THE UNIVERSITY OF CHICAGO

SUPERSTAR FIRMS AND CONSUMER WELFARE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
ZHONGLIN LI

CHICAGO, ILLINOIS
JUNE 2022

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation for my dissertation committee: Milena Almagro, Marianne Bertrand, Chad Syverson (chair) and Thomas Wollmann. I benefited tremendously from their continuous guidance, support, and invaluable advice during the course of my PhD study. I would also like to thank numerous faculty members at the University of Chicago for their helpful suggestions and plentiful insights, especially Eric Budish, Pradeep Chintagunta, Giovanni Compiani, Michael Dinerstein, Jean-Pierre Dubé, Ali Hortaçsu, Emir Kamenica, Sarah Moshary, Matthew Notowidigdo, and Canice Prendergast. I am particularly grateful to Justin Hilyin Leung, my co-author and friend, for many fruitful discussions, his kind co-operation and support for our shared work and beyond.

My appreciation also goes out to my friends and classmates, for their encouragement and a cherished time spent together. I wish to express my deepest thanks to my parents, whose contributions to my life can never be acknowledged with a few words.

ABSTRACT

This dissertation comprises two essays on superstar firms and consumer welfare in the US retail sector. The first essay investigates how the entry of big-box stores affects household consumption and welfare. It presents empirical evidence that after supercenters and warehouse clubs enter, households change various dimensions of their shopping behavior in ways that are strongly consistent with store characteristics. It further provides a novel multi-store multi-category choice model to quantify and disentangle the effects of product variety, prices, and other store characteristics on consumer welfare, highlighting the importance of variety and one-stop shopping experience.

The second essay focuses on market concentration in retailing. It documents a rise in household retail concentration and increased one-stop shopping over the past two decades and explores the driving forces behind. On the supply side, increasing local availability of superstar retailers, rises in product variety, and changes in pricing explain a portion of these trends. On the demand side, increases in households' opportunity cost of time are also key drivers. A model is developed to rationalize these results and provide the implications for market power and welfare.

INTRODUCTION

Superstar firms have continually attracted the attention of researchers, policymakers, and the public. A large literature has documented increases in market concentration in the US economy over the past several decades, meaning a small fraction of firms - superstars, have been accounting for a greater share of economic outcomes. One strand points to changes in technology and economies of scale for firms as the dominant explanation.[1] Another strand mainly attributes these trends to firms exercising their rising market power.[2] As these views lead to divergent welfare implications, it is crucial to understand the driving forces behind the rising superstar firms, and how it links to market power and welfare.

This dissertation explores these questions by zooming into the US retail sector. Specifically, it focuses on retailers that sell consumer packaged goods (CPG), and provides direct evidence on how they interact with households and affect consumer welfare by utilizing detailed household-level data.

Similar to the broader economy, the US retail sector has witnessed significant changes in recent decades. National market concentration has risen substantially. From 2004 to 2019, the market share of the top four retail chains has increased from 25% to over 37%.[3] This trend coincides with a boom of certain physical formats of retailers such as warehouse clubs and supercenters (Hortaçsu and Syverson 2015), and the expansion of superstar national retail chains into more geographic locations. As a result, concentration measures at finer levels of geography have only increased slightly or decreased substantially (Hsieh and Rossi-Hansberg 2019; Rossi-Hansberg, Sarte and Trachter 2021).

However, there can still be a gap between a low market-level concentration and a competitive market. A unique feature of the retail sector is the ability to engage in one-stop

1. For example, see Autor et al. (2020), Benkard, Yurukoglu and Zhang (2021), Ganapati (2021), and Kwon, Ma and Zimmermann (2021) among others.

2. For example, see Covarrubias, Gutirrez and Philippon (2020), De Loecker, Eeckhout and Unger (2020), and Grullon, Larkin and Michaely (2019) among others.

3. These numbers are calculated using the Nielsen Consumer Panel data. It covers major categories in the retail sector, including drug, grocery, and merchandise stores. Details are in Section 1.3.

shopping. Over the past two decades, households have been doing more one-stop shopping, with greater spending and more varieties of products purchased per shopping trip. They are also spending more in their preferred stores, i.e. a rise in the retail concentration within households. The increasing heterogeneity in the choice of retailers across households allows the negligible change in market-level concentration over time. [4] It is yet ambiguous whether more one-stop shopping or higher household retail concentration is pro- or anti-competitive.[5]

These shopping patterns are potentially related to the expansion of superstar chains like Walmart and Costco. These supercenters and warehouse clubs are big-box stores that sell a wide variety of products and allow households to do one-stop shopping. There has been a striking growth in this retail format over the past few decades. From 2004 to 2019, the number of supercenters grew by about 240% and warehouse clubs by 35%. In 2019, supercenters and warehouse clubs account for 9% of the total retail sales in the US. Chapter 1 explores in detail how the entry of this retail format alters households' shopping behavior and affects consumer welfare, with an emphasize on the multi-category feature of this retail format and response in household demand.

In Chapter 2, we move beyond the superstar firms. We first document detailed facts on household-level consumption, including rising household retail concentration and increased one-stop shopping. Then we investigate both supply- and demand-side factors to provide causal evidence to quantify various underlying mechanisms. Supply-side factors include the entry of superstar firms, changes in prices and varieties in stores, while demand-side factors include opportunity cost of time. We further develop a model to rationalize these results and discuss implications on market power and welfare.

The two essays in this dissertation provide some insights on superstar firms, market power and welfare in the retail sector. Moving forward, I think we can explore more in the following three directions for the future research on this topic. First, features beyond

---

4. See details in 2.4.

5. Thomassen et al. (2017) discusses one-stop shopping and competition, and show one-stop shopping is pro-competition in the UK grocery industry.

the price in competition. Especially for new superstars or retail formats, features other than prices such as convenience, variety, can play important roles in affecting consumer behavior and competition. Second, demand-driven revolution and firm growth. For example, households' growing opportunity cost of time may boost the growth of firms that provide more convenience, which would change the market structure. It would be interesting to quantify welfare changes due to such demand shifts. Thirdly, distributional effects in the evolution of retail formats. As households exhibit great heterogeneity in choosing stores, the trade-off across household groups in the development of a particular superstar firm may closely relates to winners and losers in the process.

CHAPTER 1

BIG-BOX STORE EXPANSION AND CONSUMER WELFARE

written jointly with Justin Leung

1.1 Abstract

Supercenters and warehouse clubs have grown rapidly in the US in recent decades. These big-box retail establishments are physically large to enable one-stop shopping, offering a broad range of product categories with relatively low prices. In this paper, we study how the entry of these big-box stores affect household consumption and welfare. We first present an event study of the store entries of four major big-box retail chains to provide empirical evidence that households change various dimensions of their shopping behavior, such as product varieties per shopping trip and prices paid, in ways that are strongly consistent with store characteristics. We then develop a novel multi-store multi-category choice model to quantify and disentangle the effects of product variety, prices, and other store characteristics on consumer welfare. We show that households benefit substantially from consuming in supercenters relative to competing retailers, highlighting the importance of the store format.

1.2 Introduction

The US has experienced striking changes in big-box store formats over the past few decades. From 2004 to 2019, the number of one major type of big-box store, supercenters, grew by about 240% to more than 4000 (Figure 1.1a). Another type of big-box store, warehouse clubs, also stands out with a 35% growth in the number of stores during the same period (Figure 1.1b).[1]

1. Major supercenters include Walmart Supercenter, Super Target, Meijer etc. Major warehouse clubs include Costco, BJ's, Sam's Club etc. Hortaçsu and Syverson (2015) also documents the growth of both supercenters and warehouse clubs since 1990 using employment data.

What makes these big-box stores unique? These retail establishments are physically large and are usually designed for one-stop shopping, offering a broad range of product categories with relatively low prices. For example, supercenters combine general merchandise with groceries, providing a large assortment of consumer packaged goods (CPG). Club stores usually operate membership warehouses that offer low prices on a wide range of categories despite a relatively limited assortment, aiming to generate high sales and rapid inventory turnover.

Despite the conveniences and low prices that big-box stores provide, members of local communities often organize to block proposals of big-box store entries, citing reasons such as negative effects on local businesses, competition, and employment. Many local governments have enacted store cap ordinances to constrain store sizes (Zhou 2017).[2] However, given that variety is highly correlated with store size, these regulations would limit consumers' one-stop-shopping experience that precisely differentiates big-box stores and potentially affects the prices consumers face.

How do households actually respond to the entry of big-box stores? How does big-box stores' ability to provide substantial variety mediate these responses and affect consumer welfare? In this paper, we address these questions by first providing new empirical evidence on the effects of big-box store expansion on households. We focus on supercenters and warehouse clubs because these two store formats have greatly shaped the retail sector in recent decades (Hortaçsu and Syverson 2015). Our results have implications for related types of big-box stores that are driven by similar mechanics. We then use these reduced-form results to motivate a novel structural demand model to disentangle the welfare impact of store entries, with an emphasis on product variety.

We utilize an event-study approach to estimate the impact of big-box store entries on households' shopping behavior. Combining hundreds of thousands of households' shopping trip records from the Nielsen Consumer Panel (henceforth HMS) with data on opening dates

2. Bertrand and Kramarz (2002) and Sadun (2014) study the effects of entry regulation in Europe.

and locations of four major big-box chains, we present three facts on the effect of big-box store entries.

First, households substitute toward big-box stores from other stores. We find a sharp hike in average households' spending share at a big-box store after its entry, for both supercenter and warehouse club entries. Concurrently, households' spending share in other stores declines, with grocery stores being the most affected. This finding suggests a reallocation of expenditure toward big-box stores, which is consistent with Arcidiacono et al. (2020), who show revenues drop in incumbent stores after the entry of Walmart Supercenters.

Second, households change the number of product categories purchased per shopping trip. After a supercenter entry, households purchase more categories per trip from the supercenter, but fewer categories per trip from other stores. This finding implies complementarities across categories within store due to one-stop shopping. Surprisingly, we find households do more multi-stop shopping after a club store enters. This change may relate to the strategy of warehouse clubs, which is to provide a limited assortment within a wide range of categories.

Third, for all categories, households pay lower prices than the national average, with food seeing the most notable drop. This effect lines up with the low-price strategy of big-box stores, which is supported by previous research (Hausman and Leibtag, 2007; Ellickson, Misra and Nair, 2012; Ailawadi, Ma and Grewal, 2018).

To quantify the effects of big-box store entry in a way that simultaneously incorporates substitution across stores, cross-category complementarities, and price changes that our empirical analysis reveals, we develop a multi-store multi-category demand model. We build on a category-choice model in Mehta and Ma (2012) by adding a multi-store choice that nests category-level choices. Our model allows households to visit up to two stores each week to fulfill their shopping needs.[3] At the category level, for each choice of stores, households determine the quantity for each product category based on price, store-category characteristics,

3. This assumption is not very restrictive, because households rarely visit more than two stores per week in our sample.

and budget. We use a non-homothetic indirect translog utility form that allows for flexible complementarity patterns across categories and income effects. Households weigh the utility generated at the category level against trip costs for each store choice, and pick the stores that achieve the highest total utility.

We estimate the model with a sample of households from the HMS. We obtain quality-adjusted price measures that allow us to compare prices for a given category across stores, holding quality fixed. We find that supercenters have relatively low quality-adjusted prices in all four major categories: food, non-food grocery, health and beauty care, and general merchandise. Club stores have relatively low quality-adjusted prices in food and non-food grocery, but not in the other categories.

Next, we quantify cross-category complementarities by using a counterfactual analysis on category exits in big-box stores. We show that different categories in the same store are substitutes, holding store choice fixed, but they exhibit more complementarities when households can freely choose between stores. This effect is stronger in supercenters than in club stores, implying the convenience of one-stop shopping is more prevalent in supercenters.

Both quality-adjusted prices and cross-category complementarities determine the value of each category and its contribution to consumer welfare. To quantify the value of these categories, we run two counterfactuals. First, we remove all categories from a store and calculate the welfare loss, which we denote as the consumption value of a store. Second, we remove each category separately from the store and calculate the welfare loss from each category exit. Our estimates suggest each category accounts for a substantial share of the consumption value of a supercenter. For example, general merchandise in a supercenter accounts for 29% of the total consumption value, even though the spending share in this category is only 7.6%. We also notice that the contribution of each category in a supercenter sums up to greater than 1, which suggests a diminishing effect of removing each category when more categories are removed. The reason is that removing a category also lowers spending in other categories and the overall probability of visiting that store. For club stores, we find this amplification effect only applies to food. The contribution of each category sums up to less than 1, which implies diminishing returns of adding each category. In addition, we compare the effect of category provision and price increase. We find that welfare loss due to a 10% increase in the price of a category is only 1.3% of the welfare loss due to the exit of a category.

Finally, we compare the total value of big-box stores with other store types. A supercenter generates a consumption value about twice that of most store types, with one-third of the gain from savings in trip costs. A club store mostly generates the same value as other stores. Therefore, even though both supercenters and club stores offer a wide range of product categories, supercenters generate substantially higher welfare gains for consumers through their lower quality-adjusted prices and cross-category complementarities, highlighting the importance of a one-stop-shopping experience.

Our paper adds to several strands of literature. First, it enriches a literature on the impact of big-box stores, as summarized in Carden and Courtemanche (2016) and Ellickson (2016). These papers study the effects of supercenters, and to a lesser extent warehouse clubs, on a wide variety of outcomes such as prices, competition, welfare, market structure, labor markets, sociocultural effects, and health.[4] However, less is known about how household shopping behavior changes at big-box stores and competing retailers after big-box store entries.[5] In this paper, we provide new empirical evidence on how households change various dimensions of their shopping behavior, such as product varieties per trip, in response to both supercenter and club store entries. We further develop a novel structural model that provides a microfoundation for store-category choice, which allows us to simultaneously quantify

4. For example, Jia (2008) and Holmes (2011) assess the density economies of Walmart with a focus on competition and market structure. Ailawadi et al. (2010) examine incumbent retailers reactions to a Wal-Mart entry and the impact of these reactions on the retailers sales. More recent work includes Arcidiacono et al. (2020), who estimate competitive price effects. Atkin, Faber and Gonzalez-Navarro (2017) provide empirical evidence on the impact of Walmart entry in Mexico and estimate large welfare gains for households.

5. One exception is Hwang and Park (2016), who estimate the impact of Walmart supercenter conversions on household shopping behavior and find an increase in per-visit expenditures drives revenue gains in Walmart. They also find evidence of increases in category-level spending in nine pre-existing product groups, particularly for food categories.

various sources of welfare gains from these store entries. This model can also explain the differential effects of supercenters and club stores.

Second, we contribute to a literature highlighting the importance of cross-category effects in consumer demand using structural models. Microfounded demand systems that accommodate cross-category complementarities typically do not model the store-choice process (Song and Chintagunta, 2007; Bhat, 2005; Mehta and Ma, 2012; Mehta, 2015). To incorporate our empirical evidence on household shopping behavior across stores, we build on Mehta and Ma (2012) and develop a unified multi-store multi-category demand model by adding a multi-store choice that nests category-level choices. The model is discrete in store choice but continuous in quantity choice for each category, while allowing for corner solutions.[6] To the best of our knowledge, the only paper that has a similar store-category-choice design is Thomassen et al. (2017).[7] They allow for a single-store choice for each category and estimate the model using UK consumer data to quantify cross-category pricing effects due to one-stop shopping, finding that internalizing them substantially reduces market power. In contrast to their model, our model allows multi-store choice for each category and income effects.[8] Also, we focus on the effects of category exit and consumer welfare.

Third, our paper contributes to a rich literature on concentration and market power as summarized in papers such as Berry, Gaynor and Scott Morton (2019) and Syverson (2019). These papers highlight that many economists and policymakers are expressing concern over the possibility of increasing monopoly power in the US and the world economy. A large and growing literature documents rising market concentration since 2000 or earlier.[9] How-

6. This allows us to account for the fact that many households do not purchase all categories in the same week.

7. Seo (2019) uses a purely discrete store-category-choice model to study the welfare impact of allowing liquor sales at grocery stores and shows large gains for consumers due to one-stop shopping. We develop a discrete-continuous two-level choice model while including broader categories and provide a general framework to study the impact of increased firm scope.

8. Multi-store choice for each category fits better with our data because we use four departments as our categories, whereas they analyze product groups within the grocery department. We also consider income effects as they directly affect welfare calculations.

9. One strand points to changes in technology and economies of scale for firms as the dominant expla-

ever, the theoretical relationship between market concentration and average market power is ambiguous. Many empirical studies find patterns of simultaneous concentration and productivity growth. Berry, Gaynor and Scott Morton (2019) and Syverson (2019) argue that the case for large and general increases in market power is not yet dispositive. These papers call for a surge in industry-level research to characterize heterogeneity more fully both across and within markets, suggesting that sources of these patterns may be multi-causal, all with potential implications for market power in possibly different directions. We complement Leung and Li (2021*b*), who use detailed micro-data on firms and consumers in the retail sector to decompose rising concentration. Whereas that paper provides causal evidence to quantify various underlying mechanisms driving retail concentration such as store entry, variety, and pricing, this paper focuses on the welfare impact of superstar retailers that sell CPG.

The rest of the paper is organized as follows. Section 1.3 describes the data. Section 1.4 presents reduced-form analysis on the entry of big-box stores. Section 2.6 estimates a multi-store multi-category choice model of households. Section 1.6 provides counterfactuals and welfare analysis. Section 2.8 concludes.

1.3 Data

1.3.1 Nielsen Consumer Panel

The Nielsen Consumer Panel Dataset (HMS)[10] represents a longitudinal panel of approximately 40,000 to 60,000 US households from 2004 to 2019 who continually provide information to Nielsen about their households and what products they buy, as well as when and

nation. For example, see Autor et al. (2020), Benkard, Yurukoglu and Zhang (2021), Ganapati (2021), and Kwon, Ma and Zimmermann (2021) among others. Another strand mainly attributes these trends to firms exercising their rising market power. For example, see Covarrubias, Gutirrez and Philippon (2020), De Loecker, Eeckhout and Unger (2020), and Grullon, Larkin and Michaely (2019) among others.

10. Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

where they make purchases.[11] Panelists use in-home scanners to record all their purchases, from any outlet, intended for personal, in-home use. Products include all Nielsen-tracked categories of food and non-food items, across all retail outlets in the US. Nielsen samples all states and major markets. Panelists are geographically dispersed and demographically balanced.

Panelists report the products they purchase in each shopping trip. For each product as defined by its universal product code (UPC), we know the quantity purchased and total price paid for all units. Over 5 million products are classified into about 1,100 product modules, 125 product groups, and 10 product departments, which allows us to calculate varieties at various levels. We further group products into five departments: food, non-food grocery, health and beauty care, general merchandise, and others.

A de-anonymized retail-chain identifier is specified for each trip and a channel type of each retail chain is provided. Major channel types in our analysis are discount store, warehouse club, grocery store, drug store and dollar store. We also observe where the household resides at various geographic levels from the Nielsen Scantrack market level (Nielsen classifies regions into around 50 market areas) down to the level of county and 5-digit zip code. Because the location of each shopping trip is not revealed, we assume households visit the closest store of each retail chain.

1.3.2 Store Locations

We obtain the store locations and opening dates of four big-box chains including Walmart supercenters and three warehouse club chains: Costco, Sam's Club, and BJ's. Data of Walmart supercenter openings are from Arcidiacono et al. (2020) and covers 2004-2013. Data of club store openings are from Coibion, Gorodnichenko and Koustas (2021) and covers 2004-

11. The data are available through a partnership between NielsenIQ and the James M. Kilts Center for Marketing at the University of Chicago Booth School of Business. Information on access to the the consumer panel data as well as the retail scanner data described below is available at `http://research.chicagobooth.edu/nielsen/`.

2015. These data allow us to conduct event studies to study the impact of big-box stores.

We use the National Establishment Time Series (NETS) dataset from 1990-2019, which covers the location of the universe of stores in the US to construct distance measures for structural estimation.[12]

We also use the Nielsen TDLinx data from 2004-2019 to obtain monthly-level store counts for each retail chain at the market level.

1.4 Empirical Analysis on Big-Box Store Entry

In this section, we utilize an event-study approach to present empirical evidence on how households change various dimensions of their shopping behavior when big-box stores enter. We document three main facts. First, households substitute toward big-box stores from other stores. Second, households change the number of product categories purchased per shopping trip. Third, households pay lower prices. These facts are consistent with the assortment and pricing strategies of supercenters and warehouse clubs.

1.4.1 Empirical Strategy

We utilize an event-study approach to estimate the impact of the entry of big-box stores. Our baseline independent variable measures the number of stores for each chain within the 5-digit zip code of each household. As shown in equation (2.1), we then regress outcomes of interest Y_{it} for household i in quarterly period t, on the number of stores $Num_{m(i)t}$ for the 5-digit zip code $m(i)$ that household i resides in, and add household fixed effects to control for fixed household characteristics, as well as period fixed effects to control for national time

12. For each household-store pair, we calculate the distance between store locations and the population centroid of the 5-digit zip area where the household lives. We match retail chains in the HMS to stores in the NETS conditioning on channel type and geographical distribution.

trends:

$$Y_{it} = \beta \times Num_{m(i)t} + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{1.1}$$

A store entering in periods when unobservable local household characteristics change or households anticipating these openings by changing patterns in significant ways would be a threat to our identification. A priori, we believe households or stores have difficulty exactly timing sharp changes in unobservables with store entry. To further alleviate these concerns, we estimate the trends before and after the entry event by adding leads and lags of the independent variable $Num_{m(i)t}$. If the pre-trends are parallel, we argue this finding provides additional evidence that stores or households cannot align changes in unobservables to the precise timing of the entry.

1.4.2 Results

We document three main facts on the effects of big-box store entries. First, we find households substitute toward big-box stores from other stores. We show the effects of supercenter and club store entry on spending share in Table 1.1 and Table 1.2, respectively. We find a sharp hike of about 6 percentage points (p.p.) in average households' spending share for supercenters after its entry, and about 3.5 p.p. for clubs. Concurrently, households' spending share in other stores declines, with grocery stores being the most affected, decreasing by 3.6 p.p. and 2.3 p.p. for supercenter and club entries, respectively. This finding suggests a reallocation of expenditure toward big-box stores.

In Figure 1.2 and Figure 1.3, we show pre-trends are parallel around store entries roughly two years before the event, whereas the effects are dynamic and continue to rise for an extended period after the event. This observation is consistent with both households taking time to learn about the presence of new stores and adjusting their purchasing habits.[13]

13. We also show in Appendix Figure 1.13 that spending shares are increasing in both supercenters and club stores in all product departments over the sample period.

Second, we find households change the number of product categories purchased per shopping trip when big-box stores enter. We show the effects of supercenter and club store entry on varieties per trip, as measured by the number of product departments, by store type in Table 1.3 and Table 1.4, respectively. After a supercenter entry, households increase product departments per trip from supercenters by about 0.006. Households also purchase fewer varieties per trip from other stores. This finding suggests complementarities across categories within supercenters due to one-stop shopping. Surprisingly, we find households purchase fewer varieties per trip when club stores enter. This finding may relate to the strategy of warehouse clubs, which is to provide a limited assortment within a wide range of categories.

We also measure varieties at different levels, starting from the number of UPCs, which is the lowest level, to the number of product departments, which is the highest level. We show the effects of store entry on the number of UPCs per trip by store type in Appendix Tables 1.17 and 1.18, which show the same patterns qualitatively. We display results for each level of variety in Appendix Tables 2.3 and 1.20. Consistent with the fact that varieties per trip decrease when club stores enter, we find households do more multi-stop shopping, with the number of trips increasing significantly by about 0.023 per quarter. We also find the number of retailers visited decreases when supercenters enter, but increases when club stores enter in Appendix Table 1.21 and Table 1.22, respectively. We show that pre-trends are again parallel in Figures 1.4, 1.5, and 1.14.

To provide suggestive evidence of why supercenters generate a different effect from club stores, we show how supercenters and club stores differ in product variety. In Figure 2.8, we show the average number of UPCs and departments per household-quarter for supercenters, club stores, and other channel types over the sample period. The number of UPCs can capture variety depth of a store, whereas the number of departments can capture variety breadth. Households buy far more varieties in supercenters than any other channel type, whereas club stores sell fewer UPCs but a similar number of departments compared with grocery stores and supercenters. The number of UPCs in club stores is closer to other channel types such as discount stores and dollar stores, but higher than drug stores and other miscellaneous channel types. This would be consistent with the hypothesis that supercenters with larger variety depth and breadth allow households to engage in more one-stop shopping, whereas club stores focus mostly on variety breadth but less on depth.

Third, we find that households pay lower prices when big-box stores enter. We calculate the relative price index (RPI) of each household following Aguiar and Hurst (2007*a*):

$$RPI_{it} = \frac{\sum_{j \in J_{it}} p_{jit} q_{jit}}{\sum_{j \in J_{it}} \bar{p}_{jt} q_{jit}}, \tag{1.2}$$

where p_{jit} and q_{jit} are the price and quantity for UPC j for household i at time t, respectively, and $\bar{p}_{jt}$ is the average national price for UPC j. In other words, to construct a household RPI, we calculate the ratio between total expenditure and the counterfactual expenditure of each good at its average price in the reference region.

We show the effects of supercenter and club store entry on household RPIs in Table 1.5 and Table 1.6, respectively. RPIs decrease by about 0.5% when supercenters enter. This drop holds for all product categories, with health and beauty care and food seeing the most notable drop. RPIs decrease by about 0.15% when club stores enter, although this result is statistically insignificant. We show pre-trends are again parallel in Figures 1.7 and 1.8. Several reasons can explain these price decreases. First, given the low-price strategy of big-box stores, which is supported by previous research (Hausman and Leibtag, 2007; Ellickson, Misra and Nair, 2012; Ailawadi, Ma and Grewal, 2018), households may now be able to enjoy the lower prices of supercenters for the same products they previously consumed. Second, households may shift their consumption bundle to products with lower RPIs. Third, prices of other stores may decrease as a competitive response to supercenter entry. However, Arcidiacono et al. (2020) find supercenter entry has no causal effect on incumbent prices.

To provide suggestive evidence of why supercenters generate a different effect from club stores, we show how supercenters and club stores differ in prices. In Figure 2.9, we calculate the RPI of each retailer or channel type. We find supercenters consistently offer lower prices than their competitors nationally, although their price advantage has been decreasing. Although club stores generally have an RPI below one, they tend to have higher RPIs than supercenters over the sample period, with the exception of Club 2 offering lower prices in the last periods, which our entry data do not capture. These findings would be consistent with our finding that supercenters generate larger price decreases due to their lower prices.

1.5 Structural Model and Estimation

To quantify the effects of big-box store entry in a way that incorporates substitution across stores, cross-category complementarities, and price changes that our empirical analysis reveals, we develop a multi-store multi-category demand model for households and use maximum likelihood estimation (MLE) to recover the parameters.

1.5.1 Model

Consider a household's weekly shopping decision that includes a store-choice decision for each shopping trip and a quantity-choice decision for each product category. The decision process for the household is as follows:

1. For each given store choice, the household chooses the quantity for each product category such that the utility of purchasing in this store choice is maximized.

2. Households weigh the utility generated at the category level against trip costs for each store choice, and pick the stores that achieve the highest total utility.

Let $\mathcal{S}$ denote the set of available stores that the household faces. We allow households to choose one store or two stores to shop per week. Let $\mathcal{R}$ be the set of one- and two-store choices. For each element $r \in \mathcal{R}$, it can be one store $\{s\}, s \in \mathcal{S}$, or a set of two stores $\{s_1, s_2\}, s_1, s_2 \in \mathcal{S}$. The possible choices for the weekly store-visit decision thus include $|\mathcal{S}|$ one-store choices and $\frac{|\mathcal{S}|(|\mathcal{S}|-1)}{2}$ two-store choices.

We consider M focal categories of products that households may have shopping needs for. Each store sells at least one product category and heterogeneity exists for the same category across different stores. Let $\mathcal{M}_r$ denote the set of store-product categories that are available in store-choice option $r \in \mathcal{R}$. If r only includes one store, $\mathcal{M}_r$ is the categories available in that single store. If r includes two stores, $\mathcal{M}_r$ includes the categories in both stores. We use $j \in \mathcal{M}_r$ to denote each store-category (e.g., "Costco-Food"). $c = c(j)$ denotes the product category of j (e.g., "Food") and $s = s(j)$ denotes the store of j (e.g., "Costco").

We introduce category 0 as a composite good of products outside our M focal categories that households may purchase in the same week. We also introduce store 0 as an outside option to allow households to shop outside focal stores $\mathcal{S}$. We assume all the products sold in store 0 belong to category 0.

For a given store choice $r \in \mathcal{R}$, the household makes purchase decisions across categories $\mathcal{M}_r$ to maximize the utility with a fixed weekly budget y:

$$\begin{aligned} V_r = & \max_{Q_j} \; U\left(\psi_1 Q_1, ..., \psi_{M_r} Q_{M_r}, Q_0\right) \\ & s.t. \sum_{j \in \mathcal{M}_r} P_j Q_j + P_0 Q_0 = y, Q_j \geq 0, j \in \mathcal{M}_r, Q_0 > 0. \end{aligned} \tag{1.3}$$

Category j's price is given as P_j, and Q_j is the quantity to be determined. ψ_j indicates the quality for category $j \in \mathcal{M}_r$, which we further explain below.

The household's overall utility from shopping weekly is determined by both the utility from consuming at the chosen store(s), namely V_r, and both observed and unobserved costs of shopping trips to the store(s), Γ_r and ν_r, respectively. The household chooses the store set $r \in \mathcal{R}$ that maximizes the overall utility:

$$\begin{aligned}
&\max_{r \in \mathcal{R}} \ \gamma^v \ln V_r + \Gamma_r + \nu_r \\
\text{where } V_r = &\max_{Q_j} \ U\left(\psi_1 Q_1, ..., \psi_{M_r} Q_{M_r}, Q_0\right) \\
&s.t. \sum_{j \in \mathcal{M}_r} P_j Q_j + P_0 Q_0 = y, Q_j \geq 0, j \in \mathcal{M}_r, Q_0 > 0.
\end{aligned} \tag{1.4}$$

1.5.1.1 Category-Level Decision

We first solve the category-choice problem within each store choice following Mehta and Ma (2012). We define quality-adjusted prices and quantity as $P_j^* = P_j/\psi_j$, $Q_j^* = Q_j\psi_j$, for $j \in \mathcal{M}_r$ and $P_0^* = P_0$, $Q_0^* = Q_0$. The category-choice problem within store choice r hence becomes

$$\begin{aligned}
V_r = &\max_{Q_j^*} \ U\left(Q_1^*, ..., Q_{M_r}^*, Q_0^*\right) \\
s.t. &\sum_{j \in \mathcal{M}_r} P_j^* Q_j^* + P_0^* Q_0^* = y, Q_j^* \geq 0, Q_0^* > 0.
\end{aligned} \tag{1.5}$$

Note we allow zero consumption of the focal categories and assume the composite good 0 is always purchased.[14] The household thus needs to make two decisions at the category-choice level: purchase incidence, that is which categories to purchase, and quantity for purchased categories.

As mentioned in Mehta and Ma (2012), the utility-maximization problem can be solved using two approaches. In the first approach, we can specify a strictly increasing and quasi-concave functional form for direct utility U and solve for a set of Kuhn-Tucker (KT) conditions to get the optimal quantities and purchase incidence, which is employed by Thomassen

14. In the baseline model, we do not allow zero consumption of category 0. A small proportion of observations in the data violate this assumption and we manually adjust the share of category 0 to be 0.001 for these observations. The model could be extended to accommodate zero consumption of category 0, similar to how we allow zero quantity for focal categories, but doing so would add extra computational burden for optimization and is thus not adopted in the baseline model.

et al. (2017). In the second approach, we can give a functional form for the indirect utility V that corresponds to a strictly increasing and quasi-concave direct utility U, and derive optimal quantities and purchase incidence using methods introduced by Lee and Pitt (1986). We follow Mehta and Ma (2012) to use the latter approach and apply a nonhomothetic indirect translog utility (let category $M_r + 1 =$ be category 0):

$$\ln V_r = -\sum_{j=1}^{M_r+1} a_{c(j)} \ln \frac{P_j^*}{y} + \frac{1}{2} \sum_{j=1}^{M_r+1} \sum_{k=1}^{M_r+1} b_{c(j)d(k)} \ln \frac{P_j^*}{y} \ln \frac{P_k^*}{y}. \tag{1.6}$$

This indirect utility form has several advantages. First, it allows us to solve for demand functions explicitly and remains flexible enough to approximate general utility functions. Second, it allows flexible complementarity patterns across different product categories $c(j)$ and $d(k)$. Third, the non-homothetic design allows income effects from budget changes, which can be important in welfare estimation and also allows us to explore heterogeneous effects across different income levels.

Assume focal categories $j \in \{m+1, ..., M_r\}$ are purchased and $j \in \{1, ..., m\}$ are not purchased. For the purchased categories, we invoke Roy's identity to solve for the budget share $\{S_j\}$:

$$S_j(\{P_j^*\}_{j=1}^{M_r}, y) \equiv -\frac{\partial \ln V_r(\{P_j^*\}_{j=1}^{M_r+1}, y)/\partial \ln P_j^*}{\partial \ln V_r(\{P_j^*\}_{j=1}^{M_r+1}, y)/\partial \ln y}, \ \forall j \in \{m+1, ..., M_r\}. \tag{1.7}$$

Quantity can be calculated using $Q_j^* = y \times S_j$. For non-purchased categories, we use virtual price T_j instead of P_j^* in the indirect utility function. The virtual price T_j is defined as the price such that $S_j = 0$ is an interior solution for category j. We solve for $\{T_j\}_{j=1}^m$ from the following equation:

$$S_j(\{T_j\}_{j=1}^m, \{P_j^*\}_{j=m+1}^{M_r}, P_0^*, y) = 0, \ j = 1, ..., m. \tag{1.8}$$

Given purchase incidence $\{I_j = 0\}_{j=1}^{m}, \{I_j = 1\}_{j=m+1}^{M_r}$, where $I_j = 1$ indicates purchased categories, the budget shares and quality-adjusted prices should satisfy the following conditions:

$$S_j(\{T_j(\{P_j^*\}_{j=m+1}^{M_r}, P_0^*, y)\}_{j=1}^{m}, \{P_j^*\}_{j=m+1}^{M_r}, P_0^*, y) > 0, \qquad j = m+1, ..., M_r. \tag{1.9}$$

$$T_j(\{P_j^*\}_{j=m+1}^{M_r}, P_0^*, y) \leq P_j^*, \qquad j = 1, ..., m. \tag{1.10}$$

The first condition says the solution for purchased categories should be strictly greater than zero, and the second condition requires the actual quality-adjusted price to be greater than or equal to the calculated virtual price such that it is too high for households to purchase any products from the category.

Using the functional form of indirect utility, we are able to specify the solutions for budget shares and purchase incidence conditions of all combinations of purchase and non-purchase decisions. The household chooses the purchase incidence that gives the highest indirect utility.

Next, we specify a detailed functional form for quality-adjusted price P^*. Recall that $P_j^* = P_j/\psi_j$. We assume quality ψ_j depends on both store-category observed and unobserved characteristics. For each store-category j in category c and store s, we define

$$\ln \psi_j = (\alpha_{s(j)} + \lambda_{c(j)} + \rho X_j + \varepsilon_j)/\mu_{c(j)}. \tag{1.11}$$

Store fixed effect $\alpha_{s(j)}$ accounts for store-specific characteristics including store amenities, retailer reputation, common features of products sold in the store (e.g., products in club stores in general have larger package size), etc. Category fixed effect $\lambda_{c(j)}$ captures common categorical tastes of consumers. For example, households usually have higher demand for food products than for health and beauty care products. X_j captures store-category characteristics. In the baseline model, we define X_j as the number of UPCs per store-category, which measures the variety depth of each store-category. We expect $\rho > 0$ because households usually enjoy more varieties of products.

ε_j is store-category-household-week specific and describes households' needs and tastes that are not observed by researchers. We assume ε_j is i.i.d. and follows a standard extreme-value distribution. $\mu_{c(j)} > 0$ is the scale parameter that varies across categories to allow different spreads of tastes in each category.

Hence, we can write the quality-adjusted prices with full subscripts for household i in week t and store-category j as follows[15]:

$$\ln P^*_{ijt} = \ln P_{jt} - \frac{\alpha_{s(j)} + \lambda_{c(j)} + \rho X_j + \varepsilon_{ijt}}{\mu_{c(j)}}. \tag{1.12}$$

We also assume for category 0 that $\ln P^*_{i0t} = \varepsilon_{i0t}$, where ε_{i0t} is i.i.d. and follows a standard normal distribution. This outside-option category 0 serves as a benchmark for other store-categories.

1.5.1.2 Store-Level Decision

Given indirect utility V_{irt} for store choice $r \in \mathcal{R}$ from the category-level decision, household i in week t chooses r that gives the highest total utility:

$$\max_{r \in \mathcal{R}} \mathcal{U}_{irt} = \gamma^v \ln V_{irt} + \Gamma_{ir} + \nu_{irt}. \tag{1.13}$$

For the observed trip costs Γ_r, we specify the following functional form:

$$\Gamma_r = \gamma_g(\gamma^1 D_r + \gamma^2 I^{two} + \gamma^3 I^{same}). \tag{1.14}$$

D_r is the distance between stores and households. When choice r has two stores, we use the sum of the distances for each store-household pair. γ^1 is expected to be less than

15. We can add another set of shocks $\{\epsilon_{ict}\}$ to address household-week-category-specific needs that are common for all stores. We have estimated the model with one set of such shocks, and our reported results are not altered significantly. We thus exclude it from our baseline model for computational considerations.

0 because a greater travelling distance suggests higher trip costs. I^{two} is 1 when a two-store option is chosen. Besides longer traveling distances, extra fixed costs may arise if the household chooses to visit two stores, so we expect $\gamma^2 < 0$. Because households may choose a two-store option because the two stores are close or are convenient to visit together, we include a dummy I^{same} equal to 1 if any households in the same market visit the store pair on the same day. Thus, γ^3 should be greater than 0, and $\gamma^2 + \gamma^3$ is the benefit when two stores can be visited together. Different households may have different sensitivity to trip costs. We use the set $\{\gamma_g, g = 1, ..., G\}$ to account for household-group heterogeneity, where we normalize $\gamma_1 = 1$ for group 1. For the outside option store 0, we assume its trip costs to be γ^0.

We include ν_{irt} as the unobserved trip costs for each store choice r of household i in week t. $\{\nu_r\}_{r \in \mathcal{R}}$ are i.i.d. across household-week-store choices, and follow a standard extreme-value distribution.

1.5.2 Estimation and Empirical Strategy

In this section, we introduce the sample for estimation and a two-stage approach to estimate the parameters.

1.5.2.1 Sample for Estimation

To overcome computational burdens, we choose households' purchase records in Texas, the state with the most observations in our data, in 2012 from HMS. To restrict supply-side changes, we exclude county-quarters with big-box store entries and counties with substantial changes in the number of retailers observed during the year. Because we assume each household-week is independent of each other in our model, we randomly sample three weeks per quarter per household. Our final sample includes 11,376 household-week observations from 1,137 households across 57 counties. Each household has at least three weeks of observations.

We have 17 focal chains in our analysis: eight grocery chains, four discount chains (including the major supercenter), two club chains, two drug store chains and one dollar store chain by assuming all the dollar stores are the same. Total spending share in the 17 chains is greater than 80% for all counties, suggesting these chains cover the majority of households' store choice options. All the spending outside the 17 focal chains belongs to outside option store 0.

We define a choice set for each household. We assume households only visit the closest store within the same chain, and thus do not differentiate between "chain" and " store" in the following analysis. A chain is included in the choice set of a household if any household living in the same county ever purchases in the chain during the year. For two-store combinations in the store-choice set $\mathcal{R}$, we include store combinations that have been visited in the same week by any household. As a result, the number of different chains ranges from 4 to 13 and number of elements in choice set $\mathcal{R}$ is between 10 and 91.

Our analysis includes products from four categories: food, non-food grocery, health and beauty care, and general merchandise. These four categories accounts for more than 96% of households' total spending in HMS. Any spending outside the four focal categories is treated as the outside option and goes to category 0. For the baseline estimation, we use a national average of biweekly prices for each chain-category, which is assumed to be exogenous from household-week-store-category unobservables. Specifically, category prices are calculated using the average price for each UPC in the category weighted by sales, and only UPCs that are sold in all biweeks in 2012 are included. Within each store-category, price variation over time solely comes from price changes within products as products and weights are fixed. Thus, we are able use this variation to estimate price elasticities. We also count the number of UPCs per store-category for the 17 chains because a measure of variety depth per store-category. A summary table of prices and variety depth are provided in Appendix Table 1.23.

Table 1.7 summarizes household demographics and shopping behavior. Households' weekly budget for CPG is \$270.6 on average.[16] One store is visited in 42% of the household-weeks and two stores are visited in 47%. The spending share of total budget is 88% when households visit one store and 90% when they visit two stores. If it is a two-store shopping week, the spending share in the second store is substantially smaller than the first store, accounting for only 25% of the weekly budget.[17] About 11% of the observations have no shopping trips to the focal stores. We thus assume the outside-option store 0 has been chosen for these observations. The outside option means a household does not visit any of the 17 focal stores during a week. We group households into eight groups by income quartile interacted with whether they have children under age 18.

