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

El Hadi Caoui; University of Toronto, elhadi.caoui@rotman.utoronto.ca
Brett Hollenbeck, UCLA Anderson, : brett.hollenbeck@anderson.ucla.edu
Matthew Osborne, University of Toronto, matthew.osborne@rotman.utoronto.ca

Invited Paper prepared for presentation at the **2023 AEA/ASSA Annual Meeting, January 6-8, 2023, New Orleans, LA.**

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

El Hadi Caoui — U of Toronto
Brett Hollenbeck — UCLA Anderson
Matthew Osborne — U of Toronto

October 27, 2022*

Abstract

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

Keywords: retail industry, dollar store, food access, dynamic games

JEL Classification: L13, D43, L51

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

We are grateful to Emek Basker, Paul Ellickson, and Mathieu Marcoux for helpful discussions of this paper. We thank Victor Aguirregabiria, Limin Fang, Paul Grieco, Sylvia Hristakeva, Yao Luo, and seminar and conference participants at the University of Toronto, UCLA, Rochester, UVA, IIOC, SICS, CEA, ISMS Marketing Science Conference, the Food Access at Dollar Stores (FADS) Workshop for their helpful comments and suggestions. The findings and conclusions in this paper have not been formally disseminated by the U.S. Department of Agriculture (USDA) and should not be construed to represent any agency determination or policy. The analysis, findings, and conclusions expressed in this paper also should not be attributed to Information Resources, Inc. (IRI). This research was conducted in collaboration with USDA under a Third Party Agreement with IRI.

1 Introduction

The dollar store retail format has grown rapidly over the past several decades, with broad and dramatic effects on the retail landscape in the United States. Following previous waves of growth by large retail chains, the three main dollar store chains (Dollar General, Dollar Tree, and Family Dollar) have become in some ways the dominant mode of retailer in many markets, with implications for competition, affordability, convenience, and food accessibility for much of the population.

The distinguishing features of these chains are the use of single or limited price points, particularly selling most goods for $1, and assortments consisting of small serving-size basic consumables, clearance or irregular goods, and a lack of fresh produce. Beginning in the 1950's, these chains exhibited slow but steady growth over the decades that followed, establishing themselves primarily in small towns in rural areas. Following the 2008 recession, several events combined to accelerate their growth. The recession itself may have made low-price, small-format consumables more attractive for many by worsening household finances. In 2007, the largest chain, Dollar General, was bought out by a private equity firm who rationalized location strategy, cut costs, and set out on a rapid growth strategy. In 2015, the two smaller chains, Dollar Tree and Family Dollar, merged and sought to expand their presence to compete with Dollar General.

Consequently, the growth of this format has been and continues to be exceptionally rapid.[1] In 2021, there were more of these stores operating than all the Walmarts, CVS, Walgreens, and Targets combined by a large margin. During the period 2018-2021, roughly half of all retail stores that opened in the U.S. were dollar stores. The growth of dollar store chains has raised a number of policy issues, in particular, many local policymakers have expressed concerns that the rapid entry of dollar store chains in their cities have forced out local independent retailers, including small grocery stores. The latter may be especially concerning to the extent it reduces access to produce and other perishable food items for low income residents, creating "food deserts." These concerns and others have lead many localities to ban dollar store chains from entering or pass dispersal regulations limiting their store density.[2]

Broadly speaking, the arguments for and against dollar store chains fall along these lines. Proponents argue that they introduce additional choice into under-served retail markets,

[1] The top three chains were collectively opening stores at the rate of 3.75 stores a day over the past decade (authors' calculations).

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

that they offer lower prices than their competitors, and that their strategy of entering in low-rent areas and opening many stores in the same market results in greater convenience for customers who can make short trips for specific items rather than long trips to the nearest big box store, which could be a large distance away.

Opponents maintain that the aggressive entry strategy of dollar store chains has forced out local independent retailers and reduced access to fresh food products as a result. Even if consumers value the convenience of the dollar stores, their strategy of offering basic consumables and household products efficiently using a low fixed cost model and avoiding the costly and complex provision of perishable food results in them capturing a large share of the profits of nearby grocers. This leaves grocers unprofitable on the basis of food sales alone and can lead to exit. Opponents also argue that their pricing and small-item product format masks the extent to which their prices and margins are in many cases higher than other retailers who sell larger volume products with quantity discounts. Yet despite the widespread public and policy debate, the academic literature has yet to study these claims or the effects of dollar store chain expansion more generally.

Our research objectives are: first, to document the extent and nature of growth of dollar store chains and place this in the context of the broader retail landscape and the economic study of retail chains' growth over the past decades. Second, we measure the effect of dollar store entry in a location on the number of independent retailers and small grocers. Third, we use detailed data on consumer purchases to measure the effects of dollar store entry on expenditures, prices paid, convenience and travel costs, and food access of nearby households, and to characterize the distributional effects across household types. Fourth, we evaluate what effects the anti-dollar store policies implemented by some localities would have if implemented more broadly: in particular, we study proposed bans on dollar store expansion to understand how they would affect the broader retail landscape and retail proximity.

We use a combination of methods to answer these questions. We begin by compiling multiple datasets to comprehensively document dollar store chain growth. We use reduced form event study methods to test hypothesized effects of dollar store entry on local market structure and consumer behavior. To understand why the dollar store format has been so successful and how they compete against rivals, we estimate a dynamic model of store entry, expansion and exit choices to measure entry costs, density economies, and competition and cannibalization effects. We then use the model estimates to solve and simulate the dynamic oligopoly game played between retailers to evaluate proposed policies and quantify the long-term equilibrium effects of dollar store expansion.

We leverage data from several sources. We track the number and type of retail stores, including dollar stores, across the U.S. using the Supplemental Nutrition Assistance Program

(SNAP) Retailer panel, a yearly panel of SNAP-authorized retailers from 2008 to 2019. An advantage of this dataset is that it covers small independent retail stores, which are typically absent from other retail census used in the literature. We combine this with data on dollar store distribution center locations and opening dates, as well as data from Nielsen TDLinx on average per-store revenue. Finally, we match the data on store openings and closings to the IRI Consumer Network panel, which contains household-level data on all retail purchases for a large nationwide sample of consumers.

We first use static and event study designs to study the effects of dollar store entries on local market structure. We control for market-level demographics and demographic trends, as well as market-year fixed effects, and consistently find that dollar store entries are associated with a significant decrease in the number of independent local grocery stores. The effect size is roughly the loss of one grocery store for every three dollar stores, when measured in the area 0-2 miles around the entry location. In the area 2 to 5 miles away, the effect size becomes substantially smaller, suggesting the effects are local.

When studying how consumer shopping behavior changes when dollar stores enter nearby, we find that consumers shift a share of their purchases away from grocery and convenience stores to the dollar stores. Consumers also decrease their spending (and volumes purchased) on produce by modest but significant amounts (4% to 7.3%). The reduction in spending due to dollar store entry explains 9% to 16% of the difference in fresh produce spending between the top and bottom quartiles of the income distribution. This reduction in produce spending grows as the number of dollar stores nearby increases, and is larger for low-income consumers.

Next, we estimate a structural model of the dynamic game played by dollar store chains and their local competitors. The goal of this model is twofold: to provide estimates of the size of dynamic entry and investment costs and the competitive effects between different store formats; and to evaluate how spatial market structure and retail proximity would evolve under counterfactual entry regulation policies. The latter allows us to quantify the cumulative reduction in the number of grocery stores, convenience stores, etc., that has resulted from dollar store chain expansion across different market types. Because the effects identified above are spatial in nature and cumulative, modeling the long-term market structure impacts are necessary to understand the net effects of dollar store expansion on consumers and local firms. We model each store type's entry and exit decisions as a dynamic oligopoly game following Ericson and Pakes (1995) but with a spatial component along the lines of Seim (2006). The key challenges in modeling this game in a tractable way are the complex nature of spatial competition and the dynamics that result from the growth over time in dollar stores' distribution networks, which reduces the fixed costs of operating a given store.

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

Our estimation results suggest that dollar store chains have substantially lower costs of opening a new store than their independent rivals. They are also substantially more profitable and grocery store profits are significantly harmed by the presence of nearby dollar stores and convenience stores, with most of the effects for stores located within 0-2mi. Estimates also point to strong demand cannibalization within chains in the 0-2mi radius, but chains tend to benefit from scale economies when locating stores in moderate proximity (2-5mi radius), likely working through lower operating costs.

We use these estimates to evaluate the impact on local market structure of a hypothetical ban on dollar store expansion beginning in 2010. We find that, in the counterfactual scenario, markets have on average more than 50% more independent stores, with on average 1.65 grocery stores per market as compared to just over 1. This includes 54% more grocery stores and 47% more convenience stores. We also uncover the distributional effects of these changes by controlling for market heterogeneity. We find that the largest impacts on the number of grocery stores materialize in lower income markets, those with larger shares of minority populations, those with higher poverty rates, and fewer households with access to a vehicle.

Related Literature. This paper contributes to three lines of research in economics. The first is the study of the evolution of the U.S. discount retail sector. This literature has focused on the impact of big box retailers (e.g., Walmart, K-Mart) and the supercenter format on market structure and competition (Jia (2008), Zhu and Singh (2009), Basker and Noel (2009), Ellickson and Grieco (2013), Grieco (2014)), on labor markets (Basker (2005)), and the role of chain and density economies (Holmes (2011), Ellickson et al. (2013)). A key finding in this literature is that Walmart's entry primarily harmed larger chain retailers in the local area within two miles of entry (Ellickson and Grieco (2013)). Small, local retailers were not substantially harmed due to travel costs and horizontal and vertical differentiation between firms. Our results suggest that, whereas small local retailers were not in direct competition with big-box retailers like Walmart, the dollar store format is more of a direct competitor to these stores and has had a large impact on local retail markets as a result. The rise of the dollar store format, therefore, has a distinctive impact on retail and raises unique policy questions.

Second, this paper is related to the study of consumers' grocery shopping behavior and food accessibility. There is an extensive literature studying nutritional inequality in the U.S., with studies focusing on price-, access-, and nutrition education-based interventions (Levi et al. (2019)). Studies of food access have focused on introduction of grocery stores to markets designated as "food deserts," with case studies around individual store entries having found mixed results (Cummins et al. (2005), Cummins et al. (2014), Elbel et al. (2015), Dubowitz et al. (2015), Liese et al. (2014), Rose and Richards (2004), Ver Ploeg and Rahkovsky (2016), Weatherspoon et al. (2013).) Notably, Allcott et al. (2019) study a large number of grocery store entries and find they have only small effects on nutrition of nearby consumers and that nutritional inequality in the U.S. is largely explained by demand factors rather than limited food access, with differences in access and prices explaining only about 10% of nutritional inequality. Levi et al. (2020) find that access to grocery stores impacts fruit and vegetable spending by affecting shopping frequency, but only among households with a low value of nutrition and at distances of less than 1 mile. Hristakeva and Levine (2022) leverage hurricane-induced temporary grocery store closures to detect supply-side effects and find that households affected by these closures shift the location and nutritional value of their purchases for a substantial period of time even after the grocery has reopened. Closer to our study, Chenarides et al. (2021) find that dollar store entry in a food desert area increases the likelihood that it remains without access to a supermarket.

We contribute to this literature by studying large numbers of dollar store and grocery store entry events, as well as large numbers of grocery store exits, to measure the impact of these events on a range of outcomes including spending on produce and frequency of shopping trips. While previous research has shown that entry of a grocery store or supermarket in a "food desert" has a limited impact on shopping behavior, we find that the exit of existing grocery stores and the entry of large numbers of dollar stores do have significant impacts on food purchases and availability. If consumers shopping behavior is characterized by inertia, as suggested by the store choice literature in marketing, this could generate asymmetric effects between grocery store openings and closings.

Finally, this paper is related to the literature using dynamic games to study the market structure impacts of retail chains (Arcidiacono et al. (2016), Zheng (2016), Igami and Yang (2016), Hollenbeck (2017), Beresteanu et al. (2019), Fang and Yang (2022)) We depart from the existing literature in two ways. First, we account for the fact that over the sample period, dollar store chains have been growing their networks of distribution centers; incorporating this dynamic aspect of the industry is clearly important to better match observed entry patterns. Second, most of previous dynamic game studies (Zheng (2016) being an exception) abstract from the spatial nature of retail competition. Because retail location choices are crucial in shaping the competitive environment (Ellickson et al. (2020)), we model firms' entry decisions into spatially differentiated location as in Seim (2006).

The rest of the paper proceeds as follows: Section 2 describes the data and institutional details, and provides descriptive statistics. Section 3 presents reduced-form static and dynamic event study results for the impact of dollar store entry on local market structure and consumers' shopping behavior. Section 4 introduces the dynamic entry and exit game. Section 5 motivates the market definition used for the structural analysis. Section 6 studies the model identification, estimation, and shows the estimation results. Section 7 presents the results of the counterfactual analysis. Section 8 concludes.

2 Industry Background, Data, and Descriptive Statistics

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

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

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

The dollar store chains have grown rapidly over the past several decades, and particularly so after the recession of 2008. By 2021, Family Dollar operated roughly 7,100 stores, Dollar General operated 18,000 stores, and Dollar Tree owned 4,350 stores. The combined nearly 30,000 stores are substantially more than the number of Wal-Marts (5,300 stores), Targets (1,900 stores), CVS (9,900 stores), and Walgreens (9,300 stores) combined and is significantly larger than the number of Subway restaurants (21,000 restaurants), the largest U.S. restaurant chain and is similar to the number of worldwide Starbucks locations. The three chains earned a combined $47 billion in revenue in 2019.

In 2015, the two smaller chains, Dollar Tree and Family Dollar, merged citing several potential complementarities between the two businesses: e.g., targeting broader ranges of customers, optimizing their combined real estate portfolio, exploiting synergies in sourcing, procurement, and distribution networks.[3] Nonetheless, integration of the two chains has been slow over the following years: the two chains have continued to be run independently, both from a store operations and supply chain perspective. For example, store support centers were not consolidated by 2019, and distribution networks were largerly ran independently until 2020 (Dollar Tree (2018) and Figure A5).[4] This motivates our treatment of the two chains as separate over our sample period ending in 2019.

