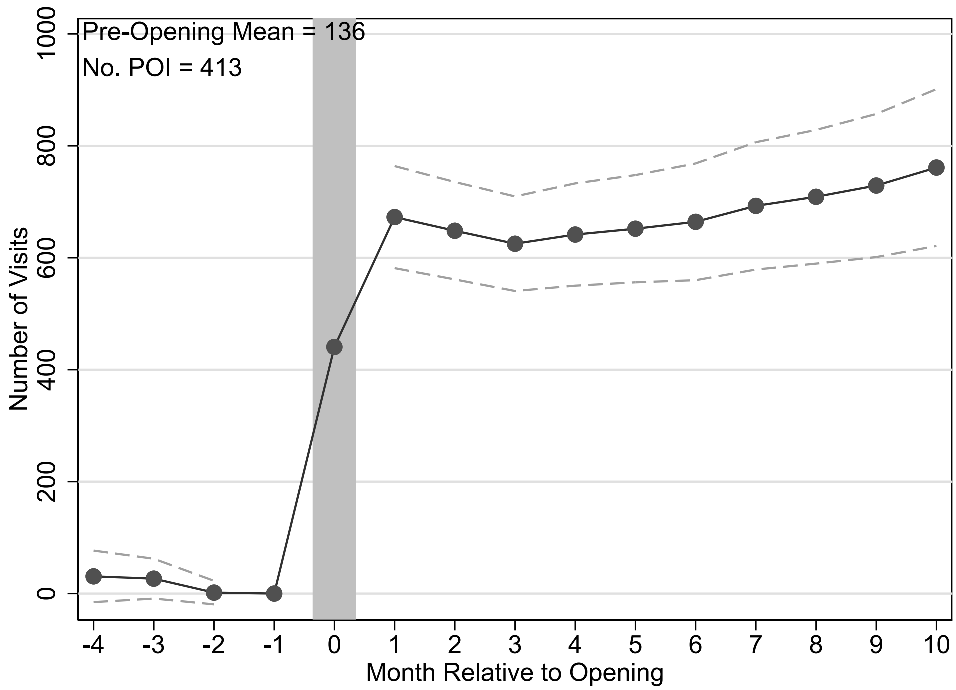


Figure 4: Validation of Opening Dates using Foot Traffic Measure

Notes: Sample consists of all grocery stores opening in our sample. This figure plots the estimated coefficients of μ_τ of equation (2), 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.

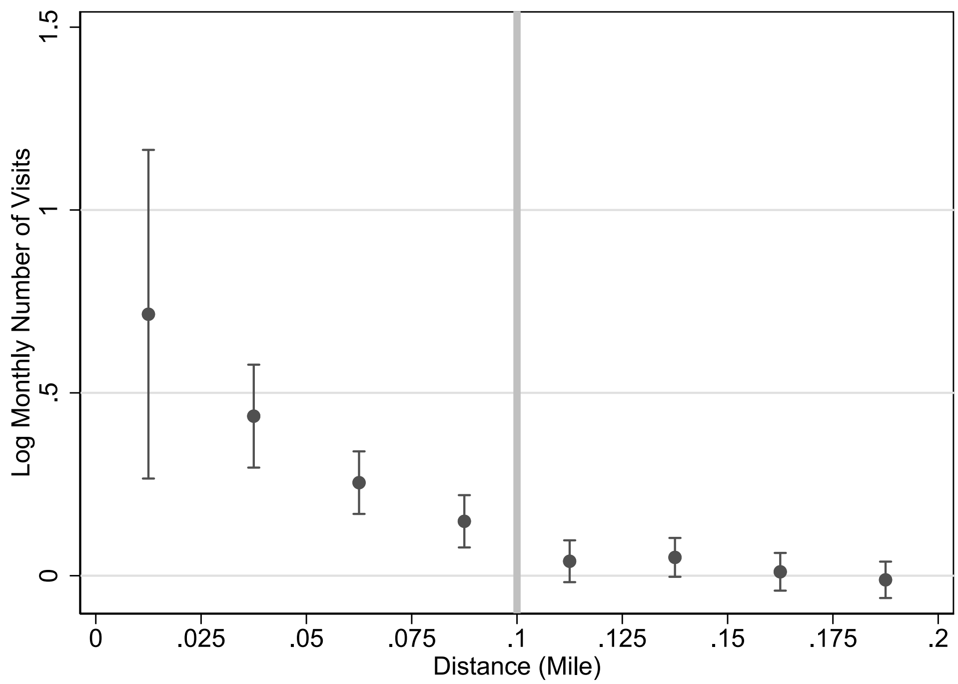
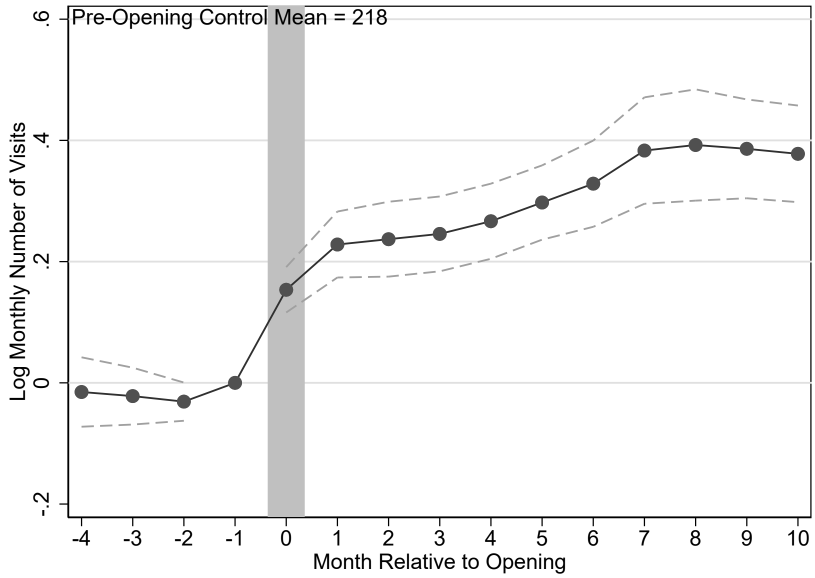
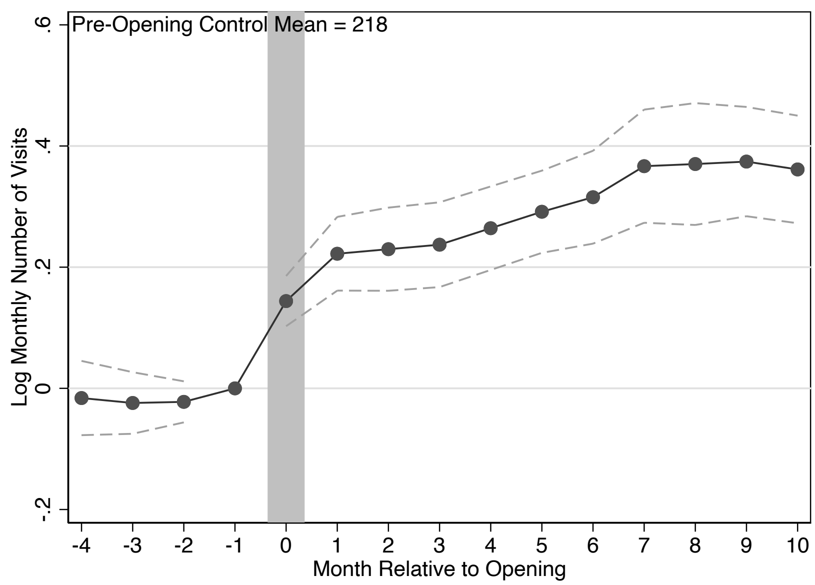


