

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

El Hadi Caoui

Brett Hollenbeck

Matthew Osborne

U of Toronto

UCLA Anderson

U of Toronto

July 1, 2025\*

## **Abstract**

We incorporate spatial differentiation into a dynamic oligopoly game of retail competition. We propose a tractable estimation strategy based on the ECCP estimator of [Kalouptsi et al. \(2020\)](#), adapted to games. The model is used to study dollar store chains' expansion in the US 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. Dispersal policies restricting dollar store clustering can reduce retail variety by limiting rivals' ability to spatially differentiate. Our results highlight the importance of accounting for spatial dynamics when evaluating retail entry regulation.

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

*JEL Classification: L13, D43, L51*

---

\*Caoui: [elhadi.caoui@rotman.utoronto.ca](mailto:elhadi.caoui@rotman.utoronto.ca). Hollenbeck: [brett.hollenbeck@anderson.ucla.edu](mailto:brett.hollenbeck@anderson.ucla.edu). Osborne: [matthew.osborne@rotman.utoronto.ca](mailto: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

Over the past several decades, the dramatic rise of dollar store chains has reshaped the US retail sector. 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 and 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 concerned with their impact on retail markets. Concerns center on whether their rapid proliferation 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.<sup>1</sup>

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

---

<sup>1</sup>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 7th, 2025].

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 of firms’ strategy space. Local competitors include grocery and convenience stores.<sup>2</sup> 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 substitute 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.

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,

---

<sup>2</sup>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 \(2013\)](#).

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 Kalouptsi 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.<sup>3</sup> 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. 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 censuses used in the literature. We combine this with data on dollar store distribution center locations and opening dates and demographic information at the census-tract level from the US Census.

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

---

<sup>3</sup>Kalouptsi 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.

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 cannibalization effects 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.<sup>4</sup>

The estimated model is used to evaluate the impact of dollar store expansion on the retail landscape. We consider two main counterfactual scenarios: (i) an entry tax that raises dollar stores’ entry costs (with varying tax rates), and (ii) a 1-mile dispersal policy that prevents new dollar store entry near existing ones. This analysis allows us to quantify changes in the spatial distribution of different retail formats (dollar, grocery, and convenience stores), the extent of within-market reallocation of retail activity, and the resulting impact on policy-relevant consumer outcomes, e.g. travel costs and retail proximity.

In response to these entry regulation policies, grocery and convenience store counts increase modestly, especially in lower-income and densely populated markets and more so under the entry tax than the dispersal policy. Under the entry tax rate that fully halts dollar store expansion, markets experience on average a 18% and 7% increase in the number of grocery and convenience stores respectively. The dispersal policy’s effect on independent stores is more mixed. By shifting dollar store entry toward unserved locations, it leads to a reduction in grocery or convenience store counts in some markets, by limiting these stores’ ability to spatially differentiate.

To isolate the spatial spillovers from the dispersal policy, we decompose changes in store counts into direct and indirect effects. Direct effects capture the additional entry of grocery

---

<sup>4</sup>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.

and convenience stores in locations where dollar store entry is restricted by the dispersal policy. Indirect effects reflect the displacement of grocery and convenience stores, as dollar stores are diverted into previously unserved or less contested locations. For every one-unit reduction in dollar store presence in constrained locations, 0.34 dollar stores enter in unconstrained locations. Conversely, for grocery and convenience stores, 20–27% of the direct gains are offset by losses elsewhere within the same market. The dispersal policy reallocates dollar store entry to locations with higher incomes and lower population density, indicating that such policies may generate uneven distributional effects across neighborhoods within a market.

These findings underscore the importance of accounting for spatial differentiation and within-market reallocation when evaluating the net impact of entry regulation. The counterfactual analysis also highlights how different policy tools—entry taxes and dispersal rules—can reshape the competitive landscape, sometimes producing unintended consequences for the retail format mix.

**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. \(2024\)](#) use a dynamic model and data from Texas to show that dollar stores generally benefit supermarkets by displacing smaller competitors. Their approach differs from the one adopted in this paper as they assume stationarity and model spatial differentiation as a choice of market-level store density. 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 or the effect of dispersal policies. Lastly, [Schneier \(2023\)](#) 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 consumer benefits from private-label products.

In studying the impact of dispersal policies, we contribute to the urban and IO literature showing that place-based regulations can distort entry, reduce competition, and misallocate resources in the retail sector (e.g., [Schivardi and Viviano \(2010\)](#), [Suzuki \(2013\)](#)). Our findings add a new dimension: even policies aimed at redistributing entry—such as dispersal rules that limit store clustering—can create negative spillovers. By forcing dollar stores into less desirable or contested locations, dispersal rules may crowd out single-store retailers in those locations and, in some cases, reduce overall grocery availability.

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 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.<sup>5</sup> A dollar store typically occupies 8,000-12,000 sq ft and stocks

---

<sup>5</sup>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.



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, 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. 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.<sup>6</sup> 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).<sup>7</sup> 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.<sup>8</sup> The primary advantages of this public dataset are its comprehensiveness and annual frequency. Crucially, the panel includes small independent stores, which are typically absent from other retail Census data used in the literature.

---

<sup>6</sup>“*Dollar Tree completes acquisition of Family Dollar.*” Dollar Tree, Press Release, July 6, 2015.

<sup>7</sup>In 2020, two new distribution centers serving both brands opened in Ocala, FL, and Rosenberg, TX.

<sup>8</sup>The dataset has been used in the geography literature studying retail proximity (e.g., Shannon et al. (2018)).

A drawback of this dataset is that entry into the SNAP program may not necessarily indicate the start of operation of a physical store. 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.<sup>9</sup>

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.

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.<sup>10</sup>

Figure 1 shows the evolution of dollar store chains’ distribution centers from 2000 to 2019. The number of distribution centers increased from 14 to 42. Over our sample period (2010-2019), this expansion 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.<sup>11</sup> Appendix Figure A3 shows the locations of distribution centers in 2019.

We use Census demographic data to document consumer heterogeneity across locations

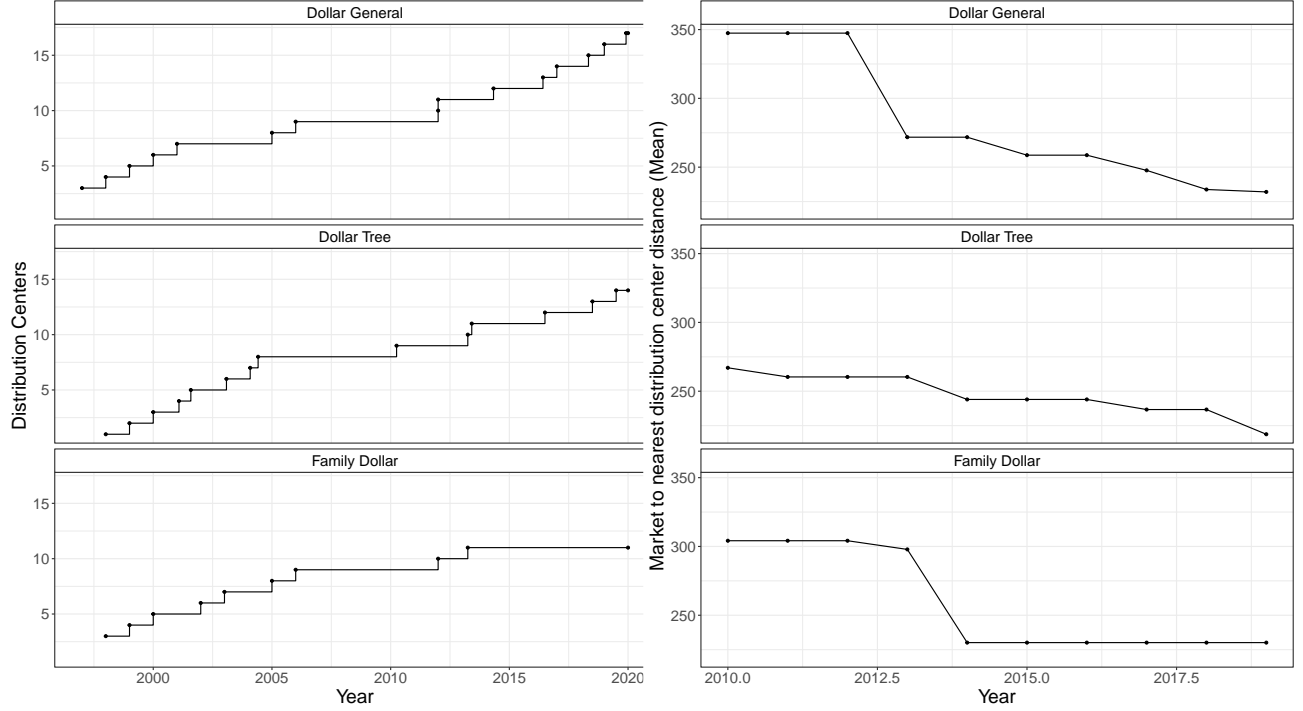
---

<sup>9</sup>Several factors suggest that delayed enrollment is unlikely to pose a major concern in the remaining sample (2010-2019). We model entry at the annual level, so short lags are unlikely to introduce measurement error. The SNAP authorization process is streamlined, with decisions typically made within 45 days (Byrne et al. (2024)). Participation among grocery retailers is near-universal in our period, and the financial incentives to enroll are strong—SNAP benefits accounts for 5–10% of annual sales for most grocery stores, and up to 20% in low-income areas.

<sup>10</sup>Appendix Figure A2 shows the total number of stores at the national level from 2010 to 2019.

<sup>11</sup>Details about the market definition are provided below.

Figure 1: Distribution centers



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

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. 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.

**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 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

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 (2193.5)	4962.8 (2566.1)	4263.9 (2295.8)
Mean Income	22686.3 (7520.7)	23538.8 (8178.7)	31315.2 (16696.5)
Mean Residential Rents	753.9 (252.6)	785.9 (275.4)	1064.8 (455.8)
Mean Share White	.738 (.24)	.739 (.253)	.713 (.254)
Mean Share Black	.166 (.225)	.162 (.234)	.127 (.210)
Mean Share in Poverty	.176 (.108)	.165 (.106)	.136 (.118)
Share HH w/ Vehicle Access	.911 (.084)	.917 (.091)	.904 (.135)
Mean Distance to DG DC	157.8 (134.1)	171.8 (142.5)	227.9 (292.9)
Mean Distance to DT DC	188.2 (111.2)	191.4 (119.7)	189.5 (239.8)
Mean Distance to FD DC	190.1 (124.2)	207.2 (126.1)	275.2 (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.*

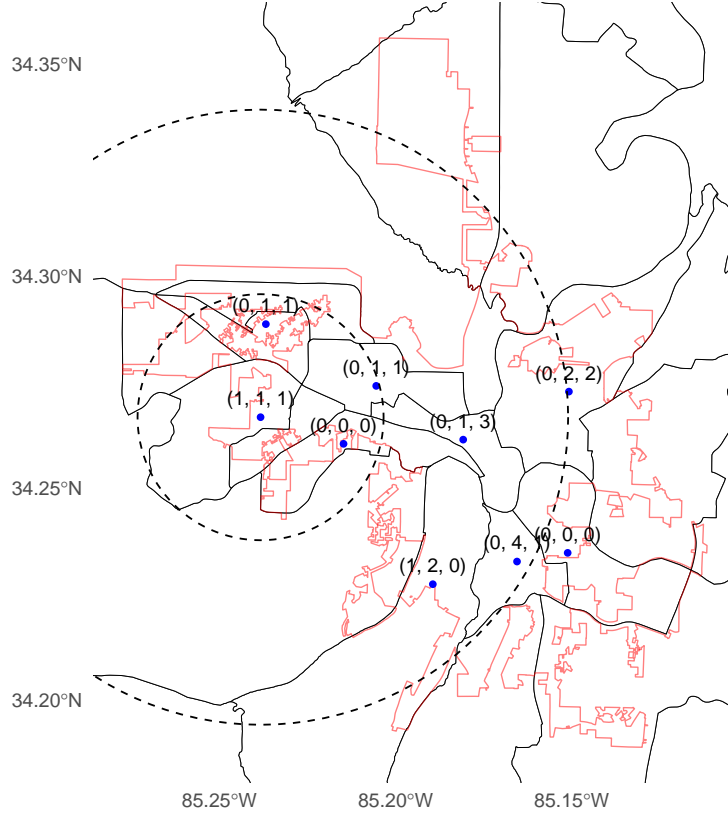
10 miles of a city with a population greater than 5,000 or within 20 miles of a city with a population greater than 25,000. Each market is partitioned into locations, which we define at the Census tract level.<sup>12</sup>

We focus on competition between dollar store chains and independent grocery and convenience stores. The previous literature on retail competition has shown that competition 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. Figure 2 illustrates the spatial distribution of locations, retail store counts per tract, and distance bands in the market Rome, GA, as of 2010.