1.5.2.2 Estimation Strategy

Two sets of parameters are estimated. From the category level decision, we have $\Theta = \{\alpha_s, \lambda_c, \rho, B = \{b_{cd}, b^2_{cd}\}, \mu_c, s \in \mathcal{S}, c, d = 1, ..., M+1\}$. From the store-level decision, we have $\Theta^S = \{\gamma^v, \gamma^1, \gamma^2, \gamma^3, \{\gamma_g\}, \gamma^0\}$. Ideally, we want to estimate all the parameters simultaneously because the random shocks at both levels affect both store-level and category-level decisions. However, due to the computational burden that mainly comes from solving for the optimal quantity and purchase incidence for each store choice in each iteration during the optimization, we first use a two-stage estimation to get estimates for our baseline analysis.[18]

In a two-stage estimation of the parameters, we first estimate parameters at the category level decision: $\Theta = \{\alpha_s, \lambda_c, \rho, B = \{b_{cd}, b^2_{cd}\}, \mu_c, s \in \mathcal{S}, c, d = 1, ..., M+1\}$. For store fixed effects $\{\alpha_s\}$, we normalize the fixed effect of chain 17 to 0, that is $\alpha_{17} = 0$, and estimate 16 parameters. For category fixed effects $\{\lambda_c\}$, we estimate four parameters for the

16. Weekly budget is calculated based on the annual household income bracket reported in HMS and the ratio of consumer goods expenditure to total expenditure from the Consumer Expenditure Survey in 2012 divided by 52 weeks.

17. A small share of household-week observations have more than two stores visited. We only include the top two store visits in terms of spending and the spending share in other stores are typically less than 10%.

18. We write the likelihood function for simultaneous estimation in Appendix 1.10.1.2. Based on our results using simulated data, the estimates from the two-stage approach are only slightly biased.

four categories. Note we normalize $\sum_{c=0}^{M} a_c = 1$ in equation (1.6), because the spending share of all categories including category 0 sums up to 1. Because λ_c and a_c cannot be separately identified for each category, we further set $a_c = 0, c = 1, ..., M$ to allow only one category fixed effect in each category. Thus, we have $a_0 = 1$ for category 0. Matrix B includes parameters in equation (1.6) that describe the complementarity across store-categories. For categories within the same store, we capture the complementarity patterns using the parameters $\{b_{cd}, c, d = 1, ..., M+1\}$, where $M+1$ stands for category 0, and assume these parameters are the same for all stores. For categories in two different stores that can be purchased in the same week, we capture the complementarity patterns using the parameters $\{b^2_{cd}, c, d = 1, ..., M\}$, and assume these parameters are the same for all two-store combinations. $\{\mu_c > 0, c = 1, ..., M\}$ are scale parameters for each category, which are greater than 0 by definition. We provide more details on the parameters in Appendix 1.10.1.1.

We use MLE to recover the estimates. At the category-level decision, the observed data for each household i and week t are the purchase incidence in the chosen store r ($\{\{I_j = 0\}_{j=1}^{m}, \{I_j = 1\}_{j=m+1}^{M_r}\}_{it}$) and spending share for purchased categories in the chosen store r ($\{\{S_j\}_{j=m+1}^{M_j}\}_{it}$). Purchase incidence and spending share are functions of the unobserved $\{\varepsilon_{ijt}\}$ for $j \in \mathcal{M}_r$. Their relationship is derived from equations (1.9) and (1.10). Using the distribution of $\{\varepsilon_{ijt}\}$ for $j \in \mathcal{M}_r$, we can write the log-likelihood function for each household i and week t observing purchase incidence $\{\{I_j = 0\}_{j=1}^{m}, \{I_j = 1\}_{j=m+1}^{M_r}\}_{it}$ and budget share $\{\{S_j\}_{j=m+1}^{M_j}\}_{it}$[19]:

$$l_{it}\left(\{\{I_j\}_{j=1}^{M_r}, \{S_j\}_{j=1}^{M_r}\}_{it}|\Theta\right) = \ln\left(\int_{\varepsilon_0=-\infty} L_{r(it)}\left(\{\{I_j\}_{j=1}^{M_r}, \{S_j\}_{j=1}^{M_r}\}_{it}|\Theta\right)\phi(\varepsilon_{0,it})d\varepsilon_{0,it}\right). \tag{1.15}$$

In the estimation, we use Gauss-Kronrod quadrature to integrate out the ε_0's. The

19. The details of the likelihood function at the category level are provided in Appendix 1.10.1.3.

log-likelihood for the entire sample with N_{obs} observations is

$$l(\Theta) = \frac{1}{N_{obs}} \sum_{i=1}^{N} \sum_{t=1}^{T} l_{it} \left(\{ \{I_j\}_{j=1}^{M_r}, \{S_j\}_{j=1}^{M_r} \}_{it} | \Theta \right). \tag{1.16}$$

At the second stage, we estimate $\Theta^S = \{\gamma^v, \gamma^1, \gamma^2, \gamma^3, \{\gamma_g, g = 2, ..., 8\}, \gamma^0\}$ given estimates from the first stage $\hat{\Theta}$ using MLE. The indirect utility $\ln V_r$ depends on both $\hat{\Theta}$ and random shocks $\{\varepsilon_{0,it}\}$ and $\{\boldsymbol{\varepsilon}_{it}\}$ for all the store-categories. We use Monte Carlo methods to draw all the shocks from the truncated distribution of $\{\boldsymbol{\varepsilon}_{it}\}$ and $\{\varepsilon_{0,it}\}$ and predict $\ln \hat{V}_r = \ln \hat{V}_r(\boldsymbol{\varepsilon})$ for each store choice $r \in \mathcal{R}$ for each household-week.[20] Given the distribution of $\{\nu_{irt}\}$, the likelihood of observing store choice $\{I_r^S = 1, I_{r'}^S = 0, r' \neq r\}$, purchase incidence $\{I_j = 0\}_{j=1}^{m}, \{I_j = 1\}_{j=m+1}^{M_r}$ and budget share $\{S_j\}_{j=m+1}^{M_j}$ for each observation can be written in logistic form:[21]

$$\hat{L}_{it} \left(\{I_{r'}^S\}_{r' \in \mathcal{R}}, \{I_j\}_{j \in \mathcal{M}_r}, \{S_j\}_{j \in \mathcal{M}_r} | \hat{\Theta}, \Theta^S \right) = \frac{\exp(\gamma^v \ln \hat{V}_r(\boldsymbol{\varepsilon_{it}}) + \Gamma_r)}{\sum_{r' \in \mathcal{R}} \exp(\gamma^v \ln \hat{V}_{r'}(\boldsymbol{\varepsilon_{it}}) + \Gamma_{r'})}. \tag{1.17}$$

The log-likelihood for the entire sample with N_{obs} observations is

$$l(\Theta^S) = \frac{1}{N_{obs}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r \in \mathcal{R}} I_r^S \ln \left(\hat{L}_{it} \left(\{I_{r'}^S\}_{r' \in \mathcal{R}}, \{I_j\}_{j \in \mathcal{M}_r}, \{S_j\}_{j \in \mathcal{M}_r} | \hat{\Theta}, \Theta^S \right) \right). \tag{1.18}$$

1.5.3 Estimates and Analysis

From stage one, we report the estimates for $\hat{\Theta}$ at the category level in two parts. First, we present parameters in the quality-adjusted price in Table 1.8. We see the coefficient ρ for the variety-depth measure is greater than 0. Quality-adjusted price for the category decreases when variety depth improves. This finding suggests that when a category has more product varieties, the overall quality of the category improves such that households are more willing

20. The distribution is truncated because category-level data impose restrictions on these shocks.

21. The details of the likelihood function at the store level are provided in Appendix 1.10.1.4.

to purchase it with a given category price.

We calculate the expected quality-adjusted price for each store-category averaging across household-weeks. This quality-adjusted price has two useful features. First, it is comparable between stores within each category. In the definition of store-category prices in section 1.5.2.1, the magnitude of store-category prices is not comparable across store-categories, because different products are included in different store-categories for the price measure. However, after we make a quality adjustment by using store fixed effects, category fixed effects, and store-category variety depth, the quality of products in different stores within a category are accounted for. Second, the quality-adjusted price changes at the same rate as the store-category price, because the log store-category price enters the log quality-adjusted price linearly with coefficient 1 (equation (1.12)). Percentage changes are the same for those two terms. This finding provides convenience in calculating price elasticity and analyzing price effects.

Figure 1.10 presents the relative log quality-adjusted price $\ln P^*$ across stores for each category when $\ln P^*$ of the supercenter is normalized to 1. Results suggest that relative costs of product categories after controlling for quality across store types line up with our understanding of these store types. For the food category, both supercenter and club stores offer products at a relatively low quality-adjusted price. The only type that provides an even lower quality-adjusted price is grocery stores. Other types, including regular discount stores, have a higher quality-adjusted price. For health and beauty care products, the two drug stores dominate all other stores. Note that for categories other than food, supercenter and other discount stores have similar quality-adjusted prices. This similarity can be explained by the fact that a supercenter is a regular discount store plus a grocery department. The results also show that even though all four categories are sold in different types of stores, the price for households can be very different, given quality. For big-box stores, supercenters have relatively low quality-adjusted prices in all categories, whereas club stores only have a relatively low quality-adjusted price for food and non-food groceries.

Second, we show the estimated parameters that describe complementarity patterns across categories. Table 1.9 displays $\{b_{cd}, c, d = 1, ..., M + 1\}$ for categories within the same store. Positive numbers for a category pair suggest the two categories are complements and negative numbers suggest substitutes. We further calculate average cross-category price elasticities for individual demand conditioning on store choice and purchase incidence (Table 1.10). The relatively small own-price elasticities suggest households are price elastic in all categories. Note the price elasticity for food is smaller than -1, whereas past research often suggests less elastic food demand (Andreyeva, Long and Brownell, 2010). Two causes are possible. One is that we allow substitution across categories.[22] The other is related to the time frame of our analysis, which is weekly. Households may substitute intertemporally across weeks. For cross-category patterns, the majority are substitutes for each other, except that non-food grocery complements other categories. All the categories are substitutes for the outside option, which is consistent with the definition of the outside option, that is, any spending outside the four focal categories. We report complementarity patterns for categories across stores in Appendix Tables 1.24 and 1.25.

Estimates from the second stage are displayed in Table 1.11. Households' utility declines when distance increases and when they need to visit two stores separately. However, if two stores can be visited on the same day, no extra cost is incurred. The fact that estimates of $\gamma^3 - \gamma^2$ are greater than 0 suggests households may visit multiple stores on the same trip. High-income households have higher trip costs and are more sensitive to distance than low-income households. This finding may imply a higher value of time for high-income households. Trip costs and the effect of distance for households with no children do not differ significantly from those for households with children.

22. Our food-category price elasticity is comparable to unconditional elasticities in Okrent and Alston (2012), in which substitution across categories is allowed.

1.6 Consumer Welfare

In this section, we estimate the value of each product category in a store and disentangle the welfare impact of big-box store entries. The value of each product category depends not only on the quality-adjusted price of the category, but also on complementarity patterns across categories. In section 1.6.1, we first study cross-category complementarities in big-box stores by examining quantity changes across categories after one category is removed. In section 1.6.2, we then compare the value of a category with the value of a store with two counterfactuals. First, we remove all categories from a store and calculate the welfare loss, which we denote as the consumption value of a store. Second, we remove each category separately from the store and calculate the welfare loss from each category exit. In section 1.6.3, we compare the total consumption value of big-box stores with other store types. Lastly, we discuss the assumptions and potential extensions of the consumer welfare analysis in section 1.6.4.

1.6.1 Cross-category Complementarity

We quantify cross-category complementarities by counterfactually removing categories in big-box stores. We show that different categories in the same store are mostly substitutes holding store choice fixed, but they exhibit more complementarities when households can freely choose between stores.

Table 1.12 exhibits the percentage change in quantity in each category after one category is removed from a supercenter. Panel A shows the change conditioning on the fact that the supercenter is chosen by households. Households are thus only allowed to adjust their quantity for each category in the supercenter. Panel B gives the change when households are further allowed to switch their consumption to other stores.[23] For a supercenter, if its general merchandise category is removed, which makes the supercenter more similar to its grocery

23. Note that the diagonal cells are the percentage quantity change of the exiting category, which is -1 by definition.

competitors, the quantity of food purchase by a household in the supercenter increases by 1.1% on average. This increase suggests food and general merchandise are substitutes. However, if we also consider the probability of choosing the supercenter, expected food demand decreases by -0.9%, which makes food and general merchandise complements. This complementarity comes from the fact that households one-stop shop for these categories in the same store. Once the probability of choosing the store becomes lower due to the exit of a category, the expected quantity purchased from other categories also declines.[24] For category pairs that remain substitutes in Panel B, we also see a decline in their substitutability. Total quantity purchased in the supercenter also drops more for the unconditional case (7.4%) than the conditional case (5.8%).

Panel C presents households' expected consumption in all the other stores under the unconditional case. When the general merchandise category in the supercenter is removed, surprisingly, households also reduce the spending in the same category in other stores. This suggests that the general merchandise category in the supercenter and other stores are on average complements instead of substitutes. This makes sense as households can have their light bulbs and lamps purchased in different stores. Once lamps become unavailable, they do not need to purchase light bulbs. Note that it is a novel feature that our model allows for complementarity for the same category across stores. Otherwise, such complementarity will not be captured. We also notice that the overall spending in all the stores declines once one category is removed. Households would rather spending more on outside goods or save than switching to other focal stores to make a purchase.

The cross-category complementarity patterns are similar for a club store, but we see a higher level of substitution across most categories within the store in general (Table 1.13). We also notice that the difference between the conditional case and the unconditional case for a club store is smaller than that of a supercenter. For general merchandise exit in a club

24. Thomassen et al. (2017) show a similar cross-category complementarity for product groups in food due to price changes, while our results on category exit can be interpreted as a large price increase such that no one would purchase the category.

store, the percentage change in total demand from the club store drops by 7% conditional on store choice but the number is only slightly larger, 7.1%, when households can switch stores. This finding means the exit of general merchandise has little impact on the probability of choosing the store and the convenience of one-stop shopping, due to the provision of this category being limited.

1.6.2 Value of a Category

To quantify the value of each category in a store, we run two counterfactuals. First, we remove all categories from a store and calculate the welfare loss, which we denote as the consumption value of a store (EV^s). When all the categories of a store are removed, the utility from the category-level decision $\ln V$ becomes the same as that for the outside store s_0, which provides the outside category only. Because the outside option is always available for all households, adding a store that gives the same $\ln V$ as the outside option does not offer households a higher value in terms of purchasing and consuming products. Thus, in this case, such a store has zero consumption value.[25] Second, we remove each category separately from the store and calculate the welfare loss from each category exit (EV^c). Both welfare losses are calculated as the equivalent variation (EV): that is, we calculate the percentage reduction in the budget of a household that is required to achieve the same loss in utility as removing categories. We then calculate the share of the value of a category in the total consumption value of a store ($EV\ Share = EV^c/EV^s$).

We present the value of the categories of a supercenter in the first row of Table 1.14. Each category accounts for a substantial share in the consumption value of a supercenter. For example, general merchandise accounts for 29% of the total consumption value, even though the spending share in this category is only 7.6%. Food has the highest spending share among the four categories (58%), and it is also the most valuable category of a supercenter,

25. In our model, the value of the outside store s_0 is not zero, because it also sells product category 0 that is valued by households. By defining the consumption value, we are still making the outside option the benchmark.

accounting for 86% of the total consumption value. We also notice the contribution of each category in a supercenter sums up to greater than 1, which suggests a diminishing effect of removing each category when more categories are removed. The reason is that removing a category also lowers spending in other categories and the overall probability of visiting that store.

We also examine the impact of incorporating income effects in our model by comparing *EV Share* and $\Delta \ln V$ *Share*. The latter is the ratio of the utility loss when one category exits to the utility loss when all categories exit, which is shown in the third row of Table 1.14. In a model with no income effects and the budget going linearly into the utility such as a quasi-linear utility, *EV Share* and $\Delta \ln V$ *Share* should be the same, which implies the difference between these two shares reflects additional adjustments driven by income effects. For illustration, consider a budget reduction with income effects. Households can adjust both their quantity choice at the category level and their store choice to re-optimize given the new budget and achieve a lower utility reduction than the scenario without income effects. Hence, *EV* becomes larger in order to have the same amount of utility reduction. This adjustment occurs for both of the counterfactuals that we consider, which determines both the numerator and the denominator of these shares. Because *EV Share* is greater than $\Delta \ln V$ *Share* for all categories, this finding suggests such an adjustment is more substantial for one category's exit than for all categories' exits. This finding is consistent with the intuition that when three categories are still available, more room exists for adjustment when the budget declines.

For club stores, we see that only the value of food takes a considerable share of the total consumption value (73%) and exceeds its spending share (63%); Table 1.15). The contribution of the other three categories ranges from 1.1% to 11%, whereas their spending share ranges from 7.7% to 16%. Additionally, the contribution of each category sums up to less than 1, which implies diminishing returns of adding each category. In other words, a category is more valuable when it is added as the first category than when it is added as the fourth category. This finding can be related to the substitutability of the categories. When categories are more substitutable, the importance of a single category is smaller given that the other categories are already provided.

In addition, to compare the value of providing a category with the effect of a price adjustment, we run another counterfactual that increases the price of a category by 10%. We calculate the EV for this price shock and examine its magnitude relative to the EV of the exit of the same category. Results for both supercenters and club stores are shown in Table 1.16. We find that welfare loss due to a 10% increase in the price of general merchandise in a supercenter is equivalent to a 0.04% reduction in households' budget. This effect is only as large as 0.11% of the welfare loss due to the exit of the category. For club stores, we see the price effect is greater for non-food grocery and general merchandise than for supercenters. This finding suggests that lowering prices will be relatively more effective for these categories. Overall, providing a category generates considerably higher benefits for consumers than lowering the price for the same category.

1.6.3 Store Value Comparison

We have shown that big-box stores such as supercenters offer lower quality-adjusted prices and exhibit a higher level of cross-category complementarities. How do these characteristics affect the value of stores for consumers? Connecting to our previous analysis on the consumption value of a store, we further conduct two exercises to compare the consumption value across stores.

Recall that the consumption value of a store is defined as the welfare loss when all the categories in a store are removed. We can view the consumption value as a major component of the total value of a store. The residual is the welfare loss generated by further removing from the choice set the store with no consumption value. We focus on the consumption value because prices and product variety affect only the total welfare directly through this term.[26]

26. The value of a store can be defined as the welfare loss for a household when the store is removed from

In the first exercise, we calculate the welfare loss when a supercenter is replaced by a different store (EV^L). It simulates size restrictions that eliminate the entry of large stores such as supercenters, while allowing the entry of smaller stores. The new store inhabits the same location of the original supercenter (i.e. the same Γ before and after replacement), but offers its own quality-adjusted prices. We compare the welfare loss to the consumption value of a supercenter (EV^s) and report its share (EV^L/EV^s). A greater share implies a higher welfare loss and a lower value of the replacement. Figure 1.11 displays the results of replacing supercenters by each chain store respectively. The loss in value lies between 20% and 90%, with an average at around 50%, suggesting that half of the consumption value is gone once a supercenter is replaced by another store.

The change in welfare is driven by two forces. First, households may face higher quality-adjusted prices and purchase fewer products after the replacement (denoted as "Consumption"). Second, households may pay higher trip costs as they switch across stores (denoted as "Trip Costs"). We calculate the change from both sources. As shown in Figure 1.11, the increase in trip costs accounts for 20-40% of the total welfare loss for most of the replacement stores. Two grocery stores see a higher share of loss from trip costs. This is because they offer lower quality-adjusted prices of food than the supercenter and the overall loss from the "Consumption" part is small given food accounts for around 60% of the spending in the HMS.

In the second exercise, we simulate the exits of retail chains from their current markets, and compare these impacts to the exit of a supercenter chain. We quantify the consumption value of each existing chain ($EV^{s'}$) and normalize the value for a supercenter (EV^s) to 1. Figure 1.12 includes the comparison of store value for both exercises. On the horizontal axis, we plot the relative value (R), calculated as $1 - EV^L/EV^s$, from the first exercise. On the vertical axis, we plot the "Relative Value (E)", calculated as $EV^{s'}/EV^s$, from the

the choice set. As the value of the outside store is not zero, removing an entire choice from the discrete choice model would incur some mechanical utility loss. Focusing on consumption value avoids this problem.

second exercise. For most of the stores, their relative values calculated from the two exercises are close. Some grocery stores show a higher value using the second method, implying that these existing stores can be of great importance for the households they serve in their current markets.

From both exercises, the consumption value of an average store is only 50-60% of the value of a supercenter, whereas a club store mostly generates the same value as other stores. Therefore, even though both supercenters and club stores offer a wide range of product categories, supercenters generate substantially higher welfare gains for consumers. Recall that a supercenter offers lower quality-adjusted prices in all the categories and exhibits a higher level of cross-category complementarities. These are the underlying sources of the larger welfare a supercenter generates for consumers. Notably, because supercenters are essentially a discount store combined with a grocery department, the greater the value that the supercenter generates, the more its grocery department will attract households. Although club stores provide all categories, their quality-adjusted price for each category is not as low, because club stores provide limited assortment per category. A lower level of cross-category complementarity also restricts the value of an additional category in the store. This finding highlights the importance of a one-stop-shopping experience with categories of higher quality.

1.6.4 Discussion

We discuss our major assumptions and potential future extensions of the welfare analysis. First, we assume the absence of any supply-side response after a big-box store entry. Stores in the choice set, prices, and product varieties are taken as given. However, other stores may respond to the entry of big-box stores in various aspects, such as changing prices, adjusting quality, and exiting the market. We justify our assumption with three main reasons. First, using IRI store data, Arcidiacono et al. (2020) show that a supercenter entry has no effect on incumbent prices in the short- and medium-run. Thus, the price response of other stores may be limited when a big-box store enters. Second, Atkin, Faber and Gonzalez-Navarro (2017) estimate a welfare gain of 5.5% due to the availability of a new store and a loss of only 0.7% due to the exit of other stores. Thus, the welfare impact from store exit may be smaller than the direct welfare increase from shopping in a new big-box store. Third, both exiting the market and adjusting qualities usually take time for a store. We estimate the model using a sample of observations within one year, and limited the change in the number of stores in the choice set. Thus, our analysis can be viewed as a short-run welfare analysis. Nevertheless, seeing changes in supply-side competition after big-box stores enter and how stores determine prices, product variety, and product quality in response is still interesting. Our demand-side estimation is a starting point for further supply-side analyses, and our welfare analysis can be a benchmark for a consumer welfare analysis that incorporates competitive responses.

Second, our baseline estimates give the welfare impact of an average household. However, just as different household groups have different sensitivity to trip costs, households can also differ in terms of the tastes and shopping needs for stores and categories. Stores can also choose their location accordingly and serve households with selected characteristics. For example, compared with club stores with more stores in the cities, supercenters serve relatively less populated areas. Emphasizing the heterogeneous effects across households may have important implications for inequality. Our current model, which allows for income effects, can facilitate analysis across different income groups. We can also further extend our model to allow for household-specific preferences for stores and categories by adopting a similar methodology to Mehta and Ma (2012) by using a mixture distribution for household preferences.

1.7 Conclusion

In this paper, we investigate the impact of the rapid expansion of two types of big-box stores, supercenters and warehouse clubs, in the US over the past few decades. Using detailed consumer scanner data and an event-study approach, we document three main facts about households' responses to big-box store entries. First, households substitute toward big-box stores from other stores. Second, households change the number of product categories purchased per shopping trip, increasing the variety per trip when supercenters enter but decreasing the variety per trip when club stores enter. This finding is consistent with the fact that supercenters provide larger assortments across a broad range of product categories, whereas club stores tend to provide more limited assortments despite a similarly broad range of product categories. Third, households pay lower prices. This response is stronger for supercenters, consistent with the fact that supercenters offer lower prices than club stores.

To quantify the effects of big-box store entry in a way that incorporates our empirical findings, we develop a multi-store multi-category demand model. We find that both quality-adjusted prices and the degree of cross-category complementarities determines the value of each product category in stores and its contribution to consumer welfare. Supercenters have relatively low quality-adjusted prices and generate stronger cross-category complementarities. Through our counterfactual analysis, we find these factors lead households to derive more welfare from supercenters than from club stores and other store types, highlighting the importance of a one-stop-shopping experience. This implies that regulations that constrain store sizes could substantially limit these benefits for consumers.

1.8 Tables

Table 1.1: Spending Share: Supercenter Entry

Store type	(1) Supercenter	(2) Grocery	(3) Discount Store	(4) Warehouse Club	(5) Drug Store	(6) Dollar Store
Number of supercenters	0.0607***	-0.0357***	-0.0085***	-0.0067***	-0.0040***	-0.0023***
	(0.0036)	(0.0036)	(0.0017)	(0.0016)	(0.0012)	(0.0006)
Observations	1531361	1531361	1531361	1531361	1531361	1531361
Adj R-squared	0.819	0.769	0.637	0.775	0.653	0.694
Within R-squared	0.002	0.001	0.000	0.000	0.000	0.000
Prob > F	0.000	0.000	0.000	0.000	0.001	0.000
Number of clusters	106458	106458	106458	106458	106458	106458
Household FE	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X
Spending Share	16.2%	47.3%	7.7%	10.4%	4.6%	1.8%

Notes: This table uses 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The sample only includes households who have never moved during the period. The dependent variable spending share for each store type is the percentage to the total expenditure in CPG products for each household-quarter observations in the HMS. Discount Store includes discount stores other than the supercenter. The reported independent variable is the number of supercenters in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.2: Spending Share: Clubs Entry

Store type	(1) Clubs	(2) Grocery	(3) Discount Store	(4) Warehouse Club	(5) Drug Store	(6) Dollar Store
Numer of clubs	0.0353***	-0.0232***	-0.0157***	0.0005	0.0007	-0.0008
	(0.0028)	(0.0040)	(0.0034)	(0.0004)	(0.0013)	(0.0007)
Observations	1865246	1865246	1865246	1865246	1865246	1865246
Adj R-squared	0.766	0.759	0.792	0.466	0.640	0.689
Within R-squared	0.001	0.000	0.000	0.000	-0.000	0.000
Prob > F	0.000	0.000	0.000	0.118	0.644	0.291
Number of clusters	120135	120135	120135	120135	120135	120135
Household FE	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X
Spending Share	8.9%	47.3%	23.9%	1.5%	4.6%	1.8%

Notes: This table uses 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The sample only includes households who have never moved during the period. The dependent variable spending share for each channel type is the percentage to the total expenditure in CPG products for each household-quarter observations in the HMS. Warehouse Club includes club stores other than the three focal clubs. The reported independent variable is the total number of club stores in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.3: ln(Departments per Trip): Supercenter Entry

Store type	(1) All	(2) Supercenter	(3) Grocery	(4) Discount Store	(5) Warehouse Club	(6) Drug Store	(7) Dollar Store
Number of supercenters	0.0063**	0.0286***	-0.0162***	-0.0181***	-0.0091	-0.0115**	-0.0008
	(0.0031)	(0.0084)	(0.0040)	(0.0065)	(0.0074)	(0.0047)	(0.0064)
Observations	1531362	817370	1485107	900027	606703	905639	656488
Adj R-squared	0.780	0.570	0.712	0.423	0.537	0.354	0.375
Within R-squared	0.000	0.000	0.000	0.000	0.000	0.000	-0.000
Prob > F	0.041	0.001	0.000	0.005	0.219	0.015	0.905
Number of clusters	106458	72827	104913	84332	55496	81656	65363
Household FE	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X

Notes: This table uses 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The sample only includes households who have never moved during the period. Dependent variables are ln(number of departments per trip) for each store type, with 5 departments in total. Discount Store includes discount stores other than the supercenter. The reported independent variable is the number of supercenters in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.4: ln(Departments per Trip): Club Entry

Store type	(1) All	(2) Clubs	(3) Grocery	(4) Discount Store	(5) Warehouse Club	(6) Drug Store	(7) Dollar Store
Number of clubs	-0.0188***	-0.0428***	-0.0177***	-0.0203***	0.0268	-0.000824	0.000644
	(0.00363)	(0.0106)	(0.00492)	(0.00707)	(0.0521)	(0.00542)	(0.00785)
Observations	1865248	719103	1805407	1560123	51782	1078176	807881
Adj R-squared	0.775	0.533	0.701	0.599	0.473	0.352	0.375
Within R-squared	0.000	0.000	0.000	0.000	-0.000	-0.000	-0.000
Prob > F	0.000	0.000	0.000	0.004	0.608	0.879	0.935
Number of clusters	120135	60277	118277	112671	8048	90535	74641
Household FE	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X

Notes: This table uses 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The sample only includes households who have never moved during the period. Dependent variables are ln(number of departments per trip) for each store type, with 5 departments in total. Warehouse Club includes club stores other than the three focal clubs. The reported independent variable is the total number of club stores in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.5: ln(Relative Price Index): Supercenter Entry

Departments	(1) All	(2) Health & Beauty Care	(3) Food	(4) Non-food Grocery	(5) General Merchandise
Number of supercenters	-0.0050***	-0.0061***	-0.0059***	-0.0022*	-0.0026
	(0.0007)	(0.0014)	(0.0008)	(0.0012)	(0.0021)
Observations	793868	430754	792485	587517	290797
Adj R-squared	0.571	0.303	0.574	0.376	0.162
Prob > F	0.000	0.000	0.000	0.065	0.211
Number of clusters	79829	59428	79736	70592	49987
Household FE	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X

Notes: This table uses 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The sample only includes households who have never moved during the period. The dependent variable relative price index (RPI) is defined in Equation 1.2. Column (1) reports RPI including all products. Column (2)-(5) report RPI including proudcts in each indicated departments respectively. The reported independent variable is the number of supercenters in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.6: ln(Relative Price Index): Clubs Entry

Departments	(1) All	(2) Health& Beauty Care	(3) Food	(4) Non-food Grocery	(5) General Merchandise
Number of clubs	-0.0015	-0.0006	-0.0020	-0.0043**	0.0048*
	(0.0012)	(0.00222)	(0.0013)	(0.0019)	(0.0028)
Observations	1842295	1199219	1839768	1578261	920118
Adj R-squared	0.688	0.409	0.709	0.433	0.137
Within R-squared	0.000	-0.000	0.000	0.000	0.000
Prob > F	0.205	0.778	0.108	0.023	0.091
Number of clusters	119428	101377	119382	113974	92877
Household FE	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X

Notes: This table uses 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The sample only includes households who have never moved during the period.The dependent variable relative price index (RPI) is defined in Equation 1.2. Column (1) reports RPI including all products. Column (2)-(5) report RPI including proudcts in each indicated departments respectively. The reported independent variable is the total number of club stores in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.7: Summary of Households

Number of observations	11376		
Average weekly budget	270.6		
	One store	Two stores	Outside option
Obs Share	0.42	0.47	0.11
	One store	Two stores	
Spending share		Store 1	Store 2
	0.88	0.65	0.25
Obs share	No kids	With kids	Row total
Income Q1	0.17	0.05	0.22
Income Q2	0.19	0.04	0.23
Income Q3	0.26	0.07	0.33
Income Q4	0.17	0.04	0.21
Column Total	0.79	0.21	1

Notes: This table presents summary statstics of the sample for structural estimation. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). We randamly sample 3 weeks per quarter per household and each household has at least 3 weeks' observation. Weekly budget is calculated based on the annual household income bracket reported in HMS and the ratio of consumer goods expenditure from Consumer Expenditure Survey in 2012 divided by 52 weeks. Outside option means a household does not visit any of the 17 focal stores during a week. Spending share is the ratio of expenditure to weekly budget. Household demographics are provided in HMS. We group households to 8 groups by income quartile interacted with whether they have children under 18. Income Q4 has the highest income level.

Table 1.8: Estimates in $\ln P^*$ from Category Level

Parameters	Estimates	s.e.	Parameters	Estimates	s.e.
log(likelihood)	-0.5653		ρ	1.1001	0.0206
λ_1	-7.0214	0.1445	$\ln \mu_1$	-1.7297	0.1172
λ_2	-7.5044	0.1577	$\ln \mu_2$	-2.0257	0.1069
λ_3	-6.4501	0.1771	$\ln \mu_3$	-2.1234	0.3343
λ_4	-7.3043	0.1473	$\ln \mu_4$	-1.2385	0.0964
α_1	-1.6781	0.0617	α_9	-0.744	0.0481
α_2	0.1993	0.0629	α_{10}	-0.7777	0.0536
α_3	0.0099	0.0737	α_{11}	-0.0614	0.048
α_4	-2.2011	0.0618	α_{12}	-1.3789	0.0605
α_5	-1.9473	0.0681	α_{13}	-1.0601	0.0886
α_6	-1.7985	0.0552	α_{14}	-1.4764	0.0512
α_7	-1.8891	0.0804	α_{15}	-1.2293	0.0555
α_8	-0.9055	0.0645	α_{16}	0.4267	0.0428

Notes: This table presents Maximum Likelihood estimates from the first stage for parameters in 1.12. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). $\alpha's$ are store fixed effects with α_{17} for a club store being normalized to 0. $\lambda's$ are category fixed effects. $\mu's$ are scale parameters. ρ is the coefficent for the number of UPCs per store-category. s.e. denotes standard errors.

Table 1.9: Estimates on Cross-category Complementarity within Stores

	Health & Beauty Care	Food	Non-food Grocery	General Merchandise	Outside option
Health & Beauty Care	0.204	-0.0082	0.0046	-0.0001	-0.0258
Food	-0.0082	0.3633	0.0021	-0.0093	-0.0291
Non-food Grocery	0.0046	0.0021	0.1122	0.0005	-0.016
General Merchandise	-0.0001	-0.0093	0.0005	0.2753	-0.0447
Outside	-0.0258	-0.0291	-0.016	-0.0447	0.0153

Notes: This table presents Maximum Likelihood estimates from the first stage for $\{b_{cd}, c, d = 1, \ldots, M+1\}$, which describes compelmentaries across categories within the same store from Equation 1.6. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). Positive numbers for a category pair suggest the two categories are complements and negative numbers suggest substitutes. Estimates to generate this matrix that are described in Appendix 1.10.1.1 are presented in Appendix Table 1.26

Table 1.10: Conditional Cross-category Price Elasticities for Individual Demand

	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Health & Beauty Care	-1.2654	0.0101	-0.0051	0.0005
Food	0.0064	-1.2844	-0.0012	0.0071
Non-food Grocery	-0.0062	-0.0029	-1.1636	-0.0006
General Merchandise	0.0008	0.0200	-0.0008	-1.5534

Notes: This table shows average cross-category price elasticities for individual demand conditioning on store choice and purchase incidence for categories within the same store. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). Each cell is elasticity of row demand with respect to column price.

Table 1.11: Estimates at Store Level Decision

parameters	estimates	s.e.	parameters	estimates	s.e.
log(likelihood)	-2.9963	0.0008	γ^v	0.0026	0.0001
γ^1	-0.0631	0.0001	γ^2	-2.283	0.0027
γ^3	3.6063	0.0035	γ^0	0.611	0.0029
γ_{g1}	1		γ_{g5}	1.0136	0.002
γ_{g2}	1.041	0.0013	γ_{g6}	1.0346	0.0025
γ_{g3}	1.1324	0.0016	γ_{g7}	1.0422	0.002
γ_{g4}	1.1682	0.0015	γ_{g8}	1.1426	0.0026

Notes: This table presents Maximum Likelihood estimates for the second stage parameters in $\mathcal{U}_{irt} = \gamma^v \ln V_{irt} + \gamma_g(\gamma^1 D_{ir} + \gamma^2 I_i^{two} + \gamma^3 I_i^{same}) + \nu_{irt}$. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). $\gamma_1, ..., \gamma_4$ are households with no kids from Income Q1 to Income Q4 descrived in Table 1.7. $\gamma_5, ..., \gamma_8$ are households with kids from lowest income group to the highest. γ_1 is normalized to 1. γ^0 is the trip cost for outside option store 0. Bootstrap standard errors are reported in colunms s.e.

Table 1.12: Cross-category Complementarity of a Supercenter

	Exiting Categories			
	A: Conditional			
	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Health & Beauty Care	-1.000	0.558	0.072	0.039
Food	-0.148	-1.000	0.042	0.011
Non-food Grocery	1.417	1.924	-1.000	0.040
General Merchandise	0.259	1.138	0.683	-1.000
All	-0.058	-0.107	-0.055	-0.056
	B: Unconditional			
	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Health & Beauty Care	-1.000	0.412	0.041	0.019
Food	-0.169	-1.000	0.012	-0.009
Non-food Grocery	1.359	1.650	-1.000	0.020
General Merchandise	0.229	0.938	0.635	-1.000
All	-0.080	-0.190	-0.082	-0.074
	C: Other Stores			
	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Health & Beauty Care	0.119	0.030	-0.015	-0.015
Food	0.039	0.014	0.004	0.001
Non-food Grocery	-0.399	-0.313	0.049	-0.004
General Merchandise	-0.046	-0.002	-0.151	-0.030
All	-0.003	-0.020	-0.003	-0.004

Notes: This table presents quantity change in column categories when row category is removed from a supercenter. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). Panel A reports the change in the Supercenter conditioning on the fact that the supercenter is chosen by households. Panel B reports the change in the Supercenter when households are allowed to choose other stores. Panel C reports the total change in the other stores except for the Supercenter when households are allowed to choose other stores.

Table 1.13: Cross-category Complementarity of a Club Store

	Exiting Categories			
	A: Conditional			
	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Health & Beauty Care	-1.000	0.870	0.223	0.170
Food	-0.171	-1.000	-0.010	-0.045
Non-food Grocery	1.237	1.848	-1.000	0.066
General Merchandise	0.338	1.473	0.534	-1.000
All	-0.075	-0.133	-0.064	-0.070
	B: Unconditional			
	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Health & Beauty Care	-1.000	0.691	0.208	0.169
Food	-0.178	-1.000	-0.022	-0.046
Non-food Grocery	1.220	1.575	-1.000	0.065
General Merchandise	0.327	1.237	0.515	-1.000
All	-0.082	-0.216	-0.075	-0.071
	C: Other Stores			
	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Health & Beauty Care	0.024	-0.002	-0.010	-0.010
Food	0.007	0.002	-0.006	-0.007
Non-food Grocery	-0.097	-0.075	0.037	0.018
General Merchandise	-0.259	-0.252	-0.279	-0.234
All	-0.020	-0.024	-0.020	-0.019

Notes: This table presents quantity change in column categories when row category is removed from a supercenter. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). Panel A reports the change in the Club Store conditioning on the fact that the supercenter is chosen by households. Panel B reports the change in the Club Store when households are allowed to choose other stores. Panel C reports the total change in the other stores except for the Club Store when households are allowed to choose other stores.

Table 1.14: Value of a Category vs. Consumption Value of a Supercenter (=1)

	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
EV Share	**34.6%**	**86.3%**	**40.4%**	**29.0%**
Spending Share	19.7%	58.2%	14.5%	7.6%
$\Delta \ln V$ Share	16.4%	65.8%	20.0%	13.2%

Notes: This table presents a comparision between the value of each category and the total consumption value of a supercenter as defined in Section 1.6.2. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). EV share indicates the ratio of the value of a category to the consumption value of a supercenter. Spending share is the percentage of average households' expenditure in the supercenter given households choose the supercenter. $\Delta \ln V$ share indicates the ratio of the utility change when the corresponding category is removed from the store to the utility change when all categories are removed.

Table 1.15: Value of a Category vs. Consumption Value of a Club Store (=1)

	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
EV Share	**6.5%**	**73.1%**	**11.3%**	**1.1%**
Spending Share	15.81%	63.10%	13.38%	7.71%
$\Delta \ln V$ Share	5.13%	65.81%	8.55%	0.85%

Notes: This table presents a comparision between the value of each category and the total consumption value of a club store as defined in Section 1.6.2. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). EV share indicates the ratio of the value of a category to the consumption value of a club store . Spending share is the percentage of average households' expenditure in the club store given households choose the club store. $\Delta \ln V$ share indicates the ratio of the utility change when the corresponding category is removed from the store to the utility change when all categories are removed.

Table 1.16: 10% Price Increase vs. Category Exit (=1)

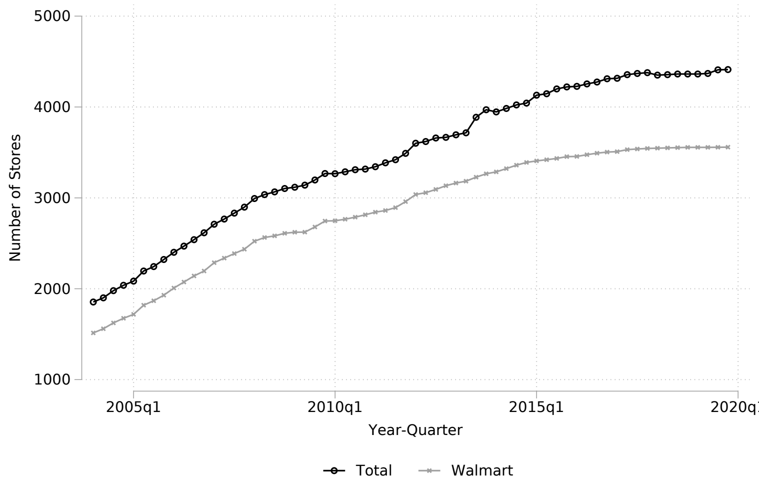
	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
	Supercenter			
EV	0.28%	1.00%	0.25%	0.04%
% of Category Exit	0.88%	1.27%	0.74%	0.11%
	Club Store			
EV	0.004%	0.350%	0.060%	0.004%
% of Category Exit	0.15%	1.15%	1.28%	0.89%

Notes: This table presents the equivalent variation for 10% increase in price for each category respectively (EV) and its comparision with the value of a category (% of Category Exit). The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). The first panel shows results for a supercenter and the second panel shows results for a club store.