Figure 5: Treatment Effects on Foot Traffic to Surrounding Businesses as a Function of Distance

Notes: This figure plots the average treatment effects of grocery store openings on nearby businesses in each concentric ring in the first 10 months after the openings by estimating equation (1). 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 site 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.



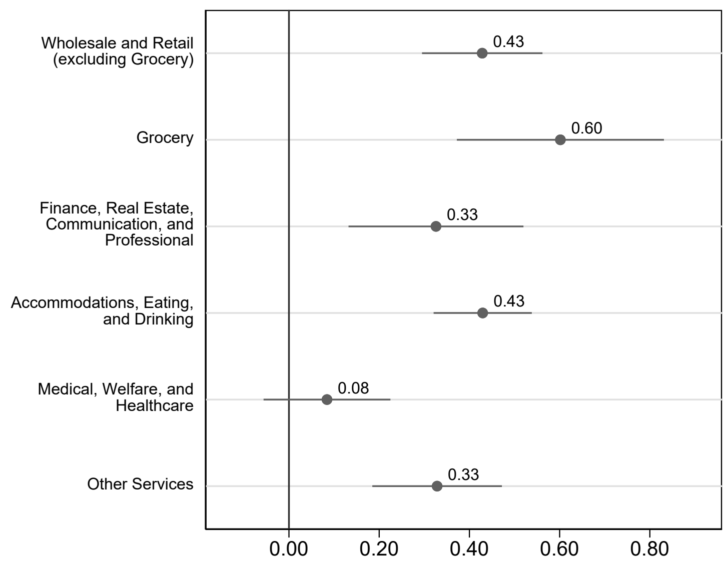
(a) Event Study Results on Surrounding POIs: No Weights



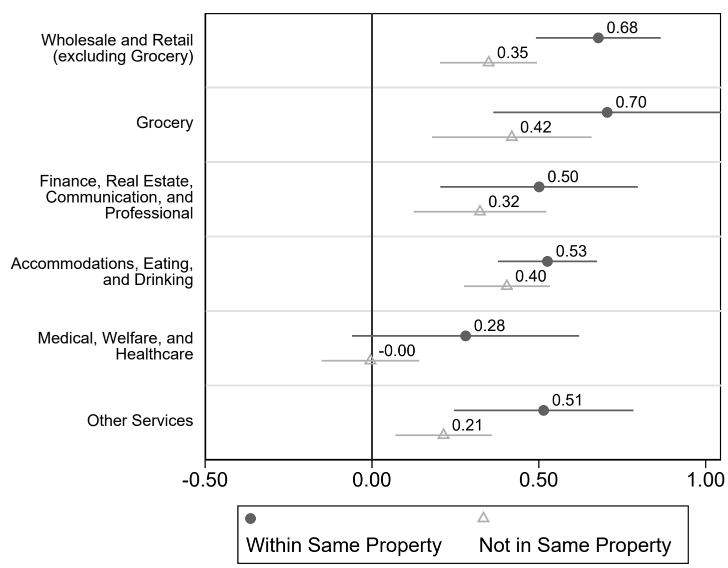
(b) Event Study Results on Surrounding POIs: IPW

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 (1). 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 the coefficients β_τ^g and corresponding 95% confidence intervals for each treatment group by estimating equation (1) and implementing an inverse propensity score weighting.