We combine several data sources to study dollar store expansion and the effects on consumers and local market structure.

The first is the SNAP Retailer panel, a yearly panel of SNAP-authorized retailers from 2008 to 2019. This dataset contains information on over 400,000 U.S. retailers including their chain affiliation and store type, as well as small independent retailers. The SNAP retailer panel contains any store that accepts SNAP benefits. In addition to dollar stores, this includes convenience stores, combination stores (stores selling a combination of general merchandise and food products), grocery stores, drugstores, gas stations, supermarkets, and supercenters. Table 1 shows store counts by type in the SNAP panel. As far as we know, the SNAP retailer data is novel in the economics literature.[5] The primary benefits of this public data source are that it is an annual measure and contains nearly the full universe of retailers in this industry. Crucially for this study, the panel includes small independent stores, which are typically absent from other retail census used in the literature.

A drawback of this dataset is that entry into the SNAP program may not necessarily indicate the start of operation of a physical store. In particular, as the SNAP program

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

[4] In 2020, two new distribution centers were opened serving both banners, e.g., Ocala, FL and Rosenberg, TX.

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

debuted in 2008, there may have been delays in stores joining the program for the first few years. We alleviate this concern in two ways. For chains, we can compare store counts in the SNAP panel against publicly disclosed store counts in chains' annual reports to investors. We do not find any significant discrepancies between the two sources. For independent stores, this approach is not possible: instead, we drop the first few years in the sample and restrict our analysis to the period from 2010 to 2019.

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

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

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

We compile data from the IRI Consumer Network panel, which we complement with the IRI MedProfiler dataset. The Consumer Network data contains household-level panels on all retail purchases for a nationwide sample of consumers. The MedProfiler contain nutritional information for food purchases (e.g., sugar, sodium) and consumer health metrics (e.g., BMI). Although the household scanner data is standard in the IO literature, we include informative summary statistics of spending by retail channel and food category in Table A2 of the Appendix.[6] These statistics show that spending at dollar stores is relatively low compared to other retail channels (e.g., supermarkets, supercenters). The largest expenditure share at dollar stores is for soda, snacks, candy, and crackers; whereas fresh produce spending is close to zero.

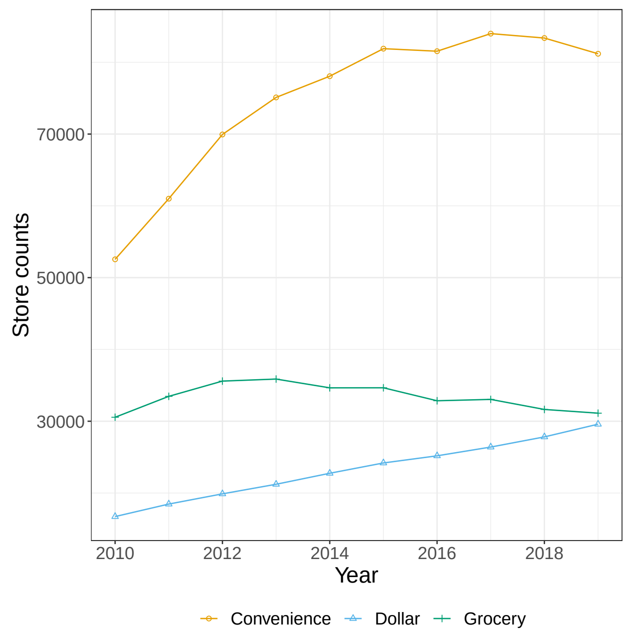
Finally, we collect market-level data on demographic characteristics from the Census and ACS at the census tract level. This allows us to study how market characteristics and consumer demographics affect dollar stores and other retailers' entry behavior and profits. We also collect data on distribution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

Figure 1 shows the total number of stores at the national level over the period 2010-2019. The total number of stores operated by the three major dollar store chains increases

[6] We note that the "Grocery" retail channel in the IRI data includes both grocery stores and supermarkets. In this paper, when studying the impact on market structure, we distinguish between small independent grocery stores and supermarket chains.

by 12,870 during this period. This increase is the net effect from 14,554 store entry events and 1,684 store exit events. The number of independent retail stores also increases, with this growth driven by convenience stores (71,010 store entries and 42,363 store exits). The number of grocery stores falls by 13% from its high in 2012.

Figure 1: Store counts by firm type



We present the evolution of dollar store chains' distribution centers over time, in Figure A4 of the Appendix. Between 2000 and 2020, the number of distribution centers increases substantially. This leads to a decrease in the average distance between a market and a distribution center, which falls by roughly 125mi for Dollar General and slightly smaller amounts for the other two chains.[7] The locations of these in 2019 are shown in Figure A5.

We use demographic information from the Census and the IRI Consumer Network to document consumer heterogeneity across locations with varying dollar store densities. Table 2 shows summary statistics of census demographic information for the locations entered by dollar store chains prior to 2010, during the 2010-2019 period, and locations never entered. A location is defined at the Census Tract level. Dollar store entry occurs in locations that have significantly lower incomes (per capita) and rents than other locations, and a significantly higher share of the population that is black or below the poverty line. Entered locations are also significantly closer to distribution centers than non-entered locations.

The patterns that emerge using the Census or ACS data are consistent with household

[7] We provide details about the market definition used in Section 5.

Table 2: Market Summary Statistics

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

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

demographics in the IRI Consumer Network panel. Table 3 shows demographic information for all panelists, broken down by the number of dollar stores entries within 2 miles of the household. Households experiencing many dollar stores entries have lower (household) income, are more likely to have a female household head, without children, and are less likely to be white, married, employed, with access to a vehicle.

To better understand which types of households display preferences for dollar stores, we further break down the sample of households who experience entry over the sample period, as shown in Table 4. In columns (1) and (2), we compare households with large spending share at the dollar store channel (top 95th percentile and above) post-dollar store entry to households with no spending at the dollar store channel post-entry. These two groups differ on several dimensions: households with high dollar store spending shares are significantly more likely to be low income (household income below \$35 thousands), with a female household head, from a minority group, unmarried, unemployed and without access to a vehicle. The latter group also spends less on fresh produce (\$33 versus \$64 per household member) and at the grocery retail channel.

In column (4), we focus on households with no dollar store spending in the time period between their first dollar store entry and their first subsequent grocery store exit. Arguably, these households do not directly benefit from dollar store entry but still experience a loss

Table 3: Household Demographics by Number of DS Entries in IRI Consumer Network

		By Number of DS Entries			
Variable	All Panelists	0	1	2	3+
Income (000s)	66.81 (43.90)	68.67 (44.64)	64.89 (43.01)	62.77 (41.99)	59.30 (40.07)
Low-Income (<35k)	0.27 (0.45)	0.26 (0.44)	0.29 (0.45)	0.30 (0.46)	0.32 (0.46)
High-Income (>100k)	0.17 (0.38)	0.18 (0.39)	0.16 (0.36)	0.14 (0.35)	0.12 (0.32)
Years education	14.64 (2.15)	14.67 (2.14)	14.60 (2.16)	14.62 (2.17)	14.63 (2.18)
No female head	0.10 (0.30)	0.09 (0.28)	0.11 (0.31)	0.13 (0.34)	0.15 (0.36)
No male head	0.26 (0.44)	0.24 (0.43)	0.28 (0.45)	0.31 (0.46)	0.36 (0.48)
With children	0.22 (0.41)	0.26 (0.44)	0.17 (0.38)	0.13 (0.34)	0.13 (0.33)
Age	56.17 (13.12)	54.72 (13.42)	57.98 (12.49)	59.36 (11.96)	60.04 (11.27)
Household Size	2.38 (1.29)	2.49 (1.33)	2.25 (1.21)	2.11 (1.17)	2.08 (1.21)
White	0.82 (0.39)	0.84 (0.37)	0.82 (0.39)	0.75 (0.43)	0.65 (0.48)
Black	0.10 (0.30)	0.08 (0.27)	0.10 (0.31)	0.17 (0.38)	0.28 (0.45)
Married	0.64 (0.48)	0.67 (0.47)	0.62 (0.48)	0.56 (0.50)	0.50 (0.50)
Employed	0.56 (0.50)	0.59 (0.49)	0.53 (0.50)	0.48 (0.50)	0.45 (0.50)
No vehicle	0.03 (0.18)	0.03 (0.16)	0.03 (0.18)	0.06 (0.23)	0.09 (0.29)
Observations	618,621	381,672	158,875	51,341	26,733

Note: The unit of observation is the household-year. The table shows mean values and standard errors are shown in parenthesis. The four right-most columns show the subsample of households who experienced a given number of entries over the sample period (2010-2019). For households with two household heads, we use the mean of age, employment hours, educational attainment for male and female household heads.

in the number of nearby grocery stores.[8] This group, like that of column (1), is on average high-income, white, employed, married, with vehicle access.

Table 4: Demographics of households experiencing at least one dollar store entry over the sample period

	By DS Spending Share Post-DS Entry			By DS Spending Share Post-DS Entry and Pre-GS Exit		
Variable	(1) No Spending	(2) High Spending (Top 5%)	(3) Δ (t-stat)	(4) No Spending	(5) High Spending (Top 5%)	(6) Δ (t-stat)
Income (000s)	75.08 (45.68)	41.02 (32.38)	96.6	74.86 (45.72)	43.01 (32.99)	46.0
Low-Income (<35k)	0.21 (0.41)	0.55 (0.50)	-69.6	0.22 (0.41)	0.52 (0.50)	-30.9
High-Income (>100k)	0.22 (0.42)	0.04 (0.20)	71.7	0.22 (0.41)	0.05 (0.22)	33.7
Years education	15.13 (2.18)	13.91 (2.08)	57.5	15.22 (2.19)	14.00 (2.12)	28.9
No female head	0.17 (0.38)	0.14 (0.34)	9.6	0.19 (0.39)	0.11 (0.32)	11.8
No male head	0.24 (0.42)	0.42 (0.49)	-36.9	0.28 (0.45)	0.44 (0.50)	-16.7
With children	0.17 (0.37)	0.12 (0.33)	13.8	0.12 (0.33)	0.11 (0.31)	2.5
Age	57.00 (13.10)	59.47 (11.41)	-20.8	58.63 (12.09)	60.90 (11.03)	-10.2
Household Size	2.21 (1.22)	2.00 (1.17)	17.8	2.05 (1.14)	1.98 (1.08)	3.0
White	0.83 (0.37)	0.65 (0.48)	38.1	0.82 (0.38)	0.61 (0.49)	22.2
Black	0.08 (0.26)	0.26 (0.44)	-44.1	0.09 (0.29)	0.30 (0.46)	-23.8
Married	0.60 (0.49)	0.45 (0.50)	30.3	0.55 (0.50)	0.46 (0.50)	8.9
Employed	0.57 (0.50)	0.41 (0.49)	32.3	0.52 (0.50)	0.40 (0.49)	12.7
No vehicle	0.04 (0.20)	0.10 (0.30)	-8.5	0.05 (0.22)	0.10 (0.29)	-3.5
Pre-Entry Fresh Produce Spending	142.10 (140.42)	66.64 (69.54)	88.2	132.50 (136.39)	68.67 (72.97)	38.5
Pre-Entry GS Channel Spending	1929.54 (1434.27)	1186.75 (965.85)	69.6	1892.96 (1412.82)	1178.17 (993.68)	34.1
Pre-Entry DS Channel Spending	2.20 (24.76)	223.87 (267.30)	-88.7	2.22 (40.34)	217.97 (265.03)	-43.6
Observations	63,535	11,452		20,794	2,874	

Note: The unit of observation is the household-year, for the subsample of households who experience at least one dollar store entry within 2 miles over the sample period. Columns (4) and (5) further restrict the sample to household who experience at least one grocery store exit following the first dollar store entry. The table shows mean values and standard errors are in parenthesis. Columns (2) and (5) show households with a share of spending at the dollar channel in the 95th percentile or above (8 to 9% of total spending). For households with two household heads, we use the mean of age, employment hours, educational attainment for male and female household heads. Household Income is available in 9 bins. Mean income is computed by taking the mid-range of each bin. The top bin (>$100k) is coded as $150,000.

[8] There could be indirect benefits from dollar store entries if other stores respond by lowering prices. However, we do not find significant support for this using the IRI data.

3 Reduced Form Analysis

3.1 Effects on Market Structure

In this section, we present evidence on the impact of dollar store chain entry on local retail markets. Our primary goal is to evaluate whether or not dollar store chain entry is associated with decreases in the number of local independent retailers and grocery stores. To do so, we rely on the data described above containing the annual universe of retailers and study the rapid expansion of dollar store chains between 2010 and 2019.

During this time period, we observe 14,554 dollar store chain entries in the United States. To study the local effects of these entries, we break markets into locations based on Census tracts as described above. For each location, we obtain the population-weighted centroid, and define distance bands around each location using radii of 0-2mi, 2-5mi and 5-10mi. Our focus is on the number of independent grocery stores as the main outcome of interest.

Our identification strategy for measuring the effects of dollar store chain entry on these outcomes is to use tract-level fixed effects to account for time-invariant unobserved market characteristics and county-year fixed effects to account for time-varying trends at the market level. We also incorporate time-varying demographic variables at the census tract level. These are intended to control for local trends in population, income, or business activity associated with economic shocks. We include population, median income, and level of residential rents as well as the annual growth rate in each of these variables.

We estimate effects using the following specification:

$$Y_{lt} = \beta \mathbf{X}_{lt} + \delta DS_{lt} + \lambda_l + \alpha_{mt} + \epsilon_{lt} \tag{1}$$

Here λ_l and α_{mt} represent location and county-time fixed effects, and the objects of interest δ are the coefficients on the number of dollar store chain entry events in location l. The outcome variable is the number of independent grocery stores in the 0-2m radius around the entry location. We also define $\mathbf{X}_{lt}$, the local demographics (including growth rates), at this level.

