Dynamic Entry & Spatial Competition: An Application to Dollar Store Expansion

El Hadi Caoui — U of Toronto

Brett Hollenbeck — UCLA Anderson

Matthew Osborne — U of Toronto

June 21, 2024*

Abstract

We incorporate spatial differentiation into a dynamic oligopoly game played by multi-store and single-store retailers. We propose a tractable estimation strategy based on the ECCP estimator of Kalouptsidi et al. (2020), adapted to games. The model is used to study dollar store chains' expansion in the US since 2008 and its impact on spatial market structure. We show that accounting for rival stores' spatial reallocation is crucial to quantify the net impact of dollar stores. Grocery stores differentiate from dollar stores by strategically relocating or entering new locations with higher income and lower population density.

Keywords: retail competition, spatial differentiation, dollar store, dynamic games

JEL Classification: L13, D43, L51

---

*Caoui: elhadi.caoui@rotman.utoronto.ca. Hollenbeck: brett.hollenbeck@anderson.ucla.edu. Osborne: matthew.osborne@rotman.utoronto.ca.

Portions of this article were previously included in an earlier draft titled "The Impact of Dollar Store Expansion on Local Market Structure and Food Access." We are grateful to Emek Basker, Paul Ellickson, and Mathieu Marcoux for helpful discussions of this article. We thank Victor Aguirregabiria, Hunt Allcott, JP Dubé, Limin Fang, Paul Grieco, Sylvia Hristakeva, Yao Luo, Paul Scott, Chuck Weinberg and seminar and conference participants at the University of Toronto, UCLA, Rochester, UVA, Yale, JHU, UT Austin, U of Melbourne, Washington University of Saint Louis, University of Iowa, Stanford GSB, Chicago Booth, BLS, IIOC, SICS, CEA, ISMS Marketing Science Conference, ET Symposium, and the Food Access at Dollar Stores (FADS) Workshop for their helpful comments and suggestions.

1 Introduction

For many decades, new retail formats have emerged and reshaped the US retail sector. The most recent development in this ongoing evolution is the dramatic rise of dollar store chains. What distinguishes this expansion is not only the speed of their growth—with the top three chains, Dollar General, Dollar Tree, and Family Dollar, collectively opening stores at a rate of 3.75 stores per day over the past decade—but also their ubiquity. By 2021, over 75% of the US population lives within five miles of a dollar store, underscoring the extensive footprint and market penetration of these retailers.

Originating in the 1950s, these chains initially targeted small towns and rural areas. Their growth accelerated following the 2008 recession as worsening household finances increased the demand for low-priced, small-format items. Notable subsequent developments include the 2007 acquisition of Dollar General by a private equity firm, which enhanced logistics and growth, and the 2015 merger of Dollar Tree and Family Dollar to compete more effectively with Dollar General. The format's expansion has been rapid; from 2018 to 2021, dollar stores constituted about half of all new US retail openings, significantly outpacing the combined stores of Walmart, CVS, Walgreens, and Target.

The expansion of dollar store chains has come under increasing scrutiny from policymakers, particularly concerning their impact on retail markets. Concerns center on whether the rapid proliferation of these stores displaces existing retailers or deters new competitors, thereby limiting the variety of goods available to consumers. This issue is especially critical in areas where the entry of dollar stores leads to the closure of local grocery outlets, impacting access to perishable foods. In response, many municipalities have enacted regulations to curb the entry of new dollar stores or have implemented dispersal policies to limit their store density.[1]

Despite ongoing public and policy debates, empirical evidence on the impact of dollar

[1] Cities that have banned or restricted dollar store entry include Birmingham, AL; Atlanta, GA; New Orleans, LA; Akron, OH; Oklahoma City, OK; Tulsa, OK; and Fort Worth, TX. See https://ilsr.org/dollar-store-restrictions/ [Last accessed: May 15, 2024].

stores is still emerging. This article uses new data and a structural approach to examine their effects on market structure. Our analysis centers around two key aspects of this expansion. The first is spatial competition: dollar stores and rival store formats are spatially differentiated. We explicitly model the strategic responses of competing retailers, in terms of location choices, to the entry of dollar stores. One would expect the impact of dollar stores to extend beyond the locations they enter, as in equilibrium, retail activity may relocate to other parts of the market. The second aspect is the dynamic nature of this expansion: over the sample period, dollar stores' costs of operating stores decreases substantially, largely due to a denser network of distribution centers. Accounting for such sources of non-stationarities is crucial to accurately model dollar stores' expansion strategy.

We build a dynamic structural model of the entry and exit choices of dollar store chains and their local competitors and incorporate location choices as an integral part in firms' strategy space. Local competitors include grocery and conveniences stores.[2] In practice, each store's entry and exit decisions are modelled as a dynamic oligopoly game following Ericson and Pakes (1995) but with a spatial component along the lines of Seim (2006).

The purpose of this model is twofold. First, it allows us to account for equilibrium effects from dollar store expansion. While the number of rival stores may decrease in locations – defined as census tracts – near a dollar store, new entrants may subtitute to locations further away from the dollar store, in a given market. Modeling the long-term spatial market structure is therefore necessary to understand the *net* effects and distributional impact of this expansion. Second, the model provides estimates of the size of dynamic entry and investment costs, as well as the competitive effects between different store formats. This enables us to explore various explanations for the success of the dollar store format.

[2]We follow the definition of "grocery stores," which are distinct from "supermarkets and supercenters," used by the USDA in its SNAP retailer panel. Grocery stores are primarily focused on selling food and consumable products, carry all four staple food categories, have annual revenue below $2m, and are generally independently owned. Supermarkets/centers have annual revenue above $2m, carry all four staple food categories, are part of a retail chain, and typically have ten or more checkout lanes with registers, bar code scanners, and conveyor belts This definition has been used to define supermarkets in previous studies, e.g., Ellickson and Grieco (2013).

The key challenges in estimating this game in a tractable way are the complex nature of spatial competition, which results in a high-dimensional state space, and the non-stationary dynamics stemming from the growth over time in dollar stores' distribution center networks, which reduces their fixed costs of operating stores.

We take advantage of the fact that firms face a terminal choice when deciding whether or not to exit, a special case of finite dependence (Arcidiacono and Miller (2011), Arcidiacono and Ellickson (2011), Arcidiacono and Miller (2019)). This property simplifies estimation of the game substantially as it allows us to recover 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). The latter article combines insights from the finite dependence approach and the GMM-Euler approach (Aguirregabiria and Magesan (2018)) to propose a method (ECCP) that circumvents integration over the high-dimensional state space.[3] Scott (2013) uses this approach to estimate a dynamic model of agricultural land use, leveraging renewal actions. As a methodological contribution, we extend the ECCP estimator from single-agent problems to dynamic games with terminal actions, highlighting and addressing a selection problem arising in games, as well as extending it to a setting with long-lived chain entrants. This extension of the ECCP estimator is useful, more generally, because it simplifies significantly the estimation of high-dimensional dynamic games with finite-dependence.

To estimate this model, we rely on data from several sources. We track the number and type of retail stores, including dollar stores, across the US using the Supplemental Nutrition Assistance Program (SNAP) Retailer panel, a yearly panel of SNAP-authorized retailers from 2008 to 2019. An advantage of this dataset is that it covers small independent retail stores, which are typically absent from other retail census used in the literature. We combine this

[3] Kalouptsidi et al. (2020) show how to incorporate serially correlated market-level unobserved heterogeneity in dynamic discrete choice problems, without having to specify the transition process for the market-level variables. The latter is achieved by invoking rational expectations and replacing expected behavior with observed (realized) behavior in the data. One benefit of this approach is that it does not require integrating over the state space when evaluating period-ahead value functions.

with data on dollar store distribution center locations and opening dates and demographic information at the census-tract level from the US Census.

We restrict our analysis to small and medium-sized urban and rural markets. We focus on these markets because they are central to the policy debate on dollar store entry restrictions: e.g., approximately 80 percent of Dollar General stores serve communities of 20,000 or fewer people. Moreover, the relatively small number of retail players in these markets make them especially susceptible to closure of existing stores. Finally, when accounting for retailers' location choices, computational reasons limit the size of markets for which we can solve for an equilibrium.

Our estimation results provide a number of direct findings. Dollar store chains have significantly lower costs of opening a new store than their independent rivals and are substantially more profitable. Grocery and convenience stores suffer significant losses when located within 0-2 miles (mi) of a dollar store. Estimates also point to strong demand cannibalization within chain in the 0-2mi radius. Further, dollar stores benefit from scale economies when locating stores in moderate proximity (2-5mi radius), through lower operating costs. The growth in dollar stores' network of distribution centers leads to a reduction in store operating costs over our sample period. The results suggest that dollar stores are able to enter cheaply and operate with low fixed costs (increasingly so over time) due to store-level economies of density and a larger network of distribution centers.[4]

The model is used to evaluate how dollar store expansion has impacted the retail landscape. We quantify the spatial reallocation of retail activity and cumulative reduction in the number of grocery and convenience stores resulting from dollar store chains' expansion across different market types, as well as the impact on consumer outcomes that are central to the current policy debate, e.g. travel costs and retail proximity. We find that markets experience on average a 31% to 33% decrease in the number of grocery and convenience

[4] These findings are consistent with stores' product assortments: by focusing sales on high-margin consumables, dollar stores compete strongly with nearby grocery stores who are left relying on the sale of low-margin produce with high operating costs. Our companion article, Caoui et al. (2024), uses scanner data to document changes in spending by retail format and food categories.

stores. The largest reductions materialize in markets with larger populations, lower average income, and higher share of minority groups. Within market, we find that locations with higher populations, lower incomes, and a higher share of households with no vehicle (a proxy for transportation costs) experience a larger reduction in the number of grocery and convenience stores.

These effects are, however, not uniform across all locations in a given market. Some locations see an increase in the number of grocery and convenience stores in response to dollar store entry. Indeed, entry by dollar store chains changes the *relative* attractiveness of locations in a market. We decompose the aggregate market-level impact of dollar stores into direct and indirect effects. Direct effects corresponds to the dollar-store induced exits of incumbent stores. Indirect effects refer to the reallocation of retail activity to locations that are less attractive to dollar stores, with the intent to spatially differentiate. Indirect effects are larger in markets with more locations, which offer greater scope for spatial differentiation. The net impact of dollar store entry on the grocery store format at the market level is 14% lower when indirect effects (reallocation) are taken into account. Grocery stores tend to shift to locations with lower population and higher income. On the other hand, convenience stores' ability to spatially differentiate is more limited: this format favors locations that are similar to dollar stores' most preferred locations.

Related Literature. This article contributes to three areas of economic research. The first focuses on the evolution of the US discount retail sector. Studies have analyzed the impact of large retailers like Walmart and K-Mart on market structure (Jia (2008), Zhu and Singh (2009), Basker and Noel (2009), Igami (2011), Ellickson and Grieco (2013), Grieco (2014)), labor markets (Basker (2005)), and the role of economies of scale and density (Holmes (2011), Ellickson et al. (2013)). Notably, Walmart's entry has primarily affected larger chain retailers within two miles (Ellickson and Grieco (2013)), whereas smaller retailers were less impacted due to travel costs and horizontal differentiation. By contrast, the entry strategy and market positioning of dollar stores suggest they may directly compete with small local retailers, significantly impacting local retail markets and prompting distinct policy considerations.

Our findings contribute to emergent research on the dollar store format. Through event study analysis of large numbers of dollar store entries, Caoui et al. (2024) demonstrate that dollar store expansion is associated with a decline in the number of grocery stores and reduced fresh produce consumption among low-income households with high travel costs. Feng et al. (2023) highlight that, despite their limited food selections, dollar stores are capturing an increasing share of food purchases, particularly in smaller markets. Lopez et al. (2023) shows that dollar store entry is associated with grocery store closures, lower retail employment and sales, with more pronounced effects in rural areas. Concurrently, Chenarides et al. (Forthcoming) use a dynamic model and data from Texas to show that dollar stores generally benefit supermarkets by displacing smaller competitors. Their approach differs from ours as they assume stationarity and model spatial differentiation as a choice of market-level store density (rather than location choices). Explicitly modelling firms' location choices is important because dollar store entry may lead to a spatial reallocation of rival store activity: e.g., the number of grocery stores may decrease near dollar stores but increase further away in a market. These equilibrium effects need to be accounted for when assessing the net impact of this expansion. Lastly, Schneier et al. (2023) examines the effects of the first dollar store entry on prices paid and shopping behavior, and Cao (2022) studies their impact on retail variety and the availability of private-label products.

Finally, this article 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), Fang and Yang (2022)). We depart from the existing literature in two ways. First, we account for non-stationary dynamics inherent to the discount store industry: over the sample period, dollar store chains have been expanding their networks of distribution centers; incorporating this dynamic aspect of the industry is clearly important to better match observed entry patterns. Second, most previous dynamic game studies (Zheng (2016) and Bodéré (2023) being exceptions) abstract from the spatial nature of competition. Because retail location choices are crucial in shaping the competitive environment (Ellickson et al. (2020)), we model firms' entry decisions into spatially differentiated locations as in Seim (2006) and Datta and Sudhir (2013).

The rest of the article proceeds as follows: Section 2 describes the data and institutional details and provides descriptive statistics. Section 3 introduces the dynamic oligopoly model. Section 4 discusses the identification and estimation of the dynamic game. Section 5 shows the estimation results. Section 6 presents the counterfactual analysis. Section 7 concludes.

2 Industry Background, Data, and Descriptive Statistics

In this section, we discuss the evolution of dollar store chains, outline our data sources, and present descriptive statistics on the industry and our sample.

The dollar store concept, pioneered by Dollar General in 1955, featured a broad array of low-cost basic goods at a $1 price point. This model gained popularity, prompting similar strategies from competitors like Family Dollar, established in 1959. By the 2000s, the market consolidated around three main players: Dollar General, Family Dollar, and Dollar Tree. These chains compete by offering low prices, typically through a single or few fixed price points.