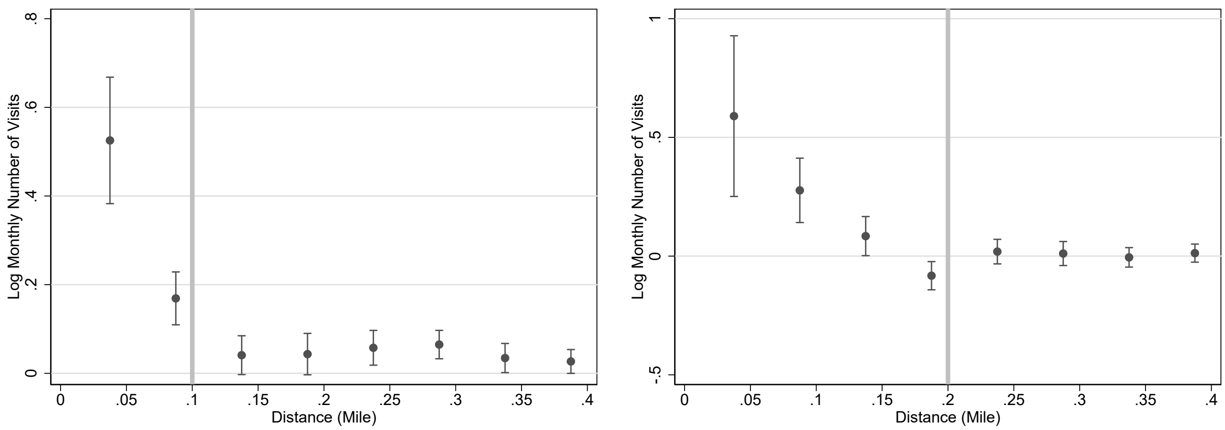
(a) Heterogeneity by Business Category



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

Figure 7: Heterogeneous Treatment Effects on Foot Traffic

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. We report the treatment effects on different business categories by row. We include "Grocery Stores" as a separate category. Examples of category "Finance, Real Estate, Communication, and Professional" include real estate brokerages and banks. Examples of category "Accommodations, Eating, and Drinking" include restaurants and bars. Examples of category "Medical, Welfare, and Healthcare" include dentists and primary care providers. Examples of 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.

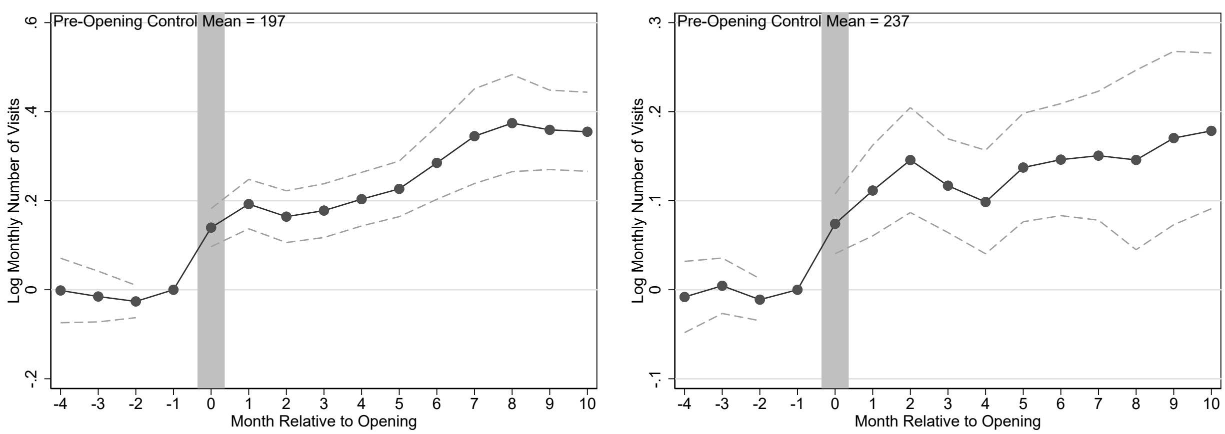


(a) DID Results in Urban Areas

(b) DID Results in Non-Urban Areas

Figure 8: Heterogeneous Treatment Effects on Foot Traffic by Urban-Rural Status

Notes: Panel (a) shows the treatment effects of grocery store openings in urban areas. Panel (b) presents the treatment effects of grocery openings in non-urban areas.



(a) Event Study Results in Urban Areas

(b) Event Study Results in Non-Urban Areas

Figure 9: Event Study Results by Urban-Rural Status

Notes: Panel (a) shows the event study results of grocery store openings in urban 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 (b) event study results of grocery store openings in non-urban 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.

