The Impact of Dollar Store Expansion on Local Market Structure and Food Access

El Hadi Caoui, U of Toronto [*]

Brett Hollenbeck, UCLA Anderson [†]

Matthew Osborne, U of Toronto [‡]

March 15, 2022[§]

Abstract

This paper studies the expansion of dollar store chains in the U.S. retail landscape following the Great Recession (2008–2019). This expansion has been accompanied by growing public concern over the impact on local retail markets and food accessibility in local communities. We develop an empirical framework to evaluate the efficiency of the free entry equilibrium and impact of entry regulation on spatial market structure. A dynamic game of entry, exit and investment into spatially differentiated locations is specified, allowing for chain-level economies of density. Reduced-form evidence and counterfactual simulations indicate that dollar store chains compete strongly with the grocery and convenience segments and that dollar store expansion has led to a significant decline in the number of grocery stores in many markets.

Keywords: retail industry, dollar store, dynamic games

JEL Classification:

[*] elhadi.caoui@rotman.utoronto.ca

[†] brett.hollenbeck@anderson.ucla.edu

[‡] matthew.osborne@rotman.utoronto.ca

[§] Acknowledgements and notes.

1 Introduction

The rapid growth of the dollar store retail format over the past several decades has had broad and dramatic effects on the retail landscape in the United States. Following previous waves of growth by large retail chains, the three main dollar store chains (Dollar General, Dollar Tree, and Family Dollar) have become in some ways the dominant mode of retailer in many markets, with implications for competition, affordability, convenience, and food accessibility for much of the population. These chains distinguishing features are the use of single or limited price points, particularly selling most goods for $1, and assortments consisting of small serving-size basic consumables, clearance or irregular goods, and a lack of perishable grocery items.

Beginning in the 1950's, these chains exhibited slow but steady growth over the decades that followed, establishing themselves primarily in small towns in rural areas. Following the 2008 recession, several events combined to accelerate their growth. The recession itself may have made low-price, small-format consumables more attractive for many by worsening household finances. In 2007, the largest chain, Dollar General, was bought out by a private equity firm who rationalized location strategy, cut costs, and set out on a rapid growth strategy. In 2015, the two smaller chains, Dollar Tree and Family Dollar, merged and sought to expand their presence to compete with Dollar General.

Consequently, the growth of this format has been and continues to be exceptionally rapid. In 2021, there were more of these stores operating than all the Walmarts, CVS, Walgreens, and Targets combined by a large margin. During the period 2018-2021, roughly half of all retail stores that opened in the U.S. were dollar stores and these chains were collectively opening stores at the rate of 3.75 stores a day over the past decade. The growth of these chains has raised a number of policy issues, in particular, many local policymakers have expressed concerns that the rapid entry of dollar store chains in their cities have forced out local independent retailers, including small grocery stores. The latter is often considered especially concerning to the extent it reduces access to produce and other perishable food items for low income residents, creating "food deserts." These concerns and others have lead many localities to ban dollar store chains from entering or pass dispersal regulations limiting the number of dollar stores that can enter.[1]

Broadly speaking, the arguments for and against dollar store chains fall along these lines. Proponents argue that they introduce additional choice into under-served retail markets, that they offer lower prices than their competitors, and that their strategy of entering in

[1] A partial list of cities that have banned dollar store entry or passed ordinances restricting the number of dollar stores that may enter includes: Birmingham AL, Atlanta GA, New Orleans LA, Akron OH, Oklahoma City OK, Tulsa OK, and Fort Worth TX. See https://ilsr.org/rule/dollar-store-dispersal-restrictions/

low-rent areas and opening many stores in the same market results in greater convenience for customers who can make short trips for specific items rather than long trips to the nearest big box store, which could be a large distance away. Opponents argue that the aggressive entry strategy of dollar store chains has led to an inefficiently large number of these stores in many markets, that they force out local independent retailers, that they reduce access to fresh food products as a result, and that their pricing and small-item product format masks the extent to which their prices are in many cases higher than other retailers who sell larger volume products with quantity discounts. Yet despite the widespread public and policy debate, the academic literature has yet to study these claims or the effects of dollar store chain expansion more generally.

Our research objectives are: first, to document the extent and nature of growth of dollar store chains and place this in the context of the broader retail landscape and the economic study of the growth of chain retailers over the past decades. Second, we measure the effect of dollar store entry in a location on the number of independent retailers and small grocers. Third, we use detailed data on consumer purchases to measure the effects of dollar store entry on expenditures, prices paid, convenience and travel costs, and food access. Fourth, we evaluate what effects the anti-dollar store policies implemented by some localities would have if scaled up nationally, in particular we study proposed bans on dollar store expansion to understand how they would effect the broader retail landscape. Finally, we examine if the amount of dollar store expansion is inefficient from a total surplus perspective.

We use a combination of methods to answer these questions. We begin by compiling multiple datasets to comprehensively document dollar store chain growth. We use reduced form event study methods to test hypothesized effects of dollar store entry on local market structure and consumer behavior. To evaluate proposed policies and quantify the long-term equilibrium effects dollar store expansion, as well as to measure entry costs, density economies, and cannibalization effects, we estimate a dynamic model of store entry, expansion and exit choices. We then use the model estimates to solve and simulate the dynamic oligopoly game played between retailers to evaluate counterfactual policy scenarios.

We use data from several sources. We track the number and type of retail stores, including dollar stores, across the U.S. using the Supplemental Nutrition Assistance Program (SNAP) Retailer panel, a yearly panel of SNAP-authorized retailers from 2008 to 2019. We combine this with data on dollar store distribution center locations and opening dates, as well as data from Nielsen TDLinx on average per-store revenue of different store formats. We also match data on store openings and closings to the Neilsen and IRI homescan panel, which contains individual-level panels on all retail purchases for a nationwide sample of consumers.

We first use an event study design to study the effects on local market structures of dollar store entries. We control for market-level demographics and demographic trends, as well as market-year fixed effects, and consistently find that dollar store entries are associated with a significant decrease in the number of independent local grocery stores. The effect size is equivalent to roughly 25% of the pre-entry number of grocery stores, when measured in the area 0-2 miles around the entry location. For the region 5-10 miles away from the entry location, by contrast, there is no change in the number of grocery stores.

Next, we estimate a structural model of the dynamic game being played between dollar store chains and their local competitors. The goal of this model is to provide estimates of the size and nature of competitive effects between different store types and to evaluate how market structure would evolve under counterfactual policy scenarios. We can thus quantify the total reduction in the number of grocery stores, convenience stores, etc., across different market types if dollar store chain expansion had not occurred. We model each store type's entry and exit decisions in a dynamic oligopoly game following Ericson and Pakes (1995) but with a spatial component along the lines of Seim (2006). We therefore follow a tradition in the study of market structure impacts of retail chains that includes Jia (2008), Holmes (2011), Ellickson et al. (2013), Zheng (2016), Igami and Yang (2016), and Hollenbeck (2017). The key challenges in modeling this game in a tractable way are the complex nature of spatial competition and the non-stationarity that results from the large growth over time in the dollar store chains.

We take advantage of the fact that firms face a terminal choice when deciding whether or not to exit, which generates a type of finite dependence (Arcidiacono and Miller (2011), Arcidiacono and Miller (2019)). This property simplifies estimation of the game substantially as it allows us to represent the firms' value functions directly in terms of the period-ahead probability of making the terminal choice. We leverage this property and estimate the model using the linear IV strategy of Kalouptsidi et al. (2020). This paper combines insights from the finite dependence approach and the GMM-Euler approach of Aguirregabiria and Magesan (2018) to propose a method that circumvent integration over the high-dimensional state space. We extend this method from single-agent problems to dynamic games, highlighting and addressing an important selection problem arising in games, and apply it in a setting with long-lived chain entrants.

Our estimation results suggest that dollar store chains have substantially lower costs of opening a new store than their independent rivals. They are also substantially more profitable and grocery store profits are significantly harmed by the presence of nearby dollar stores and convenience stores, with most of the effects for stores in the 0-2mi radius. We also find that, within dollar store chains, in the 0-2mi radius there is a strong demand cannibalization effect but in the 2-5mi range this effect is reversed and chains benefit from scale economies, likely working through lower operating costs.

We use these estimates to evaluate the impact on local market structures of a hypothetical ban on dollar store expansion beginning in 2010. We find that, in the counterfactual scenario, markets have on average more than 50% more independent stores, with 1.54 stores per market as compared to just under 1. This includes 56% more combination and grocery stores and 54% more convenience stores. We can also examine how these changes vary across different market types. We find that the largest impacts on number of grocery stores comes in lower income markets, those with larger shares of minority populations, those with higher poverty rates, and fewer households with access to vehicles.

Related Literature. This paper contributes to three strands of the literature. The first strand studies the evolution of the U.S. discount retail sector. This literature has documented the impact of big box retailers (e.g., Walmart, K-Mart) and the supercenter format on market structure and competition (Jia (2008), Zhu and Singh (2009), Basker and Noel (2009), Ellickson and Grieco (2013), Grieco (2014)), on labor markets (Basker (2005)), and the role of chain and density economies (Holmes (2011), Ellickson et al. (2013)). We contribute to this literature by studying the more recent rise of the dollar store format, which has outpaced other discounters by a large margin and who raise distinct policy questions, in particular regarding food accessibility.[2]

Second, this paper is related to the literature studying consumers' grocery shopping behavior and food accessibility. There is an extensive literature studying nutritional inequality in the U.S., with studies focusing on price-, access-, and nutrition education-based interventions (Levi et al. (2019)). Studies of food access have focused on introduction of grocery stores to markets designated as "food deserts", with case studies around individual store entries having found mixed results (Cummins et al. (2005), Cummins et al. (2014), Elbel et al. (2015), Dubowitz et al. (2015), Liese et al. (2014), Rose and Richards (2004), Ver Ploeg and Rahkovsky (2016), Weatherspoon et al. (2013).) Notably, Allcott et al. (2019) study a large number of grocery store entries and find they have only small effects on nutrition of nearby consumers and that nutritional inequality in the U.S. is largely explained by demand factors rather than limited food access, with differences in access and prices explaining only about 10% of nutritional inequality. Levi et al. (2020) finds that access to grocery stores impacts fruit and vegetable spending by affecting shopping frequency, but only among households with a low value of nutrition and at distances of less than 1 mile. We contribute to this literature by studying large numbers of dollar store and grocery store entry events, as well

[2]Hard discounters, such as Aldi, have also been expanding more recently although their presence in the U.S. remains quite limited. These stores differ from dollar stores by offering fresh produce, more limited product assortments, and a higher fraction of private labels. Chenarides et al. (2021) study the price impact of hard discounters' entry. In our paper, we include hard discounters in the supermarket/supercenter category.

as large numbers of grocery store exits, to measure the impact of these events on a range of outcomes including spending on produce and frequency of shopping trips.

Finally, this paper is related to the literature using dynamic games to study the market structure impacts of retail chains (Arcidiacono et al. (2016), Zheng (2016), Igami and Yang (2016), Hollenbeck (2017), Beresteanu et al. (2019)) We depart from the existing literature in two ways. First, we account for non-stationarities inherent to the discount store industry. Over the sample period, dollar store chains have been growing their networks of distribution centers; incorporating this dynamic aspect of the industry is clearly important to better match observed entry patterns. Second, most of previous dynamic game studies (Zheng (2016) being an exception) abstract from the spatial nature of retail competition. Because retail location choices are crucial in shaping the competitive environment (Ellickson et al. (2020)), we model firms' entry decisions into spatially differentiated location as in Seim (2006).

2 Data, Industry Background, and Descriptive Statistics

In this section we describe the history and nature of the dollar store chains, describe our data sources, and provide some descriptive statistics on the industry.

Dollar General originated the dollar store concept in 1955, selling a wide selection of low-cost basic goods at a single $1 price point. The format became popular and a number of competing variety retailers adopted it, including Family Dollar, founded in 1959. Through decades of steady growth and consolidation among competing chains, by the 2000s there remained three major dollar store chains: Dollar General, Family Dollar, and Dollar Tree. These chains distinguish themselves from other retailers by offering low prices in the form of a single price point or a limited number of round number price points.

Unlike other discount retailers like Aldi, they do not achieve their discounts by offering small selections and a large share of private labels. Instead, they offer moderately sized selections and a mix of major brand products and private labels. The stores are buil in the 8,000-12,000 sq ft range and carry 10,000-12,000 SKUs. They also save costs by employing few employees and not offering perishable goods. They primarily sell basic consumables in small formats, seasonal products, and irregular or outdated products off-loaded by major brands. Another distinguishing feature is their market entry strategy, with a focus on small and low income markets under-served by big box retailers. We discuss these markets in greater detail below.

The dollar store chains have grown rapidly over the past several decades, and particularly so after the recession of 2008. By 2021, Family Dollar operated roughly 7,100 stores, Dollar General operated 18,000 stores, and Dollar Tree owned 4,350 stores. The combined nearly 30,000 stores are substantially more than the number of Wal-Marts (5300 stores), Targets (1900 stores), CVS (9900 stores), and Walgreens (9300 stores) combined and is significantly larger than the number of Subway restaurants (21000 restaurants), the largest U.S. restaurant chain and is similar to the number of worldwide Starbucks locations.