Unlike other discount retailers like Aldi, dollar stores achieve savings not by limiting selection and focusing on private labels but by offering a moderate assortment of both major brands and private labels.[5] A dollar store typically occupies 8,000-12,000 sq ft and stocks 10,000-12,000 SKUs. They reduce costs by employing few employees and by limiting their perishable food offerings. Their inventory mainly includes basic consumables, seasonal items,

[5]For instance, private labels accounted for 12% of Dollar General's merchandise mix (Shih et al. (2019)). Hristakeva (2022) discusses retailers' strategic assortment choices to secure preferential supplier contracts.

and irregular or outdated products from major brands. Their entry strategy involves entering small, low-income markets overlooked by larger retailers, which we will explore in more detail below.

Dollar store chains have expanded rapidly, especially since the 2008 recession. By 2021, Family Dollar, Dollar General, and Dollar Tree operated approximately 7,100, 18,000, and 4,350 stores respectively, totaling nearly 30,000 stores. This count surpasses the combined totals of Wal-Mart (5,300 stores), Target (1,900 stores), CVS (9,900 stores), and Walgreens (9,300 stores) and is larger than Subway, the largest US restaurant chain with 21,000 locations, and comparable to the number of Starbucks worldwide. In 2019, these chains generated a combined revenue of $47 billion.

In 2015, Dollar Tree and Family Dollar merged, citing potential benefits such as targeting a broader customer base, optimizing their real estate, and exploiting synergies in sourcing and distribution.[6] However, integration has been slow; the chains operated independently in terms of store management and supply chains through 2019. For instance, store support centers remained separate, and distribution networks operated independently until 2020 (Dollar Tree (2018) and Figure A3 discussed below).[7] This justifies treating the chains as distinct entities in our study period ending in 2019.

Data. We combine several data sources to study dollar store chains' expansion and its effect on local market structure.

The SNAP Retailer panel is a yearly dataset of SNAP-authorized retailers from 2008 to 2019, encompassing over 400,000 US retailers. It includes information on store location, chain affiliation, store type, and covers stores such as dollar stores, convenience stores, grocery stores, drugstores, gas stations, supermarkets, and supercenters. We classify stores by type using the store type variable from this panel. To our knowledge, the SNAP retailer data is novel in the economics literature.[8] The primary advantages of this public dataset are its

[6] "*Dollar Tree completes acquisition of Family Dollar.*" Dollar Tree, Press Release, July 6, 2015.

[7] In 2020, two new distribution centers serving both brands opened in Ocala, FL, and Rosenberg, TX.

[8] The dataset has been used in the geography literature studying retail proximity (e.g., Shannon et al. (2018)).

comprehensiveness and annual frequency. Crucially for this study, the panel includes small independent stores, which are typically absent from other retail Census data used in the literature.

A drawback of this dataset is that entry into the SNAP program may not necessarily indicate the start of operation of a physical store.[9] As the SNAP program began in 2008, there may have been delays in stores joining the program in the initial years. We address this concern in two ways. For chains, we compare store counts in the SNAP panel against publicly disclosed counts in chains' annual reports, finding no significant discrepancies. For independent stores, this approach is not possible. Instead, we drop the first two years from the sample, restricting our analysis to 2010-2019.

We collect market-level data on demographic characteristics from the Census and ACS at the Census tract level. This lets us study how market characteristics and consumer demographics affect dollar stores' and other retailers' entry behavior and profits. We also gather data on the locations and opening dates of distribution centers for the three major dollar store chains over time.

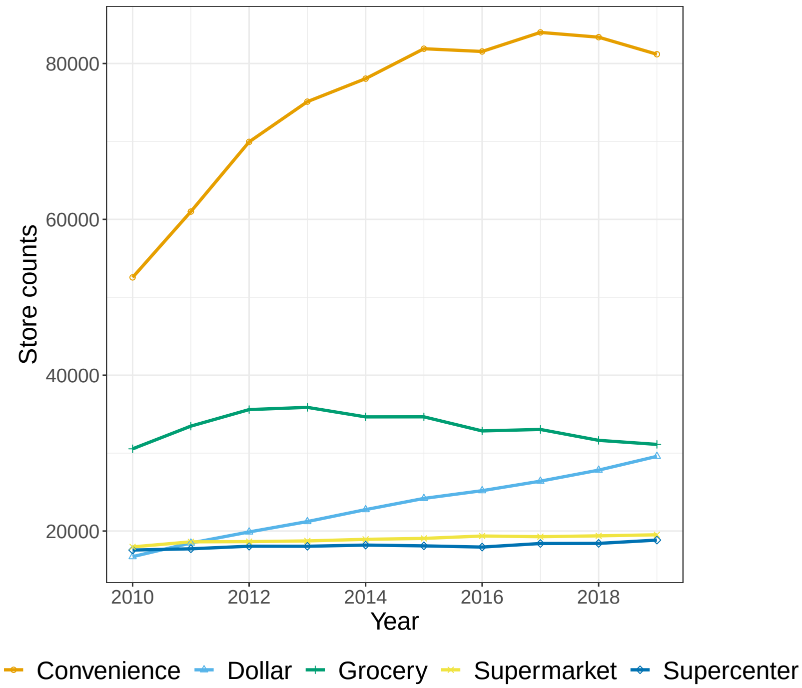
Figure 1 shows the total number of stores at the national level from 2010 to 2019. The three major dollar store chains added 12,870 stores, resulting from 14,554 store entries and 1,684 exits. Independent retail stores also grew, driven by convenience stores with 71,010 entries and 42,363 exits. The number of grocery stores decreased by 13% from its peak in 2012. The number of supermarkets and supercenters remains relatively stable over the sample period.

Appendix Figure A2 shows the evolution of dollar store chains' distribution centers from 2000 to 2019. The number of centers increased from 14 to 42. This reduced the average distance between a market and the nearest distribution center by roughly 125 miles for Dollar General and slightly less for the other two chains.[10] Appendix Figure A3 shows the locations

[9] Byrne et al. (2022) study the impact of SNAP adoption around the 2008 recession on store sales and SNAP-eligible household purchases.

[10] Details about the market definition are provided below.

Figure 1: Store counts by firm type



Notes: This figure shows store counts for the different store types in the SNAP Retailer panel data. Gas stations and drug stores are excluded.

of distribution centers in 2019.

We use Census demographic data to document consumer heterogeneity across locations with varying dollar store densities. Table 1 shows summary statistics of Census demographics for locations entered by dollar store chains before 2010, during 2010-2019, and those never entered. Locations are defined at the Census tract level. Dollar store entry occurs in areas with lower income per capita and rents, and a higher share of the population that is Black or below the poverty line. Entered locations are also closer to distribution centers than non-entered locations.

Market Definition. In our empirical application, we restrict the sample to small and medium-sized isolated markets for several reasons. First, dollar store entry strategy has historically targeted smaller urban and more rural markets. As a result, these retail markets have been most impacted by dollar store growth. Second, these markets are the primary target of the policy debate around restricting dollar store expansion due to their greater

Table 1: Market Summary Statistics

	(1) Pre-2010 Entry Only	(2) 2010-2019 Entry	(3) Never Entered
N	9778	12872	50378
Mean Population	4689.9	4962.8	4263.9
	(2193.5)	(2566.1)	(2295.8)
Mean Income	22686.3	23538.8	31315.2
	(7520.7)	(8178.7)	(16696.5)
Mean Residential Rents	753.9	785.9	1064.8
	(252.6)	(275.4)	(455.8)
Mean Share White	.738	.739	.713
	(.24)	(.253)	(.254)
Mean Share Black	.166	.162	.127
	(.225)	(.234)	(.210)
Mean Share in Poverty	.176	.165	.136
	(.108)	(.106)	(.118)
Share HH w/ Vehicle Access	.911	.917	.904
	(.084)	(.091)	(.135)
Mean Distance to DG DC	157.8	171.8	227.9
	(134.1)	(142.5)	(292.9)
Mean Distance to DT DC	188.2	191.4	189.5
	(111.2)	(119.7)	(239.8)
Mean Distance to FD DC	190.1	207.2	275.2
	(124.2)	(126.1)	(296.3)

Notes: Unit of observation is the Census Tract. Means are computed using 2019 data. Standard deviation across tracts appears in parentheses below each row.

susceptibility to concerns around food access. Third, when accounting for spatial differentiation, computational reasons limit the size of markets for which we can solve the dynamic game. Therefore, we follow the approach in Seim (2006) and define markets as cities and incorporated places with populations between 5,000 and 150,000, and exclude markets within 10 miles of a city with a population greater than $5,000$ or within 20 miles of a city with a population greater than $25,000$. Each market is partitioned into locations, which we define at the Census tract level.[11]

We focus on competition between dollar store chains and independent grocery and convenience stores.[12] The previous literature on retail competition has shown that competition

[11] Census tracts may cross the boundaries of a Census place (city or town), in such instances, we define a location as the intersection between the Census tract and the Census place. These intersections are obtained using the Census Bureau's geographic correspondence engine Georr. See https://mcdc.missouri.edu/applications/geocorr2014.html.

[12] The grocery store type includes stores that the USDA classifies as "Grocer" or as "Combination Grocery/Other." Stores that the USDA classifies as "Supermarket/Supercenter" are in the vast majority part of a chain (e.g., Walmart, Safeway).

between grocery stores operates at relatively close range (1–2mi). As a result, in our empirical specification, we allow the impact of competition to differ across distance bands. We exclude gas stations, drugstores, and supermarket chains from the set of players. In the case of supermarkets and big box retailer, previous literature (Ellickson and Grieco (2013), Grieco (2014)) has shown that their impact on small horizontally differentiated retailers is limited. Nonetheless, we do control for the presence of the latter three formats (gas stations, drugstores, supermarkets and supercenters) as potential determinants in players' payoffs, but we treat their evolution as exogenous.

Table 2 shows market and location-level demographic characteristics for our sample of 846 markets and Table 3 shows statistics on their market structure. These markets are small in terms of population and with low average incomes. Average income per capita is $20,352 compared to roughly $58,000 for the US as a whole. A market contains 5.8 locations in total, 4.3 of which are "commercial" location while the remainder are "residential." We define "commercial" locations as locations in which at least one store (of any type, e.g., dollar store, gas stations, drugstores, supermarkets) was active in any year between 2008 and 2019.

A typical market contains 2 dollar stores, 1 grocery store, 3 convenience stores, 2 gas stations, and 3 supermarkets, but with wide variation across markets. We note the relatively high number of supermarkets given our market definition. Supermarkets' catchment areas are in general much larger than our definition of a market (i.e., Census place). Our market definition is motivated by a focus on spatial competition between dollar stores and other small retailers. For supermarkets, markets are usually defined at larger geographic units, e.g., at the county or MSA level.

Table 4 shows descriptive statistics on the entry and exit dynamics of dollar store chains and single-store competitors. For each action, the table shows the proportion of firm-market-year observations. Single-store firms (grocery and convenience stores) have higher turnover than dollar store chains, with entry rates of 12% versus 6%, and exit rates of 10% versus

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 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.

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 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 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				

1.1%. Occasionally, dollar store chains open or close more than one store in the same market-period. An incumbent chain may also open a second store in a market where it already operates (2.7% of observations).

Table 4: Descriptive Statistics: Entry and Exit Dynamics

	Potential Entrants			Incumbents	
	Chains	Single-store		Chains	Single-store
Don't Enter	0.938	0.879	Close 2 stores	0.000	
Build 1 store	0.061	0.121	Close 1 store	0.011	0.105
Build 2 stores	0.001		Do nothing	0.961	0.895
			Build 1 store	0.027	
			Build 2 stores	0.001	
			Build 3 stores	0.000	
Observations	9,578	58,758		15,802	52,102

Note: An observation is a firm-market-year. The table presents the proportion of observations for each particular action. Chains are assumed to be long-lived and are potential entrants for all markets where they have not yet entered. The number of single-store potential entrants (for grocery and convenience stores) is set to the total number of unique stores which have operated at any point in the market over the period 2008-2019.

3 Industry Model

In this section, we describe a structural model of the entry and exit game played by rival retailers over time. This model serves two purposes. First, when we take this model to the data, we can recover parameters informative on why the dollar store format has been so successful at expanding and why dollar stores cause nearby grocery and convenience stores to exit. Second, as we expect the effects of dollar store entry to be non-uniform across locations in a market, we incorporate equilibrium effects (i.e., the reallocation of rival retailers away from dollar store entry locations) into the model to evaluate the *net* impact of this expansion via counterfactual simulations.

We start by presenting the basic structure of the model, as well as the equilibrium concept. In the next section, we describe the empirical implementation and results.

Players. Each market contains two types of potential entrants: multi-store firms (i.e., chains) and a set of independent single-store firms. Markets are assumed to be independent of each other.[13] 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 $t = 1, \ldots, \infty$.

In what follows, we consider a market m that is partitioned 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 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 $\mathbf{n}_{mt} = (\mathbf{n}_{it})_{i \in I}$. 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 the number of other store types (e.g., drug stores, supermarkets, and gas stations) which are not players in the game but can be payoff-relevant.[14] 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: $f(\mathbf{x}_{m,t+1} | \mathbf{x}_{mt})$ and $h_t(d_{im,t+1} | d_{imt})$. The latter transition matrix is deterministic.[15]

Every period, the vector of public information variables includes the spatial market structure $\mathbf{n}_{mt}$ and market and location level characteristics. All these variables are publicly ob-

[13] Actions by players in one market do not impact the equilibrium played in another market.

[14] This helps capture, for instance, dollar store chains' tendency to locate away or near big-box stores such as Walmart.

[15] We assume that players have perfect foresight over the evolution of distribution networks during the sample period. While endogenizing distribution centers' openings would be an interesting addition to the model, this is complicated in practice because of the small number of distribution centers which prevents the precise estimation of these choice probabilities.

served 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

$$\mathcal{M}_{j,i,t} = (\mathbf{n}_{it}, \mathbf{n}_{-it}, \mathbf{x}_{mt}, \mathbf{d}_{mt}) \tag{1}$$

Actions.