In Table 5 we show results for different specifications of controls and fixed effects where the outcome variables are the number of independent local grocery stores. We find that, once location FE or demographic controls are included, there is a consistent negative effect on the number of independent local grocery stores that is increasing in the number of dollar store entries. This effect is roughly -.32 for two entries, increasing to a decrease of more than 1 grocery stores in locations with 3+ entries.

In Table 6 we study how the effects of dollar store chain entry vary across different regions

Table 5: Effects of DS Entry (0-2mi) on Number of Grocery Stores

	(1) All Markets	(2) All Markets	(3) All Markets	(4) All Markets
First DS Entry	3.121***	-0.395***	-0.107***	-0.0706***
	(0.0612)	(0.0439)	(0.0147)	(0.0137)
Two DS Entries	6.889***	0.0138	-0.554***	-0.318***
	(0.124)	(0.0876)	(0.0300)	(0.0267)
Three+ DS Entries3	12.46***	1.347***	-1.712***	-1.027***
	(0.253)	(0.174)	(0.0636)	(0.0512)
Constant	2.446***	3.199***	4.961***	2.504***
	(0.0273)	(0.0734)	(0.248)	(0.261)
Year FE	Yes	Yes	Yes	Yes
Demographic Controls		Yes	Yes	Yes
Census Tract FE			Yes	Yes
Market*Year FE				Yes
Observations	699,980	558,783	558,030	555,819
F-stat	1413.8	1561.7	127.8	89.4
R^2	0.072	0.59	0.98	0.99
Adjusted R^2	0.072	0.59	0.98	0.98
Mean Pre-Entry	2.73	2.73	2.73	2.73

Notes: Unit of observation is the location-year. Standard errors (in parenthesis) clustered at the location level. Controls for time-varying local demographics (income, population, residential rents, and the one-year percent change in each) are included. Low-income locations are those where per capita income is less than $25,000 in 2010.

defined by the distance from the entry location. We study regions defined by radii of 0-2mi, 2-5mi, and 5-10mi from the entry location.[9] The results show a substantial fall in number of grocery stores in the 2 mile radius around where the dollar store entry occurs. In the area 2-5 miles from the dollar store entry there is a smaller but still significant effect when there are more than one entry, and in the area 5-10 miles away there are no detectable negative effects of dollar store entry.

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

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

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

Table 6: Effect of DS Entry (by distance band) on Number of Grocery Stores

	(1) 0-2m	(2) 2-5m	(3) 5-10m
First DS Entry	-0.0706***	-0.0985***	0.0292
	(0.0137)	(0.0171)	(0.0168)
Two DS Entries	-0.318***	-0.209***	0.0578
	(0.0267)	(0.0319)	(0.0333)
Three+ DS Entries	-1.027***	-0.501***	0.282***
	(0.0512)	(0.0569)	(0.0601)
Constant	2.504***	6.454***	6.853***
	(0.261)	(0.325)	(0.351)
Demographic Controls	Yes	Yes	Yes
Census Tract FE	Yes	Yes	Yes
Market*Year FE	Yes	Yes	Yes
	(0.248)	(0.276)	(0.236)
Observations	555,819	337,291	283,555
F-stat	89.4	25.3	5.06
R^2	0.99	0.99	0.99
Adjusted R^2	0.98	0.99	0.99
Mean Pre-Entry	6.80	12.0	19.0

Notes: Unit of observation is the location-year. Standard errors (in parenthesis) clustered at the location level. Time and location fixed effects are included. Controls for time-varying local demographics (income, population, residential rents, and the one-year percent change in each) are included, each is calculated for the radius area corresponding to the dependent variable.

$$Y_{lt} = \beta \mathbf{X}_{lt} + \sum_{\tau=-8}^{8} \delta_\tau DS_{l,t-\tau} + \lambda_l + \alpha_{mt} + \epsilon_{lt} \tag{2}$$

This differs from the specification in equation (1) in that now the coefficients on dollar store entry are subscripted by τ, the difference in years measured relative to the entry date. This allows for both dynamic policy effects, such as a delay in the effect on local markets as the dollar store's sales ramp up, and for detecting the presence of pre-trends in grocery store activity prior to the dollar store enters. For this analysis we also focus only on a binary dummy for whether or not an entry occurs, rather than the number of entries as used above.

This design is essentially a Difference-in-Difference with TWFE and a staggered roll-out of treatment. This type of empirical design has been shown to have the potential for biased results (Goodman-Bacon (2021), Sun and Abraham (2021)). Consequently, we present results based on heterogeneity-robust estimators, e.g., Callaway and Sant'Anna (2021).[10] We plot the results in Figure 2 in the manner suggested by Freyaldenhoven et al. (2021), with confidence intervals adjusted for multiple hypothesis testing. The left panel shows the effects of dollar store entry on the number of grocery stores in the 0-2 mile radius and the right panel shows effects in the 2-5 mile radius. In both panels we see no pre-trend in the number of grocery stores, followed by a clear downward trend after the time of dollar store entry. The effects in the 2-5m radius are about half as large as in the 0-2m radius.

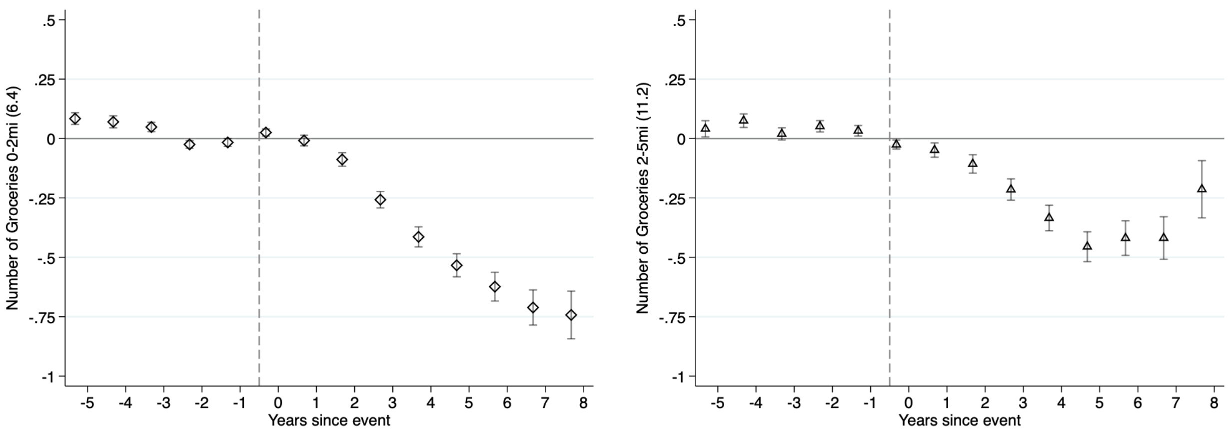


Figure 2: The effects of dollar store entry on local independent grocery stores measured in the 0-2 mile radius around entry (left panel) and in the 2-5 mile radius around entry (right panel). Results are from an event study analysis using a heterogeneity-robust estimator proposed by Callaway and Sant'Anna (2021). Confidence bands show the uniform sup-t confidence intervals adjusted for multiple hypothesis testing.

3.2 Effects on Consumers' Shopping Behavior

This section studies how retail entry (dollar stores) and exit (grocery stores) affects households' shopping behavior. We exploit variation in local supply within household in an event study framework, as in Allcott et al. (2019). We begin by analyzing how the changes in market structure highlighted in the previous section affect aggregate food spending. Next, we show that the adverse impact of dollar store entry on the number of grocery stores is reflected in spending and trip diversion from the grocery to the dollar store retail channel. Finally, we examine how the above market structure changes translate into spending on various food categories and the nutritional value of households' shopping baskets. We aim attention at spending and volume of purchases of fresh produce, specifically, because access

[10] In Appendix Figure A2 we present results from the OLS-TWFE implementation as well as several other bias correction approaches that have been suggested in the literature and they provide consistent results.

to this particular food group has been at the center of the dollar store policy debate.

In the remainder of this section, the unit of analysis is the household–year. We consider both static and dynamic (event study design) specifications. Let T_{bct} denote a measure of changes in market structure: for instance, the number of dollar store entries that have occurred within distance band b (e.g., 0-2mi, 2-5mi) from census tract c by period t. Let $\mathbf{X}_{it}$ denote time-varying household characteristics (e.g., age, education, employment, marital status). Let Y_{ict} denote the outcome variable for household i living in census tract c in period t. We consider the following static specification

$$Y_{ict} = \tau T_{bct} + \beta \mathbf{X}_{it} + \gamma_i + \eta_t + \epsilon_{ict} \tag{3}$$

where γ_i and η_t are household and time fixed effects. To capture potential non-linear effects of changes in market structure, we include T_{bct} as a categorical variable.

As the effects we estimate are likely persistent and heterogeneous over time, we also consider a dynamic specification in an event study framework. Let E_{bct} denote a dummy for whether a change in market structure (e.g., one dollar store entry) has occurred within distance band b (e.g., 0-2mi, 2-5mi) from census tract c by period t. We consider the following dynamic specification

$$Y_{ict} = \sum_{\tau} \tau_l E_{bc,t-\tau} + \beta \mathbf{X}_{it} + \gamma_i + \eta_t + \epsilon_{ict} \tag{4}$$

The omitted category is $\tau = -1$, so that all cumulative effects are relative to the period before entry.

Table 7 shows the effect of dollar store entry (within 2mi of the household) on spending and the number of yearly trips by retail channel.[11] The results indicate that dollar stores tend to divert away spending primarily from the grocery and convenience retail channel, but not from supercenters or club stores. Entry is associated with increased number of trips to the dollar channel, from 4.8 to 5.5 trips following one entry and 5.8 following two entries. Household in low-income counties tend to make more frequent trips to the dollar channel (6.48 versus 4.34 for the balanced panel) and less frequent trips to the grocery channel (42.6 versus 47.5 for the balanced panel). The impact of entry on the number grocery and convenience trips is negative and significant.

Next, we turn to the effect on spending per food category. Table 8 shows the effect of dollar store entry within 2mi and from 2 to 5mi of the household on fresh produce spending. Overall, we find a negative and significant effect of entry within 2mi of the household on

[11] Retail channels, as defined by IRI, correspond to broad retail categories. The "Grocery" channel includes both supermarkets and grocery stores but excludes supercenters and club stores.

Table 7: Effect of DS Entry (0-2mi) on Spending and Number of Trips by Retail Channel

		Spending by Channel				Trips by Channel			
	(1) Aggregate	(2) Grocery	(3) Dollar	(4) SC/Club	(5) Conv	(6) Grocery	(7) Dollar	(8) SC/Club	(9) Conv
First DS Entry	-30.82***	-20.85***	5.063***	-5.612	-0.991*	-0.913***	0.670***	-0.141	-0.162***
	(6.590)	(5.048)	(0.493)	(3.845)	(0.413)	(0.169)	(0.0509)	(0.102)	(0.0478)
Two DS Entries	-6.746	-8.324	5.949***	-1.104	0.108	-0.888**	1.060***	0.416*	-0.158
	(12.28)	(9.946)	(0.952)	(7.219)	(0.702)	(0.339)	(0.100)	(0.201)	(0.0989)
Three+ DS Entries	40.52	30.05	10.48***	-7.074	-2.105	-0.450	1.572***	1.269***	-0.539**
	(21.03)	(16.79)	(1.644)	(11.19)	(1.409)	(0.625)	(0.187)	(0.338)	(0.184)
Observations	570,689	570,689	570,689	570,689	570,689	570,689	570,689	570,689	570,689
F-stat	838.8	426.7	4.14	928.2	2.24	48.0	6.84	674.5	2.79
R^2	0.83	0.84	0.77	0.84	0.68	0.84	0.81	0.84	0.75
Adjusted R^2	0.79	0.80	0.71	0.80	0.61	0.81	0.76	0.80	0.70
Mean Pre-Entry	2917.3	1768.3	32.3	777.2	14.6	57.8	4.85	22.4	2.16

Notes: Unit of observation is the household-year. Standard errors (in parenthesis) clustered at the household level. Year and household fixed effects are included. Controls for time-varying household demographics (income, education, age, household size, marital status, occupation, weekly hours worked) are included. Results are shown for all panelists in the sample. The first column shows aggregate spending. SC/Club stands for the Supercenter and Club store channel. Conv stands for the convenience store channel.

produce spending across samples. The magnitude of the effect increases with the number of dollar store entries and is higher for households living in low-income counties. For the sample of low-income household reporting complete purchases of random-weight products (e.g., fresh produce), households that experience one entry reduce their produce spending by $5.4 or 4% of their pre-entry spending, whereas households that experience three entries or more reduce their spending by $10 or 7.3% of their pre-entry spending.

To put these effects into a broader perspective, we compute the difference in fresh produce spending between the bottom and top quartile of the per-capita income distribution, controlling for household size, age, and year indicators. This inter-quartile difference is $60 over the sample period (for an average total spending on fresh produce of $130 per year across all panelists). Therefore, the reduction in spending due to dollar store entry corresponds to between 9% and 16% of the difference in fresh produce spending between the top and bottom quartile of the income distribution.

Table 8 shows that the effects are localized. Entry of dollar stores in the 2-5mi distance band does not significantly affect produce spending. Figure 3 shows the corresponding event study plots for the balanced sample, using a two-way fixed effect as well as a heterogeneity-robust estimator (Callaway and Sant'Anna (2021)).[12] Consistent with the static analysis, we find that the effects are localized and negative only at the 0-2mi range. Moreover, the negative impact on fresh produce spending is dynamic, with the magnitude increasing in the

[12] We experiment with several other heterogeneity-robust estimators based on imputation (Borusyak et al. (2021)) and manual aggregation (Sun and Abraham (2021)). Results are shown in Figure A3 of the Appendix.