We exclude gas stations, drugstores, and supermarket chains from the set of strategic players. In the case of supermarkets and big box retailers, previous literature (Ellickson and

<sup>12</sup>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>.

Figure 2: Census tracts and store counts in the market Rome, GA



Notes: The plot displays the boundaries of census tracts in Floyd County (black outlines), with the boundary of the city of Rome, GA shown in red. Blue dots represent the population-weighted centroids of each tract in the city, calculated using 2010 population. Each centroid is labeled with the number of stores in that tract in 2010, by retail format, in the form (Grocery, Convenience, Dollar). Dashed black circles denote distance bands of 0-2 and 2-5 miles centered around one tract.

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” locations 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 in 2008-2019.

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

Variable	Mean	Median	Std.Dev	Min	Max
<i>Market-level characteristics</i>					
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				
<i>Location-level characteristics</i>					
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 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 3: Descriptive Statistics: Stores (2010-2019)

Variable	Mean	Median	Std.Dev	Min	Max
<i>Market-level characteristics</i>					
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				
<i>Location-level characteristics</i>					
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.<sup>13</sup> In what follows, we consider a market  $m$  that is partitioned into locations  $l = 1, \dots, L$ . We index firms by  $i = 1, \dots, I_m$ , and assume that market  $m$  has  $I_{c,m}$  chains, the remaining firms being single-store retailers. Time is discrete and denoted by  $t = 1, \dots, \infty$ .

**State space.** At the beginning of period  $t$  a chain's network of stores is represented by the vector  $\mathbf{n}_{it} = (n_{i1t}, \dots, n_{iLt})$ , 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  $\bar{n}$  stores in a location, such that  $n_{ilt} \in \{0, 1, \dots, \bar{n}\}$ . Single-store firms can operate only one store per market. The spatial market structure at period  $t$  is represented by the vector  $\mathbf{n}_{mt} = (\mathbf{n}_{it})_{i \in I_m}$ . 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,

---

<sup>13</sup>Actions by players in one market do not impact the equilibrium played in another market.



supermarkets, and gas stations) that are not players in the game but can be payoff-relevant.<sup>14</sup> For multi-store firms, let  $d_{imt}$  denote market  $m$ 's distance from  $i$ 's closest distribution center and  $\mathbf{d}_{mt} = (d_{imt})_{i \in I_{c,m}}$  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.<sup>15</sup> The rollout of distribution centers over time is a key source of non-stationarity in our setting. As chains expand their networks, the resulting decrease in distance to distribution centers reduces the fixed cost of operating a store, directly affecting entry decisions.<sup>16</sup>

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 observed and collected, from the perspective of firm  $i$ , into the vector  $\mathcal{M}_{j,i,t}$ , with particular realization  $j$  at time  $t$ . That is

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

## Actions.

*Multi-store firms* We assume that a chain may open or close at most one store per period per market.<sup>17</sup> 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

---

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

<sup>15</sup>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.

<sup>16</sup>We test whether this evolution can be captured by a stationary first-order Markov process by regressing, for each chain, current distance (from a market to the nearest distribution center) on lagged distance and a linear time trend, including market fixed effects. The time trend is highly significant, indicating systematic time dependence that supports our modeling assumption of non-stationarity.

<sup>17</sup>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.

$t$ , denoted  $A(\mathbf{n}_{it})$ , is such that  $A(\mathbf{n}_{it}) = \{0\} \cup \{l_+ : n_{ilt} < \bar{n}\} \cup \{l_- : n_{ilt} > 0\}$ . Note that this choice set can have more than  $L + 1$  choice alternatives (if, for some  $l$ ,  $0 < n_{ilt} < \bar{n}$ ). 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.<sup>18</sup>

*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 stay out are replaced by a new set of potential entrants in the next 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 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 that are firm-specific and private information.

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

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

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 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

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

---

<sup>18</sup>Exit by chains from a market is observed to be permanent in the sample and time horizon we consider. This assumption implies that in the long run, the probability that "all chains have exited 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.

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

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

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., a 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

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

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

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

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:

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

Similarly to entry costs, the exit value is assumed to depend 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  $\epsilon_{it} = \{\epsilon_{it}(a)\}_{a \in A(\mathbf{n}_{it})}$ . We assume that the shocks  $\epsilon_{it}$  are independently distributed across firms and over time and have a cumulative distribution function  $G(\cdot)$  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  $\epsilon_{it}$ , including spatially correlated shocks. In our application, these shocks will be distributed Type 1 extreme value, scaled by a parameter  $\theta^\epsilon$ .

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

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

**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 retailers have arguably accurate information on store counts and demographics in each location. Therefore, for this type of market, 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  $(\mathcal{M}_{j,i,t}, \epsilon_{it})$ .

A strategy profile is denoted

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

Given strategy profile  $\alpha$ , firm  $i$ 's value function satisfies

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

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

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

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  $\Pi_{j \neq i} g(\epsilon_{j,t})$  and strategies  $\alpha_j$  for  $j \neq i$ .

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

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

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

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

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

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

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

of the private shock  $\epsilon_{it}$

$$\begin{aligned} \bar{V}_{i,t}^{\mathbf{P}}(\mathcal{M}_{j,i,t}) = & \int \max_{a_{it} \in A(n_{it})} \left\{ \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \right. \\ & \left. + \beta \sum_{a_{-it}} \int \bar{V}_{i,t+1}^{\mathbf{P}}(\mathcal{M}_{j,i,t+1}) dF_t(\mathcal{M}_{j,i,t+1} | \mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it} | \mathcal{M}_{j,i,t}) \right\} dG(\epsilon_{it}). \end{aligned} \quad (14)$$

If private shocks are distributed Type 1 extreme value (with scale parameter  $\theta^\epsilon$ ), an optimal strategy for firm  $i$  will map into conditional choice probabilities of the form

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

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), Kaloupt-sidi 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).

Lemmas 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

$$\bar{V}_{i,t+1}^{\mathbf{P}}(\mathcal{M}_{j,i,t+1}) = v_{i,t+1}^{\mathbf{P}}(\tilde{a}, \mathcal{M}_{j,i,t+1}) + \gamma - \ln(P_{i,t+1}(\tilde{a}|\mathcal{M}_{j,i,t+1})) \quad (16)$$

where  $\gamma$  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)

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

This allows us to replace the ex-ante value function  $\bar{V}_{i,t+1}^{\mathbf{P}}(\mathcal{M}_{j,i,t+1})$  by known functions of the structural parameters and CCPs. Equation (16) is used extensively in the estimation approach presented below. Importantly, chain entrants differ from single-store entrants in that they are long-lived: they can delay entry into a market without being replaced by a new potential entrant. As a result, the only terminal choice for a chain is exit from incumbency.

## 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  $(\beta, G)$  are assumed to be known.<sup>19</sup>

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.<sup>20</sup> When estimating the model, we normalize the exit value to zero. A consequence of

---

<sup>19</sup>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. Identification in this non-stationary setting requires a large cross-section of markets to be observed at each specific time period to provide sufficient variation.

<sup>20</sup>More recent contributions include Kalouptsi et al. (2021b) and Kalouptsi et al. (2021a).

this “normalization” restriction is that the estimated entry costs will reflect the true *sunk* costs (entry cost net of exit value), and estimated fixed costs will reflect the true fixed costs in addition to the exit value scaled by  $(1 - \beta)$ .

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

Estimation methods such as policy evaluation ([Aguirregabiria and Mira \(2007\)](#)) or forward simulation ([Bajari et al. \(2007\)](#)), are not applicable to non-stationary games without restrictive assumptions about CCPs beyond the sample period (post-2019).<sup>21</sup> Furthermore, location decisions and multi-store firms generate a high-dimensional state space, making these methods computationally infeasible without approximations.

We address these challenges in two ways. First, we exploit the finite dependence property ([Arcidiacono and Miller \(2011, 2019\)](#)). This allows us to express the period- $t$  choice-specific

---

<sup>21</sup>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.



value function in terms of known period- $t + 1$  CCPs and the structural profit function. However, this does not eliminate the need to integrate these value functions over a high-dimensional state space. To address this, we adopt the linear IV strategy of [Kalouptsi et al. \(2020\)](#), which combines insights from finite dependence and the GMM-Euler approach of [Aguirregabiria and Magesan \(2018\)](#). The resulting method—ECCP—invokes rational expectations and replaces expected behavior with observed outcomes, yielding linear estimating equations that bypass numerical integration.

We extend the ECCP estimator of [Kalouptsi et al. \(2020\)](#) from single-agent dynamic discrete choice problems to dynamic games. To our knowledge, this marks the first application of this linear regression approach in a game-theoretic context. This extended estimator is particularly useful for settings with high-dimensional state spaces and finite dependence, or for analyzing entry dynamics in other non-stationary environments, such as those characterized by information waves ([Chi \(2025\)](#)). The estimation strategy is also useful in settings where the estimated structural objects can serve as sufficient statistics to assess policy impacts, or where researchers wish to explore multiple alternative specifications efficiently.

We derive the estimating equations and outline the intuition behind our approach using the case of a chain potential entrant. The case of single-store firms and chain incumbents is presented in [Appendix A.1](#).