Multi-store firms We assume that a chain may open or close at most one store per period.[16] 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. 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 (if, for some l, $0 < n_{ilt} < \overline{n}$). Multi-store firms are *long-lived*, that is, they can delay entry into the market. This feature allows us to capture delays in entry because a chain expects to open a distribution center closer to the market in a future period. 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.[17]

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

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

[16] There is a small number of observations (chain-market-period) where a chain closes two stores simultaneously or opens two stores or more simultaneously (see Table 4) We exclude these observations in the estimation procedure described in Section 4.2.2.

[17] Exit by chains from a market is observed to be permanent in the sample and time period we consider. This assumption implies that in the long-run, the probability that "all chains have existed from the market" converges to one. In practice, given annual exit rates of the order of 1-2% for chains, this would not be expected to occur for many decades.

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. Firm i's current profits, net of private information shocks, are

$$\pi_{it}(a_{it}, \mathcal{M}_{j,i,t}) = VP_i(\mathcal{M}_{j,i,t}) - FC_{it}(\mathcal{M}_{j,i,t}) - EC_{it}(a_{it}) + EV_{it}(a_{it}), \tag{2}$$

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, that is

$$VP_i(\mathcal{M}_{j,i,t}) = \sum_{l=1}^{L} n_{ilt} vp_{i,l}(\mathcal{M}_{j,i,t}) \tag{3}$$

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, $b = 1, \ldots, B$, around location l

$$vp_{i,l}(\mathcal{M}_{j,i,t}) = \sum_{b=1}^{B} \alpha_i^b x_{mlt}^b + \sum_{b=1}^{B} \beta_{io}^b n_{ilt}^b + \sum_{b=1}^{B} \sum_{f=1}^{F} \beta_{if}^b n_{flt}^b \tag{4}$$

where f denotes the type of competitors (i.e., dollar store, grocery 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 rival stores.

Importantly, for each store type f, profits depend on population at various distance bands around the store. This specification allows us to account for the fact that retail formats may differ in their trade areas: i.e., grocery store may draw consumers from a wider radius than convenience stores.

For chains, fixed operating costs depend on the distance to the closest distribution center and capital costs (proxied by residential rents). If a chain is operating at least one store in the market, fixed costs are

$$FC_{it}(\mathcal{M}_{j,i,t}) = \theta_{1,c}^{FC} d_{imt} + \sum_{l=1}^{L} \theta_{2,c}^{FC} rent_{mlt} \tag{5}$$

For a single-store firm operating in location l, fixed costs depend only on capital costs: $FC_{it} = \theta_s^{FC} rent_{mlt}$.

The specification of entry costs is

$$EC_{it} = \sum_{l=1}^{L} 1\{a_{it} = l_+\}\theta_i^{EC}. \tag{6}$$

In estimation, we will restrict entry costs to depend only on the type of the firm f but not the identity of firm i.

The exit value is specified as:

$$EV_{it} = \sum_{l=1}^{L} 1\{a_{it} = l_-\}\theta_i^{EV}. \tag{7}$$

Similarly to entry costs, the exit value θ_f^{EV} 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}.

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

$$\Pi_{it}(a_{it}, \mathcal{M}_{j,i,t}, \boldsymbol{\epsilon}_{it}) = \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}). \tag{8}$$

Value function and Equilibrium concept. We focus on Markov-Perfect Bayesian Nash Equilibria (MPE). Other equilibrium concepts can be considered in this setting: e.g., the partially and/or nonstationary oblivious equilibrium of Weintraub et al. (2008) and Benkard et al. (2015). However, for the small and medium-sized isolated markets we consider—where the average number of commercial locations is four—retailers have arguably accurate information on store counts and demographics in each location. Therefore, for this type of markets, we believe the MPE concept better captures the information set of both single-store firms and chains.

We first define firm strategies, value functions, and then the equilibrium conditions.

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

$$V_{i,t}^{\alpha}(\mathcal{M}_{j,i,t}, \boldsymbol{\epsilon}_{it}) = \max_{a_{it} \in A(n_{it})} \left\{ v_{i,t}^{\alpha}(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \right\} \tag{9}$$

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

$$\begin{aligned} v_{i,t}^{\alpha}(a_{it}, \mathcal{M}_{j,i,t}) = {} & \pi_i(a_{it}, \mathcal{M}_{j,i,t}) \\ & + \beta \int V_{i,t+1}^{\alpha} \left(\mathcal{M}_{j,i,t+1}, \boldsymbol{\epsilon}_{i,t+1} \right) dG(\boldsymbol{\epsilon}_{i,t+1}) dF_t(\mathcal{M}_{j,i,t+1} | a_{it}, \mathcal{M}_{j,i,t}) \end{aligned} \tag{10}$$

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 $j \neq i$.

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

$$\alpha_{i,t}^{*}(\mathcal{M}_{j,i,t}, \epsilon_{it}) = \arg \max_{a_{it} \in A(n_{it})} \left\{ v_{i,t}^{\alpha^*}(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \right\} \tag{11}$$

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

$$P_t^{\alpha}(a_{it}|\mathcal{M}_{j,i,t}) = \Pr(\alpha_{i,t}(\mathcal{M}_{j,i,t}, \epsilon_{it}) = a_{it}|\mathcal{M}_{j,i,t}) \tag{12}$$

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

$$v_{i,t}^{\mathbf{P}}(a_{it}, \mathcal{M}_{j,i,t}) = \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \beta \sum_{a_{-it}} \int \overline{V}_{i,t+1}^{\mathbf{P}}(\mathcal{M}_{j,i,t+1})\, dF_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t}) \tag{13}$$

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

$$\overline{V}_{i,t}^{\mathbf{P}}(\mathcal{M}_{j,i,t}) = \int \max_{a_{it} \in A(n_{it})} \Bigg\{ \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) + \beta \sum_{a_{-it}} \int \overline{V}_{i,t+1}^{\mathbf{P}}(\mathcal{M}_{j,i,t+1})\, dF_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t}) \Bigg\} dG(\epsilon_{it}). \tag{14}$$

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

$$P_t(a_{it}|\mathcal{M}_{j,i,t}, \mathbf{P}) = \frac{\exp\left(\frac{v_{i,t}^{\mathbf{P}}(a_{it}, \mathcal{M}_{j,i,t})}{\theta^{\epsilon}}\right)}{\sum_{a' \in A(n_{it})} \exp\left(\frac{v_{i,t}^{\mathbf{P}}(a', \mathcal{M}_{j,i,t})}{\theta^{\epsilon}}\right)}. \tag{15}$$

Simultaneity in players' moves can cause multiplicity of equilibria in this context. The CCP-based estimation approaches we use circumvent this difficulty by relying on the best-response mapping as estimating equations (Pesendorfer and Schmidt-Dengler (2008), Kalouptsidi et al. (2020), Bugni and Bunting (2021)). For our counterfactual analysis, we initialize the computation algorithm at a large number of starting values and iterate to a fixed point. We found no evidence of multiple equilibria in the counterfactual exercise.

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 can be expressed directly in terms of the period-ahead probabilities of choosing the terminal choice. Moreover, it allows us to incorporate non-stationarities into the model without making out-of-sample assumptions about players' actions for periods beyond the sample horizon (which is the year 2019).

Lemma 1 and 2 in Arcidiacono and Miller (2011) show that the ex-ante value function can be written as the sum of the choice-specific value function evaluated at any arbitrary action ($\tilde{a}$) and a known function of the CCPs. In particular, if the $\epsilon_{it}(a_{it})$ are independent type 1 extreme value, then

$$\overline{V}^{\mathbf{P}}_{i,t+1}\left(\mathcal{M}_{j,i,t+1}\right) = v^{\mathbf{P}}_{i,t+1}(\tilde{a}, \mathcal{M}_{j,i,t+1}) + \gamma - \ln\left(P_{i,t+1}(\tilde{a}|\mathcal{M}_{j,i,t+1})\right) \tag{16}$$

where γ is the Euler constant. A natural choice for the action $\tilde{a}$ is exit. If this terminal choice is chosen, the choice-specific value function is known up to the structural parameters and given by (where e refers to exit)

$$v^{\mathbf{P}}_{i,t+1}(a'_i = e, \mathcal{M}_{j,i,t+1}) = \begin{cases} \pi_i(a'_i = e, \mathcal{M}_{j,i,t+1}) & \text{if } i \text{ is an incumbent in } l \\ 0 \quad \text{if } i \text{ is potential entrant} \end{cases} \tag{17}$$

This allows us to replace the ex-ante value function $\overline{V}^{\mathbf{P}}_{i,t+1}\left(\mathcal{M}_{j,i,t+1}\right)$ by known functions of the structural parameters and CCPs. Equation (16) is used extensively in the estimation approach presented below.

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 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 (13)) can be rewritten for all $a_{it} \neq e$

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

where Equation (16) is used with $\tilde{a}$ set to exit. The choice-specific value function corresponding to the exit decision $v^{\mathbf{P}}_{i,t+1}(a'_i = e, \mathcal{M}_{j,i,t+1})$ is given by Equation (17).

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 who remains active ($a_{it} = 0$) or a potential entrant i who enters into location l ($a_{it} = l_+$), the choice-specific value function is identical to Equation (18).

For a 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

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

The first equality uses the identity in Equation (16) with $\tilde{a}$ set arbitrarily to l_+, in period $t+1$. The second equality uses the identity in Equation (16) with $\tilde{a}$ set of l_- (exit), in period $t+2$.

4 Estimation of the Dynamic Game

4.1 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.[18]

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.[19] 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 char-

[18]Markets are independent, therefore, identification is based on a cross-section of market-paths assuming that markets with the same observable characteristics feature the same equilibrium.

[19]More recent contributions include Kalouptsidi et al. (2021b) and Kalouptsidi et al. (2021a).

acteristics (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 incumbency status of rival single-store firms.

4.2 Estimation Approach

4.2.1 Baseline approach

We follow a two-step approach. In a first step, consistent estimates of the CCPs are obtained. We discuss this first step and the treatment of unobserved heterogeneity in detail in Section 4.2.2. 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 include policy evaluation (Aguirregabiria and Mira (2007), Pesendorfer and Schmidt-Dengler (2008)) or forward simulation (Bajari et al. (2007)). These methods are not applicable when estimating non-stationary games without imposing strong assumptions about the CCPs outside of the sample period (after the year 2019 in our case).[20] In addition, firms' location choices and the presence of multi-store firms generate a high-dimensional state space. Implementing these approaches 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 (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 Sec-

[20]For example, forward simulation requires knowing the CCPs far into the future. These CCPs are not available if the game is non-stationary and the probabilities are indexed by time.

tion 3, this allows us to express period-t choice-specific value function as a function of known period-$t+1$ (and $t+2$) CCPs and structural profit function. This property, however, does not address the computational challenge that these period-ahead value functions need to be integrated over a high-dimensional state space. To address this second issue, we estimate the model using the linear IV strategy of Kalouptsidi et al. (2020). The latter paper combines insights from the finite dependence approach and the GMM-Euler approach of Aguirregabiria and Magesan (2018) to propose a method (ECCP) that, among other advantages, circumvents (numerical) integration over the high-dimensional state space. This is achieved by invoking rational expectations, and using observed rather than expected behavior in successive periods to derive linear estimating equations, making this approach computationally light.

We show that the ECCP 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 and can prove useful in other settings characterized by high-dimensional state spaces and finite-dependence. In dynamic games with one-period finite dependence, the ECCP estimator applies directly. For games with τ-period finite dependence, with τ greater than two, a dynamic selection problem arises: the distribution of competitors' states in $t+2$ depends on the action of the focal player at time t.[21] This is ruled out by Kalouptsidi et al. (2020), who focus on single-agent problems where the evolution of market-level states is exogenous. We show that this selection problem can be accounted for by appropriately reweighting the data using CCP ratios, and we apply it in our setting with two-period finite dependence due to the presence of long-lived chain entrants.

We discuss how moment restrictions are constructed for single-store and multi-store firms next.

Single-store firms. Differences in choice-specific value function for potential entrants and

[21] Arcidiacono and Miller (2019) develop an algorithm for verifying whether finite dependence holds in a large class of single-agent decision problems and dynamic games.

incumbents can be derived as follows. A potential entrant can either stay out ($a_{it} = 0$) or enter by building a store in any of the locations ($a_{it} = l_+$). The corresponding choice-specific value functions are given by

$$v_{i,t}^{\mathbf{P}}(a_{it} = 0, \mathcal{M}_{j,i,t}) = 0 \tag{20}$$

$$v_{i,t}^{\mathbf{P}}(a_{it} = l_+, \mathcal{M}_{j,i,t}) = -\theta_i^{EC} + \beta\, \mathsf{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{21}$$

where the expectation is over $\mathcal{M}_{j,i,t+1}$ conditional on $(a_{it} = l_+, \mathcal{M}_{j,i,t})$ and we use $t+1$ CCP (of exiting) to express the entrant's continuation value in period $t+1$. Combining these two equations gives

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = -\theta_i^{EC} + \beta\, \mathsf{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{22}$$

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

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) \tag{23}$$

Combining Equation (22) and Equation (23), we obtain an optimality condition that involves only CCPs at t and $t+1$ and the single-period payoff function

$$\ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) = -\theta_i^{EC} + \beta\, \mathsf{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{24}$$

This equation includes expected profits and CCPs at $t+1$, and therefore, it appears that numerical integration over the state space is required. However, the equation can be used to 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.[22] Therefore, for any function of period-t information set

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

$h(\mathcal{M}_{j,i,t})$, we have

$$\mathsf{E}\left[h(\mathcal{M}_{j,i,t})u_{it}\right] = 0 \tag{25}$$

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

$$\begin{aligned} u_{it}(\mathcal{M}^*_{j,i,t+1}) &= \beta\,\mathsf{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \\ &\quad - \beta\left(vp_{i,l}(\mathcal{M}^*_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}^*_{j,i,t+1})\right) \\ &= \ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) + \theta_i^{EC} \\ &\quad - \beta\left(vp_{i,l}(\mathcal{M}^*_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}^*_{j,i,t+1})\right) \end{aligned} \tag{26}$$

where the second equation is obtained by using the expression in Equation (24) to eliminate the expectation term. The moment conditions (Equation (25)) 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, under linearity of payoffs, 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.[23] Define the left-hand side variable for potential entrant and incumbent (respectively) as[24]

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

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

[23] The endogeneity problem occurs because $t+1$ covariates may correlate with forecast errors. One can use instruments in the contemporaneous information set at t for these next-period covariates.