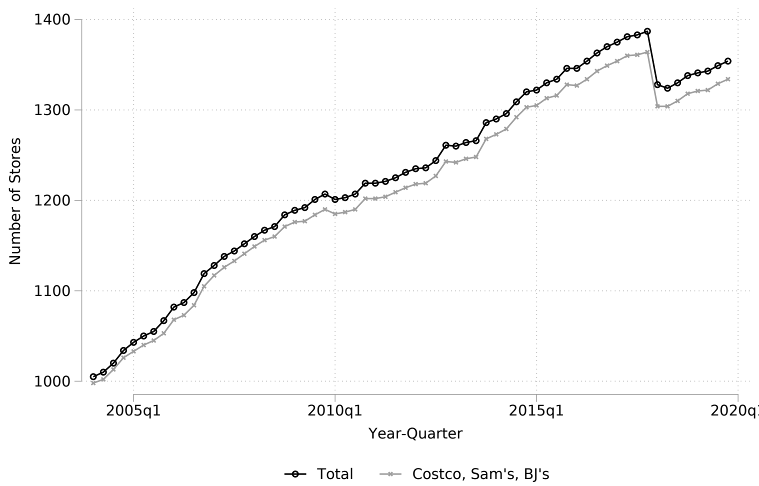
1.9 Figures

Figure 1.1: Number of Major Big-box Stores in the U.S.

(a) Supercenters



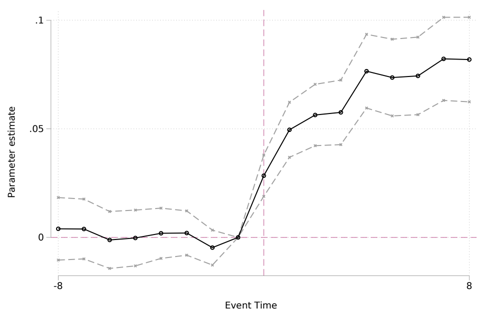
(b) Warehouse Clubs



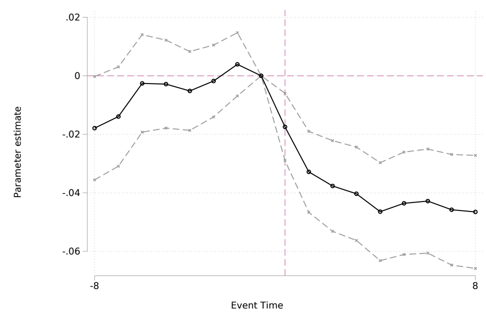
Notes: These figures present the nubmer of major big-box stores in the US from 2004 to 2019 using TDLinx Data. Figure 1.1a is the number of supercenters in the U.S. and Figure 1.1b is the number of club stores in the U.S. The drop in the number of warehouse clubs in 2018 reflects closures of Sam's Clubs (https://www.businessinsider.com/why-sams-club-is-closing-stores-2018-1).

Figure 1.2: Event Study Graph: Supercenter Entry on Spending Share

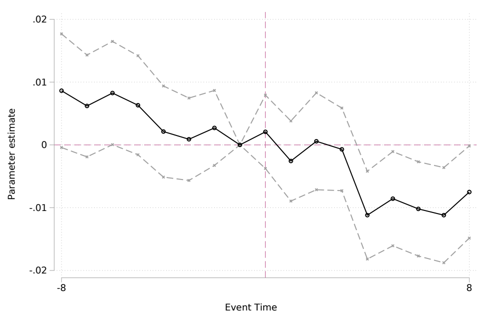
(a) Supercenter



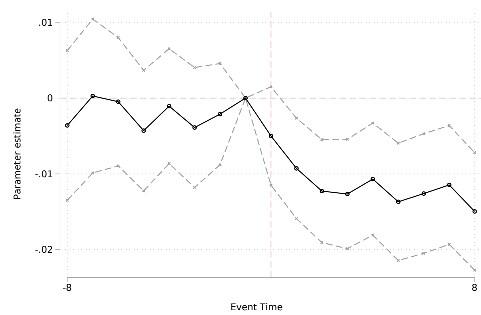
(b) Grocery



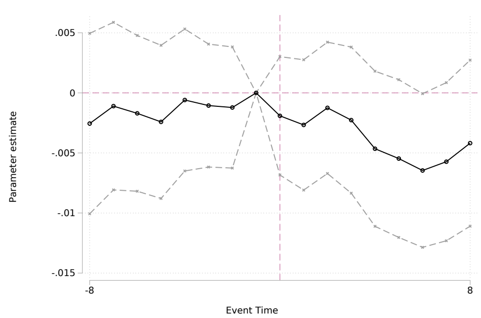
(c) Discount Store



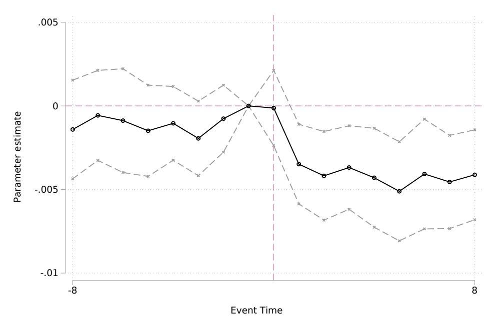
(d) Warehouse Club



(e) Drug Store



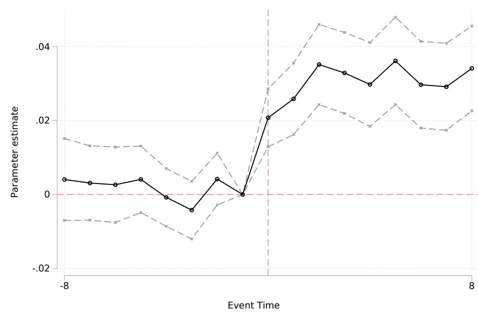
(f) Dollar Store



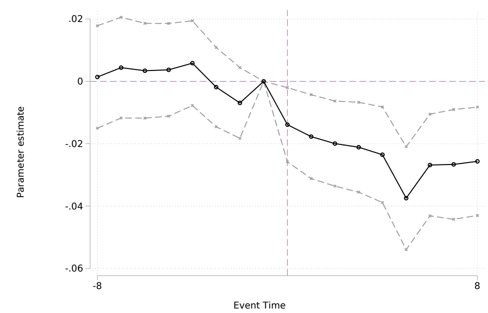
Notes: These figures use 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The figures present coefficients for eight leading and lagging periods of supercenter entries, and 95% confidence intervals from estimates of the event study on supercenter entries. The dependent variable spending share for each store type is the percentage to the total expenditure in CPG products for each household-quarter observations in the HMS. Discount Store includes discount stores other than the supercenter. All regressions control for year-quarter indicators and household fixed effects.

Figure 1.3: Event Study Graph: Clubs Entry on Spending Share

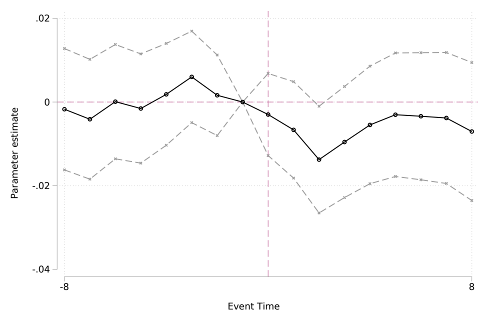
(a) Club Stores



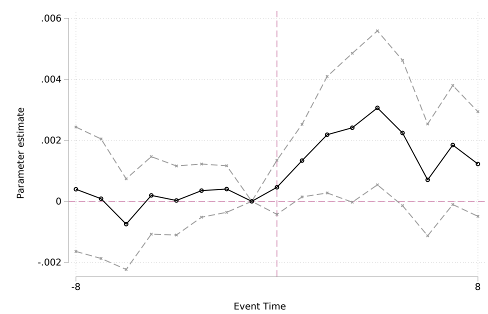
(b) Grocery



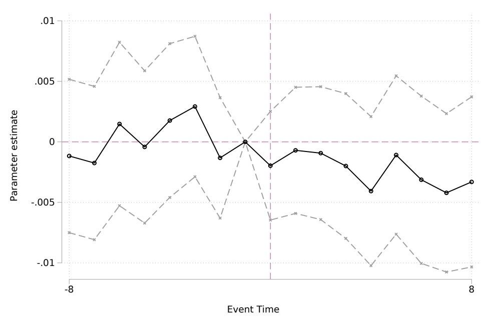
(c) Discount Store



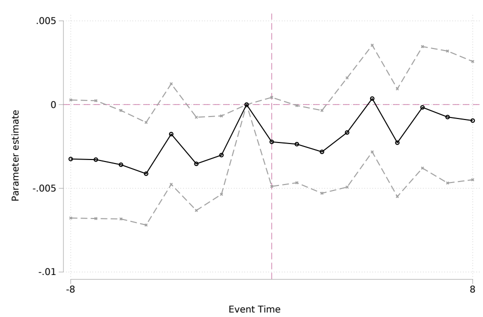
(d) Warehouse Club



(e) Drug Store



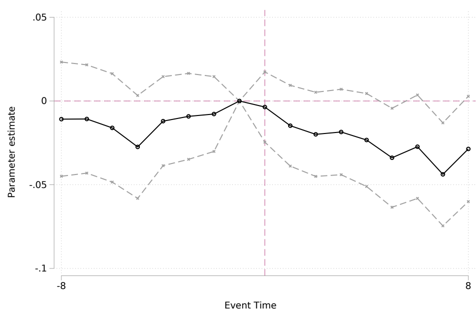
(f) Dollar Store



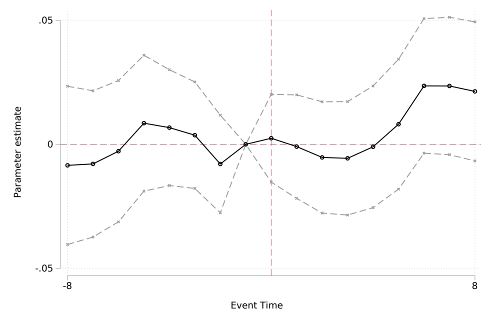
Notes: These figures use 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The figures present coefficients for eight leading and lagging periods of club store entries, and 95% confidence intervals from estimates of the event study on club store entries. The dependent variable spending share for each channel type is the percentage to the total expenditure in CPG products for each household-quarter observations in the HMS. Warehouse Club includes club stores other than the three focal clubs. All regressions control for year-quarter indicators and household fixed effects.

Figure 1.4: Event Study Graph: Supercenter Entry on Trips and Varieties per Trip

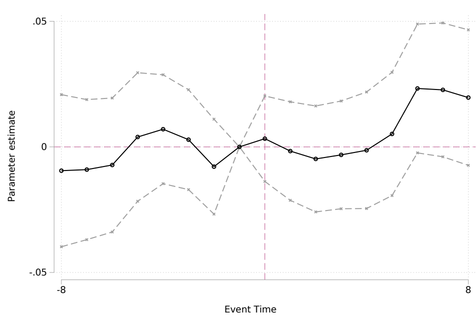
(a) Trip number



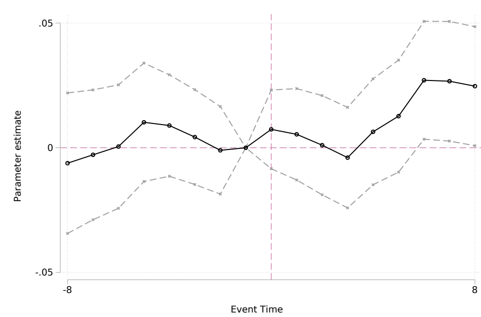
(b) UPCs



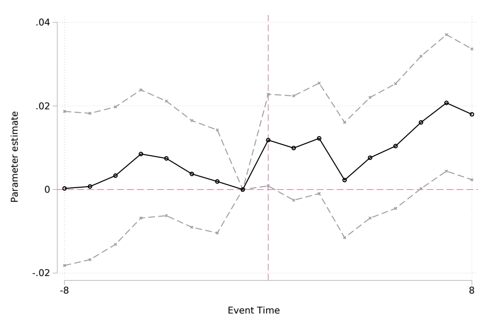
(c) Brands



(d) Product group



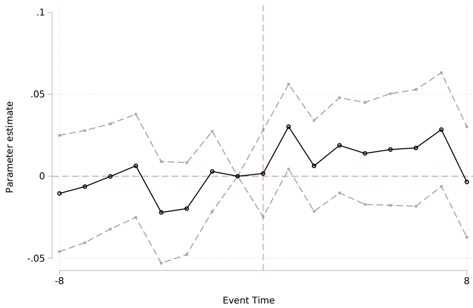
(e) Department



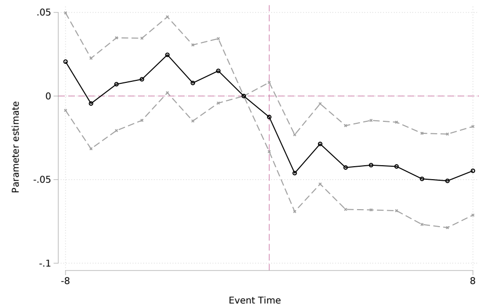
Notes: These figures use 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The figures present coefficients for eight leading and lagging periods of supercenter entries, and 95% confidence intervals from estimates of the event study on supercenter entries. The dependent variable from (a)-(e) are total number of shopping trips, number of UPCs per trip, number of brands per trip, number of product groups per trip, and number of departments per trip. Discount Store includes discount stores other than the supercenter. All regressions control for year-quarter indicators and household fixed effects.

Figure 1.5: Event Study Graph: Clubs Entry on Trips and Varieties per Trip

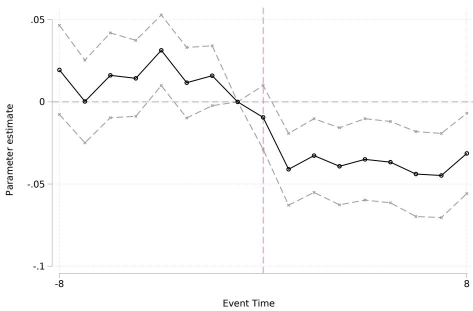
(a) Trip number



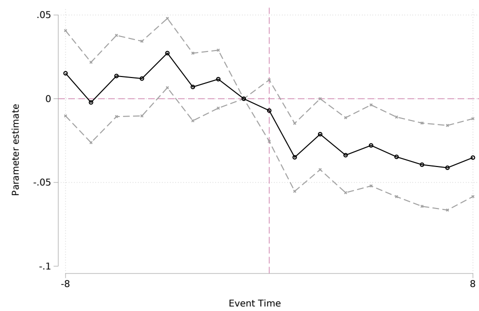
(b) UPCs



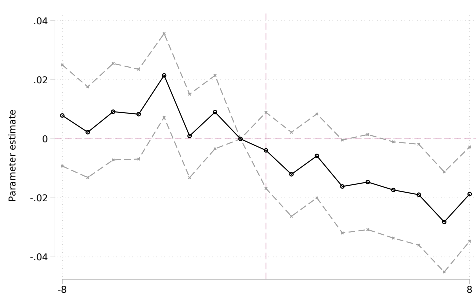
(c) Brands



(d) Product group



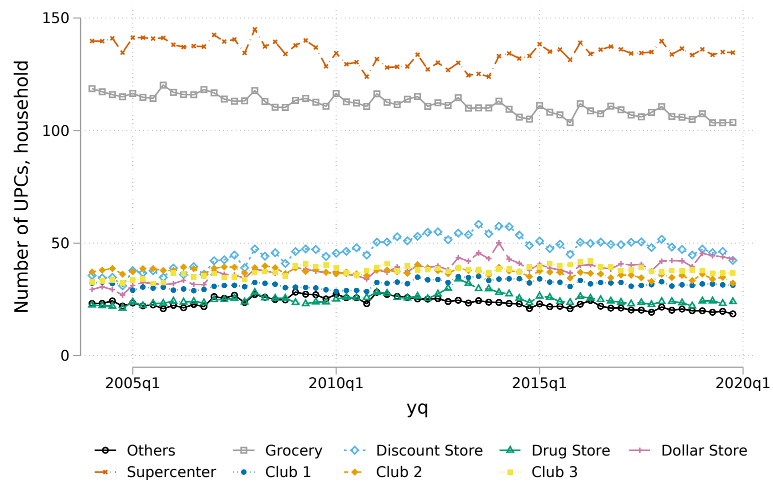
(e) Department



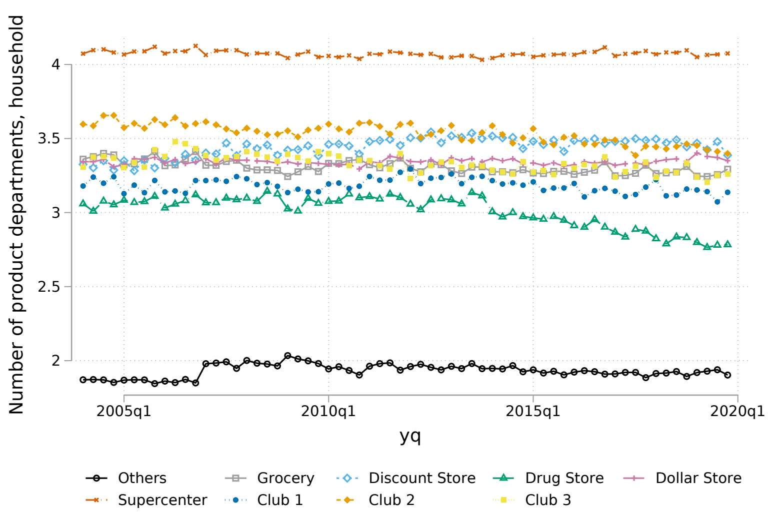
Notes: These figures use 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The figures present coefficients for eight leading and lagging periods of club store entries, and 95% confidence intervals from estimates of the event study on club store entries. The dependent variable from (a)-(e) are total number of shopping trips, number of UPCs per trip, number of brands per trip, number of product groups per trip, and number of departments per trip. Warehouse Club includes club stores other than the three focal clubs. All regressions control for year-quarter indicators and household fixed effects.

Figure 1.6: Product Assortment in Different Store Types

(a) UPCs



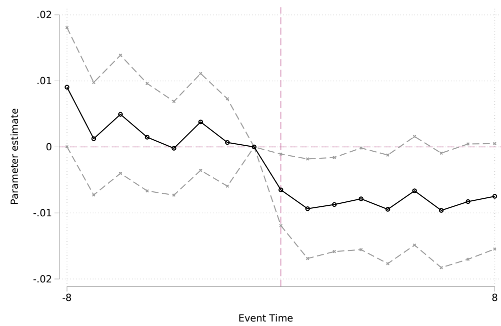
(b) Departments



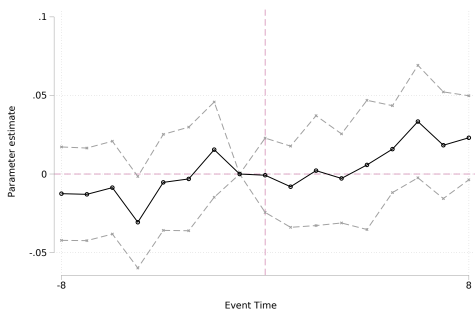
Notes: These figures present product varieties that households purchased from each store type using 2004-2019 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level. Figure 2.8a shows the number of UPCs and Figure 2.8b shows the number of departments.

Figure 1.7: Event Study Graph: Supercenter Entry on Relative Price Index

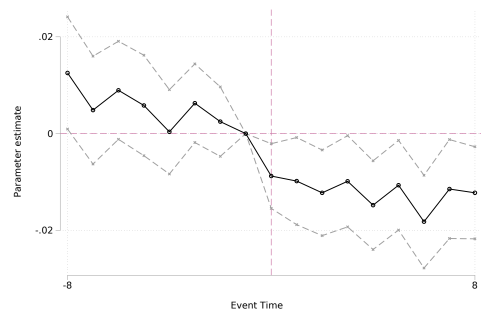
(a) All



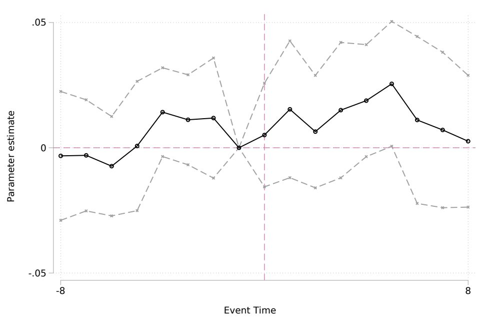
(b) Health & Beauty



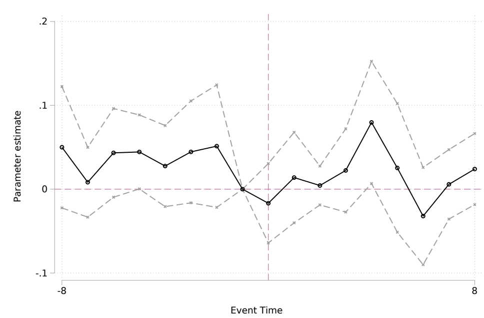
(c) Food



(d) Non-food



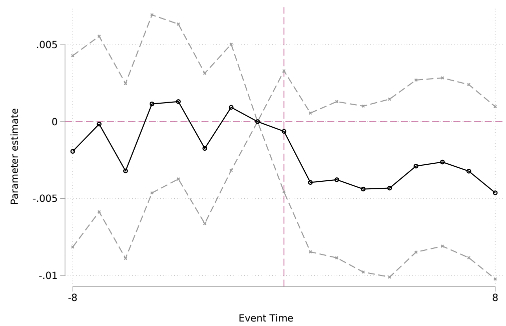
(e) General Merchandise



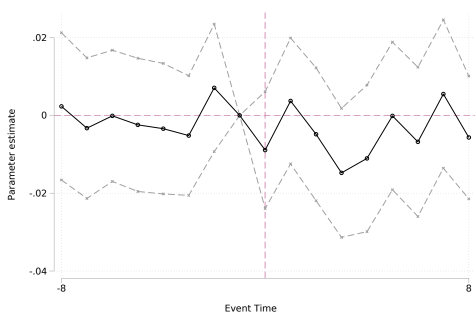
Notes: These figures use 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The figures present coefficients for eight leading and lagging periods of supercenter entries, and 95% confidence intervales from estimates of the event study on supeercenter entries. The dependent variables are log relative price index (RPI) for all products 1.7a and for each department 1.7b-1.7e are defined in Equation 1.2. All regressions control for year-quarter indicators and household fixed effects.

Figure 1.8: Event Study Graph: Club Entry on Relative Price Index

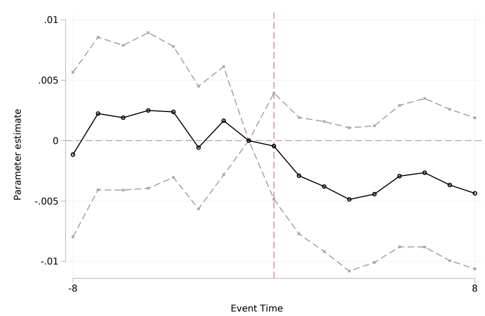
(a) All



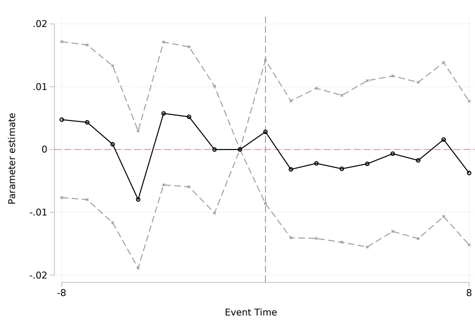
(b) Health & Beauty



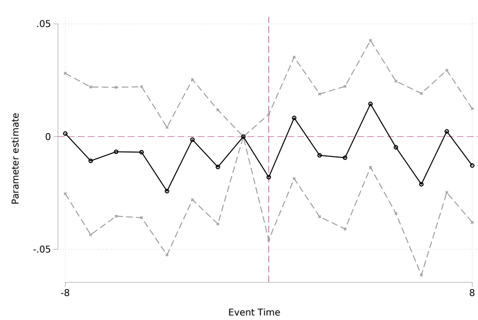
(c) Food



(d) Non-food



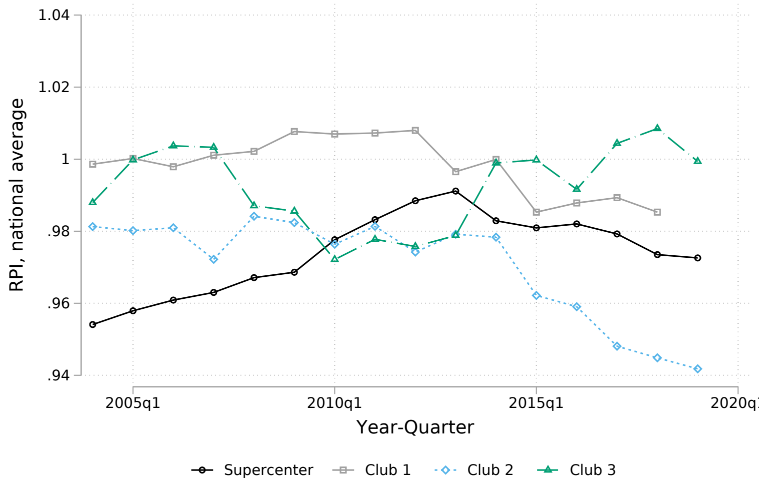
(e) GM



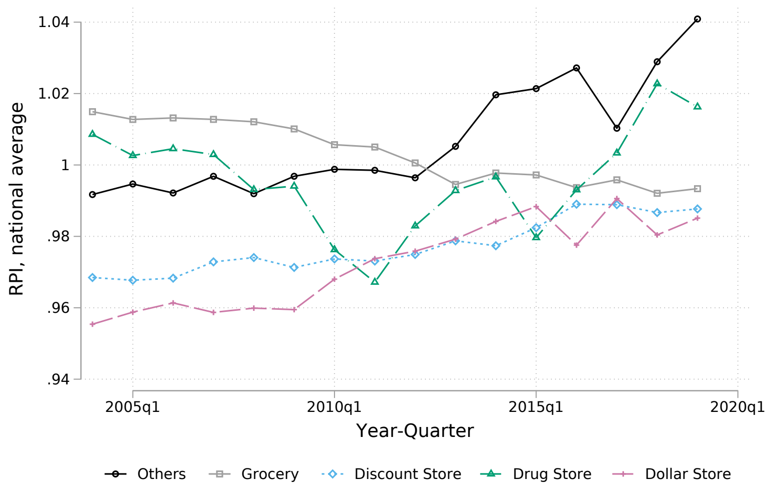
Notes: These figures use 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The figures present coefficients for eight leading and lagging periods of club store entries, and 95% confidence intervals from estimates of the event study on club store entries. The dependent variables are log relative price index (RPI) for all products 1.7a and for each department 1.7b-1.7e are defined in Equation 1.2. All regressions control for year-quarter indicators and household fixed effects.

Figure 1.9: Relative Price Index in Big-box Stores and Other Channel Types

(a) Big-box Stores



(b) Other Channel Types

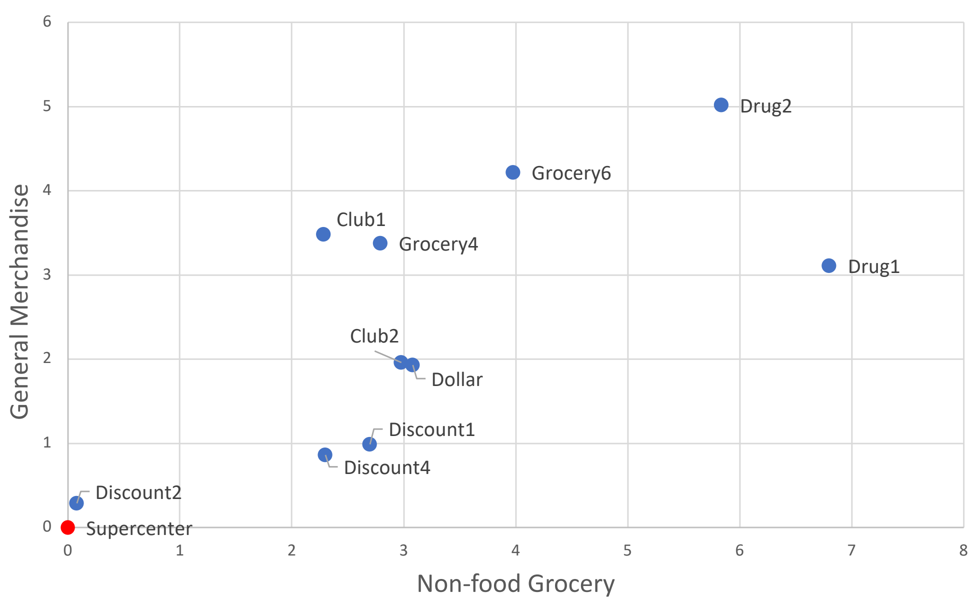


Notes: These figures present relative price index (RPI) for each store type defined as the quantity weighted averge of the ratio between total expenditure for each good and the counterfactual expenditure of each good at its national average price within a store, using 2004-2019 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level. Figure 2.9a shows the RPI for supercenters and club stores and Figure 2.9b shows the RPI for other store types.

Figure 1.10: Relative $\ln P^*$ across Stores

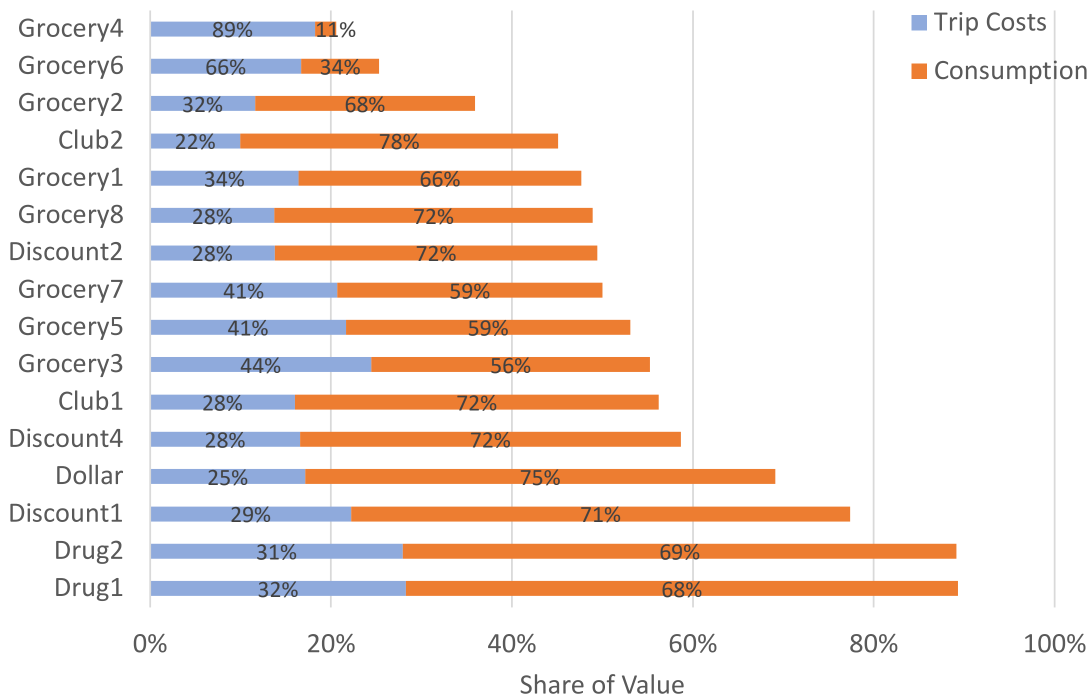
(a) Food and Health & Beauty Care

(b) Non-food Grocery and General Merchandise



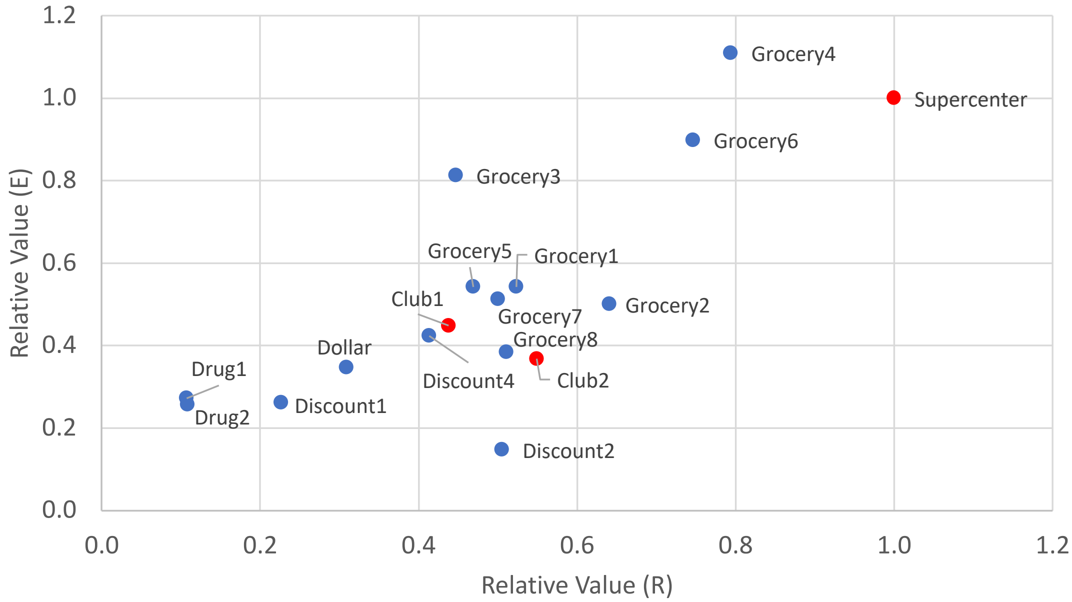
Notes: These two figures show the expected quality-adjusted price $\ln P^*$ as define in Equation 1.12 across all the weeks. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). $\ln P^*$ for a supercenter is normalized to 0 in each category. High $\ln P^*$ suggests higher cost for households to purchase a category given the same quality of the category across stores.

Figure 1.11: Welfare Loss of Replacing a Supercenter



Notes: This figure shows the welfare loss when a supercenter is replaced by a different store. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). When a supercenter is replaced, the quality-adjusted prices are replaced by that of another store. The horizontal axis shows the share of value, which is the resulting welfare loss of an average household (EV^L) compared to the consumption value of a supercenter (EV^s). The consumption value of a store is the welfare loss when all the categories are removed from the store. The vertical axis lists the stores that replace the supercenter. The total welfare loss of a replacement is decomposed into two parts. The"Trip Costs" part is the increase in trip costs due to switching across stores. The "Consumption" part is the loss from households purchasing less products with higher quality-adjusted prices after the replacement.

Figure 1.12: Store Value Comparison



Notes: This figure compares the consumption value across stores from two exercises. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). Consumption value is the welfare loss when all the categories are removed from a store. The horizontal axis shows the relative value of a store that replaces a supercenter. Relative value (R) is calculated as $1 - EV^L/EV^s$, where EV^L is the welfare loss when a supercenter is replaced by the store and EV^s is the consumption value of a supercenter. The vertical axis displays the relative consumption value of a store compared to a supercenter. Relative value (E) is defined as $EV^{s'}/EV^s$, where $EV^{s'}$ is the welfare loss when all the categories are removed in a store s' that exists in the current choice set.

1.10 Appendix

1.10.1 Structural Model

1.10.1.1 Category-Level Parameters

$$\Theta = \{\alpha_s, \lambda_c, \rho, B = \{b_{cd}, b^2_{cd}\}, \mu_c, s \in \mathcal{S}, c, d = 1, ..., M+1\}$$

.

- $\{\alpha_s, s = 1, ..., 16\}$: store fixed effects, α_{17} for chain 17 is normalized to 0.
- $\{\lambda_c, c = 1, .., 4\}$: category fixed effects.
- ρ: coefficient for variety depth, that is, the number of UPCs per store-category.
- $B = \{b_{cd}, b^2_{cd}, c, d = 1, ..., M+1\}$: complementarity across store-categories.
  - $B1 = \{b_{cd}, c, d = 1, ..., M+1\}$: complementarity across categories within stores.
  - $B2 = \{b^2_{cd}, c, d = 1, ..., M\}$: complementarity across categories from two different stores.
  - Let $Bf = \{b_{cd}, c, d = 1, ..., M\}$ and $B_{M+1} = [b_{1,M+1}, ..., b_{4,M+1}]$. For two-store options with $M \times 2$ categories, we have

$$B = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{cccc} \text{Store 1} & \text{Store 2} & \text{Category 0} & \\ \left(\begin{array}{c} Bf \\ B2 \\ B1_{M+1} \end{array}\right. & \begin{array}{c} B2 \\ Bf \\ B1_{M+1} \end{array} & \left.\begin{array}{c} B1^T_{M+1} \\ B1^T_{M+1} \\ b_{M+1,M+1} \end{array}\right) & \begin{array}{l} \text{Store 1} \\ \text{Store 2} \\ \text{Category 0.} \end{array} \end{array}$$

  - Matrices B, $B1$, $B2$ are all symmetric. Matrices B and $B1$ are positive semi-definite with Bf being positive definite. Give the structure and properties of B, we generate it in the following two steps:

    * We generate the $B1$ matrix as $B1 = Ch \times Ch^T$, where Ch is a lower triangular $(M+1) \times (M+1)$ Cholesky with elements $\{Ch_{cd}\}$. The last diagonal element $Ch_{M+1,M+1}$ is normalized to 0 because we do not have information on category 0. There are 14 parameters to be estimated in Ch.
    * We generate $B2$ by estimating the upper triangle of an $M \times M$ matrix Cho. Other elements in the low triangular part of Cho are functions of the elements in the upper triangle such that the structure of B is as defined. $B2 = Cho \times Ch^T$. There are 10 parameters to be estimated in Cho.

- $\mu_c > 0, c = 1, ..., M$: scale parameters for each category:
  - We estimate $\ln \mu_c$ to ensure $\mu_c > 0$ for $c = 1, ..., M$.

1.10.1.2 Likelihood Function for Simultaneous Estimation

Given category-level parameters Θ and store-level parameters Θ^S, the likelihood of observing store choice $\{I_r^S = 1, I_{r'}^S = 0, r' \neq r\}$, purchase incidence $\{I_j = 0\}_{j=1}^m, \{I_j = 1\}_{j=m+1}^{M_r}$, and budget share $\{S_j\}_{j=m+1}^{M_r}$ for household i week t is

$$
\begin{aligned}
&L_{it}\left(\{I_{r'}^S\}_{r'\in\mathcal{R}}, \{I_j\}_{j\in\mathcal{M}_r}, \{S_j\}_{j\in\mathcal{M}_r}|\Theta, \Theta^S\right) \\
&= \int_{\varepsilon_0=-\infty}^{\infty} \prod_{r'\in\mathcal{R}} \left[\int_{\boldsymbol{\varepsilon}_{r'\neq r}=-\infty}^{\infty} \int_{-\infty}^{-\mathbf{H}_{r,np}(\varepsilon_0)} \frac{\exp(\gamma^v \ln V_r(\boldsymbol{\varepsilon}) + \Gamma_r)}{\sum_{r'\in\mathcal{R}} \exp(\gamma^v \ln V_{r'}(\boldsymbol{\varepsilon}) + \Gamma_{r'})}\right. \\
&\left.\times \boldsymbol{\phi}(-\mathbf{H}_{r,p}(\varepsilon_0))\boldsymbol{\phi}(\boldsymbol{\varepsilon}_{r,np})\boldsymbol{\phi}(\boldsymbol{\varepsilon}_{r'\neq r})\mathbf{J}d\boldsymbol{\varepsilon}_{r,np}d\boldsymbol{\varepsilon}_{r'\neq r}\right]^{I_{r'}^S} \phi_N(\varepsilon_0)d\varepsilon_0,
\end{aligned} \tag{1.19}
$$

$$
H_j = \alpha_{s(j)} + \lambda_{c(j)} + \rho X_j + \mu_{c(j)}\kappa_j \ln y + \mu_{c(j)}(\kappa_j - 1)\varepsilon_0 - \mu_{c(j)} \left(\sum_{k=1}^{M_r} \delta_{jk} S_k\right) \left(1 - \sum_{j=1}^{M_r} \kappa_j S_j\right)^{-1}
$$

- $\Theta = \{\alpha_s, \lambda_c, \rho, B = \{b_{cd}, b_{cd}^2\}, \mu_c, s \in \mathcal{S}, c, d = 1, ..., M+1\}$.

- $\Theta^S = \{\gamma^v, \gamma^1, \gamma^2, \gamma^3, \{\gamma_g\}, \gamma^0\}$.

- $\phi(\cdot)$ is the pdf of joint T1EV distribution and $\phi_N(\cdot)$ is the pdf of standard normal distribution.

- $\mathbf{J} = D_{r'} \times \left(1 - \sum_{j\in\mathcal{M}_r'} \kappa_j S_j\right)^{-1-\sum_{j\in\mathcal{M}_r'} I_j} \prod_{j\in\mathcal{M}_r'} (\mu_{c(j)})^{I_j}$.

- $\boldsymbol{\varepsilon}_{r,np}$ is the shocks of non-purchased store-categories in the chosen store set.

- $\boldsymbol{\varepsilon}_{r'\neq r}$ is the shocks of store-categories in the non-chosen stores.

- $\{\delta_j\}_{j=1}^{M_r}$ and $\{\kappa_j\}_{j=1}^{M_r}$ are the reformulated parameters of the original parameters B, which is explained in Appendix 1.10.1.1:

  - B_f is a submatrix of B, consisting of first M_r rows and M_r columns of B.

  - $C = [C_1, ...C_{M_r}]$, $C_j = \sum_{k=1}^{M_r+1} b_{jk}$.