We combine several data sources to study dollar store expansion and the effects on consumers and local market structure. The first is the SNAP Retailer panel, a yearly panel of SNAP-authorized retailers from 2008-2019. This data contains information on 400,000 U.S. retailers including their chain affiliation and store type, including for small independent retailers. We also collect data on distribution centers of dollar store chains over time, namely the locations and opening dates for the three major chains.

The SNAP retailer panel contains any store that accepts the SNAP program. In addition to dollar stores, this includes convenience stores, combination stores (stores selling general merchandise and food products), grocery stores, and supermarkets. Table 1 shows store counts by type in the SNAP panel. The primary benefits of this data source are that it is an annual measure and contains nearly the full universe of retailers in this industry.

Table 1: Number of SNAP retailers by type (all U.S.)

Store type	Number of stores
Grocer	65,240
Supermarket/center	51,695
Small retail	283,140
Combination Grocery/Other	78,174
Convenience Store	204,966

Note: Combination grocery/Other includes dollar stores and drug stores. Convenience stores include gas stations.

We also compile data from the Neilsen and IRI homescan panel, in particular we use the IRI MedProfiler data from the FDA. This contains individual-level panels on all retail purchases for a nationwide sample of consumers. The MedProfiler also contain nutritional information for food purchases and consumer health metrics, such as BMI.

We collect market-level data on demographic characteristics from the Census and ACS at the Census Tract level. This allows us to study the types of markets dollar stores enter, how this evolves over time, and how the effects of dollar store entry vary across different demographic groups. We focus our later analysis on small and isolated markets, which we define as cities and incorporated places with populations between 5,000 and 200,000, and excluding markets within 10 miles of a city with population $\geq$ 5000 or within 20 miles of a city with population $\geq$ 25000. Within each market are locations, which we define at the Census Tract level.

Table 2 below shows the average demographics markets that contain dollar stores and Table 3 shows statistics on their market structure. As Table 2 shows, these markets are small in terms of population and with low average incomes. Average income per capita is $20,350 compared to roughly $58,000 for the U.S. as a whole. A typical market contains 2 dollar stores but with wide variation across markets. Markets contain 4.5 locations, on average, in which a store can enter.

Table 2: Descriptive Statistics: Markets and Locations (2010-2019)

Variable	Mean	Median	Std.Dev	Min	Max
Market-level characteristics					
Population	14,146	10,430	11,714	3,160	124,950
Income per capita (past 12 months)	20,352	19,779	4,617	7,796	86,593
Residential rents	624.8	593.6	135.8	318.1	1,801.0
Land area (sq mi)	15.2	10.1	20.0	1.6	301.7
Distance to closest distribution center (mi)	262.7	188.2	188.6	31.6	1,132.3
Number of locations	5.8	5.0	4.3	1.0	30.0
Number of commercial locations	4.3	3.0	3.2	1.0	28.0
Observations (Market-Year)	8,460				
Location-level characteristics					
Population	2,435	2,327	1,953	1	13,586
Income per capita (past 12 months)	21,121	20,546	6,661	2,183	112,495
Residential rents	640.6	611.7	161.4	189.4	2,134.7
Land area (sq mi)	2.6	1.6	5.5	0.0	165.0
Observations (Market-Location-Year)	49,150				

Note: Distance to closest distribution center is the average over the top three chains. "Number of locations" corresponds to both residential and commercial locations. Commercial locations are those in which at least one store (including gas stations, drugstores, and supermarkets) was active in any year between 2008 and 2019.

In Table 2 we show how the total number of stores evolves over time in these markets over the period 2010-2019. The first column shows the total number of stores in the major dollar store chains, which increases by 36% during this period. The number of independent retail stores (noted "single-store") also increases, with this growth driven by convenience stores. The number of combination stores is roughly flat and the number of grocery stores falls by 11%. The number of "Other" dollar stores that are not in the major chains is very small and so we exclude these from our analysis of dollar stores.

Table 3: Descriptive Statistics: Stores (2010-2019)

Variable	Mean	Median	Std.Dev	Min	Max
Market-level characteristics					
Dollar stores (DG, DT, FD)	2.75	2	2.25	0	21
Grocery and Combination stores	1.42	1	1.86	0	20
Convenience stores	4.71	3	5.34	0	52
Gas stations	2.85	2	2.76	0	24
Drug stores	1.41	1	1.42	0	11
Supermarkets/Supercenters	3.50	3	2.47	0	21
Observations (Market-Year)	8,460				
Location-level characteristics					
Dollar stores (DG, DT, FD)	0.64	0	0.82	0	5
Grocery and Combination stores	0.33	0	0.66	0	7
Convenience stores	1.10	1	1.36	0	11
Gas stations	0.67	0	0.87	0	8
Drug stores	0.33	0	0.61	0	3
Supermarkets/Supercenters	0.82	0	1.05	0	7
Observations (Market-Commercial Location-Year)	36,230				

Table 4: Store counts by firm type in selected markets

Year	Dollar Store	Single-store	Grocery	Combination	Convenience	Other DS
2010	2003	4270	880	270	3088	32
2011	2137	4841	933	321	3556	31
2012	2217	5234	938	372	3883	41
2013	2333	5371	879	364	4089	39
2014	2432	5663	880	376	4367	40
2015	2466	5602	823	339	4407	33
2016	2555	5810	839	352	4609	10
2017	2615	5849	846	359	4634	10
2018	2691	5739	846	339	4544	10
2019	2724	5368	781	312	4265	10

Note: Chains correspond to the top 3 dollar store chains. Single-store firms are broken down into: grocery, combination stores, convenience stores, and other dollar stores.

3 Reduced Form Results on Market Structure

In this section, we present evidence on the impact of dollar store chain entry on local retail markets. Our primary goal is to evaluate whether or not dollar store chain entry is associated with decreases in the number of local independent retailers and grocery stores. To do so, we rely on the data described above containing the annual universe of retailers and study the rapid expansion of dollar store chains between 2008 and 2019.

During this time period, we observe 721 dollar store chain entry events across 846 isolated markets. Using small and isolated markets produces a clean setting without significant spillovers across markets in competitive interactions and strategic behavior. We break markets into sub-units based on Census tracts as described above, as well as areas defined by a geographic proximity to each entry location using radii of 0-2m, 2-5m and 5-10m. Our outcome of interest in these regions is the number of independent retailers falling into the combination store category and the number of independent grocery stores.

Our identification strategy for measuring the effects of dollar store chain entry on these outcomes is to use tract-level fixed effects to account for time-invariant unobserved market characteristics and city-year fixed effects to account for time-varying trends at the market level. We also incorporate demographic variables defined at either the tract level or market level. These are intended to control for local trends in population or income associated with economic shocks. We use the tract-level and market-level population, median income, and level of residential rents. We also test a specification that incorporates the annual growth rate in these variables in addition to their level.

We estimate effects using the following specification:

$$y_{mt} = \beta x_{mt} + \delta DS_{mt} + \lambda_m + \alpha_t + \epsilon_{mt} \tag{1}$$

Here x_{mt} are local demographics, λ and α represent market and time fixed effects, and the object of interest δ is the coefficient on a dummy for whether or not a dollar store chain is active in location m in period t.

In Table 5 we show results for different specifications of controls and fixed effects where the outcome variables are the number of local combination retailers (columns 1-4) and the number of independent local grocery stores (columns 5-8.) We find a consistent null result of dollar store entry on independent combination stores across specifications. By contrast, we find a consistent negative effect on the number of independent local grocery stores. For all specifications, entry of a dollar store is associated with a decrease in grocery stores of roughly -.027.

In Table 6 we focus on outcomes for independent grocers and study how the effects of dollar store chain entry vary across different regions defined by the distance from the entry location. In addition to our baseline effect where the outcome is the number of grocers in the location itself, where location is defined as a Census tract, we also include regions defined by radii of 0-2m, 2-5m, and 5-10m.[3] The results show a substantial fall in number of grocery

[3] Each demographic variable is calculated for the region where effects are being measured.

Table 5: Effects of Dollar Store Entry on Independent Local Retailers

	# Combination Stores			# Grocery Stores		
	(1)	(2)	(3)	(4)	(5)	(6)
Dollar Store Active	0.0048	0.0013	0.0031	-0.026*	-0.031*	-0.031*
	(0.008)	(0.009)	(0.010)	(0.013)	(0.014)	(0.015)
Location Population		-0.0000013	-0.000012		-0.0000067	-0.000020
		(0.000)	(0.000)		(0.000)	(0.000)
Residential Rent		0.000061	0.000073		0.000074	0.00011*
		(0.000)	(0.000)		(0.000)	(0.000)
Income per Capita		-0.0000017	-0.0000018		0.0000013	0.0000011
		(0.000)	(0.000)		(0.000)	(0.000)
Population % Change		-0.0076	-0.0059		0.038	0.087
		(0.031)	(0.034)		(0.043)	(0.048)
Rents % Change		-0.023	-0.026		-0.036	-0.033
		(0.015)	(0.019)		(0.022)	(0.027)
Income % Change		0.011	0.0028		-0.049*	-0.030
		(0.018)	(0.022)		(0.025)	(0.030)
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes		Yes	Yes	
Market*Year FE			Yes			Yes
Observations	36230	32607	32040	36230	32607	32040
R^2	0.668	0.687	0.759	0.777	0.794	0.840

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

stores in the 2 mile radius around where the dollar store entry occurs. In the area 2-5 miles from the dollar store entry there is a smaller and statistically insignificant negative effect and in the area 5-10 miles away there is an insignificant positive effect.

Three conclusions follow from these results. First, the negative effect of dollar store entry on grocery stores that we find is not spuriously driven by larger market-level or regional economic shocks. Second, shopping patterns for dollar stores and independent grocery stores seem to take place primarily over fairly small distances. And third, in the local area in which a dollar store entry takes place the effects on grocery stores is quite substantial. The fall in number of stores of .08 is roughly 10% of the mean number of stores in a 0-2 mile area, which is .91.

Next, we perform an event study analysis to visualize these results over the years before an after the dollar store chain entry occurs. In this specification, we estimate:

Table 6: Effects of Dollar Store Entry on Grocery Stores by Distance

	# Grocery Stores			
	(1)	(2)	(3)	(4)
	Tract	0-2m	2-5m	5-10m
Dollar Store Active	-0.031*	-0.063**	0.025	-0.0057
	(0.014)	(0.024)	(0.028)	(0.029)
Population	-0.0000063	0.000051**	-0.000045	-0.00014*
	(0.000)	(0.000)	(0.000)	(0.000)
Residential Rent	0.000074*	0.000048	0.000035	0.000018
	(0.000)	(0.000)	(0.000)	(0.000)
Income per Capita	0.0000014	0.00000023	0.0000036	0.0000036
	(0.000)	(0.000)	(0.000)	(0.000)
Population % Change	0.038	-0.23	0.24*	0.089
	(0.034)	(0.136)	(0.115)	(0.111)
Rents %Change	-0.036	-0.038	-0.081	0.064
	(0.020)	(0.042)	(0.044)	(0.052)
Income % Change	-0.050*	-0.32***	-0.43***	-0.22**
	(0.023)	(0.073)	(0.073)	(0.077)
Census Tract FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	32607	32607	26316	10152
R^2	0.794	0.856	0.895	0.876

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

$$y_{mt} = \beta x_{mt} + \sum_{\tau=-8}^{8} \delta_\tau DS_{m,t-\tau} + \lambda_m + \alpha_t + \epsilon_{mt} \tag{2}$$

This differs from the specification in equation 1 in that now the coefficients on dollar store entry are subscripted by τ, the difference in years measured relative to the entry date. This allows for both dynamic policy effects, such as a delay in the effect on local markets as the dollar store's sales ramp up, and for detecting the presence of pre-trends in grocery store activity prior to the dollar store enters.

We plot the results in Figure 1 in the manner suggested by Freyaldenhoven et al. (2021), with confidence intervals adjusted for multiple hypothesis testing. The left panel shows the effects of dollar store entry on the number of grocery stores at the Census tract level and the right panel shows effects in the 0-2 mile radius. In the left panel we see a clear and nearly linear upwards pre-trend in the number of grocery stores that stops and reverses at the time of dollar store entry. In the right panel there is no pre-trend but there is a drop starting at time 0.