[24] An incumbent single-store firm chooses between staying active ($a_{it} = 0$) or exiting ($a_{it} = l_-$). Differences in the choice-specific value function are given by

$$v^{\mathbf{P}}_{i,t}(0, \mathcal{M}_{j,i,t}) - v^{\mathbf{P}}_{i,t}(l_-, \mathcal{M}_{j,i,t}) = \beta\,\mathsf{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{27}$$

where period-t profits cancel out.

We can obtain the structural parameters via the regression model

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

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

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 A.1, that an appropriate reweighting using CCP ratios addresses the selection bias, extending the approach of Kalouptsidi et al. (2020) from single-agent to dynamic games.

To evaluate the robustness of our results, we also implement an alternative estimation approach, in Appendix A.2, that does not rely on finite dependence and instead solves directly for the ex-ante value function. We recover similar parameter estimates and payoff functions.

4.2.2 Location-level unobserved heterogeneity and first-step estimates

The presence of unobserved heterogeneity is a common concern in many empirical settings and can introduce an endogeneity problem 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 drugstores, supermarkets, and gas stations) *simultaneously* operating in a given location over the period 2008-2019.[25]

The importance of controlling for unobserved heterogeneity is illustrated in Table 5. This table shows estimates of the CCPs for dollar store chains via a flexible multinomial logit regression.[26] 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.[27] We control for location-level demographic variables, cost shifters (e.g., the distance between the market and the closest distribution 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 roll-out

[25] The use of a proxy variable shares similarities with the control function method (see Wooldridge (2015) for a recent overview). In our setting, the number of rival stores operating in location l at time t is correlated with the unobserved heterogeneity term, inducing an endogeneity problem. If there is a monotonic relationship between (persistent) location-level unobserved heterogeneity and the maximum number of stores operating in a location, then one can think of the latter variable as proxying for the unobserved heterogeneity factors (amenities such as available parking, access to busy thoroughfare), that are correlated with the number of rival stores operating in location l at time t.

[26] In an ideal world, CCP would be estimated nonparametrically. For instance, Kalouptsidi et al. (2020) assume that the CCP are identified from a large cross-section of agents in each market-period. This approach is not possible in our setting given the size of the state space, the large number of choices that each firm has, and the size of the observed sample. For example, a market with four locations, three potential grocery stores, three potential convenience store, and the three chains (which can open at most 2 stores per location) has a spatial market structure $\mathbf{n}_{mt}$ with more than 650 million possible states. Location-level demographic states in $\mathbf{x}_{mt}$ further increase the size of the state space.

[27] The small number of observations where a chain opens more than one store in a period are not included when calculating the likelihood.

of distribution centers, we also include year dummies.

The effect of competition on the likelihood of building a store is 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.[28] 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 and convenience stores) via flexible multinomial logit regressions, controlling for business density. We include the regression results in Appendix Table A2 for completeness.

5 Estimation Results

This section presents our estimates of the structural parameters entering single-period payoffs for dollar store chains, grocery stores, and convenience stores. Of particular interest are the magnitude of strategic interactions across store formats, the presence and size of economies of density or cannibalization, and the role of chains' expanding network of distribution centers in driving the reduction in store operating costs.

Table 6 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

[28]This upward bias persists even when year FE are included.

Table 5: 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 (0-2 mi)	-0.071 (0.153)	0.026 (0.066)	-0.103 (0.155)	-0.311 (0.066)
Number of rival grocery (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.

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.[29] Table 7 shows mean store

[29]Specifically, we convert the revenue data from TDLinx 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

Table 6: Estimates of store profits and costs

	Chains		Grocery 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 stores (0-2 mi)	-0.074	(0.022)	-0.030	(0.021)	-0.033	(0.011)
Number of rival grocery 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.

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

We find that, consistent with anecdotal reporting on dollar store growth, dollar store 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 stores in the 0-2mi radius.[30] Second, the presence of dollar stores also significantly

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

[30]The magnitude of the business stealing effects are consistent with anecdotal evidence from grocery store

Table 7: Mean store profits and marginal effects

	Chain	Grocery	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 store (0-2 mi)	-10.30	-7.05	-7.64
One additional rival grocery 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.

harms convenience store profits, by as much as an additional convenience store. Third, 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. The location of the market relative to a chain's closest distribution center is also an important determinant of profits: a one standard deviation increase (130mi) from the mean distance (190mi) raises operating costs and reduces store profits by 6.13%.

We also implement alternative specifications for chains' dynamic investment costs. In particular, chains may benefit from network economies at the regional level by operating multiple stores in neighboring markets, which can reduce distribution and restocking costs (Jia (2008), Holmes (2011)). To capture these economies of density at the regional level,

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)

we allow the entry cost of the first store in the "region" (defined as a 100mi radius around the focal market) to differ from the entry cost of subsequent stores. The results are shown in Table A3 in Appendix B, which compares our baseline specification to the alternative described above. We find that the entry cost of the first store in the region is approximately twice as large as the entry costs of subsequent stores (4.68 versus 2.31). These estimates suggest that entry costs are drastically reduced when stores are opened near existing regional distribution networks, consistent with the previous literature studying Walmart's expansion strategy.[31] Finally, Appendix A.3 presents robustness checks with respect to the discount factor and assumptions about the number of potential entrants.

6 Counterfactual Analysis

This section uses the structural estimates for counterfactual simulations. We evaluate the impact of dollar store expansion on spatial market structure by comparing retailers' entry and exit choices with and without dollar store entry. For each market, we simulate the counterfactual MPE where dollar store chains do not enter starting in 2010.[32] The counterfactual CCPs are used to simulate each market from 2010 to 2019.

This counterfactual exercise quantifies changes in the spatial distribution of retail stores due to dollar stores' expansion and the within-market reallocation of retail activity. Accounting for these indirect effects is crucial to correctly assess the net market impact. We also quantify changes in policy-relevant outcomes, such as average household travel costs to different types of retailers. The method for solving the counterfactual MPE is detailed in Appendix A.4.[33]

[31] Density economies may affect profits through both entry and fixed costs. Because the two effects cannot be separately identified, we capture density economies via entry costs.

[32] We evaluate a counterfactual where dollar store expansion is halted in 2010, rather than the more extreme counterfactual where dollar stores are removed from the market entirely, to keep the results more in-sample. This counterfactual is also more closely aligned with policy proposals that restrict dollar store entry rather than calling for existing dollar stores to exit.

[33] Due to computational limitations, we restrict the exercise to markets with up to four commercial locations, covering 532 out of 846 markets in our sample.

6.1 The Impact on Spatial Market Structure

We define the following outcome variables that will be used throughout this section. Recall that the spatial market structure in period t and market m is represented by the vector $\mathbf{n}_{mt}$, which encodes the number of stores operated by each firm in each location. Let $\mathsf{E}[n^g_{lT}|\mathbf{n}_{m0}]$ and $\mathsf{E}[n^c_{lT}|\mathbf{n}_{m0}]$ denote the expected number of grocery (g) and convenience (c) stores in location l of market m in the last period T (2019), conditional on the observed market structure in the first period (2010), under the equilibrium played in the data. Similarly, let $\mathsf{E}[\tilde{n}^g_{lT}|\mathbf{n}_{m0}]$ and $\mathsf{E}[\tilde{n}^c_{lT}|\mathbf{n}_{m0}]$ denote these expectations under the counterfactual equilibrium with no dollar store entry. To obtain the factual and counterfactual expected market structures, we take the expectation over all possible realizations of spatial market structure given equilibrium factual and counterfactual CCPs.

The impact of dollar store entry on the number of stores of type $j \in \{g, c\}$ operating in location l is computed as

$$\Delta n^j_l = \mathsf{E}[\tilde{n}^j_{lT}|\mathbf{n}_{m0}] - \mathsf{E}[n^j_{lT}|\mathbf{n}_{m0}]. \tag{28}$$

The impact of dollar store entry on the number of stores of type $j \in \{g, c\}$ operating in market m (with L locations) is obtained by summing over locations,

$$\Delta n^j_m = \sum_{l=1}^{L} \Delta n^j_l. \tag{29}$$

Market-level analysis. First, we consider the market-level impact of dollar store chains. Figure 2 shows histograms of the change in grocery and convenience stores (Δn^g_m and Δn^c_m) in levels and percentages. Due to dollar store chains' expansion post-2010, markets have 31% to 33% fewer grocery and convenience stores on average. There is significant heterogeneity, with many markets experiencing over a 50% reduction in stores. The impact on convenience stores is larger than on grocery stores, partly due to the higher baseline number of convenience stores.[34]

[34] We note the presence of five outlier markets where Δn^g_m is positive but very small (on average equal to 0.02). Some of these outlier markets experienced a decrease in the number of dollar stores in the factual scenario, whereas in the counterfactual, this number is kept constant at its 2010 level.

Figure 3 examines how these changes vary with market characteristics. We regress Δn_m^g and Δn_m^c on market-level and geographic controls: market population, income per capita, a dummy for low racial minority share (below 25%), a dummy for high vehicle access (above 89%, the first quartile), and the number of commercial locations. Continuous variables (population and income) are normalized.

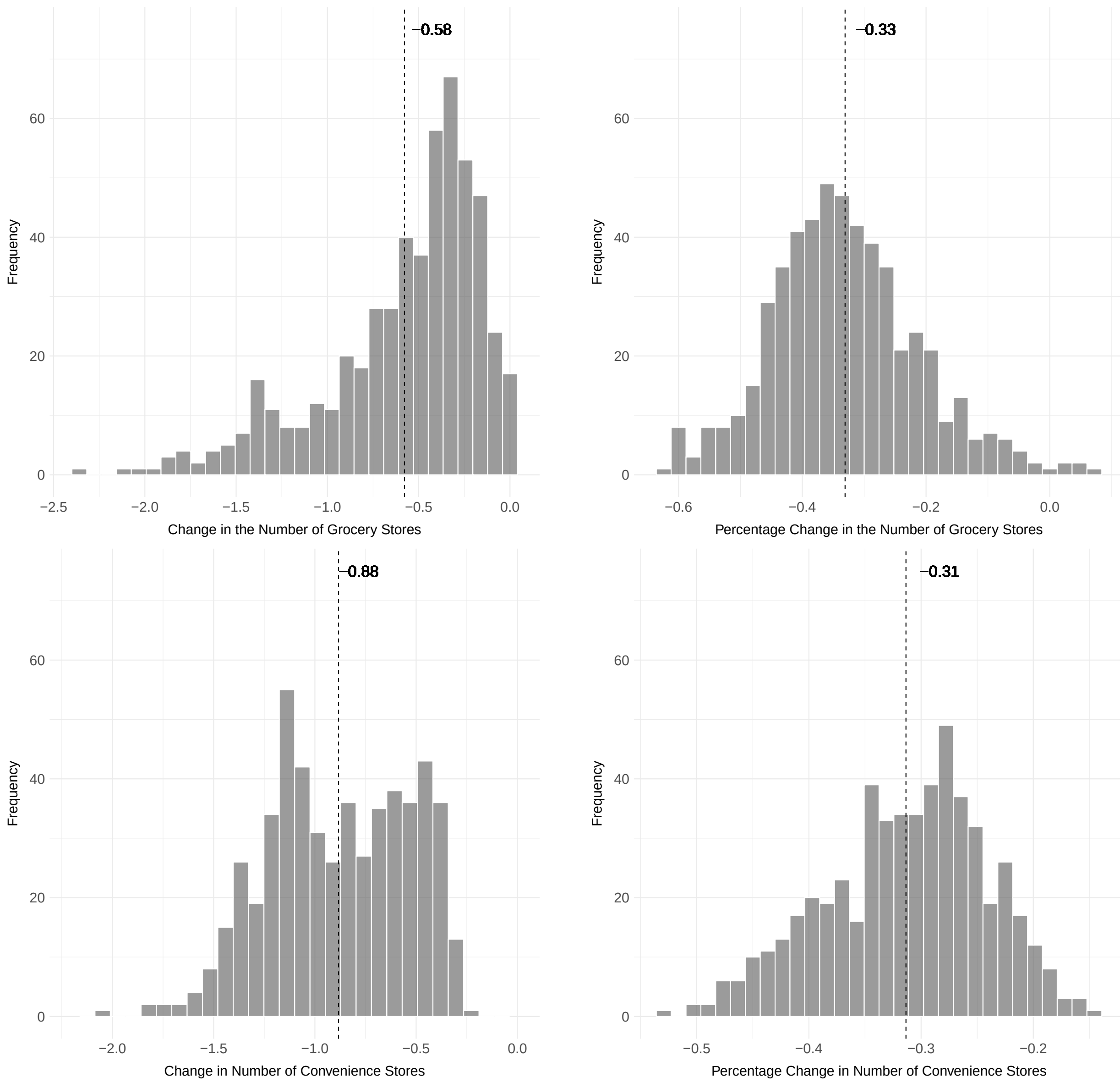
The largest effect on store numbers occurs in markets with larger populations, lower incomes, and higher minority shares. A one standard deviation increase in population (3,742 from a mean of 9,142) is associated with a 17% decrease in Δn_m^g.

Location-level analysis. Next, we direct focus on the impact of dollar stores' expansion at the location (Census tract) level within a given market. Figure 4 shows histograms of the change in the number of grocery and convenience stores at the location level (Δn_l^g and Δn_l^c). The average reduction in the number of stores per location is smaller than the average reduction per market. Some locations start with a small number of stores and see little change due to dollar store entry.

We examine how these changes correlate with location-level demographics. As in the market-level analysis, we regress Δn_l^g and Δn_l^c on the same controls as in Figure 3, but now defined at the location level, and include market fixed effects. The results, shown in Figure 5, indicate that the largest reduction in the number of stores occurs in locations with higher population, lower income, and lower share of households with vehicle access (i.e., higher transportation costs). Locations with a higher share of minority groups in a given market experience a larger reduction in the number of grocery stores. In Appendix Table A4, we provide additional details on average effects broken down by various location characteristics: e.g., the number of grocery stores decreases by 19% in locations with low vehicle access versus 6% for location with high vehicle access.