A chain entrant is long-lived and can delay entry to a later period (e.g., if the chain anticipates opening a distribution center closer to the market in the future). The choice-specific value function from entering into location  $l$  ( $a_{it} = l_+$ ) is given by

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) = -\theta_i^{EC} + \beta \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln P_{i,t+1}(l_- | \mathcal{M}_{j,i,t+1}) | l_+, \mathcal{M}_{j,i,t}] \quad (18)$$

where Equation (16) is used with  $\tilde{a}$  set to exit ( $l_-$ ) in  $t + 1$ . The choice-specific value function

from staying out ( $a_{it} = 0$ ) is given by

$$\begin{aligned}
v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) &= \beta \mathbb{E} \left( v_{i,t+1}^{\mathbf{P}}(a'_i = l_+, \mathcal{M}_{j,i,t+1}) + \gamma - \ln(P_{i,t+1}(l_+ | \mathcal{M}_{j,i,t+1}) | 0, \mathcal{M}_{j,i,t}) \right) \\
&= \beta \mathbb{E} \left( -\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+ | \mathcal{M}_{j,i,t+1}) \right) \\
&\quad + \beta \mathbb{E} [vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_- | \mathcal{M}_{j,i,t+2}) | 0, \mathcal{M}_{j,i,t}]
\end{aligned} \tag{19}$$

The first equality uses Equation (16) with  $\tilde{a}$  set arbitrarily to  $l_+$ , in period  $t + 1$ . The second equality uses Equation (16) with  $\tilde{a}$  set to  $l_-$  (exit) in period  $t + 2$ . Equation (19) states that, if an entrant stays out, they internalize the option value from entering at a later period. Differences in these choice-specific value functions can alternatively be expressed using current-period CCPs as

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \log \left( \frac{P_{i,t}(l_+ | \mathcal{M}_{j,i,t})}{P_{i,t}(0 | \mathcal{M}_{j,i,t})} \right) \tag{20}$$

Plugging in Equations (18) and (19) into the LHS of Equation (20) yields an optimality condition that can be re-expressed as moment conditions that avoid explicit integration. Under rational expectations, the conditional expectation of future CCPs and profits equals their realized value plus an expectational error orthogonal to the period- $t$  state variables.<sup>22</sup>

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). Correspondingly, the expectational errors ( $w_{it}, u_{i,t+1}$ ) are defined as

$$\begin{aligned}
w_{it} &= \mathbb{E} \left[ -\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+ | \mathcal{M}_{j,i,t+1}) | 0, \mathcal{M}_{j,i,t} \right] \\
&\quad - \left( -\theta_i^{EC} + \gamma - \ln P_{i,t+1}(l_+ | \mathcal{M}_{j,i,t+1}^*) \right)
\end{aligned} \tag{21}$$

$$\begin{aligned}
u_{i,t+1} &= \mathbb{E} [vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_- | \mathcal{M}_{j,i,t+2}) | 0, \mathcal{M}_{j,i,t}] \\
&\quad - \left( vp_{i,l}(\mathcal{M}_{j,i,t+2}^*) - FC_i + \gamma - \ln P_{i,t+2}(l_- | \mathcal{M}_{j,i,t+2}^*) \right)
\end{aligned} \tag{22}$$

---

<sup>22</sup>This idea originates in the estimation of continuous choice dynamic models using Euler equations (e.g., Hansen and Singleton (1982)). Other approaches to simplify the computation of the expected value function in high-dimensional settings include approximating this function using the averages of value functions of past iterations (Pakes and McGuire (2001), Imai et al. (2009), Ishihara and Ching (2016)).

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 (Equation (18)), we define the expectational error  $v_{it}$  as

$$v_{it} = \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})|l_+, \mathcal{M}_{j,i,t})] \\ - (vp_{i,l}(\mathcal{M}_{j,i,t+1}^*) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1}^*))) \quad (23)$$

These errors satisfy, for any function  $g(\cdot)$  of period- $t$  states, the orthogonality conditions:

$$\mathbb{E}[g(\mathcal{M}_{j,i,t})'[w_{it}, v_{it}, u_{i,t+1}]] = \mathbf{0}_3 \quad (24)$$

The moment conditions (Equation (24)) do not require integration over the state space 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.

In a game setting, replacing these moment conditions by their sample counterparts (in the form of a linear IV regression as in Kalouptside et al. (2020)) will not yield consistent estimates of the structural parameters. The key issue is that rivals' states are endogenous: they are affected by the focal firm's past action, a feature absent in single-agent problems. Specifically, the error  $u_{i,t+1}$  (Equation (22)) involves an expectation over  $\mathcal{M}_{j,i,t+2}$ , conditional on  $a_{it}$  being 0: e.g., rivals' states in  $t+2$  are conditional on focal firm  $i$  choosing to stay out in period  $t$ . Since the empirical distribution of rivals' states in  $t+2$  reflects realized actions rather than counterfactual ones (i.e.,  $a_{it} = 0$ ), sample averages will not generally converge to the relevant conditional expectation.<sup>23</sup>

To address this selection problem and ensure that the empirical counterpart of the moment condition (Equation (24)) yields consistent estimates of the structural parameters, we use an importance sampling approach (Kloek and van Dijk (1978), Geweke (1989)). Define

---

<sup>23</sup>This selection problem only concerns players' states in  $t+2$ . In  $t+1$ , rivals' states do not depend on  $a_{it}$  (conditional on the current state) because firms take their actions simultaneously between  $t$  and  $t+1$ . For example, the observed distribution of stores operated by Dollar Tree and Family Dollar in  $t+2$  depends on Dollar General's entry choice in period  $t$ .

the importance weights

$$\psi_{a_1, a_2}(\mathbf{n}_{-i, t+2} | \mathcal{M}_{j, i, t}) \equiv \frac{P(\mathbf{n}_{-i, t+2} | a_{it} = a_1, \mathcal{M}_{j, i, t})}{P(\mathbf{n}_{-i, t+2} | a_{it} = a_2, \mathcal{M}_{j, i, t})}. \quad (25)$$

Intuitively, the weight  $\psi_{a_1, a_2}(\mathbf{n}_{-i, t+2} | \mathcal{M}_{j, i, t})$  is a likelihood ratio: it reflects how much more likely the period- $t+2$  state of rivals would be if focal firm  $i$  had taken action  $a_1$  rather than  $a_2$  in period  $t$ , conditional on the current state  $\mathcal{M}_{j, i, t}$ . These weights are well-defined for any pair  $(a_1, a_2)$  because the support of  $\mathbf{n}_{-i, t+2}$  is independent of  $a_{it}$ .<sup>24</sup> The following lemma is a direct consequence of the definition of conditional expectation.

**Lemma 1.** *Let  $f(\cdot)$  be a measurable function such that  $\mathbb{E}[|f(\mathcal{M}_{j, i, t+2})| | a_{it} = a, \mathcal{M}_{j, i, t}] < \infty$  for  $a \in \{0, \tilde{a}\}$ . Then, for every alternative action  $\tilde{a}$ ,*

$$\mathbb{E}[f(\mathcal{M}_{j, i, t+2}) | a_{it} = 0, \mathcal{M}_{j, i, t}] = \mathbb{E}\left[f(\mathcal{M}_{j, i, t+2}) \psi_{0, \tilde{a}}(\mathbf{n}_{-i, t+2} | \mathcal{M}_{j, i, t}) \mid a_{it} = \tilde{a}, \mathcal{M}_{j, i, t}\right].$$

*Proof.* Let  $\mathcal{M}_{j, i, t+2} = (\mathbf{n}_{i, t+2}, \mathbf{n}_{-i, t+2}, \mathbf{x}_{m, t+2}, \mathbf{d}_{m, t+2})$ . The vectors  $\mathbf{x}_{m, t+2}$  and  $\mathbf{d}_{m, t+2}$  evolve exogenously and  $\mathbf{n}_{i, t+2}$  is a deterministic vector because firm  $i$  operates one store in location  $l$  in period  $t+2$  (Equation (19)). We have that

$$\begin{aligned} & \mathbb{E}[f(\mathcal{M}_{j, i, t+2}) | a_{it} = 0, \mathcal{M}_{j, i, t}] \\ &= \sum_{\mathbf{n}_{-i, t+2}, \mathbf{x}, \mathbf{d}} f(\mathbf{n}_{i, t+2}, \mathbf{n}_{-i, t+2}, \mathbf{x}, \mathbf{d}) P(\mathbf{n}_{-i, t+2} | a_{it} = 0, \mathcal{M}_{j, i, t}) P_{x, d}(\mathbf{x}, \mathbf{d} | \mathcal{M}_{j, i, t}) \\ &= \sum_{\mathbf{n}_{-i, t+2}, \mathbf{x}, \mathbf{d}} f(\mathbf{n}_{i, t+2}, \mathbf{n}_{-i, t+2}, \mathbf{x}, \mathbf{d}) \psi_{0, \tilde{a}}(\mathbf{n}_{-i, t+2} | \mathcal{M}_{j, i, t}) P(\mathbf{n}_{-i, t+2} | a_{it} = \tilde{a}, \mathcal{M}_{j, i, t}) P_{x, d}(\mathbf{x}, \mathbf{d} | \mathcal{M}_{j, i, t}) \\ &= \mathbb{E}[f(\mathcal{M}_{j, i, t+2}) \psi_{0, \tilde{a}}(\mathbf{n}_{-i, t+2} | \mathcal{M}_{j, i, t}) \mid a_{it} = \tilde{a}, \mathcal{M}_{j, i, t}], \end{aligned} \quad (26)$$

which proves the identity.  $\square$

In constructing the sample moment counterparts, functions of  $\mathcal{M}_{j, i, t+2}$  are scaled by the importance weights  $\psi_{0, \tilde{a}}$ , where  $\tilde{a}$  is the action played in the data. Define the re-weighted

---

<sup>24</sup>The logit shocks imply that CCPs have full support conditional on  $\mathcal{M}_{j, i, t}$ .

version of the expectational error  $u_{i,t+1}$ , for each  $a_{it} = \tilde{a}$  as

$$\begin{aligned} \tilde{u}_{i,t+1} = & \mathbb{E} [vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_- | \mathcal{M}_{j,i,t+2}) | 0, \mathcal{M}_{j,i,t}] \\ & - \psi_{0,\tilde{a}}(\mathbf{n}_{-i,t+2}^* | \mathcal{M}_{j,i,t}) (vp_{i,l}(\mathcal{M}_{j,i,t+2}^*) - FC_i + \gamma - \ln P_{i,t+2}(l_- | \mathcal{M}_{j,i,t+2}^*)) \end{aligned} \quad (27)$$

Lemma 1 ensures that  $\mathbb{E}[\tilde{u}_{i,t+1} | \mathcal{M}_{j,i,t}]$ —and a fortiori the moment condition  $\mathbb{E}[\tilde{u}_{i,t+1} g(\mathcal{M}_{j,i,t})]$ —equals zero for each  $a_{it} = \tilde{a}$ .<sup>25</sup> Substituting Equations (18)–(20) and replacing expectations with observed values and expectational errors (Equations (21), (23), and (27)), we estimate structural parameters via the following regression:

$$\begin{aligned} Y_{it}^{entrant} = & [-\theta_i^{EC} + \beta(vp_{i,l}(\mathcal{M}_{j,i,t+1}^*) - FC_i)] \\ & - [-\beta\theta_i^{EC} + \psi_{0,\tilde{a}}(\mathbf{n}_{-i,t+2}^* | \mathcal{M}_{j,i,t})\beta^2(vp_{i,l}(\mathcal{M}_{j,i,t+2}^*) - FC_i)] \\ & + (v_{it} - w_{it} - \beta\tilde{u}_{i,t+1}) \end{aligned} \quad (28)$$

where the left-hand side variable collects known functions of the CCPs

$$\begin{aligned} Y_{it}^{entrant} = & \log \left( \frac{P_{i,t}(l_+ | \mathcal{M}_{j,i,t})}{P_{i,t}(0 | \mathcal{M}_{j,i,t})} \right) \\ & - \beta[\gamma - \ln P_{i,t+1}(l_- | \mathcal{M}_{j,i,t+1}^*, a_{it} = l_+)] \\ & + \beta[\gamma - \ln P_{i,t+1}(l_+ | \mathcal{M}_{j,i,t+1}^*, a_{it} = 0)] \\ & + \beta^2\psi_{0,\tilde{a}}(\mathbf{n}_{-i,t+2}^* | \mathcal{M}_{j,i,t})[\gamma - \ln P_{i,t+2}(l_- | \mathcal{M}_{j,i,t+2}^*)], \end{aligned} \quad (29)$$

and regressors entering the profit function in  $t + 1$  and  $t + 2$  (which may be correlated with the expectational errors) are instrumented using the values of these regressors in period  $t$ .