  - $\{\delta_j\}_{j=1}^{M_r}$ are the elements of the $M_r \times M_r$ matrix $\Delta_f = (B_f)^{-1}$.

- $\{\kappa_j\}_{j=1}^{M_r}$ are the elements in the $M_r \times 1$ vector K, where $K = (B_f)^{-1}C$.

- $D_r = 1$ if none of the focal categories are purchased. If at least one is purchased, D_r takes the value of the determinant of the submatrix of the matrix Δ_f after removing none purchased store-categories.

The log-likelihood function for the entire sample is:

$$l(\Theta, \Theta^S) = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{r \in \mathcal{R}} I_r^S \ln \left(L_{it} \left(\{I_{r'}^S\}_{r' \in \mathcal{R}}, \{I_j\}_{j \in \mathcal{M}_r}, \{S_j\}_{j \in \mathcal{M}_r} | \Theta, \Theta^S \right) . \right) . \tag{1.20}$$

1.10.1.3 Likelihood Function for Category Level

Given store choice r for household i week t, the likelihood of observing purchase incidence $\{I_j = 0\}_{j=1}^{m}, \{I_j = 1\}_{j=m+1}^{M_r}$ and budget share $\{S_j\}_{j=m+1}^{M_j}$ given parameter Θ and random shock ε_0 is

$$\begin{aligned}
&L_r\left(\{I_j\}_{j=1}^{M_r}, \{S_j\}_{j=1}^{M_r} | \Theta, \varepsilon_0\right) \\
=&D_r \times \left(1 - \sum_{j=1}^{M_r} \kappa_j S_j\right)^{-1-\sum_{j=1}^{M_r} I_j} \times \prod_{j=1}^{M_r} (\mu_{c(j)})^{I_j} \exp(-\exp(H_j))(\exp(H_j))^{I_j} \qquad (1.21)
\end{aligned}$$

$$H_j = \alpha_{s(j)} + \lambda_{c(j)} + \rho X_j + \mu_{c(j)} \kappa_j \ln y + \mu_{c(j)} (\kappa_j - 1) \varepsilon_0 - \mu_{c(j)} \left(\sum_{k=1}^{M_r} \delta_{jk} S_k\right) \left(1 - \sum_{j=1}^{M_r} \kappa_j S_j\right)^{-1}.$$

Notations are the same as in Appendix 1.10.1.2. Derivation of the likelihood function refers to online appendix of Mehta and Ma (2012).

1.10.1.4 Likelihood Function for Store Level

Given estimates from stage one $\hat{\Theta}$, the likelihood of observing purchase incidence $\{I_j = 0\}_{j=1}^{m}, \{I_j = 1\}_{j=m+1}^{M_r}$ and budget share $\{S_j\}_{j=m+1}^{M_j}$ given parameter Θ^S is

$$
\begin{aligned}
& L_{it}\left(\{I_{r'}^S\}_{r'\in\mathcal{R}}, \{I_j\}_{j\in\mathcal{M}_r}, \{S_j\}_{j\in\mathcal{M}_r}|\hat{\Theta}, \Theta^S\right) \\
&= \int_{\varepsilon_0=-\infty}^{\infty} \prod_{r'\in\mathcal{R}} \left[\int_{\boldsymbol{\varepsilon}_{r'\neq r}=-\infty}^{\infty} \int_{-\infty}^{-\hat{\mathbf{H}}_{r,np}(\varepsilon_0)} \frac{\exp(\gamma^v \ln \hat{V}_r(\boldsymbol{\varepsilon}) + \Gamma_r)}{\sum_{r'\in\mathcal{R}} \exp(\gamma^v \ln \hat{V}_{r'}(\boldsymbol{\varepsilon}) + \Gamma_{r'})}\right. \\
& \quad \left. \times \boldsymbol{\phi}(-\hat{\mathbf{H}}_{r,p}(\varepsilon_0))\boldsymbol{\phi}(\boldsymbol{\varepsilon}_{r,np})\boldsymbol{\phi}(\boldsymbol{\varepsilon}_{r'\neq r})\hat{\mathbf{J}} d\boldsymbol{\varepsilon}_{r,np} d\boldsymbol{\varepsilon}_{r'\neq r}\right]^{I_{r'}^S} \phi_N(\varepsilon_0) d\varepsilon_0,
\end{aligned} \tag{1.22}
$$

$$
\hat{H}_j = \hat{\alpha}_{s(j)} + \hat{\lambda}_{c(j)} + \hat{\rho} X_j + \hat{\mu}_{c(j)} \hat{\kappa}_j \ln y + \hat{\mu}_{c(j)} (\hat{\kappa}_j - 1)\varepsilon_0 - \hat{\mu}_{c(j)} \left(\sum_{k=1}^{M_r} \hat{\delta}_{jk} S_k\right) \left(1 - \sum_{j=1}^{M_r} \hat{\kappa}_j S_j\right)^{-1}.
$$

Notations are the same as in Appendix 1.10.1.2. We draw $\{\varepsilon_{0,it}\}$ and $\{\boldsymbol{\varepsilon}_{it}\}$ from their distribution given category level decision. The steps of drawing one set of random errors for each household-week are as follows:

1. Draw ε_0 from standard normal distribution $\mathcal{N}(0, 1)$.
2. For store choice r that are not chosen, draw $\{\varepsilon_j, j \in \mathcal{M}_r\}$ from standard extreme value distribution.
3. For the chosen store choice r, we set $\{\varepsilon_j, j \in \mathcal{M}_r\}$ based on purchase incidence and spending share:
   - If j is purchased, ε_j is the value such that the spending share of j is the observed S_j. After some derivation, $\varepsilon_j = -\hat{H}_j(\varepsilon_0)$.
   - If j is not purchased, ε_j needs to satisfy the condition such that virtual price for j is smaller than observed quality-adjusted price P_j^*. After some derivation, ε_j is drawn from standard extreme value distribution with upper bound $-\hat{H}_j(\varepsilon_0)$.

For each set of $\{\varepsilon_{0,it}\}$ and $\{\boldsymbol{\varepsilon}_{it}\}$, the likelihood thus becomes:

$$\hat{L}_{it}\left(\{I^S_{r'}\}_{r'\in\mathcal{R}}, \{I_j\}_{j\in\mathcal{M}_r}, \{S_j\}_{j\in\mathcal{M}_r}|\hat{\Theta}, \Theta^S\right) = \frac{\exp(\gamma^v \ln \hat{V}_r(\boldsymbol{\varepsilon_{it}}) + \Gamma_r)}{\sum_{r'\in\mathcal{R}} \exp(\gamma^v \ln \hat{V}_{r'}(\boldsymbol{\varepsilon_{it}}) + \Gamma_{r'})}. \quad (1.23)$$

1.10.2 Tables

Table 1.17: ln(UPCs per Trip): Supercenter Entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Store type	All	Supercenter	Grocery	Discount Store	Warehouse Club	Drug Store	Dollar Store
Number of supercenters	0.0008	0.0249*	-0.0289***	-0.0430***	-0.0150	-0.0268***	-0.0217**
	(0.0057)	(0.0141)	(0.0073)	(0.0114)	(0.0104)	(0.0079)	(0.0106)
Observations	1531362	817494	1485110	900542	606764	905899	656557
Adj R-squared	0.823	0.636	0.763	0.492	0.616	0.418	0.473
Within R-squared	-0.000	0.000	0.000	0.000	0.000	0.000	0.000
Prob > F	0.890	0.078	0.000	0.000	0.150	0.001	0.041
Number of clusters	106458	72837	104913	84354	55502	81671	65365
Household FE	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X

Notes: This table uses 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The sample only includes households who have never moved during the period. Dependent variables are ln(number of UPCs per trip) for each store type. Discount Store includes discount stores other than the supercenter. The reported independent variable is the number of supercenters in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.18: ln(UPCs per Trip): Club Entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Store type	All	Clubs	Grocery	Discount Store	Warehouse Club	Drug Store	Dollar Store
Number of clubs	-0.0449***	-0.0931***	-0.0310***	-0.0293**	0.00528	-0.00828	-0.0115
	(0.00637)	(0.0156)	(0.00865)	(0.0123)	(0.0885)	(0.00921)	(0.0135)
Observations	1865248	719160	1805410	1560387	51785	1078435	807950
Adj R-squared	0.816	0.614	0.752	0.651	0.545	0.415	0.471
Within R-squared	0.000	0.000	0.000	0.000	-0.000	0.000	0.000
Prob > F	0.000	0.000	0.000	0.017	0.952	0.369	0.394
Number of clusters	120135	60281	118277	112681	8048	90549	74643
Household FE	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X

Notes: This table uses 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The sample only includes households who have never moved during the period. Dependent variables are ln(number of UPCs per trip) for each store type. Warehouse Club includes club stores other than the three focal clubs. The reported independent variable is the total number of club stores in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.19: Number of Trips and Varieties per Trip: Supercenter Entry

Variables	(1) Trip	(2) UPC	(3) Brand	(4) Product Group	(5) Department
Number of supercenters	-0.0039	0.0008	0.0027	0.0023	0.0063**
	(0.0064)	(0.0057)	(0.0054)	(0.0049)	(0.0031)
Observations	1531362	1531362	1531362	1531362	1531362
Adj R-squared	0.759	0.823	0.825	0.819	0.780
Within R-squared	0.000	0.000	0.000	0.000	0.000
Prob > F	0.542	0.890	0.621	0.633	0.041
Number of clusters	106458	106458	106458	106458	106458
Household FE	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X

Notes: This table uses 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The sample only includes households who have never moved during the period. Dependent variables from Column (1)-(5) are log number of total shopping trips, log number of UPCs per trip, log number of brands per trip, log number of product groups per trip, and log number of departments per trip. The reported independent variable is the number of supercenters in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.20: Number of Trips and Varieties per Trip: Clubs Entry

Variables	(1) Trip	(2) UPC	(3) Brand	(4) Product Group	(5) Department
Number of clubs	0.0233***	-0.0449***	-0.0386***	-0.0367***	-0.0188***
	(0.0078)	(0.0064)	(0.0059)	(0.0057)	(0.0036)
Observations	1865248	1865248	1865248	1865248	1865248
Adj R-squared	0.748	0.816	0.817	0.813	0.775
Within R-squared	0.000	0.000	0.000	0.000	0.000
Prob > F	0.003	0.000	0.000	0.000	0.000
Number of clusters	120135	120135	120135	120135	120135
Household FE	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X

Notes: This table uses 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The sample only includes households who have never moved during the period. Dependent variables from Column (1)-(5) are log number of total shopping trips, log number of UPCs per trip, log number of brands per trip, log number of product groups per trip, and log number of departments per trip. The reported independent variable is the total number of club stores in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.21: ln(Number of Retailers Visited): Supercenter Entry

	(1)	(2)	(3)	(4)	(5)	(6)
Store type	All	Grocery	Discount Store	Warehouse Club	Drug Store	Dollar Store
Number of supercenters	-0.0356***	-0.0361***	-0.0609***	-0.0176***	-0.0326***	-0.0108**
	(0.0050)	(0.0048)	(0.0056)	(0.0040)	(0.0052)	(0.0046)
Observations	1531362	1531362	1531362	1531362	1531362	1531362
Adj R-squared	0.710	0.659	0.534	0.680	0.552	0.609
Within R-squared	0.000	0.000	0.000	0.000	0.000	0.000
Prob > F	0.000	0.000	0.000	0.000	0.000	0.020
Number of clusters	106458	106458	106458	106458	106458	106458
Household FE	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X

Notes: This table uses 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). The sample only includes households who have never moved during the period. Dependent variables are ln(number of retailers visited) for each store type. Discount Store includes discount stores other than the supercenter. The reported independent variable is the number of supercenters in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.22: ln(Number of Retailers Visited): Club Store Entry

	(1)	(2)	(3)	(4)	(5)	(6)
Store types	All	Grocery	Discount Store	Warehouse Club	Drug Store	Dollar Store
Number of clubs	0.0267***	0.0090	-0.0037	0.0035*	-0.00302	0.0089
	(0.0059)	(0.0059)	(0.0055)	(0.0019)	(0.0062)	(0.0056)
Observations	1865248	1865248	1865248	1865248	1865248	1865248
Adj R-squared	0.706	0.649	0.544	0.410	0.548	0.604
Within R-squared	0.000	0.000	0.000	0.000	-0.000	0.000
Prob > F	0.000	0.129	0.509	0.069	0.627	0.108
Number of clusters	120135	120135	120135	120135	120135	120135
Household FE	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X

Notes: This table uses 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The sample only includes households who have never moved during the period. Dependent variables are ln(number of retailers visited) for each store type. Warehouse Club includes club stores other than the three focal clubs. The reported independent variable is the total number of club stores in the zip code area where each household lives. All regressions control for household and year-quarter fixed effects. Observations are not weighted for national representativeness. Robust standard errors, clustered by household, are in parentheses. *, **, ***: statistically significant with 10, 5, and 1 percent confidence, respectively.

Table 1.23: Summary of Prices and Variety Depth by Store Types and Categories

Type	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
	Average Price			
Grocery	2.49	2.09	2.75	2.55
Discount	2.91	2.14	3.53	3.67
Club	13.91	7.40	12.75	14.91
Drug	6.57	1.75	3.70	4.46
Dollar	1.42	1.28	2.10	1.16
	log(Average Number of UPCs)			
Grocery	6.19	8.49	6.14	5.14
Discount	6.67	7.83	6.40	6.23
Club	5.15	6.77	5.02	4.71
Drug	6.75	6.46	5.35	4.85
Dollar	4.78	6.51	5.01	4.42

Notes: This table shows summary statistics for prices and variety depth for each store-category. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). For both price and variety depth measure, we use the average value across all the stores in the US within each chain and further aggregate to store type level for the summary. Category price is calculated using the average price for each UPC in the category weighted by sales and only UPCs that are sold in all biweeks in 2012 are included. Thus, a high category price result from two reasons: 1. same product is sold at a higher price in the store, 2. the store sells more high-priced items, (for example, clubs sell items with larger size and thus have higher category prices).

Table 1.24: Estimates on Cross-category Complementarity within Stores

	Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Health & Beauty Care	0.0045	0.0033	0.0005	0.0078
Food	0.0033	0.025	-0.0005	0.0164
Non-food Grocery	0.0005	-0.0005	0.0036	0.0047
General Merchandise	0.0078	0.0161	0.0047	0.0218

Notes: This table displays $\{b^2_{cd}, c, d = 1, ..., M\}$ for categories from two different stores from Equation 1.6. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). Positive numbers for a category pair suggest the two categories are complements and negative numbers suggest substitutes. Estimates to generat this matrix described in Appendix 1.10.1.1 are presented in Appendix Table 1.26

Table 1.25: Conditional Cross-category Price Elasticities for Individual Demand

		Store 2			
		Health & Beauty Care	Food	Non-food Grocery	General Merchandise
Store 1	Health & Beauty Care	-0.0039	-0.0038	-0.0003	-0.0064
	Food	-0.0032	-0.0221	0.0004	-0.0111
	Non-food Grocery	-0.0005	0.0007	-0.0042	-0.0056
	General Merchandise	-0.0103	-0.0311	-0.0063	-0.0231

Notes: This table shows average cross-category price elasticities for individual demand conditioning on store choice and purchase incidence for categories from two different stores. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS). The complementarity may result from some co-movement in price change of the same category across different stores

Table 1.26: Estimates of Elements in Cholesky Matrix

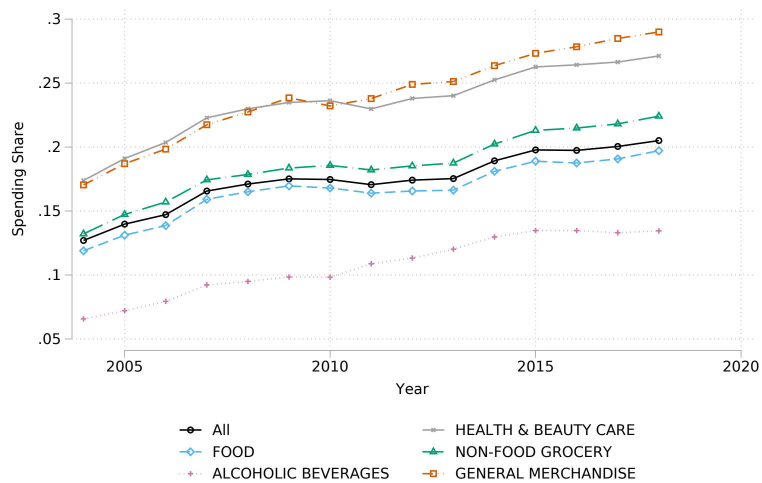
Parameters	Estimates	s.e.	Parameters	Estimates	s.e.
Ch_{11}	0.4517	0.0268	Cho_{11}	0.01	0.0066
Ch_{21}	-0.0181	0.0031	Cho_{12}	0.0059	0.0029
Ch_{22}	0.6025	0.0325	Cho_{13}	0.0011	0.0047
Ch_{31}	0.0101	0.0028	Cho_{14}	0.015	0.0051
Ch_{32}	0.0038	0.0017	Cho_{22}	0.0417	0.005
Ch_{33}	0.3347	0.0558	Cho_{23}	-0.0023	0.004
Ch_{41}	-0.0002	0.0046	Cho_{24}	0.0326	0.0047
Ch_{42}	-0.0155	0.0026	Cho_{33}	0.0109	0.0047
Ch_{43}	0.0018	0.0042	Cho_{34}	0.0088	0.0035
Ch_{44}	0.5245	0.0259	Cho_{44}	0.0424	0.0105
Ch_{51}	-0.0572	0.0038			
Ch_{52}	-0.05	0.0029			
Ch_{53}	-0.0456	0.0072			
Ch_{54}	-0.0866	0.0048			

Notes: This table presents the estimates of parameters that generate B matrix. The definition of the parameters are described in Appendix 1.10.1.1. The sample is 1137 households across 57 counties with 11376 household-week observations in Texas in 2012 from Nielsen Consumer Panel Dataset (HMS).

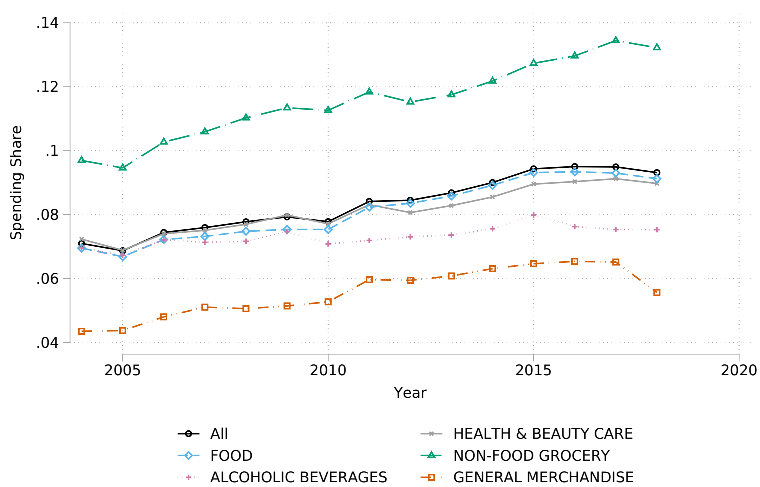
1.10.3 Figures

Figure 1.13: Spending Share in Supercenters and Club Stores

(a) Supercenters



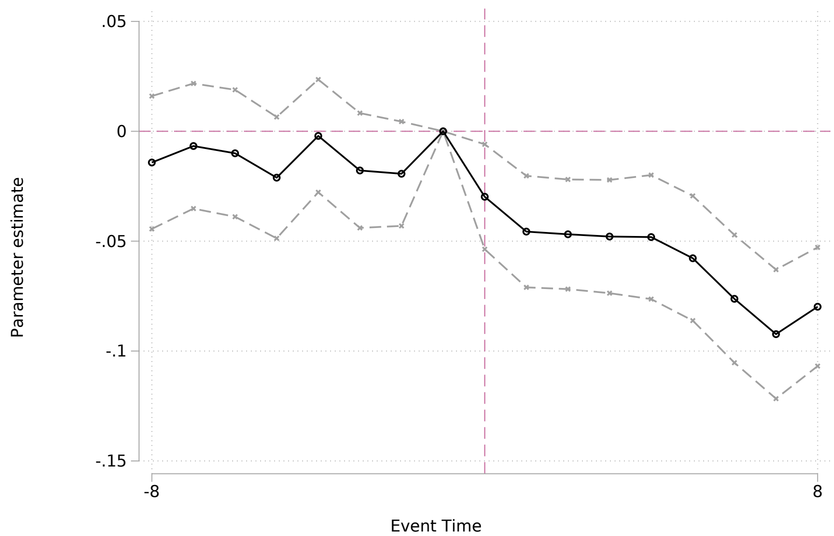
(b) Club Stores



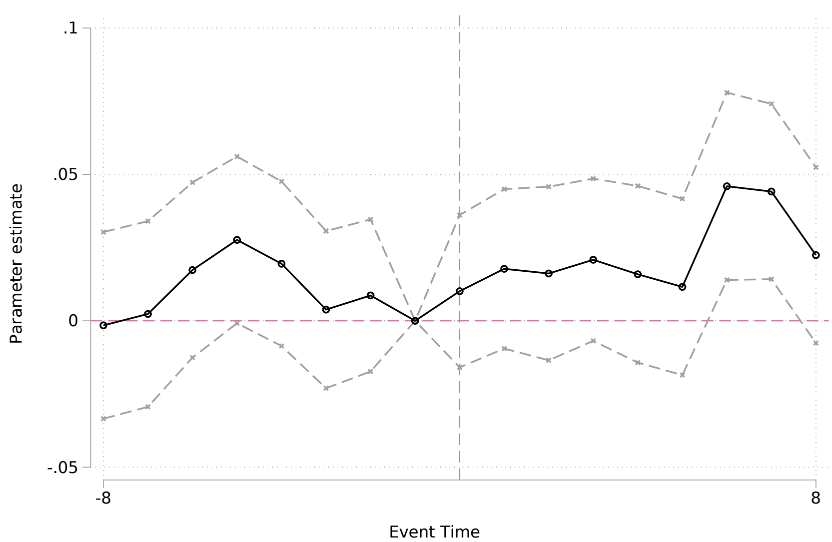
Notes: These figures present average household spending shares in supercenters and club stores across product departments, using 2004-2018 Nielsen Consumer Panel Dataset (HMS) at the household-by-year level.

Figure 1.14: Event Study Graph: Number of Retailers Visited

(a) Supercenter Entry



(b) Club Store Entry



Notes: Figure 1.14a uses 2004-2013 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of the Walmart Supercenter from Arcidiacono et al. (2020). Figure 1.14b uses 2004-2015 Nielsen Consumer Panel Dataset (HMS) at the household-by-quarter level and opening records of Costco, Sam's Club, and BJ's, from Coibion, Gorodnichenko and Koustas (2021). The figures present coefficients for eight leading and lagging periods of supercenter/club store entries, and 95% confidence intervals from estimates of the event study. The dependent variable is the total number visited by households. All regressions control for year-quarter indicators and household fixed effects.

CHAPTER 2

RISING RETAIL CONCENTRATION: SUPERSTAR FIRMS AND HOUSEHOLD DEMAND

written jointly with Justin Leung

2.1 Abstract

This paper investigates market concentration in the US retail sector. We use store- and household-level consumption micro-data from 2004-2019 to document four facts linked by a decomposition: 1) rising national concentration, 2) negligible change in regional concentration, 3) rising household concentration and heterogeneity across households, and 4) increased one-stop shopping. On the supply side, we find that increasing local availability of superstar retailers, rises in product variety, and changes in pricing explain a portion of these trends. On the demand side, we find that increases in households' opportunity cost of time are also key drivers. We develop a model that can rationalize these results to highlight the implications for market power and welfare.

2.2 Introduction

Superstar firms have continually attracted the attention of researchers, policymakers, and the public. Market concentration measures are used as a key metric to gauge the impact of these firms on the economy.[1] A large literature has documented increases in market concentration in the US economy over the last three to four decades. Are these trends mostly driven by firms on the supply side? Are they related to evolutions on the demand side driven by consumers? How do interactions between firms and consumers contribute to these trends and what are the resulting implications? In this paper, we attempt to address

1. For example, see the Horizontal Merger Guidelines by U.S. Department of Justice and Federal Trade Commission (2010).

these questions by focusing on the retail sector with detailed consumer and retail scanner data from 2004-2019.

Similar to the broader economy, the US retail sector has witnessed significant changes in recent decades. National market concentration has risen substantially. However, a unique feature of the retail sector is the ability to engage in one-stop shopping. Such distinctive retailing environments can be defined as a product market itself, separate from the actual goods purchased (Balto 2001). This implies that even market-level concentration may be an insufficient metric, since it may not fully reflect the extent of one-stop shopping.

How do these different levels of market concentration relate to each other? We show that they can all be linked by a decomposition. The national retail Herfindahl-Hirschman index (HHI) of spending across retailers can be decomposed into a revenue-weighted average of regional HHIs minus the cross-region variance, which captures how different the market share distributions are across regions. The expansion of chains into more geographic locations decreases this variance. Similarly, regional HHI can be decomposed into the average household HHI minus cross-household variance. The variance term reflects heterogeneity across households in their choice of retailers. Household HHI can be further decomposed into the average household-product department HHI minus the cross-department variance. The variance term can reflect one-stop shopping, since the market share distribution across product categories within household becomes more similar as one-stop shopping increases. We implement this decomposition using the Nielsen Consumer Panel Dataset from 2004-2019 to document four facts in the retail sector: 1) rising national concentration, 2) negligible change in regional concentration, 3) rising household concentration and heterogeneity across households, and 4) increased one-stop shopping.

We find that households are shopping more in their preferred stores. From 2004 to 2019, the average household retail HHI has increased by 17% from about 0.35 to 0.41. This rise holds within different demographic groups. In accordance with these facts, households visited fewer retail chains annually. The number of retail chains that an average household visited per year dropped by 15% despite a growing number of retail establishments in local markets. These facts imply that households are not constrained to fewer choices of retailers. Decomposing the trend in local retail concentration, we find that cross-household variance has increased by a comparable magnitude to that of within-household retail concentration. This signifies that different households are actually concentrating on different retailers over this period.

Furthermore, households were not only concentrating purchases more in some retailers within product departments, but also shopping for products in different departments increasingly in the same retailer. Decomposing the changes in household retail HHI, we find that around 60% of the increase is driven by growth in retail concentration within product departments and the remaining 40% comes from a decrease in cross-department variance. Households also made fewer shopping trips while real expenditure remained stable. Correspondingly, households have been spending more and purchasing more varieties of products per trip, pointing to more one-stop shopping.

How are these facts about household consumption related to evolutions in supply and demand in the retail sector? We attempt to provide causal empirical evidence to address this question. On the supply side, we first investigate the impact of the increasing availability of superstar big-box retailers such as supercenters and club stores. We utilize an event-study approach to estimate the impact of the entry of these superstar retailers. We find that an additional supercenter raises household HHI by about 1 percentage point and increases varieties per trip, while an additional club store actually lowers household HHI by about 1 percentage point and decreases varieties per trip. We show that our results exhibit parallel pre-trends and pass a series of robustness checks. Using these estimates, we calculate a back-of-the-envelope estimate (BOTE) for how much these entries explain the rise in household HHI. We find that supercenter entries can explain about 17% of the rise in household HHI within regions for which supercenter entries take place, but many regions did not experience a supercenter entry, such that supercenter entries only explain about 2% of the rise in household HHI across the US.

To offer potential explanations for why supercenters and club stores exhibit different effects, we document that households buy far more varieties than any other channel type, including club stores, and pay relatively low prices. This suggests that superstar retailers with larger assortments may allow households to benefit from demand-side economies of scope and engage in one-stop shopping. They may also achieve lower prices by benefiting from economies of scale to attract more customers.

We directly investigate these hypotheses by estimating the effect of changes in product variety and prices within existing stores. We follow the recent literature on uniform pricing and assortment similarity within retail chains (DellaVigna and Gentzkow (2019)) and utilize an instrumental variable (IV) based on the variety and pricing of products in other stores of the same retail chain. We estimate that as stores increase their variety, households in the region tend to increase their HHI and varieties per trip. When stores charge lower prices relative to its competitors, households in the region also increase their HHI. Our BOTEs imply that increases in variety over this sample period due to national chain-level changes can explain 5-20% of the increase in household HHI, while changes in prices did not substantially account for the rise in household HHI, since superstar retailers were not charging increasingly lower prices. This provides additional support to the hypothesis that demand-side economies of scope and supply-side economies of scale can both increase retail concentration.

After showing that supply-side extensive and intensive-margin decisions by retailers can explain a portion of the rise of household HHI and one-stop shopping, we now turn to the demand side by studying the impact of changes in time costs among households. Following the literature on opportunity costs of time, we use variation in certain household characteristics over time and proxies of labor demand shocks across demographic groups as IVs for the amount of time spent shopping, as measured by the number of shopping trips. We estimate that as shopping time decreases, household HHI increases, while cross-household variance increases and varieties per trip increases. We document that the opportunity cost of time, as proxied by average hourly wages and unemployment rate, has risen over the sample period across demographic groups. We find that proxies for labor demand shocks by age and education group do increase the opportunity cost of time and can explain up to 30% of the rise in household HHI.

Do these facts reflect a rise in the market power of superstar chains and thus a decline in consumer welfare? We develop a model that can rationalize our results and supplement it with findings from the existing literature. Based on our model, supply-side changes that increase household retail concentration such as an increase in product variety by firms, lower prices, and the increasing prevalence of supercenters will increase consumer welfare. Demand-side changes that increase the opportunity cost of time and increase household concentration may either increase or decrease welfare, depending on the source of these changes. While increases in the cost of travel will decrease welfare, wage increases, improvements in leisure technology, and rises in preference for leisure can increase welfare. The literature has generally found evidence that the supply-side developments we document that increase retail concentration tend to raise consumer welfare through economies of scale and scope. However, the demand-side developments are again more ambiguous, as a rise in one-stop shopping could be a result of demand-side changes that either increase or decrease consumer welfare. Our model also shows how aggregate market power changes as a result of our empirical findings is ambiguous and requires a more detailed quantitative analysis.

First, our paper contributes to a rich literature on concentration and market power as previously mentioned and summarized in papers such as Berry, Gaynor and Scott Morton (2019) and Syverson (2019). These papers highlight that many economists and policymakers are expressing concern over the possibility of increasing monopoly power in the US and the world economy. A large and growing literature documents rising market concentration since 2000 or earlier. However, the theoretical relationship between market concentration and average market power is ambiguous. Many empirical studies find patterns of simultaneous concentration and productivity growth, and Syverson (2019) argues that the case for large and general increases in market power is not yet dispositive. These papers call for a surge in industry-level research to characterize heterogeneity more fully both across and within markets, suggesting that sources of these patterns may be multi-causal, all with potential implications for market power in possibly different directions. Complementing Hsieh and Rossi-Hansberg (2019), who highlight a new industrial revolution in services, retail, and wholesale due to the availability of a new set of fixed-cost technologies that lower marginal costs in all markets, we use detailed micro-data on firms and consumers in the retail sector to decompose rising concentration and provide causal evidence to quantify various underlying mechanisms such as store entry, variety, and pricing. We show that these supply-side mechanisms interact with evolutions in consumer demand due to the rising value of time. We also complement Neiman and Vavra (2021), who find similar patterns of increasing household concentration and heterogeneity in product markets. They conclude that increasing product variety drives these trends. By contrast, we document patterns in retail markets and highlight increases in one-stop shopping.

Second, our paper adds to a literature on the impact of big-box stores, as summarized in Carden and Courtemanche (2016) and Ellickson (2016). For example, Hwang and Park (2016) document the impact of Walmart supercenter conversion, as opposed to entry, on varieties per trip, similarly finding that consumers increase one-stop shopping. In contrast, we jointly consider the entry of both supercenters and club stores and attempts to highlight their differential effects. Atkin, Faber and Gonzalez-Navarro (2017) provide reduced-form evidence on the impact of Walmart entry in Mexico and estimate large welfare gains for households. Leung and Li (2021*a*) similarly provide reduced-form evidence on the impact of big-box store entry in the US and quantify various sources of welfare gains using a different model to show product variety is a key differentiating factor between supercenters and club stores. In contrast, this paper focuses on how big-box store entry by superstar firms, along with within-store changes in product variety and pricing by national retail chains, contribute to the rise in retail concentration. Basker, Klimek and Van (2012) show that general merchandisers that added the most stores also made the biggest increases to their product offerings, and explain these facts with a stylized model in which a retailer's scale economies interact with consumer gains from one-stop shopping to generate a complementarity between a retailer's scale and scope. We show that this interaction between retailers and consumers can be used to explain a different set of trends.

Third, our paper contributes to a large literature on time use in economics and marketing. In economics, a wide variety of papers focus on the substitution between time and market goods, highlighting how households trade off shopping time for lower prices through increased shopping effort over the lifecycle and business cycles (e.g. Aguiar and Hurst 2007*b*; Aguiar, Hurst and Karabarbounis 2013; Nevo and Wong 2019). Coibion, Gorodnichenko and Koustas (2021) also document a drop in shopping frequency and highlights the implications for measurement of consumption inequality. In marketing, Bronnenberg (2018) summarizes the literature on how structural changes in consumers' time allocation impact retail strategy, and conversely, how retail innovations that make purchasing and home production more convenient impact purchasing habits and time use of consumers.[2] We add causal evidence using panel variation and micro-data on how increases in opportunity cost of time raise demand for one-stop shopping and affects retail concentration in recent decades. A closely related paper to ours is Bronnenberg, Klein and Xu (2020), who study how the availability of additional time shifts a households' shopping bundle towards more time-intensive goods in the Netherlands. In contrast, we study how retail strategy and household production interacts to increase market concentration in recent decades in the US retail sector, using empirical strategies developed in this literature.

2. Early work includes Messinger and Narasimhan (1997), who develop a model of retail formats based on consumers' increased demand for one-stop shopping and estimate it using time-series variation in aggregate data from 1961-1986. They argue that growing demand for time-saving convenience drives increasing assortment.

2.3 Data

2.3.1 Nielsen Consumer Panel

The Nielsen Consumer Panel Dataset (henceforth HMS) represents a longitudinal panel of approximately 40,000 to 60,000 US households from 2004 to 2019 who continually provide information to Nielsen about their households and what products they buy, as well as when and where they make purchases.[3] Panelists use in-home scanners to record all their purchases, from any outlet, intended for personal, in-home use. Products include all Nielsen-tracked categories of food and non-food items, across all retail outlets in the US. Nielsen samples all states and major markets. Panelists are geographically dispersed and demographically balanced. Each panelist is assigned a projection factor, which enables purchases to be projectable to the entire US.

Panelists report the products they purchase in each shopping trip. For each product as defined by its universal product code (UPC), we know the quantity purchased and total price paid for all units. Over 5 million products are further classified into about 1100 product modules, 125 product groups, and 10 product departments, which allows us to calculate varieties at various levels. A de-anonymized retail chain identifier is specified for each trip so that we are able to calculate the market share of each retail chain. We also observe where the household resides at various geographic levels from the Nielsen Scantrack market level (Nielsen classifies regions into around 50 market areas) down to the level of county and 5-digit zip code.

3. The data are available through a partnership between NielsenIQ and the James M. Kilts Center for Marketing at the University of Chicago Booth School of Business. Information on access to the the consumer panel data as well as the retail scanner data described below is available at http://research.chicagobooth.edu/nielsen/.

2.3.2 Nielsen Retail Scanner

The Nielsen Retail Scanner Dataset (henceforth RMS) consists of weekly pricing, volume, and store merchandising conditions generated by participating retail store point-of-sale systems across the US from 2006 to 2019. Data are included from approximately 30,000-50,000 participating stores and include store types such as drug, grocery, and mass merchandise stores, covering around 53-55% of national sales in food and drug stores and 32% of national sales in mass merchandise stores. The finest location of each store is given at the county level. We use this data to supplement our main analysis using HMS whenever needed, since RMS contains richer information at the store-level, recording every UPC that had non-zero weekly sales in each covered store.

2.3.3 Store Locations

We obtain the store locations and opening dates of several superstar retail chains from 2004-2013 using data from Arcidiacono et al. (2020) and Coibion, Gorodnichenko and Koustas (2021). These include Walmart supercenters and three club chains: Costco, Sam's Club, and BJ's. This allows us to conduct event studies to study the impact of superstar big-box retailers, which we describe in detail in Section 2.5.1.

2.4 Descriptive Evidence

We document several motivating facts using the HMS. First, we calculate the HHI at four decreasing levels of aggregation: nationally, regionally at the Scantrack market and county level, and at the household level. We show the specific formulas for measuring HHI in Appendix 2.11.1. Figure 2.1 shows that national HHI has been rising throughout the entire sample period from 2004 to 2019 by over 3 percentage points. We see a similar but more moderate increase at the market level of about 2 percentage points, but at the county level, the trend becomes almost flat with a roughly 1 percentage point increase. However, a trend that shows the largest increase is recovered at the household level, increasing from around 0.35 in 2004 to almost 0.41 in 2019, a 17% increase (6 percentage points).

To reconcile these facts, we show that changes in HHI at different levels of aggregation can be linked by a decomposition following Radaelli and Zenga (2002) (RZ). We show in Appendix 2.11.1.1 the exact formulas for this decomposition. In short, the national HHI can be decomposed into the revenue-weighted average of regional HHIs minus the cross-region variance, which captures how different the market share distributions are across regions. A larger variance implies a large difference across regions. Hence, a rising national HHI along with a flat county HHI implies that counties are becoming increasingly similar in their market share distributions even as the weighted average of HHI within each county is roughly flat, as shown in Figure 2.2a.[4]

Likewise, this decomposition can be applied at each lower level of aggregation. A rising household HHI is consistent with a flat county HHI when households increasingly buy at their preferred retailers while different households increasingly concentrate on different retailers, leading to an increase in the cross-household variance as shown in Figure 2.2b. We then further decompose household HHI into household-product-category HHI and cross-category variance in Figure 2.2c, where we define each product category at the product-department level. We find that the household-category HHI is increasing while the cross-category variance is decreasing, implying that households are increasingly buying different product categories at the same retailer.[5]

Next, we investigate whether the increase in household retail concentration is driven by households with certain demographic characteristics. Figure 2.3 plots the changes in household HHI for various household income groups, households living in more urban vs. rural counties, and various age and employment status groups by the gender of the household

4. We show these decompositions for the market level in Appendix Figure 2.12.

5. We show that these results are not driven by compositional changes in our dataset by also using a Dynamic Olley-Pakes decomposition (Melitz and Polanec 2015) on top of the RZ decomposition in Appendix Table 2.7, with the details shown in Appendix 2.11.1.1.

head. While we find differences in the levels of HHI across different demographic groups, we find that the HHI is increasing in nearly every demographic group albeit at different speeds.

Does the increase in national and household HHI imply a decreasing availability of retailers for households? We suggest that this is not the case in Figure 2.4. While the rise in household HHI is indeed driven by a 15% decrease in the number of retailers visited each period, the real expenditure has barely contracted over the sample period. The number of drug, grocery, and mass merchandise retail establishments per county has also risen fairly substantially over this period. We also find that households are decreasing their frequency of shopping trips, spending less days per week shopping.

To further unpack the rise in cross-category variance for households, we show in Figure 2.5 that households are indeed spending more per trip. They do this partly by increasing the number of varieties per trip, whether measured by the number of UPCs, brands, product modules, or product groups.

These changes in the number of trips, retailers visited, and real expenditures by households can be linked as shown in Figure 2.6. The number of trips can be decomposed as the number of retailers multiplied by the number of trips per retailer, while real expenditure can be decomposed as either the number of retailers multiplied by the real expenditure per retailer or the number of trips multiplied by the real expenditure per trip. We show that the drop in the number of trips is driven entirely in the drop in retailers visited as opposed to a drop in trips per retailer. The real expenditure per retailer and trip has increased, consistent with our previous findings.

Why are households increasingly shopping in their preferred retailers but concentrating on different retailers? Why are they decreasing their shopping trips while increasing their expenditure and varieties per trip even as the number of retail establishments has increased? We turn to providing plausibly causal evidence to explain these facts in the next section.

2.5 Reduced-Form Evidence

In this section, we present reduced-form evidence for a series of potential explanations for the trends we observe in the previous section. In each subsection, we list out both the empirical strategy as well as the results. On the supply side, we first analyze the effect of the entry of several big-box superstar retailers in Section 2.5.1. We then investigate the impact of changes in product variety and prices by retail stores in Section 2.5.2. On the demand side, we study the impact of changes in time costs among households in Section 2.5.3.

2.5.1 Entry of Superstar Big-box Retailers

2.5.1.1 Empirical Strategy

To study the impact of the increasing availability of superstar big-box retailers such as supercenters and club stores, we utilize an event-study approach to estimate the impact of the entry of these superstar retailers. Our baseline independent variable measures the number of stores for each chain within the 5-digit zip code of each household. We also calculate alternative distance measures. As shown in equation (2.1), we then regress our outcome of interest for household i in quarterly period t, for example the household retail HHI, on the number of stores, and add household fixed effects to control for fixed household characteristics, as well as period fixed effects to control for national time trends.

$$Y_{it} = \beta \times Num_{it} + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{2.1}$$

If a store enters in periods when unobservable local household characteristics change, or households anticipate these openings by changing patterns in significant ways, then this would be a threat to our identification. A priori, we believe that it is difficult for households or stores to exactly time sharp changes in unobservables with store entry. To further alleviate these concerns, we estimate the trends before and after the entry event by adding leads and lags of the independent variable Num_{it}. If the pre-trends are parallel, we argue that this gives additional evidence to suggest that stores or households cannot align changes in unobservables to the precise timing of the entry.