Table 8: Effect of DS Entry (by distance band) on Fresh Produce Spending

	0-2mi			2-5mi		
	(1) All	(2) Bal	(3) Bal/RW/LI	(4) All	(5) Bal	(6) Bal/RW/LI
First DS Entry	-0.945	-1.609	-5.464**	0.0319	0.478	-1.138
	(0.588)	(0.951)	(1.735)	(0.504)	(0.872)	(1.800)
Two DS Entries	-3.431**	-5.366***	-7.634*	0.315	-0.492	-0.308
	(1.148)	(1.621)	(3.047)	(0.782)	(1.197)	(2.540)
Three+ DS Entries	-5.209**	-8.323***	-9.939*	-1.292	-2.692	-4.525
	(1.934)	(2.410)	(4.595)	(1.098)	(1.545)	(3.236)
Observations	570,689	246,851	59,088	570,689	246,851	59,088
F-stat	55.3	6.36	2.83	55.2	6.25	2.77
R^2	0.76	0.74	0.78	0.76	0.74	0.78
Adjusted R^2	0.71	0.72	0.74	0.71	0.72	0.74
Spending Pre-Entry	130.9	146.4	128.3	126.1	137.9	125.0

Notes: Unit of observation is the household-year. Standard errors (in parenthesis) clustered at the household level. Year and household fixed effects are included. Controls for time-varying household demographics (income, education, age, household size, marital status, occupation, weekly hours worked) are included. The balanced panel corresponds to households observed for at least 9 consecutive years. The RW panel reports complete purchases of random weight products. LI stands for low-income, or household located in a county where per capita income is less than $25,000 in 2010.

time since entry.

We provide evidence that the impact of dollar store entry on fresh produce purchases acts in part through exits of grocery stores. Table 9 shows the effect of grocery store exists on fresh produce spending. The results indicate that the exit of two or more grocery stores near the household reduces spending on produce by up to 7% of pre-exit spending.

Table 10 show spending on other food categories, ranging from Soda and Snacks to Meals (soup, rice, and pasta). The majority of changes in spending are insignificant, except for frozen produce, on which spending increases following three or more entries.

Finally, we investigate whether these changes in fresh produce spending translate into meaningful changes in volumes consumed. Figure 4 shows a event study plot of entry on volume of fresh produce purchased in ounces. The results indicates that entry within 2mi of the household leads to a significant reduction in volumes of fresh produce purchased of approximately 5%. For completeness, we include the corresponding static analysis in Table A3 of the Appendix.

On the whole, the findings in sections 3.1 and 3.2 point to a large impact of dollar store expansion on market structure and modest but significant impact on consumers' dietary

Table 9: Effect of Grocery Exits (0-2mi) on Fresh Produce Spending

	Dependent Variable: Spending on Fresh Produce		
	(1) All panelists	(2) Balanced	(3) Balanced/RW/LI
First GR Exit	-0.816	-1.730	-0.612
	(0.585)	(0.956)	(2.007)
Two GR Exits	-2.208*	-4.082**	-4.102
	(1.018)	(1.447)	(3.335)
Three+ GR Exits	-4.054*	-6.899***	-10.03*
	(1.587)	(2.086)	(5.003)
Observations	570,689	246,851	59,088
F-stat	55.3	6.31	2.76
R^2	0.76	0.74	0.78
Adjusted R^2	0.71	0.72	0.74

Notes: Unit of observation is the household-year. Standard errors (in parenthesis) clustered at the household level. Year and household fixed effects are included. Controls for time-varying household demographics (income, education, age, household size, marital status, occupation, weekly hours worked) are included. The balanced panel corresponds to households observed for at least 9 consecutive years. The RW panel reports complete purchases of random weight products. LI stands for low-income, or household located in a county where per capita income is less than $25,000 in 2010.

Table 10: Effect of DS Entry (0-2mi) on Spending by Product Category

	Dependent Variable: Spending by Product Category						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First DS Entry	-0.521	0.209	1.171	-0.588	-2.685*	-0.900	1.604
	(0.399)	(0.344)	(1.939)	(0.709)	(1.301)	(1.076)	(1.681)
Two DS Entries	0.190	0.920	-0.457	0.887	-1.554	-0.987	3.915
	(0.681)	(0.637)	(3.388)	(1.247)	(2.216)	(1.840)	(2.939)
Three+ DS Entries	1.551	2.767**	1.789	1.881	-0.572	2.152	7.898
	(1.010)	(0.990)	(5.065)	(1.843)	(3.353)	(2.989)	(4.542)
Observations	246,851	246,851	246,851	246,851	246,851	246,851	246,851
F-stat	10.9	5.66	14.8	17.8	22.3	15.8	8.67
R^2	0.66	0.67	0.79	0.73	0.79	0.75	0.75
Adjusted R^2	0.62	0.63	0.76	0.70	0.77	0.72	0.72
Spending Pre-Entry	64.8	39.9	491.5	153.8	351.1	214.7	331.7

Notes: Unit of observation is the household-year. Standard errors (in parenthesis) clustered at the household level. Year and household fixed effects are included. Controls for time-varying household demographics (income, education, age, household size, marital status, occupation, weekly hours worked) are included. Results are shown for the balanced panel. Each column corresponds to a product category: (1) is canned produce, (2) is frozen produce, (3) is soda, snacks, candy and crackers, (4) is meals (incl. rice, pasta, soup), (5) is dairy, (6) is refrigerated and frozen meats, (7) is refrigerated and frozen baked goods, desserts, dough, and beverages.

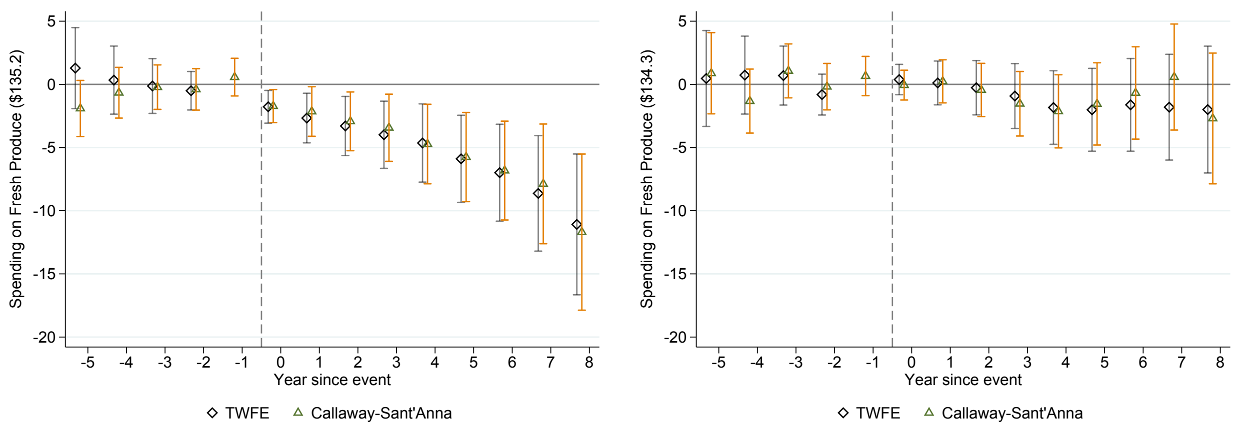


Figure 3: The effects of dollar store entry at the 0-2mi (left panel) and 2-5mi (right panel) bands around the household on spending on fresh produce. Results are from an event study analysis on the balanced panel, using a TWFE and a heterogeneity-robust estimator proposed by Callaway and Sant'Anna (2021). Household and year fixed effects are included. Confidence bands show the uniform sup-t confidence intervals adjusted for multiple hypothesis testing.

choices. While previous research (e.g., Allcott et al. (2019)) has shown that entry of grocery stores and supermarkets has economically small effects on healthy grocery shopping, the analysis above demonstrates that (1) grocery stores are driven out by dollar store entry in close proximity; (2) the exit of many grocery stores is associated with a moderate but significant reduction in spending on and volumes of fresh produce. This asymmetric impact of grocery store entries and exits on healthy grocery shopping may be tied to consumer inertia in store and brand choices (Dubé et al. (2010)). Inertia in households' shopping may create asymmetries in the response to positive or negative shocks to their retail environment.

4 Industry Model

In this section, we describe a model of the entry and exit game played by rival retailers over time. The goal of this model is to provide estimates of the size and nature of competitive effects between different store types and to evaluate how market structure would evolve under counterfactual policy scenarios. We can thus quantify the cumulative reduction in the number of grocery stores, convenience stores, etc., across different market types that has resulted from dollar store chains' expansion.

To effectively answer our research questions, the model must be able to generate the counterfactual market structure under alternative policies with respect to dollar store entry, including a total ban on dollar store entry in the sample period. We consider the actions

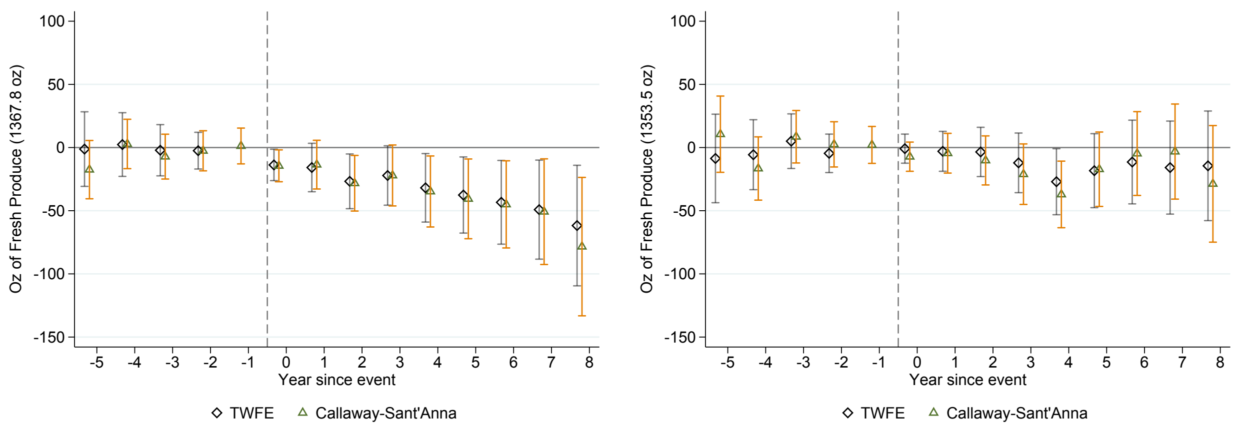


Figure 4: The effects of dollar store entry at the 0-2mi (left panel) and 2-5mi (right panel) bands on fresh produce volume purchased. Results are from an event study analysis on the balanced panel, using a TWFE and a heterogeneity-robust estimator proposed by Callaway and Sant'Anna (2021). Household and year fixed effects are included. Confidence bands show the uniform sup-t confidence intervals adjusted for multiple hypothesis testing.

of each type of store, how they interact, and how they depend on exogenous market characteristics. The main innovations in the model are that we model a situation that contains long-lived chains as central players and combines dynamics with spatial differentiation.

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

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

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

There are market and location characteristics that evolve exogenously over time, denoted $\mathbf{x}_{mt} = \{x_{mlt}\}_{l \in L}$. These include the population, income per capita, and rents in each location. Market-level characteristics include the number of other store types (e.g., drug stores, supermarkets, and gas stations) which are not players in the game, but can be payoff-relevant.[13]

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

For multi-store firms, let d_{imt} denote market m's distance from i's closest distribution center and $\mathbf{d}_{mt} = (d_{imt})_{i \in I_c}$ the vector collecting this variable for all chains. This vector evolves deterministically over time, as the chains expand their network of distribution centers. The transition matrices for these variables are denoted: $f(x_{ml,t+1}|x_{ml,t})$ and $h_t(d_{im,t+1}|d_{imt})$. The latter transition matrix is deterministic and the source of non-stationarities in the model.[14]

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

$$\mathcal{M}_{j,i,t} = (\mathbf{n}_{it}, \mathbf{n}_{-it}, \mathbf{x}_{mt}, \mathbf{d}_{mt}) \tag{5}$$

Actions

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

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

We represent the transition rule of market structure as $\mathbf{n}_{t+1} = \mathbf{n}_t + 1[a_t]$, where $1[a_t]$ is a vector such that its (i, l)-element is equal to $+1$ when $a_{it} = l_+$, to -1 when $a_{it} = l_-$, and to zero otherwise. Firms' choices are dynamic because of partial irreversibility in the decision to open a new store, i.e., sunk costs. At the end of period t firms simultaneously

as Walmart.

[14] 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 precise estimation of these choice probabilities.

choose their network of stores $\mathbf{n}_{t+1}$ with an understanding that they will affect their variable profits at future periods. We model the choice of store location as a game of incomplete information, so that each firm i has to form beliefs about other firms' choices of networks. More specifically, there are components of the entry costs and profits of a store which are firm-specific and private information.

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

$$\pi_{it}(a_{it}, \mathcal{M}_{j,i,t}) = VP_i(\mathcal{M}_{j,i,t}) - FC_{it}(\mathcal{M}_{j,i,t}) - EC_{it}(a_{it}) + EV_{it}(a_{it}), \tag{6}$$

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

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

$$VP_i(\mathcal{M}_{j,i,t}) = \sum_{l=1}^{L} n_{ilt} vp_{i,l}(\mathcal{M}_{j,i,t}) \tag{7}$$

where $vp_{i,l}(\mathcal{M}_{j,i,m,t})$ are per-store profits. For a store in location l, variable profits are a function of the exogenous characteristics and the number of (own and rival) stores in location l and surrounding locations. Following Seim (2006), we capture this dependence by defining these variables for various distance bands, $b = 1, \ldots, B$, around location l

$$vp_{i,l}(\mathcal{M}_{j,i,t}) = \sum_{b=1}^{B} \alpha_i^b x_{mlt}^b + \sum_{b=1}^{B} \beta_{io}^b n_{ilt}^b + \sum_{b=1}^{B} \sum_{f=1}^{F} \beta_{if}^b n_{flt}^b \tag{8}$$

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

$$FC_{it}(\mathcal{M}_{j,i,t}) = \theta^{FC}_{1,c} d_{imt} + + \sum_{l=1}^{L} \theta^{FC}_{2,c} rent_{mlt} \tag{9}$$

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

The specification of entry costs is

$$EC_{it} = \sum_{l=1}^{L} 1\{a_{it} = l_{+}\}\theta^{EC}_{i}. \tag{10}$$

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

The exit value is specified as:

$$EV_{it} = \sum_{l=1}^{L} 1\{a_{it} = l_{-}\}\theta^{EV}_{i}. \tag{11}$$

Similarly to entry costs, the exit value θ^{EV}_{f} depends only on the type of the firm f.