<sup>26</sup>

---

<sup>25</sup>Omitting the reweighting would bias the estimated choice-specific value of staying out,  $v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t})$ . The expected number of rival firms operating in  $t + 2$  would be under-estimated, because we would be assuming the focal firm stayed out in period  $t$ , whereas in the data, the firm may have entered. To match observed CCPs, the model would then require a stronger (more negative) effect of the number of active rivals on profits, biasing the estimated parameters away from zero.

<sup>26</sup>To evaluate the robustness of our estimation 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 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.<sup>27</sup>

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.<sup>28</sup> An entrant chain can either build a store in one of the locations in the

---

<sup>27</sup>Our proxy-variable approach to addressing endogeneity in controlled state variables is closely related to recent IV-based methods such as [Berry and Compiani \(2023\)](#), as well as the finite mixture framework of [Kasahara and Shimotsu \(2009\)](#). Like IV methods, our approach aims to address the correlation between observed state variables (e.g., number of incumbents) and unobserved location heterogeneity. Instead of using instruments, we include a proxy variable—business density—constructed from long-run market outcomes, which we assume captures the relevant heterogeneity. This offers a tractable solution in settings with high-dimensional heterogeneity, where each location differs in its appeal, though at the cost of relying on the strength of the proxy. Our method also aligns with the logic in [Kasahara and Shimotsu \(2009\)](#), in that it leverages observed variation in covariates to control for latent unobserved location heterogeneity.

<sup>28</sup>In an ideal world, CCP would be estimated nonparametrically. For instance, [Kalouptside et al. \(2020\)](#) assume that the CCPs 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 stores, 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.

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.<sup>29</sup> 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 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, reflecting agglomeration effects due to unobserved location-level amenities.<sup>30</sup> 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, 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 pay-offs. 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. 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

---

<sup>29</sup>The small number of observations where a chain opens more than one store in a period are not included when calculating the likelihood.

<sup>30</sup>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)
<i>Location-level characteristics</i>				
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)
<i>Cost shifters</i>				
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)
<i>Measures of competition</i>				
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)
<i>Market-level characteristics</i>				
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.*

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



Table 6: Estimates of store profits and costs

Parameters	Chains		Grocery Store		Convenience Store	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
Constant	2.616	(0.444)	-1.177	(0.485)	-0.876	(0.220)
<i>Location-level characteristics</i>						
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)
<i>Fixed cost components</i>						
Median residential rent	-0.072	(0.056)	0.050	(0.050)	0.009	(0.028)
Distance to own distribution center	-0.058	(0.020)				
<i>Measures of competition and cannibalization</i>						
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)				
<i>Market-level characteristics</i>						
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)
<i>Dynamic investment costs</i>						
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.*

scale parameter of firm shocks  $\theta^\epsilon$  to match revenue data for all dollar stores operating in the markets under consideration, obtained from Nielsen TDLinX.<sup>31</sup> Table 7 shows mean store 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

<sup>31</sup>Specifically, we convert the revenue data from TDLinX into profits (deflated to 2010), assuming a 5% net profit rate, and calibrate the scale parameter  $\theta^\epsilon$  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\$. Our approach to scaling model-predicted profits using external revenue data is related in spirit to Ellickson and Misra (2012), though we do not estimate structural revenue functions; we do adjust for selection by incorporating the expected value of the choice-specific shock conditional on remaining active  $\mathbb{E}[\epsilon_{it}|a_{it} = 1]$  when constructing model-predicted profits.

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
<i>Percentage change in mean store profits from</i>			
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.*

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.<sup>32</sup> Second, the presence of dollar stores also significantly 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.<sup>33</sup> The location of the market relative to a chain's closest

<sup>32</sup>The magnitude of the business-stealing effects is consistent with anecdotal evidence from grocery store owners. For instance, the owner of the Foodliner store in Haven, KS reports,

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

<sup>33</sup>This pattern aligns with existing evidence that cannibalization effects decay significantly with distance; see Pancras et al. (2012).

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, 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.<sup>34</sup> 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 our model estimates for counterfactual simulations. We evaluate the impact of dollar store expansion using two sets of entry regulation policies: an entry tax and a dispersal policy. For each market, we simulate the counterfactual where dollar store chains are subject to entry regulations and compare them to the baseline with free entry. The equilibrium CCPs under each scenario are used to simulate each market from 2010 to 2019.

The counterfactual exercise quantifies changes in the spatial distribution of retail stores from regulating dollar store entry, as well as the within-market reallocation of retail activity.

---

<sup>34</sup>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.

Accounting for within-market spillover effects is crucial to correctly assess the net market impact of regulation. We also quantify changes in policy-relevant outcomes, such as average household travel costs to different types of retailers. Due to computational limitations, we restrict the exercise to markets with up to four commercial locations, covering 458 out of 846 markets in our sample.<sup>35</sup> The method used for solving for counterfactual MPE and dealing with the large dimensional state space is presented in Appendix A.4.

## 6.1 The Impact on Spatial Market Structure

In the first exercise, we assume that local authorities impose a tax on new dollar stores, which we implement by increasing dollar store chains’ entry cost by a factor  $\tau \in \{50\%, 100\%, 150\%, 200\%\}$ . In the second exercise, we assume that local authorities impose dispersal restrictions on dollar store chains: i.e., a dollar store may not enter within 1 mile of an existing dollar store.<sup>36</sup> This policy applies to any dollar store regardless of whether the potential entrant and incumbent are owned by the same chain. We set the 1-mile cut-off to mimic typical dispersal policies used in various cities across the U.S.<sup>37</sup>

We note two caveats with these counterfactual exercises. First, the identification of counterfactuals in dynamic discrete games is an active area of research (Kalouptside et al. (2017), Kalouptside et al. (2021b)). For the entry tax counterfactual, we experiment with multiplicative and additive taxes and find broadly consistent results. Second, entry is modeled at the census tract level and, under the dispersal policy, we measure distances between census tract centroids rather than between the actual store locations. For instance, if a dollar store is present in a given census tract, we assume that no other dollar store can enter in census

---

<sup>35</sup>This subsample consists of all markets with one to three locations and 20 markets with 4 locations.

<sup>36</sup>The dispersal policy is dynamic in the sense that dollar stores’ strategy space at time  $t$  is a function of spatial market structure at time  $t$ .

<sup>37</sup>See <https://ilsr.org/dollar-store-restrictions/> for a list of cities with such restrictions. The definition used most commonly for a dollar store is: “retail sales uses with floor area less than 12,000 square feet that offer for sale a combination and variety of convenience shopping goods and consumer shopping goods; and continuously offer a majority of the items in their inventory for sale at a price less than \$10.00 per item.” In practice, the stores concerned by these restrictions are operated by the three main dollar store chains.

tracts within 1-mile of the focal census tract. While measuring distances between census tract centroids rather than exact store locations introduces approximation, the exercise remains informative about the directional effects of dispersal policies on entry patterns and market structure.

In what follows, we refer to the factual free entry equilibrium as “Baseline,” the entry tax counterfactual is referred to by level of the tax  $\tau \in \{50\%, 100\%, 150\%, 200\%\}$ , and the dispersal policy counterfactual is referred to as “Dispersal.”

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  $\mathbb{E}[n_{lT}^g | \mathbf{n}_{m0}]$ ,  $\mathbb{E}[n_{lT}^c | \mathbf{n}_{m0}]$ , and  $\mathbb{E}[n_{lT}^d | \mathbf{n}_{m0}]$  denote the expected number of grocery ( $g$ ), convenience ( $c$ ), and dollar ( $d$ ) stores in location  $l$  of market  $m$  in the last period  $T$  (2019) under the “Baseline” scenario. Similarly, let  $\mathbb{E}[\tilde{n}_{lT}^g | \mathbf{n}_{m0}]$ ,  $\mathbb{E}[\tilde{n}_{lT}^c | \mathbf{n}_{m0}]$ , and  $\mathbb{E}[\tilde{n}_{lT}^d | \mathbf{n}_{m0}]$  denote these expectations under one of the counterfactuals. 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 effect of a counterfactual entry regulation policy on the number of stores of type  $j \in \{g, c, d\}$  operating in location  $l$  is computed as

$$\Delta n_l^j = \mathbb{E}[\tilde{n}_{lT}^j | \mathbf{n}_{m0}] - \mathbb{E}[n_{lT}^j | \mathbf{n}_{m0}]. \quad (30)$$

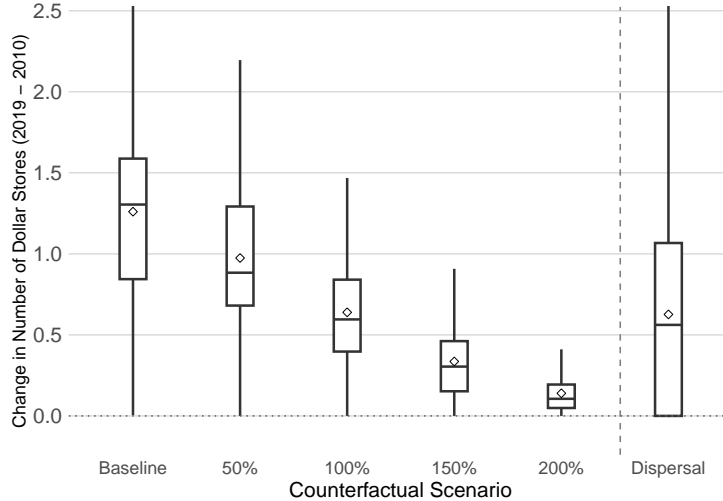
The effect of a counterfactual entry regulation policy on the number of stores of type  $j \in \{g, c, d\}$  operating in market  $m$  (with  $L$  locations) is obtained by summing over locations,

$$\Delta n_m^j = \sum_{l=1}^L \Delta n_l^j. \quad (31)$$

**Effectiveness of Counterfactual Policies.** We first analyze how effective the various counterfactual policies are at curbing dollar store entry. For each scenario and each market, we compute the difference between the (expected) number of dollar stores in the last period

(2019) and the observed number of dollar stores in the first period (2010). Figure 3 shows boxplots of the distribution of this difference over markets, for each scenario. Under the baseline, a typical market adds 1.3 new dollar stores. As expected, higher levels of the entry tax lead to lower dollar store entry. When the tax reaches 200%, entry is nearly eliminated in most markets. The high level of entry tax required to block all dollar store entry is consistent with our estimates showing that dollar store chains earn profits about twice those of single-store retailers while facing entry costs roughly half as large. The dispersal policy also dampens entry but appears less effective than a high entry tax in curbing expansion.

Figure 3: Dollar Store Expansion under Free and Regulated Entry



*Notes:* The box spans the interquartile range (IQR), from the 25th to the 75th percentile of the distribution. The horizontal line inside the box marks the median, and the diamond marker denotes the mean. Whiskers extend to the largest and smallest values within 1.5 times the IQR from the box; observations beyond this range are omitted from display.

**Market-level analysis.** We consider the market-level impact of entry regulation policies. Figure 4 shows boxplots of the change in the number of grocery, convenience, and dollar stores ( $\Delta n_m^g$ ,  $\Delta n_m^c$ ,  $\Delta n_m^d$ ) in levels and percentages.