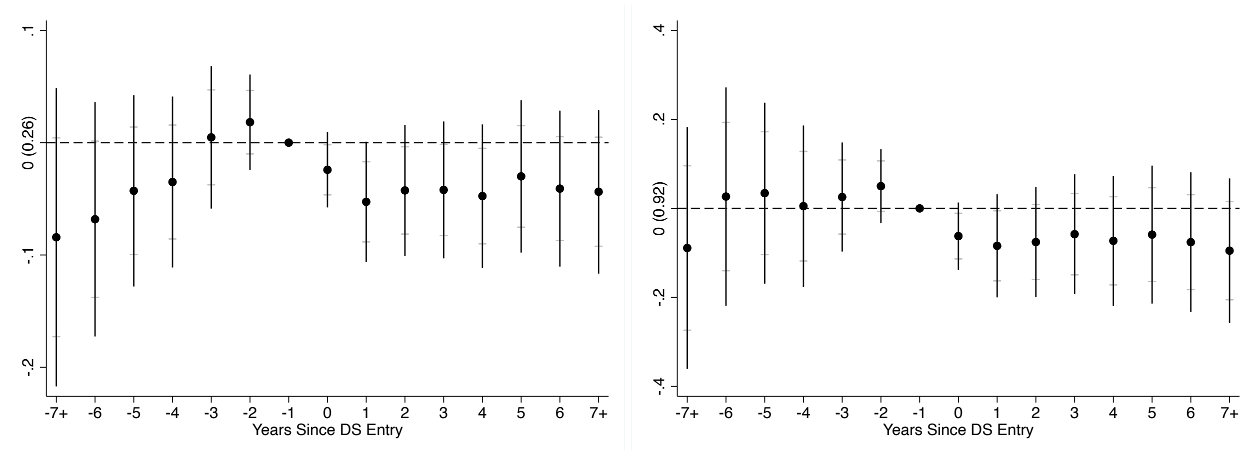

Figure 1: The effects of dollar store entry on local independent grocery stores measured at the Census tract level (left panel) and in the 0-2 mile radius around entry (right panel). Results are from regressions including all demographic controls and fixed effects. Confidence bands show the uniform sup-t confidence intervals adjusted for multiple hypothesis testing, with notches showing the standard pointwise confidence intervals.

4 Industry Model

In this section, we describe a model of the entry and exit game played by rival retailers over time. To effectively answer our research questions, the model must be able to generate the counterfactual market structure under alternative policies with respect to dollar store entry, including a total ban on dollar store entry in the sample period. To do so, we consider the actions of each type of store, how they interact, and how they depend on exogenous market characteristics. The main innovations in the model are that we model a situation that is non-stationary due to the rapid growth in the dollar store chains over time, contains long-lived chains as central players, and combines dynamics with spatial differentiation.

Players Two types of entrants can potentially operate in the market: multi-store firms (i.e., chains) and a set of independent single-store firms. Markets are assumed to be completely independent of each other. We index firms by $i = 1, \ldots, I_m$, and assume that market m has $I_{s,m}$ single-store firms, the remaining $I_{c,m} = I_m - I_{s,m}$ firms being chains (abusing notation, we also use $I_{s,m}$ and $I_{c,m}$ to denote sets of firms). Time is discrete and denoted by $t = 1, \ldots, \infty$. Each market $m = 1, \ldots, M$ is partitioned into locations denoted by l.

In what follows, we consider a market m that is partioned into locations $l = 1, \ldots, L$.

State space At the beginning of period t a chain's network of stores is represented by the vector $\mathbf{n}_{it} = (n_{i1t}, \ldots, n_{iLt})$, where n_{ilt} is a positive integer representing the number of stores that firm i operates in location l at period t. For simplicity, we assume that a chain can have up to $\overline{n}$ stores in a location, such that $n_{ilt} \in \{0, 1, \ldots, \overline{n}\}$. Single-store firms can operate only one store per market. The spatial market structure at period t is represented by the vector $\mathbf{n}_t = (\mathbf{n}_{it})_{i \in I}$. Let $\mathbf{n}_{-it}$ denote the network of stores of all firms other than i.

There are market and location characteristics that evolve exogenously over time, denoted $\mathbf{x}_{mt} = \{x_{mlt}\}_{l \in L}$. These include the population, income per capita, and rents in each location. Market-level characteristics include population, number of drug stores, supermarkets, and gas stations. For multi-store firms, let d_{imt} denote market m's distance from i's closest distribution center and $\mathbf{d}_{mt} = (d_{imt})_{i \in I_c}$ the vector collecting this variable for all chains. This vector evolves deterministically over time, as the chains expand their network of distribution centers. The transition matrices for these variables are denoted: $f(x_{ml,t+1}|x_{ml,t})$ and $h_t(d_{im,t+1}|d_{imt})$. The latter transition matrix is deterministic and the source of non-stationarities in the model.

Every period, the vector of public information variables includes the spatial market structure $\mathbf{n}_t$ and market and location level characteristics. All these variables are publicly observed and collected, from the perspective of firm i, into the vector $\mathcal{M}_{j,i,t}$, with particular realization j at time t. That is

$$\mathcal{M}_{j,i,t} = (\mathbf{n}_{it}, \mathbf{n}_{-it}, \mathbf{x}_{mt}, \mathbf{d}_{mt}) \tag{3}$$

Actions

Multi-store firms We assume that a chain may open or close at most one store per period. Let a_{it} be the decision of firm i at period t such that: $a_{it} = l_+$ represents the decision to open a new store at location l; $a_{it} = l_-$ means that a store at location l is closed; and $a_{it} = 0$ the firm chooses to do nothing. Some choice alternatives are not feasible for a firm given its current network $\mathbf{n}_{it}$. In particular, a firm cannot close a store in a location where it has no stores. The set of feasible choices for firm i at period t, denoted $A(\mathbf{n}_{it})$, is such that $A(\mathbf{n}_{it}) = \{0\} \cup \{l_+ : n_{ilt} < \overline{n}\} \cup \{l_- : n_{ilt} > 0\}$. Note that this choice set can have more than $L + 1$ choice alternatives. Multi-store firms are long-lived, that is, they can delay entry into the market. Exit from a market (that is, $\mathbf{n}_{it} = \mathbf{0}_L$ given that firm i was operating a store in $t - 1$) is a terminal action.

Single-store firms A single-store firm can enter if it is a potential entrants: $A_e = \{0\} \cup \{l_+\}$; or it can exit if it is an incumbent: $A_b = \{0\} \cup \{l_- : n_{ilt} = 1\}$. Firms that exit or potential entrants that decide to stay out are replaced by a new set of potential entrants in the following period.

We represent the transition rule of market structure as $\mathbf{n}_{t+1} = \mathbf{n}_t + 1[a_t]$, where $1[a_t]$ is a vector such that its (i,l)-element is equal to $+1$ when $a_{it} = l_+$, to -1 when $a_{it} = l_-$, and to zero otherwise. Firms' choices are dynamic because of partial irreversibility in the decision to open a new store, i.e., sunk costs. At the end of period t firms simultaneously choose their network of stores $\mathbf{n}_{t+1}$ with an understanding that they will affect their variable profits at future periods. We model the choice of store location as a game of incomplete information, so that each firm i has to form beliefs about other firms' choices of networks. More specifically, there are components of the entry costs and profits of a store which are firm-specific and private information.

Flow profits (and components) Firm i's current profits (net of private information shocks) are

$$\pi_{it}(a_{it}, \mathcal{M}_{j,i,t}) = VP_i(\mathcal{M}_{j,i,t}) - FC_{it}(\mathcal{M}_{j,i,t}) - EC_{it}(a_{it}) + EV_{it}(a_{it}), \tag{4}$$

where $VP_i(\mathcal{M}_{j,i,t})$ are variable profits, FC_{it} is the fixed cost of operating all the stores of firm i, EC_{it} is the entry or set-up cost of a new store, and EV_{it} is the exit value of closing a store.

Variable profits $VP_i(\mathcal{M}_{j,i,t})$ are obtained as the sum of profits over all stores firm i is operating in the market at time t,[4] that is

$$VP_i(\mathcal{M}_{j,i,t}) = \sum_{l=1}^{L} n_{ilt} vp_{i,l}(\mathcal{M}_{j,i,m,t}) \tag{5}$$

where $vp_{i,l}(\mathcal{M}_{j,i,m,t})$ are per-store profits. For a store in location l, variable profits are a function of the exogenous characteristics and the number of (own and rival) stores in location l and surrounding locations. Following Seim (2006), we capture this dependence by defining these variables for various distance bands around location l

$$vp_{i,l}(\mathcal{M}_{j,i,t}) = \sum_{b=1}^{B} \alpha_i^b x_{mlt}^b + \sum_{b=1}^{B} \beta_{io}^b n_{ilt}^b + \sum_{b=1}^{B} \sum_{f=1}^{F} \beta_{if}^b n_{flt}^b \tag{6}$$

where f denotes the type of competitors (i.e., dollar store, grocery or combination store, convenience store), and b are distance bands around location l (e.g., 0-2 miles, 2-5 miles). The variables x_{mlt}^b, n_{ilt}^b, and n_{flt}^b correspond to exogenous location characteristics, own stores, and rival stores of type f in distance band b around location l. The second term captures cannibalization and/or economies of density, the third term captures business stealing between

[4] The variable profit is equal to the difference between revenue and variable costs. It varies continuously with the firm's output and it is equal to zero when output is zero.

rival stores.

For chains, fixed operating costs depend on the number of stores and distance to the closest distribution center, capital costs (proxied by residential rents). Given that the firm is operating at least one store in the market, fixed costs are

$$FC_{it}(\mathcal{M}_{j,i,t}) = \theta^{FC}_{1,c}|\mathbf{n}_{it}| + \theta^{FC}_{2,c} d_{imt} + + \sum_{l=1}^{L} \theta^{FC}_{3,c} rent_{mlt} \tag{7}$$

For single-store firms, fixed costs depend on real estate costs: $FC_{it} = \theta^{FC}_{s} x_{mlt}$.

The specification of entry/investment cost is:

$$EC_{it} = \sum_{l=1}^{L} 1\{a_{it} = l_+\}\theta^{EC}_{i} \tag{8}$$

Entry costs θ^{EC}_{f} are a constant that depends only on the type of the firm f (dollar store, convenience store, grocery and combination stores).

The exit value is specified as:

$$EV_{it} = \sum_{l=1}^{L} 1\{a_{it} = l_-\}\theta^{EV}_{i} \tag{9}$$

The exit value θ^{EV}_{f} depends only on the type of the firm f.

At the beginning of period t, each firm draws a vector of private information shocks associated with each possible action $\boldsymbol{\epsilon}_{it} = \{\epsilon_{it}(a)\}_{a\in A(\mathbf{n}_{it})}$. We assume that the shocks ϵ_{it} are independently distributed across firms and over time and have a cumulative distribution function $G(.)$ that is strictly increasing and continuously differentiable with respect to the Lebesgue measure. These two assumptions allow for a broad range of specifications for the ϵ_{it}, including spatially correlated shocks. In our application, these shocks will be distributed Type 1 extreme value, scaled by a parameter θ^{ϵ}.[5]

It will be convenient to distinguish two additive components in the current profit function:

$$\Pi_{it}(a_{it}, \mathcal{M}_{j,i,t}, \boldsymbol{\epsilon}_{it}) = \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \tag{10}$$

Value function and Equilibrium concept We focus on Markov-Perfect Bayesian Nash Equilibria (MPE). We begin by defining firm strategies, value functions, and then the equilibrium conditions.

[5]These shocks can be thought of as representing the firms' idiosyncratic conditions in terms of real estate information, corporate finance, and other managerial or organizational climate for store-development activities (Igami and Yang (2016)).

A firm's strategy, at time t, depends only on its payoff relevant state variables $(\mathcal{M}_{j,i,t}, \boldsymbol{\epsilon}_{it})$. A strategy profile is denoted

$$\alpha = \{\alpha_{i,t}(\mathcal{M}_{j,i,t}, \boldsymbol{\epsilon}_{it}))\}_{i\in I, t\geq 0}.$$

Given strategy profile α, firm i's value function satisfies

$$V^{\alpha}_{i,t}(\mathcal{M}_{j,i,t}, \boldsymbol{\epsilon}_{it}) = \max_{a_{it}\in A(n_{it})} \left\{ v^{\alpha}_{i,t}(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \right\} \tag{11}$$

where $v^{\alpha}_{i,t}(a_{it}, \mathcal{M}_{j,i,t})$ are choice-specific value functions, defined as

$$\begin{aligned} v^{\alpha}_{i,t}(a_{it}, \mathcal{M}_{j,i,t}) &= \pi_i(a_{it}, \mathcal{M}_{j,i,t}) \\ &\quad + \beta \int V^{\alpha}_{i,t}\left(\mathcal{M}_{j,i,t+1}, \boldsymbol{\epsilon}_{i,t+1}\right) dG(\boldsymbol{\epsilon}_{i,t+1}) dF_t(\mathcal{M}_{j,i,t+1} | a_{it}, \mathcal{M}_{j,i,t}) \end{aligned} \tag{12}$$

where the next-period state $\mathcal{M}_{j,i,t+1}$ is formed of the next-period spatial market structure $\mathbf{n}_{t+1}$, and market and firm-level covariates $(\mathbf{x}_{m,t+1}, \mathbf{d}_{m,t+1})$. The distribution over next-period states is given by the transition probabilities $f(\mathbf{x}_{m,t+1}|\mathbf{x}_{m,t})$ and $h_t(\mathbf{d}_{m,t+1}|\mathbf{d}_{m,t})$ of exogenous states, and the distribution of rivals' shocks $\Pi_{j\neq i} g(\boldsymbol{\epsilon}_{j,t})$ and strategies α_j for $j \neq i$. Note that the roll-out of distribution centers $(f_t(\mathbf{d}_{m,t+1}|\mathbf{d}_{m,t})))$ introduces non-stationarities in the model.