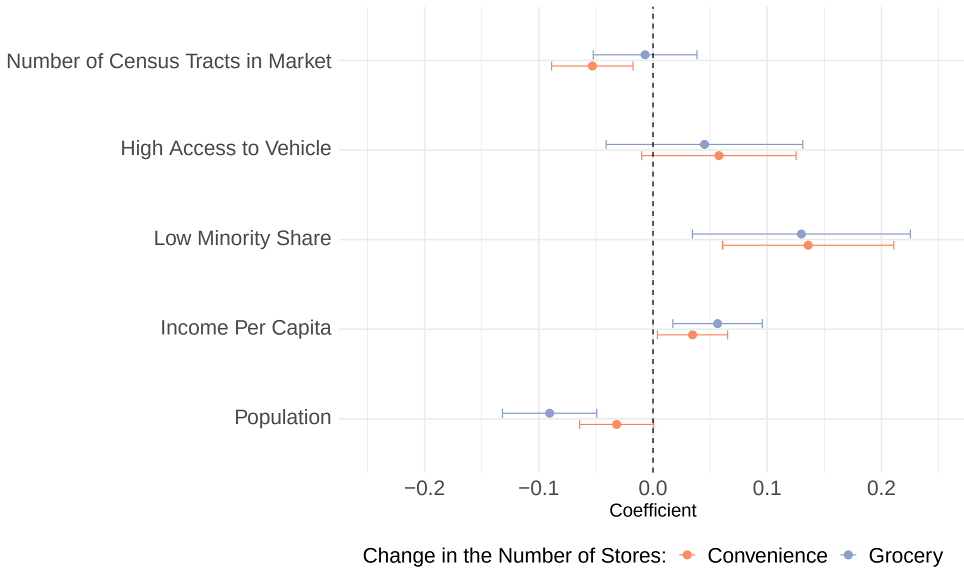
Reallocation. Finally, to study the potential for spatial spillovers, we decompose the market-level impact of dollar store expansion into direct and indirect effects. The direct effect refers to dollar store-induced exits of incumbent grocery and convenience stores. The

Figure 2: Change in the Expected Number of Stores per Market



Notes: These figures show histograms of the change in the expected number of grocery and convenience stores in level (Δn_m^g and Δn_m^g) (left) and in percentage terms (right). Average values are shown in each plot using dashed vertical lines.

Figure 3: Correlation between the Change in Stores and Market Characteristics



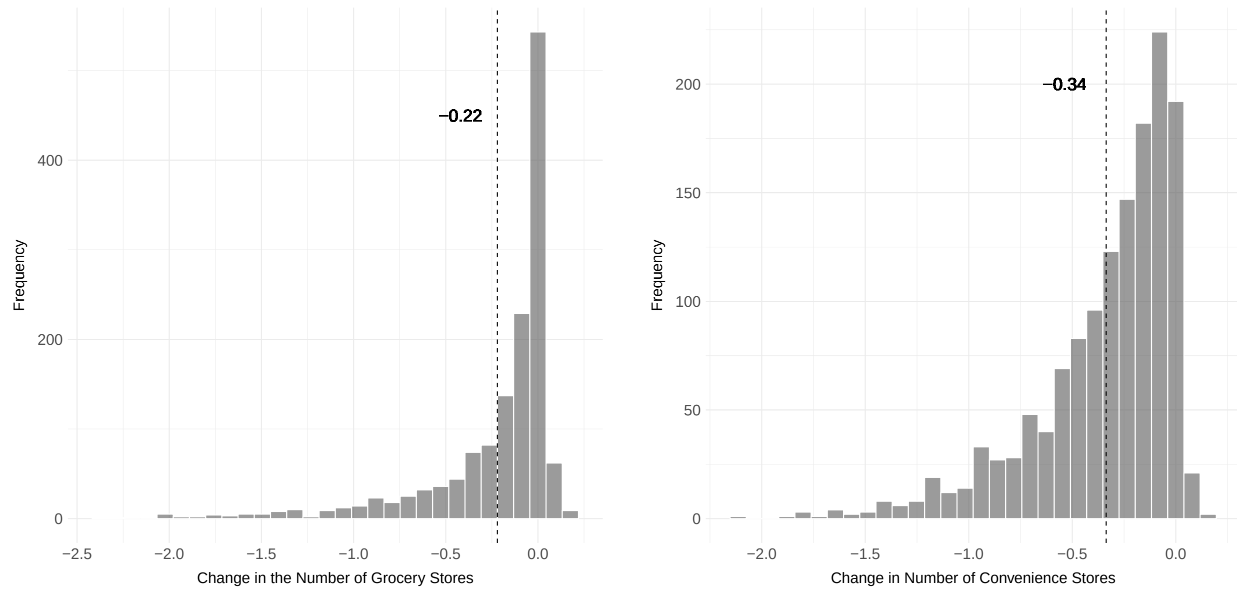
indirect effect involves the additional entry generated by reallocation of grocery and convenience stores to locations further from dollar stores. Indirect effects occur if stores can profitably differentiate spatially in response to dollar store entry, which changes the relative attractiveness of locations in the market. We emphasize that "reallocation" here means new stores entering rather than existing ones relocating. Grocery and convenience stores differ in the extent to which they are harmed by other retail formats, affecting their ability to spatially differentiate.

To operationalize this decomposition, we write the market-level change in the number of each store type $j \in \{g, c\}$ as the sum of location-level changes between the counterfactual and factual scenarios, distinguishing between locations that see an increase or decrease in the number of stores, as follows

$$\begin{aligned} \Delta n_m^j &= \sum_{l=1}^{L} \Delta n_l^j \\ &= \underbrace{\sum_{l:\Delta n_l^j < 0} \Delta n_l^j}_{\text{direct effects}} + \underbrace{\sum_{l:\Delta n_l^j \geq 0} \Delta n_l^j}_{\text{indirect effects}} \end{aligned} \tag{30}$$

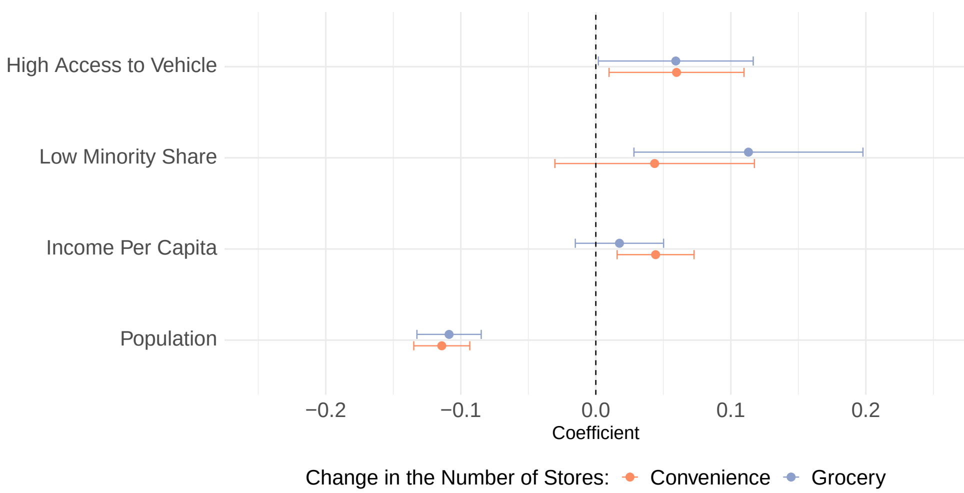
For spillovers or indirect effects to be present, there needs to be more than one location in

Figure 4: Change in the Expected Number of Stores per Location



Notes: These figures show histograms of the change in the expected number of grocery and convenience stores per location (Δn_l^g and Δn_l^g) in levels. Average values are shown in each plot using dashed vertical lines.

Figure 5: Correlation between the Change in Stores and Location Characteristics



the market. We restrict the analysis below to markets with two to four commercial locations (469 markets out of a total of 846 markets in our sample).

The prevalence of indirect effects differs for grocery and convenience stores. For grocery stores, indirect effects occur in 54% of the markets: i.e., for roughly half of markets, the number of grocery stores increases in some location(s) within the market. For convenience stores, only 20% of markets have indirect effects, with most markets experiencing a decrease in convenience stores in all locations.

What types of markets are most likely to have indirect effects? Which locations are likely to see an increase in grocery or convenience stores in response to dollar store entry? To answer these questions, we examine predictors of indirect effects. For the first question, we regress a dummy for whether $\sum_{l:\Delta n_l^j \geq 0} \Delta n_l^j > 0$ on market characteristics, as in Figure 3. For the second question, we regress a dummy for whether a location experiences an increase in grocery or convenience store counts on location characteristics, controlling for market fixed effects, as in Figure 5.

Figure 6 presents the results. The top panel shows that market population and the number of locations are the main predictors of indirect effects. Reallocation is less likely in densely populated markets but more likely in markets with more entry locations. Large population markets attract dollar stores across multiple locations, while markets with many locations offer more scope for spatial differentiation. The bottom panel shows that grocery stores (and to a lesser extent convenience stores) are more likely to relocate to locations with lower population and higher income. This suggests grocery stores shift to locations that are less attractive to dollar stores.

To assess the magnitude of within-market reallocation, we compute the ratio of indirect to direct effects (in absolute value) for each market m

$$\left| \frac{\sum_{l:\Delta n_l^j \geq 0} \Delta n_l^j}{\sum_{l:\Delta n_l^j < 0} \Delta n_l^j} \right| .$$

For grocery stores, the ratio of indirect to direct effects averages 14%, meaning the impact of dollar stores on grocery stores at the market level is 14% lower when reallocation is taken into account. For convenience stores, this ratio is only 3%. These differences stem from varying payoffs between store types (Table 6). Grocery stores generate more profit per unit of population than convenience stores, aligning with higher household spending at grocery stores. Thus, grocery stores can profitably enter locations with lower population density. By contrast, convenience stores favor locations that are similar to dollar stores' most preferred locations, and therefore their ability to spatially differentiate is more limited.

The magnitude of reallocation is arguably small but is based on markets with two to four locations. These figures likely underestimate reallocation on a national scale, where larger markets with more locations offer greater scope for spatial differentiation.

In the next subsection, we show how predicted changes in market structure can be used to compute consumer-level outcomes such as retail proximity and distance to the nearest store for different store types.

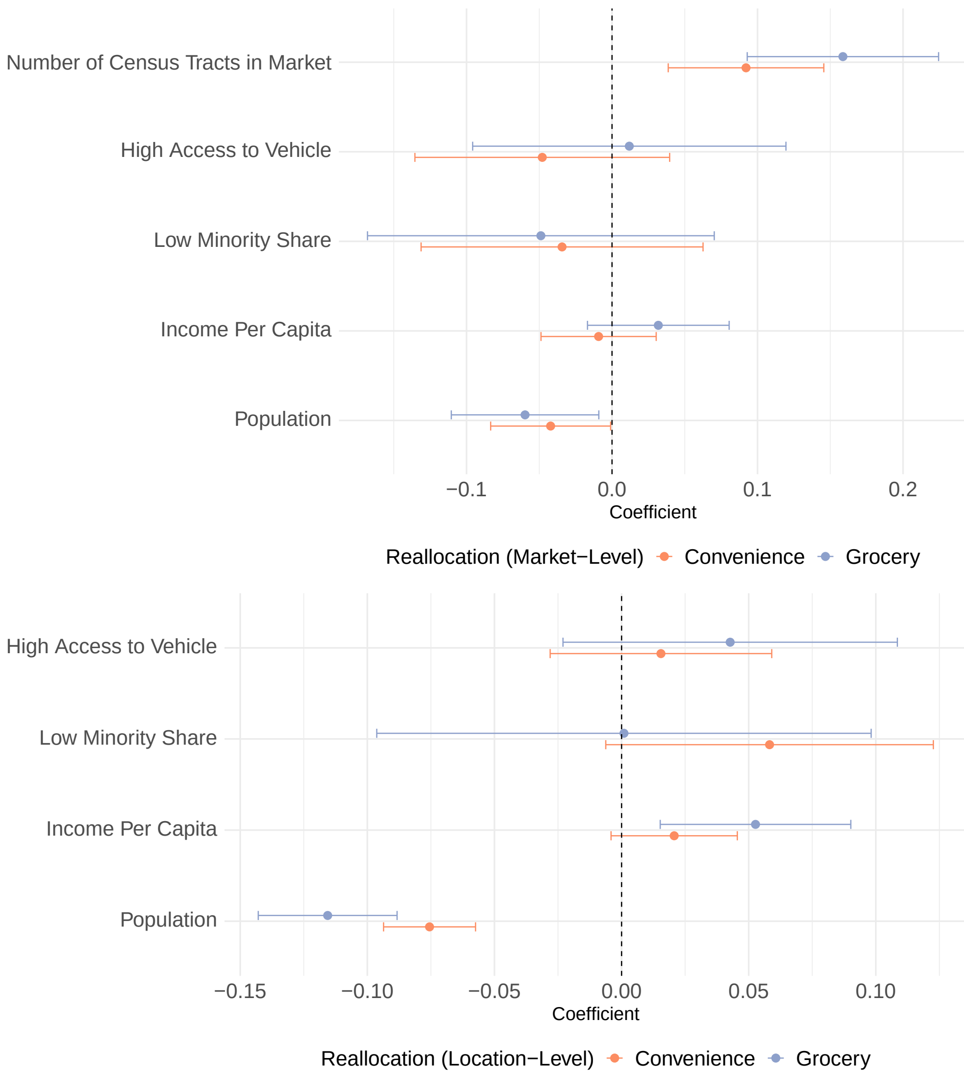
6.2 The Impact on Retail Proximity and Travel Costs

We consider how changes in market structure translate into changes in retail proximity for households in 2019. All measures of retail proximity are constructed from data on population at the census block group, i.e., one level down from our definition of locations (census tract). Mean and median statistics are obtained by taking the (population-weighted) mean and median over all census block groups in a market.[35] Table 8 shows the average distance between consumers and the nearest store, for the different retail formats, under the factual and counterfactual scenarios.

