Three Essays on Empirical Industrial Organization in Grocery Retailing

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum (dr. rer. pol.)
im Fach Volkswirtschaftslehre

eingereicht an der Wirtschaftswissenschaftlichen Fakultät
Heinrich-Heine-Universität Düsseldorf

von: Anna W. Lu
geboren am 01.03.1989 in München

Erstgutachter: Prof. Dr. Tomaso Duso
Zweitgutachter: Prof. Dr. Justus Haucap

Abgabedatum: 22. Mai 2017

Acknowledgment

This thesis was written during my time as a research assistant at the Düsseldorf Institute for Competition Economics (DICE) and at the German Institute for Economic Research (DIW Berlin). I am lucky to have been a part of two such vibrant research environments, and I thank my colleagues for many insightful discussions.

Most of all, I have to thank my first advisor Tomaso Duso for the tremendous amount of time, energy, and support he has given me. He introduced me to the field of empirical industrial organization, always encouraged me to learn new things, and gave me countless opportunities to present and discuss my work. I am also grateful to my second advisor Justus Haucap who gave me the opportunity to do a Ph.D., guided me through my discovery of a fascinating industry and taught me a lot about research dissemination and the potential policy impact of my research.

My thesis was vastly improved through conversations with Ulrich Doraszelski, Jean-Pierre Dubé, Matthew Osborne, and Stephan Seiler. I also received helpful comments and suggestions from Jana Friedrichsen, Floran Heiss, Ali Hortaçsu, Nadine Riedel, Hannes Ullrich, and Lars Zeigermann.

I am thankful to Federico Ciliberto and Céline Bonnet who invited me to the University of Virginia and the Toulouse School of Economics, respectively, and taught me much about academic research in general and my field in particular. I would also like to thank Hans-Theo Normann for his always insightful advice on how to navigate the early stages of a research career.

I want to thank my colleagues and friends for many good times in and out of the office. In particular, I want to thank Germain Gaudin for his thoughtful comments on numerous drafts, for always making me laugh, and for being by my side. Last but not least, I am grateful to my parents and my sister for their encouragement, love, and inspiring curiosity in the world and how it works.

Contents

List of Tables

List of Figures

Chapter 1

General Introduction

When we look at modern-day grocery retailing, we get a multi-faceted picture of how firms and consumers make strategic decisions: On the supply side, retailers compete not just in prices, but also in various other dimensions, such as product quality, assortment variety, store location, customer service, marketing campaigns, and advertising. On the demand side, consumers have to choose from increasingly large assortments, navigate promotional activities and loyalty programs, and make a multitude of decisions about where, when, and how much to buy.

Understanding these strategic decisions is important for multiple reasons: It helps regulators to design optimal interventions, e.g. in competition policy, consumer protection, or taxation; it helps firms to improve their business strategies; and it helps researchers to gain general insights into human decision-making. All of this is particularly relevant in the supermarket industry which occupies a vital role both in the economy and in the everyday life of consumers: In the recent years, the sector has had an annual turnover of ca. 1,100 EUR billion in the European Union[1] and ca. 650 USD billion in the United States.[2] Supermarkets constitute the primary channel in consumer access to nutrition – in the EU for example, the average household spends almost 13% of its budget on food and non-alcoholic beverages (Gerstberger and Yaneva 2013).

Consequently, a vast body of literature studies the supermarket industry.[3] In this thesis, I build primarily on the economics literature but I also draw from the marketing literature. The two literatures are closely related and often look at the same questions from different angles – economists are mostly interested in social welfare and in implications for public policy whereas marketers focus on firm performance. In the past two decades, the empirical analysis of the supermarket industry has been advanced by two important, interdependent innovations; one pioneered by the economics literature and one pioneered by the marketing literature.

The first innovation is that large-scale scanner panel data became available. Scanner-panel data generally comes in two forms: Store panels that collect sales information from stores' check-out barcode scanners, and household panels in which participants use handheld barcode scanners to scan their purchases after each shopping trip. Both types of data typically include detailed information on products,

[1] `www.ec.europa.eu/competition/sectors/agriculture/overview_en.html`. Last accessed on 30 March 2017.

[2] `www.fmi.org/research-resources/supermarket-facts`. Last accessed on 1 February 2017.

[3] Recent surveys include Basker (2007) and Ellickson (2016).

prices, and marketing-mix variables, thus providing rich variation across time and panelists. While marketers explored such scanner data already in the 1980s, economists only started to use them in the mid to late 1990s. Since then, estimation using scanner data has become an important part of empirical industrial organization.

The second innovation was the rise of structural models in the estimation of demand and supply. Structural models are directly derived from economic theory and impose a set of model assumptions on the data in order to identify the underlying policy-invariant parameters of firm and consumer decision-making. Structural models are a powerful tool to simulate previously unseen environments and evaluate welfare effects. Although they have a long tradition in economics (e.g. Marschak 1953), they only gained wide popularity when computing power improved and bigger and better data became available.[4] Today, structural models do not only play a major role in empirical industrial organization but have also gained traction in quantitative marketing (for surveys see Dube et al. (2002) or Chintagunta et al. (2006)). They co-exist with the more traditional "reduced-form" models that provide statistical estimation without reference to specific economic models and are thus less suited to make predictions and simulations. Both structural and reduced-form models are valuable tools in their own right; ultimately, the choice between them depends on the research question and the data structure.

In this thesis, I combine large-scale scanner data with the estimation tools of empirical industrial organization to provide insights into the strategic decisions of retailers and consumers. I explore issues that have become particularly relevant in the modern landscape of retailing. Specifically, I study research questions related to two trends that have been persistent in the industry since its beginnings in the early 20th century. The first trend is a continuous increase in market concentration: In the U.S. for example, the eight largest chains had less than 30% market share in 1992 but about 50% in 2013,[5] consequently raising attention from competition authorities. The second trend is a steady increase in the number of available products. Back in the 1920s, groceries were mostly sold in individually owned stores with limited variety and low turnover. With the rise of chain stores, assortments grew and store size increased; modern mass-retailing continued this trend with the introduction

[4] For a survey of how structural models have been used in empirical industrial organization see Einav and Levin (2010).

[5] `www.ers.usda.gov/topics/food-markets-prices/retailing-wholesaling/retail-trends/`. Last accessed on 30 March 2017.

of supercenters and hypermarkets in the early 1990s. In the U.S., the number of products in an average supermarket rose from 8,948 to almost 47,000 between 1975 and 2008.[6]

In Chapter 2, titled "Retail Mergers and Assortment Repositioning", I study how supermarkets compete in product variety and what this means for merger control. Authorities previously focused their attention on price effects of mergers. However, the recent years saw a shift in attention: For example, the new U.S. Horizontal Merger Guidelines (2010), issued by the Department of Justice and the Federal Trade Commission, now emphasize the importance of merger effects on product variety. So far, the literature has found ambiguous effects of mergers on store level assortment, i.e. the average number of products carried by a store (Pires and Trindade 2015; Argentesi et al. 2016).

I extend the analysis to look at *strategic repositioning* between stores. I study a series of local mergers in the U.S. and find that, while mergers increased store-level assortment by 7.4%, they also increased the assortment overlap between stores by 3%, suggesting that gains from business-stealing exceeded losses from cannibalization. Effectively, a merger raised the number of available products in the average market by 6.9%, i.e. mergers, on average, improved consumer access to variety. The analysis is based on a data-driven market definition: Unlike the existing literature, I use machine-learning methods to cluster stores into local markets according to their geographic location, thus yielding highly flexible markets.

All in all, the findings in this chapter support the notion that product variety is an important factor in merger control. Merger effects on store-level assortments tell only one part of the story – in order to get the full picture, regulators should also look at the merger effect on substitutability between stores. This is particularly relevant in industries in which consumers switch frequently between stores and in which store assortment is one of the major drivers of shopper patronage (Hoch, Bradlow and Wansink 1999; Fox, Montgomery and Lodish 2004; Briesch, Chintagunta and Fox 2009).

Chapter 3, titled "Consumer Stockpiling and Sales Promotions", looks at consumer stockpiling and retailer pricing in storable goods markets. It builds on the well-established finding that consumers tend to stockpile for future use when they

[6] `www.consumerreports.org/cro/magazine/2014/03/too-many-product-choices-in-supermarkets/index.htm`. Last accessed on 30 March 2017.

face price promotions, anticipating that prices will go up in later periods. While the literature has developed sophisticated models to describe and estimate such forward-looking consumer behavior (e.g. Erdem, Imai and Keane 2003; Hendel and Nevo 2006; Seiler 2013), there is very little empirical work on how retailers should respond to it, and in particular, how they should best design promotions.[7]

I am the first to compare the effects of promotion length with the effects of promotion depth on long-run consumer purchases and seller revenues. I study the U.S. laundry detergent market which is characterized by frequent price promotions and substantial consumer stockpiling. I estimate a dynamic discrete-choice model of strategic stockpiling in which consumers are forward-looking with rational expectations over future prices. Consumers can keep storage but incur storage costs. I use the estimates from this model to simulate pricing policies of varying promotion length and depth. I find that, in the detergent market, the revenue elasticity with respect to promotion depth is about four times larger than with respect to promotion length. My results suggest that retailers should increase promotion depth rather than length in markets with steady consumption rates, few demand shocks, and large heterogeneity in storage costs and price sensitivity.

Chapter 4, titled "Inference of Consumer Consideration Sets", investigates how consumers make decisions when they face a large number of alternatives and how the underlying, typically unobserved decision process can be inferred from observed choice data. It is motivated by the fact that, in the modern market place, consumers often have to choose from an overwhelmingly large variety of products. In order to simplify this decision problem, consumers tend to reduce the number of objectively available alternatives to a subset of relevant alternatives. The literature refers to this subset as a "consideration set".[8] In structural models of demand, we typically have to make assumptions about consideration sets because they are usually unobserved. However, these assumptions are not innocuous: If the consideration model is misspecified, the demand estimates can be severely biased (Sovinsky 2008; Draganska and Klapper 2011; Conlon and Mortimer 2013).

I propose a novel framework to formally test any two competing models of consideration against one another in order to determine which model fits the data best. My test follows the intuition of a menu approach (e.g. Nevo 2001; Villas-Boas 2007)

[7] Notable exceptions include Nair (2007), Osborne (2010), and Hendel and Nevo (2013).

[8] For a review of the literature on consideration sets see Roberts and Lattin (1997).

and uses supplemental data on marginal cost shifters to construct overidentifying restrictions. To illustrate my approach, I show an application to German grocery retailing in the categories of coffee and milk. Specifically, I test two popular models against each other. The first model is the workhorse model in empirical industrial organization; it assumes that consumers consider all products in the market. The second model is more popular in quantitative marketing research and models a two-stage consideration process in which consumers first choose a store and then choose a product within this store. I find that consideration sets differ fundamentally across product categories: The two-stage model outperforms the model of global consideration in the milk category but not in the coffee category. I relate this finding to differences in demand and supply conditions of the two product markets.

Lastly, I conclude in Chapter 5 and summarize the main results and insights of my thesis.

Bibliography

Argentesi, Elena, Paolo Buccirossi, Roberto Cervone, Tomaso Duso, and Alessia Marrazzo. 2016. “The Effect of Supermarket Mergers on Variety.” Working Paper.

Basker, Emek. 2007. “The Causes and Consequences of Wal-Mart’s Growth.” *Journal of Economic Perspectives*, 21(3): 177–198.

Briesch, Richard A., Pradeep K. Chintagunta, and Edward J. Fox. 2009. “How Does Assortment Affect Grocery Store Choice?” *Journal of Marketing Research*, 46(2): 176–189.

Chintagunta, Pradeep, Tülin Erdem, Peter E. Rossi, and Michel Wedel. 2006. “Structural Modeling in Marketing: Review and Assessment.” *Marketing Science*, 25(6): 604–616.

Conlon, Christopher T., and Julie Holland Mortimer. 2013. “Demand Estimation under Incomplete Product Availability.” *American Economic Journal: Microeconomics*, 5(4): 1–30.

Draganska, Michaela, and Daniel Klapper. 2011. “Choice Set Heterogeneity and the Role of Advertising: an Analysis with Micro and Macro Data.” *Journal of Marketing Research*, 48(4): 653–669.

Dube, Jean-Pierre, Pradeep Chintagunta, Amil Petrin, Bart Bronnenberg, Ron Goettler, P.B. Seetharaman, K. Sudhir, Raphael Thomadsen, and Ying Zhao. 2002. “Structural Applications of the Discrete Choice Model.” *Marketing Letters*, 13(3): 207–220.

Einav, Liran, and Jonathan Levin. 2010. “Empirical Industrial Organization: A Progress Report.” *Journal of Economic Perspectives*, 24(2): 145–162.

Ellickson, Paul B. 2016. “The Evolution of the Supermarket Industry: From A&P to Walmart.” In *Handbook on the Economics of Retailing and Distribution.* , ed. Emek Basker. Edward Elgar Publishing, Northhampton.

Erdem, Tülin, Susumu Imai, and Michael P. Keane. 2003. "Brand and Quantity Choice Dynamics Under Price Uncertainty." *Quantitative Marketing and Economics*, 1(1): 5–64.

Fox, Edward J., Alan L. Montgomery, and Leonard M. Lodish. 2004. "Consumer Shopping and Spending Across Retail Formats." *Journal of Business*, 77(2): 25–60.

Gerstberger, Christine, and Daniela Yaneva. 2013. "Analysis of EU-27 Household Final Consumption Expenditure: Baltic Countries and Greece Still Suffering Most from the Economic and Financial Crisis." *Eurostat: Statistics in Focus*, 2: 1–7.

Hendel, Igal, and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, 74(6): 1637–1673.

Hendel, Igal, and Aviv Nevo. 2013. "Intertemporal Price Discrimination in Storable Goods Markets." *American Economic Review*, 103(7): 2722–2751.

Hoch, Stephen J., Eric T. Bradlow, and Brian Wansink. 1999. "The Variety of an Assortment." *Marketing Science*, 18(4): 527–546.

Marschak, J. 1953. "Economic Measurements for Policy and Prediction. Studies in Econometric Method." In *Studies in Econometric Method.* , ed. William C. Hood and Tjalling C. Koopmans. John Wiley and Sons Inc., New York.