A MPE is a strategy profile α^* such that for every player, state, and period

$$\alpha^*_{i,t}(\mathcal{M}_{j,i,t}, \epsilon_{it}) = \arg \max_{a_{it}\in A(n_{it})} \left\{ v^{\alpha^*}_{i,t}(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \right\} \tag{13}$$

The probability that firm i chooses action a_{it} in period t given state $\mathcal{M}_{j,i,t}$ (hereafter, the conditional choice probability or CCP) is defined as

$$P^{\alpha}_t(a_{it}|\mathcal{M}_{j,i,t}) = P(\alpha_{i,t}(\mathcal{M}_{j,i,t}, \epsilon_{it}) = a_{it}|\mathcal{M}_{j,i,t}) \tag{14}$$

We find it convenient to express the choice-specific value function as a function of CCPs instead of strategies. That is,

$$v^{\mathbf{P}}_{i,t}(a_{it}, \mathcal{M}_{j,i,t}) = \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \beta \sum_{a_{-it}} \int \overline{V}^{\mathbf{P}}_{i,t+1}\left(\mathcal{M}_{j,i,t+1}\right) dF_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t}) \tag{15}$$

where $a_t = (a_{it}, a_{-it})$ and $\overline{V}^{\mathbf{P}}_{i,t}$ is the *ex-ante* value function expressed before the realization of the private shock $\boldsymbol{\epsilon}_{it}$

$$\overline{V}_{i,t}^{\mathbf{P}}(\mathcal{M}_{j,i,t}) = \int \max_{a_{it}\in A(n_{it})} \Bigg\{ \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \\ + \beta \sum_{a_{-it}} \int \overline{V}_{i,t+1}^{\mathbf{P}}(\mathcal{M}_{j,i,t+1})\, dF_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t}) \Bigg\} dG(\epsilon_{it}) \tag{16}$$

The best-response mapping can also be defined in the space of CCPs as $\alpha_t^{BR}(\mathcal{M}_{j,i,t}, \epsilon_{it}, \mathbf{P}) = \arg\max_{a_{it}\in A(n_{it})} \left\{ v_i^{\mathbf{P}}(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \right\}$ and a MPE can be represented as a fixed point of the best-response mapping (Aguirregabiria and Mira (2007), Presendorfer and Schmidt-Dengler (2008))

$$\Phi_{i,t}(a_{it}|\mathcal{M}_{j,i,t}, \mathbf{P}) = \int 1 \left(a_{it} = \arg\max_{a_{it}\in A(n_{it})} \left\{ v_{i,t}^{\mathbf{P}}(a_{it}, \mathcal{M}_{j,i,t}) + \theta^{\epsilon}\epsilon_{it}(a_{it}) \right\} \right) dG(\epsilon_{it})$$

If private shocks are distributed Type 1 extreme value (with scale parameter θ^{ϵ}), an optimal strategy for firm i will map into conditional choice probabilities of the form

$$P_t(a_{it}|\mathcal{M}_{j,i,t}, \mathbf{P}) = \frac{\exp\left(\frac{v_{i,t}^{\mathbf{P}}(a_{it}, \mathcal{M}_{j,i,t})}{\theta^{\epsilon}}\right)}{\sum\limits_{a'\in A(n_{it})} \exp\left(\frac{v_{i,t}^{\mathbf{P}}(a', \mathcal{M}_{j,i,t})}{\theta^{\epsilon}}\right)}. \tag{17}$$

Finite dependence The model features a terminal choice—exit without the possibility of re-entry—a special case of finite dependence (Altuğ and Miller (1998), Arcidiacono and Miller (2011)). Finite dependence eases the calculation of ex-ante and choice-specific value functions because these too can be expressed directly in terms of the period-ahead probabilities of choosing the terminal choice. Moreover, it allows us to incorporate nonstationarities into the model without making out-of-sample assumption about players' actions for periods beyond the sample horizon (which is the year 2019).

Single-store firms If a single-store incumbent exits or a single-store potential entrant stays out, the continuation value is zero. The choice-specific value function from staying active (either entering or remaining in the market) can therefore be expressed relative to the exit choice, denoted e (for an incumbent, it is $a_{it} = l_-$ if $n_{ilt} = 1$, or for a potential entrant $a_{it} = 0$). The choice-specific value function (Equation (15)) can be rewritten for all $a_{it} \neq e$

$$
\begin{aligned}
v^{\mathbf{P}}_{i,t}(a_{it}, \mathcal{M}_{j,i,t}) = \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \beta \sum_{a_{-it}} \int [v^{\mathbf{P}}_{i,t+1}(a'_i = e, \mathcal{M}_{j,i,t+1}) \\
+ \gamma - \ln(P_{i,t+1}(e|\mathcal{M}_{j,i,t+1})] dF_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t})
\end{aligned} \tag{18}
$$

where γ is the Euler constant. The choice-specific value function corresponding to the exit decision is given by

$$
v^{\mathbf{P}}_{i,t+1}(a'_i = e, \mathcal{M}_{j,i,t+1}) = \begin{cases} \pi_i(a'_i = e, \mathcal{M}_{j,i,t+1}) & \text{if } i \text{ is an incumbent in } l \\ 0 & \text{if } i \text{ is potential entrant} \end{cases}
$$

Multi-store firms. The problem for multi-store entrants differs from that of single-store ones because chains are long-lived: they can delay entry into a market without being replaced by a new potential entrant. Therefore, the only terminal choice for a multi-store firm is exit from incumbency. Because multi-store firms are restricted to close or open only one store per period, exit can occur only when the firm is operating a single store. Finite dependence still holds, but in the number of periods it takes to bring the firm to operating a single-store in the market. For instance, a firm operating three stores in period t and choosing to do nothing, will be in a state that features three-period finite dependence.

For an incumbent operating a single-store in location l ($a_{it} = 0$) or a potential entrant i who enters into location l ($a_{it} = l_+$), the choice-specific value function is identical to Equation (18).

For potential entrant i who stays out in period t, that is $a_{it} = 0$, the choice-specific value function can be expressed, given entry into an arbitrary location $a'_i = l_+$ in period $t+1$ as follows

$$
\begin{aligned}
v^{\mathbf{P}}_{i,t}(0, \mathcal{M}_{j,i,t}) &= \pi_i(0, \mathcal{M}_{j,i,t}) + \beta \sum_{a_{-it}} \int [v^{\mathbf{P}}_{i,t+1}(a'_i = l_+, \mathcal{M}_{j,i,t+1}) \\
&\quad - \ln(P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})] f_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t}) \\
&= \pi_i(0, \mathcal{M}_{j,i,t}) + \beta \sum_{a_{-it}} \int ([\pi_i(l_+, \mathcal{M}_{j,i,t+1}) - \ln(P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})] \\
&\quad + \beta \sum_{a_{-i,t+1}} \int [v^{\mathbf{P}}_{i,t+2}(a''_i = l_-, \mathcal{M}_{j,i,t+2}) - \ln(P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})] \\
&\quad \times dF_{\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t+1}} P_{-i,t+1}) dF_{\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}} P_{-i,t}
\end{aligned} \tag{19}
$$

For an incumbent chain operating a network of stores $\mathbf{n}_{it}$ and choosing action a in period t, define a sequence of choices $\{a_{t+\tau}\}_{\tau=1}^{|\mathbf{n}_{it}+a|}$, such that in every period $t+\tau$, the chain closes one of its operating stores. In the last period $|\mathbf{n}_{it}+a|$, the chain chooses the terminal action (exit). Then, the choice-specific value function can be expressed following the aforementioned sequence of choices to obtain ($|\mathbf{n}_{it}+a|+1$ finite dependence) as follows

$$v_{i,t}^{\mathbf{P}}(a,\mathcal{M}_{j,i,t}) = \pi_i(a,\mathcal{M}_{j,i,t}) + \sum_{\tau=1}^{|\mathbf{n}_{it}+a|-1} \beta^{\tau-1}\,\mathbb{E}[\pi_i(a_{i,t+\tau},\mathcal{M}_{j,i,t+\tau}) - \ln(P_{i,t+\tau}(a_{i,t+\tau}|\mathcal{M}_{j,i,t+\tau})|\mathcal{M}_{j,i,t}] + \beta^{|\mathbf{n}_{it}+a|}\,\mathbb{E}[v_{i,t+|\mathbf{n}_{it}+a|}^{\mathbf{P}}(a=e,\mathcal{M}_{j,i,t+|\mathbf{n}_{it}+a|}) - \ln(P_{i,t+|\mathbf{n}_{it}+a|}(e|\mathcal{M}_{j,i,t+|\mathbf{n}_{it}+a|})] \tag{20}$$

5 Identification and Estimation

5.1 Identification

As is standard in the literature on the identification of dynamic decision problems (Rust (1994), Magnac and Thesmar (2002), Bajari et al. (2015)), the discount factor and the distribution of firm shocks (β, G) are assumed to be known.[6]

Aguirregabiria and Suzuki (2014) study the identification of market entry and exit games. They show that the level of fixed costs, entry costs and exit value are not separately identified. When estimating the model, we normalize the exit value to zero. A consequence of this "normalization" restriction is that the estimated entry costs will reflect the true *sunk* costs (entry cost net of exit value), and estimated fixed costs will reflect the true fixed costs in addition to the exit value scaled by $(1-\beta)$.

Variable profits are identified from exogenous variation in market and location-level characteristics (i.e., income, population, rents) and the geographic layout of markets (i.e, the distance between each pair of locations in a market) creating variation in these exogenous variables by distance bands around each location. The effects of rivals' stores on profits (i.e., competitive effects) are identified in two ways: for chains, we rely on exogenous variation in the distance to the closest (rival) distribution center which shifts rival chains' entry decisions without directly affecting own variable profits; for single-store firms, competitive effects are identified from variation in the number of potential entrants.

[6]Markets are independent, therefore, identification is based on a cross-section of market-paths assuming they all feature the same equilibrium.

5.2 Estimation

5.2.1 Estimation approach

We follow a two-step approach. In a first step, consistent estimates of the CCPs, denoted $\widehat{\mathbf{P}}_0$ are obtained. We discuss this first step and the treatment of unobserved heterogeneity in detail in Section 5.2.2. In the remainder of this section, we focus on the estimation of the structural parameters given first-step estimates of the CCPs.

Existing methods for the estimation of dynamic games such as policy evaluation (Aguirregabiria and Mira (2007), Presendorfer and Schmidt-Dengler (2008)) or forward simulation (Bajari et al. (2007)) require stationarity, i.e., firms' strategies and transition processes should not depend on time. This requirement does not hold in our application, because we account for the nonstationary nature of the retail industry (and in particular, its discount retail segment), through the expansion of dollar store chains' network of distribution centers. Moreover, firms' dynamic locational choices within a market and the presence of multi-store firms generate a high-dimensional state space. Implementing the approaches cited above would require a combination of state space discretization, approximation of the value functions, and numerical integration over the state space.

We tackle these challenges in two ways. First, we leverage the finite dependence property to estimate nonstationary games (Arcidiacono and Miller (2011, 2019)). In our setting, firms have a terminal choice (exit without the possibility of re-entry), a special case of finite dependence. As detailed in Section 4, this allows us to express period-t choice-specific value function as a function of period-$t+1$ (known) CCPs and structural profit function. Second, we avoid numerical integration over the high-dimensional state space by using the linear IV regression approach of Kalouptsidi et al. (2020). The latter paper combines insights from the finite dependence approach and the GMM-Euler approach (Aguirregabiria and Magesan (2013, 2018)) to construct moment restrictions for the structural parameters. These moment restrictions do not require explicit integration over the state space but only averaging over the sample observations. Estimation of the structural parameters amounts to a linear regression equation which substantially eases the computational burden.

We show that the estimator of Kalouptsidi et al. (2020) developed for *single-agent* dynamic discrete choice problems can be extended to dynamic games. As far as we know, this is the first application of the linear regression estimation approach to dynamic games. In dynamic games with simultaneous moves, a focal firm i's decision in period t affects rivals' decisions in $t+1$ and their endogenous states in $t+2$: this aspect of games, absent in single-agent models, creates a dynamic selection problem for endogenous states. For instance, the number of grocery stores operating in location l at time $t+2$ will depend on a dollar store chains' entry and exit decisions in period t (in the data).[7] The estimator of Kalouptsidi et al. (2020) is modified to account for this important selection problem and obtain consistent estimates of the structural parameters. Next, we discuss how moment restrictions are constructed for single-store and multi-store firms.