At the beginning of period t, each firm draws a vector of private information shocks associated with each possible action $\boldsymbol{\epsilon}_{it} = \{\epsilon_{it}(a)\}_{a \in A(\mathbf{n}_{it})}$. We assume that the shocks ϵ_{it} are independently distributed across firms and over time and have a cumulative distribution function $G(.)$ that is strictly increasing and continuously differentiable with respect to the Lebesgue measure. These two assumptions allow for a broad range of specifications for the ϵ_{it}, including spatially correlated shocks. In our application, these shocks will be distributed Type 1 extreme value, scaled by a parameter θ^{ϵ}.[15]

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

$$\Pi_{it}(a_{it}, \mathcal{M}_{j,i,t}, \boldsymbol{\epsilon}_{it}) = \pi_{i}(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}). \tag{12}$$

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

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

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

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

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

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

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

$$\begin{aligned} v_{i,t}^{\alpha}(a_{it}, \mathcal{M}_{j,i,t}) &= \pi_i(a_{it}, \mathcal{M}_{j,i,t}) \\ &\quad + \beta \int V_{i,t}^{\alpha}\left(\mathcal{M}_{j,i,t+1}, \boldsymbol{\epsilon}_{i,t+1}\right) dG(\boldsymbol{\epsilon}_{i,t+1}) dF_t(\mathcal{M}_{j,i,t+1} | a_{it}, \mathcal{M}_{j,i,t}) \end{aligned} \tag{14}$$

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

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

$$\alpha_{i,t}^*(\mathcal{M}_{j,i,t}, \epsilon_{it}) = \arg \max_{a_{it} \in A(n_{it})} \left\{ v_{i,t}^{\alpha^*}(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it}) \right\} \tag{15}$$

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

$$P_t^{\alpha}(a_{it}|\mathcal{M}_{j,i,t}) = \Pr(\alpha_{i,t}(\mathcal{M}_{j,i,t}, \epsilon_{it}) = a_{it}|\mathcal{M}_{j,i,t}) \tag{16}$$

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

$$v_{i,t}^{\mathbf{P}}(a_{it}, \mathcal{M}_{j,i,t}) = \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \beta \sum_{a_{-it}} \int \overline{V}_{i,t+1}^{\mathbf{P}}\left(\mathcal{M}_{j,i,t+1}\right) dF_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t}) \tag{17}$$

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

$$\overline{V}^{\mathbf{P}}_{i,t}(\mathcal{M}_{j,i,t}) = \int \max_{a_{it}\in A(n_{it})} \Bigg\{ \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \epsilon_{it}(a_{it})$$

$$+\beta \sum_{a_{-it}} \int \overline{V}^{\mathbf{P}}_{i,t+1}(\mathcal{M}_{j,i,t+1})\, dF_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t}) \Bigg\} dG(\epsilon_{it}). \tag{18}$$

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

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

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

$$P_t(a_{it}|\mathcal{M}_{j,i,t}, \mathbf{P}) = \frac{\exp\left(\frac{v_{i,t}^{\mathbf{P}}(a_{it}, \mathcal{M}_{j,i,t})}{\theta^{\epsilon}}\right)}{\sum\limits_{a'\in A(n_{it})} \exp\left(\frac{v_{i,t}^{\mathbf{P}}(a', \mathcal{M}_{j,i,t})}{\theta^{\epsilon}}\right)}. \tag{19}$$

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

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

Single-store firms If a single-store incumbent exits or a single-store potential entrant stays out, the continuation value is zero. The choice-specific value function from staying active (either entering or remaining in the market) can therefore be expressed relative to the exit choice, denoted e (for an incumbent, it is $a_{it} = l_-$ if $n_{ilt} = 1$, or for a potential entrant $a_{it} = 0$). The choice-specific value function (Equation (17)) can be rewritten for all $a_{it} \neq e$

$$v_{i,t}^{\mathbf{P}}(a_{it}, \mathcal{M}_{j,i,t}) = \pi_i(a_{it}, \mathcal{M}_{j,i,t}) + \beta \sum_{a_{-it}} \int [v_{i,t+1}^{\mathbf{P}}(a_i' = e, \mathcal{M}_{j,i,t+1}) + \gamma - \ln(P_{i,t+1}(e|\mathcal{M}_{j,i,t+1})] dF_t(\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}, a_t) P_{-i,t}(a_{-it}|\mathcal{M}_{j,i,t}) \tag{20}$$

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

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

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

For an incumbent operating a single-store in location l ($a_{it} = 0$) or a potential entrant i who enters into location l ($a_{it} = l_+$), the choice-specific value function is identical to Equation (20).

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

$$
\begin{aligned}
v^{\mathbf{P}}_{i,t}(0,\mathcal{M}_{j,i,t}) &= \pi_i(0,\mathcal{M}_{j,i,t}) + \beta\sum_{a_{-it}}\int[v^{\mathbf{P}}_{i,t+1}(a'_i = l_+,\mathcal{M}_{j,i,t+1}) + \gamma \\
&\qquad - \ln(P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})]dF_{\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}}P_{-i,t} \\
&= \pi_i(0,\mathcal{M}_{j,i,t}) + \beta\sum_{a_{-it}}\int([\pi_i(l_+,\mathcal{M}_{j,i,t+1}) + \gamma - \ln(P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})] \\
&\qquad + \beta^2\sum_{a_{-i,t+1}}\int[v^{\mathbf{P}}_{i,t+2}(a''_i = l_-,\mathcal{M}_{j,i,t+2}) + \gamma - \ln(P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})] \\
&\qquad \times dF_{\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t+1}}P_{-i,t+1})dF_{\mathcal{M}_{j,i,t+1}|\mathcal{M}_{j,i,t}}P_{-i,t}
\end{aligned} \tag{21}
$$

The last equality shows that in period $t+2$, one can obtain an expression for the ex-ante value function by chosing the exit choice $a''_i = l_-$.[16]

5 Market Definition for the Dynamic Game

In our structural analysis, we restrict the sample to small and medium-sized isolated markets for several reasons. First, due to their relatively small number of stores, these markets are most impacted by closure of an existing store. This makes such markets more susceptible to issues of food accessibility. Second, when accounting for spatial differentiation, computational reasons limit the size of markets for which we can solve the dynamic game. Following Seim (2006), we define markets as cities and incorporated places with populations between 5,000 and 150,000, and exclude markets within 10 miles of a city with population greater than $5,000$ or within 20 miles of a city with population greater than $25,000$. Each market is partitioned into locations, which we define at the census tract level.[17]

[16] More generally, for an incumbent chain operating a network of stores $\mathbf{n}_{it}$ and choosing action a in period t, define a sequence of choices $\{a_{t+\tau}\}_{\tau=1}^{|\mathbf{n}_{it}+a|}$, such that in every period $t+\tau$, the chain closes one of its operating stores. Then, the choice-specific value function can be expressed as follows ($|\mathbf{n}_{it}+a|+1$ finite dependence)

$$
\begin{aligned}
v^{\mathbf{P}}_{i,t}(a,\mathcal{M}_{j,i,t}) &= \pi_i(a,\mathcal{M}_{j,i,t}) + \sum_{\tau=1}^{|\mathbf{n}_{it}+a|-1}\beta^{\tau-1}\,\mathbb{E}[\pi_i(a_{i,t+\tau},\mathcal{M}_{j,i,t+\tau}) - \ln(P_{i,t+\tau}(a_{i,t+\tau}|\mathcal{M}_{j,i,t+\tau})|\mathcal{M}_{j,i,t}] \\
&\quad + \beta^{|\mathbf{n}_{it}+a|}\,\mathbb{E}[v^{\mathbf{P}}_{i,t+|\mathbf{n}_{it}+a|}(a=e,\mathcal{M}_{j,i,t+|\mathbf{n}_{it}+a|}) - \ln(P_{i,t+|\mathbf{n}_{it}+a|}(e|\mathcal{M}_{j,i,t+|\mathbf{n}_{it}+a|})]
\end{aligned} \tag{22}
$$

[17] 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 correspondance engine Georr. See https://mcdc.missouri.edu/applications/geocorr2014.html.

The above market definition is motivated by our focus on competition between the three major dollar store chains and three store formats, as defined in the USDA SNAP retailer panel: grocery stores, combination grocery/other stores (stores selling a combination of general merchandise and grocery products), and convenience stores.[18] We combine "grocery stores" and "combination grocery/other" into a single category because the two formats tend to overlap: both sell grocery items but may or may not offer general merchandise as well. Previous literature on retail competition has shown that competition between grocery stores operates at relatively close range (1–2mi). We exclude gas stations, drugstores, and supermarket chains from the set of players. In the case of supermarkets and big box retailer, previous literature (Ellickson and Grieco (2013), Grieco (2014)) has shown that their impact on small horizontally differentiated retailers is limited. Nonetheless, we do control for the presence of the latter three formats (gas stations, drugstores, supermarkets) as potential determinants in players' payoffs, but will treat their evolution as exogenous.

Table 11 shows market and location-level demographic characteristics for our sample of 846 markets and Table 12 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 U.S. as a whole. A market contains 5.8 locations in total, 4.3 of which are "commercial" location while the remainder are "residential." We define "commercial" locations as locations in which at least one store (of any type, e.g., dollar store, gas stations, drugstores, supermarkets) was active in any year between 2008 and 2019. Markets are on average 262 miles away from the closest dollar store chain distribution center.

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. Close examination of the data reveals that 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.

[18]The "combination grocery/other" category includes grocery stores selling other general merchandise, dollar stores and drugstores. We only select the first store format.

Table 11: Descriptive Statistics: Markets and Locations (2010-2019)

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

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

Table 12: Descriptive Statistics: Stores (2010-2019)

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

6 Identification and Estimation

6.1 Identification

As is standard in the literature on the identification of dynamic decision problems (Rust (1994), Magnac and Thesmar (2002), Bajari et al. (2015)), the discount factor and the distribution of firm shocks (β, G) are assumed to be known.[19]

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.[20] When estimating the model, we normalize the exit value to zero. A consequence of this "normalization" restriction is that the estimated entry costs will reflect the true *sunk* costs (entry cost net of exit value), and estimated fixed costs will reflect the true fixed costs in addition to the exit value scaled by $(1 - \beta)$.

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

6.2 Estimation

6.2.1 Baseline estimation approach

We follow a two-step approach. In a first step, consistent estimates of the CCPs, denoted $\widehat{\mathbf{P}}_0$ are obtained. We discuss this first step and the treatment of unobserved heterogeneity in detail in Section 6.2.2. In the remainder of this section, we focus on the estimation of the structural parameters given first-step estimates of the CCPs.

Existing methods for the estimation of dynamic games include policy evaluation (Aguirregabiria and Mira (2007), Pesendorfer and Schmidt-Dengler (2008)) or forward simulation (Bajari et al. (2007)). In our setting, however, firms' dynamic locational choices within a market and the presence of multi-store firms generate a high-dimensional state space. Implementing these approaches would require a combination of state space discretization,

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

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

approximation of the value functions, and numerical integration over the state space.

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

We show that the estimator of Kalouptsidi et al. (2020) developed for *single-agent* dynamic discrete choice problems can be extended to dynamic games. As far as we know, this is the first application of the linear regression estimation approach to dynamic games. In dynamic games with simultaneous moves, a focal firm i's decision in period t affects rivals' decisions in $t+1$ and their endogenous states in $t+2$: this aspect of games, absent in single-agent models, creates a dynamic selection problem for endogenous states. For instance, the number of grocery stores operating in location l at time $t+2$ will depend on a dollar store chains' entry and exit decisions in period t (in the data).[21] The estimator of Kalouptsidi et al. (2020) is modified to account for this selection problem and obtain consistent estimates of the structural parameters. Next, we discuss how moment restrictions are constructed for single-store and multi-store firms.

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

$$v^{\mathbf{P}}_{i,t}(a_{it} = 0, \mathcal{M}_{j,i,t}) = 0 \tag{23}$$

$$v^{\mathbf{P}}_{i,t}(a_{it} = l_+, \mathcal{M}_{j,i,t}) = -\theta^{EC}_i + \beta\,\mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{24}$$

where the expectation is over $\mathcal{M}_{j,i,t+1}$ conditional on $(a_{it} = l_+, \mathcal{M}_{j,i,t})$ and we use finite

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

dependence to express the entrant continuation value in period $t+1$. Combining these two equations gives

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = -\theta_i^{EC} + \beta\,\mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{25}$$

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

$$v_{i,t}^{\mathbf{P}}(l_+, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) = \ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) \tag{26}$$

Combining Equation (25) and Equation (26), we obtain an optimality condition that involves only CCPs at t and $t+1$ and the single-period payoff function

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

This equation includes expected profits and CCPs at $t+1$, and therefore, it appears that numerical integration over the state space is required. However, the equation can be used to construct moment conditions that do not require explicit integration over the space of state variables. Under rational expectations, the conditional expectation at period t of CCPs and profits at $t+1$ is equal to these variables minus an expectational error that is orthogonal to the state variables at period t.[22] Therefore, for any function of period-t information set $h(\mathcal{M}_{j,i,t})$, we have

$$\mathbb{E}\left[h(\mathcal{M}_{j,i,t})u_{it}\right] = 0 \tag{28}$$

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

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

$$
\begin{aligned}
u_{it}(\mathcal{M}^*_{j,i,t+1}) &= \beta\,\mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \\
&\qquad - \beta\left(vp_{i,l}(\mathcal{M}^*_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}^*_{j,i,t+1})\right) \\
&= \ln\left(\frac{P_{i,t}(l_+|\mathcal{M}_{j,i,t})}{P_{i,t}(0|\mathcal{M}_{j,i,t})}\right) + \theta_i^{EC} \\
&\qquad - \beta\left(vp_{i,l}(\mathcal{M}^*_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}^*_{j,i,t+1})\right)
\end{aligned} \tag{29}
$$

where the second equation is obtained by using the expression in Equation (27) to eliminate the expectation term. The moment conditions (Equation (28)) do not require integration over the space of state variables but only averaging over the sample observations. The computational cost of estimating the structural parameters using GMM based on these moment conditions does not depend on the dimension of the state space.