In panel A, we condition on the subset of markets in which there is at least one store format operating in both scenarios. The column $Pr(n > 0)$ gives the population-weighted probability that a market has at least one store of that format operating. The results suggest that the expansion of dollar store chains led to a reduction in "access to grocery stores," in that 85% of markets contain a grocery store in the counterfactual, whereas in the factual only 70% do.

[35] To compute distance to the nearest store, we assume that stores are located at the population-weighted centroid of their census tract. This assumption is required to keep the structural model tractable. We can, however, compare distances to the nearest store if the actual location (latitude and longitude) of each store in the factual scenario is used. Using population-weighted centroids instead of the actual store location does not quantitatively affect the results for the factual scenario.

Figure 6: Reallocation Effects



In panels B and C, we use all markets, including those where no store is present in one of the scenarios. Because proximity is inherently not well-defined in these cases, we compute the upper (and lower) bounds on the distance to the nearest store by assuming a distance of 5 (and 7) miles if there are no stores of a particular format operating in the market. We find that the distance to the nearest grocery store decreases by about 30% under the counterfactual scenario without dollar store expansion: from 2.30 to 1.68 miles (lower bound), or 2.91 to 2.00mi (upper bound). Given that prior research suggests consumers do not travel more than 1 or 2 miles for grocery trips, a change in average distance of .9 miles amounts to a significant increase in travel costs to the nearest grocery store. By contrast, in the counterfactual without dollar store expansion, the distance to the nearest dollar store only increases by .18 miles suggesting that, due to their initial market penetration in 2010, relatively fewer consumers experience significant changes in access to dollar stores by 2019.

Table 8: Predicted retail proximity: distance to nearest store

	Factual			Counterfactual		
	Mean	Median	Pr($n>0$)	Mean	Median	Pr($n>0$)
Panel A. Conditional on ($n>0$)						
Distance to nearest (in miles)						
Grocery store	1.14	0.92	0.70	1.09	0.86	0.85
Convenience store	0.99	0.79	0.89	0.90	0.72	0.97
Dollar store	0.96	0.69	0.94	0.99	0.75	0.82
Any store format	0.80	0.62		0.81	0.62	
Panel B. All Markets (lower bound)						
Distance to nearest (in miles)						
Grocery store	2.30	2.35		1.68	1.56	
Convenience store	1.41	1.12		1.02	0.83	
Dollar store	1.21	0.74		1.73	0.92	
Any store format	0.88	0.67		0.84	0.66	
Panel C. All Markets (upper bound)						
Distance to nearest (in miles)						
Grocery store	2.91	3.02		2.00	1.81	
Convenience store	1.64	1.22		1.10	0.85	
Dollar store	1.34	0.74		2.10	0.92	
Any store format	0.92	0.67		0.86	0.66	

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. Unconditional panels assume that if there are no stores (of a given retail format) operating in the market, the distance travelled is 5mi (lower bound) or 7mi (upper bound).

7 Conclusion

Dollar store chains have significantly transformed the US retail landscape. This expansion has sparked policy debates, with supporters highlighting increased consumer convenience and critics warning of negative impacts on local businesses and food access. Despite these debates, empirical evidence evaluating these arguments is only beginning to emerge. This article quantifies the impact of dollar stores' expansion on spatial market structure, specifically the geographic layout and number of retail stores by format. This outcome is particularly important as it affects retail proximity and convenience.

We specify a dynamic model of spatial competition between different store formats. The model explicitly incorporates equilibrium effects (i.e., the spatial reallocation of rival store activity in response to dollar store entry) to assess the net impact of this expansion. The spatial nature of competition introduces non-trivial complexities when estimating and solving this game. Methodologically, we deal with the high-dimensionality of the firms' problem by extending the ECCP estimator of Kalouptsidi et al. (2020) from single-agent problems to games with finite dependence.

The model provides structural estimates that are useful in understanding how the dollar store format grew so rapidly. Estimates indicate that dollar store chains benefit from lower entry costs and from operating at higher store density. Their increasingly wide network of distribution centers allows them to reduce store-level fixed costs over time and support their expanding retail footprint. Grocery and convenience stores, on the other hand, are harmed by the presence of dollar stores at close proximity.

The counterfactual exercise establishes that dollar store chains' expansion led to a reduction in the number of grocery and convenience stores operating in our sample by 31% to 33%. The largest impact occurs in larger, lower-income markets with significant minority populations. Within these markets, store count reductions are most notable in densely populated, lower-income areas with limited vehicle access (i.e., a proxy for transportation costs). We break down the aggregate impact at the market level into direct and indirect effects. In particular, the grocery format responds to dollar store entry by spatially differentiating into locations with lower population density and higher income. Accounting for such reallocation of rival stores is important to accurately measure the net impact of dollar store expansion. These changes in market structure are associated with a significant population-weighted increase in retail distance to grocery stores.

This article centers around spatial competition between dollar stores and their local competitors. The welfare implications of dollar store expansion are arguably multifaceted. Dollar store entry may affect consumer welfare through price changes, store convenience, product availability, and ultimately, by altering the composition of consumers' shopping baskets (Caoui et al. (2024)). In the medium to long run, these changes can significantly impact consumers' dietary choices and health outcomes. Although limitations in data and complexity prevent an estimation approach that accounts for all these dimensions in a single model, this article aims to improve our understanding of dollar stores' impact and informs the policy debate around this retail format.

References

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.

Aguirregabiria, Victor, and Gustavo Vicentini. 2016. "Dynamic Spatial Competition Between Multi-Store Retailers." *THE JOURNAL OF INDUSTRIAL ECONOMICS* LXIV (4): 710–754.

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 Paul B. Ellickson. 2011. "Practical Methods for Estimation of Dynamic Discrete Choice Models." *Annual Review of Economics* 3 (Volume 3, 2011): 363–394.

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.

Benkard, Lanier, Przemyslaw Jeziorski, and Gabriel Weintraub. 2015. "Oblivious equilibrium for concentrated industries." *RAND Journal of Economics* 46 (4): 671–708.

Beresteanu, Arie, Paul B. Ellickson, Sanjog Misra, and James C. Reeder. 2019. "The dynamics of retail oligopoly." 1–27.

Bodéré, Pierre. 2023. "Dynamic Spatial Competition in Early Education: an Equilibrium Analysis of the Preschool Market in Pennsylvania."

Bugni, Federico A, and Jackson Bunting. 2021. "On the Iterated Estimation of Dynamic Discrete Choice Games." *The Review of Economic Studies* 88 (3): 1031–1073.

Byrne, Anne, Xiao Dong, Erik James, Jessie Handbury, and Katherine Meckel. 2022. "Welfare Implications of Increased Retailer Participation in SNAP." https://www.ers.usda.gov/amber-waves/2019/august/.

Cao, Yue. 2022. "The Welfare Impact of Dollar Stores."

Caoui, El Hadi, Brett Hollenbeck, and Matthew Osborne. 2024. "The Impact of Dollar Store Expansion on Local Market Structure and Food Access." *SSRN Electronic Journal*.

Chenarides, Lauren, C Metin, and Timothy J Richards. Forthcoming. "Dynamic Model of Entry: Dollar Stores." *American Journal of Agricultural Economics*.

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

Datta, Sumon, and K. Sudhir. 2013. "Does reducing spatial differentiation increase product differentiation? Effects of zoning on retail entry and format variety." *Quantitative Marketing and Economics (QME)* 11 (1): 83–116.

Dollar Tree. 2018. "Dollar Tree to Consolidate Its Dollar Tree and Family Dollar Store Support Centers in Chesapeake, Virginia." https://www.dollartreeinfo.com/news-releases/news-release-details/dollar-tree-consolidate-its-dollar-tree-and-family-dollar-0.

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.

Fang, Limin, and Nathan Yang. 2022. "Measuring Deterrence Motives in Dynamic Oligopoly Games." *SSRN Electronic Journal* 1–47.

Feng, Wenhui, Elina T. Page, and Sean B. Cash. 2023. "Dollar Stores and Food Access for Rural Households in the United States, 2008–2020." *American Journal of Public Health* 113 (3): 331–336, PMID: 36657091.

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.

Hristakeva, Sylvia. 2022. "Determinants of Channel Profitability: Retailers' Control over Product Selections as Contracting Leverage." *Marketing Science* 41 (2): 315–335.

Igami, Mitsuru. 2011. "Does Big Drive Out Small?." *Review of Industrial Organization* 38 (1): 1–21.

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 Barwick, Panle, and Parag A. Pathak. 2015. "The costs of free entry: An empirical study of real estate agents in Greater Boston." *RAND Journal of Economics* 46 (1): 103–145.

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, Yuichi Kitamura, Eduardo Souza-Rodrigues, and Lucas Lima. 2021a. "Counterfactual Analysis for Structural Dynamic Discrete Choice Models." (June): .

Kalouptsidi, Myrto, Paul Scott, and Eduardo Souza-Rodrigues. 2021b. "Identification of Counterfactuals in Dynamic Discrete Choice Models." *Quantitative Economics.*

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

Lopez, Rigoberto, Keenan Marchesi, and Sandro Steinbach. 2023. "Dollar Store Expansion, Food Retail Competition, and Rural Employment."

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

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

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.

Schneier, Camilla, Yixin Sun, and Fern Ramoutar. 2023. "Buck Wild : The Impact of the Dollar Store on Households and Local Retail Competition."

Scott, Paul T. 2013. "Dynamic Discrete Choice Estimation of Agricultural Land Use."

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

Shannon, Jerry, Grace Bagwell-Adams, Sarah Shannon, Jung Sun Lee, and Yangjiaxin Wei. 2018. "The mobility of food retailers: How proximity to SNAP authorized food retailers changed in Atlanta during the Great Recession." *Social Science and Medicine* 209 (November 2017): 125–135.

Shih, Willy, Stephen Kaufman, and Rebecca Mckillican. 2019. "Dollar General (A)." *Harvard Business School* Case 607-1 1–28.

Sweeting, Andrew. 2013. "Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry." *Econometrica* 81 (5): 1763–1803.

Weintraub, Gabriel, C. Lanier Benkard, Przemyslaw Jeziorski, and Benjamin Van Roy. 2008. "Nonstationary Oblivious Equilibrium." *Manuscript, Columbia*, http://econ.jhu.edu/wp-content/uploads/pdf/papers/wp568.pdf.

Wooldridge, Jeffrey M. 2015. "Control Function Methods in Applied Econometrics." *The Journal of Human Resources* 50 (2): 420–445, http://www.jstor.org/stable/24735991.

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 Estimation Details

A.1 Estimation Approach for Multi-Store Firms

This section describes the baseline estimation approach for dollar 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 ($a_{it} = 0$) and entering into location l ($a_{it} = l_+$) are given, respectively, by

$$v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \beta\, \mathsf{E}\big(-\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1}) \\ + \beta^2\, \mathsf{E}[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})]|0, \mathcal{M}_{j,i,t}\big) \tag{31}$$

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) = -\theta_i^{EC} + \beta\, \mathsf{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})|l_+, \mathcal{M}_{j,i,t}] \tag{32}$$

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

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \log\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) \tag{33}$$

Combining Equation (31), Equation (32), and Equation (33), 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 potential entrants who stay out, there are two expectations (over $t+1$ and $t+2$ states), therefore, the expectational errors ($w_{it}, u_{i,t+1}$) are defined as

$$w_{it} = \mathsf{E}\left[-\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})|0, \mathcal{M}_{j,i,t}\right] \\ - \left(-\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}^*_{j,i,t+1})\right) \tag{34}$$

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

where the starred variables $\mathcal{M}^*_{j,i,t+1}$ and $\mathcal{M}^*_{j,i,t+2}$ are realized (observed) states in the data. For potential entrants who enter (and become incumbents in $t+1$), we define the expectational error v_{it} as

$$\begin{aligned} v_{it} = \mathsf{E}[&vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t}, l_+] \\ &- \big(vp_{i,l}(\mathcal{M}^*_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}^*_{j,i,t+1})\big) \end{aligned} \tag{36}$$

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

$$\mathsf{E}\left[g(\mathcal{M}_{j,i,t})'[w_{it}, v_{it}, u_{i,t+1}]\right] = \mathbf{0}_3 \tag{37}$$

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 $a_{it} = 0$. However, 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.[36][37]

To address this selection problem, we define the weights

$$\psi_{a_1,a_2}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t}) = \frac{P(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t}, a_{it} = a_1)}{P(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t}, a_{it} = a_2)} \tag{39}$$

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*

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

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

$$\begin{aligned} \widetilde{u}_{i,t+1} = \mathsf{E}&\left[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})|0, \mathcal{M}_{j,i,t}\right] \\ &- \psi_{0,\widetilde{a}}(\mathcal{M}^*_{j,i,t+2}|\mathcal{M}_{j,i,t})\left(vp_{i,l}(\mathcal{M}^*_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}^*_{j,i,t+2})\right) \end{aligned} \tag{40}$$

---

[36] 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 (a deterministic function of a_{it}).

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

$$\begin{aligned} E[u_{i,t+1}|\mathcal{M}_{j,i,t}] = \mathsf{E}&\left[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})|0, \mathcal{M}_{j,i,t}\right] \\ - \mathsf{E}&\left[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})|\tilde{a}, \mathcal{M}_{j,i,t}\right] \end{aligned} \tag{38}$$

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

Defining the left-hand side variables for entrant chains as

$$\begin{aligned} Y_{it}^{entrant} = \log\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) \\ - \beta[\gamma - \ln P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \\ + \beta[\gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})] \\ + \beta^2 \psi_{0,\widetilde{a}}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t})[\gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})] \end{aligned} \tag{41}$$

we obtain the structural parameters for entrants via the regression model

$$\begin{aligned} Y_{it}^{entrant} = \left[-\theta_i^{EC} + \beta(vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i)\right] \\ - \left[-\beta\theta_i^{EC} + \psi_{0,\widetilde{a}}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t})\beta^2(vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i)\right] \\ + (v_{it} - u_{it} - \beta\widetilde{u}_{i,t+1}) \end{aligned} \tag{42}$$

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

$$v_{i,t}^{\mathbf{P}}(l_-^*, \mathcal{M}_{j,i,t}) = vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \tag{43}$$

$$\begin{aligned} v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \\ + \beta\,\mathsf{E}[vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln P_{i,t+1}(l_-^*|\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t}, 0] \end{aligned} \tag{44}$$

$$\begin{aligned} v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) = vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i - \widetilde{\theta}_i^{EC} \\ + \beta\,\mathsf{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) + vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \\ + \gamma - \ln P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t}, l_+] \\ + \beta^2\,\mathsf{E}[vp_{i,l^*}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-^*|\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t}, l_+] \end{aligned} \tag{45}$$

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

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \log\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) \tag{46}$$

$$v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(l_-^*, \mathcal{M}_{j,i,t}) = \log\left(\frac{P_{i,t}(0|\mathcal{M}_{j,i,t})}{P_{i,t}(l_-^*|\mathcal{M}_{j,i,t})}\right) \tag{47}$$

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 (45), 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.