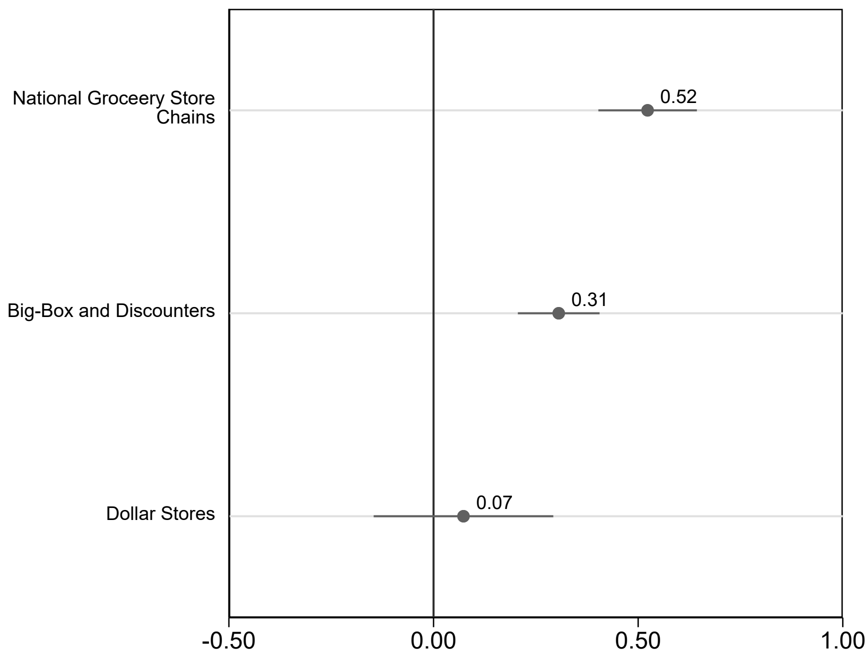


Figure 10: Heterogeneous Treatment Effects on Foot Traffic: by Type of Grocery Stores

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. We report the treatment effects of “National Grocery Store Chains”, “Big-Box and Discounters”, and “Dollar Stores” by row. Standard errors are clustered at the real or CF site level.

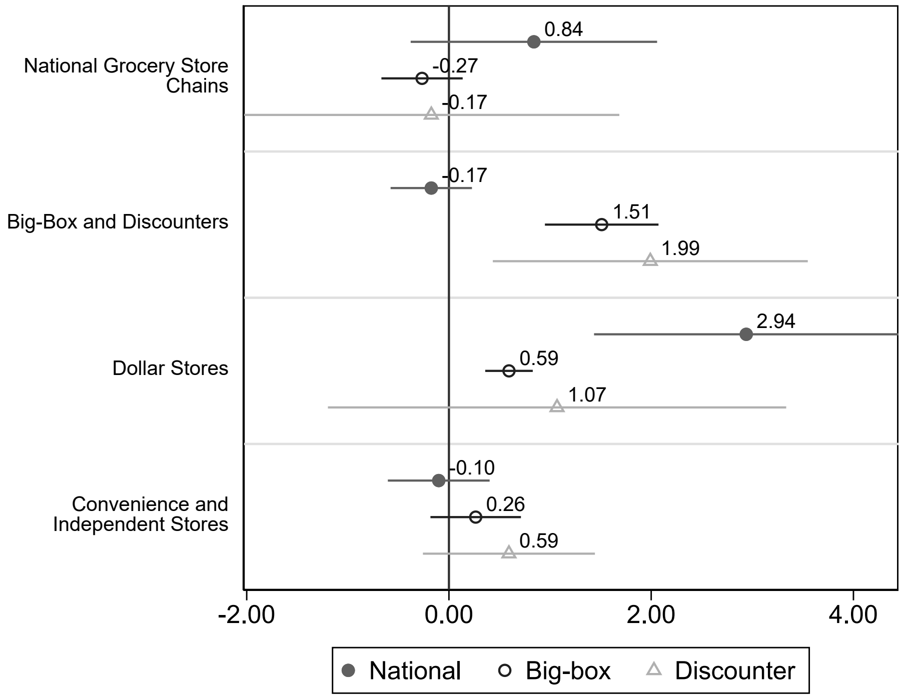


Figure 11: Heterogeneity by Opening Grocery Store Type and Surrounding Grocery Store Type

Notes: This figure plots the heterogeneous treatment effects by the type of grocery openings on the log of monthly visit counts to different types of nearby grocery stores within 0–0.1 miles from the grocery store openings 6–10 months after opening. 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.

Variable	(1)	(2)	(3)	(4)
Post × Treatment	0.262^{**}	0.238^{*}	0.319^{**}	0.297^{**}
	(0.143)	(0.134)	(0.146)	(0.146)
Lease-level Controls	No	Yes	No	Yes
Weights	No	No	Yes	Yes

Table 2: Capitalization of Demand Spillovers on CRE Rents

Notes: This table reports the coefficient β_2 by estimating equation (3). Coefficients β_2 summarize the average effects of an opening on rents negotiated for new leases signed after the opening. The regression sample consists of a cross-section of all leases within 0.5 miles of the real grocery store openings and their matched CF sites, observed from 4 quarters prior to the openings to 8 quarters after the openings. The treatment group consists of all leases that are 0–0.5 miles from the real grocery store openings. The control group consists of all leases that are 0–0.5 miles from the matched CF sites. In all regressions, we include the fixed effects η_{it} which is an interaction between the opening case and the calendar year-quarter of signing the lease. Controls are a set of lease characteristics, including the lease term, type of lease (net vs. gross), the purpose of space use (retail, office, industrial, etc.), and a dummy for whether the rent is asking or effective rent. Columns (1) and (2) report the estimated effects without adding any weights, columns (3) and (4) report the estimated effects after applying the inverse propensity score weighting. For the leases in the treatment group, the weights are inversely proportional to the estimated propensity scores for the opening sites from Section 3. For the control group, the weights are inversely proportional to one minus the estimated propensity scores for the CF sites. Standard errors are clustered at the real or CF site level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Appendix For Online Publication