Nair, Harikesh. 2007. "Intertemporal Price Discrimination with Forward-Looking Consumers: Application to the US Market for Console Video-Games." *Quantitative Marketing and Economics*, 5(3): 239–292.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-To-Eat Cereal Industry." *Econometrica*, 69(2): 307–342.

Osborne, Matthew. 2010. "Frequency Versus Depth: How Changing the Temporal Process of Promotions Impacts Demand for a Storable Good." Working Paper.

Pires, Tiago, and Andre Trindade. 2015. "Ex-Post Evaluation of Mergers in the Supermarket Industry." Working Paper.

Roberts, John H., and James M. Lattin. 1997. "Consideration: Review of Research and Prospects for Future Insights." *Journal of Marketing Research*, 34(3): 406–410.

Seiler, Stephan. 2013. "The Impact of Search Costs on Consumer Behavior: A Dynamic Approach." *Quantitative Marketing and Economics*, 11(2): 155–203.

Sovinsky, Michelle. 2008. "Limited Information and Advertising in the US Personal Computer Industry." *Econometrica*, 76(5): 1017–1074.

Villas-Boas, Sofia B. 2007. "Vertical Relationships Between Manufacturers and Retailers: Inference with Limited Data." *Review of Economic Studies*, 74(2): 625–652.

Chapter 2

Retail Mergers and Assortment Repositioning

2.1 Introduction

Supermarkets are the primary channel for sales of food at home. In 2014, consumers in the U.S. spent 650 billion dollars on food groceries, accounting for about 5.5% of their disposable income.[1] Due to this central role in consumers' access to nutrition, the supermarket industry has attracted close attention of government agencies. In the U.S. for example, the Federal Trade Commission challenged supermarket mergers affecting 134 antitrust markets and investigated an additional 19 between 1998 and 2007 (Ellickson 2016).

In the recent years, variety effects of mergers have become an important concern in merger evaluation. The new U.S. Horizontal Merger Guidelines (2010), issued by the Department of Justice and the Federal Trade Commission, explicitly state that the authorities should focus not only on price effects of mergers but also on variety effects.[2] This dimension is particularly relevant in the supermarket industry in which assortment – more than price – drives consumer store choice decisions (e.g. Hoch, Bradlow and Wansink 1999; Fox, Montgomery and Lodish 2004; Pan and Zinkhan 2006; Briesch, Chintagunta and Fox 2009).

From a theoretical standpoint, the variety effect of a merger is ambiguous (Berry and Waldfogel 2001). A merger may increase variety due to cost efficiencies from economies of scale and scope (Ellickson 2007), improved bargaining power towards the suppliers (Horn and Wolinsky 1988; Davis 2010; Gowrisankaran, Nevo and Town 2014), because merging firms move away from each other in the product space in order to avoid cannibalization (Gandhi et al. 2008; Draganska, Mazzeo and Seim 2009; Sweeting 2010), or because merging firms face more customers such that previously marginal products may become profitable. At the same time, variety may go down because newly merged firms internalize negative externalities from introducing additional products (Mankiw and Whinston 1986), strategically delist products in order to enhance their bargaining position towards suppliers (Inderst and Shaffer 2007), or reposition their assortment closer to competitors (Sweeting 2010). In the end,

[1] `www.fmi.org/research-resources/supermarket-facts`. Last accessed on 1 February 2017.

[2] "The Agencies also consider whether a merger is likely to give the merged firm an incentive to cease offering one of the relevant products sold by the merging parties. [...] If the merged firm would withdraw a product that a significant number of customers strongly prefer to those products that would remain available, this can constitute a harm to customers over and above any effects on the price or quality of any given product." (U.S. Horizontal Merger Guidelines, 2010, page 24)

the variety effect of a merger is an empirical question and depends on the demand and supply conditions in a market.

The literature has mostly studied variety effects in media markets (Berry and Waldfogel 2001; George 2007; Sweeting 2010; Fan 2013; Jeziorski 2014) but has not paid much attention to retailing markets, with a particular lack of work on grocery retailing. We are aware of only two papers: Both use *ex-post* merger evaluations but find very different results. Pires and Trindade (2015) study a wave of local mergers in the U.S. between 2003 and 2005 and find that store-level assortment size increases by 1% whereas Argentesi et al. (2016) find that, after a national merger in the Netherlands, it decreases by 4.7% . We add to this literature by estimating the merger effect not only on store-level assortment but also on assortment *repositioning*. Specifically, we study a series of local supermarket mergers that occurred in the U.S. in 2010.

We make three contributions to the literature: Firstly, we study how mergers affect assortment similarity between stores. Doing so, we add to a growing literature that documents how a merger may give firms an incentive to reposition themselves relative to their rivals (Berry and Waldfogel 2001; Draganska, Mazzeo and Seim 2009; Sweeting 2010; Fan 2013), for example in order to avoid cannibalization or to steal business. Following the work of Hwang, Bronnenberg and Thomadsen (2010), we construct a variable of similarity between pairs of stores, based on which products they carry and how often these products are available. We find evidence of strategic assortment repositioning: On average, a merger increased the overlap between assortments of merging firms and their competitors by about 3%, suggesting that gains from business stealing exceed losses from cannibalization. We do not find evidence that mergers affect substitutability within pairs of merging parties or pairs of competitors.

Our second contribution is to investigate how mergers affect product variety in the entirety of a local market. The previous literature focuses on store-level assortment alone. This is not sufficient to make statements about consumer access to variety because consumers do not shop at only one store. Instead, they tend to switch between the stores in their neighborhood (Rhee and Bell 2002; Cleeren et al. 2010), i.e. more than store assortment size, the market-level product count is an approximation of actual consumer choice sets. The results from our difference-in-differences estimation suggest that mergers increased the total number of available products in a market by 6.9% and, thus, improved consumer access to variety.

Our third contribution lies in our market definition. The previous literature mostly uses predefined administrative areas, e.g. metropolitan areas or counties. This approach does not take into account that cities and metropolitan areas are further clustered into neighborhoods and that consumers tend to shop within or close to their neighborhoods (Pinkse, Slade and Brett 2002; Eizenberg, Lach and Yiftach 2015). In contrast, we cluster stores based on their geographical location. This results in local markets that a) can span administrative borders, b) cluster cities into neighborhoods, and c) vary in size depending on the density of competitors, thus reflecting how consumer willingness to travel differs, e.g. between rural and urban areas.

The remainder of the paper is structured as follows: We give an overview of the literature in Section 2.2. In Section 2.3, we present the data. We describe the estimation and the identification strategy in Section 2.4. We discuss the results in Section 2.5 and conclude in Section 2.6.

2.2 Literature

We contribute to the literature on *ex-post* merger evaluation, i.e. studies that use data on past mergers to directly estimate merger effects.[3] These studies are typically case studies and focus on price effects of mergers, such as mergers between hospitals (Vita and Sacher 2001), in banking (Sapienza 2002; Focarelli and Panetta 2003), in the airline industry (Borenstein 1990; Kim and Singal 1993; Peters 2006), in media and publishing (Chandra and Collard-Wexler 2009; Aguzzoni et al. 2016), in the consumer goods industry (Ashenfelter and Hosken 2010; Weinberg and Hosken 2013; Friberg and Romahn 2015), and in grocery retailing (Hosken, Olson and Smith 2015; Allain et al. 2013; Argentesi et al. 2016).

Within the literature on retrospective merger studies, a small but growing strand focuses on variety effects. One of the first studies of variety effects is by Berry and Waldfogel (2001) who analyze a merger wave in the U.S. radio industry after the Telecommunications Act of 1996. They find that mergers between radio stations increased the number of programming formats. Further research finds that merging

[3]For an overview of retrospective merger analysis see Hunter, Leonard and Olley (2008).

stations reduced their playlist overlap and simultaneously positioned their playlists closer to their competitors (Sweeting 2010; Jeziorski 2014). George (2007) uses data from the U.S. newspaper industry to study how a merger wave in the early 1990s affected variety. She finds that a decrease in the number of owners led to an increase in the number of topics. Ashenfelter, Hosken and Weinberg (2013) study the merger between the home appliance manufacturers Whirlpool and Maytag, and find that the number of products in the market decreased.

Given the relevance of grocery retailing for the vast majority of households, it is surprising that few studies to date have estimated the variety effect of supermarket mergers. We are aware of only two papers: Pires and Trindade (2015) look at a series of U.S. supermarket mergers between 2003 and 2005, and study its effects on the beverage category. They find that mergers did not affect prices but increased the number of offered products by 1%. A shortcoming of this study is that it uses assortment data from only one product category. In contrast, Argentesi et al. (2016) use data on stores' entire assortment. They investigate a national supermarket merger in the Netherlands and find that the merger did not affect overall prices but decreased product variety by 4.7%. Both Argentesi et al. and Pires and Trindade look at store-level assortment but ignore the fact that stores may strategically adjust overlaps with rivals' assortments. This is important because the two effects may offset each other: Even with an increase in store-level assortments, the overall number of available products in a market can decrease if store assortments become very similar. We fill this gap in the literature by investigating how mergers affect substitutability between the assortments of stores and how this affects overall consumer access to variety.

Lastly, our paper is related to the literature on endogenous product choice. Papers in this literature typically use structural models to quantify welfare effects of mergers, i.e. they model firms' optimal product choice and simulate how firms reoptimize their assortment after a change in market structure. For example, Fan (2013) simulates a merger between two Minneapolis newspapers that was blocked by the Department of Justice. She finds that after the merger, both newspapers would have reduced content variety and content quality. Draganska, Mazzeo and Seim (2009) simulate a merger between two national ice-cream manufacturers and find that the number of offered flavors would decrease after the merger.

2.3 Data

2.3.1 IRI Store Panel

We use weekly store-level scanner data from the U.S., provided by the market research company IRI. The data spans the years 2008-2011 and includes an unbalanced sample of stores, 1,195 of which participate in all years.[4] The stores in the sample are grocery stores with an annual total sales value of at least USD 2 million, as opposed to drug stores (e.g. CVS or Walgreens) and mass retailers (e.g. Walmart or Costco). We observe weekly sales quantities and prices of all products in 27 different product categories (see category list in Table A.2). We also observe brand names and variety names of all products. In the carbonated beverage category for example, "Coca Cola" would be the brand name and "Classic", "Light", and "Zero" would be product varieties. For each store in the IRI sample, we observe the ZIP code. Most stores are located in the North-East and in the South. The Mid West and West are sparsely covered, with the exception of the West Coast (see Figure A.1).

In order to gather more information on the competitive structure of the local markets in which the IRI stores operate, we collect locations and chain affiliations for a sample of 23,683 U.S. supermarkets (see Figure A.1).[5] The chains in our sample reflect the structure in the U.S. retail landscape: Grocery retailing is a mature industry with well-established players and a high industry concentration. Consequently, about half of the stores belong to the 10 largest chains (see Table A.3). Around a third of the outlets are independent stores, "mom and pop" stores, or belong to small chains with at most three outlets.

2.3.2 Local Markets and Mergers

Retail competition is inherently local. Cities are clustered into neighborhoods, and consumers tend do shop close to their neighborhood (Pinkse, Slade and Brett 2002; Eizenberg, Lach and Yiftach 2015; Ellickson, Grieco and Khvastunov 2016). Distance to the store has been found to be a major – if not the main – determinant of consumer store choice (Hoch, Bradlow and Wansink 1999; Smith 2004; Pan and

[4]Table A.1 shows the number of participating stores in each year.

[5]We collected the addresses from `www.supermarketpage.com/supermarktlist.php` (retrieved on 14 December 2016).

Zinkhan 2006; Briesch, Chintagunta and Fox 2009). It is therefore crucial to carefully define local markets.

Previous research on the competitive structure in the supermarket industry uses administrative units to delineate markets, such as cities (Pires and Trindade 2015), counties (Jia 2008), or market-research units which span whole metropolitan areas (Draganska, Mazzeo and Seim 2009; Hosken, Olson and Smith 2015; Hristakeva 2016). These definitions typically neglect cross-border competitive effects. In addition, geographical units such as counties tend to increase in size from East to West due to a combination of climatic conditions (e.g. more mountain ranges and deserts in the West) and the historical development of colonization.

Allain et al. (2013) define French markets by drawing catchment areas of 20 km around hypermarkets and 10 km around supermarkets. Similar definitions are used by competition authorities, such as the FTC (Ellickson, Grieco and Khvastunov 2016) or the German Cartel Office.[6] While this market definition takes into account that households are willing to travel farther for some stores than for others, it does not allow for catchment areas to vary with the local environment. Specifically, it does not reflect that consumers in a neighborhood with many supermarkets will be less willing to travel 10 km than consumers living in a rural area.

In contrast, we cluster stores into markets based on their location (for a similar approach see Ellickson and Misra (2008)). Specifically, we use k-means clustering, a machine learning technique that is popular in real-world market segmentation. It partitions the data into a pre-specified number of disjoint, exhaustive clusters in such a way that each store belongs to the cluster with the nearest mean coordinates (for a detailed description of the clustering procedure see Appendix 2.9.2). The algorithm produces non-overlapping clusters, i.e. we do not have to worry that markets overlap and contaminate our identification.[7]

It is important to use as many store locations as possible for the clustering procedure. If the number of store locations is small, it is more likely that stores are falsely clustered together. In an extreme example, a store on the East Coast and a store on the West Coast could be clustered together if we had only these two observed store locations. Consequently, we pool IRI stores and self-collected stores for the

[6] `www.bundeskartellamt.de/SharedDocs/Entscheidung/DE/Entscheidungen/Fusionskontrolle/2015/B2-96-14.pdf`. Last accessed on 21 March 2017.

[7] We use a specification with 3,500 clusters. We try different specifications with 2,500, 4,500 and 5,000 clusters and find that our results are robust.

clustering step, resulting in a full sample of 25,205 stores. Once we have clustered the stores, we want to keep only clusters that carry some information about store assortment because we are ultimately interested in product variety. Since we observe assortment only in the IRI sample, we drop all clusters in which we do not observe at least one IRI store.