Single-store firms. Differences in choice-specific value function for potential entrants and incumbents can be derived as follows. A potential entrant can either stay out ($a_{it} = 0$) or enter by building a store in location l ($a_{it} = l_+$). The corresponding choice-specific value functions are given by

$$v^{\mathbf{P}}_{i,t}(a_{it} = 0, \mathcal{M}_{j,i,t}) = 0 \tag{21}$$

$$v^{\mathbf{P}}_{i,t}(a_{it} = l_+, \mathcal{M}_{j,i,t}) = -\theta^{EC}_i + \beta\, \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{22}$$

where the expectation is over $\mathcal{M}_{j,i,t+1}$ conditional on $(a_{it} = l_+, \mathcal{M}_{j,i,t})$ and we use finite dependence to express the entrant continuation value in period $t+1$. Combining these two equations gives

$$v^{\mathbf{P}}_{i,t}(l_+, \mathcal{M}_{j,i,t}) - v^{\mathbf{P}}_{i,t}(0, \mathcal{M}_{j,i,t}) = -\theta^{EC}_i + \beta\, \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{23}$$

Differences in choice-specific value functions can alternatively be expressed using current period CCPs as

$$v^{\mathbf{P}}_{i,t}(l_+, \mathcal{M}_{j,i,t}) - v^{\mathbf{P}}_{i,t}(0, \mathcal{M}_{j,i,t}) = \ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) \tag{24}$$

Combining Equation (23) and Equation (24), we obtain an optimality condition that involves only CCPs at t and $t+1$ and the single-period payoff function

$$\ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) = -\theta^{EC}_i + \beta\, \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{25}$$

This equation includes expected profits and CCPs at $t+1$, and therefore, it seems that it requires numerical integration over the state space. However, this equation can be used to

[7] Note due to simultaneous moves, a firm's action in t impacts rivals' states in $t+2$ but not in $t+1$.

construct moment conditions that do not require explicit integration over the space of state variables. Under rational expectations, the conditional expectation at period t of CCPs and profits at $t+1$ is equal to these variables minus an expectational error that is orthogonal to the state variables at period t.[8] Therefore, for any function of period-t information set $h(\mathcal{M}_{j,i,t})$, we have

$$\mathbb{E}\left[h(\mathcal{M}_{j,i,t})u_{it}\right] = 0 \tag{26}$$

where u_{it} is the expectational error (also known as forecast errors). It is defined, for any realization $\mathcal{M}^*_{j,i,t+1}$ of the random variable $\mathcal{M}_{j,i,t+1}$ as

$$\begin{aligned} u_{it}(\mathcal{M}^*_{j,i,t+1}) &= \beta\,\mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \\ &\qquad - \beta\left(vp_{i,l}(\mathcal{M}^*_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}^*_{j,i,t+1})\right) \\ &= \ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) + \theta_i^{EC} \\ &\qquad - \beta\left(vp_{i,l}(\mathcal{M}^*_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}^*_{j,i,t+1})\right) \end{aligned} \tag{27}$$

where the second equation is obtained by using the expression in Equation (25) to eliminate the expectation term. The moment conditions (Equation (26)) do not require integration over the space of state variables but only averaging over the sample observations. The computational cost of estimating the structural parameters using GMM based on these moment conditions does not depend on the dimension of the state space.

Kalouptsidi et al. (2020) show that these moment conditions (replaced by their sample counterparts) can form the basis of a linear IV regression, where period-t variables are used as instruments for period $t+1$ variables. Define the left-hand side variable for potential entrant and incumbent (respectively) as

$$\begin{aligned} Y_{it}^{entrant} &= \ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) - (\gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})) \\ Y_{it}^{incumbent} &= \ln\left(\frac{P_{i,t}(0|\mathcal{M}_{j,i,t})}{P_{i,t}(l_-|\mathcal{M}_{j,i,t})}\right) - (\gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})) \end{aligned}$$

We can obtain the structural parameter via the regression model

$$Y_{it}^{entrant} = -\theta_i^{EC} + \beta\left[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i\right] + u_{it}$$

[8]This idea has been first used in the estimation of continuous choice dynamic structural models using Euler equations (e.g., Hansen and Singleton (1982)).

$$Y_{it}^{incumbent} = \beta[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i] - \theta_i^{EV} + u_{it}$$

where regressors entering the variable profit function in $t+1$ $vp_{i,l}(\mathcal{M}_{j,i,t+1})$ (population, income, etc.) are instrumented using the values of these regressors in period t.

Multi-store firms. The estimation procedure for multi-store chains is conceptually similar but somewhat more complicated due to the fact that they are long-lived and can delay entry. This implies that finite dependence holds in two (or more) periods. In dynamic games, a focal firm i's action in period t will affect its rivals' actions in $t+1$ and their states in $t+2$. This creates a selection problem in the data: endogenous state variables in period $t+2$ are observed only conditional on the action a_{it} played by firm i *in the data*. We show, in Appendix B, that an appropriate reweighting using CCP ratio addresses the selection bias, extending the approach of Kalouptsidi et al. (2020) from single-agent to dynamic games.

5.2.2 Location-level unobserved heterogeneity and first-step estimates

The presence of unobserved heterogeneity is a common concern in many empirical settings, and can introduce a serious endogeneity problems in the context of dynamic games of market entry and exit as it leads to biased estimates of competition. If unobserved heterogeneity is not controlled for, firms may appear to favor locations and markets with large numbers of competitors, which ultimately will yield economically implausible estimates of competitive effects.

We incorporate location-level unobserved heterogeneity via a proxy variable. This approach has been used in previous studies of market entry, e.g., Collard-Wexler (2013), and has the advantage of being computationally light. This is particularly important as a market is partitioned into multiple locations, which may differ in their attractiveness, yielding multi-dimensional unobserved heterogeneity. The proxy variable strikes a balance between granularity in the level of unobserved heterogeneity and computational feasibility. We define a location-level proxy for unobserved heterogeneity as the maximum number of establishments (of all types, including drug stores, supermarkets, and gas stations) *simultaneously* operating in a given location over the period 2008-2019.

The importance of controlling for unobserved heterogeneity is illustrated in Table 7.[9] This table shows estimates of the CCPs for dollar store chains via a flexible multinomial logit regression. An entrant chain can either build a store in one of the locations in the market or stay out. An incumbent chain can do nothing, build an additional store in one of the locations, or close one of its existing stores.[10] We control for location-level demographic

[9] In an ideal world, CCPs would be estimated nonparametrically, but this is not possible given the size of the state space, the large number of choices that each firm has, and the size of the observed sample.

[10] The small number of observations where a chain opens more than one store in a period are not included

variables, cost shifters (e.g., distance between the market and the closest distributin center), the location-level competitive environment, and market-level characteristics (e.g., other store types such as gas stations and supermarkets). We allow the parameters to differ for the decision to open and close a store. The first two columns correspond to a specification without unobserved heterogeneity. The last two columns include the proxy for unobserved heterogeneity ("Business Density"). To allow strategies to depend on the nonstationary roll-out of distribution centers, we also include year dummies.

The effect of competition on the likelihood of building a store are biased upward when business density is not controlled for (column 2) relative to when it is included (column 4). In column 2, many competition coefficients are in fact positive (e.g., for the number of grocery stores and convenience stores within 2mi), reflecting agglomeration effects due to unobserved location-level amenities. This is not the case when location-level business density is included (column 4).

Similarly to chains, we estimate the CCP for single-store firms (grocery/combination stores and convenience stores) via flexible multinomial logit regressions, controlling for business density. We include the regression results in Appendix F for completeness.

5.2.3 Estimation results

The section presents our estimates of the structural parameters entering single-period payoffs for dollar store chains, and independent grocery and convenience stores.

Table 8 shows estimates of normalized store profits and entry costs. We include a constant term in the profit function to capture the level of fixed costs and/or any baseline level of profits. The effect of most variables decays with distance from the store location, highlighting the importance of spatial differentiation in retail competition. For all retailers, profits are increasing with the population within 2 miles of the store location. Dollar store chains' favor locations with lower income. Profits for chains are decreasing in the distance to the closest distribution center. The majority of competition effects are precisely estimated and with the expected magnitude.

To help interpretation, we convert our profit estimates into dollars by calibrating the scale parameter of firm shocks θ^{ϵ} to match revenue data for all dollar stores operating in the markets under consideration, obtained from Nielsen TDLinx.[11] Table 9 shows mean store

when calculating the likelihood.

[11]Specifically, we convert the revenue data into profits (deflated to 2010), assuming a 5% net profit rate, and calibrate the scale parameter θ^{ϵ} to match the model-predicted profits and the observed profits for all operating dollar stores in 2019. We use the calibrated scale parameters to convert all estimates into 2010$. For many stores, revenue data is imputed by Nielsen. Due to these imputations and the absence of revenue data for single-store firms, we do not use these data to estimate a demand model. This conversion is just

Table 7: Multinomial logit of multi-store firms' choice

	Multi-store firms		Multi-store firms	
	Close store in l	Build store in l	Close store in l	Build store in l
Entrant		11.905 (1.773)		2.906 (1.726)
Incumbent	-19.433 (4.553)	9.898 (1.770)	-17.237 (4.670)	1.044 (1.726)
Location-level characteristics				
Population (0-2 mi)	0.552 (0.315)	0.209 (0.064)	0.608 (0.313)	0.190 (0.066)
Population (2-5 mi)	-0.160 (0.053)	0.056 (0.028)	-0.155 (0.054)	0.040 (0.026)
Income per capita (0-2 mi)	1.162 (0.464)	-0.817 (0.170)	0.857 (0.470)	-0.415 (0.168)
Income per capita (2-5 mi)	0.124 (0.033)	-0.015 (0.018)	0.124 (0.034)	-0.012 (0.017)
Cost shifters				
Distance to own distribution center	-0.055 (0.120)	-0.205 (0.050)	-0.055 (0.128)	-0.203 (0.047)
Distance to distribution center (rival 1)	0.069 (0.137)	0.024 (0.049)	0.095 (0.147)	0.020 (0.047)
Distance to distribution center (rival 2)	0.447 (0.136)	-0.039 (0.052)	0.467 (0.144)	-0.040 (0.049)
Median residential rent	0.045 (0.393)	-0.442 (0.172)	-0.075 (0.396)	-0.242 (0.177)
Number of own chain stores in market	-0.407 (1.135)	1.190 (0.311)	-0.531 (1.100)	0.942 (0.319)
Measures of competition				
Number of rival chain stores (0-2 mi)	0.345 (0.190)	-0.132 (0.089)	0.227 (0.191)	-0.394 (0.084)
Number of rival chain stores (2-5 mi)	0.164 (0.247)	-0.199 (0.102)	0.101 (0.245)	-0.176 (0.093)
Number of rival grocery/combination (0-2 mi)	-0.071 (0.153)	0.026 (0.066)	-0.103 (0.155)	-0.311 (0.066)
Number of rival grocery/combination (2-5 mi)	0.296 (0.240)	-0.222 (0.087)	0.351 (0.240)	-0.273 (0.085)
Number of rival convenience (0-2 mi)	-0.044 (0.122)	0.030 (0.060)	-0.187 (0.133)	-0.243 (0.056)
Number of rival convenience (2-5 mi)	0.268 (0.157)	0.025 (0.074)	0.235 (0.158)	0.002 (0.070)
Number of own chain stores (0-2 mi)	0.343 (1.071)	-1.112 (0.242)	0.425 (1.038)	-1.191 (0.246)
Number of own chain stores (2-5 mi)	-0.052 (0.648)	0.012 (0.229)	-0.041 (0.633)	0.159 (0.228)
Market-level characteristics				
Population	-0.522 (0.386)	-0.662 (0.119)	-0.485 (0.392)	-0.290 (0.111)
Number of gas stations	-0.016 (0.133)	-0.039 (0.065)	-0.109 (0.136)	0.096 (0.061)
Number of drug stores	-0.068 (0.219)	0.271 (0.083)	0.036 (0.230)	0.159 (0.076)
Number of supermarkets	0.035 (0.277)	0.398 (0.101)	0.163 (0.282)	0.318 (0.089)
Year FE	No		Yes	
Business Density	No		Yes	
Observations	24,923		24,923	
Log Likelihood	-6,289.707		-5,937.582	

Note: Standard errors are clustered by market. The baseline alternative is "do nothing." Dollar figures are in 2010\$. Business density is defined as the maximum number of establishments simultaneously operating in location l over the period 2008-2019. Distance to distribution center is at the market level, residential rent is at the location level. All continuous variables and store counts are in log.

profits and entry costs expressed in 2010$, as well as marginal effects.

We find that, consistent with anecdotal reporting on dollar store growth, these chains have substantially lower costs of opening a new store than their independent rivals. They are also substantially more profitable. When we examine the competitive effects of nearby rivals on profits, several results stand out. First, grocery store profits are significantly harmed by the presence of nearby dollar stores and convenience stores, with most of the effects for

for interpretation purposes and is not used in the counterfactuals that follow.