A Additional Tables **43**

B Additional Figures **47**

C Determination of Opening Month **54**

C.1 Manual Validation of Opening Month 54

C.2 Use Structural Breaks to Identify Opening Month 54

D Use Safegraph Placekeys to Identify POIs in the Same Property **56**

E GAN-based CNN Model for Finding Counterfactual Locations **58**

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

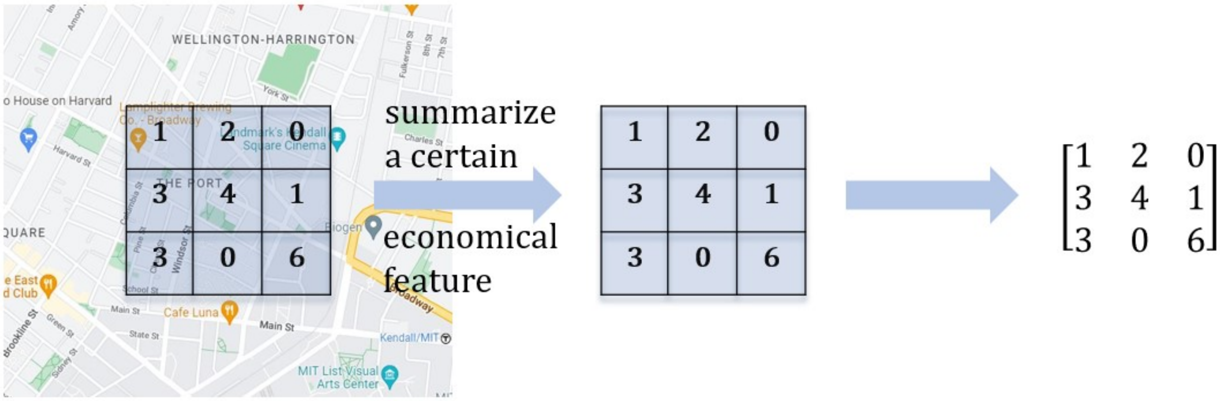
Table A3: Grocery Store Openings by Chain

B Additional Figures

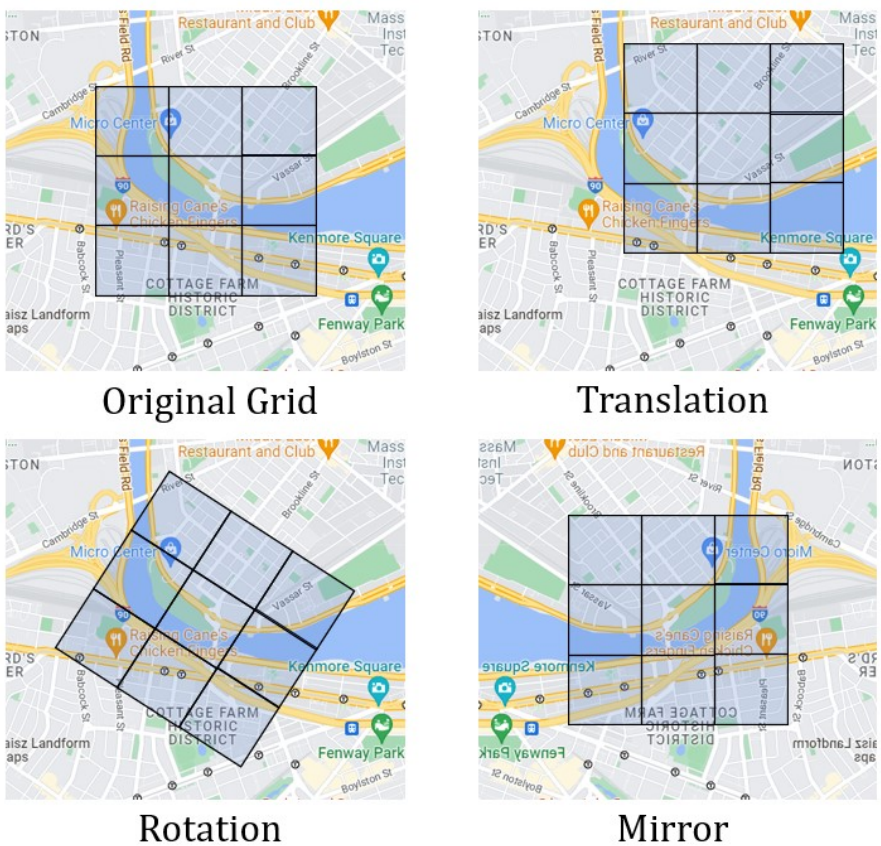
Shopkeeper-concerning Elements	Sub-elements	Variable Description	Source
Neighborhood Demographics and Characteristics	the number of people living in the area	population	Census American Community Survey (ACS) data
		working population	not included
		employment/residential population	not included
		the 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 each characteristics in detail accordingly. The ACS-sourced data is block group level data.