2.5.1.2 Results

We show the results of estimating equation (2.1) in Table 2.1. We estimate the effect of an additional supercenter or club store respectively in a household's 5-digit zip code. We also separately estimate the effect for only zip codes with at least one entry event or all zip codes. We find that an additional supercenter raises household HHI by 0.008 while an additional club store lowers household HHI by 0.012. These results are statistically significant at the 1% level.

We calculate a back-of-the-envelope estimate (BOTE) of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household HHI in the sample period. Leung and Li (2021*a*) and Hortaçsu and Syverson (2015) both document the dramatic rise of supercenters and club stores over this period. We find that in zip codes with at least one entry, the rise in the number of supercenters explains about 17% of the rise in household HHI, while the rise in club stores decreases HHI, explaining about -8% of the rise in household HHI. Since many zip codes did not experience an entry event, the BOTE is much smaller at about 2% and -1% respectively for all regions.

In Figure 2.7, we show that pre-trends are parallel around store entries roughly two years before the event, while the effects are dynamic and continue to rise for an extended period after the event. This is consistent with both households taking time to learn about the presence of new stores and adjusting their purchasing habits. We also show that our results are robust to using alternative measures of distance in Appendix Table 2.8 such as distance to the nearest store or the number of stores within a certain mile radius. The fact that the effect dissipates as the distance from each household rises increases our confidence in a causal interpretation of our estimates.

To provide suggestive evidence of why supercenters generate a different effect from club stores, we show how supercenters and club stores differ in two characteristics that typically define a retailer: product variety and prices.

In Figure 2.8, we show the average number of UPCs and product modules per household-quarter for supercenters, club stores, and other channel types over the sample period. Households buy far more varieties in supercenters than any other channel type, whereas club stores sell fewer UPCs and product modules compared with grocery stores and supercenters, with the number of varieties close to other channel types such as discount stores and dollar stores, but higher than drug stores and other miscellaneous channel types. This would be consistent with the hypothesis that superstar retailers with larger assortments allow households to benefit from demand-side economies of scope and engage in more one-stop shopping, increasing their household HHI. We directly investigate the effect of variety in Section 2.5.2.

In Figure 2.9, we calculate the relative price index (RPI) of each retailer or channel type following Aguiar and Hurst (2007*b*) (hereafter AH). To construct a retailer RPI, we calculate the ratio between total expenditure for each good and the counterfactual expenditure of each good at its average price in the reference region. We then take the weighted average across goods and counties to calculate a national RPI for each retailer that uses national averages as reference prices. We find that supercenters consistently offer lower prices than its competitors nationally, although their price advantage has been decreasing. While club stores generally have an RPI below one, they tend to have higher RPIs than supercenters over the sample period, with the exception of Club 2 offering lower prices in the last periods, which our entry data does not capture. This would be consistent with the hypothesis that superstar retailers with lower prices may attract more households to benefit from supply-side economies of scale, increasing their household HHI. However, ex-ante it is difficult to predict how prices mediate the effect of big-box store entry on household HHI, since household HHI may be concentrated before store entry precisely because low-price alternatives are not available. We directly investigate the effect of prices in Section 2.5.2.

We investigate the effect of entry on other measures of concentration in Table 2.2. We find that the entry of supercenters decreases the number of retailers visited by households while club stores have the opposite effect. The rise in household HHI due to supercenter entry is driven both by a rise in household-category HHI and a drop in cross-category variance, while the drop in household HHI due to club entry is driven mostly by a drop in household-category HHI. Cross-household variance decreases but the change is statistically insignificant.

We estimate the effect of entry on the number of trips per quarter and varieties per trip for households in Table 2.3. Supercenters actually do not decrease the number of trips by a significant amount while clubs increase the number of trips by a significant amount. Nonetheless, supercenters do increase the number of varieties per trip, in particular the number of departments per trip, while club stores have the opposite effect. This is once again consistent with larger assortments in supercenters relative to club stores.

How do these effects vary across households? We use tools from Chernozhukov, FernndezVal and Luo (2018) to explore heterogeneous treatment effects across households. To do this, we add interaction terms to equation 2.1 using variables such as household income, size, age, and region, to estimate sorted effects. We then analyze whether characteristics are different across the most affected and least affected households. We plot the sorted effects in Appendix Figure 2.13. While the effects are indeed heterogeneous, we find that most of the differences in household characteristics between the most and least affected groups are not statistically significant. The point estimates do suggest that lower-income, younger, single households living in more rural areas are more likely to increase concentration when facing a supercenter entry. However, even for these households, BOTEs suggest that supercenter entry does not fully explain the rise in household HHI.

Overall, we find that supercenters do increase household HHI but they do not explain the entire increase over the sample period. This is true both because supercenter entries can explain only about 17% of the rise in household HHI within regions for which supercenter entries take place, and because many regions did not experience a supercenter entry. On the other hand, club stores actually work in the opposite direction. This suggests that supply-side changes in the retail landscape as measured by entry of superstar big-box retailers contribute only partly to the rise in household HHI.

2.5.2 Variety and Prices

2.5.2.1 Empirical Strategy

To estimate the effect of changes in product variety and prices by retailers on households, we follow the recent literature on uniform pricing and assortment similarity within retail chains (DellaVigna and Gentzkow 2019) and utilize an instrumental variable (IV) that is based on the variety and pricing of products of other stores in a given retail chain. This IV strategy relates to the one in Hausman and Bresnahan (2008) and has been employed recently in DellaVigna and Gentzkow (2019) and Allcott et al. (2019) among others.

Specifically, our estimating equation is as follows:

$$Y_{it} = X'_{rt}\gamma + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{2.2}$$

We regress our outcomes of interest such as household retail HHI on a vector of variety and prices for each region-period rt. This vector includes three variables which are logged region-level revenue-weighted averages: (1) the number of products (as measured by UPCs) per store, denoted as variety depth (2) the number of product modules per store, denoted as variety breadth, and (3) the price index for each store. Each variable x_{rt} is constructed as the revenue-weighted average for each store s in region r:[6]

$$x_{rt} = \sum_{s \in r} w_{st} x_{st} \tag{2.3}$$

To isolate variation coming from supply-side changes that are plausibly exogenous to unobservables that affect local household outcomes, we construct an IV z_{rt} excluding all stores in region r:

$$z_{rt} = \sum_{c \in r} w_c \frac{\sum_{s \in c} \left(w_{st} x_{st} - \sum_{s' \in r} w_{s't} x_{s't} \right)}{\sum_{s \in c} \left(w_{st} - \sum_{s' \in r} w_{s't} \right)} \tag{2.4}$$

For each retail chain c, we first construct its revenue-weighted national average of the variable leaving out the region of interest. We then weight it by the revenue share it earns in that region in the entire sample period w_c, allowing us to hold the weight fixed across time. Therefore, the identification assumption is that retailers price and stock products similarly across their chains, such that national supply shocks to the chain affect local prices and assortment, but are plausibly exogenous to unobservable demand shocks that affect our household outcomes.

2.5.2.2 Results

We show the results of estimating equation 2.2 in Table 2.4. We estimate the effect of a percentage change in variety depth, variety breadth, and prices respectively. We construct these variables using both the HMS and the RMS for comparison, harnessing the strength of each dataset. For the HMS, we construct an RPI as described in Section 2.5.1.2 since it has broader cross-sectional coverage nationally. We use both region (county) and national reference prices. For the RMS, we construct a store price index following Leung (2021) due to its ability to observe products at higher frequencies in each store. Likewise, for the variety

6. In the HMS, we only observe each store as a retail chain-region pair cr.

measure, the HMS will be able to capture a broader set of retailers while the RMS is able to capture variety within store more precisely for the set of stores it contains.

We find that increasing variety as well as lower prices both lead to a rise in household HHI. While variety depth and breadth work in opposite directions when controlling for each other, the net effect is positive given our BOTEs.[7] This is consistent with results in the previous section, where entry of supercenters with both more variety and lower prices increases household HHI. This also implies that demand-side economies of scope and supply-side economies of scale can both increase retail concentration.

Our BOTEs imply that increases in variety over this sample period due to national chain-level changes can explain 5-20% of the increase in household HHI, while changes in prices did not substantially account for the rise in household HHI.[8] This is because as seen in Figure 2.9, the largest retailers were not offering increasingly lower prices relative to competitors. We show that nationwide, stores are indeed offering more varieties in Figure 2.14.

We also estimate equation 2.2 for other measures of concentration in Appendix Table 2.11 and 2.12. These results remain consistent with Table 2.4. Increases in variety and lower prices decreases the number of retailers visited. More variety and lower prices increases household-category HHI and decreases cross-category variance, although less so for variance. The effect on cross-household variance is small overall.

We estimate the effect of variety and prices on the number of trips per quarter and varieties per trip for households in Appendix Table 2.13 and 2.14. Increasing variety significantly increases the number of varieties per trip, while changing prices has a more mixed and much smaller effect. The effect on the number of trips is mixed across the datasets, but is roughly consistent with a rise in variety contributing to the decrease in the number of trips, as shown

7. Increasing variety depth holding variety breadth constant would be increasing the number of UPCs within each product module holding the number of product modules constant, while increasing variety breadth holding variety depth constant would be increasing the number of product modules while stocking fewer UPCs per module.

8. To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively.

by the BOTE in Table 2.14.

2.5.3 Time Costs

2.5.3.1 Empirical Strategy

To estimate the effect of changes in the opportunity cost of time on households, we follow AH and use the number of shopping trips per household in each time period as a proxy for the amount of shopping time spent. Figure 5 in Coibion, Gorodnichenko and Koustas (2021) offers support for the use of this proxy, since it shows similar trends in shopping time using the American Time Use Survey (ATUS). We then evaluate how plausibly exogenous changes in shopping time affect household retail concentration. Specifically, our estimating equation is as follows:

$$Y_{it} = \beta \times N_{it} + X_{it}'\gamma + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{2.5}$$

We regress our outcomes of interest such as household retail HHI on the number of shopping trips N_{it} for each household-period it along with a vector of control variables. We then instrument for the number of shopping trips with a series of IVs following AH and Aguiar, Hurst and Karabarbounis (2013).

First, we use household characteristics as IVs. By including household and time fixed effects, we can extend beyond using cross-sectional variation as in AH and leverage within-household variation in household characteristics, using household age, size, income, and unemployment as IVs. A caveat is that these IVs themselves may reflect changes in unobservables and shopping needs. We address these threats to identification in several ways. First, we directly control for shopping needs X_{it} as in AH by including the log of the quantity index derived from the RPI as well as the log number of UPCs and product groups purchased per period. Second, we show that our results are robust to using a wide range of IVs, which assuages concerns that the IVs themselves may be reflecting changes in unobservables.

Second, we use proxies of labor demand shocks as IVs to further address threats to identification. We first document that based on the Current Population Survey Merged Outgoing Rotation Groups (CPS MORG), average hourly wages have been rising over the sample period and unemployment rate rose substantially during the recession but gradually dropped to levels below those before the recession. This is consistent with the fall in shopping trips shown in Figure 2.4, which stagnated during the recession before falling further during the recovery. We show the trends for average wages and unemployment rate by gender and age group in Figure 2.10, and by gender and education group in Figure 2.11. As seen in these figures, trends do differ by gender, age, and education.

We follow Aguiar, Hurst and Karabarbounis (2013) and Aguiar et al. (2021) by using these differential changes as proxies of labor demand shocks that affect the opportunity cost of time, using either state- or national-level changes in both average wages and unemployment rates by household group as IVs. Our identifying assumption is that these changes reflect differential shifts in labor demand across various household demographic characteristics that are plausibly exogenous to unobservables affecting individual household shopping behavior. We construct these proxies at various levels of aggregation across household characteristics in the CPS MORG, such as age, education, age-education, and age-education-occupation cells by state or across the US. We match these proxies to households in the HMS using the mode of their characteristics over the sample period to avoid using variation induced by shifts in characteristics over time.

2.5.3.2 Results

We show our results using household characteristics as IVs in Table 2.5. We first show OLS results with and without controls for shopping needs. We find that adding controls increases the effect of shopping trips on household HHI. A 1% decrease in the number of trips increases household HHI by about 0.2 percentage points. This is consistent with the fact that unobservables affecting household HHI, such as changes in shopping needs, may also affect the number of shopping trips. For the remaining specifications, we continue to control for shopping needs to isolate the effect of time costs on household HHI separate from the effect it also has on shopping quantities.

We find that using household age, size, and employment as IVs generate nearly identical coefficients. These results are also robust to adding a county-time fixed effect to control for potential supply-side changes that vary over time within a county. Using household income as an IV roughly halves the coefficient, but the estimate remains statistically and economically significant.[9] The size of the first-stage F-stat implies the IV is relevant and the signs of the first-stage coefficients are consistent with a priori reasoning, i.e. increased household age, decreased household size, and unemployment all increase the amount of shopping trips.

Using the average change in shopping trips over the sample period to calculate our BOTE, we find that shopping time has the potential to explain the entire rise in household HHI, with BOTEs ranging from 40%-100%. However, this BOTE can reflect both endogenous and exogenous factors affecting shopping time.

Therefore, we now focus on a potentially exogenous force decreasing shopping time by presenting results in Table 2.6 using national- or state-level changes in average wages and unemployment by household group as IVs. These groups include (1) age, (2) education, (3) interactions between age and education, (4) interactions between age, education, and occupation, and (5) age plus education combined. While the magnitudes of the estimated coefficients do vary by IV, the magnitudes are more similar for IVs that are relevant, and nearly all coefficients are statistically and economically significant. In most of the specifications, the first-stage F-stat passes the usual thresholds and the signs of the first-stage coefficient are consistent with a priori reasoning, i.e. rises in wages and drops in unemployment decrease shopping time. Focusing on the relevant IVs with a first-stage F-stat above 10, the BOTE using changes in shopping trips range from about 70% to 170%, while the

9. This may be because changes in income are particularly likely to also reflect changes in shopping quantity, which may not be fully captured by our controls.

BOTE using only changes in the IVs range from about 0-30%. This suggests that labor demand shocks by age and education group do increase the opportunity cost of time and can explain up to 30% of the rise in household HHI.

We show that results are qualitatively similar under a series of robustness checks, which include adding county-time fixed effects in Appendix Table 2.15, allowing for changes in household characteristics over time in Appendix Table 2.16, using median wages and unemployment rate in Appendix Table 2.17, and isolating the effect of average wages, median wages, and unemployment rate in Table 2.18, 2.19, and 2.20 respectively. In particular, we find that unemployment rates are stronger IVs than average or median wages.

We also show results using other measures of concentration and varieties per trip in Appendix Table 2.21 and 2.22. We find that decreases in shopping time increases the number of retailers visited, increases household-category HHI, decreases cross-category variance, increases cross-household variance, and increases varieties per trip, all consistent with the trends we see in Section 2.4.

In Appendix Section 2.11.2, we discuss the rise of online shopping as a related hypothesis, which may also induce households to engage in more one-stop shopping. We find that concentration trends barely change when considering only offline retailers, given that the share of online shopping remains small. This is consistent with Hortaçsu and Syverson (2015), who show evidence that online share of retail sales remains very small in the product categories we consider. They argue that although online retail will surely continue to be a force shaping the sector going forward and may yet emerge as the dominant mode of commerce in the retail sector in the US, its time for supremacy has not yet arrived. We also consider a few empirical strategies and find that the effect of online shopping is again negligible.

2.6 Model

In this section, we develop a model to rationalize our results and highlight the implications for market power and welfare. We model the demand-side as a two-layer nested CES utility function following Hottman, Redding and Weinstein (2016) among others. The first nest contains firms and the second nest contains products, which enables consumers to choose from any number of multi-product firms.[10] We then introduce a cost to visiting each firm following Bronnenberg (2015), allowing consumers to derive utility from leisure with a Cobb-Douglas consumption-leisure utility function. Given the model, we derive comparative statics for each of the variables of interest, showing the conditions under which evolutions in both supply and demand can increase household concentration and highlight the resulting welfare implications. All derivations are shown in Appendix Section 2.11.3.

2.6.1 Demand

Consumers have a Cobb-Douglas utility function and derive utility from (1) quantities of a composite consumption good $X(\mathcal{V})$, (2) the total variety of these goods $\mathcal{V}$, and (3) leisure $L(\mathcal{V})$, with a preference for leisure of ρ as shown in equation (2.6).[11] They maximize utility by choosing $X(\mathcal{V})$, $\mathcal{V}$, $L(\mathcal{V})$, and labor supply h, subject to a time constraint and budget constraint. The amount of time they have is T, which they split between shopping time $\tau(\mathcal{V})$, leisure L, and labor supply h. Shopping time $\tau(\mathcal{V})$ is the integral of the shopping cost for each variety $\mu(\nu)$ over all varieties $\mathcal{V}$. Consumers have income Y from supplying labor h at wage w and non-labor income K, which is used to buy consumption goods with price $p(\nu)$ for each variety ν. We can combine the time and budget constraints into the full income

10. We use the terms consumers and households interchangeably.

11. Results are nearly identical and give the same intuition using a separable consumption-leisure constant relative risk aversion (CRRA) utility function that is common in labor supply life-cycle models.

constraint in equation (2.7).

$$\max_{X,\mathcal{V},L,h} U(X(\mathcal{V}), L(\mathcal{V})) = X(\mathcal{V})^{1-\rho} L(\mathcal{V})^{\rho} \tag{2.6}$$

$$s.t. \quad \tau(\mathcal{V}) + L + h = T, \text{ where } \tau(\mathcal{V}) = \int_{\nu\in\mathcal{V}} \mu(\nu) d\nu$$

$$s.t.\ Y = wh + K = \int_{\nu\in\mathcal{V}} p(\nu)x(\nu)d\nu = PX(\mathcal{V})$$

$$s.t.\ wT + K = \int_{\nu\in\mathcal{V}} p(\nu)x(\nu)d\nu + w(\tau(\mathcal{V}) + L) \tag{2.7}$$

Consumption good $X(\mathcal{V})$ contains two CES nests for firms and products as shown in equation (2.8). The first nest consists of firms ν, which are retailers in our context. Each firm has a taste parameter $\varphi^F(\nu)$ and quantity consumed $x^F(\nu)$ with an elasticity of substitution σ^F between firms. The second nest contains products u, which enables consumers to choose a set of products $\mathcal{U}_f$ from firm f, with consumers choosing the set of multiproduct firms $\mathcal{V}$.[12] Each product has a taste parameter $\varphi^U(u)$ and quantity consumed $x^U(u)$ with an elasticity of substitution σ^U between products.

$$X(\mathcal{V}) = \left(\int_{\nu\in\mathcal{V}} \left(\varphi^F(\nu)x^F(\nu)\right)^{\frac{\sigma^F-1}{\sigma^F}} d\nu \right)^{\frac{\sigma^F}{\sigma^F-1}},\ x^F(\mathcal{U}_f) = \left(\int_{u\in\mathcal{U}_f} \left(\varphi^U(u)x^U(u)\right)^{\frac{\sigma^U-1}{\sigma^U}} du \right)^{\frac{\sigma^U}{\sigma^U-1}} \tag{2.8}$$

Solving for the demand at the firm and product level following Hottman, Redding and Weinstein (2016), we have the following equations:

$$x^F(\nu) = A(\mathcal{V}) p^F(\nu)^{-\sigma^F} \varphi^F(\nu)^{\sigma^F-1} \tag{2.9}$$

$$A(\mathcal{V}) = YP^{\sigma^F-1},\ P = \left(\int_{\nu\in\mathcal{V}} \left(\frac{p^F(\nu)}{\varphi^F(\nu)} \right)^{1-\sigma^F} d\nu \right)^{\frac{1}{1-\sigma^F}}$$

12. To maintain consistency with previous literature, we use ν and f interchangeably for firms.

$$x^U(u) = \left(\varphi_f^F\right)^{\sigma^F - 1} (\varphi_u^U)^{\sigma^U - 1} Y P^{\sigma^F - 1} (P_f^F)^{\sigma^U - \sigma^F} (P_u^U)^{-\sigma^U} \tag{2.10}$$

$$P_f^F = p^F(\nu) = \left(\int_{u \in \mathcal{U}_f} \left(\frac{P_u^U}{\varphi_u^U} \right)^{1-\sigma^U} du \right)^{\frac{1}{1-\sigma^U}}$$

$$= \left(\underbrace{N_f}_{\text{Scope}} \underbrace{\frac{1}{N_f} \int_{u \in \mathcal{U}_f} \left(\frac{P_u^U}{\varphi_u^U} \right)^{1-\sigma^U} du}_{\text{Average product taste-adjusted prices}} \right)^{\frac{1}{1-\sigma^U}} \tag{2.11}$$

The price index for each firm can be written as its scope N_f multiplied by the average product taste-adjusted prices.[13] Next, we can use the first-order condition (FOC) to derive the optimal composite good consumed and leisure as follows:

$$L(\mathcal{V}) = \frac{\rho \left(w \left(T - \tau(\mathcal{V}) \right) + K \right)}{w} \tag{2.12}$$

$$X(\mathcal{V}) = \frac{(1-\rho) \left(w \left(T - \tau(\mathcal{V}) \right) + K \right)}{P} = \frac{Y}{P} = \frac{w(T - \tau(\mathcal{V}) - L) + K}{P} \tag{2.13}$$

Substituting these expressions into the utility function, we can follow Bronnenberg (2015) and derive that the optimal cutoff variety ν_D in the set of varieties D satisfies the following condition:

$$\frac{A(D) \left(\frac{p(\nu_D)}{\varphi(\nu_D)} \right)^{1-\sigma^F}}{\sigma^F - 1} - w\mu(\nu_D) = 0 \tag{2.14}$$

$$A(D) = Y P^{\sigma^F - 1} = (1-\rho) \left(w \left(T - \tau(D) \right) + K \right) \left(\int_{\nu \in D} \left(\frac{p^F(\nu)}{\varphi^F(\nu)} \right)^{1-\sigma^F} d\nu \right)^{-1}$$

Equation 2.14 equalizes the marginal benefit and cost of each additional variety. The marginal benefit increases as the taste-adjusted price of firm f decreases relative to the price index P the household faces, decreases with elasticity of substitution σ^F, and increases with

13. We could further add a cost per product to generate zero purchases for certain products. However, we abstract away from this feature for tractability.

income Y. The marginal cost increases with wage w, which also represents the opportunity cost of time, and shopping cost μ. Each additional variety will be consumed whenever the marginal benefit exceeds the marginal cost.

Next, we simplify the model with the following assumptions for expositional purposes and as a result of our empirics, which find that variety varies far more than prices at the firm level, and prices do not drive changes in household retail HHI much but variety has a substantial impact. First, we assume that firms have identical taste parameters φ^F equal to one. Second, we assume that products are symmetric such that products have identical taste parameters φ^U equal to one and identical prices p_f in each firm f. We can index each firm f by their price index P_f^F, which is a function of the number of products it sells N_f and price p_f, abstracting away from idiosyncratic tastes for each product and firm.[14] Assume that P_f^F lies on a continuum $[\underline{P^F}, \overline{P^F}]$.[15] The firm price index for a firm, denoted as the cutoff variety ν_D, and N_D products can be written more simply as follows:

$$P^{F*} = p(\nu_D) = \left(\int_0^{N_D} p_D^{1-\sigma^U} dn \right)^{\frac{1}{1-\sigma^U}} = p_D N_D^{\frac{1}{1-\sigma^U}}$$

$$P_f^F = p_f N_f^{\frac{1}{1-\sigma^U}} = \frac{p_f}{N_f^{\frac{1}{\sigma^U - 1}}}, \; P = \left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \right)^{\frac{1}{1-\sigma^F}}$$

We assume that the shopping cost at each firm is a function of the price index P_f^F in each firm, with $\mu'(P_f^F) < 0$. This is consistent with the fact that firms with larger variety and lower prices, i.e. lower P_f^F, tend to be located farther away from consumers relative to firms with lower variety and higher prices, due to higher costs of land in areas with higher population density and the need for stores with larger square footage in order to stock

14. Allowing for firm tastes simply requires us to index each firm by their taste-adjusted price index and divide P_f^F by φ_f^F. Allowing for product tastes requires us to write the price index as in equation (2.11).

15. Assuming the firms lie on a continuum indexed by P_f^F allows for higher analytical tractability and the use of integrals. We can also allow for discrete P_f^F and use summations instead with similar intuitions for all of our derivations.

higher variety and lower prices through economies of scale. We can show using equation (2.14) that if $\mu'(P_f^F)$ is small enough in absolute value, the net marginal gain of shopping at an additional firm is monotonically decreasing in P_f^F, and there exists a unique cutoff firm P^{F*} that satisfies the following condition, such that consumers only buy from firms that have a price index within the set $[\underline{P^F}, P^{F*}]$:

$$\left(P^{F*}\right)^{1-\sigma^F} \frac{1}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} \frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{P^{F*}} \mu(p)dp\right) + K\right)}{w(\sigma^F - 1)} - \mu(P^{F*}) = 0 \tag{2.15}$$

2.6.2 Household Retail Concentration

Given our derivations, the market share for household i for each firm can be written as follows:

$$S_{fi}^F = \frac{P_f^F x^F(\nu)}{Y} = \begin{cases} \left(P \frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F - 1} & \text{if } \frac{Y P^{\sigma^F - 1} \left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}}{w(\sigma^F - 1)} \geq \mu(P_f^F) \\ 0 & \text{otherwise} \end{cases}$$

Given our simplifying assumptions, we can further simplify the market share expression and write the household retail HHI for household i as follows:

$$H_i = \int_f \left(S_{fi}^F\right)^2 df = \frac{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{2(1-\sigma^F)} dP_f^F}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \tag{2.16}$$

We can then show some comparative statics of household retail HHI in response to changes in various parameters. First, we show that any parameters, in this case denoted by t, that decrease the cutoff firm price index P^{F*} will increase household retail HHI. This is intuitive since the set of retailers that households consume from $[\underline{P^F}, P^{F*}]$ will decrease in size as

P^{F*} decreases.

$$\frac{dP^{F*}}{dt} < 0 \Rightarrow \frac{dH_i}{dt} > 0 \tag{2.17}$$

Given this fact, we can use equation (2.15) to derive comparative statics of the cutoff firm price index P^{F*} in response to changes in various parameters to see how these parameters change household retail HHI. These comparative statics can then be compared against our empirical findings.

First, consider changes on the demand side. Let the shopping cost be $\mu(P^F) = t\delta(P^F)$, where t is the time cost per distance traveled and $\delta(P^F)$ is the distance of the consumer from each firm indexed by P^F. We derive that if the condition required for the existence of a unique cutoff price index holds, we have the following results:

$$\frac{dH_i}{dt} > 0 \tag{2.18}$$
$$\frac{dH_i}{dw} > 0 \tag{2.19}$$
$$\frac{dH_i}{d\rho} > 0 \tag{2.20}$$

As the time cost per distance traveled t increases, household retail HHI increases. As wage w increases, household retail HHI also increases. As the preference for leisure ρ increases, household retail HHI also increases. Since the preference for consumption is written as $1-\rho$, we can instead write the preference for leisure as $\theta\rho$, where θ is a form of leisure technology. By varying θ, we can keep the preference for consumption unchanged. This would be consistent with increases in leisure technology highlighted in Aguiar et al. (2021). We find that similar to changes in ρ, we have

$$\frac{dH_i}{d\theta} > 0 \tag{2.21}$$

As leisure technology θ increases, household retail HHI increases. All of these results are intuitive and consistent with our empirical results that a rise in the opportunity cost of time raises household retail HHI. As the shopping cost increases or the relative cost of leisure decreases, households spend less time shopping and visit fewer retailers, raising household retail HHI.

Second, consider changes on the supply side such as the opening of new firms, which decrease μ. Denoting such changes as an increase in β, we can write $\mu = \mu(P^F, \beta)$ and $\frac{\partial \mu}{\partial \beta} < 0$. A rise in β generates the opposite effect as a rise in t, decreasing household retail HHI. This is because households can now spend more time shopping and visit more retailers, decreasing household retail HHI. This is consistent with our empirical results that the opening of club stores, which provide similar amounts of product variety to many other stores, decreases household retail HHI. Alternatively, consider the introduction of supercenters, which we model as a fall in $\underline{P^F}$, the lower limit for the range of firm price indices, since supercenters represent a shopping format that provides an unprecedented number of products at low prices in a single store. We find that a decrease in $\underline{P^F}$ raises household retail HHI if there is a sufficient decrease in P^{F*} as shown in Appendix Section 2.11.3, consistent with our empirical results. Intuitively, a fall in $\underline{P^F}$ represents the introduction of a new firm that lowers the overall price index, increasing the relative price of the initial cutoff price index P^{F*} such that P^{F*} falls.

$$\frac{dH_i}{d\beta} < 0 \tag{2.22}$$

$$\frac{dH_i}{d\underline{P^F}} < 0 \tag{2.23}$$

Third, consider changes on the supply side such as increasing economies of scale or scope, which decrease P^F by lowering prices or raising variety. Denoting such changes as an increase in α, we can write $P^F = P^F(\alpha)$ and $\frac{dP^F}{d\alpha} < 0$. A rise in α could either increase or decrease household retail HHI, depending on whether the firm price index decreases favor firms with a larger or smaller initial firm price index. Let $f(P^F, \alpha) = (P^F(\alpha))^{1-\sigma^F}$. We derive a condition under which household retail HHI would increase as α increases, which can be interpreted as economies of scale or scope that disproportionately favor firms with lower prices and larger variety, consistent with our empirical results.

$$\int f \frac{df}{d\alpha} > \int f \int \frac{df}{d\alpha} \Rightarrow \frac{dH_i}{d\alpha} > 0 \tag{2.24}$$

2.6.3 Welfare

Substituting the optimal quantity consumed X, variety $\mathcal{V}$, and leisure L into the utility, we can derive consumer welfare in this model. Differentiating utility with respect to various parameters in our model and using the envelope theorem, we can show how welfare changes in response to various changes on the demand side and supply side.

First, consider changes on the demand side. We have the following results:

$$\frac{dU}{dt} < 0 \tag{2.25}$$

$$\frac{dU}{dw} > 0 \tag{2.26}$$

$$L > \frac{Y}{P} \Rightarrow \frac{dU}{d\rho} > 0 \tag{2.27}$$

$$\frac{dU}{d\theta} > 0 \tag{2.28}$$

A rise in the time cost per distance traveled t lowers welfare, since more time is spent shopping, decreasing both income Y and leisure L. A rise in wage w raises welfare as it increases income Y. A rise in the preference for leisure ρ may increase (decrease) welfare depending on whether leisure L is greater (lower) than consumption $\frac{Y}{P}$. On the other hand, a rise in leisure technology θ unambiguously raises welfare by increasing the value of leisure. Therefore, the source of the rise in opportunity cost of time is crucial in determining whether welfare increases or decreases.

Second, consider changes on the supply side. We have the following results:

$$\frac{dU}{d\beta} > 0 \tag{2.29}$$

$$\frac{dU}{d\underline{P^F}} < 0 \tag{2.30}$$

$$\frac{dU}{d\alpha} > 0 \tag{2.31}$$

A rise in β which decreases shopping cost μ increases welfare in the same way as a reduction in t. A fall in minimum firm price index $\underline{P^F}$ increases welfare. A rise in α which decreases price indices within firms increases welfare. Therefore, all of these supply-side changes should increase welfare.

2.6.4 Heterogeneity

To account for changes in cross-household variance, we need to allow for heterogeneity across consumers. To model such heterogeneity in the simplest way possible, we first allow for two groups of consumers. The first group of consumers only buys from firms that provide firm price indices below the cutoff firm price index $\overline{P^{F*}}$ that we previously used, consuming from the set of firms $[\underline{P^F}, \overline{P^{F*}}]$. We introduce a second group of consumers that only buys from firms that offer a price index above the cutoff firm price index $\underline{P^{F*}}$, consuming from the set of firms $[\underline{P^{F*}}, \overline{P^F}]$. We find that as long as $\mu'(P^F)$ is large enough in absolute value, the net marginal gain of shopping at an additional firm is monotonically increasing in P^F_f, and there exists such a unique cutoff firm $\underline{P^{F*}}$. We can consider such consumers as those who find it far more costly to visit firms with large variety and low prices that are generally located away from population centers, e.g. consumers that do not have cars.

We then derive comparative statistics on household retail concentration for this second group of consumers. All of the derivations have the same intuition. First, household retail HHI increases as the number of retailers visited decreases, which in this case corresponds to an increase in $\underline{P^{F*}}$. Second, the results on the comparative statistics remain unchanged. A decrease in $\underline{P^{F*}}$ now corresponds to an increase in $\underline{P^{F*}}$, both of which increase household retail HHI.

How does the introduction of this second group of consumers account for a rise in cross-household variance? If $\underline{P^{F*}} < \overline{P^{F*}}$, this implies that there are a set of firms that both groups will consume from. Demand and supply-side changes that increase household retail HHI will decrease $\overline{P^{F*}}$ for one group and increase $\underline{P^{F*}}$ for the other group, decreasing the set of firms $[\underline{P^{F*}}, \overline{P^{F*}}]$ that both groups consume from. In other words, the first group will polarize towards low price index firms with larger variety, lower prices, and higher shopping costs, while the second group will polarize towards high price index firms with low variety, higher prices, and lower shopping costs, increasing cross-household variance. Allowing for a continuum of consumer groups gives the same intuition, in which households may polarize towards their preferred firms when shrinking the set of firms they consume from.[16]

Therefore, we find that demand-side changes such as a rise in t, ρ, and θ will all increase household retail HHI and increase cross-household variance by decreasing $\overline{P^{F*}}$ and increasing $\underline{P^{F*}}$. For supply-side changes, a rise in β will affect $\overline{P^{F*}}$ and $\underline{P^{F*}}$ as long as it affects the set of firms consumers visit. For example, if the second group of consumers does not visit club stores, then the entry of club stores do not change their household retail HHI, and any change in cross-household variance will only result from changes in consumption of the first group, i.e. an increase in $\overline{P^{F*}}$ and not a drop in $\underline{P^{F*}}$. Likewise, a fall in $\underline{P^{F}}$ will not affect $\underline{P^{F*}}$ as long as $\underline{P^{F*}} > \underline{P^{F}}$, so any change in cross-household variance will result from a fall in $\overline{P^{F*}}$ and not a rise in $\underline{P^{F*}}$. Hence, these changes may have a smaller effect on cross-household variance, consistent with our findings that the entry of supercenters and club stores do not have a statistically significant effect on cross-household variance. Likewise, whether an increase in α increases cross-household variance also depends on which group of

16. We can further allow households to consume from a set of firms $[\underline{P^{F*}}, \overline{P^{F*}}]$, i.e. their preferred set may not contain the lower or upper limits, by assuming that the second derivative of the cutoff equation is negative, which holds when the shopping cost function is sufficiently convex. In this case, we would not need to assume the net marginal gain from shopping at an additional firm is monotonically increasing or decreasing in P_f^F.

firms are disproportionately changing their price indices.[17]

As for changes in welfare, our previous derivations are general and allow for consumers that belong to both the first and second group. One exception is that a fall in $\underline{P^F}$ should again have no effect on households that do not consume from these firms.

2.6.5 Markups

What are the implications of our empirical results for market power? First, note that given profit maximization, we have the following equation for the firm-level markup $\mathcal{M}_f^F$ for firm f, which is defined as the price p_f minus marginal cost mc_f, as a function of the elasticity of the firm's residual demand ε_f^F, which is the same across all products.[18]

$$\mathcal{M}_f^F \equiv \frac{p_f}{mc_f} = \frac{\varepsilon_f^F}{\varepsilon_f^F - 1}$$

Next, we have the following identity which relates the elasticity of demand to the market share elasticity with respect to price:

$$\varepsilon_f^F = 1 - \frac{\partial S_f^F}{\partial P_f^F} \frac{P_f^F}{S_f^F}$$

We then show how the firm market share for the entire market S_f^F is related to the individual household market share s_{fi}^F. Recall that the market share of firm f for household

17. We can further explain the rise in one-stop shopping if consumers predominantly shift towards high-variety firms such as supercenters with a fall in $\underline{P^{F*}}$. One way to allow them to increase varieties per trip when polarizing towards low price index firms when increasing $\overline{P^{F*}}$ would require introducing costs per variety, which we abstract away from for tractability as mentioned.

18. This is true even without imposing product symmetry and is a property of nested demand systems, as mentioned in Hottman, Redding and Weinstein (2016).

i S^F_{fi} can be written as follows:

$$S^F_{fi} = S^F_f(D_i) = \begin{cases} \left(P\frac{\varphi^F_f}{P^F_f}\right)^{\sigma^F-1} & \text{if } \frac{YP^{\sigma^F-1}\left(\frac{P^F_f}{\varphi^F_f}\right)^{1-\sigma^F}}{w(\sigma^F-1)} \geq \mu^F_{fi} = \mu_i(P^F_f) \\ 0 & \text{otherwise} \end{cases}$$

$$P = P(D_i) = \left(\int_{f\in D_i}\left(\frac{P^F_f}{\varphi^F_f}\right)^{1-\sigma^F} dP^F_f\right)^{\frac{1}{1-\sigma^F}}$$

$$Y_i = Y(D_i) = (1-\rho)\left(w\left(T - \int_{f\in D_i}\mu^F_{fi}df\right) + K\right)$$

The market share, household price index, and income are all dependent on the specific set of firms D_i that household i consumes from, and D_i is a function of μ^F_{fi} by equation (2.14). We now allow for consumer heterogeneity more generally, allowing each household i to have a different shopping cost function μ^F_{fi} with probability density function $f(\mu^F_{fi})$, where the shopping cost function is decreasing in P^F_f as mentioned previously. Since equation (2.14) is monotonic in μ^F_{fi}, for each firm f there exists a cutoff household $i*$ such that equation (2.14) holds. Household $i*$ then represents the marginal consumer for firm f. We can then write the market share S^F_f as follows:

$$\forall f,\ \exists i^* \text{ s.t. } \frac{Y(D_{i*})S^F_{fi}(D_{i*})}{\sigma^F - 1} = w\mu^F_{fi*}$$

$$S^F_f = \frac{\int_i Y_i S^F_{fi} di}{\int_i Y_i di} = \frac{\int_0^{\mu^F_{fi*}} Y(D_i)S^F_f(D_i)f(\mu^F_{fi})d\mu^F_{fi}}{\int_i Y_i di}$$

For ease of exposition, assume that income Y_i is independent of P^F_f.[19] We can then

19. As shown above, relaxing this assumption would allow the Y_i to shift as weights in response to changes in prices, which complicates the analysis without affecting the main intuition.

rewrite the market share elasticity more simply as follows:

$$\frac{\partial S_f^F}{\partial P_f^F}\frac{P_f^F}{S_f^F} = \frac{\partial S_f^F(D_{i*})}{\partial P_f^F}\frac{P_f^F}{S_f^F(D_{i*})}\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F + \int_0^{\mu_{fi*}^F}\frac{\partial S_f^F(D_i)}{\partial P_f^F}\frac{P_f^F}{S_f^F(D_i)}\frac{S_f^F(D_i)}{S_f^F}\frac{Y(D_i)f(\mu_{fi}^F)}{\int_i Y_i di}d\mu_{fi}^F$$

Given the household market share S_{fi}^F, we can derive the following household market share elasticities under monopolistic competition and Bertrand competition:

$$\frac{\partial S_{fi}^F}{\partial P_f^F}\frac{P_f^F}{S_{fi}^F} = \begin{cases} 1-\sigma^F & \text{if monopolistic competition} \\ (1-S_{fi}^F)(1-\sigma^F) & \text{if Bertrand competition} \end{cases}$$

Under monopolistic competition, the market share elasticity is only dependent on the elasticity of substitution σ^F. Under Bertrand competition, a change in the price index of firm f affects the aggregate price index P. The market share elasticity depends on both σ^F and the market share S_{fi}^F. A larger market share lowers the absolute value of the market share elasticity, lowering the elasticity of demand and raising the markup. Substituting the household market share elasticities into the market share elasticity and hence the demand elasticity, we can derive the markup under monopolistic competition and Bertrand competition:

Monopolistic Competition

$$\mathcal{M}_f^F = \frac{\sigma^F + (\sigma^F - 1)\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F}{\sigma^F - 1 + (\sigma^F - 1)\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F} = \frac{\sigma^F + \frac{\left(S_f^F(D_{i*})\right)^2}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\frac{Y(D_{i*})}{w}}{\sigma^F - 1 + \frac{\left(S_f^F(D_{i*})\right)^2}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\frac{Y(D_{i*})}{w}} \tag{2.32}$$

Bertrand Competition

$$\mathcal{M}_f^F = \frac{\sigma^F + (\sigma^F - 1)\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F - (\sigma^F - 1)\int_0^{\mu_{fi*}^F}\frac{1}{S_f^F}\frac{Y(D_i)f(\mu_{fi}^F)}{\int_i Y_i di}\left(S_f^F(D_i)\right)^2 d\mu_{fi}^F}{\sigma^F - 1 + (\sigma^F - 1)\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F - (\sigma^F - 1)\int_0^{\mu_{fi*}^F}\frac{1}{S_f^F}\frac{Y(D_i)f(\mu_{fi}^F)}{\int_i Y_i di}\left(S_f^F(D_i)\right)^2 d\mu_{fi}^F} \tag{2.33}$$

Therefore, the markup differs from the standard CES markup $\frac{\sigma^F}{\sigma^F-1}$ in two ways. First, there is an additional extensive margin term that lowers the markup, since the existence of such marginal consumers increases the demand elasticity. This term is a function of several terms which include the share of firm sales from marginal consumers and the shopping cost at the margin. This term is similar to the additional extensive margin term in Neiman and Vavra (2021). In contrast to their paper which focuses on taste heterogeneity across products, we allow for consumer heterogeneity to work through the shopping cost μ.