Kalouptsidi et al. (2020) show that, under linearity of payoffs, these moment conditions (replaced by their sample counterparts) can form the basis of a linear IV regression, where period-t variables are used as instruments for period $t+1$ variables.[23] Define the left-hand side variable for potential entrant and incumbent (respectively) as[24]

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

We can obtain the structural parameter via the regression model

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

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

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

Multi-store firms. The estimation procedure for multi-store chains is conceptually similar

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

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

$$
v_{i,t}^{\mathbf{P}}(0, \mathcal{M}_{j,i,t}) - v_{i,t}^{\mathbf{P}}(l_-, \mathcal{M}_{j,i,t}) = \beta\,\mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+1}) - FC_i + \gamma - \ln(P_{i,t+1}(l_-|\mathcal{M}_{j,i,t+1})] \tag{30}
$$

where period-t profits cancel out.

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

To evaluate the robustness of our results, we also implement an alternative estimation approach, in Appendix B, that does not rely on finite dependence and instead solves directly for the ex-ante value function.

6.2.2 Location-level unobserved heterogeneity and first-step estimates

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

We incorporate location-level unobserved heterogeneity via a proxy variable. This approach has been used in previous studies of market entry, e.g., Collard-Wexler (2013), and has the advantage of being computationally light. This is particularly important as a market is partitioned into multiple locations, which may differ in their attractiveness, yielding multi-dimensional unobserved heterogeneity. The proxy variable strikes a balance between granularity in the level of unobserved heterogeneity and computational feasibility. We define a location-level proxy for unobserved heterogeneity as the maximum number of establishments (of all types, including drugstores, supermarkets, and gas stations) *simultaneously* operating in a given location over the period 2008-2019.[25]

The importance of controlling for unobserved heterogeneity is illustrated in Table 13.[26] This table shows estimates of the CCPs for dollar store chains via a flexible multinomial

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

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

logit regression. An entrant chain can either build a store in one of the locations in the market or stay out. An incumbent chain can do nothing, build an additional store in one of the locations, or close one of its existing stores.[27] We control for location-level demographic variables, cost shifters (e.g., 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 are biased upward when business density is not controlled for (column 2) relative to when it is included (column 4). In column 2, many competition coefficients are in fact positive (e.g., for the number of grocery stores and convenience stores within 2mi), reflecting agglomeration effects due to unobserved location-level amenities. This is not the case when location-level business density is included (column 4).

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

6.2.3 Estimation results

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

Table 14 shows estimates of normalized store profits and entry costs. We include a constant term in the profit function to capture the level of fixed costs and/or any baseline level of profits. The effect of most variables decays with distance from the store location, highlighting the importance of spatial differentiation in retail competition. For all retailers, profits are increasing with the population within 2 miles of the store location. Dollar store chains' favor locations with lower income. Profits for chains are decreasing in the distance to the closest distribution center. The majority of competition effects are precisely estimated and with the expected magnitude.

To help interpretation, we convert our profit estimates into dollars by calibrating the scale parameter of firm shocks θ^{ϵ} to match revenue data for all dollar stores operating in the

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

Table 13: Multinomial logit of multi-store firms' choice

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

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

markets under consideration, obtained from Nielsen TDLinx.[28] Table 15 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 chains have substantially lower costs of opening a new store than their independent rivals.

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

Table 14: Estimates of store profits and costs

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

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

Table 15: Mean store profits and marginal effects

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

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

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.[29] Second, the presence of dollar stores also significantly harms convenience store profits, by as much as an additional convenience store. Third, within dollar store store chains, in the 0-2mi radius there is a strong demand cannibalization effect but in the 2-5mi range this effect is reversed and chains benefit from scale economies, likely working through lower operating costs. The location of the market relative to a chain's closest distribution center is also an important determinant of profits. We find that a one standard deviation increase (about 130mi) from the mean distance (190mi) 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 A5 in Appendix E, 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.[30]

7 Counterfactual Policy Analysis

This section uses the structural estimates to conduct counterfactual policy evaluations. We focus on evaluating the impact of dollar store chains' expansion on market structure by sim-

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

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

[30]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.

ulating the evolution of the industry had chains been prevented from expanding beginning in 2010. The counterfactual exercise allows us to quantify the changes in the spatial distribution of retail stores and retail proximity induced by dollar stores' expansion. By controlling for market heterogeneity, we can identify the distributional impacts of dollar store expansion for different market and consumer types.

In practice, we simulate the counterfactual MPE where dollar store chains are prevented from expanding starting in 2010. The counterfactual CCPs obtained are used to simulate the industry forward from 2010 to 2019 in each market. We outline in detail the method used to solve for counterfactual MPE in Appendix C.

Table 16 and Table 17 show how the expected number of grocery stores and convenience stores in 2019 are predicted to change in the counterfactual (CF) relative to the market equilibrium where dollar store chains expanded (Factual) by location and by market respectively. Due to computational limitations, we restrict the exercise to markets with up to 4 locations (4 being the median number of locations per market in our sample).

We find that, in the hypothetical world in which dollar stores did not expand after 2010, markets have on average more than 50% more independent stores, including 54% more grocery stores and 47% more convenience stores. If scaled to all 846 markets we study, the net effect of these changes is that, without dollar store expansion there would be roughly 482 more grocery stores and 747 more convenience stores.

We also examine how these changes vary across different market types. The largest impact on the number of grocery stores occurs in locations (census tracts) with lower income, larger shares of minority populations, higher poverty rates, and fewer households with access to vehicles. The effects are also largest in locations in which dollar stores do not already have a presence in 2010. For convenience stores, the increase in store counts is similar across market and location types.

In Table 18 and Table 19, we consider how changes in market structure translate into changes in retail proximity in 2019. All measures of retail proximity are constructed from data on population at the census block group, i.e., one level down from our definition of locations (census tract). Mean and median statistics are obtained by taking the (population-weighted) mean and median over all census block groups in a market.[31] The first table shows the distance to the nearest store, for the different retail formats, under the factual and counterfactual scenarios. In panel A, distance is conditional on there being at least one

[31] 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.

Table 16: Expected number of stores by location

	Grocery/Combination store counts				Convenience store counts			
	Factual	CF	Δ	% Δ	Factual	CF	Δ	% Δ
All locations	0.411	0.631	0.22	0.323	0.798	1.134	0.336	0.375
By number of commercial locations								
1	0.896	1.473	0.577	0.663	1.419	2.199	0.781	0.657
2	0.483	0.738	0.255	0.351	0.93	1.34	0.411	0.429
3	0.354	0.545	0.191	0.306	0.756	1.059	0.302	0.342
4	0.347	0.521	0.175	0.272	0.649	0.901	0.252	0.329
By income and population								
Population below median, Income below median	0.312	0.441	0.129	0.208	0.757	1.027	0.271	0.3
Population above median, Income below median	0.618	0.997	0.38	0.498	1.108	1.612	0.505	0.447
Population below median, Income above median	0.237	0.317	0.081	0.142	0.53	0.72	0.19	0.292
Population above median, Income above median	0.475	0.765	0.29	0.445	0.794	1.172	0.378	0.464
By share of minority groups								
Above 0.25	0.538	0.863	0.325	0.371	1.097	1.536	0.439	0.345
Below 0.25	0.378	0.571	0.193	0.311	0.72	1.03	0.31	0.383
By share of population with access to vehicle								
Below first quartile (0.89)	0.513	0.816	0.303	0.438	1.069	1.506	0.437	0.386
Above first quartile (0.89)	0.377	0.569	0.192	0.285	0.707	1.01	0.303	0.372
By share of population under poverty line								
Below median (0.16)	0.351	0.529	0.178	0.282	0.646	0.927	0.281	0.381
Above median (0.16)	0.471	0.733	0.262	0.365	0.95	1.341	0.391	0.369
By presence of dollar stores in 2010								
No dollar stores in 2010	0.292	0.409	0.117	0.195	0.63	0.852	0.223	0.307
Dollar stores present in 2010	0.564	0.917	0.353	0.488	1.014	1.496	0.483	0.463
By presence of supermarkets in 2010								
No supermarkets in 2010	0.576	0.859	0.283	0.316	0.816	1.171	0.355	0.433
Supermarkets present in 2010	0.406	0.624	0.218	0.323	0.797	1.133	0.336	0.374

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

store format operating in the market. The column $Pr(n > 0)$ gives the population-weighted probability that a market has at least one store of that format operating.

The results suggest that the expansion of dollar store chains led to a reduction in "access to grocery stores," defined as the probability a consumer has a grocery store operating in their market, from 0.85 to 0.70. In panels B and C, we compute upper and lower bounds on the unconditional distance to the nearest store (assuming a distance of 5 and 7 miles if there are no stores operating in the market). We find that the distance to the nearest grocery store decreases by about 30% under the counterfactual scenario: from 2.30 to 1.68 miles (lower bound), or 2.91 to 2.00mi (upper bound).

Table 19 computes an alternative measure of retail proximity: the number of stores

Table 17: Expected number of stores by market

	Grocery/Combination store counts				Convenience store counts			
	Factual	CF	Δ	% Δ	Factual	CF	Δ	% Δ
All markets	1.079	1.657	0.578	0.544	2.094	2.978	0.883	0.474
By number of commercial locations								
1	0.896	1.473	0.577	0.663	1.419	2.199	0.781	0.657
2	0.966	1.476	0.51	0.53	1.86	2.681	0.821	0.492
3	1.063	1.636	0.572	0.539	2.269	3.176	0.907	0.429
4	1.387	2.086	0.698	0.507	2.596	3.603	1.008	0.409
By income and population								
Population below median, Income below median	1.067	1.648	0.582	0.555	2.299	3.244	0.945	0.468
Population above median, Income below median	1.275	1.996	0.721	0.602	2.359	3.333	0.974	0.452
Population below median, Income above median	0.912	1.374	0.462	0.485	1.668	2.406	0.738	0.502
Population above median, Income above median	1.067	1.618	0.551	0.533	2.023	2.889	0.866	0.473
By share of minority groups								
Above 0.25	1.258	1.972	0.714	0.584	2.639	3.678	1.038	0.427
Below 0.25	1.028	1.567	0.539	0.532	1.939	2.778	0.839	0.487
By share of population with access to vehicle								
Below first quartile (0.89)	1.147	1.794	0.647	0.586	2.537	3.513	0.976	0.411
Above first quartile (0.89)	1.056	1.611	0.555	0.53	1.947	2.799	0.852	0.495
By share of population under poverty line								
Below median (0.16)	0.968	1.476	0.508	0.525	1.73	2.512	0.782	0.513
Above median (0.16)	1.19	1.838	0.648	0.562	2.459	3.443	0.985	0.435
By presence of dollar stores in 2010								
No dollar stores in 2010	1.041	1.67	0.629	0.65	1.901	2.762	0.861	0.534
Dollar stores present in 2010	1.087	1.654	0.567	0.521	2.135	3.023	0.888	0.461
By presence of supermarkets in 2010								
No supermarkets in 2010	1.266	1.889	0.623	0.512	1.795	2.576	0.781	0.496
Supermarkets present in 2010	1.072	1.648	0.576	0.545	2.106	2.993	0.887	0.473

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

within various distance bands of consumers. Consistent with the changes in market structure discussed above, we find that the number of grocery and convenience stores within 2 miles of a consumer increase by 50% approximately.

8 Conclusion

The rise of dollar store chains has profoundly reshaped the U.S. retail sector. By 2021, more than 75 percent of the U.S. population lives within five miles of a dollar store. This rapid expansion has been met by growing scrutiny from policy makers. Proponents claim that dollar stores broaden shopping options and improve convenience for consumers in under-

Table 18: Predicted retail proximity: distance to nearest store

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

Note: Measures of retail proximity are constructed by taking the (population-weighted) mean and median over all census block groups. $Pr(n > 0)$ gives the (population-weighted) mean probability that at least one store is operating in the market. Retail proximity to dollar stores is measured using actual realizations in the data in 2019 for Factual and in 2010 for Counterfactual. Unconditional panels assume that if there are no stores (of a given retail format) operating in the market, the distance travelled is 5mi (lower bound) or 7mi (upper bound).

served areas. Advocates of tighter controls argue that these chains threaten local independent stores, discourage entry by full-line grocery stores, and limit consumers' access to fresh produce. In many municipalities, controls take the form of zoning ordinances that limit dollar store density, with the aim of allowing more grocery stores to enter and operate.

Despite the existing public and policy debates, there is little empirical evidence supporting the arguments advanced. This paper brings new data and methods to bear on these questions. We quantify the impacts of dollar stores' expansion and the effects of entry regulation. Our focus is on two sets of outcomes. First, we consider spatial market structure, that is, the geographic layout and number of retail stores by format. This outcome is particularly important as it affects retail proximity, convenience, and product variety. Second, we consider consumers' shopping behavior, including their spending across retail formats and food categories (e.g., fresh produce).

Our analysis demonstrates that dollar store entry leads to a significant reduction in the number of grocery stores, especially in close proximity (0–2mi) from the entry location. We find that markets lose one grocery store for every three new dollar stores. Dollar store entry and subsequent grocery store exits are associated with lower household spending and volumes purchased of fresh produce. Moreover, we highlight the distributional consequences of these findings: fresh produce consumption and grocery access decrease most for consumers who initially tend to eat less healthfully and have higher transportation costs—low income, older, from minority groups, without access to a vehicle. Overall, we find that negative shocks to food availability have a meaningful impact on consumers' dietary choices. These results add nuance to the literature on the determinants of nutritional inequality.