Table 8: Estimates of store profits and costs

	Chains		Grocery/Combination Store		Convenience Store	
Parameters	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
Constant	2.616	(0.444)	-1.177	(0.485)	-0.876	(0.220)
Location-level characteristics						
Population (0-2 mi)	0.049	(0.015)	0.156	(0.031)	0.069	(0.010)
Population (2-5 mi)	0.010	(0.005)	0.004	(0.006)	0.002	(0.004)
Income per capita (0-2 mi)	-0.175	(0.049)	-0.016	(0.048)	0.017	(0.021)
Income per capita (2-5 mi)	-0.004	(0.003)	0.001	(0.004)	-0.000	(0.002)
Fixed cost components						
Median residential rent	-0.072	(0.056)	0.050	(0.050)	0.009	(0.028)
Distance to own distribution center	-0.058	(0.020)				
Measures of competition and cannibalization						
Number of rival chain stores (0-2 mi)	-0.070	(0.023)	-0.128	(0.024)	-0.069	(0.012)
Number of rival chain stores (2-5 mi)	-0.048	(0.022)	-0.005	(0.022)	-0.018	(0.011)
Number of rival grocery/combination stores (0-2 mi)	-0.074	(0.022)	-0.030	(0.021)	-0.033	(0.011)
Number of rival grocery/combination stores (2-5 mi)	-0.073	(0.025)	-0.068	(0.025)	-0.035	(0.014)
Number of rival convenience stores (0-2 mi)	-0.073	(0.022)	-0.104	(0.017)	-0.057	(0.009)
Number of rival convenience stores (2-5 mi)	0.026	(0.022)	-0.012	(0.020)	-0.005	(0.010)
Number of own chain stores (0-2 mi)	-0.094	(0.045)				
Number of own chain stores (2-5 mi)	0.077	(0.024)				
Market-level characteristics						
Population	-0.092	(0.028)	-0.099	(0.034)	-0.059	(0.013)
Number of gas stations	0.016	(0.019)	-0.002	(0.019)	-0.052	(0.011)
Number of drug stores	0.068	(0.022)	0.009	(0.024)	-0.001	(0.014)
Number of supermarkets/centers	0.099	(0.025)	-0.036	(0.025)	0.034	(0.017)
Dynamic investment costs						
Entry cost	2.495	(0.240)	5.515	(0.063)	5.878	(0.052)
Entry cost of additional store	9.713	(0.165)				

Note: Standard errors are obtained via bootstrap of market-histories (200 replications). All continuous variables and store counts are in log. Business density and year fixed effects are controlled for. Residential rent is at the location level.

stores in the 0-2mi radius.[12] Second, the presence of dollar stores also significantly harms convenience store profits, by as much as an additional convenience store. Third, within dollar store store chains, in the 0-2mi radius there is a strong demand cannibalization effect but in the 2-5mi range this effect is reversed and chains benefit from scale economies, likely working through lower operating costs.

[12] The magnitude of the business stealing effects are consistent with anecdotal evidence from grocery store owners. For instance, the owner of the Foodliner store in Haven, KS reports,

> "We lasted three years and three days after Dollar General opened," he said. "Sales dropped and just kept dropping. We averaged 225 customers a day before and immediately dropped to about 175. A year ago we were down to 125 a day. Basically we lost 35 to 40% of our sales. I lost a thousand dollars a day in sales in three years." (The Guardian, "Where even Walmart won't go: how Dollar General took over rural America", 2018)

Table 9: Mean store profits and marginal effects

	Chain	Grocery/Combination	Convenience
Mean store profits (conditional on remaining active) in 2010$	73,074	42,719	43,937
Mean entry costs (conditional on entering) in 2010$	129,169	192,093	244,170
Percentage change in mean store profits from			
One additional rival chain store (0-2 mi)	-9.69	-30.45	-16.01
One additional rival chain store (2-5 mi)	-6.71	-1.28	-4.13
One additional rival grocery/combination store (0-2 mi)	-10.30	-7.05	-7.64
One additional rival grocery/combination store (2-5 mi)	-10.14	-16.24	-8.18
One additional rival convenience store (0-2 mi)	-10.22	-24.78	-13.22
One additional rival convenience store (2-5 mi)	3.69	-2.90	-1.09
One additional own chain store (0-2 mi)	-13.15		
One additional own chain store (2-5 mi)	10.67		
Increase in dist. to distribution center by one s.d. from mean	-6.13		

Note: Averages are over all incumbent stores (for profits) and entrants (for entry costs) over the period 2010-2019. Conditional profits and entry costs include the expectation of the structural shock. Percentage changes are relative to the monopoly case. The mean distance to the closest distribution center is 190mi and the standard deviation is 130mi.

6 Counterfactual Policy Analysis

This section uses the structural estimates to conduct counterfactual policy evaluations. We outline in detail the method used to solve for counterfactual MPE in Appendix C. We focus on evaluating the impact of dollar store chains' expansion on market structure by simulating the evolution of the industry had chains been prevented from expanding beginning in 2010.

6.1 The Impact of Dollar Stores' Expansion on Market Structure

We simulate the counterfactual MPE where dollar store chains are prevented from expanding starting in 2010. The counterfactual CCPs obtained are used to simulate the industry forward from 2010 to 2019 in each market.

Table 10 and Table 11 show how the expected number of grocery/combination stores and convenience stores are predicted to change in the counterfactual (CF) relative to the market equilibrium where dollar store chains expanded (Actual) in 2019, by location and by market respectively.

We find that, in the hypothetical world in which dollar stores did not expand after 2010, markets have on average more than 50% more independent stores, including 56% more combination and grocery stores and 54% more convenience stores.[**BH: These numbers need to be corrected:**] The net effect of these changes across the 846 markets we study is that, without dollar store expansion there would be roughly 450 more grocery and combination stores and 750 more convenience stores.

Table 10: Expected number of stores by location

	Grocery/Combination store counts				Convenience store counts			
	Factual	CF	Δ	% Δ	Factual	CF	Δ	% Δ
All locations	0.442	0.684	0.242	0.348	0.87	1.247	0.377	0.398
By number of commercial locations								
1	0.896	1.473	0.577	0.663	1.419	2.199	0.781	0.657
2	0.483	0.738	0.255	0.351	0.93	1.34	0.411	0.429
3	0.354	0.545	0.191	0.306	0.756	1.059	0.302	0.342
By income and population								
Population below median, Income below median	0.368	0.519	0.151	0.247	0.875	1.203	0.329	0.326
Population above median, Income below median	0.65	1.066	0.416	0.544	1.213	1.779	0.565	0.471
Population below median, Income above median	0.25	0.335	0.085	0.133	0.562	0.77	0.208	0.308
Population above median, Income above median	0.498	0.814	0.316	0.469	0.827	1.233	0.406	0.487
By share of minority groups								
Above 0.25	0.593	0.957	0.364	0.42	1.248	1.752	0.504	0.366
Below 0.25	0.403	0.614	0.211	0.329	0.773	1.118	0.345	0.406
By share of population with access to vehicle								
Below first quartile (0.89)	0.548	0.88	0.332	0.458	1.171	1.658	0.487	0.403
Above first quartile (0.89)	0.406	0.618	0.212	0.311	0.769	1.109	0.34	0.396
By share of population under poverty line								
Below median (0.16)	0.393	0.599	0.206	0.302	0.701	1.02	0.32	0.409
Above median (0.16)	0.491	0.769	0.278	0.394	1.039	1.474	0.435	0.386
By presence of dollar stores in 2010								
No dollar stores in 2010	0.32	0.456	0.136	0.206	0.677	0.928	0.251	0.323
Dollar stores present in 2010	0.586	0.953	0.367	0.516	1.096	1.623	0.527	0.486
Food deserts								
No supermarkets in 2010	0.586	0.875	0.289	0.3	0.85	1.215	0.365	0.422
Supermarkets present in 2010	0.436	0.676	0.24	0.35	0.87	1.248	0.378	0.397

Note: Factual corresponds to the expected number of stores under the market equilibrium. CF corresponds to the counterfactual expected number of stores. Δ (resp. % Δ) gives the difference (resp. percentage difference) between the market outcome and counterfactuals averaged over all the locations. All demographic variables and store counts are at the location level, except for the presence of supermarkets which is at the census place (or market) level.

We can also examine how these changes vary across different market types. We find that the largest impacts on number of grocery/combination stores comes in lower income markets, those with larger shares of minority populations, those with higher poverty rates, and fewer households with access to vehicles. The effects are also largest in markets in which dollar stores do not already have a presence in 2010. The reverse is true for convenience stores. The largest increase in convenience stores under the counterfactual policy comes from markets that have higher incomes, lower minority shares, less poverty, and higher access to vehicles.

Table 11: Expected number of stores by market

	Grocery/Combination store counts				Convenience store counts			
	Factual	CF	Δ	% Δ	Factual	CF	Δ	% Δ
All markets	0.995	1.54	0.545	0.554	1.958	2.807	0.849	0.491
By number of commercial locations								
1	0.896	1.473	0.577	0.663	1.419	2.199	0.781	0.657
2	0.966	1.476	0.51	0.53	1.86	2.681	0.821	0.492
3	1.063	1.636	0.572	0.539	2.269	3.176	0.907	0.429
By income and population								
Population below median, Income below median	0.976	1.527	0.551	0.573	2.226	3.139	0.913	0.479
Population above median, Income below median	1.202	1.9	0.697	0.619	2.209	3.166	0.958	0.476
Population below median, Income above median	0.853	1.265	0.412	0.461	1.574	2.269	0.695	0.508
Population above median, Income above median	0.958	1.481	0.523	0.557	1.8	2.624	0.824	0.502
By share of minority groups								
Above 0.25	1.196	1.89	0.694	0.599	2.597	3.624	1.027	0.436
Below 0.25	0.937	1.44	0.502	0.541	1.775	2.573	0.799	0.507
By share of population with access to vehicle								
Below first quartile (0.89)	1.076	1.705	0.629	0.603	2.471	3.437	0.966	0.419
Above first quartile (0.89)	0.968	1.484	0.517	0.537	1.786	2.596	0.81	0.516
By share of population under poverty line								
Below median (0.16)	0.911	1.401	0.491	0.532	1.605	2.361	0.756	0.529
Above median (0.16)	1.079	1.678	0.6	0.576	2.31	3.253	0.943	0.454
By presence of dollar stores in 2010								
No dollar stores in 2010	1.018	1.649	0.631	0.665	1.818	2.659	0.84	0.547
Dollar stores present in 2010	0.989	1.513	0.524	0.527	1.992	2.843	0.852	0.478
Food deserts								
No supermarkets in 2010	1.172	1.75	0.577	0.514	1.701	2.43	0.73	0.499
Supermarkets present in 2010	0.987	1.53	0.544	0.555	1.969	2.824	0.855	0.491

Note: Factual corresponds to the expected number of stores under the market equilibrium. CF corresponds to the counterfactual expected number of stores. Δ (resp. % Δ) gives the difference (resp. percentage difference) between the market outcome and counterfactuals averaged over all markets. All demographic variables and store counts are at the market level.

7 Conclusion

Table 12: Retail proximity using actual store location and tract centroids

	Factual (store locations)		Factual (tract centroids)	
	Mean	Median	Mean	Median
Distance to nearest (in miles)				
Convenience store	0.84	0.60	0.83	0.59
Grocery/Combination store	1.37	1.03	1.33	1.00
Dollar store	1.01	0.79	0.97	0.73
Any store format	0.65	0.48	0.67	0.51
Number of stores within 0-2mi				
Convenience stores	4.32	3	4.44	3
Grocery/Combination stores	1.29	1	1.28	1
Dollar stores	2.30	2	2.39	2
Any store format	7.91	6	8.11	7
Number of stores within 2-5mi				
Convenience stores	3.48	1	3.38	1
Grocery/Combination stores	0.79	0	0.81	0
Dollar stores	1.76	1	1.66	1
Any store format	6.04	3	5.85	2

Note: Factual (store locations) uses the true location (latitude and longitude) of stores to compute measures of retail proximity. Factual (tract centroids) assumes stores are located at the population-weighted centroids of their census tract. All three measures of retail proximity are constructed by taking the (population-weighted) mean and median over all census block groups.

Table 13: Predicted retail proximity (aggregate)

	Factual			Counterfactual		
	Mean	Median	Pr($n > 0$)	Mean	Median	Pr($n > 0$)
Distance to nearest (in miles)						
Grocery/Combination store	1.06	0.84	0.67	1.01	0.80	0.82
Convenience store	0.94	0.75	0.87	0.87	0.69	0.96
Dollar store	0.91	0.67	0.92	0.93	0.70	0.78
Any store format	0.79	0.61		0.80	0.61	
Number of stores within 0-2mi						
Grocery/Combination stores	0.92	0.71		1.42	1.12	
Convenience stores	1.72	1.67		2.48	2.46	
Dollar stores	1.75	2		1.17	1	
Any store format	4.39	4.19		5.07	4.86	
Number of stores within 2-5mi						
Grocery/Combination stores	0.12	0		0.19	0	
Convenience stores	0.24	0		0.34	0	
Dollar stores	0.24	0		0.15	0	
Any store format	0.59	0		0.67	0	

Note: Measures of retail proximity are constructed by taking the (population-weighted) mean and median over all census block groups. $Pr(n > 0)$ gives the (population-weighted) mean probability that at least one store is operating in the market. Retail proximity to dollar stores is measured using actual realizations in the data in 2019 for Factual and in 2010 for Counterfactual.