Grocery and convenience store counts rise modestly under all counterfactuals, reflecting the business-stealing effect of dollar store entry. Under a 200% entry tax, the expected number of stores per market increases on average by 18% for grocery stores and 7% for

convenience stores. By contrast, dollar store counts decline sharply: the 200% tax reduces their average number by 50% relative to baseline.<sup>38</sup>

The dispersal policy also limits dollar store expansion, though less severely: the policy appears effective, on average, at increasing grocery store counts, as intended by municipalities adopting such regulation. As can be seen in Figure 4, in a few markets, it leads to a decline in grocery and convenience stores, as the dispersal rule forces dollar stores into locations that would otherwise be attractive to single-store retailers, leading to slight declines in their counts. This underscores that dispersal policies can reconfigure the competitive landscape in ways that sometimes hurt single-store retailers by limiting their ability to spatially differentiate. We explore these location-level spillovers in more detail below.

The 200% entry tax has a larger positive effect on grocery and convenience store counts ( $\Delta n_m^g$  and  $\Delta n_m^c$ ) in markets with larger populations and lower incomes—precisely the types of markets where dollar store entry is most common. By contrast, the impact of the dispersal policy is weaker in markets with more commercial locations (census tracts), which provide greater scope for dollar store entry given the imposed dispersal constraints.

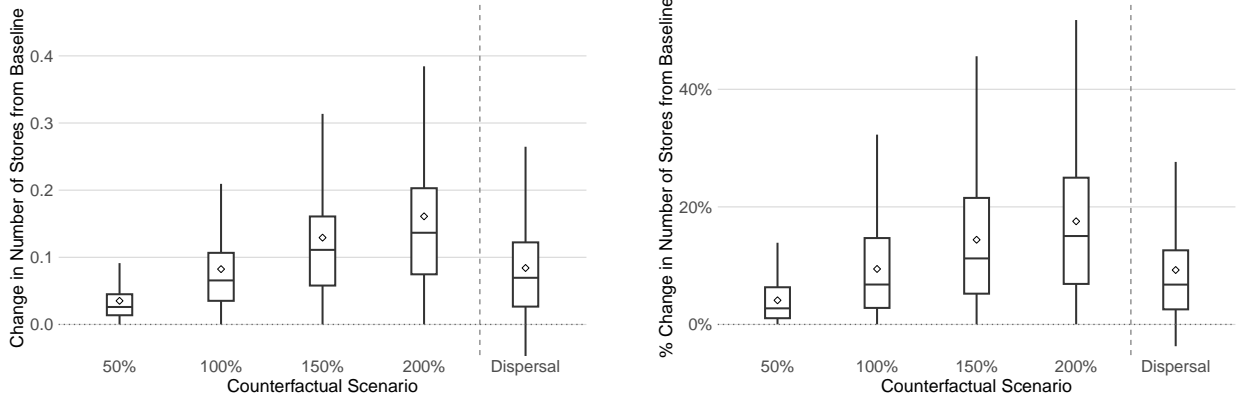
**Location-level analysis.** Next, we direct focus to the impact of the counterfactual policies across locations (census tract) within market. Figure 5 shows boxplots of the change in the number of grocery, convenience, and dollar stores per location ( $\Delta n_l^g$ ,  $\Delta n_l^c$ ,  $\Delta n_l^d$ ) in levels.<sup>39</sup>

As in the market-level analysis, the number of dollar stores decreases monotonically with the level of the entry tax. Grocery and convenience store counts also respond monotonically with the level of the tax. Importantly, some locations experience a reduction in grocery and convenience store counts. Under the entry tax counterfactuals, this reflects the endogenous spatial response of single-store retailers to dollar store expansion: in the baseline, single-store entrants can avoid direct competition by locating in areas less favored by dollar stores

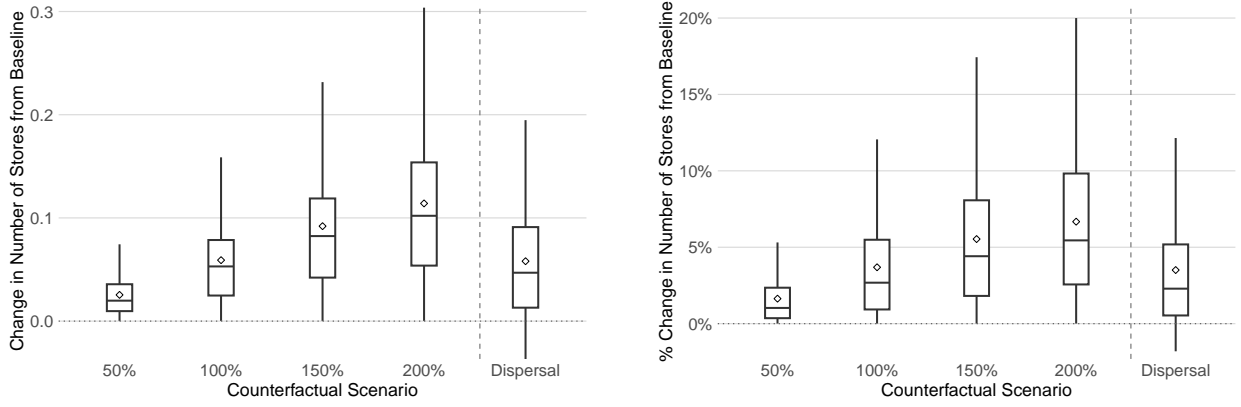
<sup>38</sup>The reduction is not 100% because a significant share of markets have dollar stores already operating in the first period, before entry regulation is imposed.

<sup>39</sup>We do not plot percentage changes—e.g.,  $\frac{\mathbb{E}[\tilde{n}_{lT}^j | \mathbf{n}_{m0}] - \mathbb{E}[n_{lT}^j | \mathbf{n}_{m0}]}{\mathbb{E}[n_{lT}^j | \mathbf{n}_{m0}]}$ —because they can be artificially inflated when the baseline number of stores in location  $l$   $\mathbb{E}[n_{lT}^j | \mathbf{n}_{m0}]$  is close to zero.

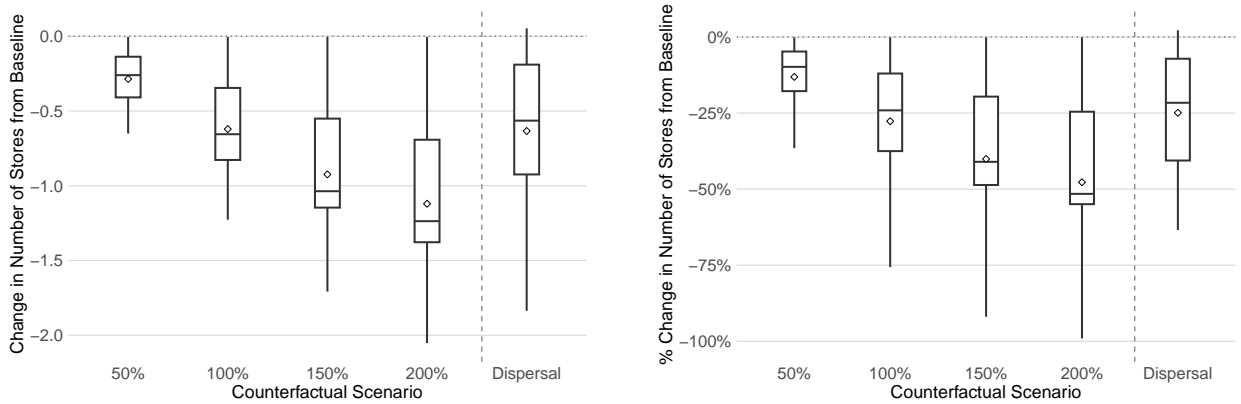
Figure 4: Change in the Expected Number of Stores per Market under Counterfactuals



(a) Grocery Stores



(b) Convenience Stores



(c) Dollar Stores

*Notes:* Each panel shows boxplots of the change in the expected number of stores per market ( $\Delta n_m^g$ ,  $\Delta n_m^c$ ,  $\Delta n_m^d$ ) under counterfactual policies—in levels (left) and percent changes (right). Boxplot interpretation as in Figure 3.



(reallocation); when dollar stores are prevented from entering through the entry tax, this reallocation is undone. Dispersal policies also cause grocery and convenience store counts to decline in 25% of locations, driven by the forced spatial spread of dollar stores into previously unserved areas, where they displace (or discourage entry by) single-store retailers.

We examine how the changes in store counts under these policies vary with location demographic characteristics. We regress  $\Delta n_l^g$  and  $\Delta n_l^c$  on location-level controls: 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 include market fixed effects and control for the number of dollar store in  $t = 0$ . We find that denser, lower-income locations experience a greater increase in the number of single-store retailers: e.g., under the dispersal policy, a one standard deviation increase in population leads to a 31% increase in  $\Delta n_l^g$ , whereas a one standard deviation increase in income per capita leads to an 18% decrease in  $\Delta n_l^g$ .

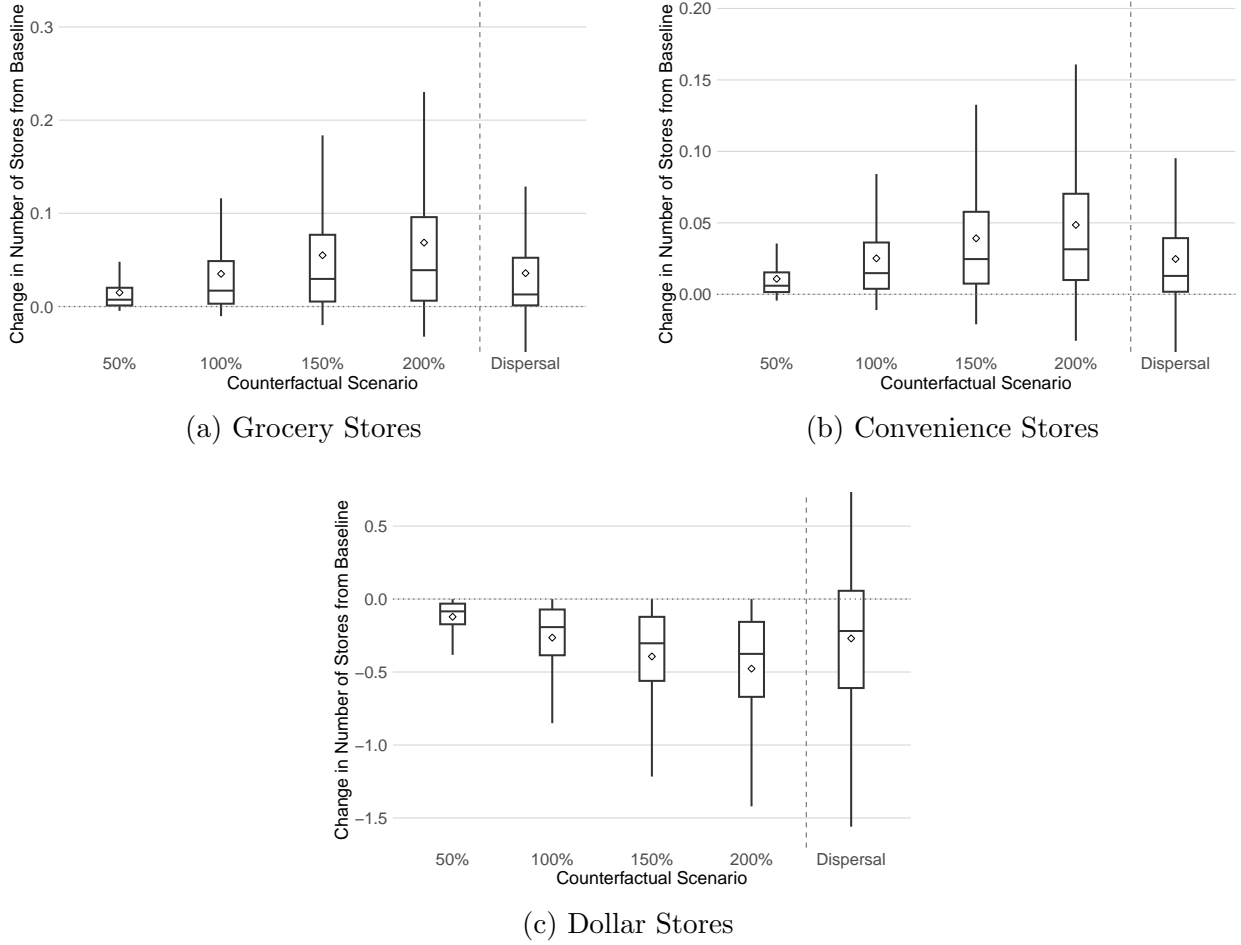
**Reallocation.** To study the potential for spatial spillovers from the dispersal policy, we decompose its market-level impact into direct and indirect effects. The direct effect of the dispersal policy refers to additional entry of grocery and convenience stores generated by the dispersal constraints imposed on dollar stores. The indirect effects involve the exits of grocery and convenience stores due to dollar stores expanding into new unserved locations (within the market) in response to the dispersal constraints.