Our final sample consists of 1,037 clusters which contain 1,522 IRI stores and 7,758 self-collected stores. In the majority (66.4%) of clusters, we only observe one IRI store; the rest contains up to 11 IRI stores per cluster (see Table A.4). Most clusters span relatively small areas. Table A.5 shows summary statistics for the average distance between stores and the corresponding cluster center. There is substantial heterogeneity across states: In New Jersey – the most densely populated state in the U.S. – the average distance to the cluster center is 2.646 km. In contrast, the distance is 16.048 km in New Mexico, the most sparsely populated state in our sample.

Table A.6 summarizes how the clusters are distributed across states: There is substantial variation both in the number of clusters and the number of chains per state. Large or densely populated states such as New Jersey, New York, or California tend to have more clusters. Most clusters contain six to eight stores, but in densely populated areas, clusters can contain considerably more stores, such as up to 50 in the State of New York. The exemplary map in Figure A.2 plots all clusters for the State of Virginia, with cluster agglomerations in metropolitan regions like Richmond, Arlington, Alexandria, or Norfolk. Figure A.3 shows clusters in and around Richmond. We see that clusters tend to grow in size the farther they are from the city center.

In our sample period from 2008 to 2011, we observe mergers only in Spring of 2010. A total of 17 stores in Virginia and New York were subject to a merger. For each merger event, we know the location of the store and the week in which the store changed chain affiliation. We drop two mergers because we do not have assortment data before *and* after the merger. We eliminate another three mergers because the acquiring firm was not active in the merger market previous to the merger, i.e. there was a change in ownership but no real increase in concentration. Our final sample contains six mergers in Virginia and six mergers in New York (see Table 2.1).

Table 2.1: Mergers in Sample Period

County	State	Date	Selling	Acquiring
Chautauqua County	New York	March 22, 2010	Chain A	Chain B
Chautauqua County	New York	March 22, 2010	Chain A	Chain B
Madison County	New York	March 29, 2010	Chain A	Chain B
Oneida County	New York	March 29, 2010	Chain A	Chain B
Onondaga County	New York	March 29, 2010	Chain A	Chain B
Oswego County	New York	March 29, 2010	Chain A	Chain B
Chesterfield County	Virginia	April 26, 2010	Chain C	Chain D
Chesterfield County	Virginia	May 3, 2010	Chain C	Chain D
Dinwiddie County	Virginia	April 19, 2010	Chain C	Chain D
Henrico County	Virginia	April 19, 2010	Chain C	Chain D
Henrico County	Virginia	April 5, 2010	Chain C	Chain D
Richmond City County	Virginia	April 5, 2010	Chain C	Chain D

The third column indicates the first day of the week in which the store changed banners. The fourth and fifth column show the coded names of the selling and acquiring chain. For confidentiality reasons we cannot disclose the identity of the chains.

2.3.3 Product Variety

When constructing variety measures, we face three caveats in our data. Firstly, we do not directly observe product availability but weekly store-level sales. If the data indicates zero sales for a product, it can be either because the product was not available in the store or because the product was available but nobody purchased it. In order to rule out the latter, we consider a product as part of a store's assortment only if it was sold in the store during at least four weeks of a year. Products that were sold less often are likely to be trial products or leftover inventories (Hwang, Bronnenberg and Thomadsen 2010).

Secondly, we observe only 27 product categories, i.e. we do not know the entirety of a store's assortment. We believe that our 27 product categories serve as a good representation of the total assortment: Our sample contains both food- and non-food categories, perishable and storable products (e.g. yogurt vs. diapers), staple and niche categories (e.g. toilet paper vs. photo film), as well as low- and high-priced products (e.g. carbonated beverages vs. razor blades). We believe that – due to the wide range of product categories in our sample – the number of observed products correlates closely with the size of the entire store assortment.

Lastly, we do not have data on the assortment in all U.S. stores. Instead, we observe assortment only in 1,522 IRI stores. The relevance of our results thus depends critically on the representativeness of the IRI sample. When comparing the sample of IRI stores to the sample of self-collected stores, we find that they follow very similar patterns.[8] When the self-collected sample indicates that a chain is the largest chain in a cluster, we observe the assortment of at least one store of this chain in 42.2% of all clusters. In 83.4% of all clusters, we observe assortment in at least one store of either the largest or the second-largest chain. Overall, our findings suggest that the assortments observed in the IRI sample provide a reasonable projection of the true assortments in the corresponding clusters.

In the following, we construct three different measures of product variety. The first measure is store-level assortment size, i.e. the total number of products carried by a store. The second measure is market-level variety, i.e. the total number of products available in the market. The third measure captures assortment similarity between stores. We do not observe how quickly assortments were adjusted after the merger, i.e. parts of 2010 may have seen a slow transition in assortments. In order to have a clean identification, we drop all variety measures for the year 2010. Instead, we are going to look only at the years 2008, 2009, and 2011.

Store Assortment Size We define store-level assortment size as the number of different brand-varieties across product categories in a store. For a given year, each store marks one observation. We have 1,430 observations in 2008, 1,408 observations in 2009, and 1,320 observations in 2011. Tables A.7 to A.9 show that there is substantial heterogeneity in the size of product assortments across categories. For example, highly differentiated categories like breakfast cereals and shampoo tend to have assortments of well over 100 products. In contrast, niche categories like photo film or sugar substitutes rarely comprise more than two dozen products. For some stores, the number of products in a category is zero because the store does not offer this category.

[8] To be precise, we also do not know whether the self-collected sample is representative. However, when we look at Figure A.1, we find that – unlike the IRI sample – the self-collected sample covers even sparsely populated areas, such as in the Midwest. This suggests that, while the self-collected sample may not be perfectly representative, it is a good representation of actual store distribution.

Although changes in store-level assortment size are important to understand retailer strategy, they do not allow us to draw conclusions about whether consumer choice sets increase. This is because consumers do not shop at only one store but frequently switch stores: Figure A.4 shows that, from 2008 to 2011, the large majority of households visits two to fourteen different stores, and only 12.9% of households visit one store exclusively. It is therefore crucial to look at how closely store assortments overlap: If stores increase their assortment size but move their assortments closer to their competitors, consumers may effectively end up with less product variety.

Assortment Similarity In the following, we construct a metric of assortment similarity. Specifically, we use a cosine similarity measure (e.g. Jaffe 1986; Hwang, Bronnenberg and Thomadsen 2010). This measure is convenient because it is easy to interpret: It ranges from zero to one, with a value of one meaning that the stores have identical assortments and a value of zero meaning that they have no overlap at all.

Firstly, we start by constructing assortment vectors which capture how long a product was carried at a given store. Let J be the total number of products in the market. Store i's assortment is a $J \times 1$ vector for which each entry is the share of weeks in which a given product was stocked at store i. For example, assume a market contain three products, i.e. $J = 3$. In this market, store i never stocked brand-variety 1 but stocked brand-variety 2 for 26 weeks and brand-variety 3 all year long. Then the assortment vector is

$$A_i = \begin{pmatrix} 0/52 \\ 26/52 \\ 52/52 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.5 \\ 1 \end{pmatrix}. \tag{2.1}$$

After having constructed assortment vectors for all stores in the sample, we compute the similarity measure for all pair-wise combinations of stores in a given market. More precisely, we compute the cosine of the angle between the two assortment vectors A_i and A_j. It is defined as:

$$S_{ij} = \frac{A_i \cdot A_j}{\parallel A_i \parallel \cdot \parallel A_j \parallel} \tag{2.2}$$

To be able to compute similarity, we require at least two IRI stores per cluster. Hence, we drop all clusters that contain only one IRI store. In our similarity sample, each store pair is one observation in a given year. We observe similarity for 341 pairs in 2008, 347 in 2009, and 349 in 2011. Table A.10 shows summary statistics for our similarity measure.

The average similarity across all years is about 84%. No store pair at any point in time has 100% similarity. However, some stores are at almost 99% similarity. We compute market similarity as the average similarity across all store pairs in a market. Figures A.5 to A.7 show the distribution of market similarity. The distribution shows the same pattern in all three years: Few markets have similarity of less than 70%, most markets have similarity between 70% and 90%, and a small share of markets has high similarity close to 100%. These relatively high levels of similarity can be explained by the existence of "must-have" products, i.e. products with strong consumer loyalty that are carried by almost all supermarkets (e.g. Coca Cola).

Market-Level Product Count We define market-level variety as the total number of different brand-varieties available across the 27 product categories and across all IRI stores in a market. Each observation corresponds to one market. Our final sample contains 834 observations in 2008, 835 observations in 2009, and 838 observations in 2011. Figures A.8 to A.10 show the distribution of market-level product counts. Over our sample period, the average market contains 1,483 products. The markets with the largest variety carry up to 2,482 products whereas markets with little variety carry less than 500 products. It is important to study the market-level product count because it is closely related to product availability in a market. Since consumers switch between stores (Rhee and Bell 2002; Cleeren et al. 2010), the market-level product count serves as a proxy of consumer choice sets and is therefore crucial in the evaluation of consumer welfare.

2.4 Estimation

2.4.1 Difference-in-Differences Estimator

The goal of this study is to determine how grocery retail mergers affect the assortments of supermarkets. What allows us to identify this effect is the co-existence of markets that experienced mergers and markets that did not experience mergers. More specifically, we compute the merger effect by exploiting both time and cross-sectional variation in store assortments. Like many recent studies that estimate the effect of mergers (Hosken, Olson and Smith 2015; Allain et al. 2013; Pires and Trindade 2015; Aguzzoni et al. 2016; Argentesi et al. 2016), we use a difference-in-differences estimator. This estimator identifies the assortment effect of a merger as the change in assortment in a market experiencing the merger minus the average change in assortment in similar comparison markets not experiencing a merger. We estimate the following equation for the merger effect on store-level assortment size:

$$\begin{aligned} \log(AssortSize_{ijt}) = \alpha &+ \theta_1 Post_{jt} + \theta_2 Merger_{jt} + \theta_3 PostMerger_{jt} \\ &+ \beta X_{jt} + \eta_j + \delta_i + \varepsilon_{ijt} \end{aligned} \tag{2.3}$$

where i denotes the store, j denotes the market and t denotes the time. $Post_{jt}$ is one if the year is 2011 and zero otherwise. $Merger_{jt}$ is one if the market experiences a merger and zero otherwise. $PostMerger_{jt}$ is the interaction term between $Post_{jt}$ and $Merger_{jt}$. θ_3 is the parameter we are mainly interested in: It captures the effect of a merger on variety when time, market, and time-market factors are controlled for. We further include time-specific market characteristics X_{jt} (e.g. population density or the number of competitors), store fixed-effects δ_i, and market-fixed effects η_j. ε_{ijt} is a zero-mean i.i.d. idiosyncratic error term.

We similarly estimate the merger effect on the total market-level product count:

$$\begin{aligned} \log(TotalCount_{jt}) = \alpha &+ \theta_1 Post_{jt} + \theta_2 Merger_{jt} + \theta_3 PostMerger_{jt} \\ &+ \beta X_{jt} + \eta_j + \varepsilon_{jt} \end{aligned} \tag{2.4}$$

and on assortment similarity:

$$\log(Simil_{pt}) = \alpha + \theta_1 Post_{pt} + \theta_2 Merger_{pt} + \theta_3 PostMerger_{pt} + \beta X_{pt} + \eta_p + \varepsilon_{pt} \tag{2.5}$$

where p denotes a pair of stores. The rest of the notation remains the same as in Equation 2.3. Note that the unit of observation differs across the three equations: In Equation 2.3, the unit of observation is a store, in Equation 2.4 it is a store pair, and in Equation 2.5 it is a market.

2.4.2 Comparison Markets

It is important to carefully construct a set of comparison markets. In order to correctly identify the effect of a merger, the set of comparison markets has to experience similar demand and supply conditions as the merger market, with the merger being the only exception. If the comparison market experienced a change in market structure, such as firm entry, the differences in variety between the merger market and the comparison market could be attributed to both the merger and firm entry. For the comparison group, we therefore exclude all 195 markets that experience any change in market structure. In the following, we call this set of comparison markets the "broad comparison group".

A crucial requirement for a difference-in-differences approach to be valid is that assortment in the comparison markets closely approximates how assortment would have developed in markets with mergers. This assumption is often referred to as the parallel trend assumption. One reason why this may be violated is the fact that stores do not randomly receive the merger treatment. Instead, firms evaluate a market and then decide whether they want to proceed with an acquisition. Therefore, markets which experience a merger are likely to be inherently different from markets which do not experience a merger. In this case, we cannot identify whether assortment changes are caused by the merger or by the underlying market characteristics that facilitated the merger.

We cannot directly test the parallel trend assumption because we do not observe counterfactual assortments. However, we can test whether assortments of merger markets and comparison markets followed similar trends *before* the merger. We use pre-merger data from 2008 and 2009 to estimate

$$\overline{Y}_{jt} = \gamma_0 + \gamma_1 t + \gamma_2 MergMkt_j + \gamma_3 tMergMkt_j + \varepsilon_{jt}, \tag{2.6}$$

where $\overline{Y}_{jt}$ is market j's average of the assortment measure of interest, i.e. store-level assortment size, assortment similarity, or market-level variety. $MergMkt_j$ is one if market j experiences a merger and zero otherwise, t is zero if the year is 2008 and one if it is 2009. $tMergMkt_j$ is the interaction term of t and $MergMkt_j$, and the associated parameter γ_3 is the parameter in which we are interested: If it is not significantly different from zero, pre-merger time trends are identical for merger markets and comparison markets.

Table A.11 shows the estimated pre-merger trends for store-level assortment size (column 1), market-level product variety (column 2), and assortment similarity (column 3). In all three regressions, the coefficient of the interaction term between the time trend and the merger market indicator is not significantly different from zero. This means that merger markets and comparison markets experienced the same time trends. It suggests that assortment in comparison markets is a good approximation of how assortment would have developed in merger markets if they had not experienced a merger.

Although the previous evidence suggests that the parallel trend assumption is valid in our application, we conduct an additional robustness check. Specifically, we use propensity-score matching (PSM) which matches merger markets with comparison markets based on their pre-merger probability of experiencing a merger. PSM collapses the multiple dimensions in which treated and control markets might differ into a scalar, the so-called propensity score. This score is defined as the probability of experiencing a merger conditional on a set of pre-treatment variables X:

$$p(X) = Pr(Merger|X). \tag{2.7}$$

We use a logistic regression to compute the propensity score. The vector of explanatory variables X includes the median income in a market, population density, the share of the hispanic population, the Herfindahl-Hirschman Index, an index of house and land prices, and the number of firms (see Table A.12). It is rare for two markets to have the exact same propensity score. Therefore, we match merger markets to the control markets with the closest propensity scores (for details on the PSM procedure see Appendix 2.9.5).