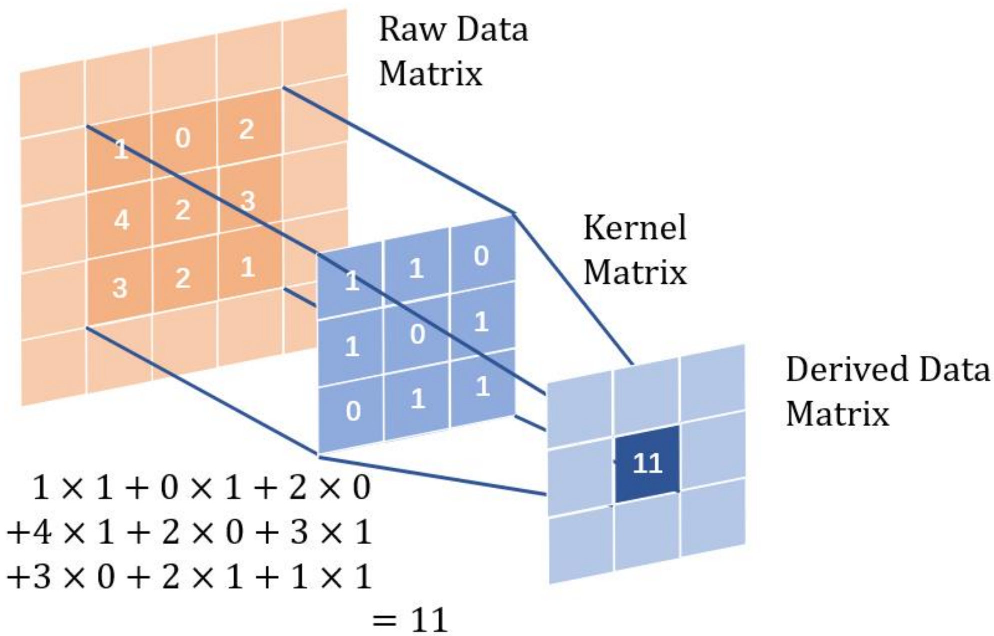
(a) CNN Input Matrix



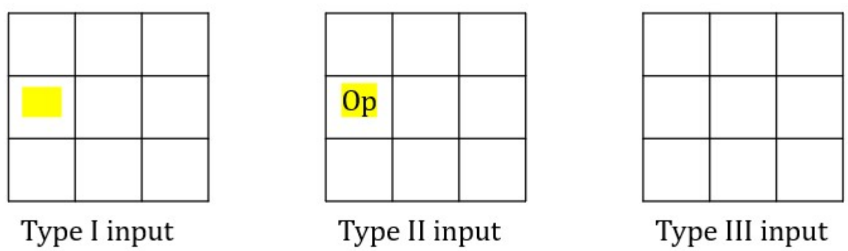
(b) Data Augmentation Methods

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 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 kinds 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 mimicking the an ideal counterfactual location. The second type also consist of real opening. The third type is randomly pick on the map, not necessarily consist any real openings.("Op" means real opening)