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 dispersal and baseline scenarios, distinguishing between locations that see an increase or decrease in the number of stores, as follows

$$\begin{aligned}\Delta n_m^j &= \sum_{l=1}^L \Delta n_l^j \\ &= \underbrace{\sum_{l: \Delta n_l^j > 0} \Delta n_l^j}_{\text{direct effects}} + \underbrace{\sum_{l: \Delta n_l^j \leq 0} \Delta n_l^j}_{\text{indirect effects}}\end{aligned}\tag{32}$$

We restrict the analysis below to markets with two to four commercial locations (395 markets

Figure 5: Change in the Expected Number of Stores per Location under Counterfactuals



Notes: Each panel shows boxplots of the change in the expected number of stores per location ( $\Delta n_l^g$ ,  $\Delta n_l^c$ ,  $\Delta n_l^d$ ) under counterfactual policies. Boxplot interpretation as in Figure 3.

out of a total of 846 markets in our sample).<sup>40</sup>

The dispersal policy causes the number of grocery and convenience stores to decrease in some location(s) within 36% of markets. Conversely, we find that 60% of markets experience an increase in the number of dollar stores in some location(s) within the market. These figures are not inconsistent with each other since single-store firms' profits are a function of the number of dollar stores within 2 miles of a focal location, which may still decrease despite there being more dollar stores in the focal location.

To assess the magnitude of these spatial spillovers, we compute the ratios of indirect to

<sup>40</sup>For spillovers or indirect effects to be present, there needs to be more than one location in the market.

direct effects (in absolute value) for each market  $m$ . For grocery and convenience stores ( $j = \{g, c\}$ ), the diversion ratio is given by

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

For dollar stores ( $j = \{d\}$ ), the diversion ratio is given by

$$\left| \frac{\sum_{l:\Delta n_l^d > 0} \Delta n_l^d}{\sum_{l:\Delta n_l^d \leq 0} \Delta n_l^d} \right|.$$

The average diversion ratio for dollar stores equals 34%, that is, for every one-unit *decrease* in dollar-store presence in constrained locations, we observe a 0.34-unit *increase* in unconstrained locations within the market. The diversion ratio is strictly less than one because the unconstrained locations that can be entered after dispersal policies are put in place are arguably less attractive to dollar stores. In the case of single-store firms, we find average diversion ratios of 20% for grocery stores and 27% for convenience stores: i.e., for every one-unit increase in single-store retailers in locations constrained by the dispersal policy, we observe a 0.2 to 0.27 decrease in unconstrained locations within the market.

We compare the characteristics of locations by whether they impose or incur negative spillovers from dispersal policies. We split the sample of locations based on whether  $\Delta n_l^d$  is negative or positive. Locations that see an increase in dollar stores ( $\Delta n_l^d > 0$ ) incur negative spillovers from the counterfactual policy. Table 8 reports average population, income per capita, and the share of locations with high vehicle access (above 89%) and low minority share (below 25%). We find that, in response to the dispersal policy, dollar store entry is reallocated from denser, lower-income areas to less populated, higher-income locations. These receiving areas also exhibit higher rates of vehicle access, consistent with higher socioeconomic status.

This implies that dispersal policies reconfigure the spatial allocation of retail activity in a given market. The aggregate impact of dispersal policies on single-store firms at the market level is, therefore, attenuated once these negative spillovers are accounted for. In

Table 8: Characteristics of Locations with Direct and Indirect Effects

Variable	$\Delta n_l^d \leq 0$		$\Delta n_l^d > 0$		t-stat
	Mean	SE	Mean	SE	
Population	3,502	1,521	2,508	1,789	8.56
Income per Capita	20,639	6,898	22,167	5,900	-3.61
High Access to Vehicle	0.70	0.46	0.85	0.36	-5.51
Low Minority Share	0.77	0.42	0.79	0.41	-0.72

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. 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.<sup>41</sup> Table 9 shows the average distance between consumers and the nearest store, for the different retail formats, under the baseline, the entry tax (with  $\tau = 200\%$ ), and the dispersal policy.

In panel A, we condition on the subset of markets in which there is at least one store format operating in each scenario. The column  $Pr(n > 0)$  gives the population-weighted probability that a market has at least one store of that format operating. The results indicate that grocery stores are present in more markets under the entry tax (82%) than in the baseline scenario (76%), suggesting that dollar store entry is associated with a decline in the prevalence of grocery stores.<sup>42</sup>

<sup>41</sup>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.

<sup>42</sup>While we document changes in retail proximity, entry may have mixed effects on consumers: e.g.,

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 lower (and upper) bound 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 14% under the entry tax: from 2.24 to 2.01 miles (lower bound), or 2.74 to 2.39 miles (upper bound). Given that prior research suggests consumers typically travel no more than 1 to 2 miles for grocery trips, this change represents a notable difference in the average distance to the nearest grocery store across scenarios. Average distance to the nearest dollar store rises—from 1.22 to 1.87 miles in Panel B, and from 1.26 to 2.22 miles in Panel C.

The dispersal policy, which limits dollar store clustering, reduces the average distance to dollar stores (to 1.10 miles) while preserving near-universal coverage ( $Pr(n > 0) = 0.99$ ). Unlike the entry tax, it redistributes dollar stores across locations without significantly affecting the extensive margin of entry. However, the distance to single-store retailers falls less than under the entry tax. By pushing dollar stores into less desirable areas, the dispersal policy may increase competition with single-store retailers, limiting their ability to spatially differentiate and potentially crowding them out in marginal locations.

## 7 Conclusion

This article aims to inform the policy debate around dollar stores by quantifying their impact on retail markets, specifically the number of retail stores by format and their geographic layout. We specify a dynamic model of spatial competition between different store types. The model explicitly incorporates equilibrium effects to assess the net impact of dollar store expansion. The spatial nature of competition introduces non-trivial complexities when estimating and solving this game. Methodologically, we deal with the high-dimensionality

---

entry may also lead to lower prices or competitive responses by incumbents. Assessing the overall welfare implications via a pricing model is beyond the scope of this paper.

Table 9: Predicted retail proximity: distance to nearest store

	Baseline			Entry Tax (200%)			Dispersal		
	Mean	Median	$Pr(n > 0)$	Mean	Median	$Pr(n > 0)$	Mean	Median	$Pr(n > 0)$
<b>Panel A. Conditional on <math>n &gt; 0</math></b>									
<i>Distance to nearest (in miles)</i>									
Grocery store	1.37	0.97	0.76	1.36	0.95	0.82	1.37	0.97	0.79
Convenience store	1.23	0.85	0.94	1.21	0.84	0.95	1.22	0.85	0.94
Dollar store	1.19	0.82	0.99	1.31	0.92	0.83	1.10	0.76	0.99
Any format	1.10	0.74		1.10	0.74		1.05	0.70	
<b>Panel B. Lower bound</b>									
<i>Distance to nearest (in miles)</i>									
Grocery store	2.24	2.22		2.01	1.90		2.12	2.03	
Convenience store	1.44	1.15		1.37	1.07		1.40	1.10	
Dollar store	1.22	0.86		1.87	1.10		1.11	0.78	
Any format	1.11	0.78		1.14	0.79		1.05	0.73	
<b>Panel C. Upper bound</b>									
<i>Distance to nearest (in miles)</i>									
Grocery store	2.74	2.71		2.39	2.28		2.56	2.44	
Convenience store	1.58	1.21		1.48	1.13		1.53	1.16	
Dollar store	1.26	0.88		2.22	1.10		1.13	0.78	
Any format	1.14	0.79		1.17	0.80		1.07	0.74	

*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. The bottom two 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).*

of the firms' problem by extending the ECCP estimator of Kalouptsi et al. (2020) from single-agent problems to games with finite dependence.

Our 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.

We use the estimated model to show that entry taxes are more effective at halting expansion than dispersal policies. These policies increase the number of grocery and convenience stores, especially in lower-income and densely populated markets. However, effects vary across locations: the dispersal policy pushes dollar stores into previously unserved areas, which can crowd out independent retailers and reduce their ability to spatially differentiate. A significant share of the gains in grocery and convenience store counts are offset by losses elsewhere in the market. These findings highlight the importance of accounting for spatial

spillovers when regulating entry in retail markets.

The full 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.
- Berry, Steven T., and Giovanni Compiani.** 2023. “An Instrumental Variable Approach to Dynamic Models.” *Review of Economic Studies* 90 (4): 1724–1758.
- 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.** 2024. “Welfare Implications of Increased Retailer Participation in SNAP.”
- 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.** 2024. “Dynamic Model of Entry: Dollar Stores.” *American Journal of Agricultural Economics*.
- Chi, Feng.** 2025. “Information Waves and Firm Investment.”
- 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.
- Ellickson, Paul B., and Sanjog Misra.** 2012. “Enriching interactions: Incorporating outcome data into static discrete games.” *Quantitative Marketing and Economics* 10 (1): 1–26.
- 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.
- Geweke, John.** 1989. “Bayesian Inference in Econometric Models Using Monte Carlo Integration.” *Econometrica* 57 (6): 1317–1339.
- 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.
- Imai, Susumu, Neelam Jain, and Andrew T. Ching.** 2009. “Bayesian Estimation of Dynamic Discrete Choice Models.” *Econometrica* 77 (6): 1865–1899.
- Ishihara, Masakazu, and Andrew T. Ching.** 2016. “Bayesian Estimation of Finite-Horizon Discrete Choice Dynamic Programming Models.” *SSRN Electronic Journal*.
- 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.
- Kalouptside, Myrto, Yuichi Kitamura, Eduardo Souza-Rodrigues, and Lucas Lima.** 2021a. “Counterfactual Analysis for Structural Dynamic Discrete Choice Models.” (June): .
- Kalouptside, Myrto, Paul Scott, and Eduardo Souza-Rodrigues.** 2021b. “Identification of Counterfactuals in Dynamic Discrete Choice Models.” *Quantitative Economics*.
- Kalouptside, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues.** 2017. “On the non-identification of counterfactuals in dynamic discrete games.” *International Journal of Industrial Organization* 50 362–371.
- Kalouptside, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues.** 2020. “Linear IV regression estimators for structural dynamic discrete choice models.” *Journal of Econometrics*.
- Kasahara, Hiroyuki, and Katsumi Shimotsu.** 2009. “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices.” *Econometrica* 77 (1): 135–175.
- Kloek, T., and H. K. van Dijk.** 1978. “Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo.” *Econometrica* 46 (1): 1–19.
- 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.
- Pakes, Ariel, and Paul McGuire.** 2001. “Stochastic algorithms, symmetric markov perfect equilibrium, and the ‘curse’ of dimensionality.” *Econometrica* 69 (5): 1261–1281.