2.5 Results

In the following, we present the estimates of the merger effect on store-level assortment size, on assortment overlap between stores, and on market-level product variety. For each estimation, we first perform a graphical analysis and then a difference-in-differences regression. We provide a detailed discussion of our results in Section 4.6.3.

2.5.1 Assortment Size

We first estimate the merger effect on store-level assortment size. Figure A.11 provides graphical evidence for the effect. It shows how the average assortment in a store develops in merger markets as opposed to the average market in the broad comparison group. We find that, while the assortment stays relatively stable in the average comparison market, assortments tend to increase in the merger markets after the merger has taken place. This preliminary finding suggests that mergers increase store-level assortment size. The same finding should also be reflected in a more formal econometric analysis.

In order to obtain a precise estimate of the merger effect, we estimate the difference-in-differences regression in Equation 2.3. Table 2.2 shows the regression results. We find that a merger increases average assortment size by 7.43% (column 1). We separately estimate the effect for merging parties and competitors and find that merging firms (7.47%) increase their assortment more than their rivals (6.94%). The average result is largely driven by merging stores because, due to the relatively small clusters and the limited number of IRI stores, we observe competitors' assortments only in three out of twelve merger markets. We apply propensity-score matching and find that the matching procedure performs well: None of the observable characteristics of treated markets and control markets differ significantly from each other (see Panel I in Table A.13).

Table 2.2: Estimated Merger Effect on Store Assortment Size

Variable	(1) Full	(2) Merging	(3) Competitors
Post	-0.00114	-0.000810	0.000508
	(0.00112)	(0.00112)	(0.00113)
PostMerger	0.0743***	0.0747***	0.0694***
	(0.00690)	(0.00759)	(0.0167)
Constant	7.190***	7.190***	7.189***
	(0.0449)	(0.0450)	(0.0456)
Observations	3,953	3,953	3,953
R-squared	0.974	0.974	0.973

The unit of observation is a store. We control for chain- and market fixed-effects. Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5% and 10% level, respectively.

2.5.2 Assortment Similarity

Next, we estimate the merger effect on assortment similarity between stores. This is important because an increase in assortment size is not necessarily beneficial for consumers: If it goes hand in hand with an increase in the assortment overlap between stores, consumers may face an overall decrease in market-level variety. It is therefore important to understand how firms reposition their assortments in the wake of a merger and how it affects substitutability between stores.

In the following, each unit of observation is a pair of stores. Because we do not have assortment data for all store pairs and years, we include only store pairs for which we have at least one year of data before and after the merger. Figure A.12 presents graphical evidence of how mergers affect assortment overlap: It compares assortment similarity in treated store pairs to assortment similarity in the average untreated store pair, and shows a weak tendency of a post-merger similarity increase.

Table 2.3 displays the results from the difference-in-differences estimation. We use only the broad comparison group because the PSM procedure performs poorly, resulting in significant differences in four out of eight observed characteristics between treated and control store pairs (see Panel II in Table A.13). Our results show that a merger induces store pairs to reposition their assortments relative to each other such that their assortment overlap increases by 2.4% (column 1). This effect is driven by pairs consisting of one merging party and one competitor (column 4). There is no significant change in similarity for pairs of merging parties (column 2) or pairs of competitors (column 3).

Table 2.3: Estimated Merger Effect on Assortment Similarity

Variable	(1) Full	(2) MM	(3) CC	(4) MC
Post	-0.00935***	-0.00897***	-0.00895***	-0.00937***
	(0.000871)	(0.000872)	(0.000872)	(0.000867)
PostMerger	0.0241***	-0.0129	-0.0172	0.0305***
	(0.00724)	(0.0187)	(0.0162)	(0.00782)
Constant	-0.0369***	-0.0370***	-0.0370***	-0.0368***
	(0.00816)	(0.00822)	(0.00822)	(0.00814)
Observations	1,037	1,037	1,037	1,037
R-squared	0.989	0.989	0.989	0.989

The unit of observation is a store pair. MM, CC, and MC denote whether the store pair consists of two merging parties (MM), two competitors (CC), or one merging party and one competitor (MC). We control for chain- and market fixed-effects. Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5%, and 10% level, respectively.

2.5.3 Market-Level Product Count

In order to determine how a) an increase in store-level assortment size and b) an increase in assortment overlap affect *market-level* product availability, we estimate the merger effect on the market-level product count. Figure A.13 provides preliminary graphical evidence and compares the total product count of merger markets with that of the average control market in the broad comparison group. We find that in three out of twelve cases, the number of available products in the market increased.

Table 2.4 shows the results from the corresponding difference-in-differences estimation. We find that, in the average market, a merger increases the number of available products by 6.9%. The propensity-score matching procedure performs well: No observed characteristics, except land price, differ significantly between treated markets and control markets (see Panel III in Table A.13). In the next section, we provide a discussion of the results presented in Sections 2.5.1 to 2.5.3.

Table 2.4: Estimated Merger Effect on Total Product Count in Market

Variable	(1)
Post	-0.00104
	(0.00173)
PostMerger	0.0688***
	(0.0140)
Constant	7.242***
	(0.0416)
Observations	2,379
R-squared	0.977

The unit of observation is a market. We control for chain- and market fixed-effects. Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5%, and 10% level, respectively.

2.5.4 Discussion

We study the merger effect on store assortment size and find that the average assortment increases by 7.4%. This finding is of the same order of magnitude as that by Pires and Trindade (2015). They use U.S. supermarket data for the beverage category from 2003 to 2005 and find that, on average, stores increased assortment size by 1% after a merger. On the other hand, Argentesi et al. (2016) find that a national merger in the Netherlands led the merging parties to *decrease* their assortments by 4.7%.

These contrasting results may be reconciled by the fact that the U.S. and the Dutch market differ strongly in supply and demand conditions. Firstly, Argentesi et al. study national mergers whereas we and Pires and Trindade study local mergers. This has implications on cost efficiencies: Before the merger, the areas of business activity of the merging firms in our sample overlapped at most partially. Consequently, the merger allowed them to acquire distribution networks in previously not served areas. In the retailing industry, where distribution constitutes a large share of the costs, this is likely to yield substantial cost efficiencies. Contrarily, each of the merging parties of the national merger studied in Argentesi et al. already had a functioning national distribution network. The efficiency gains are therefore likely to be smaller than in local mergers, and as a result allow less leeway for assortment expansion. Secondly, consumer demand differs between the U.S. and the Netherlands. Since population density in the U.S. is less than 9% of the population density in the Netherlands,[9] U.S. stores have traditionally competed more in the size dimension, with a median store size of 42,800 square feet and a strong consumer preference for the supercenter format.[10] Consequently, U.S. stores may experience more competitive pressure to offer variety than their Dutch counterparts.

In general, our findings suggest that consolidated chains in the U.S. invest their efficiency gains in assortment expansion whereas Dutch supermarkets face natural space restrictions and therefore have to invest in other features, e.g. advertising, opening hours or store design. The existence of efficiency gains from supermarket mergers has been strongly suggested by the literature, for example through economies of scale and scope (Ellickson 2007) or through an improved bargaining position towards suppliers (Davis 2010). Other supermarket studies that find results consistent with efficiency gains include Lamm (1981) and Cotterill (1986).[11]

We find that not only merging firms but also competitors increase their assortments after a merger, albeit slightly less than merging parties. Ellickson (2007) finds that supermarkets generally have an incentive to invest their efficiency gains in variety-enhancing distribution because large variety allows supermarkets to form natural oligopolies. In a similar vein, Matsa (2011) studies U.S. supermarket data from 1988 through 2004 and finds that stores that face more intense competition are more careful to avoid stock-outs. This supports our finding that rivals feel competitive pressure to extend their assortments if they face a merged firm able to invest its efficiency gains into assortment expansion.

Stores do not only increase their assortments but they do it in a way such that their assortments become more similar: A merger increases the overlap between merging firms' and competitors' assortments by 3%. This indicates that it is profitable to steal business from competitors by positioning own assortment closer to rivals' assortments. It is also in line with the findings of Sweeting (2010) who studies

[9]The average population density is 85 inhabitants/mi^2 in the U.S. and 1,067 inhabitants/mi^2 in the Netherlands.

[10]`www.fmi.org/research-resources/supermarket-facts`. Last accessed on 1 February 2017.

[11]For a review of the literature on efficiency gains see Röller, Stennek and Verboven (2006).

consolidation in the U.S. radio industry and finds that merging stations moved their programming closer to their competitors.

However, unlike Sweeting (2010), we do not find that merging parties become less similar in their assortments. Multiple forces drive similarity in merging parties' assortment: On the one hand, repositioning assortments away from each other can avoid cannibalization. On the other hand, economies of scale and scope give merging firms an incentive to align their assortments, and newly acquired stores will adjust their assortments along the assortment reputation of the chain banner (Hwang, Bronnenberg and Thomadsen 2010), e.g. with a focus on organic products, private labels, or regional brands. In our case, it is likely that all of these effects are in play but offset each other such that there is no statistically significant change in the overall similarity between merging firms.

One of the main findings of this paper is that mergers seem to have improved consumer access to variety. This is the result of two opposing effects: While store-level assortment size increases, the overlap between store assortments also increases. The former effect is dominant, leading to a market-level variety increase by 6.9%. We are the first to quantify merger effects on assortment similarity and market-level product count whereas the previous literature studies effects on store-level assortment size only.

It is difficult to make statements about whether the increase in consumer choice sets is welfare-enhancing because post-merger variety may be excessive. The introduction of products is closely related to the literature on firm entry. This literature shows that entry may lead to inefficient product proliferation because entrant firms incur substantial fixed costs and offer close substitutes to each other, ignoring negative externalities on rivals and ultimately only stealing business from each other (e.g. Mankiw and Whinston 1986; Klemperer and Padilla 1997; Berry and Waldfogel 1999). Unfortunately, without additional information it is not possible to assess whether variety is inefficiently large in a market.

Lastly, variety may also be excessive from a consumer point of view. The marketing and psychology literature finds that if the number of product alternatives exceeds a certain threshold, consumers may be less likely to make a purchase (Iyengar and Lepper 2000), find it more difficult to make a choice (Kuksov and Villas-Boas 2010), and end up less satisfied with the choice they made (Chernev 2003). In this case, retailers can increase sales by decreasing variety (Boatwright and Nunes 2001).

Nonetheless, excessive variety from the standpoint of the consumer is a secondary concern because, in the long term, stores should adjust to a level of variety that is most liked by consumers.

2.6 Conclusion

We examine a series of local supermarket mergers in the U.S. and retrospectively evaluate the effects on store assortment. We find that supermarkets obtain cost efficiencies through mergers and invest these gains in variety-enhancing distribution in order to improve their market position. This in turn exerts competitive pressure on rival supermarkets to also invest in assortment expansion. At the same time, merging parties and competitors reposition their assortments closer to each other, suggesting that gains from business-stealing exceed losses from cannibalization. Overall, we find that mergers increased the total number of products available to consumers in a market by almost 7%, thus improving consumer access to variety.

Our results provide support for the concern of the U.S. Department of Justice and the Federal Trade Commission who updated their Horizontal Merger Guidelines to emphasize that it is important to consider not only price effects but also variety effects when deciding whether to approve a merger and, if so, under which conditions. This is particularly relevant in industries in which firms compete less in prices and more in assortment size and assortment composition, a prime example being grocery retailing.

We also propose a new way to define markets that may be useful in antitrust policy. We adopt a clustering mechanism from machine learning that uses stores' geographical locations to group close-by stores into local markets. This market definition departs from the more traditional way of defining markets by administrative units or travel distance, and accounts for the fact that willingness-to-travel varies across regions depending on urban development. Our market definition yields highly flexible markets that vary in size and density, may span administrative borders, and cluster the urban landscape into neighborhoods. Doing so, it may help competition authorities to obtain a more accurate picture of local competition.

Finally, it would be interesting to explore in future research how mergers affect assortment positioning in the *characteristics* space. The existing literature typically defines variety as the number of available products in a store or market, i.e. an assortment of n very similar products is of the same variety as an assortment of n strongly differentiated products. However, consumers generally value differentiation because it allows them to find better matches. One could adopt an attribute-based approach following Boatwright and Nunes (2001) in order to capture how close the products of an assortment are located to each other in the characteristics space. The effect of mergers on the within-store level of differentiation would provide a more detailed picture of how stores reposition their assortment after a merger and how this affects consumer welfare.

2.7 Acknowledgement

I am grateful for comments and advice from Tomaso Duso, Germain Gaudin, and Mattia Nardotto. I also received helpful suggestions from Jana Friedrichsen, Mia Lu, and seminar participants at DIW Berlin. I thank Mike Kruger from IRI for his help with understanding the data, and Ivan Mitkov for excellent research assistance.

Bibliography

Aguzzoni, Luca, Elena Argentesi, Lorenzo Ciari, Tomaso Duso, and Massimo Tognoni. 2016. "Ex Post Merger Evaluation in the UK Retail Market for Books." *Journal of Industrial Economics*, 64(1): 170–200.

Allain, Marie-Laure, Claire Chambolle, Stephane Turolla, and Sofia Villas-Boas. 2013. "The Impact of Retail Mergers on Food Prices: Evidence from France." Working Paper.

Argentesi, Elena, Paolo Buccirossi, Roberto Cervone, Tomaso Duso, and Alessia Marrazzo. 2016. "The Effect of Supermarket Mergers on Variety." Working Paper.

Ashenfelter, Orley C., and Daniel S. Hosken. 2010. "The Effect of Mergers on Consumer Prices: Evidence from Five Mergers on the Enforcement Margin." *Journal of Law and Economics*, 53(3): 417–466.

Ashenfelter, Orley C., Daniel S. Hosken, and Matthew C. Weinberg. 2013. "The Price Effects of a Large Merger of Manufacturers: A Case Study of Maytag-Whirlpool." *American Economic Journal: Economic Policy*, 5(1): 239–261.