To better understand how the dollar store model grew so rapidly and the long-run effects of regulating entry, we specify a dynamic model of spatial competition between store formats. The model accounts for the growth in dollar stores' network of distribution centers, and concomitant reduction in store operating costs. The spatial nature of competition introduces non-trivial complexities when estimating and solving this game. Methodologically, we deal with the high-dimensionality of the firms' problem by extending the ECCP estimator of Kalouptsidi et al. (2020) from single-agent problems to dynamic games.

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. Independent grocery and convenience stores, on the other hand, are harmed by the presence of dollar stores at close proximity.

For small and medium-sized markets, our counterfactual exercise establishes that dollar store chains' expansion led to a significant reduction in the number of independent grocery stores, from 1.65 to just over 1 store per market. The largest impact occurs in lower income markets, those with larger shares of minority groups, those with higher poverty rates, and fewer household with access to a vehicle.

The welfare implications of dollar store expansion are arguably multifaceted. Dollar store entry may affect consumer welfare through changes in prices (both at dollar stores and through their competitors' response), changes in store convenience (or travel costs), and changes in product availability and the ensuing composition of consumers' shopping baskets. In the medium to long-run, changes in consumers' dietary choices can have important implications for health outcomes. While the existing data and the many channels above prevent an estimation approach that accounts for all these dimensions in a single model, we view this paper as a necessary first step in quantifying this impact and informing the debate around the place of the dollar store format in the U.S. retail landscape.

Table 19: Predicted retail proximity: number of stores by distance band

	Factual		Counterfactual	
	Mean	Median	Mean	Median
Panel A. All markets				
Number of stores within 0-2mi				
Grocery/Combination stores	0.96	0.74	1.48	1.17
Convenience stores	1.78	1.74	2.54	2.54
Dollar stores	1.87	2	1.24	1
Any store format	4.61	4.63	5.26	5.22
Number of stores within 2-5mi				
Grocery/Combination stores	0.16	0	0.25	0
Convenience stores	0.31	0	0.44	0
Dollar stores	0.36	0	0.22	0
Any store format	0.84	0	0.91	0
Panel B. Markets with two locations				
Number of stores within 0-2mi				
Grocery/Combination stores	0.89	0.70	1.36	1.04
Convenience stores	1.65	1.46	2.39	2.20
Dollar stores	1.74	2	1.19	1
Any store format	4.28	4.10	4.95	4.57
Number of stores within 2-5mi				
Grocery/Combination stores	0.09	0	0.14	0
Convenience stores	0.17	0	0.25	0
Dollar stores	0.16	0	0.10	0
Any store format	0.42	0	0.49	0
Panel C. Markets with three locations				
Number of stores within 0-2mi				
Grocery/Combination stores	0.96	0.75	1.48	1.15
Convenience stores	1.89	1.91	2.67	2.74
Dollar stores	1.91	2	1.25	1
Any store format	4.76	4.81	5.40	5.29
Number of stores within 2-5mi				
Grocery/Combination stores	0.16	0	0.24	0
Convenience stores	0.32	0	0.45	0
Dollar stores	0.35	0	0.20	0
Any store format	0.83	0	0.90	0
Panel D. Markets with four locations				
Number of stores within 0-2mi				
Grocery/Combination stores	1.07	0.88	1.62	1.34
Convenience stores	1.94	2.02	2.70	2.88
Dollar stores	2.15	2	1.41	1
Any store format	5.17	5.52	5.73	6.07
Number of stores within 2-5mi				
Grocery/Combination stores	0.28	0.08	0.42	0.07
Convenience stores	0.50	0.20	0.70	0.24
Dollar stores	0.66	0	0.41	0
Any store format	1.44	0.41	1.52	0.46

Note: Measures of retail proximity are constructed by taking the (population-weighted) mean and median over all census block groups. Retail proximity to dollar stores is measured using actual realizations in the data in 2019 for Factual and in 2010 for Counterfactual.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation."

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.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics* 225 (2): 200–230, Themed Issue: Treatment Effect 1.

Chenarides, Lauren, Clare Cho, Rodolfo M. Nayga, and Michael R. Thomsen. 2021. "Dollar stores and food deserts." *Applied Geography* 134 (June): 102497.

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

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

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

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.

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

Dubé, Jean Pierre, Gunter J. Hitsch, and Peter E. Rossi. 2010. "State dependence and alternative explanations for consumer inertia." *RAND Journal of Economics* 41 (3): 417–445.

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

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

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

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

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

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

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

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* 225 (2): 254–277, Themed Issue: Treatment Effect 1.

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, and Julia Levine. 2022. "Grocery Store Closures and Household Nutritional Choices." *Working Paper*.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225 (2): 175–199, Themed Issue: Treatment Effect 1.

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.

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

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

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

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

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

A Estimation Approach for Multi-Store Firms

This section describes the baseline estimation approach for dollar store chains. Differences in choice-specific value functions for chains are derived as follows.

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

$$
\begin{aligned}
v^{\mathbf{P}}_{i,t}(0, \mathcal{M}_{j,i,t}) = \beta\, \mathbb{E}\big(&-\theta^{EC}_i + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1}) \\
&+ \beta^2\, \mathbb{E}[vp_{i,l}(\mathcal{M}_{j,i,t+2}) - FC_i + \gamma - \ln P_{i,t+2}(l_-|\mathcal{M}_{j,i,t+2})]|0, \mathcal{M}_{j,i,t}\big)
\end{aligned}
\tag{31}
$$

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

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

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

Combining Equation (31), Equation (32), and Equation (33), we obtain an optimality condition that depends only on the structural parameters and the known CCPs.

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

$$
\begin{aligned}
w_{it} = \mathbb{E}&\left[-\theta^{EC}_i + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})|0, \mathcal{M}_{j,i,t}\right] \\
&- \left(-\theta^{EC}_i + \gamma - \ln P_{i,t+1}(l_+|\mathcal{M}_{j,i,t+1})\right)
\end{aligned}
\tag{34}
$$

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

For potential entrants who enter (and become incumbents in $t+1$), we define the expectational error v_{it} as

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

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

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

Replacing these moment conditions by their sample counterparts (in the form of a linear IV regression as in Kalouptsidi et al. (2020)) will not, in general, yield consistent estimates of the structural parameters. Indeed, the error $u_{i,t+1}$ involves an expectation over $t+2$ states, denoted $\mathcal{M}_{j,i,t+2}$, *conditional* on $a_{it} = 0$. But the empirical distribution of $\mathcal{M}_{j,i,t+2}$ (in particular rivals' states) is conditional on the action a_{it} that was played in the data, which may or may not be 0.[32][33]

To address this selection problem, we define the weights

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

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

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

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

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

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

Defining the left-hand side variables for entrant chains as

---

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

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

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

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

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

we obtain the structural parameters for entrants via the regression model

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

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

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

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

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

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

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

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

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

As with entrants, we can dispose of the expectations using the rational expectation assumption and derive moment conditions. As before, period $t+2$ states, appearing in Equation (45), are *conditional* on the action $a_{it} = l_+$. However, the empirical distribution of $\mathcal{M}_{j,i,t+2}$ is conditional on the a_{it} played in the data, which may or may not be l_+. To correct for this selection problem in forming the IV regression equation, any term involving $\mathcal{M}_{j,i,t+2}$ is re-weighted using $\psi_{l_+,\tilde{a}}(\mathcal{M}_{j,i,t+2}|\mathcal{M}_{j,i,t})$, where $\tilde{a}$ is the action played in the data.

B Alternative Estimation Approach

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

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

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

In practice, approximating functions $\phi_{ki}(\mathcal{M}_{j,i,t})$ include all exogenous variables and number of rival and own stores by distance bands and locations.[34] We allow the coefficients λ_k to differ by firm type (e.g., convenience, grocery, dollar store) and by incumbency status.[35]

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

[34] 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.

[35] 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

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,

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

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

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

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

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

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

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

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

be able to estimate profits, the ex-ante value function is, in practice, approximated as

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

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

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

Denote $\boldsymbol{\theta}^{(k+1)}$ the updated structural parameters.[36]

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

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

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

Table A1 shows estimates of the structural parameters based on our baseline estimator (ECCP with finite dependence) and the $1-$MD estimator presented above, using the sample of 846 markets. The two approaches yield very close estimates of entry costs, and profit parameters that are broadly of similar sign and magnitudes. For chains, we note that the magnitude of the (variable) profit coefficients are larger, but marginal effects are of similar magnitude. For example, adding one chain store within 2 miles of an exiting own store reduces mean profits by 11% under ECCP and by 14% under 1-MD, the difference being not statistically significant. If a store is added in the 2-5mi band, mean store profits increase by 9% under ECCP and 14% under 1-MD.[37]

Profits estimates based on the two sets of parameters are of 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.

[36] 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.

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

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

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

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

C Solution Method for the Dynamic Game

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

As the state space is extremely large (with continuous state variables), it is computationally prohibitive to solve for value functions and CCPs at all states. We fix the demographic state variables (income, population, etc.) to their value realized in the data and assume their transitions are deterministic. For periods outside of our sample, i.e, $t \geq T+1$ (where period $T+1$ corresponds to the year 2020), we assume that these demographic variables become stationary and equal the expected value given their realizations in period T.[38] We also maintain chains' distribution networks at their state in period T.

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

1. Initialize the vectors of CCPs for each firm and state $\mathbf{P}_{i,T+1}$. If firm i is a potential

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

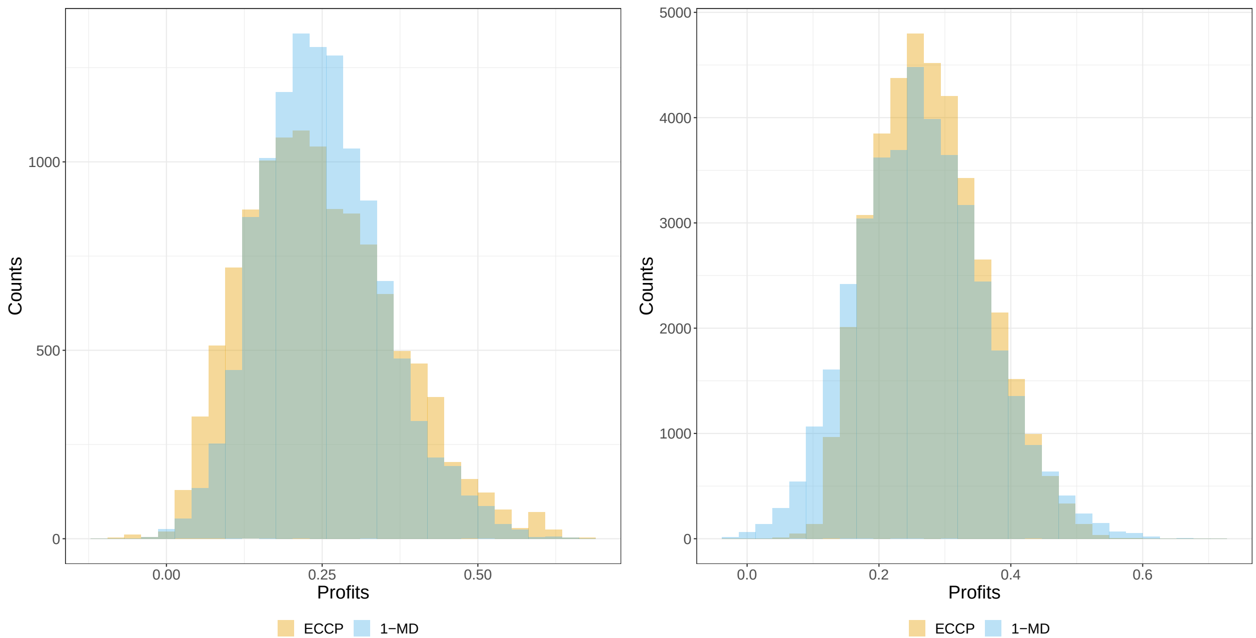


Figure A1: These figures show estimated profits for grocery stores (left) and convenience stores (right) under our baseline estimator (ECCP with finite dependence) and the alternative estimator (minimum-distance with value function approximation).

entrant, $\mathbf{P}_{i,T+1}$ is a vector indexed by the state and locations $(\mathcal{M}_{i,j,T+1}, l_+)$ giving the CCP of entry into location l in state $\mathcal{M}_{i,j,T+1}$.[39] If i is an incumbent, $\mathbf{P}_{i,T+1}$ is a vector indexed by $(\mathcal{M}_{i,j,T+1}, a_{it})$ giving the CCP of choosing action a_{it} (remaining active for single-store firms, or building an additional store/remaing active/closing an existing store for chains) in state $\mathcal{M}_{i,j,T+1}$.

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

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

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

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

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

4. Update the vectors of CCPs as

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

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

This iterative approach yields the equilibrium CCPs for periods $t \geq T+1$, denoted $\mathbf{P}^*_{T+1}$. Proceeding backwards, the equilibrium CCPs in period t, given optimal CCPs in period $t+1$ ($\mathbf{P}^*_{t+1}$) are obtained by iterating over Steps 2 to 4 above, with the exeption that Equation (53) is replaced by

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

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

D Robustness Checks

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

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

The counterfactual results remain qualitatively similar albeit the magnitudes depend on the discount factor: for instance, if dollar store chains were prevented from expanding starting in 2010, the increase in the number of independent stores (per market) ranges from 42% to 76% for grocery stores, and from 37% to 71% from convenience stores.[40]

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 independent market, and their identity is known. Independent grocery and convenience stores are "local" entrants: each firm considers entry only in a single independent market. Moreover, we observe entry

[40]For computational reasons, counterfactual simulations conducted under alternative discount factors were obtained for markets with one and two locations only.

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/combination 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/combination stores and 20% higher for convenience stores relative to the baseline specification. Store profits remain stable across specifications of the number of potential entrants.