- Pancras, Joseph, S. Sriram, and V. Kumar.** 2012. “Empirical investigation of retail expansion and cannibalization in a dynamic environment.” *Management Science* 58 (11): 2001–2018.
- 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.
- Schivardi, Fabiano, and Eliana Viviano.** 2010. “Entry Barriers in Retail Trade.” *The Economic Journal* 121 (551): 145–170.
- Schneier, Camilla.** 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.
- Suzuki, Junichi.** 2013. “Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry.” *International Economic Review* 54 (2): 495–523.
- 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*.
- 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.

# Online Appendix for “Dynamic Entry & Spatial Competition: An Application to Dollar Store Expansion”

## A Estimation Details

### A.1 Estimation Approach for Single-Store Firms and Multi-Store Incumbents

We derive estimating equations for single-store firms and chain incumbents. The use of finite-dependence and rational expectations follows the same motivation as the case of chain entrants presented in Section 4.2.1.

*Single-store firms.* 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

$$v_{i,t}^{\mathbf{P}}(a_{it} = 0, \mathcal{M}_{j,i,t}) = 0 \quad (33)$$

$$v_{i,t}^{\mathbf{P}}(a_{it} = l_+, \mathcal{M}_{j,i,t}) = -\theta_i^{EC} + \beta \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1}))] \quad (34)$$

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

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = -\theta_i^{EC} + \beta \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1}))] \quad (35)$$

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

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \ln \left( \frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})} \right) \quad (36)$$

Combining Equation (35) and Equation (36), we obtain an optimality condition that involves only CCPs at  $t$  and  $t+1$  and the single-period payoff function

$$\ln \left( \frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})} \right) = -\theta_i^{EC} + \beta \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1}))] \quad (37)$$

This 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 realized variables minus an expectational error that is orthogonal to the state variables at period  $t$ . Therefore, for any function of period- $t$  information set  $h(\mathcal{M}_{j,i,t})$ , we have

$$\mathbb{E}[h(\mathcal{M}_{j,i,t})u_{it}] = 0 \quad (38)$$

where  $u_{it}$  is the expectational error defined, for any realization  $\mathcal{M}_{j,i,t+1}^*$  of the random

variable  $\mathcal{M}_{j,i,t+1}$ , as

$$\begin{aligned}
u_{it} &= \beta \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})) \\
&\quad - \beta (vp_{i,l}(\mathcal{M}_{j,i,t+1}^*) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1}^*)) \\
&= \ln \left( \frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})} \right) + \theta_i^{EC} \\
&\quad - \beta (vp_{i,l}(\mathcal{M}_{j,i,t+1}^*) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1}^*))
\end{aligned} \tag{39}$$

where the second equation is obtained by using Equation (37) to eliminate the expectation term.

Kalouptsidei et al. (2020) show that, under linearity of payoffs, these moment conditions (replaced by their sample counterparts) can form the basis of a linear regression model. Define the left-hand side variable by collecting all known terms that depend on the CCPs

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

We can obtain the structural parameters for single-store firms by rewriting Equation (39) as

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

where regressors entering the variable profit function  $vp_{i,l}(\mathcal{M}_{j,i,t+1})$  in  $t+1$  (population, income, etc.) are instrumented using the values of these regressors in period  $t$ .<sup>43</sup>

For single-store incumbents, one can construct a similar regression model by taking the difference between the choice-specific value functions of staying active ( $a_{it} = 0$ ) or exiting ( $a_{it} = l_-$ )

$$v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(l_-, \mathcal{M}_{j,i,t}) = \beta \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1}))] \tag{40}$$

which yields the regression equation

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

where the left-hand side variable is defined as

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

*Multi-store incumbents.* The derivation of estimating equations for chain incumbents follows similar steps as in the case presented in the main text for chain entrants (Section 4.2.1). 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  $\tilde{\theta}_i^{EC}$  to be different than for that of the first store  $\theta_i^{EC}$ ). The corresponding

---

<sup>43</sup>The endogeneity problem occurs because  $t+1$  covariates may correlate with expectational error  $u_{it}$ . One can use instruments in the contemporaneous information set at  $t$  for these next-period covariates.

choice-specific value functions are given by

$$v_{i,t}^{\mathbf{P}}(l^*, \mathcal{M}_{j,i,t}) = vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \quad (41)$$

$$\begin{aligned} v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) &= vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \\ &+ \beta \mathbb{E}[vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln P_{i,t+1}(l^* | \mathcal{M}_{j,i,t+1}) | \mathcal{M}_{j,i,t}, 0] \end{aligned} \quad (42)$$

$$\begin{aligned} v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) &= vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i - \tilde{\theta}_i^{EC} \\ &+ \beta \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) + vp_{i,l^*}(\mathcal{M}_{j,i,t+1}) - FC_i \\ &+ \gamma - \ln P_{i,t+1}(l_- | \mathcal{M}_{j,i,t+1}) | \mathcal{M}_{j,i,t}, l_+] \\ &+ \beta^2 \mathbb{E}[vp_{i,l^*}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l^* | \mathcal{M}_{j,i,t+1}) | \mathcal{M}_{j,i,t}, l_+] \end{aligned} \quad (43)$$

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

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

$$v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(l^*, \mathcal{M}_{j,i,t}) = \log \left( \frac{P_{i,t}(0 | \mathcal{M}_{j,i,t})}{P_{i,t}(l^* | \mathcal{M}_{j,i,t})} \right) \quad (45)$$

As with chain entrants, we can dispose of the expectations in Equations (41) to (43) using the rational expectation assumption and derive moment conditions. As before, period  $t + 2$  states, appearing in Equation (43), are *conditional* on the action  $a_{it} = l_+$ . However, the empirical distribution of  $\mathcal{M}_{j,i,t+2}$  (in particular the vector  $\mathbf{n}_{-i,t+2}$ ) 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 equations, any term involving  $\mathcal{M}_{j,i,t+2}$  is scaled by the importance weights  $\psi_{l_+, \tilde{a}}(\mathbf{n}_{-i,t+2} | \mathcal{M}_{j,i,t})$  (Lemma 1), where  $\tilde{a}$  is the action played in the data.

## A.2 Alternative Estimation Approach

The baseline estimation approach extends the ECCP estimator of Kalouptsi 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

$$\bar{V}_{i,t}^{\mathbf{P}}(\mathcal{M}_{j,i,t}) \approx \sum_{k=1}^K \lambda_k \phi_{ki}(\mathcal{M}_{j,i,t}) \quad (46)$$

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.<sup>44</sup> We allow the coefficients  $\lambda_k$  to differ by firm type (e.g., convenience, grocery, dollar store) and by incumbency status.<sup>45</sup>

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

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,

$$\mathbf{V}_i = \boldsymbol{\lambda}' \boldsymbol{\phi}_i = \sum_a \mathbf{P}_i^{(k)}(a) \left[ \boldsymbol{\pi}_i^{(k)}(a) + \mathbb{E}[\epsilon|a] \right] + \beta \boldsymbol{\lambda}' \mathbb{E}[\boldsymbol{\phi}_i] \quad (47)$$

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

$$\boldsymbol{\lambda}' (\boldsymbol{\phi}_i - \beta \mathbb{E}[\boldsymbol{\phi}_i]) = \sum_a \mathbf{P}_i^{(k)}(a) \left[ \boldsymbol{\pi}_i^{(k)}(a) + \mathbb{E}[\epsilon|a] \right] \quad (48)$$

$\boldsymbol{\lambda}$  can be found by an OLS regression. To calculate  $\mathbb{E}[\boldsymbol{\phi}_i]$ , 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

---

<sup>44</sup>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.

<sup>45</sup>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

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

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

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  $\boldsymbol{\theta}$

$$\mathbf{v}_i^{(k)}(a|\boldsymbol{\theta}) = \boldsymbol{\pi}_i(a|\boldsymbol{\theta}) + \beta \widehat{\boldsymbol{\lambda}}' \mathbb{E}[\boldsymbol{\phi}_i|a] \quad (49)$$

where  $\boldsymbol{\pi}_i(a|\boldsymbol{\theta})$  are current-period profits parameterized as a function of candidate parameter  $\boldsymbol{\theta}$ , and  $\mathbb{E}[\boldsymbol{\phi}_i|a]$  is the expected future value of the approximating functions given action  $a$  is played, that is,  $\mathbb{E}[\boldsymbol{\phi}_{ki}(\mathcal{M}_{j,i,t+1})|a, \mathcal{M}_{j,i,t+1}]$ , 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,

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

Denote  $\boldsymbol{\theta}^{(k+1)}$  the updated structural parameters.<sup>46</sup>

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

$$\mathbf{P}_i^{(k+1)}(a) = \frac{\exp \left( \mathbf{v}_i^{(k)}(a|\boldsymbol{\theta}^{(k+1)}) \right)}{\sum_{\tilde{a}} \exp \left( \mathbf{v}_i^{(k)}(\tilde{a}|\boldsymbol{\theta}^{(k+1)}) \right)}. \quad (50)$$

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

---

<sup>46</sup>We do not attempt to calculate and use the optimal weight matrix derived in Bugni and Bunting (2021), 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.



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.<sup>47</sup>

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.

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

Parameters	ECCP with finite dependence						1-MD with value function approximation					
	Chain		Grocery		Conv.		Chain		Grocery		Conv.	
	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)
<i>Location-level characteristics</i>												
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)
<i>Fixed cost components</i>												
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)				
<i>Measures of competition and cannibalization</i>												
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)				
<i>Market-level characteristics</i>												
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)
<i>Dynamic investment costs</i>												
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.

### A.3 Robustness Checks

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

<sup>47</sup>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.