Berry, Steven T., and Joel Waldfogel. 1999. "Free Entry and Social Inefficiency in Radio Broadcasting." *RAND Journal of Economics*, 30(3): 397–420.

Berry, Steven T., and Joel Waldfogel. 2001. "Do Mergers Increase Product Variety? Evidence from Radio Broadcasting." *Quarterly Journal of Economics*, 116(3): 1009–1025.

Boatwright, Peter, and Joseph C. Nunes. 2001. "Reducing Assortment: An Attribute-Based Approach." *Journal of Marketing*, 65(3): 50–63.

Borenstein, Severin. 1990. "Airline Mergers, Airport Dominance, and Market Power." *American Economic Review*, 80(2): 400–404.

Briesch, Richard A., Pradeep K. Chintagunta, and Edward J. Fox. 2009. "How Does Assortment Affect Grocery Store Choice?" *Journal of Marketing Research*, 46(2): 176–189.

Chandra, Ambarish, and Allan Collard-Wexler. 2009. "Mergers in Two-Sided Markets: An Application to the Canadian Newspaper Industry." *Journal of Economics & Management Strategy*, 18(4): 1045–1070.

Chernev, Alexander. 2003. "When More Is Less and Less Is More: the Role of Ideal Point Availability and Assortment in Consumer Choice." *Journal of Consumer Research*, 30(2): 170–183.

Cleeren, Kathleen, Frank Verboven, Marnik G. Dekimpe, and Katrijn Gielens. 2010. "Intra- And Interformat Competition Among Discounters and Supermarkets." *Marketing Science*, 29(3): 456–473.

Cotterill, Ronald W. 1986. "Market Power in the Retail Food Industry: Evidence from Vermont." *The Review of Economics and Statistics*, 68(3): 379–386.

Davis, David E. 2010. "Prices, Promotions, and Supermarket Mergers." *Journal of Agricultural & Food Industrial Organization*, 8: 1–25.

Draganska, Michaela, Michael Mazzeo, and Katja Seim. 2009. "Beyond Plain Vanilla: Modeling Joint Product Assortment and Pricing Decisions." *Quantitative Marketing and Economics*, 7(2): 105–146.

Eizenberg, Alon, Saul Lach, and Merav Yiftach. 2015. "Retail Prices in a City: An Empirical Analysis." Working Paper 7.

Ellickson, Paul B. 2007. "Does Sutton Apply to Supermarkets?" *RAND Journal of Economics*, 38(1): 43–59.

Ellickson, Paul B. 2016. "The Evolution of the Supermarket Industry: From A&P to Walmart." In *Handbook on the Economics of Retailing and Distribution.* , ed. Emek Basker. Edward Elgar Publishing.

Ellickson, Paul B., and Sanjog Misra. 2008. "Supermarket Pricing Strategies." *Marketing Science*, 27(5): 811–828.

Ellickson, Paul B., Paul L.E. Grieco, and Oleksii Khvastunov. 2016. "Measuring Competition in Spatial Retail: An Application to Groceries." Working Paper.

Fan, Ying. 2013. "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market." *American Economic Review*, 103(5): 1598–1628.

Focarelli, Dario, and Fabio Panetta. 2003. "Are Mergers Beneficial to Consumers? Evidence from the Market for Bank Deposits." *American Economic Review*, 93(4): 1152–1172.

Fox, Edward J., Alan L. Montgomery, and Leonard M. Lodish. 2004. "Consumer Shopping and Spending Across Retail Formats." *Journal of Business*, 77(2): 25–60.

Friberg, Richard, and André Romahn. 2015. "Divestiture Requirements as a Tool for Competition Policy: A Case from the Swedish Beer Market." *International Journal of Industrial Organization*, 42: 1–18.

Gandhi, Amit, Luke Froeb, Steven Tschantz, and Gregory J. Werden. 2008. "Post-Merger Product Repositioning." *Journal of Industrial Economics*, 56(1): 49–67.

George, Lisa. 2007. "What Is Fit to Print: The Effect of Ownership Concentration on Product Variety in Daily Newspaper Markets." *Information Economics and Policy*, 19(3): 285–303.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town. 2014. "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry." *American Economic Review*, 105(1): 172–203.

Hoch, Stephen J., Eric T. Bradlow, and Brian Wansink. 1999. "The Variety of an Assortment." *Marketing Science*, 18(4): 527–546.

Horn, Henrick, and Asher Wolinsky. 1988. "Bilateral Monopolies and Incentives For Merger." *RAND Journal of Economics*, 19(3): 408–419.

Hosken, Daniel S., Luke Olson, and Loren Smith. 2015. "Do Retail Mergers Affect Competition? Evidence from Grocery Retailing." Working Paper 313.

Hristakeva, Sylvia. 2016. "How Do Vertical Contracts Affect Product Availability? An Empirical Study of the Grocery Industry." Working Paper.

Hunter, Graeme, Gregory K. Leonard, and G. Steven Olley. 2008. "Merger Retrospective Studies: A Review." *Antitrust Bulletin*, 23(1): 34–41.

Hwang, Minha, Bart J. Bronnenberg, and Raphael Thomadsen. 2010. "An Empirical Analysis of Assortment Similarities Across US Supermarkets." *Marketing Science*, 29(5): 858–879.

Inderst, Roman, and Greg Shaffer. 2007. "Retail Mergers, Buyer Power and Product Variety." *Economic Journal*, 117: 45–67.

Iyengar, Sheena S., and Mark R. Lepper. 2000. "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, 79(6): 995–1006.

Jaffe, Adam. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review*, 76(5): 984–1001.

Jeziorski, Przemysaw. 2014. "Effects of Mergers in Two-Sided Markets: The US Radio Industry." *American Economic Journal: Microeconomics*, 6(4): 35–73.

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

Kim, E. Han, and Vijay Singal. 1993. "Mergers and Market Power: Evidence from the Airline Industry." *American Economic Review*, 83(3): 549–569.

Klemperer, Paul, and A. Jorge Padilla. 1997. "Do Firms' Product Lines Include Too Many Varieties?" *RAND Journal of Economics*, 28(3): 472–488.

Kuksov, Dmitri, and J. Miguel Villas-Boas. 2010. "When More Alternatives Lead to Less Choice." *Marketing Science*, 29(3): 507–524.

Lamm, R. McFall. 1981. "Prices and Concentration in the Food Retailing Industry." *The Journal of Industrial Economics*, 30(1): 67–78.

Mankiw, N. Gregory, and Michael D. Whinston. 1986. "Free Entry and Social Inefficiency." *RAND Journal of Economics*, 17(1): 48–58.

Matsa, David A. 2011. "Competition and Product Quality in the Supermarket Industry." *Quarterly Journal of Economics*, 126(3): 1539–1591.

Pan, Yue, and George M. Zinkhan. 2006. "Determinants of Retail Patronage: A Meta-Analytical Perspective." *Journal of Retailing*, 82(3): 229–243.

Peters, Craig. 2006. "Evaluating the Performance of Merger Simulation: Evidence from the US Airline Industry." *Journal of Law and Economics*, 49(2): 627–649.

Pinkse, Joris, Margaret E. Slade, and Craig Brett. 2002. "Spatial Price Competition: A Semiparametric Approach." *Econometrica*, 70(3): 1111–1153.

Pires, Tiago, and Andre Trindade. 2015. "Ex-Post Evaluation of Mergers in the Supermarket Industry." Working Paper.

Rhee, Hongjai, and David R. Bell. 2002. "The Inter-Store Mobility of Supermarket Shoppers." *Journal of Retailing*, 78(4): 225–237.

Röller, Lars-Hendrik, Johan Stennek, and Frank Verboven. 2006. "Efficiency Gains from Mergers." In *European Merger Control: Do We Need an Efficiency Defence?.* , ed. Fabienne Ilzkovitz and Roderick Meiklejohn. Edward Elgar Publishing, Northhampton, MA.

Sapienza, Paola. 2002. "The Effects of Banking Mergers on Loan Contracts." *Journal of Finance*, 57(1): 329–367.

Smith, Howard. 2004. "Supermarket Choice and Supermarket Competition in Market Equilibrium." *Review of Economic Studies*, 71(1): 235–263.

Sweeting, Andrew. 2010. "The Effects of Mergers on Product Positioning: Evidence from the Music Radio Industry." *RAND Journal of Economics*, 41(2): 372–397.

Trindade, Andre. 2015. "Price and Variety in Supermarkets: Can Store Competition Hurt Consumers?" Working Paper.

Ver Ploeg, Michele, Lisa Mancino, Jessica E. Todd, Dawn M. Clay, and Benjamin Scharadin. 2015. "Where Do Americans Usually Shop for Food and

How Do They Travel To Get There? Initial Findings From the National Household Food Acquisition and Purchase Survey." *Economic Information Research Bulletin*, 138.

Vita, Michael G., and Seth Sacher. 2001. "The Competitive Effects of Not-For-Profit Hospital Mergers: A Case Study." *Journal of Industrial Economics*, 49(1): 63–84.

Weinberg, Matthew C., and Daniel S. Hosken. 2013. "Evidence on the Accuracy of Merger Simulations." *Review of Economics and Statistics*, 95(5): 1584–1600.

2.9 Appendix

2.9.1 Summary Statistics: Stores and Local Markets

Table A.1: Descriptives Sample

Number of stores in 2008	1,430
Number of stores in 2009	1,408
Number of stores in 2011	1,320
Number of stores present in all 3 years	1,195

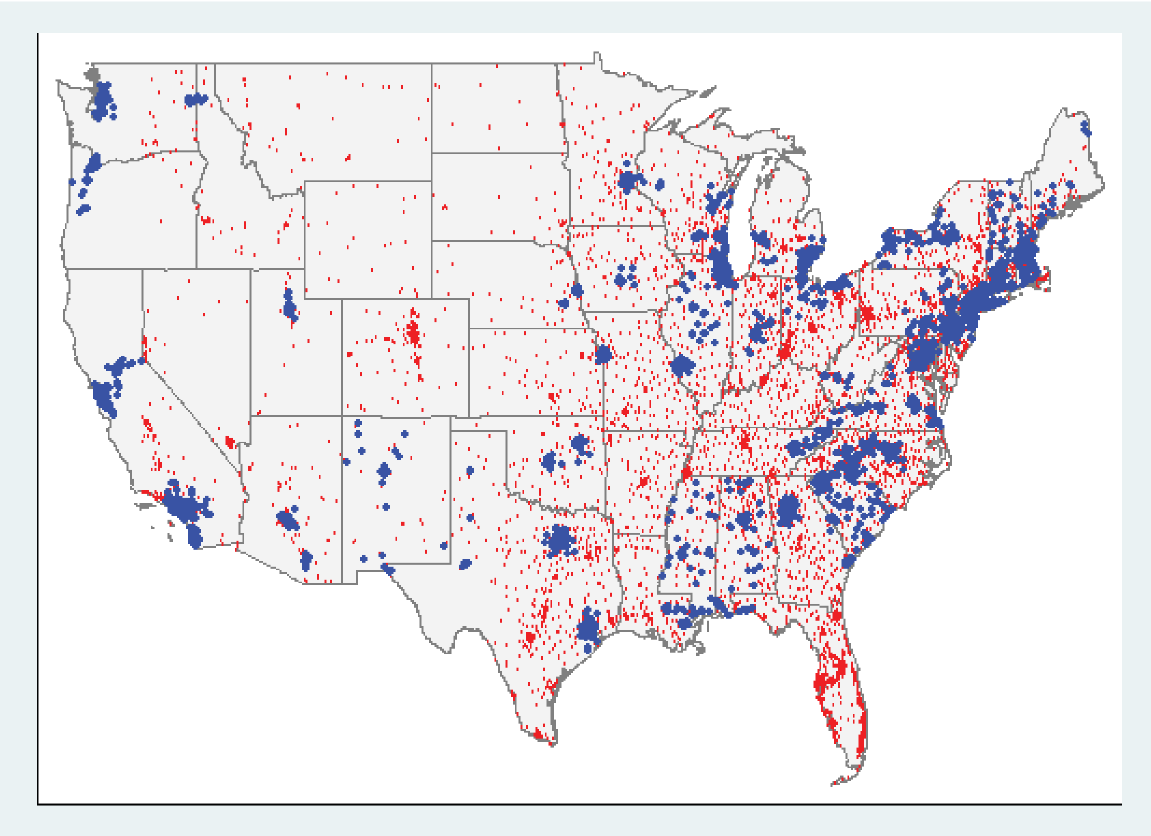
Table A.2: Product Categories

Food	Non-Food
Carbonated Beverages	Cigarettes
Cereal	Deodorants
Coffee	Diapers
Frozen Dinners	Facial Tissues
Frozen Pizza	Household Cleaner
Margarine & Butter	Paper Towels
Mayo	Photo Film
Milk	Razor Blades
Mustard & Ketchup	Shampoo
Peanut Butter	Tooth Brushes
Salty Snacks	Tooth Paste
Soup	Toilet Paper
Spaghetti Sauce	
Sugar Substitutes	
Yogurt	

Table A.3: Self-Collected Stores by Chain

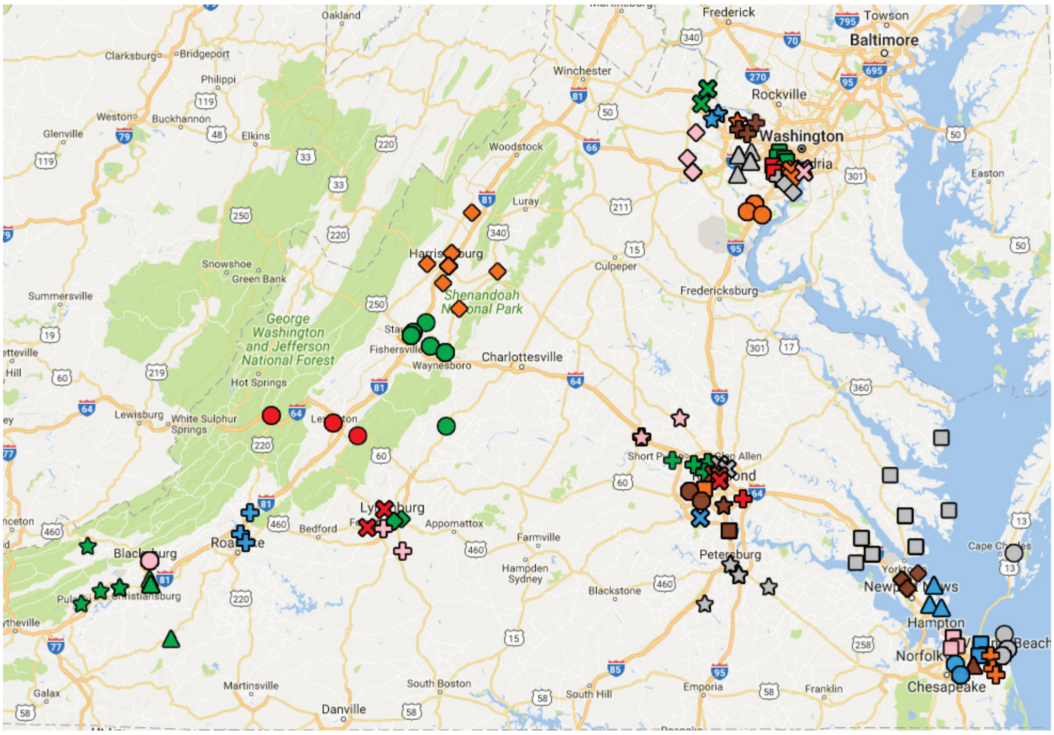
Supermarkets	# Stores	%	Cumulative
Wal-Mart Supercenter	2,592	10.94	10.94
Kroger	1,332	5.62	16.57
Food Lion	1,238	5.23	21.80
Publix	1,144	4.83	26.63
Safeway	1,074	4.53	31.16
Albertsons	979	4.13	35.30
ALDI	856	3.61	38.91
Stop & Shop	663	2.80	41.71
Costco	662	2.80	44.50
Winn Dixie	608	2.57	47.07
Others	12535	52.93	52.93
Total	23683	100	100

Figure A.1: Distribution of Stores in IRI Store Sample.