E Supplementary Tables and Figures

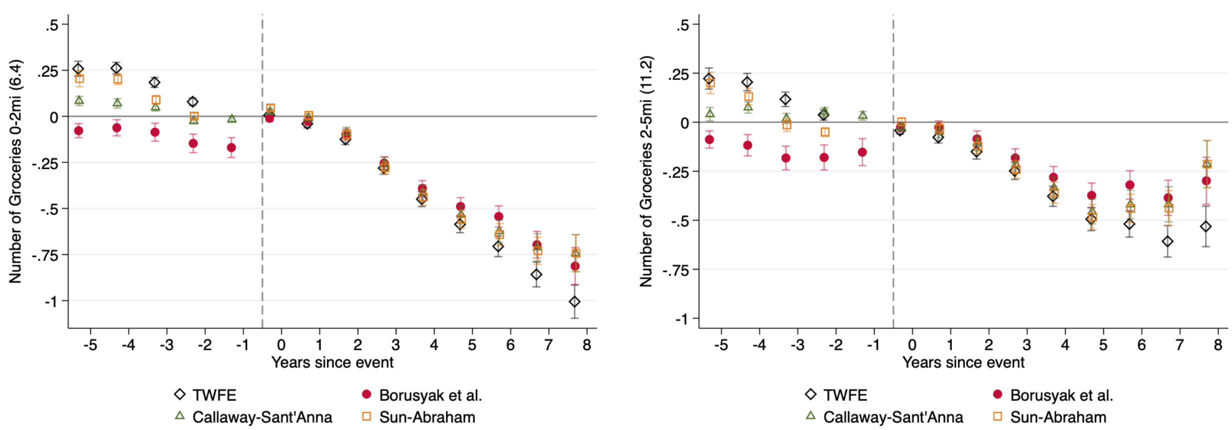


Figure A2: The effects of dollar store entry at the 0-2mi and 2-5mi bands on spending on the number of independent groceries. Results are from an event study analysis on the balanced panel, using a TWFE and heterogeneity-robust estimators proposed by Borusyak et al. (2021), Sun and Abraham (2021), and Callaway and Sant'Anna (2021). For the latter paper, confidence bands show the uniform sup-t confidence intervals adjusted for multiple hypothesis testing.

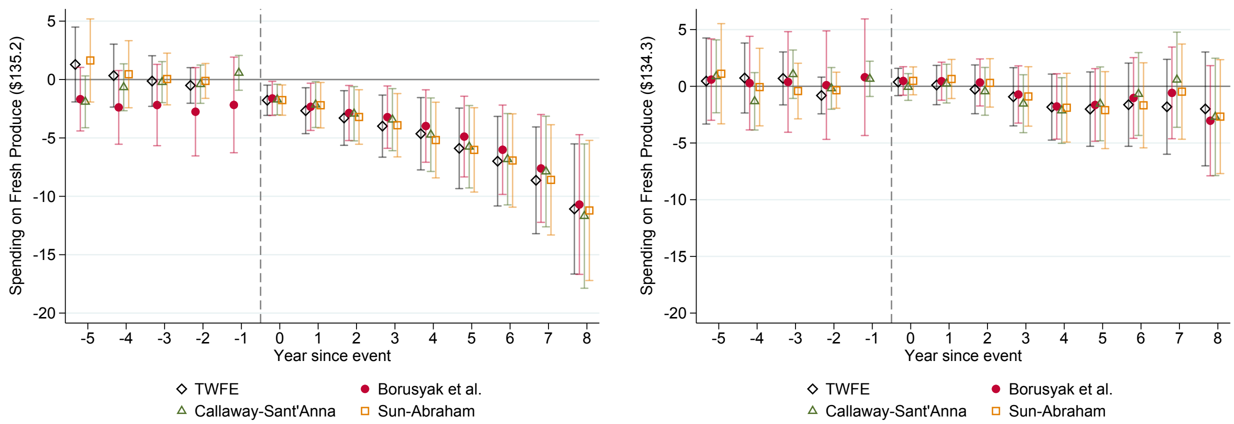
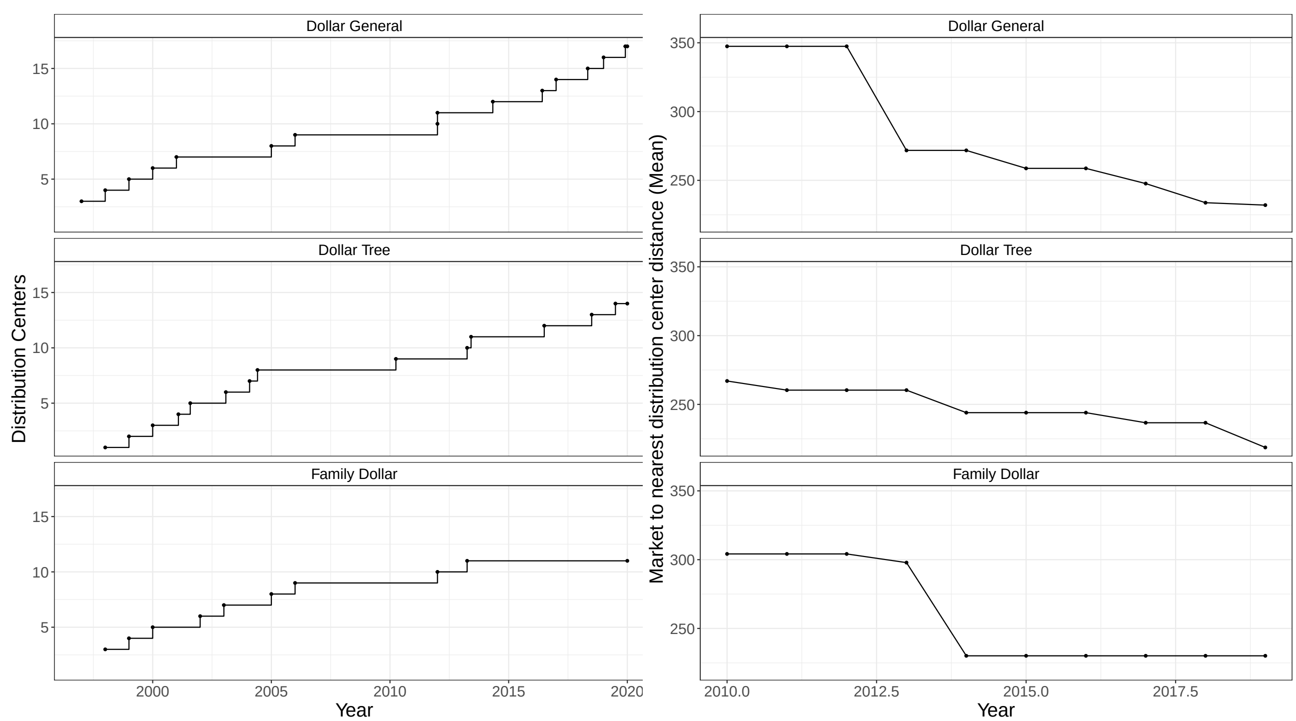


Figure A3: The effects of dollar store entry at the 0-2mi and 2-5mi bands on spending on fresh produce. Results are from an event study analysis on the balanced panel, using a TWFE and heterogeneity-robust estimators proposed by Borusyak et al. (2021), Sun and Abraham (2021), and Callaway and Sant'Anna (2021). For the latter paper, confidence bands show the uniform sup-t confidence intervals adjusted for multiple hypothesis testing.

Figure A4: Distribution centers



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

Table A2: Household Spending by Retail Channel and Food Category in IRI Consumer Network (in $ per year)

Variable	Mean	Median	Std.Dev
Aggregate Spending	2,908.3	2,560.4	1,753.1
Spending in Grocery Channel	1,765.8	1,456.4	1,422.7
Spending on Fresh Produce	88.6	52.7	112.7
Spending on Can/Frozen Produce	63.3	43.3	70.5
Spending on Soda, Snacks, Candy	227.4	159.5	235.4
Spending on Dairy	221.7	169.8	205.7
Spending on Refg./Frozen Meat	127.0	85.3	142.9
Spending on Meals	94.3	66.9	98.9
Spending on Frozen Other	145.5	92.6	175.9
Spending in Dollar Channel	33.7	1.0	108.6
Spending on Fresh Produce	0.6	0.0	7.0
Spending on Can/Frozen Produce	1.2	0.0	6.4
Spending on Soda, Snacks, Candy	14.6	0.0	45.4
Spending on Dairy	1.5	0.0	10.5
Spending on Refg./Frozen Meat	0.9	0.0	8.0
Spending on Meals	2.8	0.0	14.9
Spending on Frozen Other	1.6	0.0	13.1
Spending in SC/Club Channel	767.0	377.4	1,027.2
Spending on Fresh Produce	38.6	7.8	83.4
Spending on Can/Frozen Produce	27.3	6.9	51.8
Spending on Soda, Snacks, Candy	134.8	58.0	206.2
Spending on Dairy	87.0	27.1	142.3
Spending on Refg./Frozen Meat	71.0	20.0	123.8
Spending on Meals	37.7	11.0	66.4
Spending on Frozen Other	57.8	12.7	115.9
Spending in Convenience Channel	14.4	0.0	88.0
Spending on Fresh Produce	0.1	0.0	4.0
Spending on Can/Frozen Produce	0.1	0.0	2.0
Spending on Soda, Snacks, Candy	3.3	0.0	24.2
Spending on Dairy	1.6	0.0	13.4
Spending on Refg./Frozen Meat	0.2	0.0	3.9
Spending on Meals	0.2	0.0	3.5
Spending on Frozen Other	0.4	0.0	6.4
Observations (Household-Year)	618,621		

Note: SC/Club stands for Supercenter and Club store retail channel. Meals includes products such as pasta, rice, and soup.

Table A3: Effect of DS Entry (by distance band) on Total Oz of Fresh Produce, and Total Fat, Sugar, Sodium Purchased

	0-2mi				2-5mi			
	(1) Fresh Produce (Oz)	(2) Fat (g)	(3) Sugar (g)	(4) Sodium (mg)	(5) Fresh Produce (Oz)	(6) Fat (g)	(7) Sugar (g)	(8) Sodium (mg)
First DS Entry	-10.40	-37.04	409.9	-3086.3	0.669	-82.97	-70.19	-2923.6
	(8.656)	(126.2)	(227.8)	(6390.3)	(7.956)	(120.4)	(197.4)	(6116.1)
Two DS Entries	-38.87**	328.5	611.0	10186.6	-14.43	1.263	96.61	518.9
	(14.79)	(216.1)	(457.8)	(10539.3)	(10.90)	(166.7)	(290.6)	(8592.3)
Three+ DS Entries	-55.24*	340.1	-22.77	29150.4	-27.94*	57.51	204.6	5316.9
	(21.45)	(331.4)	(944.9)	(16584.6)	(14.13)	(209.0)	(394.4)	(10478.1)
Observations	246,851	246,851	147,512	246,851	246,851	246,851	147,512	246,851
F-stat	8.47	29.3	10.4	23.2	8.49	29.3	10.3	23.1
R^2	0.75	0.80	0.87	0.71	0.75	0.80	0.87	0.71
Adjusted R^2	0.72	0.77	0.84	0.68	0.72	0.77	0.84	0.68
Mean Pre-Entry	1,444	38,133	60,586	1,460,188	1,426	39,528	62,367	1,515,818

Notes: Unit of observation is the household-year. Standard errors (in parenthesis) clustered at the household level. Year and household fixed effects are included. Controls for time-varying household demographics (income, education, age, household size, marital status, occupation, weekly hours worked) are included. Results are shown for the balanced panel.

Figure A5: Distribution center locations in 2019

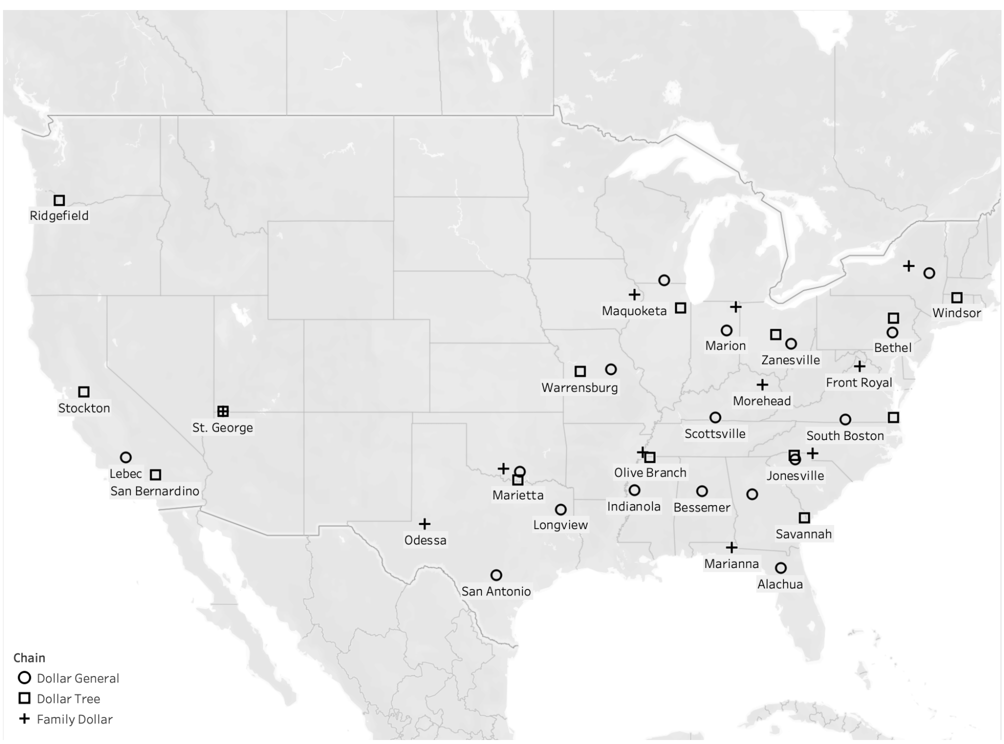


Table A4: Multinomial logit of single-store firms' choice

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

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

Table A5: Estimates of stores profits and costs: alternative specifications

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