Second, there is a term that increases markups due to consumer heterogeneity under Bertrand competition. This term has also been highlighted in Feenstra, Macedoni and Xu (2022), who again focus on taste heterogeneity across products. Intuitively, firms do not weight demand elasticities across consumers equally, but optimally use a greater weight on low-elasticity consumers, such that when consumer heterogeneity is present, they can charge higher markups and obtain higher profits, since the gains from charging higher markups to the lower elasticity consumers offsets the loss in demand on higher elasticity consumers. Crucially, this term increases as the variance of market shares across households increases, since firms can charge higher markups when there is a larger share of lower elasticity consumers.

The aggregate markup $\mathcal{M}$ is then the share-weighted average of all firm markups:

$$\mathcal{M} = \int_f S_f^F \mathcal{M}_f^F df$$

Therefore, how aggregate market power changes as a result of our empirical findings is ambiguous and requires a more detailed quantitative analysis, depending on the shopping cost distribution and market structure assumptions. As household retail concentration increases along with cross-household variance, there should be some firms with an increased variance of market shares across households, generating potentially offsetting effects on markups by raising the extensive margin term and lowering the consumer heterogeneity term, such that the effect on firms' markups, as well as the aggregate markup, is ambiguous.

2.7 Discussion

We briefly discuss some alternative models that can rationalize our results to highlight the implications for market power and welfare. Early work such as Messinger and Narasimhan (1997) estimated a stylized model to explain the growth of one-stop shopping using time-series variation from 1961-1986, and found that per capita disposable income had a significant positive effect on both supermarket assortment and store operating costs. This suggests that greater prevalence of one-stop shopping has been a response to growing demand for time-saving convenience as opposed to retail scale economies. However, this model assumes perfect competition and does not have direct implications for market power or welfare.

Basker, Klimek and Van (2012) introduce a stylized model in which a retailer's scale economies interact with consumer gains from one-stop shopping to generate a complementarity between a retailer's scale and scope. They suggest that welfare effects can be substantial, with shoppers experiencing welfare gains from increased variety at superstores, but this comes at the expense of mom-and-pop stores. Other welfare implications of these technological innovations and chains' resulting expansion and consolidation are more complex due to the presence of externalities. While consumers benefit from increases in the availability of one-stop shopping, consumers may create externalities by increasing the stores' sale volume, and as a result, fewer consumers shop at superstores than is socially optimal assuming that superstores benefit from economies of scale and scope, which induces chains to add more stores and varieties, benefiting other consumers. Anderson and Waldfogel (2015) summarize a literature on preference externalities in the context of media markets, which also applies to industries with high fixed costs and heterogeneity of consumer preferences such as the retail sector. Choices and welfare for consumers depend on the number and mix of consumers according to their preferences. In our case, the rise of one-stop shoppers influences the mix of stores in a particular region, potentially benefiting similar shoppers at the expense of multi-stop shoppers.

Bronnenberg (2015) proposes a simple general equilibrium model of demand for variety with purchasing costs, in which firms can invest in providing convenience through one-stop shopping. He argues that in his model, high prices in a marketing equilibrium may not be automatically a "bad" because they support more entry and assortment. Hence, whether welfare is improved or reduced from the introduction of investments in convenience is not clear a priori. In his model, investments in convenience tend to be underprovided relative to the social optimum.

These models imply that various models could be used to rationalize our results, but the implications for market power and welfare are not entirely clear. Instead, we discuss some of the existing literature on the welfare implications of the supply-side and demand-side evolutions we documented.

On the supply side, Atkin, Faber and Gonzalez-Navarro (2017) and Leung and Li (2021*a*) both find that big-box stores, in particular supercenters, offer more variety and lower prices, which increases consumer welfare. As a result, the entry of more big-box retailers, combined with growth in variety within existing stores, are likely to be beneficial to consumers. This is consistent with our model, which finds that supply-side changes that increase household retail concentration tend to increase welfare.

On the demand-side, Thomassen et al. (2017) use a multi-category multi-seller demand model estimated using UK consumer data to show that consumers inclined to one-stop shopping, as opposed to multi-stop shopping, have a greater pro-competitive impact because although they have lower demand elasticities, they also generate relatively large complementary cross-category effects. Since stores internalize these cross-price elasticities, market power can be reduced substantially as a result. Hence, whether market power is reduced and consumer welfare is increased hinges on the relative sizes of the changes in the cross-category effect against the fall in demand elasticities from increased one-stop shopping. In our model, we show that changes in opportunity cost of time could come from sources that either increase or decrease welfare. How all these changes impact market power and welfare would require a more detailed quantitative analysis.

2.8 Conclusion

In this paper, we document that national retail concentration has risen substantially from 2004-2019. Regional concentration has barely increased, implying that regions are becoming more similar in their market share distributions. This rise is also driven by a rise in household concentration, even as households becoming more heterogeneous in their choice of retailers. The rise in household concentration is consistent with increased one-stop shopping, rather than decrease availability of retailers to choose from.

We explain these facts using evolutions in supply and demand in the retail sector. On the supply side, we find that the number of big-box stores, namely supercenters and club stores, have increased. Utilizing an event study approach to estimate the impact of entry of these superstar retailers, we find that supercenters raise household HHI and one-stop shopping, while club stores decrease household HHI and one-stop shopping. This can be attributed to the fact that households buy far more varieties and pay lower prices in supercenters than any other channel type, including club stores. We then utilize an IV strategy to estimate the effect of changes in product variety and prices within existing stores. We find that increases in variety and lower prices increases household HHI. These supply-side changes by retailers can explain up to about 20% of the rise in household HHI.

On the demand side, we find that a rise in the opportunity cost of time decreases the number of shopping trips, which leads to higher household HHI and increased one-stop shopping, using an IV strategy. These demand-side changes by households can explain up to 30% of the rise in household HHI.

We develop a model to rationalize these results, and highlight the resulting implications for market power and welfare. While these supply-side changes tend to be associated with increased welfare, the demand-side changes have more ambiguous effects, and depend crucially on the source of these changes.

A rise in one-stop shopping could imply demand is more inelastic and higher prices, but there is a pro-competitive effect of one-stop shopping due to multiproduct retailers internalizing cross-price elasticities from shopping complementarities, which has an opposing effect on prices and consumer welfare. We leave a more quantitative analysis of how these changes impact market power and welfare to future work.

2.9 Tables

Table 2.1: Effects of Entry: Number of Stores Within Own 5-digit Zip Code

	(1)	(2)	(3)	(4)
VARIABLES	Household Retail Concentration			
NumSup	0.00812***	0.00799***		
	(0.00254)	(0.00237)		
NumClubs			-0.0121***	-0.0129***
			(0.00281)	(0.00277)
Observations	134,495	1,837,668	350,865	2,468,557
R-squared	0.681	0.698	0.682	0.686
Prob > F	0.001	0.001	0.000	0.000
Household-Quarter FE	X	X	X	X
Year-Quarter FE	X	X	X	X
Number of units	9371	125817	21749	152934
Number of clusters	9371	125817	21749	152934
BOTE	0.1688	0.019	-0.0766	-0.014

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Columns (1) and (3) include households living in zip5 with store entry during the period. Columns (2) and (4) include all households. NumSup refers to the number of supercenters and NumClubs refers to the number of club stores. BOTE refers to a back-of-the-envelope estimate of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table 2.2: Effects of Entry on Other Measures of Concentration

VARIABLES	(1) Num Retailers	(2) Within dept HHI	(3) Cross dept variance	(4) Cross HH variance	(5) Num Retailers	(6) Within dept HHI	(7) Cross dept variance	(8) Cross HH variance
NumSup	**-0.229*** **	**0.00281**	**-0.00542*** **	-0.000173				
	(0.0447)	(0.00244)	(0.00109)	(0.00263)				
NumClub3					0.246***	-0.0154***	-0.00134	-0.00264
					(0.0562)	(0.00302)	(0.00129)	(0.00325)
Observations	121,723	134,482	134,482	134,482	98,588	99,805	99,805	99,805
R-squared	0.695	0.667	0.506	0.627	0.703	0.667	0.510	0.632
Prob > F	0.000	0.249	0.000	0.947	0.000	0.000	0.299	0.416
Household-Quarter FE	X	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X	X
Number of units	9062	9371	9371	9371	6560	6323	6323	6323
Number of clusters	9062	9371	9371	9371	6560	6323	6323	6323
BOTE	0.167	0.112	0.297	-0.004	-0.180	-0.752	0.074	-0.050

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Includes only households living in zip5 with store entry during the period. BOTE refers to a back-of-the-envelope estimate of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table 2.3: Effects of Entry on Number of Trips and Varieties per Trip

VARIABLES	(1) TripNum	(2) UPC per trip	(3) ProductGroup per trip	(4) Dept per trip	(5) TripNum	(6) UPC per trip	(7) ProductGroup per trip	(8) Dept per trip
NumSup	-0.203	0.0893	0.0659*	0.0317***				
	(0.211)	(0.0602)	(0.0349)	(0.00943)				
NumClubs					0.960***	-0.281***	-0.159***	-0.0364***
					(0.283)	(0.0664)	(0.0387)	(0.0112)
Observations	134,495	134,495	134,495	134,495	109,651	109,651	109,651	109,651
R-squared	0.755	0.778	0.790	0.766	0.766	0.789	0.799	0.771
Prob > F	0.336	0.138	0.059	0.001	0.001	0.000	0.000	0.001
Household-Quarter FE	X	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X	X
Number of units	9371	9371	9371	9371	6814	6814	6814	6814
Number of clusters	9371	9371	9371	9371	6814	6814	6814	6814
BOTE	0.040	0.131	0.143	0.163	-0.126	-0.314	-0.279	-0.150

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Includes only households living in zip5 with store entry during the period. BOTE refers to a back-of-the-envelope estimate of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table 2.4: Effect of Variety and Prices on Household Retail Concentration

	(1)	(2)	(3)
Data	HMS		RMS
VARIABLES	Household HHI		
Variety Depth	-0.0847***	-0.0903***	0.124***
	(0.0187)	(0.0193)	(0.0247)
Variety Breadth	0.216***	0.218***	-0.148*
	(0.0461)	(0.0448)	(0.0848)
RPI (County)	-0.143***		
	(0.0293)		
RPI (US)		-0.225***	
		(0.0305)	
Price Index			0.00419
			(0.0345)
Observations	3491193	3605864	974443
R-squared	0.662	0.660	0.696
Prob > F	0.000	0.000	0.000
Number of units	185449	190795	82135
Number of clusters	185449	190795	82135
First stage F-stat	137.195	92.590	273.257
BOTE: All	0.034	0.030	0.204
BOTE: Variety Depth	-0.068	-0.107	0.253
BOTE: Variety Breadth	0.114	0.152	-0.050
BOTE: Prices	-0.012	-0.015	0.000

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety depth refers to the number of UPCs per store, variety breadth refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 2.5.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table 2.5: Effect of shopping trips on household HHI, IV with household characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IV	None	None: Controls	Age	Size		Income		Employment	
VARIABLES	Household HHI								
Log Trips	-0.129***	-0.201***	-0.204***	-0.203***	-0.198***	-0.0862**	-0.108***	-0.218***	-0.222***
	(0.000364)	(0.000613)	(0.00291)	(0.0141)	(0.0145)	(0.0340)	(0.0351)	(0.0134)	(0.0137)
Observations	3668938	3622992	2312242	3622992	3622926	3589733	3589130	3622992	3622926
R-squared	0.712	0.724	0.384	0.724	0.739	0.708	0.730	0.724	0.739
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	194524	193527	2650	193527	193526	192164	192140	193527	193526
Number of clusters	194524	193527	136238	193527	193526	192164	192140	193527	193526
First stage F-stat			6225.057	102.427	98.491	162.099	144.213	116.170	109.195
BOTE: Trips	0.613	0.957	0.941	0.963	0.939	0.409	0.514	1.036	1.055
County-time FE			X		X		X		X

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.6: Effect of shopping trips on household HHI, IV with region average wage and unemployment rate

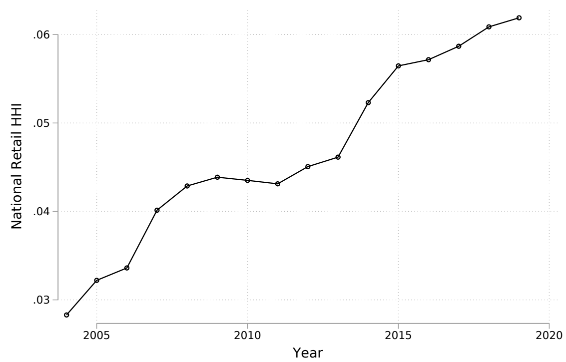
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IV group	Age		Education		Age-Education		Age-Education-Occupation		Age & Education	
IV region	All	State	All	State	All	State	All	State	All	State
VARIABLES	Household HHI									
Log Trips	-0.298***	-0.421***	-0.282***	-0.193***	-0.130**	-0.190*	-0.146*	0.0416	-0.291***	-0.244***
	(0.0361)	(0.0977)	(0.0563)	(0.0669)	(0.0608)	(0.0990)	(0.0789)	(0.260)	(0.0306)	(0.0579)
Observations	2285235	2285228	2305700	2303820	2278148	2210623	1007906	710297	2278148	2276304
R-squared	0.728	0.682	0.730	0.738	0.732	0.739	0.733	0.667	0.730	0.737
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	128280	128279	128764	128753	128038	127556	62106	54773	128038	128027
Number of clusters	128280	128279	128764	128753	128038	127556	62106	54773	128038	128027
First stage F-stat	42.073	6.581	15.757	10.598	13.308	5.524	8.423	1.076	30.737	7.462
BOTE: Trips	1.655	2.007	1.276	0.925	0.615	0.901	0.684	-0.198	1.718	1.187
BOTE: IVs	0.194	0.004	0.127	0.018	-0.011	-0.003	-0.012	0.000	0.295	0.024

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

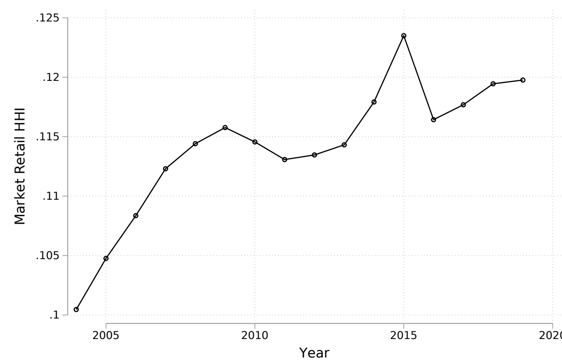
2.10 Figures

Figure 2.1: Retail Concentration Over Time

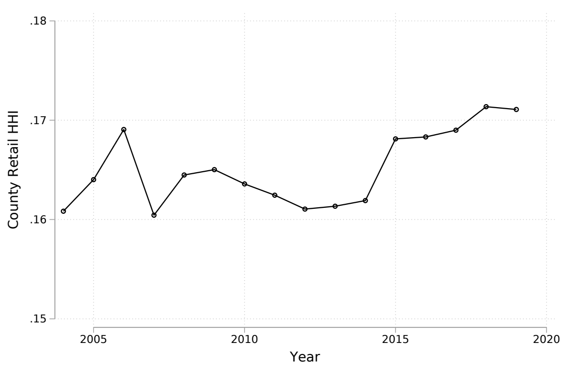
(a) National



(b) Market



(c) County



(d) Household

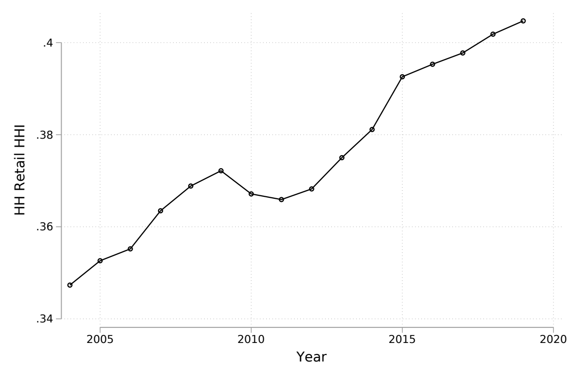
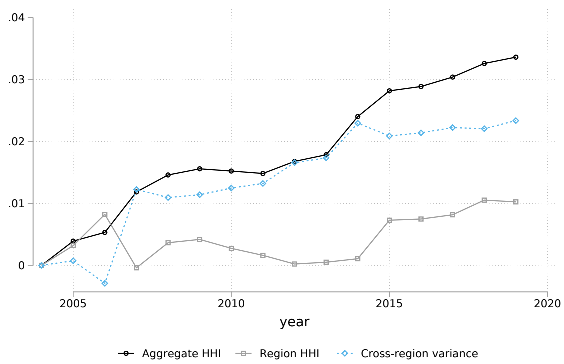
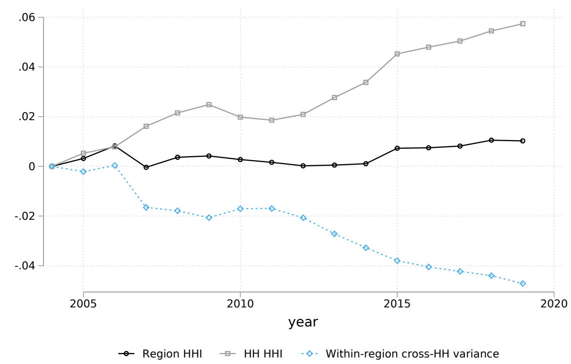


Figure 2.2: Decomposing Changes in Retail Concentration

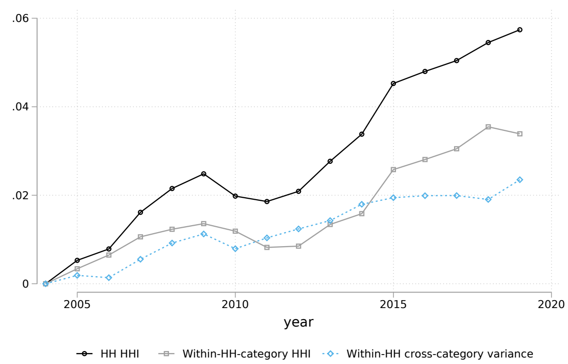
(a) National and County



(b) County and Household



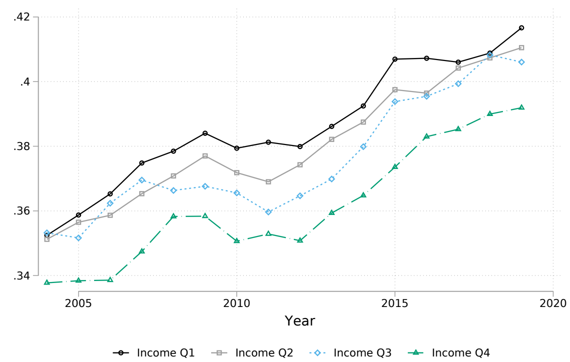
(c) Household and Category



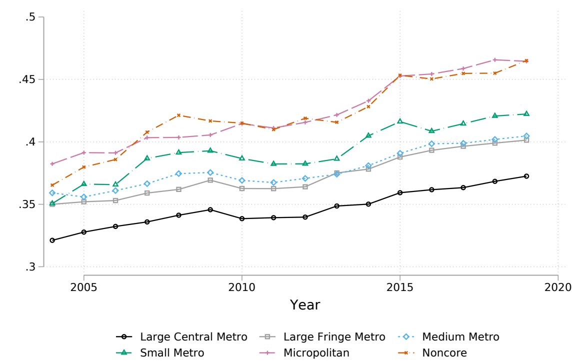
Notes: This figure shows that HHI can be decomposed at each level following Radaelli and Zenga (2002). For each HHI term, we plot the yearly change over the sample period while we plot the negative of the cross-variance since it has a negative contribution to HHI.

Figure 2.3: Household Retail Concentration by Demographic Group

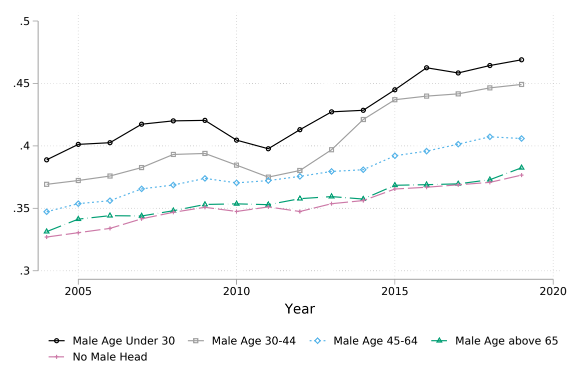
(a) Household Income



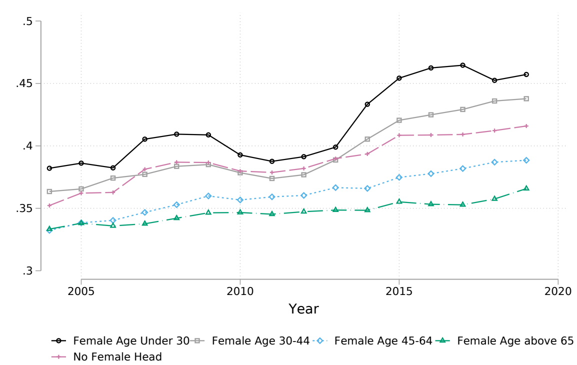
(b) Urban/Rural Counties



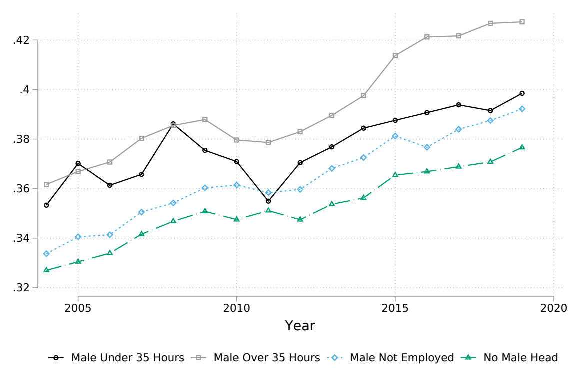
(c) Male Head Age



(d) Female Head Age



(e) Male Head Employment



(f) Female Head Employment

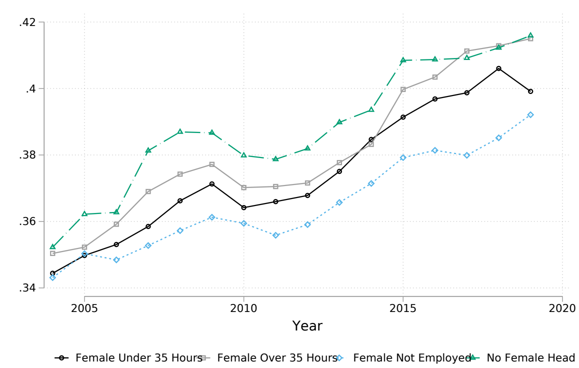
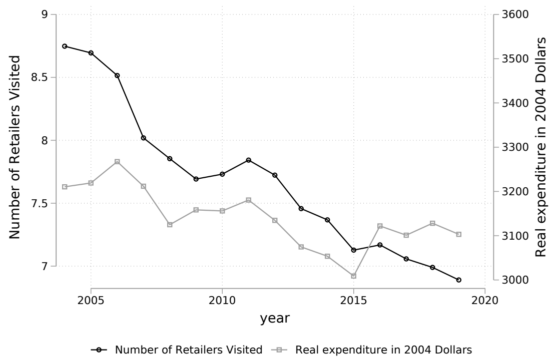
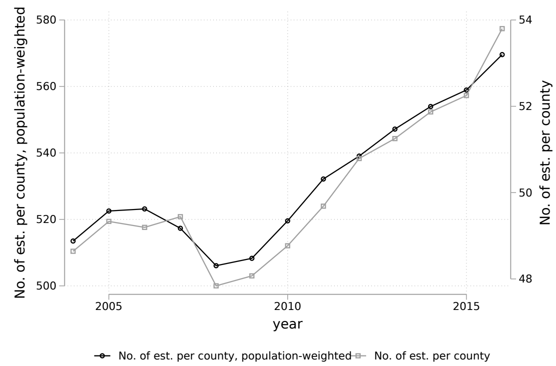


Figure 2.4: Retail Chains Visited, Number of Drug, Grocery, and Mass Merchandise Establishments per County, and Shopping Trips and Days

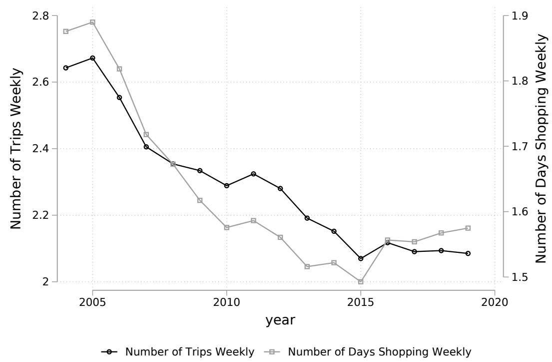
(a) Retail Chains Visited



(b) Establishments per County



(c) Shopping Trips and Days



Notes: Figure 2.4a plots the number of retail chains visited per quarter per household, aggregated first to the yearly level with a simple average and then averaged across households weighted by sampling weights. We deflate the household expenditures using the chained food-at-home price index from the BLS. We obtain the total number of drug, grocery, and mass merchandise establishments per county from the County Business Patterns data from the Economic Census. We plot these numbers only up to 2016 due to changes in the reporting thresholds since 2017.

Figure 2.5: Expenditure per Trip and Varieties per Trip

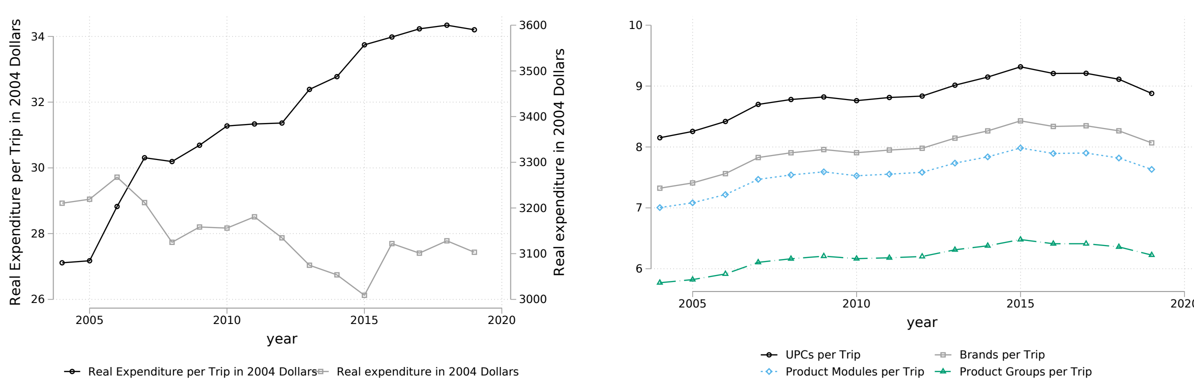
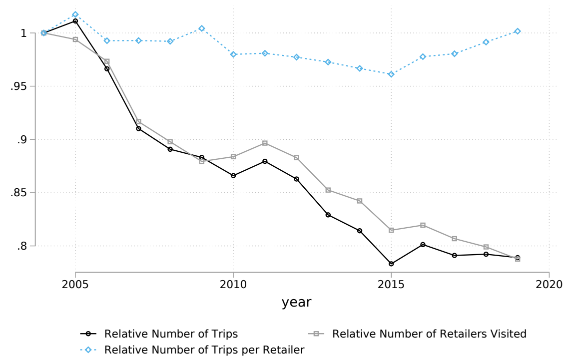
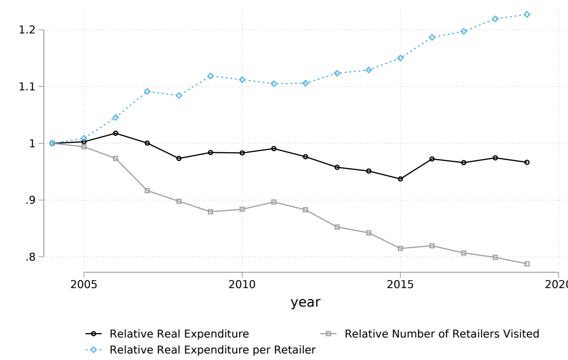


Figure 2.6: Decomposing trips, expenditures, and number of retailers per household

(a) Number of trips, number of retailers, and trips per retailer



(b) Real expenditure, number of retailers, and expenditures per retailer



(c) Real expenditure, number of trips, and expenditures per trip

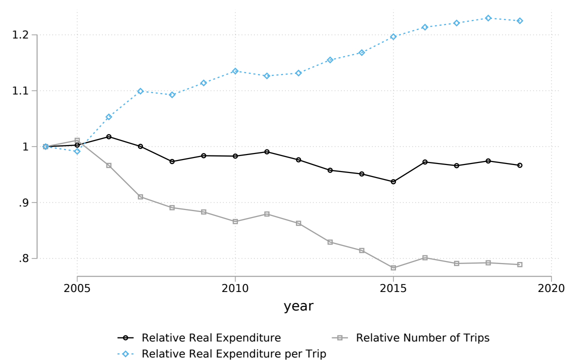
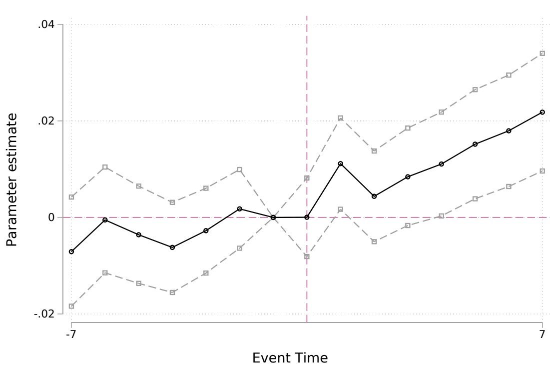


Figure 2.7: Event Study Graph for Store Entry

(a) Supercenter Entry



(b) Club Entry

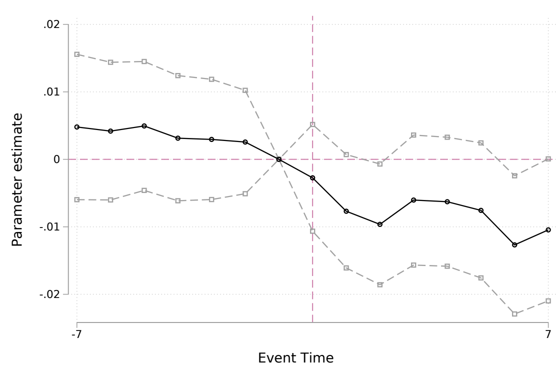
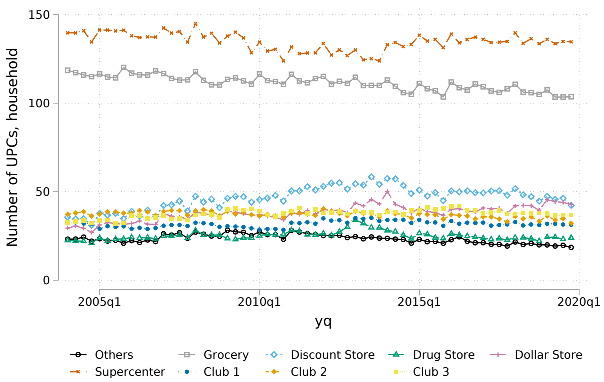


Figure 2.8: Product Assortment in Big-box Stores

(a) UPCs



(b) Product Modules

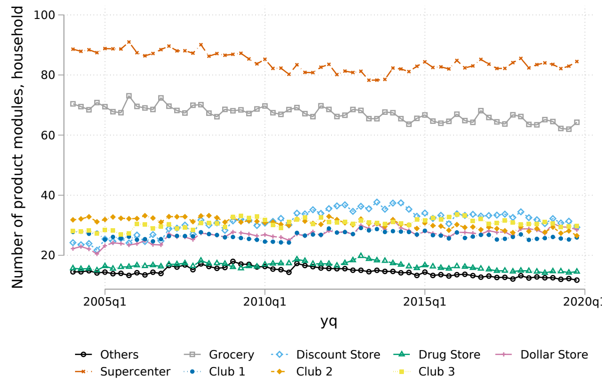
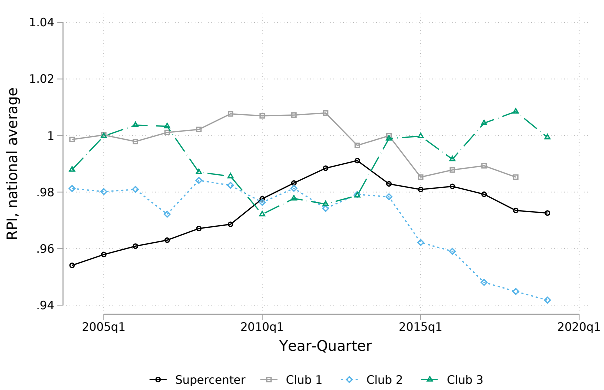


Figure 2.9: RPI in Big-box Stores and Other Channel Types

(a) Big-box Stores



(b) Other Channel Types

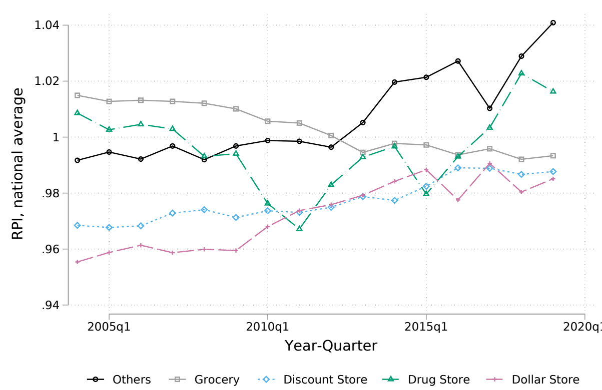
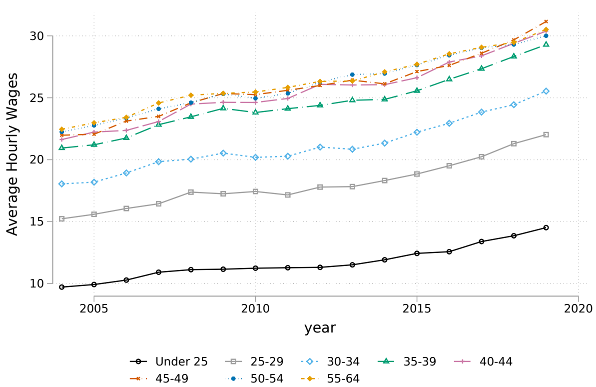
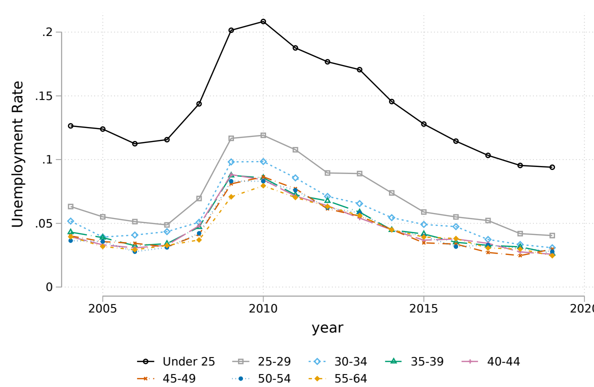


Figure 2.10: Time Costs by Age Group

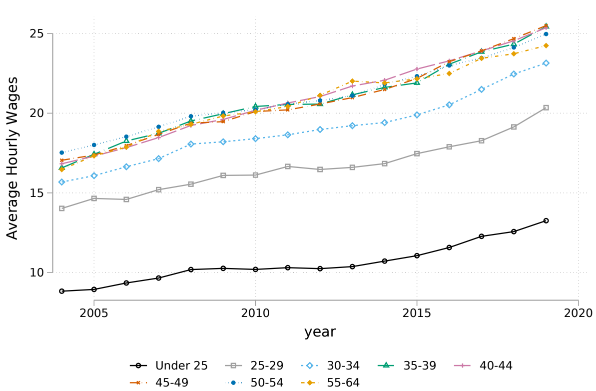
(a) Average Wages, Men



(b) Unemployment Rate, Men



(c) Average Wages, Women



(d) Unemployment Rate, Women

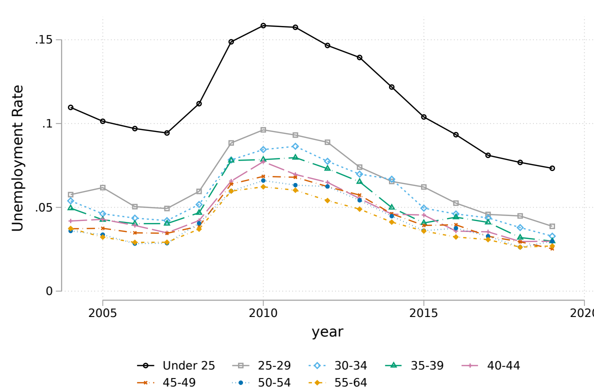
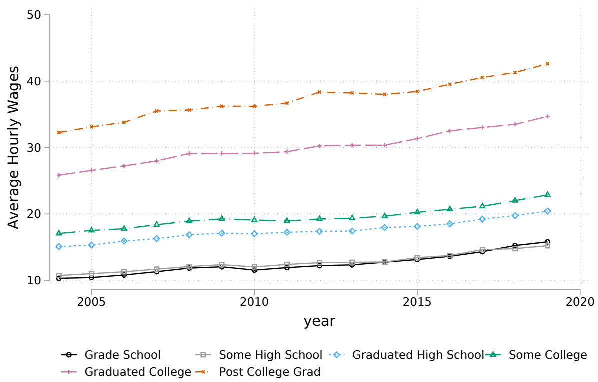
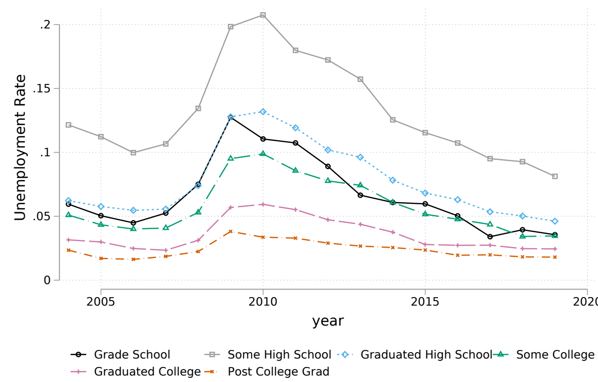


Figure 2.11: Time Costs by Education

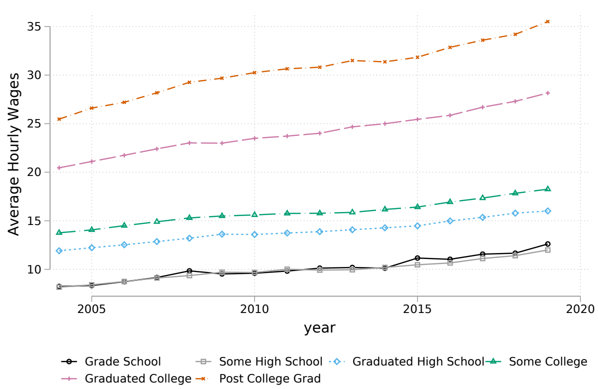
(a) Average Wages, Men



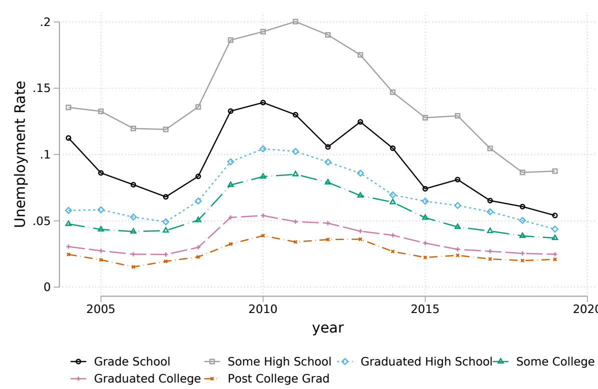
(b) Unemployment Rate, Men



(c) Average Wages, Women



(d) Unemployment Rate, Women



2.11 Appendix

2.11.1 Measuring Retail Concentration

We use the Herfindahl-Hirschman Index (HHI) as our primary measure of retail concentration at the household level and different geographical levels.

Household retail concentration is constructed in three steps. First, for each household i and retailer j, we calculate the total expenditure S_{ij}^t in period t. Second, the associated market share is

$$m_{ij}^t = \frac{S_{ij}^t}{\sum_j S_{ij}^t}.$$

Finally, household retail concentration is the HHI of the above market share:

$$H_i^t = \sum_j (m_{ij}^t)^2.$$

We further calculate the regional average household retail concentration, defined as the weighed average of household retail concentration in a given region. Let $r(i,t)$ denote the region where household i lives in period t. The set of households who lives in region r in period t is $I_{rt} = \{i : r(i,t) = r\}$. Regional average household retail concentration can then be calculated from

$$\bar{H}_r^t = \sum_{i \in I_{rt}} \alpha_{ir}^t H_i^t.$$

The household weight α_{ir}^t depends on all the households' sampling weight (projection factor) w_{ir}^t and total expenditure S_{ij}^t:

$$\alpha_{ir}^t = \frac{\sum_j (w_{ir}^t S_{ij}^t)}{\sum_{i:r(i,t)=r} \sum_j (w_{ir}^t S_{ij}^t)}.$$

We use a similar measure for retail concentration at different geographical units, such as Scantrack markets as defined by Nielsen, and nationally. Following the above definition, the market share for retailer j at time t in region r is

$$m_{jr}^t = \frac{\sum_{i \in I_{rt}} (w_{ir}^t S_{ij}^t)}{\sum_j \sum_{i \in I_{rt}} (w_{ir}^t S_{ij}^t)}.$$

The HHI of region r in period t becomes

$$H_r^t = \sum_j (m_{jr}^t)^2.$$

We call it the aggregate retail concentration of region r in period t, which is a measure of regional retail concentration.