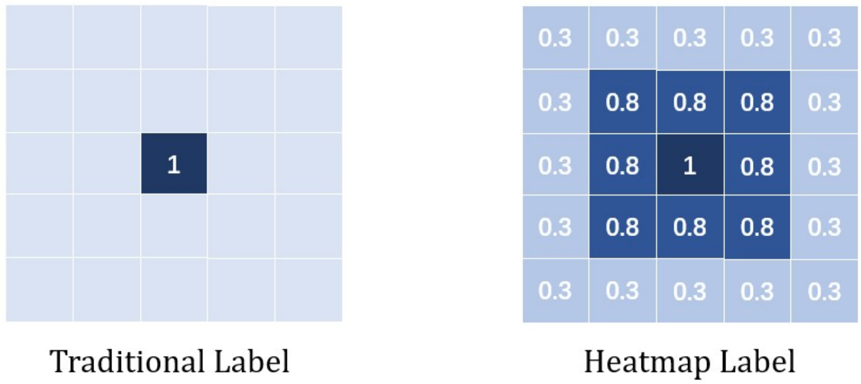
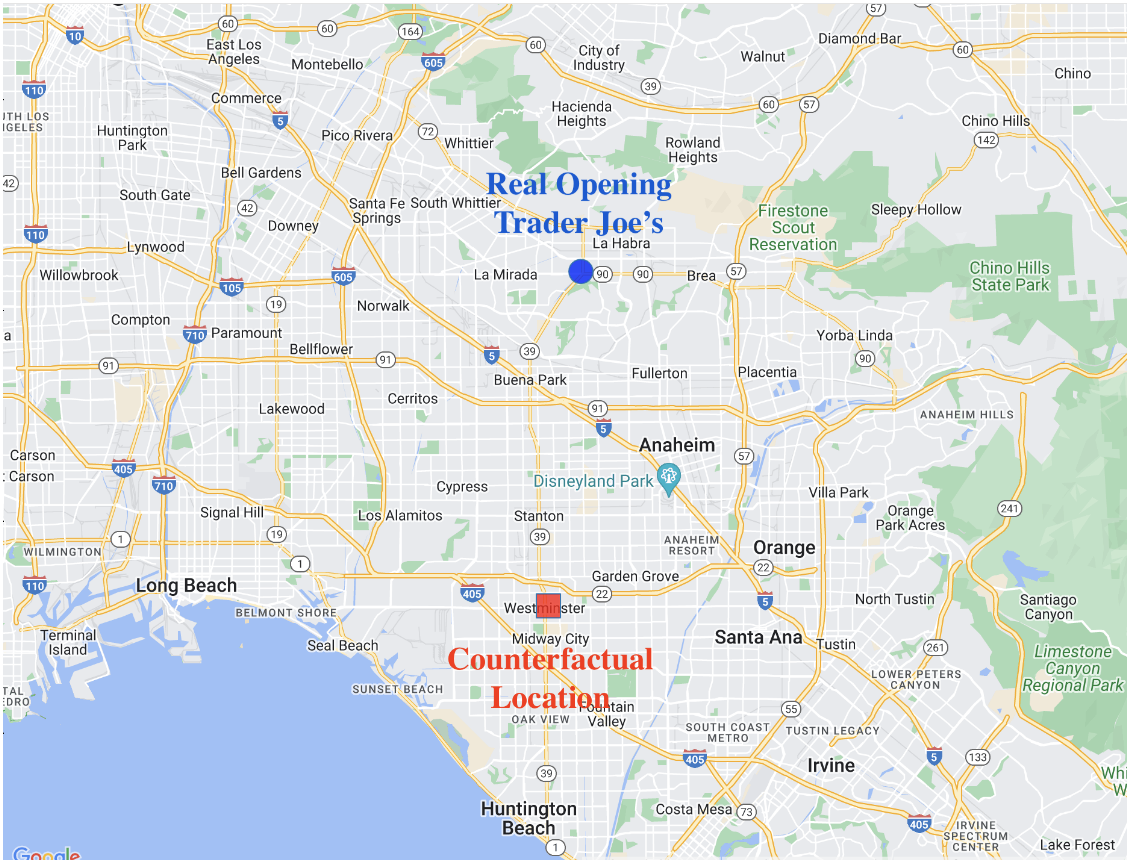
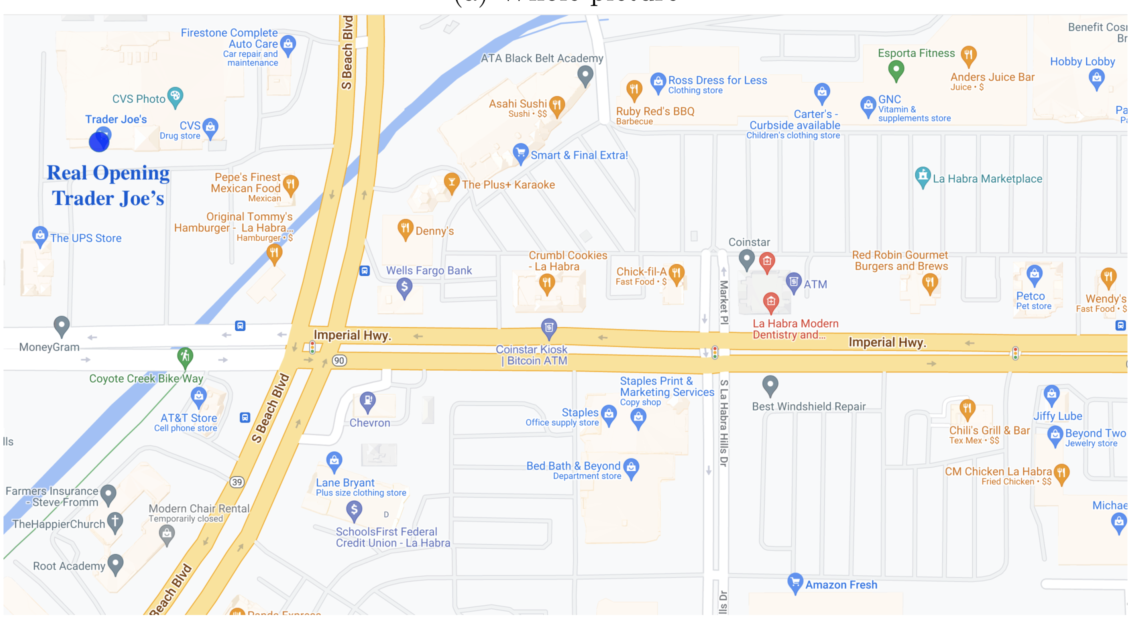