A.2 Alternative Estimation Approach

The baseline estimation approach extends the ECCP estimator of Kalouptsidi et al. (2020) from single-agent problems to dynamic games with finite dependence. The estimator leverages the finite-dependence property to express ex-ante value functions as a function of (observed) CCPs, drastically reducing computational costs. A potential caveat of this estimator is that it places a sizeable burden on particular reference probabilities (e.g., the probability of exit). If these probabilities are not precisely estimated, this can introduce bias in the estimates of dynamic investment costs (e.g., entry costs). To alleviate these concerns, we compare our baseline estimation results to ones obtained using an alternative approach that does not rely on finite dependence but instead solves directly for the ex-ante value function.

As the state space is exceptionally large and some state variables are continuous, it is impossible to solve for value functions at all states. Therefore, we approximate the ex-ante value function by a linear parametric function of K variables at a pre-specified number of states N (including all states observed in the data). Value function approximation has been implemented in other studies such as Sweeting (2013), Aguirregabiria and Vicentini (2016), Jia Barwick and Pathak (2015), and Beresteanu et al. (2019). Following the notation in Sweeting (2013), we express the ex-ante value function as

$$\overline{V}^{\mathbf{P}}_{i,t}(\mathcal{M}_{j,i,t}) \approx \sum_{k=1}^{K} \lambda_k \phi_{ki}(\mathcal{M}_{j,i,t}) \tag{48}$$

In practice, approximating functions $\phi_{ki}(\mathcal{M}_{j,i,t})$ include all exogenous variables and number of rival and own stores by distance bands and locations.[38] We allow the coefficients λ_k to differ by firm type (e.g., convenience, grocery, dollar store) and by incumbency status.[39]

Given a vector of CCPs and structural parameters in iteration (k), denoted $(\mathbf{P}^{(k)}, \boldsymbol{\theta}^{(k)})$, we iterate over the following steps.

[38] For incumbents, we sum these approximating functions over all stores the firm is currently operating. For a single-store entrant, the continuation value of staying out is zero (terminal action); whereas the continuation value of entering is equal to the ex-ante value function of being an incumbent in period $t+1$ (net of entry costs). By contrast, for a chain entrant, staying out is not a terminal action, therefore we also approximate the ex-ante value function of being an entrant, as the sum of exogenous and endogenous variables by distance bands over *all locations* in the market. We experimented with a saturated model including interactions, and/or market fixed effects, without significant improvements in fit. This is likely due to the fact that the static payoff function is linear.

[39] Because there is a time-to-build (and to exit) of one period, current-period profits for incumbents only depend on the stores operated at the beginning of the period but not on the chosen action (remain active, exit, build an additional store) and cancel out when taking differences in choice-specific value function. To be able to estimate profits, the ex-ante value function is, in practice, approximated as

$$\overline{V}^{\mathbf{P}}_{i,t}(\mathcal{M}_{j,i,t}) \approx VP_i(\mathcal{M}_{j,i,t}) + \sum_{k=1}^{K} \lambda_k \phi_{ki}(\mathcal{M}_{j,i,t})$$

To keep the exposition concise, we ignore this first term in the derivation that follows.

1. Solve for the ex-ante value function for each player i, denoted $\mathbf{V}_i$. This is a vector stacking value functions at the N pre-specified states. Under our approximation, the value function can be expressed as $\mathbf{V}_i = \boldsymbol{\lambda}'\boldsymbol{\phi}_i$, in matrix form. In equilibrium, the ex-ante value function must satisfy the following identity,

$$\mathbf{V}_i = \boldsymbol{\lambda}'\boldsymbol{\phi}_i = \sum_a \mathbf{P}_i^{(k)}(a)\left[\boldsymbol{\pi}_i^{(k)}(a) + \mathsf{E}[\epsilon|a]\right] + \beta\boldsymbol{\lambda}'\,\mathsf{E}\left[\boldsymbol{\phi}_i\right] \tag{49}$$

where $\boldsymbol{\pi}_i^{(k)}(a)$ is the vector of current-period profits given structural parameters $\boldsymbol{\theta}^{(k)}$, $E[\epsilon|a]$ is the expected firm-specific shock given action a is chosen, and $\mathsf{E}\left[\boldsymbol{\phi}_i\right]$ is the expected future value of the approximating functions with component $\mathsf{E}\left[\phi_{ki}(\mathcal{M}_{j,i,t+1})|\mathcal{M}_{j,i,t+1}\right]$. Rewriting this identity as

$$\boldsymbol{\lambda}'\left(\boldsymbol{\phi}_i - \beta\,\mathsf{E}\left[\boldsymbol{\phi}_i\right]\right) = \sum_a \mathbf{P}_i^{(k)}(a)\left[\boldsymbol{\pi}_i^{(k)}(a) + \mathsf{E}[\epsilon|a]\right] \tag{50}$$

$\boldsymbol{\lambda}$ can be found by an OLS regression. To calculate $\mathsf{E}\left[\boldsymbol{\phi}_i\right]$, we fit an AR-1 process for each exogenous state variables (location-level population, income, and rents, in logarithm), allowing for innovation shocks that are correlated across locations in a market. For the endogenous states (store counts), we simulate $1,000$ realizations of next-period spatial market structure by drawing from the current vector of CCPs $\mathbf{P}^{(k)}$.

2. Given estimates of the ex-ante value function, update the vector of choice-specific value functions for each player i, denoted $\mathbf{v}_i^{(k)}(a|\boldsymbol{\theta})$, as a function of a candidate vector of structural parameter $\boldsymbol{\theta}$

$$\mathbf{v}_i^{(k)}(a|\boldsymbol{\theta}) = \boldsymbol{\pi}_i(a|\boldsymbol{\theta}) + \beta\widehat{\boldsymbol{\lambda}}'\,\mathsf{E}\left[\boldsymbol{\phi}_i|a\right] \tag{51}$$

where $\boldsymbol{\pi}_i(a|\boldsymbol{\theta})$ are current-period profits parameterized as a function of candidate parameter $\boldsymbol{\theta}$, and $\mathsf{E}\left[\boldsymbol{\phi}_i|a\right]$ is the expected future value of the approximating functions given action a is played, that is, $\mathsf{E}\left[\phi_{ki}(\mathcal{M}_{j,i,t+1})|a, \mathcal{M}_{j,i,t+1}\right]$, which is calculated in a similar fashion as in the previous step (except for the additional conditioning on action a).

3. After pooling the data across all markets and periods, optimize the objective function with respect to the structural parameters. We implement a minimum-distance estimator (Pesendorfer and Schmidt-Dengler (2008), Bugni and Bunting (2021)). The distance between the initial CCPs and predicted CCPs (or alternatively between differences in choice-specific value functions) is minimized, that is,

$$\min_{\boldsymbol{\theta}} ||\log\left(\frac{\mathbf{P}_i^{(0)}(a)}{\mathbf{P}_i^{(0)}(a')}\right) - (\mathbf{v}_i^{(k)}(a|\boldsymbol{\theta}) - \mathbf{v}_i^{(k)}(a'|\boldsymbol{\theta})||_2$$

Denote $\boldsymbol{\theta}^{(k+1)}$ the updated structural parameters.[40]

[40] We do not attempt to calculate and use the optimal weight matrix derived in Bugni and Bunting (2021),

4. Update the CCPs, using the new structural parameters, that is,

$$\mathbf{P}_i^{(k+1)}(a) = \frac{\exp\left(\mathbf{v}_i^{(k)}(a|\boldsymbol{\theta}^{(k+1)})\right)}{\sum_{\widetilde{a}} \exp\left(\mathbf{v}_i^{(k)}(\widetilde{a}|\boldsymbol{\theta}^{(k+1)})\right)}. \tag{52}$$

The $k-$Minimum Distance ($k-$MD) estimator iterates on these steps k times. Bugni and Bunting (2021) show that the $k-$MD estimator is consistent and asymptotically normal for any $k \geq 1$, and for an initial choice of CCPs that is asymptotically equivalent to the frequency estimator (e.g., flexible logit) and a specific optimal weight matrix, the $1-$MD estimator is optimal.

Table A1 shows estimates of the structural parameters based on our baseline estimator (ECCP with finite dependence) and the $1-$MD estimator presented above, using the sample of 846 markets. The two approaches yield very close estimates of entry costs, and profit parameters that are broadly of similar sign and magnitudes. For chains, we note that the magnitude of the (variable) profit coefficients are larger, but marginal effects are of similar magnitude. For example, adding one chain store within 2 miles of an exiting own store reduces mean profits by 11% under ECCP and by 14% under 1-MD. If a store is added in the 2-5mi band, mean store profits increase by 9% under ECCP and 14% under 1-MD.[41]

Profits estimates based on the two sets of parameters are qualitatively similar. To illustrate this point, Figure A1 shows histograms of incumbents' profits (unscaled) based on the baseline and alternative estimation approaches. The magnitude and distribution of profits for single-store firms are comparable across the two specifications. Overall, these results indicate that our baseline estimates of entry costs and profits are robust to the estimation approach providing supporting evidence for our counterfactual exercise.

A.3 Robustness Checks

This section investigates how the various assumptions required by the model impact the quantified effects.

Discount factor. The main estimates use an annual discount factor β equal to 0.9025 (corresponding to 0.95 per 6 months). We examine how the estimation and counterfactual predictions change with different discount factors ranging from 0.85 to 0.95. All else equal, a lower discount factor will be offset with higher estimates of per-period profits, holding entry costs fixed. As expected, we find that estimated mean store profits (conditional on remaining active) are 32% higher with a 0.85 annual discount rate compared to the baseline (0.9025), and 25% lower with a 0.95 annual discount rate. Under these alternative values

given the size of the state space, as it would likely introduce finite-sample bias. We also experimented using the pseudo-likelihood as our objective function, as in Aguirregabiria and Mira (2007), and obtain qualitatively similar results.

[41] In experimenting with the two estimators, we noted that the value function approximation performed better for single-store firms than chains. This is likely due to the fact that the ex-ante value function is simpler for single-store incumbents (since they either stay active or exit); whereas chains have many more choices (remain active, close any of the existing stores, build an additional store in any of the locations, etc.), making it highly non-linear.

Table A1: Estimates of store profits and costs (Baseline ECCP and 1-MD estimators)

	ECCP with finite dependence						1-MD with value function approximation					
	Chain		Grocery		Conv.		Chain		Grocery		Conv.	
Parameters	Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.
Constant	2.201	(0.536)	-0.805	(0.409)	-0.354	(0.273)	12.532	(2.618)	0.613	(0.998)	2.239	(1.054)
Location-level characteristics												
Population (0-2 mi)	0.043	(0.012)	0.161	(0.025)	0.073	(0.010)	0.130	(0.189)	0.110	(0.022)	0.058	(0.007)
Population (2-5 mi)	0.010	(0.004)	0.011	(0.006)	0.005	(0.003)	0.028	(0.034)	0.023	(0.009)	0.017	(0.005)
Income per capita (0-2 mi)	-0.144	(0.048)	-0.040	(0.052)	-0.014	(0.019)	-1.196	(0.202)	-0.186	(0.075)	-0.250	(0.076)
Income per capita (2-5 mi)	-0.004	(0.003)	-0.002	(0.003)	-0.002	(0.002)	-0.023	(0.025)	-0.009	(0.005)	-0.010	(0.003)
Fixed cost components												
Median residential rent	-0.067	(0.044)	0.015	(0.048)	0.002	(0.025)	-0.173	(0.209)	-0.016	(0.038)	-0.044	(0.031)
Distance to own distribution center	-0.062	(0.026)					-0.366	(0.178)				
Measures of competition and cannibalization												
Number of rival chain stores (0-2 mi)	-0.068	(0.015)	-0.144	(0.018)	-0.080	(0.009)	-0.257	(0.086)	-0.054	(0.033)	-0.051	(0.013)
Number of rival chain stores (2-5 mi)	-0.040	(0.021)	-0.013	(0.024)	-0.028	(0.010)	-0.168	(0.094)	-0.056	(0.024)	-0.043	(0.018)
Number of rival grocery stores (0-2 mi)	-0.073	(0.025)	-0.035	(0.019)	-0.032	(0.010)	-0.221	(0.054)	-0.013	(0.018)	-0.049	(0.013)
Number of rival grocery stores (2-5 mi)	-0.080	(0.029)	-0.060	(0.026)	-0.035	(0.013)	-0.233	(0.092)	-0.053	(0.030)	-0.057	(0.017)
Number of rival convenience stores (0-2 mi)	-0.075	(0.022)	-0.123	(0.016)	-0.063	(0.010)	-0.135	(0.050)	-0.143	(0.022)	-0.094	(0.013)
Number of rival convenience stores (2-5 mi)	0.024	(0.023)	-0.037	(0.019)	-0.009	(0.010)	-0.054	(0.070)	-0.052	(0.024)	-0.014	(0.017)
Number of own chain stores (0-2 mi)	-0.091	(0.037)					-0.112	(0.066)				
Number of own chain stores (2-5 mi)	0.082	(0.022)					0.322	(0.078)				
Market-level characteristics												
Population	-0.078	(0.023)	-0.093	(0.022)	-0.059	(0.012)	0.124	(0.464)	-0.005	(0.039)	-0.029	(0.031)
Number of gas stations	0.011	(0.018)	-0.025	(0.020)	-0.057	(0.010)	0.045	(0.066)	-0.042	(0.030)	-0.103	(0.018)
Number of drug stores	0.063	(0.021)	0.019	(0.025)	0.004	(0.017)	0.202	(0.067)	0.070	(0.031)	0.025	(0.035)
Number of supermarkets/centers	0.097	(0.023)	-0.008	(0.017)	0.036	(0.019)	0.375	(0.080)	0.003	(0.029)	0.093	(0.039)
Dynamic investment costs												
Entry cost	2.111	(0.123)	5.478	(0.066)	5.862	(0.043)	2.612	(0.184)	5.496	(0.067)	5.864	(0.044)
Entry cost of additional store	9.734	(0.165)					10.027	(0.080)				

Note: Standard errors are obtained via bootstrap of market-histories for both estimation approaches (20 replications). All continuous variables and store counts are in log. Business density is controlled for. Residential rent is at the location level.

for the discount factor, the counterfactual results remain broadly similar to our baseline predictions.