An alternative measure of retail concentration is the total share of expenditure on the top n retail chains, which is generally known as concentration ratios C_n. Since results are qualitatively identical using this measure, we focus on the HHI in our main analysis.

2.11.1.1 RZ Decomposition

Moreover, we follow Radaelli and Zenga (2002) (RZ) to define the cross-household variance of retail concentration:

$$V_r^t = \sum_{i:r(i,t)=r} \alpha_{ir}^t \left(\sum_j (m_{ij}^t - m_{jr}^t)^2 \right).$$

We can formally relate local average household concentration and local aggregate concentration with the decomposition:

$$H_r^t = \bar{H}_r^t - V_r^t.$$

This decomposition suggests that if all the households' consumption patterns in each retail chain are the same, then $\bar{V}_r^t = 0$ and the local average household retail concentration is a perfect reflection of the local aggregate concentration in the region.

Therefore, beginning at the national level, we can decompose changes in national aggregate HHI $\Delta H = H^t - H^{t-n}$ into changes in the national average local HHI minus the cross-county variance:

$$\Delta H = \Delta \sum_r \alpha_r H_r - \Delta \sum_r \alpha_r V_r.$$

At the local level, we can decompose changes in the local aggregate HHI into changes in the local average household HHI minus the cross-household variance:

$$\Delta H_r = \Delta \sum_r \sum_i \alpha_{ir} H_{ir} - \Delta \sum_r \sum_i \alpha_{ir} V_{ir}.$$

At the household level, we can further classify expenditures at the household-product-category level S_{ic}. We then decompose changes in the household aggregate HHI into changes in the household average category HHI minus the cross-category variance:

$$\Delta H_{ir} = \Delta \sum_r \sum_i \sum_c \alpha_{irc} H_{irc} - \Delta \sum_r \sum_i \sum_c \alpha_{irc} V_{irc}.$$

Combining all four levels, changes in national aggregate HHI can be formally decomposed into changes in the following four terms: Household-category HHI H_{irc}, household cross-category variance V_{irc}, regional cross-household variance V_{ir}, and national cross-region variance V_r. We write out the entire decomposition as follows:

$$\begin{aligned}\Delta H = &\Delta \sum_r \sum_i \sum_c \alpha_{irc} H_{irc} - \Delta \sum_r \sum_i \sum_c \alpha_{irc} V_{irc} \\ &- \Delta \sum_r \sum_i \alpha_{ir} V_{ir} - \Delta \sum_r \alpha_r V_r,\end{aligned}$$

where

$$H_{irc} = \sum_j (m_{irjc})^2,\ V_{irc} = \sum_j (m_{irjc} - m_{irj})^2,\ \alpha_{irc} = \frac{w_{ir} S_{irc}}{\sum_i \sum_r \sum_c w_{ir} S_{irc}}.$$

To investigate whether composition changes in our dataset, such as the entry and exit of regions or households, are driving our results, we also use a Dynamic Olley-Pakes decomposition (Melitz and Polanec 2015) on top of the RZ decomposition, as shown below:

$$\begin{aligned}
&\Delta \sum_r \sum_i \sum_c \alpha_{irc} H_{irc} \\
&= (H_{S2} - H_{S1}) + \alpha_{E2} (H_{E2} - H_{S2}) + \alpha_{X1} (H_{S1} - H_{X1}) \\
&= \Delta \bar{H}_S + \Delta cov_S + \alpha_{E2} (H_{E2} - H_{S2}) + \alpha_{X1} (H_{S1} - H_{X1}) \\
&\alpha_{Gt} = \sum_{i \in G} \alpha_{it},\ H_{Gt} = \sum_{i \in G} \frac{\alpha_{it}}{\alpha_{Gt}} H_{it}
\end{aligned}$$

Applying this decomposition to each of the four terms decomposing the national aggregate HHI in Appendix Table 2.7, we find that within-survivor growth is driving most of the variation.

2.11.2 Online Shopping

Can the rise of online shopping contribute to the trends we observe? We find that concentration trends barely change when considering only offline retailers in Appendix Figure 2.15, given that the share of online shopping remains small even as it is increasing. This is consistent with Hortaçsu and Syverson (2015), who show evidence that online share of retail sales remains very small in the product categories we consider. They argue that although online retail will surely continue to be a force shaping the sector going forward and may yet emerge as the dominant mode of commerce in the retail sector in the US, its time for supremacy has not yet arrived.

In Appendix Figure 2.16, we further show the share of online shopping in the product departments in our data. While the shares are increasing in nearly every product department, they all remain below 5%. Appendix Figure 2.17 shows the household concentration trends are very similar with and without online retailers.

We investigate the impact of online shopping shares on household HHI in Appendix Table 2.23. We first regress the household HHI with offline retailers, household HHI, the number of offline shopping trips, and the number of offline retailers visited on the online shopping share, with household and time fixed effects. Given that online shopping share is likely to be endogenous, we caution against interpreting these estimates as causal and use them as a rough estimate of the potential impacts of online shopping. In fact, if some unobserved factors lead to both higher household concentration and higher share of online shopping (e.g. higher trip costs), the effect of online shopping will be overestimated. We can then treat the estimates as upper bounds. We find that since the rise in online shares remain small, the BOTEs are at most 1-2%. Given that endogeneity is unlikely to substantially downward bias our estimates, we view these results as suggestive evidence that online shopping is not a main driver of our trends.

To assuage potential endogeneity concerns, we use state-level variation in Amazon taxes to estimate the impact of online shopping on household HHI and the number of trips in Appendix Table 2.24. We find that the introduction of an Amazon tax has negligible effects on household HHI and the number of trips.

2.11.3 Derivations

2.11.3.1 Demand

We assume that the shopping cost at each firm is a function of the price index P_f^F in each firm, with $\mu'(P_f^F) < 0$. This is consistent with the fact that firms with larger variety and lower prices, i.e. lower P_f^F, tend to be located farther away from consumers relative to firms with lower variety and higher prices, due to higher costs of land in areas with higher population density and the need for larger square footage stores in order to stock higher variety and lower prices through economies of scale. We can show that if $\mu'(P_f^F)$ is small enough in absolute value, there exists a unique cutoff firm P^{F*} that satisfies the following condition using equation (2.14), such that consumers only buy from firms that have a price index within the set $[\underline{P^F}, P^{F*}]$:

$$\left(P^{F*}\right)^{1-\sigma^F} \frac{1}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} \frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{P^{F*}} \mu(p)dp\right) + K\right)}{w(\sigma^F - 1)} - \mu(P^{F*}) = 0 \tag{2.34}$$

<u>Uniqueness + existence of P^{F*}</u>

$$f(P^F) = \underbrace{\left(P^F\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^F}}^{P^F} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{P^F} \mu(p)dp\right) + K\right)}{w(\sigma^F - 1)}}_{C} - \mu(P^F)$$

$$f(\underline{P^F}) = \left(\underline{P^F}\right)^{1-\sigma^F} \frac{1}{\left(\underline{P^F}\right)^{1-\sigma^F}} \frac{(1-\rho)\left(w\left(T - \mu(\underline{P^F})\right) + K\right)}{w(\sigma^F - 1)} - \mu(\underline{P^F}) \overset{\text{Assume}}{>} 0$$

$$f(\overline{P^F}) = \left(\overline{P^F}\right)^{1-\sigma^F} \frac{1}{\int_{\underline{P^F}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} \frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{\overline{P^F}} \mu(p)dp\right) + K\right)}{w(\sigma^F - 1)} - \mu(\overline{P^F}) \overset{\text{Assume}}{<} 0$$

$$\begin{aligned}
f'(P^F) = {} & (B^{-1}C)(1-\sigma^F)\left(P^F\right)^{-\sigma^F} && < 0 \\
& + (AC)\left(-B^{-2}\right)\left(P^F\right)^{1-\sigma^F} && < 0 \\
& + (AB^{-1})\frac{1-\rho}{\sigma^F - 1}\left(-\mu(P^F)\right) && < 0 \\
& - \mu'(P^F) && > 0 \\
& \underset{\text{Assume}}{< 0} && \text{if } \mu'(P^F) < 0 \text{ and is not too large in absolute value}
\end{aligned}$$

2.11.3.2 Household Retail Concentration

Given our derivations, the market share for household i for each firm can be written as follows:

$$S_{fi}^F = \frac{P_f^F x^F(\nu)}{Y} = \begin{cases} \left(P\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F-1} & \text{if } \frac{YP^{\sigma^F-1}\left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}}{w(\sigma^F-1)} \geq \mu(P_f^F) \\ 0 & \text{otherwise} \end{cases}$$

Given our simplifying assumptions, we can further simplify the market share expression and write the household retail HHI for household i as follows:

$$H_i = \int_f \left(S_{fi}^F\right)^2 df = \frac{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{2\left(1-\sigma^F\right)} dP_f^F}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \tag{2.35}$$

We can then show some comparative statics of household retail HHI in response to changes in various parameters. First, we show that any parameters, in this case denoted by t, that decrease the cutoff firm price index P^{F*} will increase household retail HHI. This is intuitive since the set of retailers that households consume from, $[\underline{P^F}, P^{F*}]$, will decrease as P^{F*} decreases.

$$\frac{dP^{F*}}{dt} < 0 \Rightarrow \frac{dH_i}{dt} > 0 \tag{2.36}$$

$$
\begin{aligned}
\frac{dH_i}{dt} &= \int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F} \right)^2 dP_f^F \\
&\cdot \left(-2 \left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \right)^{-3} \left(\left(P^{F*}\right)^{1-\sigma^F} \frac{dP^{F*}}{dt} + \int_{\underline{P^F}}^{P^{F*}} \frac{\partial \left(P_f^F\right)^{1-\sigma^F}}{\partial t} dP_f^F \right) \right) \\
&+ \frac{1}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \right)^2} \left(\left(\left(P^{F*}\right)^{1-\sigma^F} \right)^2 \frac{dP^{F*}}{dt} + \int_{\underline{P^F}}^{P^{F*}} \frac{\partial \left(\left(P_f^F\right)^{1-\sigma^F} \right)^2}{\partial t} dP_f^F \right) \\
&= \frac{\left(P^{F*}\right)^{1-\sigma^F}}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \right)^2} \left(\frac{2 \int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F} \right)^2 dP_f^F}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} - \left(P^{F*}\right)^{1-\sigma^F} \right) \left(-\frac{dP^{F*}}{dt} \right)
\end{aligned}
$$

$$
\frac{dP^{F*}}{dt} < 0 \Rightarrow \frac{dH_i}{dt} > 0 \text{ if } \frac{2 \int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F} \right)^2 dP_f^F}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} > \left(P^{F*}\right)^{1-\sigma^F}
$$

$$
\text{Jensen's inequality} \Rightarrow \int \left(\left(P_f^F\right)^{1-\sigma^F} \right)^2 dP_f^F \geqslant \left(\int \left(P_f^F\right)^{1-\sigma^F} dP_f^F \right)^2
$$

$$
\begin{aligned}
&2 \int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F} \right)^2 dP_f^F \geqslant 2 \left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \right)^2 \\
&> \left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \right)^2 \geq \left(P^{F*}\right)^{1-\sigma^F} \int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \\
&\because \int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \geq \left(P^{F*}\right)^{1-\sigma^F}
\end{aligned}
$$

Given this fact, we can use equation (2.34) to derive comparative statics of the cutoff firm price index P^{F*} in response to changes in various parameters to see how these parameters change household retail HHI. These comparative statics can then be compared against our empirical findings.

First, consider changes on the demand side. Let the shopping cost be $\mu(P^F) = t\delta(P^F)$, where t is the time cost per distance traveled and $\delta(P^F)$ is the distance of the consumer from each firm indexed by P^F.

Extensive margin (P^{F*}, Comparative Statics)

$$f(P^{F*}) = \underbrace{\left(P^{F*}\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T-\int_{\underline{P^F}}^{P^{F*}} t\delta(p)dp\right)+K\right)}{w(\sigma^F-1)}}_{C} - t\delta(P^{F*}) = 0$$

$$\begin{aligned}
\frac{dP^{F*}}{dt} &= -\frac{\frac{\partial f(P^{F*},t)}{\partial t}}{\frac{\partial f(P^{F*},t)}{\partial P^F}} \\
&= \frac{\delta(P^{F*}) + \left(AB^{-1}\right)\frac{1-\rho}{\sigma^F-1}\left(\int_{\underline{P^F}}^{P^{F*}} \delta(p)dp\right)}{\frac{\partial f(P^{F*},t)}{\partial P^F}} \\
&- \frac{\partial f(P^{F*},t)}{\partial t} > 0, \frac{\partial f(P^{F*},t)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{dt} < 0
\end{aligned}$$

$$\begin{aligned}
\frac{dP^{F*}}{dw} &= -\frac{\frac{\partial f(P^{F*},w)}{\partial w}}{\frac{\partial f(P^{F*},w)}{\partial P^F}} \\
&= \frac{\left(AB^{-1}\right)\frac{1-\rho}{\sigma^F-1}Kw^{-2}}{\frac{\partial f(P^{F*},w)}{\partial P^F}} \\
&- \frac{\partial f(P^{F*},w)}{\partial w} > 0, \frac{\partial f(P^{F*},\rho)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{dw} < 0
\end{aligned}$$

$$\begin{aligned}
\frac{dP^{F*}}{d\rho} &= -\frac{\frac{\partial f(P^{F*},\rho)}{\partial \rho}}{\frac{\partial f(P^{F*},\rho)}{\partial P^F}} \\
&= \frac{\left(AB^{-1}\right)\frac{1}{w(\sigma^F-1)}\left(w\left(T-\int_{\underline{P^F}}^{P^{F*}} t\delta(p)dp\right)+K\right)}{\frac{\partial f(P^{F*},\rho)}{\partial P^F}} \\
&- \frac{\partial f(P^{F*},\rho)}{\partial \rho} > 0, \frac{\partial f(P^{F*},\rho)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{d\rho} < 0
\end{aligned}$$

$$f(P^{F*}) = \underbrace{\left(P^{F*}\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{P^{F*}} t\delta(p)dp\right) + K\right)}{w(\sigma^F - 1)}}_{C} \underbrace{\frac{1}{1-\rho+\theta\rho}}_{D^{-1}} - t\delta(P^{F*})$$

$$= 0$$

$$\begin{aligned} \frac{dP^{F*}}{d\theta} &= -\frac{\frac{\partial f(P^{F*},\theta)}{\partial \theta}}{\frac{\partial f(P^{F*},\theta)}{\partial P^F}} \\ &= \frac{\left(AB^{-1}C\right)\left(D^{-2}\right)\rho}{\frac{\partial f(P^{F*},t)}{\partial P^F}} \\ &- \frac{\partial f(P^{F*},\theta)}{\partial \theta} > 0, \frac{\partial f(P^{F*},\theta)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{d\theta} < 0 \end{aligned}$$

Let $\mu = \mu(N, \beta)$ and $\frac{\partial \mu}{\partial \beta} < 0$

$$\begin{aligned} \frac{dP^{F*}}{d\beta} &= -\frac{\frac{\partial f(P^{F*},\beta)}{\partial \beta}}{\frac{\partial f(P^{F*},\beta)}{\partial P^F}} \\ &= \frac{\frac{\partial \mu}{\partial \beta} + AB^{-1}\frac{1-\rho}{\sigma^F-1}\int_{\underline{P^F}}^{P^{F*}} \frac{\partial \mu}{\partial \beta} dn}{\frac{\partial f(P^{F*},\beta)}{\partial P^F}} \\ &- \frac{\partial f(P^{F*},\beta)}{\partial \beta} < 0, \frac{\partial f(P^{F*},\beta)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{d\beta} > 0 \end{aligned}$$

$$
\begin{aligned}
\frac{dP^{F*}}{d\underline{P^F}} &= -\frac{\frac{\partial f(P^{F*},\underline{P^F})}{\partial \underline{P^F}}}{\frac{\partial f(P^{F*},\underline{P^F})}{\partial P^F}} \\
&= \frac{-AB^{-1}\frac{1-\rho}{\sigma^F-1}\mu(\underline{P^F}) - AC^{-1}(B^{-2})\underline{P^F}^{1-\sigma^F}}{\frac{\partial f(P^{F*},\beta)}{\partial P^F}} \\
&- \frac{\partial f(P^{F*},\underline{P^F})}{\partial \underline{P^F}} < 0, \frac{\partial f(P^{F*},\underline{P^F})}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{d\underline{P^F}} > 0
\end{aligned}
$$

$$
H_i = \int_f \left(S_{fi}^F\right)^2 df = \frac{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{2\left(1-\sigma^F\right)} dP_f^F}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2}
$$

$$
\begin{aligned}
\frac{dH_i}{d\underline{P^F}} &= \int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F \\
&\cdot \left(-2\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^{-3} \left(\left(P^{F*}\right)^{1-\sigma^F} \frac{dP^{F*}}{d\underline{P^F}} - \left(\underline{P^F}\right)^{1-\sigma^F} + \int_{\underline{P^F}}^{P^{F*}} \frac{\partial\left(P_f^F\right)^{1-\sigma^F}}{\partial \underline{P^F}} dP_f^F\right)\right) \\
&+ \frac{1}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \left(\left(\left(P^{F*}\right)^{1-\sigma^F}\right)^2 \frac{dP^{F*}}{d\underline{P^F}} - \left(\left(\underline{P^F}\right)^{1-\sigma^F}\right)^2 + \int_{\underline{P^F}}^{P^{F*}} \frac{\partial\left(\left(P_f^F\right)^{1-\sigma^F}\right)^2}{\partial \underline{P^F}} dP_f^F\right) \\
&= -\frac{\left(P^{F*}\right)^{1-\sigma^F}}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \underbrace{\left(\frac{2\int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} - \left(P^{F*}\right)^{1-\sigma^F}\right)}_{A_1} \frac{dP^{F*}}{d\underline{P^F}} \\
&+ \frac{\left(\underline{P^F}\right)^{1-\sigma^F}}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \underbrace{\left(\frac{2\int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F}{\int_{\underline{P^F}}^{P^{F*}} \left(\underline{P^F}\right)^{1-\sigma^F} dP_f^F} - \left(\underline{P^F}\right)^{1-\sigma^F}\right)}_{A_2}
\end{aligned}
$$

$$
A_1 \geq A_2 > 0,\ \frac{dP^{F*}}{d\underline{P^F}} > \left(\frac{P^{F*}}{\underline{P^F}}\right)^{\sigma^F-1} \frac{A_2}{A_1} \Rightarrow \frac{dH_i}{d\underline{P^F}} < 0
$$

$$\text{Let } P^F = P^F(\alpha) \text{ and } f(P^F,\alpha) = P^F(\alpha)^{1-\sigma^F},\ \frac{dP^F}{d\alpha} < 0 \Rightarrow \frac{df}{d\alpha} > 0$$

$$H_i = \frac{2}{\left(\int f dp\right)^2}\left(\int f\frac{df}{d\alpha} - \int f^2\left(\int f\right)^{-1}\int\frac{df}{d\alpha}\right) > 0 \text{ if}$$

$$\int f\int f\frac{df}{d\alpha} - \int f^2\int\frac{df}{d\alpha} > 0$$

$$\int f\frac{df}{d\alpha} > \frac{\int f^2\int\frac{df}{d\alpha}}{\int f} \geq \frac{(\int f)^2\int\frac{df}{d\alpha}}{\int f} = \int f\int\frac{df}{d\alpha}$$

where we use Jensen's inequality for the $\geq$ condition

2.11.3.3 Welfare

$$U(X(\mathcal{V}), L(\mathcal{V})) = X(\mathcal{V})^{1-\rho}L(\mathcal{V})^\rho$$

$$X(\mathcal{V}) = \left(\int_{\nu\epsilon\mathcal{V}}(\varphi(\nu)x(\nu))^{\frac{\sigma-1}{\sigma}}dv\right)^{\frac{\sigma}{\sigma-1}}$$

$$\tau(D) = t\int_0^D d(\nu)d\nu$$

$$\mu(\nu) = td(\nu)$$

$$U = \left(\frac{Y}{P}\right)^{1-\rho}L\,(D)^\rho$$

$$\frac{Y}{P} = (1-\rho)\left(w\left(T - \overbrace{\int_0^D\mu(\nu)d\nu}^{\tau(D)\equiv}\right) + K\right)\overbrace{\left(\int_0^D\left(\frac{p(\nu)}{\varphi(\nu)}\right)^{1-\sigma^F}d\nu\right)^{\frac{1}{\sigma^F-1}}}^{P^{-1}}$$

$$L(D) = \frac{\rho\left(w\left(T - \int_0^D\mu(\nu)d\nu\right) + K\right)}{w}$$

Applying the envelope theorem, we have

$$\begin{aligned}
\left.\frac{dU}{dt}\right|_{D=D^*} &= (1-\rho)X(D)^{-\rho}\left[-(1-\rho)w\int_0^D d(\nu)d\nu P^{-1}\right]L(D)^{\rho} \\
&+ X(D)^{1-\rho}\rho L(D)^{\rho-1}\left[-\rho\int_0^D d(\nu)d\nu\right] < 0 \\
&= -X(D)^{-\rho}L(D)^{\rho-1}\int_0^D d(\nu)d\nu\left[(1-\rho)^2 wP^{-1}\underbrace{L(D)}_{\rho(T-\tau(D))} + \rho^2\underbrace{X(D)}_{(1-\rho)w(T-\tau(D))P^{-1}}\right] \\
&= -X(D)^{-\rho}L(D)^{\rho-1}\rho\int_0^D d(\nu)d\nu < 0
\end{aligned}$$

$$\begin{aligned}
U &= P^{\rho-1}(1-\rho)^{1-\rho}\rho^{\rho}\left[w^{1-\rho}(T-\tau(D)) + Kw^{-\rho}\right] \\
\left.\frac{dU}{dw}\right|_{D=D^*} &= P^{\rho-1}(1-\rho)^{1-\rho}\rho^{\rho}\left[(1-\rho)w^{-\rho}(T-\tau(D)) - K\rho w^{-\rho-1}\right] \\
&= P^{\rho-1}(1-\rho)^{1-\rho}\rho^{\rho}w^{-\rho-1}\left[(1-\rho)w(T-\tau(D)+K-K) - \rho K\right] \\
&= P^{\rho-1}(1-\rho)^{1-\rho}\rho^{\rho}w^{-\rho-1}\left(Y-K\right) > 0 \\
&= P^{\rho-1}(1-\rho)^{1-\rho}\rho^{\rho}w^{-\rho}\left(T-\tau(D)-L(D)\right) > 0
\end{aligned}$$

$$\begin{aligned}
lnU &= (1-\rho)ln\left(\frac{Y}{P}\right) + \rho lnL(D) \\
\left.\frac{dlnU}{d\rho}\right|_{D=D^*} &= -ln\left(\frac{Y}{P}\right) + (1-\rho)\frac{1}{\frac{Y}{P}}\left(-w\left(T-\tau(D)\right)^{\rho-1}\right) + lnL(D) + \rho\frac{1}{L(D)}(T-\tau(D)) \\
&= -ln(\frac{Y}{P}) - 1 + lnL(D) + 1 = -ln\left(\frac{Y}{P}\right) + lnL(D) > 0 \text{ if } L(D) > \frac{Y}{P}
\end{aligned}$$

$$\text{If } U\left(X(\mathcal{V}), L(\mathcal{V})\right) = X(\mathcal{V})^{1-\rho} L(\mathcal{V})^{\theta\rho}$$

$$lnU = (1-\rho) ln\left(\frac{Y}{P}\right) + \theta\rho ln L(D)$$

$$\left.\frac{dlnU}{d\theta}\right|_{D=D^*} = \rho ln L(D) > 0$$

$$\text{Let } \mu = \mu(P^F, \beta) \text{ and } \frac{d\mu}{d\beta} < 0$$

$$\left.\frac{dU}{d\beta}\right|_{D=D^*} = -(1-\rho)X(D)^{-\rho}(1-\rho)wP^{-1}L(D)^{\rho}\frac{d\mu}{d\beta} - X(D)^{1-\rho}\rho L(D)^{\rho-1}\rho\frac{d\mu}{d\beta} > 0$$

Using the FOC for variety and the conditions required for the existence of a unique cutoff price index, we have

$$\left.\frac{dU}{d\underline{P^F}}\right|_{D=D^*} = -\rho^{\rho}(1-\rho)^{1-\rho}\left(\frac{w}{P}\right)^{1-\rho}\left[-\mu(\underline{P^F}) + \frac{A(D)\left(p(\underline{P^F})\right)^{1-\sigma}}{w(\sigma-1)}\right] < 0$$

$$\text{Let } P^F = P^F(\alpha) \text{ and } f(P^F,\alpha) = P^F(\alpha)^{1-\sigma^F},\ \frac{dP^F}{d\alpha} < 0 \Rightarrow \frac{df}{d\alpha} > 0$$

$$\left.\frac{dU}{d\alpha}\right|_{D=D^*} = \left(\overbrace{(1-\rho)\left(w(T-\tau(D))+K\right)}^{Y(D)}\right)^{1-\rho}\left(\frac{\overbrace{\rho\left(w\left(T-\tau(D)\right)+K\right)}^{L(D)}}{w}\right)^{\rho}\frac{dP^{\rho-1}}{d\alpha} = ...$$

$$\frac{dP^{\rho-1}}{d\alpha} = (\rho-1)P^{\rho-2}\frac{dP}{d\alpha} = \underbrace{(\rho-1)}_{<0}P^{\rho-2}\underbrace{\frac{1}{1-\sigma^F}}_{<0}\left(\int_{\underline{P^F}}^{P^{F*}}\left(P_f^F\right)^{1-\sigma^F}\right)^{\frac{\sigma^F}{1-\sigma^F}}\underbrace{\frac{d\int_{\underline{P^F}}^{P^{F*}} f(P^F,\alpha)dP^F}{d\alpha}}_{>0} > 0$$

$$P = \left(\int_0^D \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^{\frac{1}{1-\sigma^F}}$$

2.11.3.4 Heterogeneity

If there is a group of consumers with $f'(P^F) > 0$, $f(\underline{P^F}) < 0$ and $f(\overline{P^F}) > 0$, i.e. $\mu'(P^F)$ large enough in absolute value $\Rightarrow \exists \underline{P^{F*}} s.t.\ f \in \left[\underline{P^{F*}}, \overline{P^F}\right]$

One group of consumers with $\overline{P^{F*}}\downarrow$ if $\frac{d\overline{P^{F*}}}{dt} < 0 \Rightarrow \frac{dH_i}{dt} > 0$, another group of consumers with $\underline{P^{F*}}^*\uparrow$ if $\frac{d\underline{P^{F*}}}{dt} > 0 \Rightarrow \frac{dH_i}{dt} > 0$, then cross-variance $\uparrow$ and each group polarizes into their respective set of firms.

Uniqueness + existence of $\underline{\underline{P^{F*}}}$

$$f(P^F) = \underbrace{\left(P^F\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^F}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T-\int_{P^F}^{\overline{P^F}}\mu(p)dp\right)+K\right)}{w(\sigma^F-1)}}_{C} - \mu(P^F)$$

$$f(\underline{P^F}) = \left(\underline{P^F}\right)^{1-\sigma^F} \frac{1}{\int_{\underline{P^F}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} \frac{(1-\rho)\left(w\left(T-\int_{\underline{P^F}}^{\overline{P^F}}\mu(p)dp\right)+K\right)}{w(\sigma^F-1)} - \mu(\underline{P^F}) \overset{\text{Assume}}{<} 0$$

$$f(\overline{P^F}) = \left(\overline{P^F}\right)^{1-\sigma^F} \frac{1}{\left(\overline{P^F}\right)^{1-\sigma^F}} \frac{(1-\rho)\left(w\left(T-\mu(\overline{P^F})\right)+K\right)}{w(\sigma^F-1)} - \mu(\overline{P^F}) \overset{\text{Assume}}{>} 0$$

$$\begin{aligned}
f'(P^F) = {} & (B^{-1}C)(1-\sigma^F)\left(P^F\right)^{-\sigma^F} && < 0 \\
& + (AC)\left(B^{-2}\right)\left(P^F\right)^{1-\sigma^F} && > 0 \\
& + (AB^{-1})\frac{1-\rho}{\sigma^F-1}\left(\mu(P^F)\right) && > 0 \\
& - \mu'(P^F) && > 0 \\
& \underset{\text{Assume}}{> 0} && \text{if } \mu'(P^F) < 0 \text{ and is large enough in absolute value}
\end{aligned}$$

$$H_i = \frac{\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{2\left(1-\sigma^F\right)} dP_f^F}{\left(\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \qquad f \in \left[\underline{P^{F*}}, \overline{P^F}\right]$$

2.11.3.5 Markups

First, note that given profit maximization, we have the following equation for the firm-level markup $\mathcal{M}_f^F$ for firm f, which is defined as the price p_f minus marginal cost mc_f, as a function of the elasticity of the firm's residual demand ε^F, which is the same across all products.[20]

$$\mathcal{M}_f^F \equiv \frac{p_f}{mc_f} = \frac{\varepsilon_f^F}{\varepsilon_f^F - 1}$$

Next, we have the following identity which relates the elasticity of demand to the market share elasticity with respect to price:

$$\begin{aligned} S_f^F &= \frac{P_f^F x_f}{\int_f P_f^F x_f df} \\ \frac{\partial S_f^F}{\partial P_f^F} &= \frac{x_f + P_f^F \frac{\partial x_f}{\partial P_f^F}}{\int_f P_f^F x_f df} \\ \varepsilon_f^F &= 1 - \frac{\partial S_f^F}{\partial P_f^F} \frac{P_f^F}{S_f^F} \end{aligned}$$

We then show how the firm market share for the entire market S_f^F is related to the individual household market share s_{fi}^F. Recall that the market share of firm f for household i S_{fi}^F can be written as follows:

$$\begin{aligned} S_{fi}^F = S_f^F(D_i) &= \begin{cases} \left(P \frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F - 1} & \text{if } \frac{Y P^{\sigma^F - 1} \left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}}{w(\sigma^F - 1)} \geq \mu_{fi}^F = \mu_i(P_f^F) \\ 0 & \text{otherwise} \end{cases} \\ P = P(D_i) &= \left(\int_{f \in D_i} \left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F} dP_f^F \right)^{\frac{1}{1-\sigma^F}} \\ Y_i = Y(D_i) &= (1 - \rho) \left(w \left(T - \int_{f \in D_i} \mu_{fi}^F df \right) + K \right) \end{aligned}$$

20. This is true even without imposing product symmetry and is a property of nested demand systems, as mentioned in Hottman, Redding and Weinstein (2016).

The market share, household price index, and income are all dependent on the specific set of firms D_i that household i consumes from, and D_i is a function of μ^F_{fi} by equation (2.14). We now allow for consumer heterogeneity more generally, allowing each household i to have a different shopping cost function μ^F_{fi} with probability density function $f(\mu^F_{fi})$, where the shopping cost function is decreasing in P^F_f as mentioned previously. Since equation (2.14) is monotonic in μ^F_{fi}, for each firm f there exists a cutoff household $i*$ such that equation (2.14) holds. Household $i*$ then represents the marginal consumer for firm f. We can then write the market share S^F_f as follows:

$$\forall f,\, \exists i^* \text{ s.t. } \frac{Y(D_{i*})S^F_{fi}(D_{i*})}{\sigma^F - 1} = w\mu^F_{fi*}$$

$$S^F_f = \frac{\int_i Y_i S^F_{fi} di}{\int_i Y_i di} = \frac{\int_0^{\mu^F_{fi*}} Y(D_i) S^F_f(D_i) f(\mu^F_{fi}) d\mu^F_{fi}}{\int_i Y_i di}$$

The firm market share elasticity can then be derived as follows:

$$\begin{aligned}
\frac{\partial S^F_f}{\partial P^F_f}\frac{P^F_f}{S^F_f} &= \frac{1}{\int_i Y_i di}\Bigg[Y(D_{i*})S^F_f(D_{i*})f(\mu^F_{fi})\frac{1}{w(\sigma^F-1)}\frac{\partial Y(D_{i*})S^F_f(D_{i*})f(\mu^F_{fi})}{\partial P^F_f} \\
&\quad + \int_0^{\mu^F_{fi*}} \frac{\partial Y(D_i)S^F_f(D_i)f(\mu^F_{fi})}{\partial P^F_f} d\mu^F_{fi}\Bigg]\frac{P^F_f}{S^F_f} \\
&\quad - \int_i Y_i S^F_{fi} di \left[\left(\frac{1}{\int_i Y_i di}\right)^2 \int_0^\infty \frac{\partial Y(D_i) f(\mu^F_{fi})}{\partial P^F_f} d\mu^F_{fi}\right]\frac{P^F_f}{S^F_f}
\end{aligned}$$

$$\frac{\partial Y(D_i)S^F_f(D_i)}{\partial P^F_f} = Y(D_i)\frac{\partial S^F_{fi}(D_i)}{\partial P^F_f} + S^F_f(D_i)\frac{\partial Y(D_i)}{\partial P^F_f}$$

For ease of exposition, assume that income Y_i is independent of P^F_f.[21] We can then

21. As shown above, relaxing this assumption would allow the Y_i to shift as weights in response to changes in prices, which complicates the analysis without affecting the main intuition.

rewrite the market share elasticity more simply as follows:

$$\frac{\partial S_f^F}{\partial P_f^F}\frac{P_f^F}{S_f^F} = \frac{\partial S_f^F(D_{i*})}{\partial P_f^F}\frac{P_f^F}{S_f^F(D_{i*})}\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F$$
$$+\int_0^{\mu_{fi*}^F}\frac{\partial S_f^F(D_i)}{\partial P_f^F}\frac{P_f^F}{S_f^F(D_i)}\frac{S_f^F(D_i)}{S_f^F}\frac{Y(D_i)f(\mu_{fi}^F)}{\int_i Y_i di}d\mu_{fi}^F$$

Given the household market share S_{fi}^F, we can derive the following household market share elasticities under monopolistic competition and Bertrand competition:

$$S_{fi}^F = \left(P\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F-1}$$

$$\frac{\partial S_{fi}^F}{\partial P_f^F} = (P\varphi_f^F)^{\sigma^F-1}\frac{\partial\left(\frac{1}{P_f^F}\right)^{\sigma^F-1}}{\partial P_f^F} + \left(\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F-1}\frac{\partial P^{\sigma^F-1}}{\partial P_f^F}$$
$$= (P\varphi_f^F)^{\sigma^F-1}(1-\sigma^F)(P_f^F)^{-\sigma^F} + \left(\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F-1}\frac{\partial P^{\sigma^F-1}}{\partial P_f^F}$$

$$P = \left(\int_f\left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}\right)^{\frac{1}{1-\sigma^F}} \quad \text{if monopolistic competition}$$

$$P = \left(\sum_f\left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}\right)^{\frac{1}{1-\sigma^F}} \quad \text{if Bertrand competition}$$

$$\frac{\partial P^{\sigma^F-1}}{\partial P_f^F} = \begin{cases} 0 & \text{if monopolistic competition} \\ (\sigma^F-1)P^{\sigma^F-2}\varphi^{\sigma^F-1}\left(\frac{P}{P_f^F}\right)^{\sigma^F} & \text{if Bertrand competition}\end{cases}$$

$$\frac{\partial S_{fi}^F}{\partial P_f^F}\frac{P_f^F}{S_{fi}^F} = \begin{cases} 1-\sigma^F & \text{if monopolistic competition} \\ (1-S_{fi}^F)(1-\sigma^F) & \text{if Bertrand competition}\end{cases}$$

Under monopolistic competition, the market share elasticity is only dependent on the elasticity of substitution σ^F. Under Bertrand competition, a change in the price index of firm f affects the aggregate price index P. The market share elasticity depends on both σ^F and the market share S^F_{fi}. A larger market share lowers the absolute value of the market share elasticity, lowering the elasticity of demand and raising the markup. Substituting the household market share elasticities into the market share elasticity and hence the demand elasticity, we can derive the markup under monopolistic competition and Bertrand competition:

Monopolistic Competition

$$\mathcal{M}^F_f = \frac{\sigma^F + (\sigma^F - 1)\frac{S^F_f(D_{i*})}{S^F_f}\frac{Y(D_{i*})f(\mu^F_{fi*})}{\int_i Y_i di}\mu^F_{fi*}}{\sigma^F - 1 + (\sigma^F - 1)\frac{S^F_f(D_{i*})}{S^F_f}\frac{Y(D_{i*})f(\mu^F_{fi*})}{\int_i Y_i di}\mu^F_{fi*}} = \frac{\sigma^F + \frac{\left(S^F_f(D_{i*})\right)^2}{S^F_f}\frac{Y(D_{i*})f(\mu^F_{fi*})}{\int_i Y_i di}\frac{Y(D_{i*})}{w}}{\sigma^F - 1 + \frac{\left(S^F_f(D_{i*})\right)^2}{S^F_f}\frac{Y(D_{i*})f(\mu^F_{fi*})}{\int_i Y_i di}\frac{Y(D_{i*})}{w}} \tag{2.37}$$

Bertrand Competition

$$\mathcal{M}^F_f = \frac{\sigma^F + (\sigma^F - 1)\frac{S^F_f(D_{i*})}{S^F_f}\frac{Y(D_{i*})f(\mu^F_{fi*})}{\int_i Y_i di}\mu^F_{fi*} - (\sigma^F - 1)\int_0^{\mu^F_{fi*}} \frac{1}{S^F_f}\frac{Y(D_i)f(\mu^F_{fi})}{\int_i Y_i di}\left(S^F_f(D_i)\right)^2 d\mu^F_{fi}}{\sigma^F - 1 + (\sigma^F - 1)\frac{S^F_f(D_{i*})}{S^F_f}\frac{Y(D_{i*})f(\mu^F_{fi*})}{\int_i Y_i di}\mu^F_{fi*} - (\sigma^F - 1)\int_0^{\mu^F_{fi*}} \frac{1}{S^F_f}\frac{Y(D_i)f(\mu^F_{fi})}{\int_i Y_i di}\left(S^F_f(D_i)\right)^2 d\mu^F_{fi}} \tag{2.38}$$

Therefore, the markup differs from the standard CES markup $\frac{\sigma^F}{\sigma^F - 1}$ in two ways. First, there is an additional extensive margin term that lowers the markup, since the existence of such marginal consumers increases the demand elasticity. This term is a function of several terms which include the share of firm sales from marginal consumers and the shopping cost at the margin. This term is similar to the additional extensive margin term in Neiman and Vavra (2021). In contrast to their paper which focuses on taste heterogeneity across products, we allow for consumer heterogeneity to work through the shopping cost μ.

Second, there is a term that increases markups due to consumer heterogeneity under Bertrand competition. This term has also been highlighted in Feenstra, Macedoni and Xu

(2022), who again focus on taste heterogeneity across products. Intuitively, firms do not weight demand elasticities across consumers equally, but optimally use a greater weight on low-elasticity consumers, such that when consumer heterogeneity is present, they can charge higher markups and obtain higher profits, since the gains from charging higher markups to the lower elasticity consumers offsets the loss in demand on higher elasticity consumers. Crucially, this term increases as the variance of market shares across households increases, since firms can charge higher markups when there is a larger share of lower elasticity consumers.

The aggregate markup $\mathcal{M}$ is then the share-weighted average of all firm markups:

$$\mathcal{M} = \int_f S_f^F \mathcal{M}_f^F df$$

2.11.4 Tables

Table 2.7: RZ-DOPD: Change in aggregate HHI, 2004-2015

	Category HHI	Category Variance	Household Variance	Region Variance
Survivor	0.05446	0.00363	-0.05547	0.02463
Within Survivor	0.02969	0.04744	-0.04799	0.03182
Between Survivor	0.02477	-0.04381	-0.00748	-0.00719
Entrants	0.00185	0.01687	-0.00378	-0.00218
Exiters	-0.02241	0.00302	0.01210	0.00089

Table 2.8: Effect of Entry: Minimum Distance and Within Different Distance Thresholds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Household Retail Concentration							
	Supercenters				Clubs			
min_dis	-0.000399**				0.000703***			
	(0.000173)				(0.000158)			
num_5mi		0.00245**				-0.00520***		
		(0.00115)				(0.00158)		
num_10mi			0.000719				-0.00229**	
			(0.000581)				(0.000955)	
num_15mi				0.000515				-0.00141**
				(0.000395)				(0.000704)
Observations	1,616,203	1,616,203	1,616,203	1,616,203	467,181	467,181	467,181	467,181
R-squared	0.696	0.696	0.696	0.696	0.681	0.681	0.681	0.681
Prob > F	0.021	0.033	0.216	0.192	0.000	0.001	0.017	0.046
Household-Quarter FE	X	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X	X
Number of units	109224	109224	109224	109224	29187	29187	29187	29187
Number of clusters	109224	109224	109224	109224	29187	29187	29187	29187
BOTE	0.019	0.019	0.018	0.025	-0.086	-0.064	-0.057	-0.056

Notes: Robust standard errors are in parentheses, clustered by household. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Includes only HHs living in zip5 areas that have store entry within 20 miles. min_dis refers to the distance to the nearest supercenter or club, while num_5mi, num_10mi, and num_15mi refer to the number of stores within a 5-mile, 10-mile, and 15-mile radius from the zip-code centroid of each household. BOTE refers to a back-of-the-envelope estimate of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table 2.9: Difference between the most affected households and the least affected households for supercenter entry

	Mean Diff	sd	p-value
hh_hhi_yq	0.053	0.023	0.546
tripnum_yq	-1.931	2.204	1.000
numsup	0.038	0.040	1.000
numclubs	-0.082	0.050	0.887
ind.hhincomeQ1	**0.362**	0.177	0.680
ind.hhincomeQ2	**0.344**	0.169	0.682
ind.hhincomeQ3	-0.213	0.160	0.972
ind.hhincomeQ4	-0.494	0.162	0.152
ind.hhsize1	**0.283**	0.162	0.843
ind.hhsize2	-0.673	0.169	0.014**
ind.hhsize3plus	**0.390**	0.187	0.661
ind.ageAbove55	**-0.686**	0.210	0.101
ind.CentralMetro	-0.751	0.324	0.528
ind.LFringeMetro	0.126	0.270	1.000
ind.MediumMetro	-0.362	0.252	0.949
ind.SmallMetro	**0.719**	0.209	0.058*
ind.Micropolitan	0.105	0.115	1.000
ind.Noncore	0.164	0.082	0.710

The most affected group has top 10% sorted effect. The least affected group has bottom 10% sorted effect. The difference is average value for the most affected minus the least affected.