Figure B4: A Traditional label and a Heatmap Label

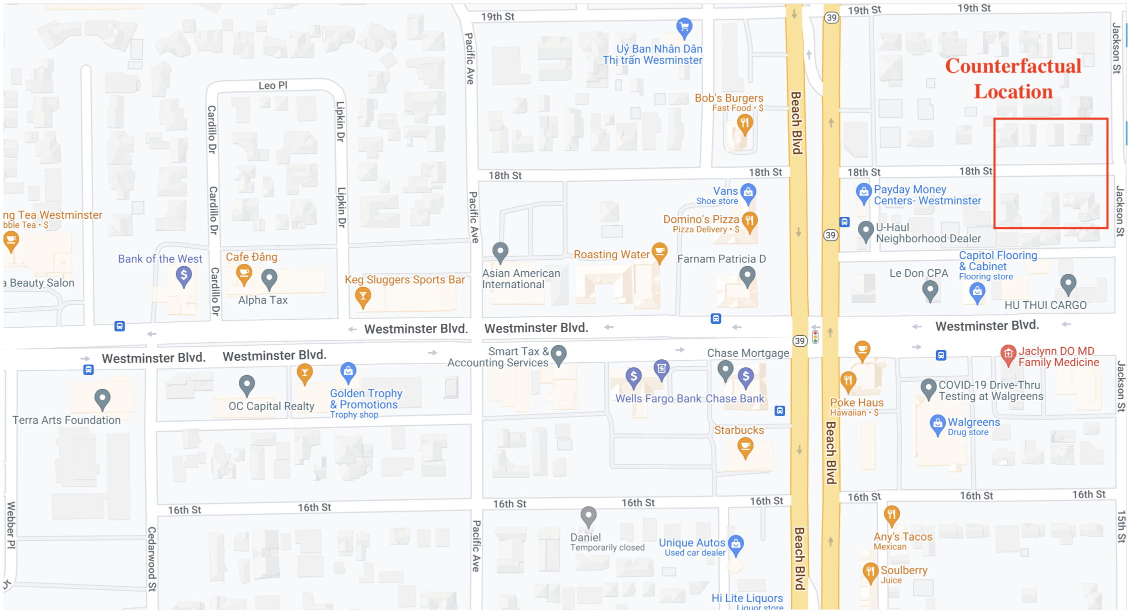
Notes: The traditional label only give a mark at the correct position, and eventually compresses the matrix into a single character, while the Heatmap label assigns a certain mark value to the entire matrix to guide CNN to reflect and learn, and finally optimizes the overall loss function based on Heatmap regression. We assign 1 to the specific location of real openings, and assign decreasing numbers as labels to the locations surround them, the number decay function can vary, here we merely presents 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 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.

C Determination of Opening Month

Noting that the opening months of POIs are inaccurate in many cases, we discuss how we correct the opening month of opening grocery stores.

C.1 Manual Validation of Opening Month

We manually searched the Internet for all new grocery stores in our raw data set for their opening month. We finally successfully found the opening month for 30% grocery stores in the sample. We introduce a method using structural breaks in foot traffic to correct the opening month for the remaining 70% grocery stores.

C.2 Use Structural Breaks to Identify Opening Month

Methodology

We use the series of the monthly number of visits to impute and correct the opening month. For each POI, 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:

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

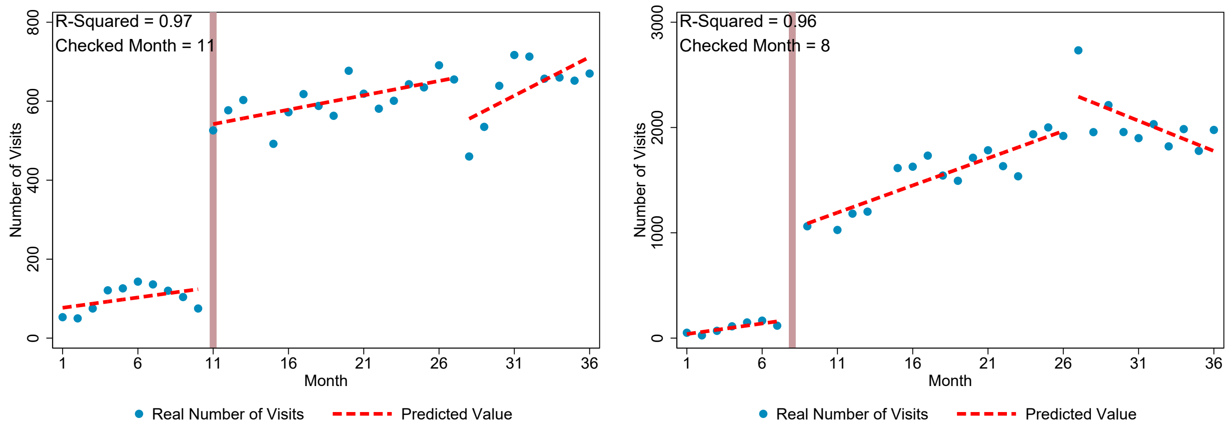
here, N_{it} is the monthly number of visits for POI i in month t, t_i^{1*} and t_i^{2*} are the months of the first and the second structural break, respectively.

After we get the imputed months of structural breaks, we assign t_i^{1*} as the opening month for grocery stores if the R-squared is higher than 0.9.

Validation

There is no guarantee that the imputed opening month is correct for all grocery stores. Thus, we also impute the opening month for the 30% opening grocery stores in our sample and compare the imputed opening dates with the real ones. Among these grocery stores, the differences between the actual and imputed opening months are less than or equal to 2 months, suggesting that the imputation is accurate.

Figure C6 plots two examples of how we use the foot traffic data to impute opening months. In the figures, the red bar represents our imputed opening months. We see that the actual opening month exactly coincides with our imputed one.



(a) Example 1

(b) Example 2

Figure C6: Examples of an Imputed Opening Month

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 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 plackeys of these POIs with the same-address plackey structure, if they exist.
- **Step 3:** we find all placekeys that share the same parent placekeys found in the 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 the 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 D7: 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 the 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.

E GAN-based CNN Model for Finding Counterfactual Locations

TBD.