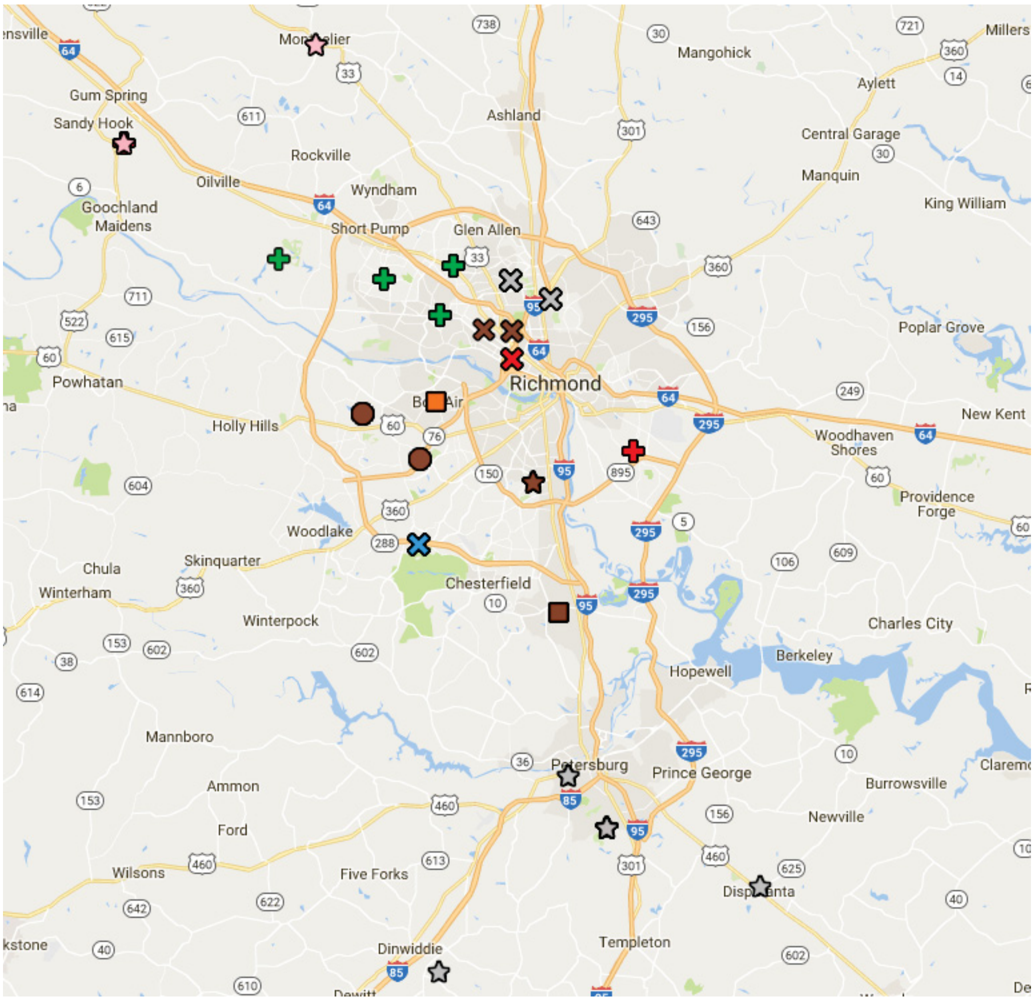
IRI stores are marked in blue, self-collected stores in red.

Figure A.2: Store Clusters in the State of Virginia



Clusters are characterized by a unique combination of symbol (e.g. triangle or star) and color.

Figure A.3: Store Clusters in Richmond, VA



Clusters are characterized by a unique combination of symbol (e.g. triangle or star) and color. The two merging stores in Richmond are marked in red.

Figure A.4: Number of Different Stores Visited by Households 2008-2011

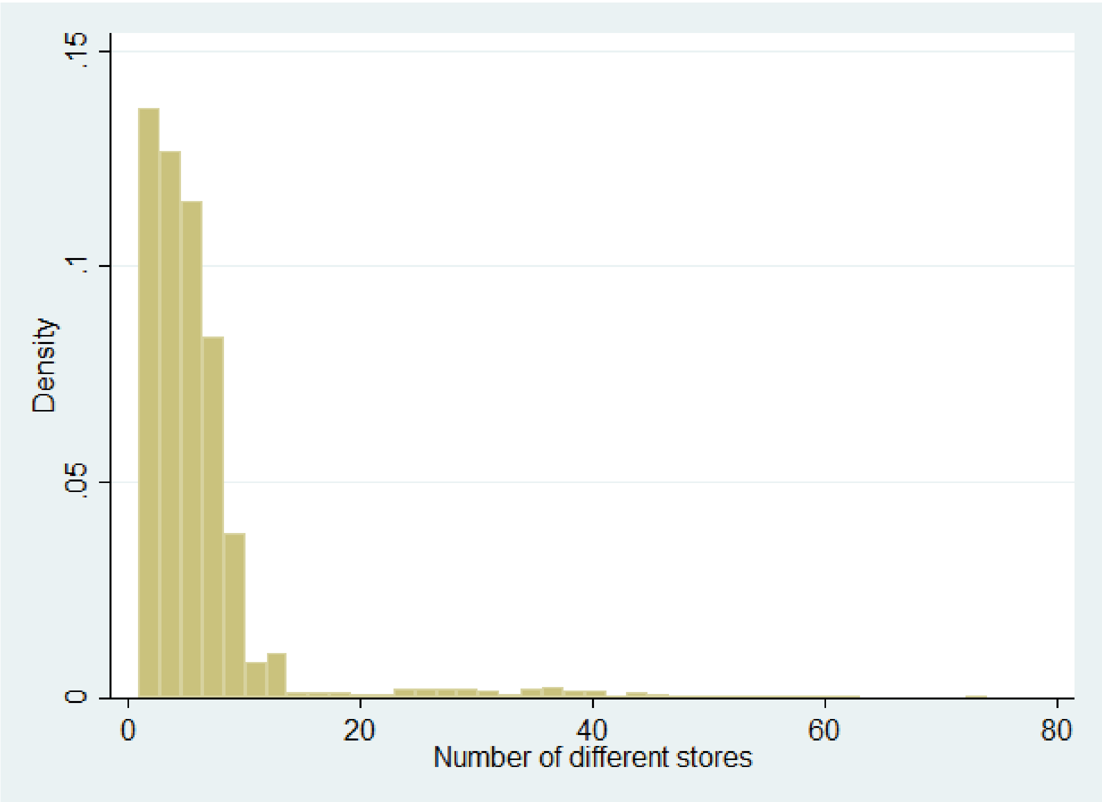


Table A.4: Number of IRI Stores per Cluster

Number of IRI Stores per Cluster	Number of Clusters	Percent	Cumulative
1	689	66.44	66.44
2	255	24.59	91.03
3	56	5.40	96.43
4	27	2.60	99.04
5	5	0.48	99.52
6	2	0.19	99.71
7	1	0.10	99.81
9	1	0.10	99.90
11	1	0.10	100.00
Total	1,037	100.00	

Table A.5: Summary Statistics: Cluster Size I

State	Mean(Distance)	SD(Distance)	Min(Distance)	Max(Distance)
Alabama	11.305	10.949	.742	71.69
Arizona	3.522	5.327	0	78.187
California	3.53	3.815	0	41.819
Connecticut	3.591	2.94	0	15.695
Delaware	1.713	1.394	0	5.687
District of Columbia	1.028	.45	.421	1.444
Florida	2.068	1.981	0	8.398
Georgia	2.244	3.325	0	31.21
Idaho	7.005	3.226	1.746	11.574
Illinois	4.981	6.511	0	45.975
Indiana	4.854	4.988	0	26.83
Iowa	13.757	11.268	1.909	43.221
Kansas	4.852	4.406	1.18	29.735
Louisiana	7.438	8.196	0	34.702
Maine	8.623	11.205	.682	44.019
Maryland	2.508	2.221	0	9.51
Massachusetts	2.892	2.499	0	19.233
Michigan	5.917	6.484	.499	44.919
Minnesota	5.113	2.93	.173	15.456
Mississippi	10.656	7.963	0	39.274
Missouri	3.467	3.358	0	22.054
Nebraska	3.995	4.197	0	26.222
New Hampshire	6.071	5.107	0	23.17
New Jersey	2.646	2.22	0	15.047
New Mexico	16.048	16.908	.95	75.05
New York	3.322	5.115	0	45.058
North Carolina	3.803	3.801	0	24.669
Ohio	6.98	5.627	0	25.469
Oklahoma	5.753	8.281	.501	48.989
Oregon	4.373	4.623	0	31.967
Pennsylvania	3.428	3.556	0	22.866
Rhode Island	3.498	2.435	.556	10.576
South Carolina	5.618	6.402	0	32.884
Tennessee	5.848	4.313	0	22.66
Texas	5.619	8.045	0	63.588
Utah	4.158	2.288	1.128	11.094
Vermont	9.484	8.252	0	37.773
Virginia	4.55	5.886	0	48.258
Washington	3.478	3.342	0	28.005
West Virginia	9.104	6.778	.739	27.812
Wisconsin	9.501	9.102	.337	60.212

Cluster size is measured by the average distance from a store to the cluster center (in km).

Table A.6: Summary Statistics: Cluster Size II

State	N	Mean(n)	SD(n)	Min(n)	Max(n)
Alabama	13	8.765	3.070	4	14
Arizona	32	5.848	2.884	1	12
California	127	6.073	3.730	1	18
Connecticut	16	7.263	3.352	1	12
Delaware	4	9.625	5.452	1	16
District of Columbia	2	7	1.095	6	8
Florida	4	7.4	2.586	4	10
Georgia	27	7.514	4.086	1	15
Idaho	1	17	0	17	17
Illinois	36	7.813	3.634	1	14
Indiana	13	7.944	3.814	1	16
Iowa	5	11.182	4.412	2	14
Kansas	6	5.9	2.006	3	8
Louisiana	6	8	3.194	3	13
Maine	8	6.3	2.879	3	11
Maryland	22	7.030	3.306	2	14
Massachusetts	36	7.5	3.953	2	15
Michigan	20	4.76	2.283	2	10
Minnesota	8	7.917	2.130	3	11
Mississippi	9	6.833	3.359	2	13
Missouri	18	6.806	2.560	2	11
Nebraska	6	7.5	2.850	2	11
New Hampshire	12	4.143	2.160	1	8
New Jersey	50	5.738	2.789	1	18
New Mexico	9	6.111	2.736	1	11
New York	65	6.3026	5.584	1	50
North Carolina	49	8.565	4.135	2	18
Ohio	15	5.263	1	10	
Oklahoma	9	6.44	1.960	3	8
Oregon	20	6.038	3.424	1	14
Pennsylvania	42	7.059	3.875	1	19
Rhode Island	9	5.6	1.831	1	8
South Carolina	35	6.978	4.919	1	19
Tennessee	12	6.052	3.533	1	12
Texas	53	6.553	2.761	1	14
Utah	5	6.167	1.917	3	9
Vermont	7	9.875	7.011	2	21
Virginia	53	7.397	3.892	1	17
Washington	36	7.212	4.539	1	18
West Virginia	2	7	1.5	5	8
Wisconsin	13	5.778	2.866	2	10

N stands for the number of clusters per State, n stands for the number of stores per cluster.

2.9.2 k-Means Clustering

k-means clustering is a method to group a set of n observations x_1, ..., $x_N \in \mathbb{R}^d$ into k cohesive clusters. Its basic idea is to sort each observation into the cluster with the nearest mean. Due to the simplicity of the objective function and the efficiency of the algorithm, it is one of the most popular clustering procedures, and widely used both in research and in real-world applications, e.g. in market segmentation. k-means clustering uses an unsupervised machine learning algorithm, i.e. the algorithm does not receive a sample of correctly clustered data to learn from.

The clustering procedure partitions the N observations into $k < N$ clusters $C = \{C_1, ..., C_k\}$. Specifically, it aims to find k cluster centers μ_1, ..., μ_k such that it minimizes

$$J(C, \mu) = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 , \tag{2.8}$$

where $\|\cdot\|$ denotes the Euclidean norm and $J(C, \mu)$ corresponds to the within-cluster least squares, i.e. the sum of the squares of the distance between each observed store and its corresponding cluster center.