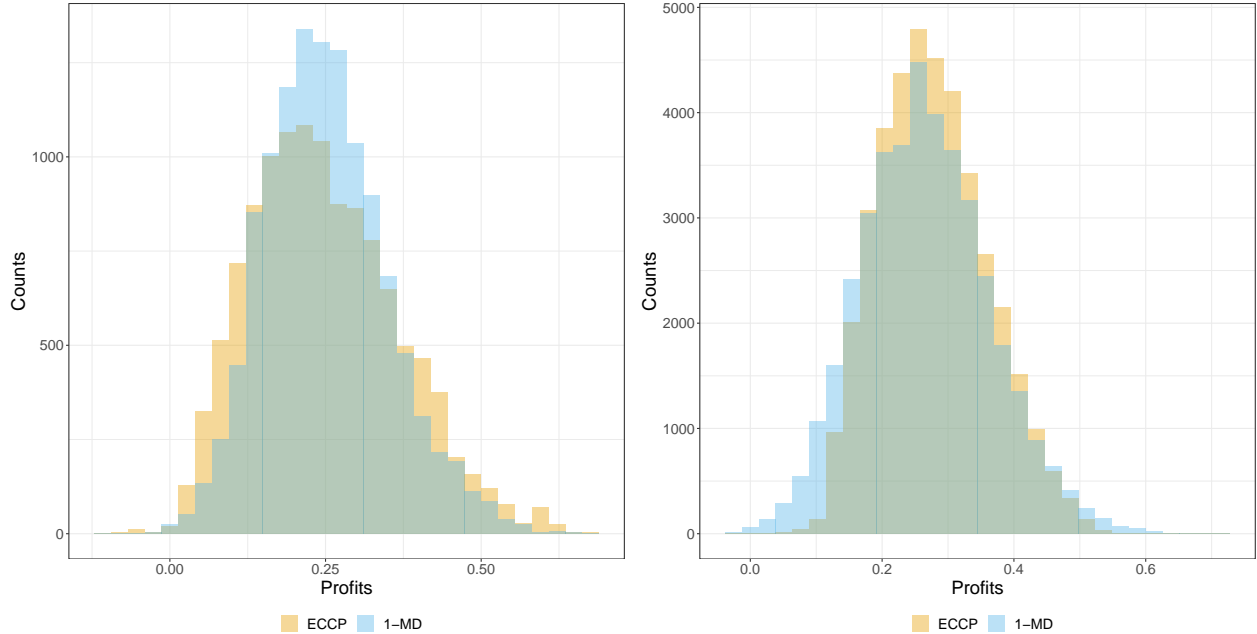


Figure A1: These figures show estimated profits for grocery stores (left) and convenience stores (right) under our baseline estimator (ECCP with finite-dependence) and the alternative estimator (minimum-distance with value function approximation).

*Discount factor.* The main estimates use an annual discount factor  $\beta$  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 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.

## A.4 Solution Method for the Dynamic Game

This section provides an 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 adopt two simplifying assumptions to maintain tractability.

First, 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  $T$ .<sup>48</sup> We also assume that chains' distribution networks remain fixed at their period- $T$  configurations for  $t \geq T + 1$ .

Second, with many single-store retailers and three chain players, the number of industry states in multi-location markets is typically too large to store in memory. To address this issue, we restrict the strategy space for dollar store chains by assuming that each chain may open at most one store over the game horizon  $T$ —in addition to any stores they were operating in the first period—and that chains do not close stores. This simplification is broadly in line with the empirical frequency of entry and exit observed in the data (see Table 4) and reflects the nature of the counterfactuals we study, which primarily constrain dollar store expansion.

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

1. Initialize the vectors of CCPs for each firm and state  $\mathbf{P}_{i,t}$ . If firm  $i$  is a potential entrant,  $\mathbf{P}_{i,t}$  is a vector indexed by the state and locations  $(\mathcal{M}_{i,j,t}, l_+)$  giving the CCP of entry into location  $l$  in state  $\mathcal{M}_{i,j,t}$ . If  $i$  is an incumbent,  $\mathbf{P}_{i,t}$  is a vector indexed by  $(\mathcal{M}_{i,j,t}, a_{it})$  giving the CCP of choosing action  $a_{it}$  (remaining active/exiting for single-store firms, or building an additional store/doing nothing for chains) in state  $\mathcal{M}_{i,j,t}$ .

---

<sup>48</sup>To compute this expectation, we assume that demographic variables in each location evolve according to AR-1 processes, allowing for geographically correlated innovation shocks within each market.

2. Form the transition matrix from state  $\mathcal{M}_{i,j,t}$  to state  $\mathcal{M}_{i,k,t+1}$  for each firm type, *conditional* on the action played  $a$ . Denote this transition matrix  $\mathbf{F}_{i,t}(a)$ . 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. Solve for dollar store chains' ex-ante value function. For period  $t \leq T$ , the ex-ante value function is obtained recursively, given period- $t + 1$  ex-ante value function, as

$$\mathbf{V}_{i,t} = \sum_a \mathbf{P}_{i,t}(a) [\boldsymbol{\pi}_{i,t}(a) + \mathbb{E}[\epsilon_{it}|a] + \beta \mathbf{F}_{i,t}(a) \mathbf{V}_{i,t+1}] . \quad (51)$$

where  $\boldsymbol{\pi}_{i,t}(a)$  are single-period profits under action  $a$ , and  $\mathbb{E}[\epsilon_{it}|a]$  is the expectation of the firm-specific shock conditional on action  $a$  being played. For period  $t \geq T + 1$ , the ex-ante value function solves the system of equations

$$\mathbf{V}_{i,T+1} = \left( I - \beta \sum_a \mathbf{P}_{i,T+1}(a) \mathbf{F}_{i,T+1}(a) \right)^{-1} \sum_a \mathbf{P}_{i,T+1}(a) (\boldsymbol{\pi}_{i,T+1}(a) + \mathbb{E}[\epsilon_{iT+1}|a]) , \quad (52)$$

where  $I$  is the identity matrix.

4. Update the conditional choice-specific value function. Let  $\mathbf{v}_{i,t}(a)$  denote a vector collecting the choice-specific value function of firm  $i$  if it plays action  $a$  for all states  $(\mathcal{M}_{i,j,t})$ . For single-store retailers, we leverage their terminal action. The choice-specific value function satisfies the equality (in matrix form)

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

$\mathbf{v}_{i,\text{exit}}$  is only a function of the single-period profits.

For dollar store chains, given period- $t + 1$  ex-ante value function, we update the choice-specific value function in period  $t$  as follows

$$\mathbf{v}_{i,t}(a) = \boldsymbol{\pi}_{i,t}(a) + \beta \mathbf{F}_{i,t}(a) \mathbf{V}_{i,t+1}. \quad (54)$$

5. Update the vectors of CCPs as

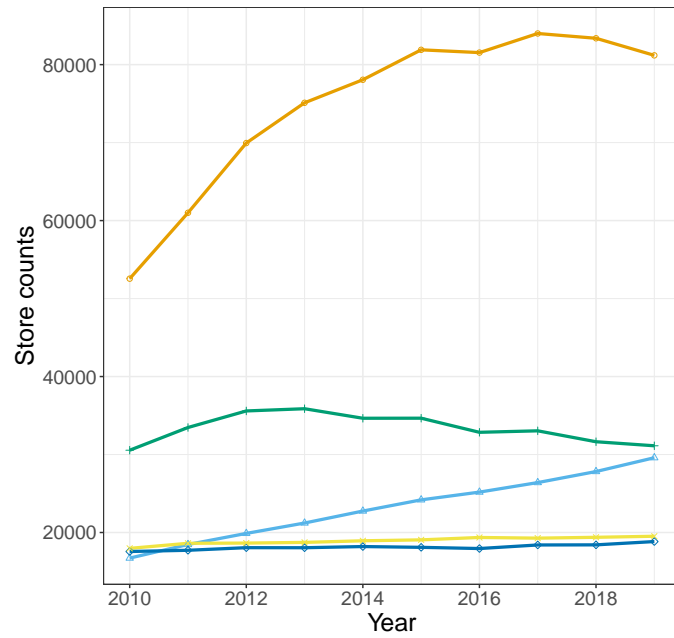
$$\mathbf{P}'_{i,t}(a) = \frac{\exp(\mathbf{v}_{i,t}(a))}{\sum_{\tilde{a}} \exp(\mathbf{v}_{i,t}(\tilde{a}))}. \quad (55)$$

If the maximum absolute difference between  $\mathbf{P}_t$  and  $\mathbf{P}'_t$  is less than the pre-defined tolerance level, the procedure stops: the equilibrium CCPs  $\mathbf{P}'_t$  and dollar store chains' ex-ante value function  $\mathbf{V}_{i,t}$  are saved, and we proceed backwards to period  $t - 1$ . If not, define updated CCPs as a convex combination of old and new CCPs  $\alpha \mathbf{P}_{i,t} + (1 - \alpha) \mathbf{P}'_{i,t}$  for each player  $i$  and return to Step 2.

As markets are independent, we solve the model for each market separately.

## B Supplementary Tables and Figures

Figure A2: Store counts by firm type



— Convenience — Dollar — Grocery — Supermarket — Supercenter

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

Figure A3: Distribution center locations in 2019

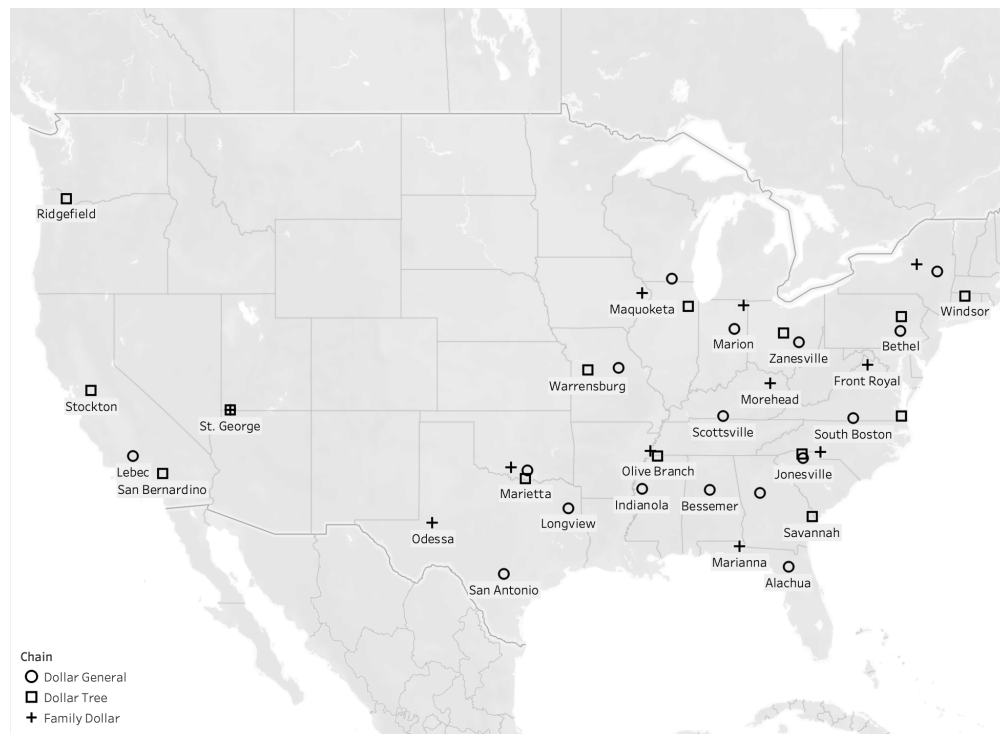


Table A2: Multinomial logit of single-store firms' choice

	Dependent variable: Firm enters or remains active in location $l$			
	Grocery (1)	Grocery (2)	Convenience (3)	Convenience (4)
Entrant	-0.613 (1.016)	-6.208 (1.116)	-0.407 (0.801)	-4.786 (0.719)
Incumbent	5.062 (1.014)	-0.673 (1.121)	5.538 (0.807)	1.101 (0.722)
<i>Location-level characteristics</i>				
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)
<i>Cost shifters</i>				
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)
<i>Measures of competition</i>				
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)
<i>Market-level characteristics</i>				
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

Parameters	Chains			
	(1)		(2)	
	Estimate	s.e.	Estimate	s.e.
Constant	2.616	(0.444)	2.318	(0.449)
<i>Location-level characteristics</i>				
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)
<i>Fixed cost components</i>				
Median residential rent	-0.072	(0.056)	-0.051	(0.057)
Distance to own distribution center	-0.058	(0.020)	-0.021	(0.021)
<i>Measures of competition and cannibalization</i>				
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)
<i>Market-level characteristics</i>				
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)
<i>Dynamic investment costs</i>				
Entry cost for first store in market	2.495	(0.240)	2.316	(0.242)
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.*