Table 14: Predicted retail proximity (by number of locations in the market)

	Factual			Counterfactual		
	Mean	Median	Pr($n > 0$)	Mean	Median	Pr($n > 0$)
Markets with one location						
Distance to nearest (in miles)						
Grocery/Combination store	0.95	0.61	0.61	0.95	0.61	0.79
Convenience store	0.95	0.61	0.75	0.95	0.61	0.91
Dollar store	0.96	0.62	0.81	0.95	0.61	0.67
Any store format	0.95	0.61		0.95	0.61	
Number of stores within 0-2mi						
Grocery/Combination stores	0.84	0.60		1.39	1.17	
Convenience stores	1.28	0.87		1.99	1.51	
Dollar stores	1.17	1		0.76	1	
Any store format	3.29	2.83		4.14	3.67	
Number of stores within 2-5mi						
Grocery/Combination stores	0.08	0		0.14	0	
Convenience stores	0.12	0		0.19	0	
Dollar stores	0.10	0		0.06	0	
Any store format	0.29	0		0.39	0	
Markets with two locations						
Distance to nearest (in miles)						
Grocery/Combination store	1.02	0.77	0.64	0.99	0.75	0.80
Convenience store	0.93	0.72	0.84	0.87	0.67	0.95
Dollar store	0.89	0.64	0.92	0.89	0.68	0.81
Any store format	0.79	0.59		0.79	0.60	
Number of stores within 0-2mi						
Grocery/Combination stores	0.89	0.70		1.36	1.04	
Convenience stores	1.65	1.46		2.39	2.20	
Dollar stores	1.74	2		1.19	1	
Any store format	4.28	4.10		4.95	4.57	
Number of stores within 2-5mi						
Grocery/Combination stores	0.09	0		0.14	0	
Convenience stores	0.17	0		0.25	0	
Dollar stores	0.16	0		0.10	0	
Any store format	0.42	0		0.49	0	
Markets with three locations						
Distance to nearest (in miles)						
Grocery/Combination store	1.11	0.94	0.70	1.05	0.88	0.85
Convenience store	0.95	0.82	0.91	0.86	0.73	0.98
Dollar store	0.92	0.72	0.96	0.96	0.75	0.79
Any store format	0.75	0.62		0.76	0.63	
Number of stores within 0-2mi						
Grocery/Combination stores	0.96	0.75		1.48	1.15	
Convenience stores	1.89	1.91		2.67	2.74	
Dollar stores	1.91	2		1.25	1	
Any store format	4.76	4.81		5.40	5.29	
Number of stores within 2-5mi						
Grocery/Combination stores	0.16	0		0.24	0	
Convenience stores	0.32	0		0.45	0	
Dollar stores	0.35	0		0.20	0	
Any store format	0.83	0		0.90	0	

Note: Measures of retail proximity are constructed by taking the (population-weighted) mean and median over all census block groups. $Pr(n > 0)$ *gives the (population-weighted) mean probability that at least one store is operating in the market. Retail proximity to dollar stores is measured using actual realizations in the data in 2019 for Factual and in 2010 for Counterfactual.*

References

Aguirregabiria, Victor, and Arvind Magesan. 2013. *Euler equations for the estimation of dynamic discrete choice structural models.* Volume 31. Advances in Econometrics, 3–44.

Aguirregabiria, Victor, and Arvind Magesan. 2018. "Solution and Estimation of Dynamic Discrete Choice Structural Models Using Euler Equations." *SSRN Electronic Journal.*

Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential estimation of dynamic discrete games." *Econometrica* 75 (1): 1–53.

Aguirregabiria, Victor, and Junichi Suzuki. 2014. "Identification and counterfactuals in dynamic models of market entry and exit." *Quantitative Marketing and Economics* 12 (3): 267–304.

Allcott, Hunt, Rebecca Diamond, Jean-pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell. 2019. "Food Deserts and the Causes of Nutritional Inequality." *The Quarterly Journal of Economics* 1–52.

Altuğ, Sumru, and Robert A. Miller. 1998. "The Effect of Work Experience on Female Wages and Labour Supply." *Review of Economic Studies* 65 (1): 45–85.

Arcidiacono, P., and R. Miller. 2011. "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity." *Econometrica* 79 1823–1867.

Arcidiacono, Peter, Patrick Bayer, Jason R. Blevins, and Paul B. Ellickson. 2016. "Estimation of dynamic discrete choice models in continuous time with an application to retail competition." *Review of Economic Studies* 83 (3): 889–931.

Arcidiacono, Peter, and Robert A. Miller. 2019. "Nonstationary dynamic models with finite dependence." *Quantitative Economics* 10 (3): 853–890.

Bajari, Patrick, Lanier Benkard, and Jonathan Levin. 2007. "Estimating dynamic models of imperfect competition." *Econometrica* 75 (5): 1331–1370.

Bajari, Patrick, Victor Chernozhukov, Han Hong, and Denis Nekipelov. 2015. "Identification and Efficient Semiparametric Estimation of a Dynamic Discrete Game."

Basker, E., and M. Noel. 2009. "The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry." *Journal of Economics and Management Strategy* 18 977–1009.

Basker, Emek. 2005. "Job creation or destruction? Labor market effects of Wal-Mart expansion." *Review of Economics and Statistics* 87 (1): 174–183.

Beresteanu, Arie, Paul B. Ellickson, Sanjog Misra, and James C. Reeder. 2019. "The dynamics of retail oligopoly." (2001): 1–27, http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Dynamics+of+Retail+Oligopoly#0.

Chenarides, Lauren, Miguel Ignacio Gomez, Timothy J. Richards, and Koichi Yonezawa. 2021. "Retail Markups and Discount-Store Entry." *SSRN Electronic Journal* (2019): .

Collard-Wexler, Allan. 2013. "Demand Fluctuations in the Ready-Mix Concrete Industry." *Econometrica* 81 (3): 1003–1037.

Cummins, S., E. Flint, and S.A. Matthews. 2014. "New neighborhood grocery store increasing awareness of food access but did not alter dietary habits or obesity." *Health Affairs* 33 (2): 283–291.

Cummins, S., M. Petticrew, C. Higgings, A. Findlay, and L. Sparks. 2005. "Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment." *J Epidemiol Community Healthy* 59 1035–1040.

Dubowitz, T., M. Ghosh-Dastidar, D. Cohen, R. Beckman, E. Steiner, G. Hunter, K. Florez, C. Huang, C. Vaughan, and J. Sloan. 2015. "Changes in diet after introduction of a full service supermarket in a food desert." *Health Affairs* 34 (11): 1858.

Elbel, B., A. Moran, LB. Dixon, K. Kiszko, J. Cantor, C. Abrams, and T. Mijanovich. 2015. "Assessment of a government-subsidized supermarket in a high-need area on household food availability and children's dietary intakes." *Public Health Nutrition* 18 (15): 2881–2890.

Ellickson, P., S. Houghton, and C. Timmins. 2013. "Estimating Network Economies in Retail Chains: A Revealed Preference Approach." *RAND Journal of Economics* 169–193.

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

Ellickson, Paul B., Paul L.E. Grieco, and Oleksii Khvastunov. 2020. "Measuring competition in spatial retail." *RAND Journal of Economics* 51 (1): 189–232.

Ericson, R., and A. Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62 53–82.

Freyaldenhoven, Simon, Christian Hansen, Jorge Pérez Pérez, and Jesse M. Shapiro. 2021. "Visualization, Identification, and Estimation in the Linear Panel Event-Study Design." NBER Working Papers 29170, National Bureau of Economic Research, Inc, https://ideas.repec.org/p/nbr/nberwo/29170.html.

Grieco, Paul L. E. 2014. "Discrete games with flexible information structures: an application to local grocery markets." *RAND Journal of Economics* 45 (2): 303–340.

Hansen, Lars Peter, and Kenneth J. Singleton. 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica* 50 (5): 1269.

Hollenbeck, B. 2017. "The Economic Advantages of Chain Affiliation." *RAND Journal of Economics* 48 1103–1135.

Holmes, T.J. 2011. "The Diffusion of Wal-Mart and Economics of Density." *Econometrica* 79 253–302.

Igami, Mitsuru, and Nathan Yang. 2016. "Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada." *Quantitative Economics* 7 (2): 483–521.

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

Judd, Kenneth L. 1998. *Numerical Methods in Economics.* Volume 1. of MIT Press Books, The MIT Press, , https://ideas.repec.org/b/mtp/titles/0262100711.html.

Kalouptsidi, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues. 2020. "Linear IV regression estimators for structural dynamic discrete choice models." *Journal of Econometrics.*

Levi, Retsef, Elisabeth Paulson, and Georgia Perakis. 2019. "Optimal Interventions for Increasing Healthy Food Consumption Among Low Income Households." *MIT Sloan Research PaperNo. 6053-19.*

Levi, Retsef, Elisabeth Paulson, and Georgia Perakis. 2020. "Fresh Fruit and Vegetable Consumption: The Impact of Access and Value." *MIT Sloan Research Paper No. 5389-18.*

Liese, A., B. Bell, T. Barnes, N. Colabianchi, J. Hibbert, C. Blake, and D. Freedman. 2014. "Environmental influences on fruit and vegetable intake: results from a path analytic model." *Public Health Nutrition* 17 (11): 2595–2604.

Magnac, Thierry, and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes." *Econometrica* 70 (2): 801–816.

Presendorfer, Martin, and Philipp Schmidt-Dengler. 2008. "Asymptotic least squares estimators for dynamic games." *Review of Economic Studies* 75 (3): 901–928.

Rose, D., and R. Richards. 2004. "Food store access and household fruit and vegetable use among participants in the US Food Stamp Program." *Public Health Nutrition* 17 (11): 2595–2604.

Rust, John. 1994. "Structural Estimation Decision Processes." *Handbook of Econometrics* Chapter 51 3139–3139.

Rust, John. 2000. "Parametric Policy Iteration : An Efficient Algorithm for Solving Multidimensional DP Problems?."Technical report.

Seim, Katja. 2006. "An empirical model of firm entry with endogenous product-type choices." *RAND Journal of Economics* 37 (3): 619–640.

Ver Ploeg, M., and I. Rahkovsky. 2016. "Recent evidence on the effects of food store access on food choice and diet quality." *USDA Economic Research Service*.

Weatherspoon, D., J. Oehmke, A. Dembele, M. Coleman, T. Satimanon, and L. Weatherspoon. 2013. "Price and expenditure elasticities for fresh fruits in an urban food desert." *Urban Studies* 50 (1): 88–106.

Zheng, Fanyin. 2016. "Spatial Competition and Preemptive Entry in the Discount Retail Industry." *Working Paper*.

Zhu, Ting, and Vishal Singh. 2009. "Spatial competition with endogenous location choices: An application to discount retailing." *Quantitative Marketing and Economics* 7 (1): 1–35.

A Institutional Details

B Estimation Approach for Multi-Store Firms

In this section, we describe how we estimate the model for multi-store chains. Differences in choice-specific value functions for chains are derived as follows.

Potential entrants are long-lived and can delay entry into a later period (e.g., if a chain anticipates opening a distribution center closer to the market in the future). The choice-specific value functions from staying out and entering into location l are given, respectively, by

$$\begin{aligned} v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \beta\, \mathbb{E}\big(&-\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1}) \\ &+ \beta^2\, \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + +\gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})]|0, \mathcal{M}_{j,i,t}\big) \end{aligned} \tag{28}$$

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) = -\theta_i^{EC} + \beta\, \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})|l_+, \mathcal{M}_{j,i,t}] \tag{29}$$

The first equation shows that, if an entrant stays out, they internalize the option value from entering at a later period. Differences in choice-specific value functions can alternatively be expressed using current period CCPs as

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \log\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) \tag{30}$$

Combining Equation (28), Equation (29), and Equation (30), we obtain an optimality condition that depends only on the structural parameters and the known CCPs.

This optimality condition involves expectations over period $t+1$ and $t+2$ states. To avoid numerical integration over the high-dimensional state space, we dispose of the expectations by invoking the rational expectations assumption. Define the expectational errors as the difference between the expectations and the realizations of the random variables. For entrants, there are two expectations (over $t+1$ and $t+2$ states), therefore, the expectational errors $(w_{it}, u_{i,t+1})$ are defined as

$$\begin{aligned} w_{it} = \mathbb{E}\left[-\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})|0, \mathcal{M}_{j,i,t}\right] \\ - \left(-\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})\right) \end{aligned} \tag{31}$$

$$\begin{aligned} u_{i,t+1} = \mathbb{E}\left[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})|0, \mathcal{M}_{j,i,t}\right] \\ - \left(vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})\right) \end{aligned} \tag{32}$$

For incumbents, we define the expectational error v_{it} as

$$v_{it} = \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t}, l_+] \\ - (vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})) \tag{33}$$

These errors satisfy, for any function $g(.)$ of period-t variable, the moment conditions:

$$\mathbb{E}\left[g(\mathcal{M}_{j,i,t})'[w_{it}, u_{i,t+1}, v_{it}]\right] = \mathbf{0}_3 \tag{34}$$

Replacing these moment conditions by their sample counterparts (in the form of a linear IV regression as in Kalouptsidi et al. (2020)) will not, in general, yield consistent estimates of the structural parameters. Indeed, the error $u_{i,t+1}$ involves an expectation over $t+2$ states, denoted $\mathcal{M}_{j,i,t+2}$, *conditional* on $a_{it} = 0$. But the empirical distribution of $\mathcal{M}_{j,i,t+2}$ (in particular rivals' states) is conditional on the action a_{it} that was played in the data, which may or may not be 0.[13][14]

To address this selection problem, we define the weights

$$\psi_{a_1,a_2}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t}) = \frac{P(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t}, a_{it} = a_1)}{P(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t}, a_{it} = a_2)} \tag{36}$$

Lemma 1. *Let $\mathcal{M}_{j,i,t+2}(\tilde{a})$ be the random vector of $t+2$ states conditional on $a_{it} = \tilde{a}$. Then, the (re-weighted) random variable*

$$\mathcal{M}_{j,i,t+2}(\tilde{a}) \times \psi_{0,\tilde{a}}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t})$$

follows the distribution of $\mathcal{M}_{j,i,t+2}$ conditional on $a_{it} = 0$.