While this problem is computationally difficult, it can be quickly solved with a simple iterative algorithm that is sometimes referred to as Lloyd's algorithm (proposed by Stuart Lloyd in 1957). The algorithm works as follows:

1. Randomly choose μ_1, ..., μ_k.
2. Assign each point x_1, ..., x_N to the cluster center to which it is closest.
3. Recompute cluster centers as the arithmetic middle of the cluster points.
4. Repeat 2. and 3. until the clusters do not change any more.

In other words, the algorithm repeatedly does two steps: It first holds cluster centers μ_1, ..., μ_k fixed and minimizes $J(C, \mu)$ with respect to cluster membership, and then it holds cluster membership fixed and minimizes $J(C, \mu)$ with respect to cluster centers. A a consequence, $J(C, \mu)$ must monotonically decrease and converge.

The clustering result critically depends on which number of clusters k the researcher chooses. The choice of k is directly related to variance in each cluster, i.e. the more clusters we have, the smaller the clusters are and the smaller the distance between the cluster members will be. In the trivial case of $k = N$, each cluster contains only one store. In our application, we choose k such that the size of the clusters corresponds largely to previous findings about how far consumers generally travel to do their shopping (e.g. Ver Ploeg et al. 2015).

2.9.3 Summary Statistics: Assortment

Table A.7: Assortment Size by Category, 2008

Variable	Mean	Std. Dev.	Min.	Max.
Carbonated Beverages	95.336	14.17	0	146
Blades	44.493	10.933	0	63
Cigarettes	58.444	17.258	0	124
Coffee	43.461	11.099	0	84
Cereal	159.769	35.926	52	269
Deodorant	73.855	16.234	0	108
Diapers	19.468	3.209	4	28
Facial Tissues	8.339	2.272	2	18
Frozen Dinner	88.976	15.69	18	131
Frozen Pizza	31.165	7.956	6	57
Household Cleaner	73.783	16.139	23	132
Margarine and Butter	23.462	3.609	8	35
Mayo	14.759	3.67	4	27
Milk	19.401	5.503	0	40
Mustard and Ketchup	26.093	8.033	6	54
Paper Towels	11.392	1.849	5	17
Peanut Butter	14.931	3.504	3	25
Photo Film	14.12	5.631	0	36
Salty Snacks	117.031	24.998	0	202
Shampoo	187.297	60.255	7	374
Soup	73.05	15.736	22	118
Spaghetti Sauce	33.771	13.19	6	77
Sugar Substitute	11.011	3.509	3	24
Toilet Paper	14.664	2.205	7	23
Tooth Brush	107.923	38.608	0	211
Tooth Paste	74.790	17.833	5	107
Yogurt	44.122	9.942	6	78
Total Assortment Size	1353.113	257.955	362	1932
N		1430		

2.9.4 Merger Markets and Comparison Markets

Table A.8: Assortment Size by Category, 2009

Variable	Mean	Std. Dev.	Min.	Max.
Carbonated Beverages	93.13	14.176	0	150
Razor Blades	44.896	12.008	0	67
Cigarettes	60.124	20.655	0	121
Coffee	42.635	10.738	0	96
Cereal	151.089	33.579	0	242
Deodorant	89.714	20.414	0	132
Diapers	22.914	4.181	0	32
Facial Tissues	8.132	2.526	3	16
Frozen Dinner	93.386	15.962	0	131
Frozen Pizza	29.972	7.371	5	57
Household Cleaner	74.824	16.058	18	135
Margarine and Butter	23.438	3.664	8	36
Mayo	14.193	3.41	3	25
Milk	22.117	6.42	0	46
Mustard and Ketchup	25.744	7.464	4	53
Paper Towels	12.097	2.454	4	19
Peanut Butter	15.241	3.346	3	25
Photo Film	10.266	4.466	0	24
Salty Snacks	120.244	25.699	0	222
Shampoo	190.882	61.369	0	387
Soup	72.856	16.072	0	121
Spaghetti Sauce	33.289	12.357	6	71
Sugar Substitute	13.504	3.712	4	27
Toilet Paper	14.107	2.534	3	22
Tooth Brush	103.03	37.552	0	218
Tooth Paste	77.388	19.312	7	114
Yogurt	45.306	10.763	0	80
Total Assortment Size	1370.081	260.928	349	2010
N		1408		

Table A.9: Assortment Size by Category, 2011

Variable	Mean	Std. Dev.	Min.	Max.
Carbonated Beverages	94.998	15.243	46	157
Razor Blades	46.469	13.235	0	68
Cigarettes	69.376	31.459	0	153
Coffee	42.561	10.776	10	75
Cereal	144.227	31.158	60	233
Deodorant	89.47	21.688	2	136
Diapers	24.707	4.903	0	35
Facial Tissues	10.562	2.37	3	18
Frozen Dinner	98.662	16.513	30	139
Frozen Pizza	28.149	7.646	6	61
Household Cleaner	72.699	16.501	23	136
Margarine and Butter	21.964	3.088	9	32
Mayo	12.889	3.342	3	24
Milk	27.689	7.143	6	49
Mustard and Ketchup	24.566	7.275	6	44
Paper Towels	10.202	2.001	2	18
Peanut Butter	16.492	3.942	4	29
Photo Film	4.852	2.964	0	18
Salty Snacks	113.641	24.529	51	214
Shampoo	168.096	59.97	1	334
Soup	72.686	16.308	24	116
Spaghetti Sauce	34.38	12.155	8	69
Sugar Substitute	14.307	4.198	4	29
Toilet Paper	15.574	2.143	5	24
Tooth Brush	93.936	33.842	3	189
Tooth Paste	76.717	18.645	3	112
Yogurt	55.105	12.929	8	99
Total Assortment Size	1358.448	265.013	498	2038
N		1320		

Table A.10: Summary Statistics: Assortment Similarity

Year	Mean	Std. Dev.	Min.	Max.	N
2008	.8390455	.079562	.56144	.98672	341
2009	.8359329	.0819808	.56001	.98538	347
2011	.8302011	.086016	.53315	.98542	349