Number of potential entrants. The set of potential entrants is an important modelling choice in entry games. In our setting, there are two types of entrants: the three dollar store chains are "global" entrants, i.e., they are potential entrant in every market, and their identity is known. Grocery and convenience stores are "local" entrants: each firm considers entry only in a single market. Moreover, we observe entry decisions only by local entrants which end up entering but not by those firms staying out. In the baseline specification, we set the number of "local" potential entrants (by retail format) to the total number of unique stores which have operated at any point in a given market over the period 2008-2019. This is arguably a lower bound on the set of local potential entrants.

We consider how increasing the set of potential entrants affects our structural estimates. The baseline number of local potential entrants (grocery and convenience stores) is increased from the baseline to twice as many as in the baseline. One would expect that, with more local entrants, the model rationalizes observed entry rates (i.e., the number of incumbents) with higher entry costs. This is indeed the case: we find that doubling the number of local potential entrants yields entry costs that are 19% higher for grocery stores and 20% higher for convenience stores relative to the baseline specification. Store profits remain stable across specifications of the number of potential entrants.

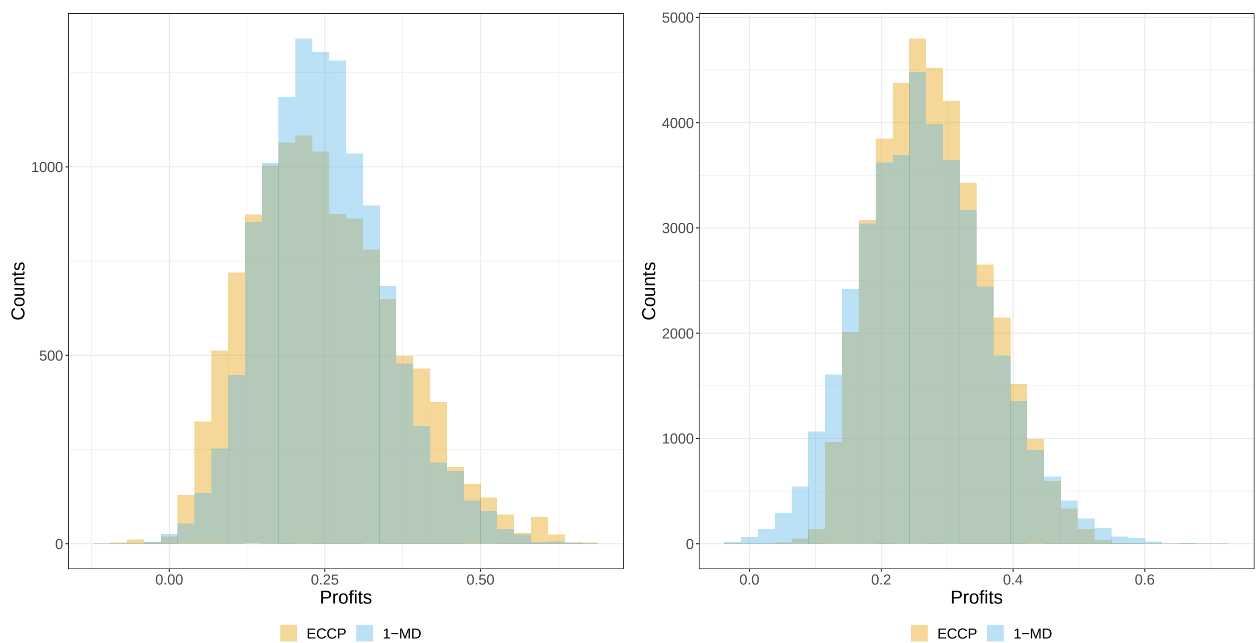


Figure A1: These figures show estimated profits for grocery stores (left) and convenience stores (right) under our baseline estimator (ECCP with finite-dependence) and the alternative estimator (minimum-distance with value function approximation).

A.4 Solution Method for the Dynamic Game

This section provides a detailed overview of the solution method used to find counterfactual equilibria of the dynamic game. 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.[42] We also maintain chains' distribution networks at their state in period T.

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

[42] 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.

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 $\mathcal{M}_{i,j,T+1}$.[43] If i is an incumbent, $\mathbf{P}_{i,T+1}$ 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)

$$\mathbf{v}_{i,T+1}(a) = \boldsymbol{\pi}_{i,T+1}(a) + \beta \mathbf{F}_{i,T+1}(a)[\mathbf{v}_{i,T+1}(exit) + \gamma - \ln(\mathbf{P}_{i,T+1}(exit))] \tag{53}$$

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

$$\mathbf{P}'_{i,T+1}(a) = \frac{\exp\left(\mathbf{v}_{i,T+1}(a)\right)}{\sum_{\widetilde{a}} \exp\left(\mathbf{v}_{i,T+1}(\widetilde{a})\right)} \tag{54}$$

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 (53) is replaced by

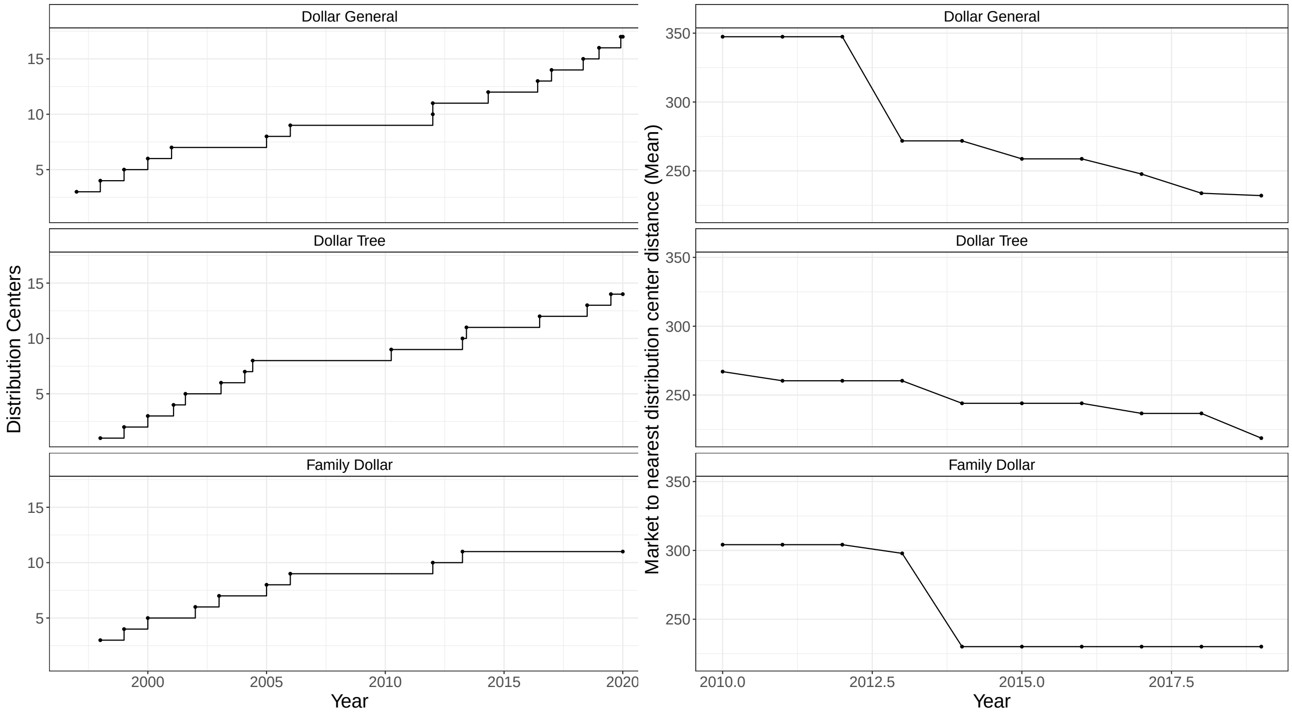
$$\mathbf{v}_{i,t}(a) = \boldsymbol{\pi}_{i,t}(a) + \beta \mathbf{F}_{i,t}(a)[\mathbf{v}_{i,t+1}(exit) + \gamma - \ln(\mathbf{P}^*_{i,t+1}(exit))] \tag{55}$$

where equilibrium CCPs in $t+1$ are used. As markets are independent, we solve the model for each market separately. For our counterfactual analysis, we initialize this algorithm at a large number of starting values and iterate to a fixed point. We found no evidence of multiple equilibria in the counterfactual exercise.

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

B Supplementary Tables and Figures

Figure A2: Distribution centers



(a) Number of distribution centers over time (b) Market to nearest distribution center distance

Figure A3: Distribution center locations in 2019

Table A2: Multinomial logit of single-store firms' choice

	Dependent variable: Firm enters or remains active in location l			
	Grocery	Grocery	Convenience	Convenience
	(1)	(2)	(3)	(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 (0-2 mi)	0.065 (0.033)	−0.060 (0.041)	0.016 (0.029)	−0.096 (0.027)
Number of rival grocery (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.

Table A3: Estimates of stores profits and costs: alternative specifications

	Chains			
	(1)		(2)	
Parameters	Estimate	s.e.	Estimate	s.e.
Constant	2.616	(0.444)	2.318	(0.449)
Location-level characteristics				
Population (0-2 mi)	0.049	(0.015)	0.050	(0.015)
Population (2-5 mi)	0.010	(0.005)	0.012	(0.006)
Income per capita (0-2 mi)	-0.175	(0.049)	-0.176	(0.050)
Income per capita (2-5 mi)	-0.004	(0.003)	-0.005	(0.003)
Fixed cost components				
Median residential rent	-0.072	(0.056)	-0.051	(0.057)
Distance to own distribution center	-0.058	(0.020)	-0.021	(0.021)
Measures of competition and cannibalization				
Number of rival chain stores (0-2 mi)	-0.070	(0.023)	-0.087	(0.023)
Number of rival chain stores (2-5 mi)	-0.048	(0.022)	-0.059	(0.022)
Number of rival grocery stores (0-2 mi)	-0.074	(0.022)	-0.074	(0.022)
Number of rival grocery stores (2-5 mi)	-0.073	(0.025)	-0.072	(0.025)
Number of rival convenience stores (0-2 mi)	-0.073	(0.022)	-0.061	(0.023)
Number of rival convenience stores (2-5 mi)	0.026	(0.022)	0.030	(0.022)
Number of own chain stores (0-2 mi)	-0.094	(0.045)	-0.091	(0.045)
Number of own chain stores (2-5 mi)	0.077	(0.024)	0.076	(0.024)
Market-level characteristics				
Population	-0.092	(0.028)	-0.096	(0.029)
Number of gas stations	0.016	(0.019)	0.013	(0.019)
Number of drug stores	0.068	(0.022)	0.072	(0.022)
Number of supermarkets/centers	0.099	(0.025)	0.104	(0.025)
Dynamic investment costs				
Entry cost for first store in market	2.495	(0.240)	2.316	(0.242)
$\mathbb{1}\{$First store in 100mi radius$\}$			2.365	(0.317)
Entry cost for second+ store in market	9.713	(0.165)	9.526	(0.161)

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 included. Residential rent is at the location level.

Table A4: Expected number of stores by location (census tract)

	Grocery store counts				Convenience store counts			
	CF	Factual	Δn_l^g	% Δn_l^g	CF	Factual	Δn_l^c	% Δn_l^c
All locations	0.631	0.411	-0.22	-0.095	1.134	0.798	-0.336	-0.24
By number of commercial locations								
1	1.473	0.896	-0.577	-0.365	2.199	1.419	-0.781	-0.386
2	0.738	0.483	-0.255	-0.114	1.34	0.93	-0.411	-0.269
3	0.545	0.354	-0.191	-0.079	1.059	0.756	-0.302	-0.222
4	0.521	0.347	-0.175	-0.061	0.901	0.649	-0.252	-0.215
By income and population								
Population below median, Income below median	0.441	0.312	-0.129	0.016	1.027	0.757	-0.271	-0.195
Population above median, Income below median	0.997	0.618	-0.38	-0.236	1.612	1.108	-0.505	-0.286
Population below median, Income above median	0.317	0.237	-0.081	0.043	0.72	0.53	-0.19	-0.186
Population above median, Income above median	0.765	0.475	-0.29	-0.205	1.172	0.794	-0.378	-0.292
By share of minority groups								
Above 0.25	0.863	0.538	-0.325	-0.096	1.536	1.097	-0.439	-0.224
Below 0.25	0.571	0.378	-0.193	-0.095	1.03	0.72	-0.31	-0.244
By share of population with access to vehicle								
Below first quartile (0.89)	0.816	0.513	-0.303	-0.197	1.506	1.069	-0.437	-0.255
Above first quartile (0.89)	0.569	0.377	-0.192	-0.062	1.01	0.707	-0.303	-0.235
By share of population under poverty line								
Below median (0.16)	0.529	0.351	-0.178	-0.08	0.927	0.646	-0.281	-0.242
Above median (0.16)	0.733	0.471	-0.262	-0.111	1.341	0.95	-0.391	-0.237

Note: "Factual" corresponds to the expected number of stores under the market equilibrium (with expansion by dollar store chains). "CF" corresponds to the counterfactual expected number of stores (with no expansion post-2010). 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.