In constructing the sample moment counterparts, the data is reweighted using the CCPs ratio $\psi_{0,\widetilde{a}}$ as follows. Define the re-weighted expectational error, for each $a_{it} = \widetilde{a}$ as

$$\widetilde{u}_{i,t+1} = \mathbb{E}\left[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})|0, \mathcal{M}_{j,i,t}\right] \\ - \psi_{0,\widetilde{a}}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t})\left(vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})\right) \tag{37}$$

Defining the left-hand side variables for entrant chains as

[13] This selection problem only concerns $t+2$ rivals' states. In period $t+1$, rivals' states do not depend on a_{it} (conditional on the current state) because all firms take their actions simultaneously between t and $t+1$. Additionally, the selection problem does not concern exogenous variables (independent of a_{it}) or the firm's own state (which a deterministic function of a_{it}).

[14] To see why the exclusion restriction fails when $a_{it} = \widetilde{a}$, note that

$$E[u_{i,t+1}|\mathcal{M}_{j,i,t}] = \mathbb{E}\left[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})|0, \mathcal{M}_{j,i,t}\right] \\ - \mathbb{E}\left[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})|\tilde{a}, \mathcal{M}_{j,i,t}\right] \tag{35}$$

equals zero only when $\tilde{a} = 0$.

$$
\begin{aligned}
Y_{it}^{entrant} = \log\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) - \\
&- \beta[\gamma - \ln P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \\
&+ \beta[\gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})] \\
&+ \beta^2 \psi_{0,\widetilde{a}}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t})[\gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})]
\end{aligned} \tag{38}
$$

we obtain the structural parameters for entrants via the regression model

$$
\begin{aligned}
Y_{it}^{entrant} = &\left[-\theta_i^{EC} + \beta(vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i+)\right] \\
&- \left[-\beta\theta_i^{EC} + \psi_{0,\widetilde{a}}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t})\beta^2(vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i)\right] \\
&+ (v_{it} - u_{it} - \beta\widetilde{u}_{i,t+1})
\end{aligned} \tag{39}
$$

where regressors entering the variable profit function in $t+1$ and $t+2$ are instrumented using the values of these regressors in period t.

For an incumbent chain with one store in location l^*, possible actions are to do nothing, build a second store, or close its existing store (note we allow the entry cost for the second store $\widetilde{\theta}_i^{EC}$ to be different than for that of the first store θ_i^{EC}). The corresponding choice-specific value functions are given by

$$
v_{i,t}^{\mathbf{P}}(l_-^*, \mathcal{M}_{j,i,t}) = vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \tag{40}
$$

$$
\begin{aligned}
v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = &\, vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \\
&+ \beta\, \mathbb{E}[vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln P_{i,t+1}(l_-^*|\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t}, l_+]
\end{aligned} \tag{41}
$$

$$
\begin{aligned}
v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) = &\, vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i - \widetilde{\theta}_i^{EC} \\
&+ \beta\, \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) + vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \\
&\quad + \gamma - \ln P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t}, l_+] \\
&+ \beta^2\, \mathbb{E}[vp_{i,l^*}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-^*|\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t}, l_+]
\end{aligned} \tag{42}
$$

We can derive two sets of optimality conditions (e.g., do nothing vs. build second store, and do nothing vs. close existing store), by taking differences in the choice-specific value functions and using their CCP representation

$$
v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \log\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) \tag{43}
$$

$$
v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(l_-^*, \mathcal{M}_{j,i,t}) = \log\left(\frac{P_{i,t}(0|\mathcal{M}_{j,i,t})}{P_{i,t}(l_-^*|\mathcal{M}_{j,i,t})}\right) \tag{44}
$$

As with entrants, we can dispose of the expectations using the rational expectation assumption and derive moment conditions. As before, period $t+2$ states, appearing in Equation (42), are *conditional* on the action $a_{it} = l_+$. However, the empirical distribution of $\mathcal{M}_{j,i,t+2}$ is conditional on the a_{it} played in the data, which may or may not be l_+. To correct for this selection problem in forming the IV regression equation, any term involving $\mathcal{M}_{j,i,t+2}$ is re-weighted using $\psi_{l_+,\tilde{a}}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t})$, where $\tilde{a}$ is the action played in the data.

C Details of Solution Method for Dynamic Game

Here we provide a detailed overview of how we solve the dynamic game for counterfactual scenarios. The dynamic game is solved via policy iteration (Judd (1998), Rust (2000)). This approach consists in iterating repeatedly between two steps: a given iteration starts by updating the ex-ante and choice-specific value functions given the current vector of CCPs (policy evaluation), then these value functions are used to update the vector CCPs (policy improvement). The algorithm iterates until value functions and CCPs converge, up to a pre-defined tolerance level.

As the state space is extremely large (with continuous state variables), it is computationally prohibitive to solve for value functions and CCPs at all states. We fix the demographic state variables (income, population, etc.) to their value realized in the data and assume their transitions are deterministic. For periods outside of our sample, i.e, $t \geq T+1$ (where period $T+1$ corresponds to the year 2020), we assume that these demographic variables become stationary and equal the expected value given their realizations in period T.[15]

The dynamic game is solved by backward induction starting from the first period outside our sample, i.e., $t = T+1$. In this period we iterate over the following steps:

1. Initialize the vectors of CCPs for each firm and state $\mathbf{P}_{i,T+1}$. If firm i is a potential entrant, $\mathbf{P}_{i,T+1}$ is a vector indexed by the state and locations $(\mathcal{M}_{i,j,T+1}, l_+)$ giving the CCP of entry into location l in state $\mathcal{M}_{i,j,T+1}$.[16] If i is an incumbent, $\mathbf{P}_i$ is a vector indexed by $(\mathcal{M}_{i,j,T+1}, a_{it})$ giving the CCP of choosing action a_{it} (remaining active for single-store firms, or building an additional store/remaing active/closing an existing store for chains) in state $\mathcal{M}_{i,j,T+1}$.

2. Form the transition matrix from state $\mathcal{M}_{i,j,T+1}$ to state $\mathcal{M}_{i,k,T+2}$ for each firm type, *conditional* on the action played a. Denote this transition matrix $\mathbf{F}_{i,T+1}(a)$. If firm i plays a terminal action (e.g, an incumbent single-store firm exits) the continuation value is zero, therefore, knowledge of this transition matrix is not necessary.

3. Update the conditional choice-specific value function, leveraging finite dependence. Let $\mathbf{v}_{i,T+1}(a)$ denote a vector collecting the choice-specific value function of firm i if it plays action a for all states $(\mathcal{M}_{i,j,T+1}$. This vector satisfies the equality (in matrix form)

[15] To compute this expectation, we assume that demographic variables in each location evolve according to AR-1 processes, where the innovation shocks are allowed to be geographically correlated across locations within a market.

[16] Demographic variables are fixed, therefore, different states correspond to different realizations of the spatial market structure (number of stores by type in each location).

$$\mathbf{v}_{i,T+1}(a) = \boldsymbol{\pi}_{i,T+1}(a) + \beta\mathbf{F}_{i,T+1}(a)[\mathbf{v}_{i,T+1}(exit) + \gamma - \ln(\mathbf{P}_{i,T+1}(exit))] \tag{45}$$

where $\boldsymbol{\pi}_i(a)$ is a vector giving single-period profits. For instance, if i is a potential entrant and $a = l_+$, then $\boldsymbol{\pi}_i(l_+) = -\theta_i^{EC}$. $\mathbf{v}_{i,exit}$ is only a function of the single-period payoff.

4. Update the vectors of CCPs as

$$\mathbf{P}'_{i,T+1}(a) = \frac{\exp\left(\mathbf{v}_{i,T+1}(a)\right)}{\sum_{\widetilde{a}} \exp\left(\mathbf{v}_{i,T+1}(\widetilde{a})\right)} \tag{46}$$

If the maximum absolute difference between $\mathbf{P}_{T+1}$ and $\mathbf{P}'_{T+1}$ is less than the pre-defined tolerance level, the procedure stops and $\mathbf{P}'_{T+1}$ is saved. If not, define updated CCPs as a convex combination of old and new CCPs $\alpha\mathbf{P}_{i,T+1} + (1-\alpha)\mathbf{P}'_{i,T+1}$ for each player i and return to Step 2.

This iterative approach yields the equilibrium CCPs for periods $t \geq T+1$, denoted $\mathbf{P}^*_{T+1}$. Proceeding backwards, the equilibrium CCPs in period t, given optimal CCPs in period $t+1$ ($\mathbf{P}^*_{t+1}$) are obtained by iterating over Steps 2 to 4 above, with the exeption that Equation (45) is replaced by

$$\mathbf{v}_{i,t}(a) = \boldsymbol{\pi}_{i,t}(a) + \beta\mathbf{F}_{i,t}(a)[\mathbf{v}_{i,t+1}(exit) + \gamma - \ln(\mathbf{P}^*_{i,t+1}(exit))] \tag{47}$$

where equilibrium CCPs in $t+1$ are used. As markets are independent, we solve the model for each market separately.

D Model Fit

E Robustness Checks

F Supplementary Tables

Table 15: Multinomial logit of single-store firms' choice

	Dependent variable: Firm is active in location l			
	Grocery/Combination (1)	Grocery/Combination (2)	Convenience (3)	Convenience (4)
Entrant	−0.613 (1.016)	−6.208 (1.116)	−0.407 (0.801)	−4.786 (0.719)
Incumbent	5.062 (1.014)	−0.673 (1.121)	5.538 (0.807)	1.101 (0.722)
Location-level characteristics				
Population (0-2 mi)	0.316 (0.038)	0.334 (0.057)	0.208 (0.025)	0.222 (0.026)
Population (2-5 mi)	−0.010 (0.017)	−0.017 (0.016)	−0.016 (0.013)	−0.013 (0.013)
Income per capita (0-2 mi)	−0.159 (0.090)	0.045 (0.105)	−0.019 (0.068)	0.145 (0.060)
Income per capita (2-5 mi)	0.011 (0.013)	0.016 (0.011)	0.010 (0.009)	0.006 (0.009)
Cost shifters				
Distance to DG distribution center	0.061 (0.027)	0.052 (0.033)	0.026 (0.029)	0.004 (0.026)
Distance to DT distribution center	0.040 (0.038)	0.053 (0.044)	−0.024 (0.030)	−0.015 (0.026)
Distance to FD distribution center	0.067 (0.034)	0.047 (0.045)	0.028 (0.033)	0.029 (0.031)
Median residential rent	−0.108 (0.102)	0.089 (0.108)	−0.123 (0.069)	−0.025 (0.065)
Measures of competition				
Number of rival chain stores (0-2 mi)	−0.125 (0.046)	−0.284 (0.053)	−0.107 (0.037)	−0.242 (0.036)
Number of rival chain stores (2-5 mi)	−0.040 (0.046)	−0.015 (0.053)	−0.110 (0.036)	−0.072 (0.036)
Number of rival grocery/combination (0-2 mi)	0.065 (0.033)	−0.060 (0.041)	0.016 (0.029)	−0.096 (0.027)
Number of rival grocery/combination (2-5 mi)	−0.082 (0.038)	−0.139 (0.044)	−0.040 (0.034)	−0.083 (0.030)
Number of rival convenience (0-2 mi)	−0.083 (0.032)	−0.235 (0.037)	−0.054 (0.024)	−0.170 (0.024)
Number of rival convenience (2-5 mi)	−0.033 (0.036)	−0.041 (0.042)	0.010 (0.026)	−0.018 (0.025)
Market-level characteristics				
Population	−0.452 (0.064)	−0.249 (0.075)	−0.416 (0.052)	−0.244 (0.049)
Number of gas stations	−0.099 (0.037)	−0.016 (0.041)	−0.168 (0.032)	−0.148 (0.030)
Number of drug stores	0.095 (0.052)	0.031 (0.059)	0.018 (0.048)	−0.023 (0.043)
Number of supermarkets	−0.0002 (0.056)	−0.103 (0.058)	0.153 (0.053)	0.087 (0.049)
Business Density	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
Observations	28,144	28,144	82,180	82,180
Log Likelihood	−13,074.510	−12,730.870	−38,249.850	−37,639.490

Note: Standard errors are clustered by market. The baseline alternative is "firm is inactive" (either by exiting or staying out). Dollar figures are in 2010$. Business density is defined as the maximum number of establishments simultaneously operating in location l over the period 2008-2019. Distance to distribution center is at the market level, residential rent is at the location level. All continuous variables and store counts are in log.