Table 2.10: Difference between the most affected households and the least affected households for clubs entry

	Mean Diff	sd	p-value
hh_hhi_yq	0.052	0.030	0.952
tripnum_yq	-7.433	2.746	0.495
num_sup	-0.176	0.116	0.983
numclubsoc	0.071	0.047	0.983
ind.hhincomeQ1	0.044	0.141	1.000
ind.hhincomeQ2	0.134	0.141	1.000
ind.hhincomeQ3	-0.552	0.175	0.319
ind.hhincomeQ4	0.374	0.228	0.962
ind.hhsize1	0.191	0.189	1.000
ind.hhsize2	-0.704	0.195	0.157
ind.hhsize3plus	0.514	0.222	0.732
ind.ageAbove55	-0.927	0.222	0.057*
ind.CentralMetro	-0.782	0.233	0.235
ind.LFringeMetro	0.849	0.306	0.480
ind.MediumMetro	-0.429	0.250	0.956
ind.SmallMetro	0.321	0.232	0.996
ind.Micropolitan	0.040	0.152	1.000
ind.Noncore	0.000	0.000	

The most affected group has top 10% sorted effect. The least affected group has bottom 10% sorted effect. The difference is average value for the most affected minus the least affected.

Table 2.11: Effect of Variety and Prices on Other Measures of Concentration, HMS

VARIABLES	(1) Num Retailers	(2) Within-dept HHI	(3) Cross-dept variance	(4) Cross-HH variance
Variety Depth	0.986***	-0.0437**	0.0466***	0.0147
	(0.335)	(0.0177)	(0.00758)	(0.0219)
Variety Breadth	-2.343***	0.117***	-0.101***	-0.0315
	(0.783)	(0.0411)	(0.0176)	(0.0507)
RPI (US)	3.114***	-0.194***	0.0316**	0.0215
	(0.547)	(0.0285)	(0.0126)	(0.0325)
Observations	3605864	3605864	3605864	3605864
R-squared	0.683	0.638	0.491	0.579
Prob > F	0.000	0.000	0.000	1.000
Number of units	190795	190795	190795	190795
Number of clusters	190795	190795	190795	190795
First stage F-stat	92.590	92.590	92.590	92.590
BOTE: All	0.008	0.027	0.035	-0.004
BOTE: Variety Depth	-0.037	-0.082	-0.150	0.020
BOTE: Variety Breadth	0.051	0.129	0.190	-0.026
BOTE: Prices	-0.007	-0.020	-0.006	0.002

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety depth refers to the number of UPCs per store, variety breadth refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 2.5.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table 2.12: Effect of Variety and Prices on Other Measures of Concentration, RMS

VARIABLES	(1) Num Retailers	(2) Within-dept HHI	(3) Cross-dept variance	(4) Cross-HH variance
Variety Depth	-0.599	0.133***	0.00908	0.0428
	(0.438)	(0.0233)	(0.0105)	(0.0262)
Variety Breadth	-2.070	-0.211***	-0.0629	-0.273***
	(1.492)	(0.0792)	(0.0385)	(0.0982)
Price Index	0.476	-0.0164	-0.0206	-0.0248
	(0.630)	(0.0326)	(0.0157)	(0.0371)
Observations	974443	974443	974443	974443
R-squared	0.718	0.681	0.522	0.631
Prob > F	0.000	0.000	0.237	0.000
Number of units	82135	82135	82135	82135
Number of clusters	82135	82135	82135	82135
First stage F-stat	273.257	273.257	273.257	273.257
BOTE: All	0.068	0.463	0.009	-0.009
BOTE: Variety Depth	0.044	0.629	-0.046	0.129
BOTE: Variety Breadth	0.025	-0.165	0.053	-0.136
BOTE: Prices	-0.001	-0.002	0.002	-0.002

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety depth refers to the number of UPCs per store, variety breadth refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 2.5.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table 2.13: Effect of Variety and Prices on Trip Measures, HMS

VARIABLES	(1) TripNum	(2) UPC per trip	(3) Groups per trip	(4) Dept per trip
Variety Depth	1.335	-3.117***	-1.751***	-0.291***
	(1.707)	(0.450)	(0.264)	(0.0342)
Variety Breadth	-3.564	7.609***	4.247***	0.685***
	(3.991)	(1.043)	(0.614)	(0.0796)
RPI (US)	-2.855	-2.638***	-1.562***	-0.234***
	(2.819)	(0.714)	(0.418)	(0.0518)
Observations	3605864	3605864	3605856	3605856
R-squared	0.726	0.763	0.776	0.688
Prob > F	0.853	0.000	0.000	0.000
Number of units	190795	190795	190795	190795
Number of clusters	190795	190795	190795	190795
First stage F-stat	92.590	92.590	92.591	92.591
BOTE: All	0.008	0.085	0.075	0.394
BOTE: Variety Depth	-0.012	-0.220	-0.201	-1.168
BOTE: Variety Breadth	0.019	0.316	0.287	1.616
BOTE: Prices	0.001	-0.010	-0.010	-0.053

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety depth refers to the number of UPCs per store, variety breadth refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 2.5.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table 2.14: Effect of Variety and Prices on Trip Measures, RMS

VARIABLES	(1) TripNum	(2) UPC per trip	(3) Groups per trip	(4) Dept per trip
Variety Depth	-5.962***	1.992***	1.197***	0.147***
	(2.094)	(0.608)	(0.339)	(0.0381)
Variety Breadth	9.297	-1.305	-1.286	0.0159
	(6.821)	(1.946)	(1.125)	(0.134)
Price Index	-8.591***	1.548**	0.732	0.0489
	(3.171)	(0.783)	(0.455)	(0.0553)
Observations	974443	974443	974443	974443
R-squared	0.768	0.795	0.807	0.722
Prob > F	0.000	0.000	0.000	0.000
Number of units	82135	82135	82135	82135
Number of clusters	82135	82135	82135	82135
First stage F-stat	273.257	273.257	273.257	273.257
BOTE: All	0.076	0.179	0.161	0.393
BOTE: Variety Depth	0.098	0.197	0.193	0.383
BOTE: Variety Breadth	-0.025	-0.021	-0.034	0.007
BOTE: Prices	0.003	0.004	0.003	0.003

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety depth refers to the number of UPCs per store, variety breadth refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 2.5.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table 2.15: Effect of shopping trips on household HHI, IV with region average wage and unemployment rate, county-time FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV group	Age		Education		Age-Education		Age-Education-Occupation	
IV region	All	State	All	State	All	State	All	State
VARIABLES	Household HHI							
Log Trips	-0.297***	-0.382***	-0.261***	-0.160	-0.126*	-0.00410	-0.176**	0.152
	(0.0365)	(0.141)	(0.0613)	(0.102)	(0.0684)	(0.152)	(0.0862)	(0.503)
Observations	2274654	2274654	2295327	2293357	2267467	2198119	978902	678245
R-squared	0.749	0.723	0.754	0.756	0.753	0.716	0.767	0.642
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	127883	127883	128376	128364	127637	127097	60857	52918
Number of clusters	127883	127883	128376	128364	127637	127097	60857	52918
First stage F-stat	41.649	3.079	12.997	4.604	10.704	2.826	7.054	0.351
BOTE: Trips	1.677	1.824	1.197	0.762	0.597	0.019	0.828	-0.723
BOTE: IVs	0.224	0.004	0.120	0.006	-0.007	-0.000	-0.012	0.001

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.16: Effect of shopping trips on household HHI, IV with region average wage and unemployment rate, allowing for changes in household characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV group	Age		Education		Age-Education		Age-Education-Occupation	
IV region	All	State	All	State	All	State	All	State
VARIABLES	Household HHI							
Log Trips	-0.252***	-0.238***	-0.282***	-0.0866	-0.196***	-0.227***	-0.174***	-0.132*
	(0.0140)	(0.0236)	(0.0722)	(0.0861)	(0.0252)	(0.0361)	(0.0576)	(0.0701)
Observations	2322244	2322230	2322228	2320253	2322228	2253376	1016942	700950
R-squared	0.740	0.741	0.736	0.727	0.743	0.742	0.750	0.750
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	136511	136510	136511	136499	136511	135986	75443	66007
Number of clusters	136511	136510	136511	136499	136511	135986	75443	66007
First stage F-stat	198.670	83.068	8.472	7.610	70.796	37.384	15.843	11.010
BOTE: Trips	1.604	1.177	1.361	0.417	0.995	1.095	0.847	0.632
BOTE: IVs	0.604	0.066	0.004	0.007	0.098	0.023	0.030	0.008

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.17: Effect of shopping trips on household HHI, IV with region median wage and unemployment rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV group	Age		Education		Age-Education		Age-Education-Occupation	
IV region	All	State	All	State	All	State	All	State
VARIABLES	Household HHI							
Log Trips	-0.234***	-0.406***	-0.277***	-0.205***	-0.0771	-0.157*	-0.158**	0.00575
	(0.0387)	(0.0986)	(0.0615)	(0.0681)	(0.0504)	(0.0841)	(0.0659)	(0.222)
Observations	2285235	2285228	2305700	2303820	2278148	2210049	1007850	709914
R-squared	0.738	0.690	0.731	0.738	0.720	0.737	0.734	0.687
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	128280	128279	128764	128753	128038	127552	62106	54767
Number of clusters	128280	128279	128764	128753	128038	127552	62106	54767
First stage F-stat	32.955	6.276	11.771	10.068	21.014	7.833	11.706	1.377
BOTE: Trips	1.224	1.932	1.304	0.983	0.361	0.746	0.742	-0.027
BOTE: IVs	0.084	0.002	-0.098	0.018	-0.020	-0.003	-0.017	0.000

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.18: Effect of shopping trips on household HHI, IV with region average wage

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV group	Age		Education		Age-Education		Age-Education-Occupation	
IV region	All	State	All	State	All	State	All	State
VARIABLES	Household HHI							
Log Trips	-0.434***	-0.401	-0.181***	-0.0236	0.419	0.280	-0.210*	0.00510
	(0.0773)	(0.355)	(0.0659)	(0.137)	(0.324)	(0.293)	(0.118)	(0.919)
Observations	2285235	2285228	2305700	2303820	2278148	2210623	1007906	710297
R-squared	0.675	0.692	0.737	0.699	0.272	0.458	0.738	0.687
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	128280	128279	128764	128753	128038	127556	62106	54773
Number of clusters	128280	128279	128764	128753	128038	127556	62106	54773
First stage F-stat	22.761	1.014	22.522	6.295	2.921	2.628	7.091	0.156
BOTE: Trips	2.348	1.912	0.793	0.113	-1.960	-1.327	0.983	-0.024
BOTE: IVs	0.199	0.001	0.079	0.002	0.049	0.005	-0.019	0.000

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.19: Effect of shopping trips on household HHI, IV with region median wage

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV group	Age		Education		Age-Education		Age-Education-Occupation	
IV region	All	State	All	State	All	State	All	State
VARIABLES	Household HHI							
Log Trips	-0.348***	0.0959	0.0758	-0.0475	0.0316	0.0532	-0.197**	-0.0975
	(0.0993)	(0.565)	(0.201)	(0.158)	(0.0764)	(0.135)	(0.0830)	(0.383)
Observations	2285235	2285228	2305700	2303820	2278148	2210049	1007850	709914
R-squared	0.714	0.630	0.643	0.708	0.672	0.660	0.737	0.726
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	128280	128279	128764	128753	128038	127552	62106	54767
Number of clusters	128280	128279	128764	128753	128038	127552	62106	54767
First stage F-stat	9.923	0.474	3.463	4.374	22.375	7.847	14.358	0.797
BOTE: Trips	1.790	-0.456	-0.345	0.227	-0.147	-0.252	0.926	0.463
BOTE: IVs	0.076	0.000	0.033	0.003	0.009	0.001	-0.023	-0.001

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.20: Effect of shopping trips on household HHI, IV with region unemployment rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV group	Age		Education		Age-Education		Age-Education-Occupation	
IV region	All	State	All	State	All	State	All	State
VARIABLES	Household HHI							
Log Trips	-0.170***	-0.423***	-0.336***	-0.229***	-0.200***	-0.319***	-0.105	-0.227
	(0.0420)	(0.101)	(0.0699)	(0.0762)	(0.0630)	(0.120)	(0.111)	(0.299)
Observations	2285235	2285228	2305700	2304252	2278148	2240530	1010522	766046
R-squared	0.737	0.681	0.717	0.737	0.739	0.723	0.725	0.740
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	128280	128279	128764	128753	128038	127707	62141	56318
Number of clusters	128280	128279	128764	128753	128038	127707	62141	56318
First stage F-stat	54.345	12.160	18.622	15.891	23.105	8.020	9.142	1.224
BOTE: Trips	0.814	2.008	1.607	1.089	0.952	1.516	0.501	1.079
BOTE: IVs	0.023	0.003	0.035	0.004	0.006	0.001	0.001	0.000

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.21: Effect of shopping trips on other measures of concentration and trip measures, IV with household employment

VARIABLES	(1) Log Num Retailers	(2) Within-dept HHI	(3) Cross-dept variance	(4) Cross-HH variance	(5) UPC per trip	(6) Groups per trip	(7) Dept per trip
Log Trips	0.861***	-0.170***	0.0475***	-0.254***	-9.168***	-5.225***	-0.506***
	(0.0280)	(0.0127)	(0.00578)	(0.0151)	(0.196)	(0.113)	(0.0180)
Observations	3622992	3622992	3622992	3622992	3622992	3622990	3622990
R-squared	0.812	0.707	0.533	0.603	0.922	0.929	0.797
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	193527	193527	193527	193527	193527	193527	193527
Number of clusters	193527	193527	193527	193527	193527	193527	193527
First stage F-stat	116.170	116.170	116.170	116.170	116.170	116.180	116.180
BOTE: Trips	0.828	1.284	0.612	1.423	2.599	2.413	8.150

Notes: Robust standard errors are in parentheses, clustered by household. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.22: Effect of shopping trips on other measures of concentration and trip measures, IV with region average wage and unemployment rate, age and education

VARIABLES	(1) Log Num Retailers	(2) Within-dept HHI	(3) Cross-dept variance	(4) Cross-HH variance	(5) UPC per trip	(6) Groups per trip	(7) Dept per trip
Log Trips	0.966***	-0.261***	0.0300**	-0.0613*	-10.16***	-5.616***	-0.695***
	(0.0675)	(0.0289)	(0.0132)	(0.0342)	(0.499)	(0.275)	(0.0444)
Observations	2278148	2278148	2278148	2278148	2278148	2278147	2278147
R-squared	0.815	0.708	0.544	0.625	0.926	0.933	0.795
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	128038	128038	128038	128038	128038	128038	128038
Number of clusters	128038	128038	128038	128038	128038	128038	128038
First stage F-stat	30.737	30.737	30.737	30.737	30.737	30.730	30.730
BOTE: Trips	1.154	2.442	0.481	0.427	3.581	3.224	13.923
BOTE: IVs	0.198	0.419	0.083	0.073	0.615	0.554	2.392

Notes: Robust standard errors are in parentheses, clustered by household. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.23: Effect of Online Shopping (Upper Bound)

VARIABLES	(1) HH Offline HHI	(2)	(3) HH HHI	(4)	(5) Offline Trip Number	(6) Offline Number of Retailers
Online Share	0.0578***	0.0565***	-0.174***	-0.176***	-14.51***	-1.774***
	(0.00582)	(0.00576)	(0.00637)	(0.00626)	(0.486)	(0.0701)
log(income)	0.00285***	0.00285***	0.00244***	0.00260***	-0.576***	-0.0287**
	(0.000657)	(0.000459)	(0.000607)	(0.000441)	(0.0680)	(0.0133)
Household Size	0.000953*	0.000357	0.000871*	0.000303	0.736***	0.0609***
	(0.000516)	(0.000380)	(0.000493)	(0.000368)	(0.0523)	(0.0110)
Observations	1432259	2409750	1433160	2410651	1432259	1432259
R-squared	0.687	0.711	0.686	0.712	0.761	0.720
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000
Number of clusters	49	49	49	49	49	49
Year-quarter FE	X	X	X	X	X	X
HH FE	X	X	X	X	X	X
HH-quarter FE	X	X	X	X	X	X
BOTE	0.013	0.012	-0.046	-0.046	0.026	0.014

Notes: Robust standard errors are in parentheses, clustered by state. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Column (2) and (4) use all households and the rest of columns use households who have reported online shopping at least once. Household online share increased by 0.012 from 2004 to 2016. Household offline HHI increased by 0.055, HH HHI 0.0458. Offline trip number decreased by 6.62 per quarter, offline number of retailers 1.54 per quarter.

Table 2.24: Effect of Amazon tax on household concentration

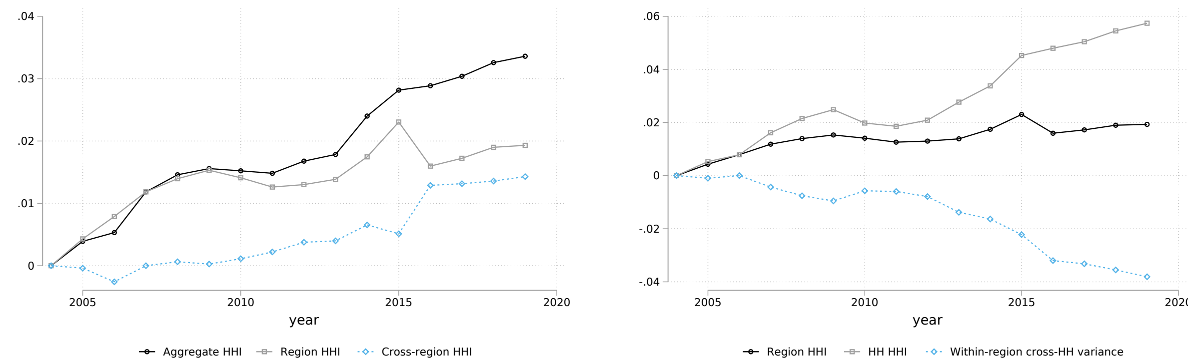
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Household Retail Concentration			Number of Trips		
$TreatedStates \times I(t \geq treatedYQ)$	-0.0006			0.2155***		
	(0.0008)			(0.0693)		
$TreatedStates \times I(t = treatedYQ - 1)$		0.0001			-0.0790	
		(0.0008)			(0.0703)	
$TreatedStates \times I(t = treatedYQ)$		-0.0007			0.0306	
		(0.0009)			(0.0606)	
$TreatedStates \times I(t = treatedYQ + 1)$		0.0012			-0.0827	
		(0.0009)			(0.0720)	
$TreatedStates \times I(t \geq treatedYQ) \times TaxRate$			0.003			2.8245**
			(0.0120)			(1.1970)
Observations	2357822	2357822	2357822	2357822	2357822	2357822
R-squared	0.6687	0.6687	0.6687	0.7367	0.7367	0.7367
Prob > F	0.001	0.0049	0.0018	0.000	0.000	0.000
Number of clusters	2927	2927	2927	2927	2927	2927
Fixed Effects	X	X	X	X	X	X
Household Income	X	X	X	X	X	X

Notes: Robust standard errors are in parentheses, clustered by county. *** p<0.01, ** p<0.05, * p<0.1. From 2006 to 2016, 21 states are treated. The treatment has no effect on household online share, household offline concentration or overal household concentration.

2.11.5 *Figures*

Figure 2.12: Decomposing Changes in Retail Concentration

(a) National and Market (b) Market and Household



Notes: This figure shows that HHI can be decomposed at each level following Radaelli and Zenga (2002). For each HHI term, we plot the yearly change over the sample period while we plot the negative of the cross-variance since it has a negative contribution to HHI.

Figure 2.13: Sorted Effects across Households

(a) Supercenter Entry (b) Club Entry

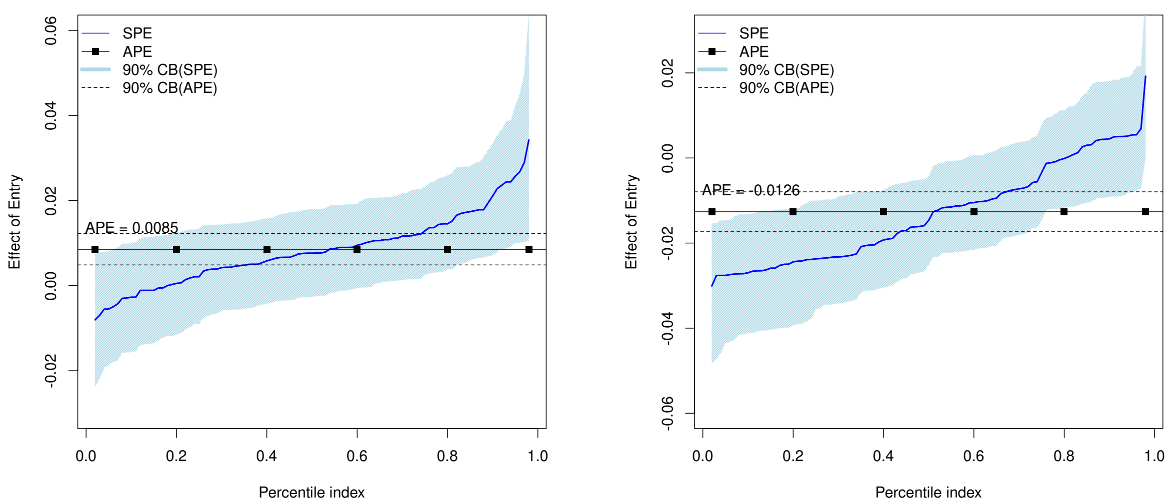
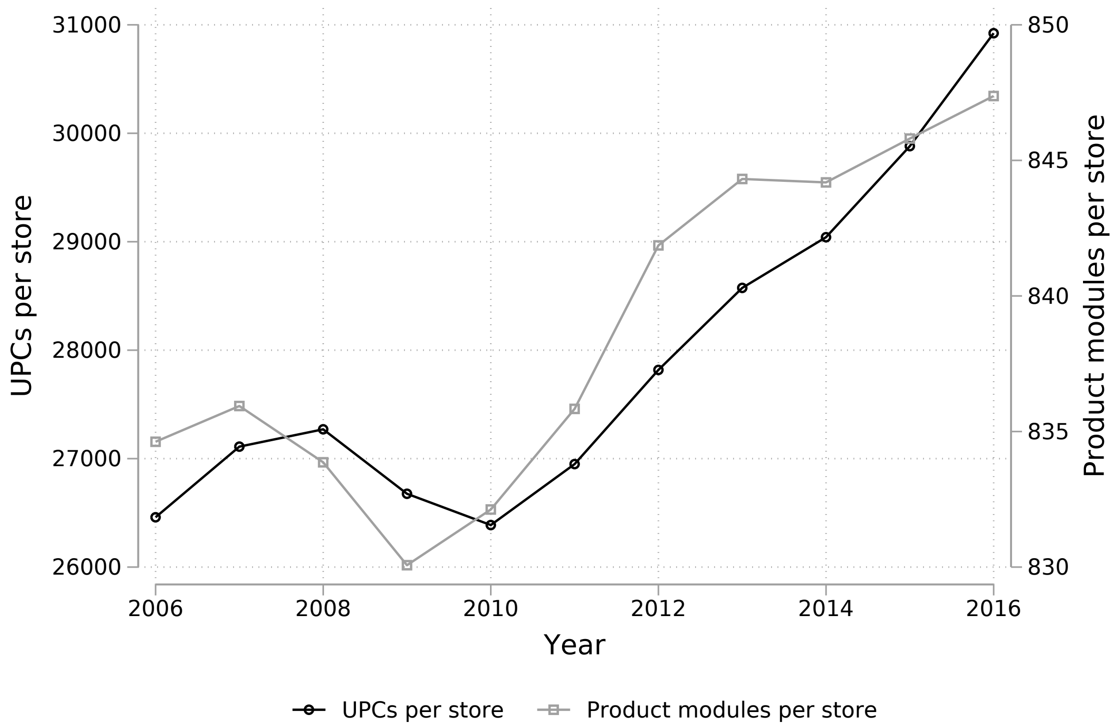


Figure 2.14: UPCs and product modules per store, RMS



Notes: This figure shows the number of UPCs and product modules per store in the RMS. The average is calculated as the weighted average across all stores using store revenue as weights.

Figure 2.15: Household Concentration and Online Shopping

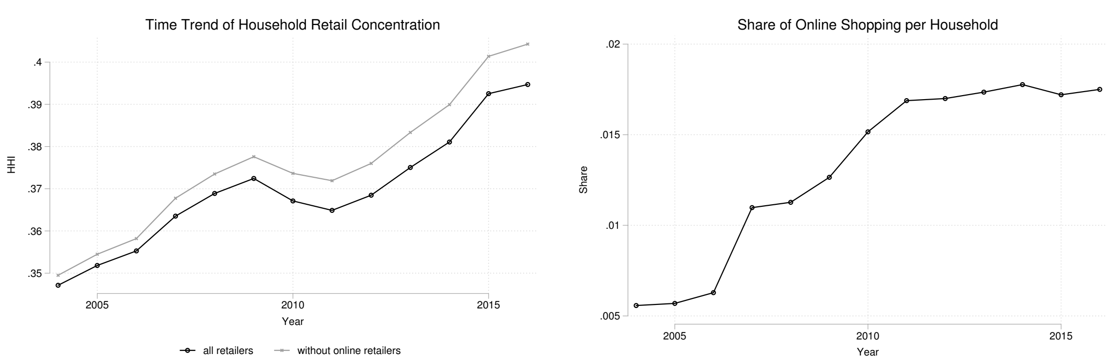


Figure 2.16: Share of Online Shopping in Each Department

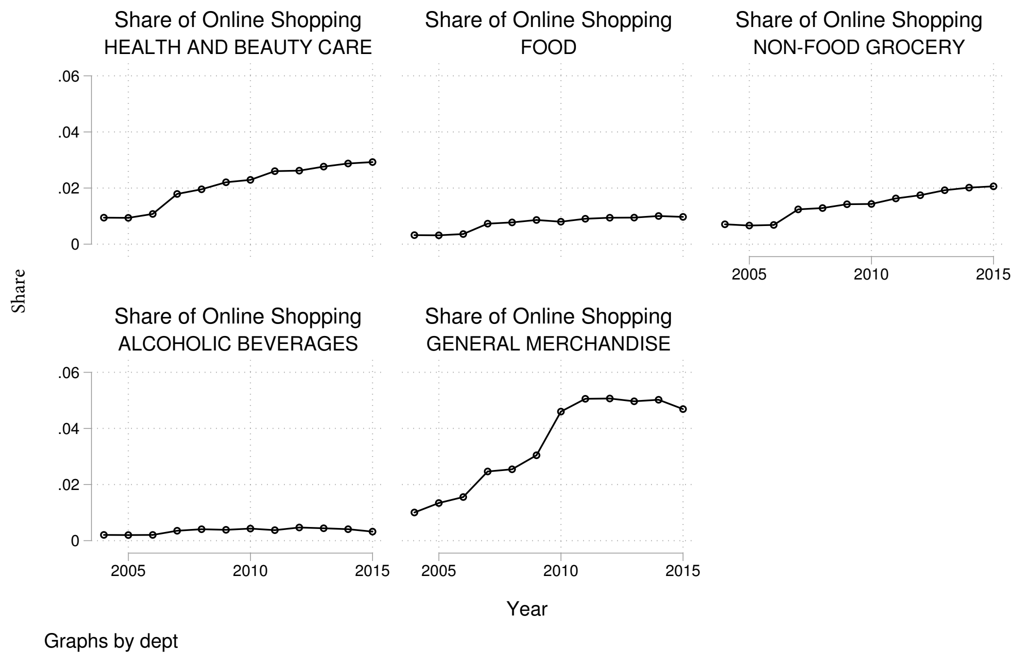
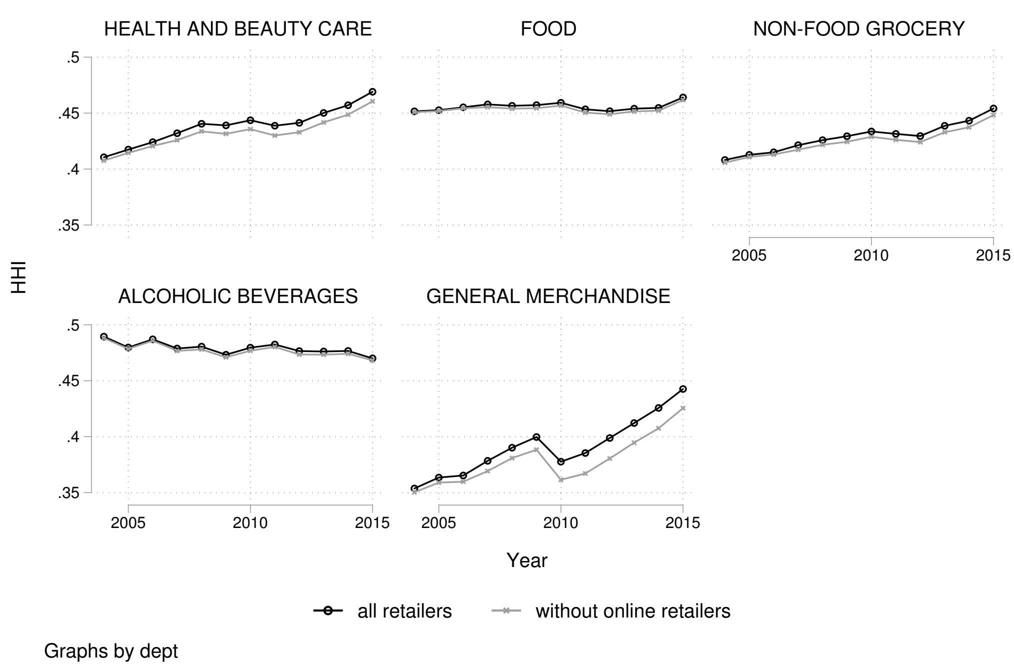


Figure 2.17: Household Concentration With and Without Online Shopping



REFERENCES

Aguiar, Mark, and Erik Hurst. 2007*a*. "Life-Cycle Prices and Production." *The American Economic Review*, 97(5): 1533–1559.

Aguiar, Mark, and Erik Hurst. 2007*b*. "Life-cycle prices and production." *American Economic Review*, 97(5): 1533–1559.

Aguiar, Mark, Erik Hurst, and Loukas Karabarbounis. 2013. "Time Use During the Great Recession." *American Economic Review*, 103(5): 1664–1696.

Aguiar, Mark, Mark Bils, Kerwin Kofi Charles, and Erik Hurst. 2021. "Leisure Luxuries and the Labor Supply of Young Men." *Journal of Political Economy*, 129(2): 337–382. Publisher: The University of Chicago Press.

Ailawadi, Kusum L., Jie Zhang, Aradhna Krishna, and Michael W. Kruger. 2010. "When Wal-Mart Enters: How Incumbent Retailers React and how this Affects their Sales Outcomes." *Journal of Marketing Research*, 47(4): 577–593. Publisher: SAGE Publications Inc.

Ailawadi, Kusum L, Yu Ma, and Dhruv Grewal. 2018. "The club store effect: Impact of shopping in warehouse club stores on consumers' packaged food purchases." *Journal of Marketing Research*, 55(2): 193–207.

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dub, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell. 2019. "Food Deserts and the Causes of Nutritional Inequality." *The Quarterly Journal of Economics*, 134(4): 1793–1844.

Anderson, Simon P., and Joel Waldfogel. 2015. "Chapter 1 - Preference Externalities in Media Markets." In *Handbook of Media Economics*. Vol. 1 of *Handbook of Media Economics*, , ed. Simon P. Anderson, Joel Waldfogel and David Strmberg, 3–40. North-Holland.

Andreyeva, Tatiana, Michael W Long, and Kelly D Brownell. 2010. "The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food." *American journal of public health*, 100(2): 216–222.

Arcidiacono, Peter, Paul B. Ellickson, Carl F. Mela, and John D. Singleton. 2020. "The Competitive Effects of Entry: Evidence from Supercenter Expansion." *American Economic Journal: Applied Economics*, 12(3): 175–206.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro. 2017. "Retail Globalization and Household Welfare: Evidence from Mexico." *Journal of Political Economy*, 126(1): 1–73.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *The Quarterly Journal of Economics*, 135(2): 645–709.

Balto, David A. 2001. "Supermarket Merger Enforcement." *Journal of Public Policy & Marketing*, 20(1): 38–50. Publisher: SAGE Publications Inc.

Basker, Emek, Shawn Klimek, and Pham Hoang Van. 2012. "Supersize It: The Growth of Retail Chains and the Rise of the Big-Box Store." *Journal of Economics & Management Strategy*, 21(3): 541–582. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1530-9134.2012.00339.x.

Benkard, C. Lanier, Ali Yurukoglu, and Anthony Lee Zhang. 2021. "Concentration in Product Markets." National Bureau of Economic Research Working Paper 28745. Series: Working Paper Series.

Berry, Steven, Martin Gaynor, and Fiona Scott Morton. 2019. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives*, 33(3): 44–68.

Bertrand, Marianne, and Francis Kramarz. 2002. "Does Entry Regulation Hinder Job Creation? Evidence from the French Retail Industry." *The Quarterly Journal of Economics*, 117(4): 1369–1413.

Bhat, Chandra R. 2005. "A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions." *Transportation Research Part B: Methodological*, 39(8): 679–707.

Bronnenberg, Bart J. 2015. "The provision of convenience and variety by the market." *The RAND Journal of Economics*, 46(3): 480–498. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12094.

Bronnenberg, Bart J. 2018. "Retailing and consumer demand for convenience." *Handbook of Research on Retailing*. ISBN: 9781786430281 Publisher: Edward Elgar Publishing Section: Handbook of Research on Retailing.

Bronnenberg, Bart J, Tobias J Klein, and Yan Xu. 2020. "Consumer Time Budgets and Grocery Shopping Behavior." Working Paper.

Carden, Art, and Charles Courtemanche. 2016. "The evolution and impact of the general merchandise sector." *Handbook on the Economics of Retailing and Distribution*. ISBN: 9781783477388 Publisher: Edward Elgar Publishing Section: Handbook on the Economics of Retailing and Distribution.

Chernozhukov, Victor, Ivn FernndezVal, and Ye Luo. 2018. "The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages." *Econometrica*, 86(6): 1911–1938.

Coibion, Olivier, Yuriy Gorodnichenko, and Dmitri Koustas. 2021. "Consumption Inequality and the Frequency of Purchases." *American Economic Journal: Macroeconomics*, 13(4): 449–482.

Covarrubias, Matias, Germn Gutirrez, and Thomas Philippon. 2020. "From Good to Bad Concentration? US Industries over the Past 30 Years." *NBER Macroeconomics Annual*, 34: 1–46. Publisher: The University of Chicago Press.

DellaVigna, Stefano, and Matthew Gentzkow. 2019. "Uniform pricing in us retail chains." *The Quarterly Journal of Economics*, 134(4): 2011–2084.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *The Quarterly Journal of Economics*, 135(2): 561–644.

Ellickson, Paul B. 2016. "The evolution of the supermarket industry: from A & P to Walmart." *Handbook on the Economics of Retailing and Distribution*. ISBN: 9781783477388 Publisher: Edward Elgar Publishing Section: Handbook on the Economics of Retailing and Distribution.

Ellickson, Paul B, Sanjog Misra, and Harikesh S Nair. 2012. "Repositioning dynamics and pricing strategy." *Journal of Marketing Research*, 49(6): 750–772.

Feenstra, Robert C., Luca Macedoni, and Mingzhi Xu. 2022. "Large Firms, Consumer Heterogeneity and the Rising Share of Profits." National Bureau of Economic Research Working Paper 29646. Series: Working Paper Series.

Ganapati, Sharat. 2021. "Growing Oligopolies, Prices, Output, and Productivity." *American Economic Journal: Microeconomics*, 13(3): 309–327.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely. 2019. "Are US Industries Becoming More Concentrated?" *Review of Finance*, 23(4): 697–743.

Hausman, Jerry A., and Timothy F. Bresnahan. 2008. "5. Valuation of New Goods under Perfect and Imperfect Competition." In *The Economics of New Goods*. 209–248. University of Chicago Press.

Hausman, Jerry, and Ephraim Leibtag. 2007. "Consumer benefits from increased competition in shopping outlets: Measuring the effect of Wal-Mart." *Journal of Applied Econometrics*, 22(7): 1157–1177.

Holmes, Thomas J. 2011. "The diffusion of Wal-Mart and economies of density." *Econometrica*, 79(1): 253–302.

Hortaçsu, Ali, and Chad Syverson. 2015. "The ongoing evolution of US retail: A format tug-of-war." *Journal of Economic Perspectives*, 29(4): 89–112.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity." *The Quarterly Journal of Economics*, 131(3): 1291–1364.

Hsieh, Chang-Tai, and Esteban Rossi-Hansberg. 2019. "The Industrial Revolution in Services." National Bureau of Economic Research Working Paper 25968.

Hwang, Minha, and Sungho Park. 2016. "The Impact of Walmart Supercenter Conversion on Consumer Shopping Behavior." *Management Science*, 13.

Jia, Panle. 2008. "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry." *Econometrica*, 76(6): 1263–1316.

Kwon, Spencer Yongwook, Yueran Ma, and Kaspar Zimmermann. 2021. "100 Years of Rising Corporate Concentration." *SSRN Electronic Journal.*

Lee, Lung-Fei, and Mark M Pitt. 1986. "Microeconometric demand system with binding nonnegativity constraints: the dual approach." *Econometrica: Journal of the Econometric Society*, 1237–1242.

Leung, Justin H. 2021. "Minimum Wage and Real Wage Inequality: Evidence from Pass-Through to Retail Prices." *The Review of Economics and Statistics*, 103(4): 754–769.

Leung, Justin H., and Zhonglin Li. 2021*a*. "Big-box Store Expansion and Consumer Welfare." Working Paper.

Leung, Justin H., and Zhonglin Li. 2021*b*. "Rising Retail Concentration: Superstar Firms and Household Demand." Working Paper.

Mehta, Nitin. 2015. "A flexible yet globally regular multigood demand system." *Marketing Science*, 34(6): 843–863.

Mehta, Nitin, and Yu Ma. 2012. "A multicategory model of consumers' purchase incidence, quantity, and brand choice decisions: Methodological issues and implications on promotional decisions." *Journal of Marketing Research*, 49(4): 435–451.

Melitz, Marc J., and Sao Polanec. 2015. "Dynamic Olley-Pakes productivity decomposition with entry and exit." *The RAND Journal of Economics*, 46(2): 362–375.

Messinger, Paul R., and Chakravarthi Narasimhan. 1997. "A Model of Retail Formats Based on Consumers' Economizing on Shopping Time." *Marketing Science*, 16(1): 1–23. Publisher: INFORMS.

Neiman, Brent, and Joseph Vavra. 2021. "The Rise of Niche Consumption." Working Paper.

Nevo, Aviv, and Arlene Wong. 2019. "The Elasticity of Substitution Between Time and Market Goods: Evidence from the Great Recession." *International Economic Review*, 60(1): 25–51. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/iere.12343.

Okrent, Abigail, and Julian Alston. 2012. "The demand for disaggregated food-away-from-home and food-at-home products in the United States." *USDA-ERS Economic Research Report*, , (139).

Radaelli, P., and M. Zenga. 2002. "A "New" Two Term Subtractive Decomposition of Herfindahl Concentration Measure." Vol. Sessioni Spontanee, 525–528. Societ Italiana di Statistica.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter. 2021. "Diverging Trends in National and Local Concentration." *NBER Macroeconomics Annual*, 35: 115–150. Publisher: The University of Chicago Press.

Sadun, Raffaella. 2014. "Does Planning Regulation Protect Independent Retailers?" *Review of Economics and Statistics*, 97(5): 983–1001.

Seo, Boyoung. 2019. "Firm scope and the value of one-stop shopping in washington state's deregulated liquor market." *Kelley school of business research paper*, , (16-70).

Song, Inseong, and Pradeep K Chintagunta. 2007. "A discrete–continuous model for multicategory purchase behavior of households." *Journal of Marketing Research*, 44(4): 595–612.

Syverson, Chad. 2019. "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives*, 33(3): 23–43.

Thomassen, Øyvind, Howard Smith, Stephan Seiler, and Pasquale Schiraldi. 2017. "Multi-category competition and market power: a model of supermarket pricing." *American Economic Review*, 107(8): 2308–51.

U.S. Department of Justice, and Federal Trade Commission. 2010. "Horizontal Merger Guidelines."

Zhou, Tingyu. 2017. "Does Entry Regulation of Big-box Stores Protect the Retail Sector? Evidence from Store Cap Ordinances in the U.S." Working Paper.