Figure A.5: Assortment Similarity 2008

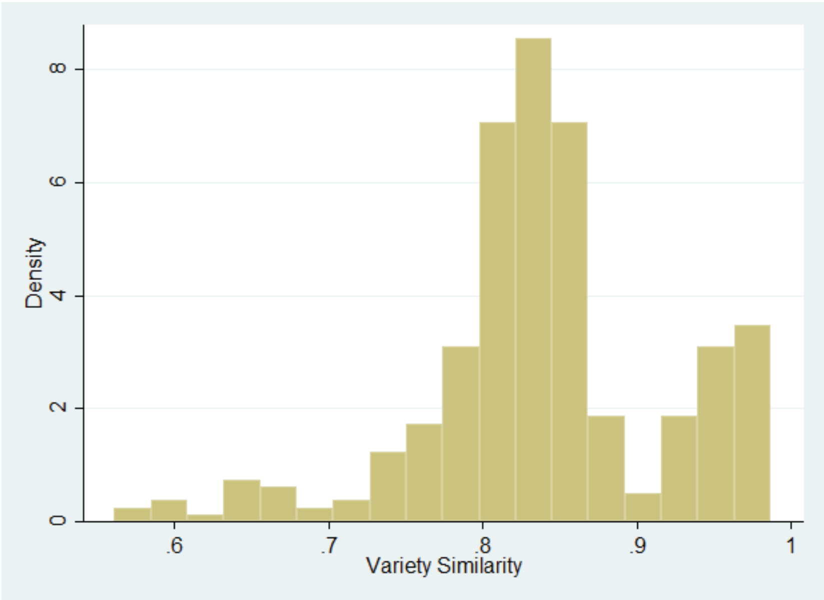


Figure A.6: Assortment Similarity 2009

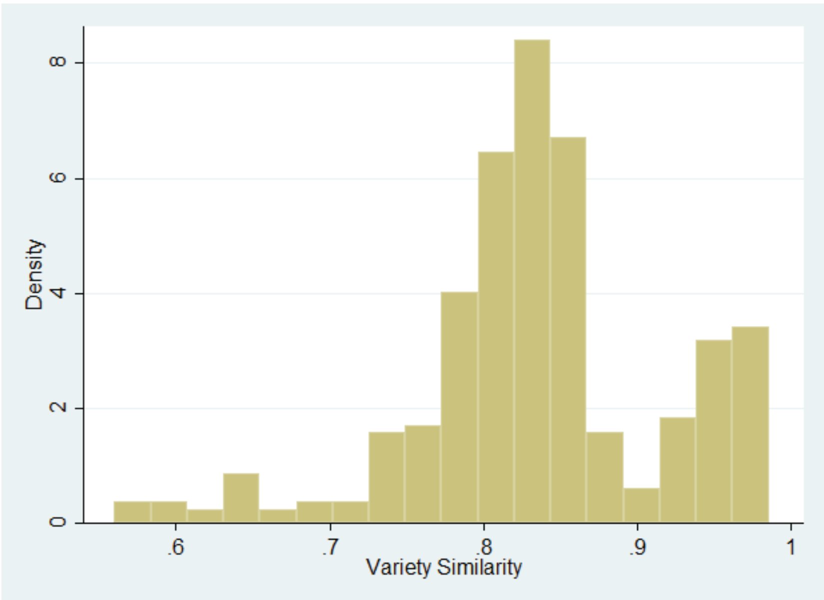


Figure A.7: Assortment Similarity 2011

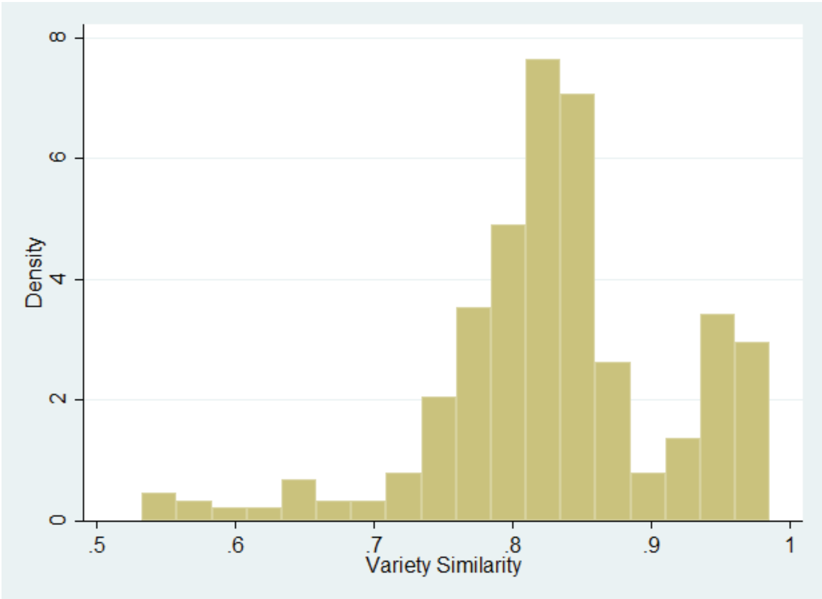


Figure A.8: Total Product Count 2008

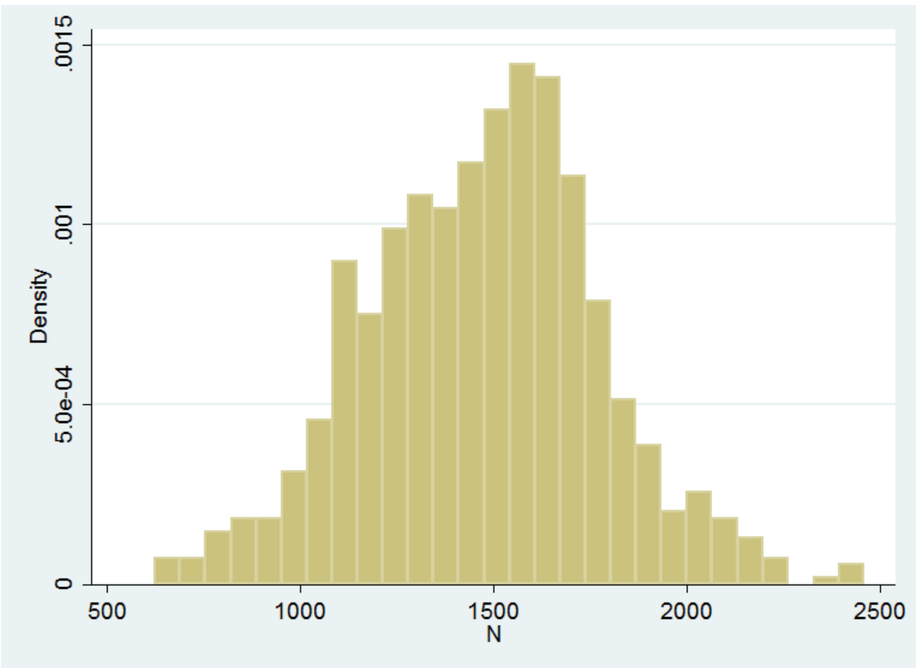


Figure A.9: Total Product Count 2009

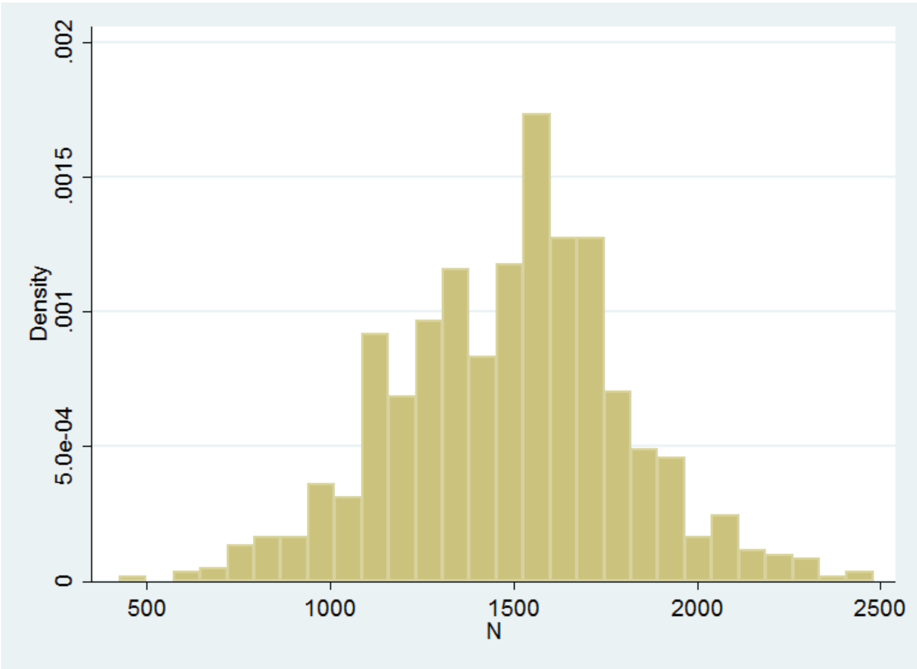


Figure A.10: Total Product Count 2011

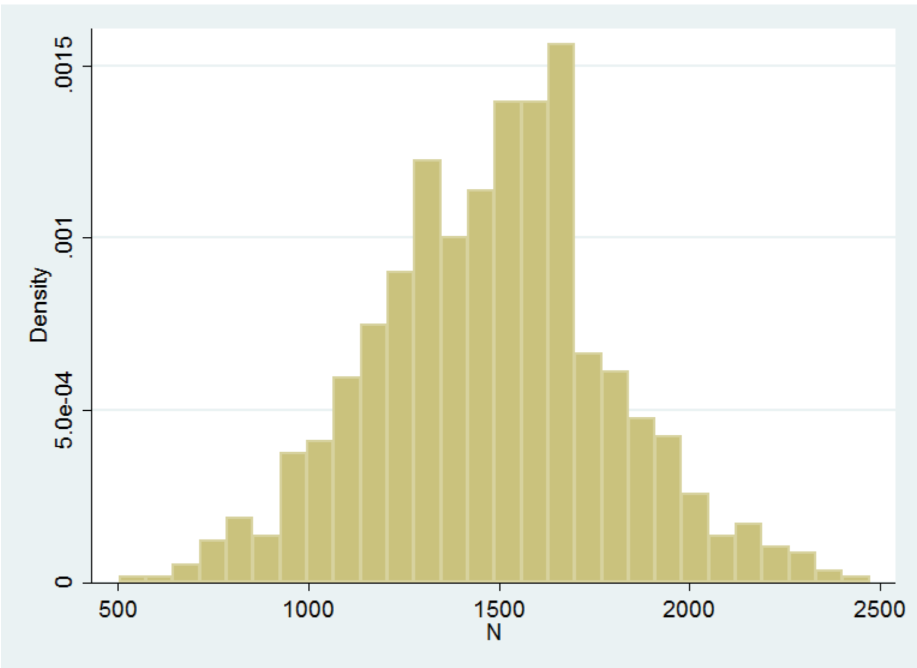
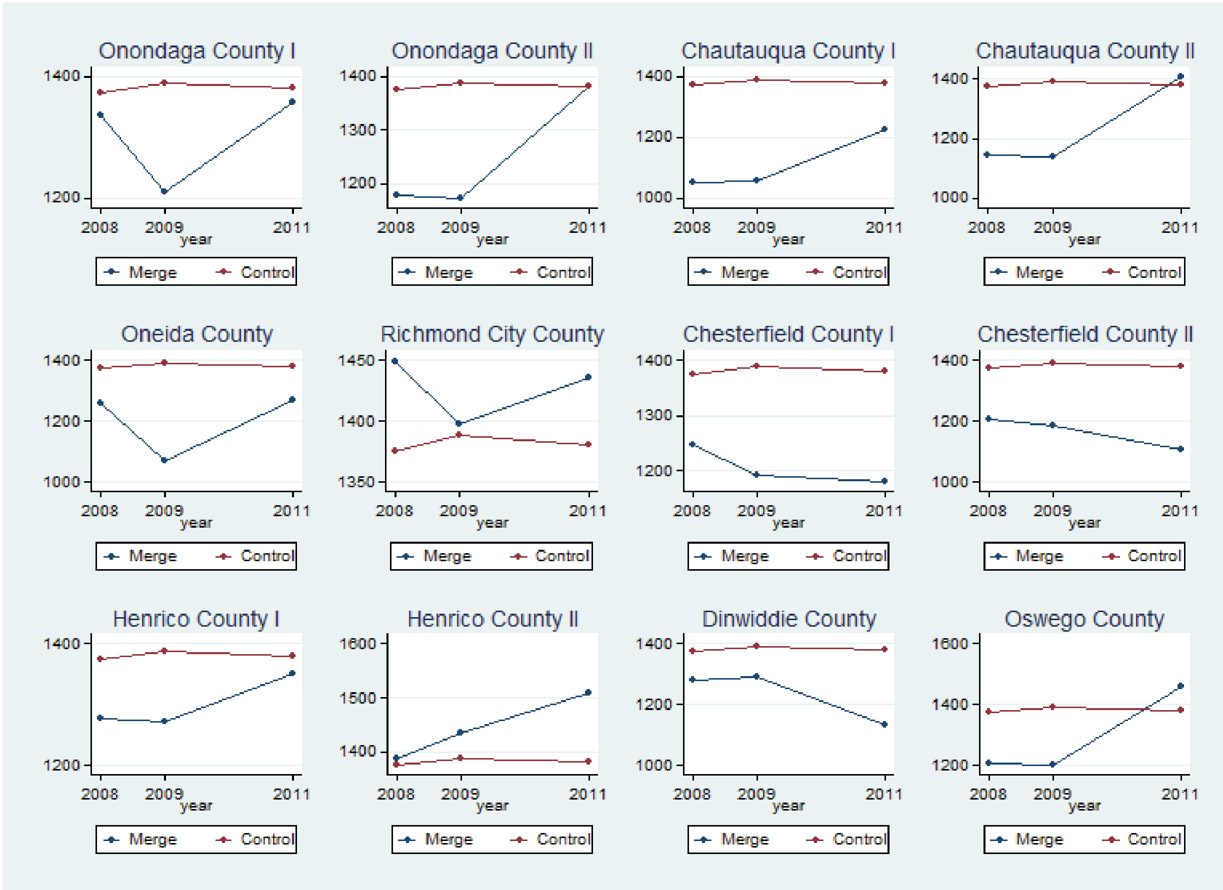


Table A.11: Pre-Merger Trends

Variable	(1) Assortment Size Store Level	(2) Assortment Size Market Level	(3) Assortment Similarity
Merger Market	-0.135***	-0.0566	-0.0782*
	(0.0299)	(0.0392)	(0.0413)
Time	-0.0114	-0.0113	0.00322
	(0.00699)	(0.00916)	(0.00769)
Merger Market × Time	0.0521	0.0196	-0.0104
	(0.0429)	(0.0562)	(0.0823)
Constant	7.395***	7.336***	-0.183***
	(0.00494)	(0.00647)	(0.00543)
R-squared	0.011	0.002	0.008

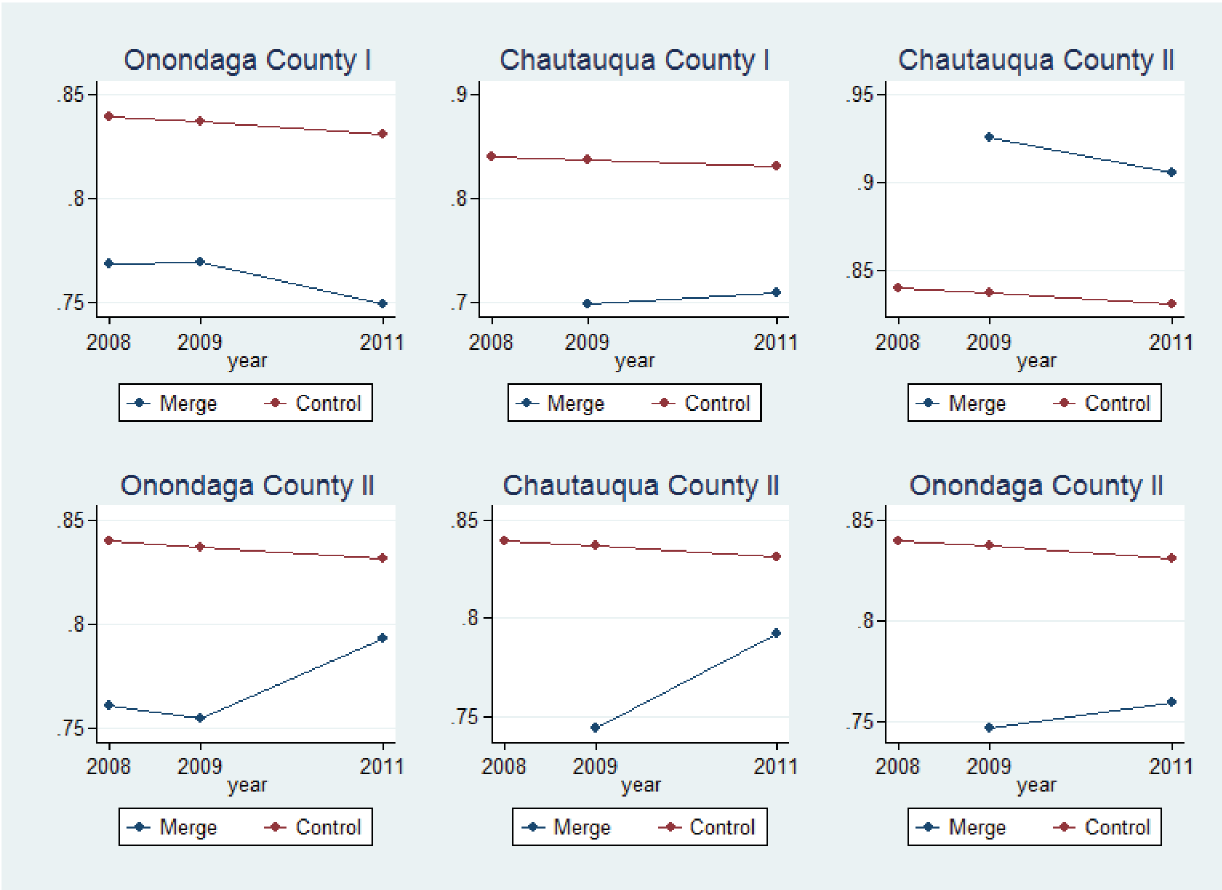
The unit of observation is a market. We control for market fixed-effects. Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5%, and 10% level, respectively.

Figure A.11: Comparison of Store Assortment in Treated and Control Markets



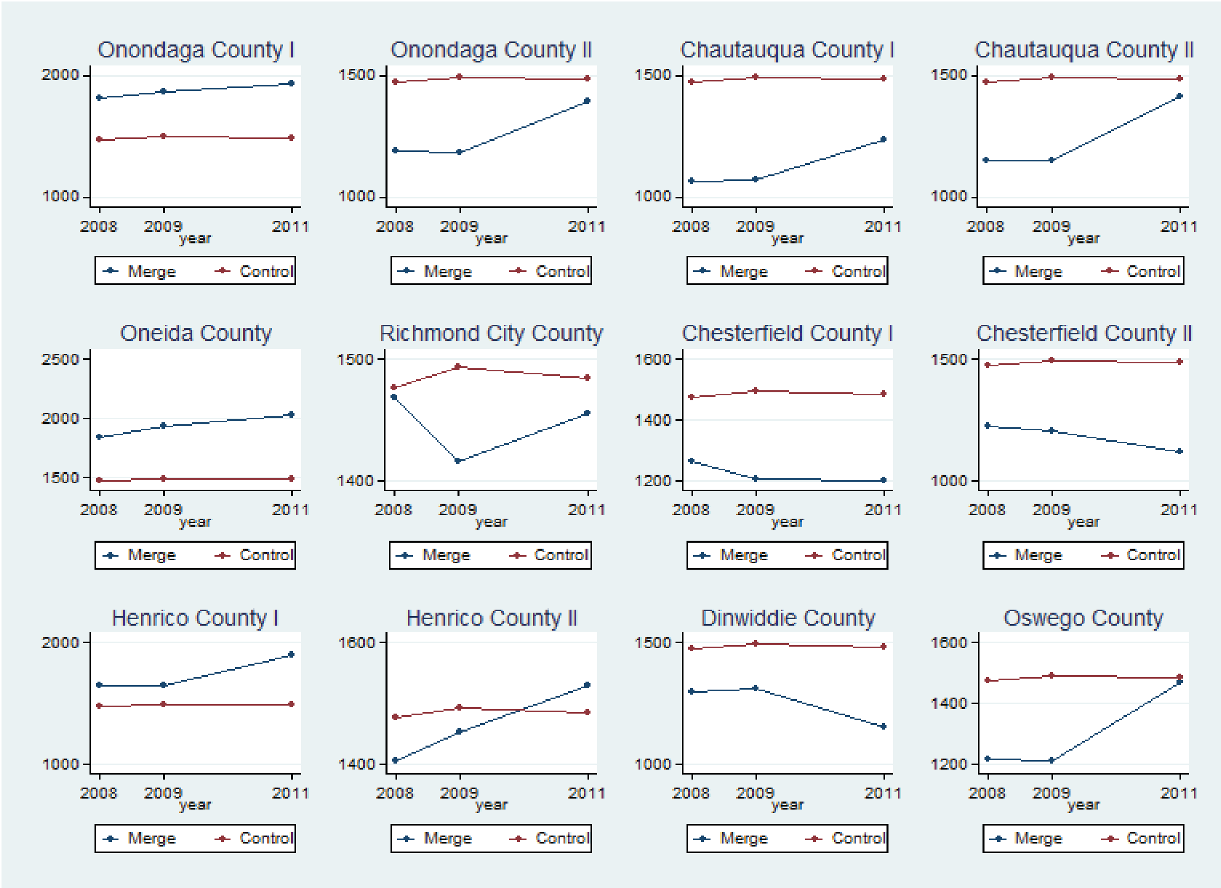
The control is the average comparison market from the broad comparison group.

Figure A.12: Comparison of Store-Pair Similarity in Treated and Control Markets



The control is the average comparison market from the broad comparison group.

Figure A.13: Comparison of Market-Level Product Count in Treated and Control Markets



The control is the average comparison market from the broad comparison group.

2.9.5 Propensity Score Matching

Table A.12 describes the control variables we use to match markets in our PSM procedure. Table A.13 shows how similar the characteristics of matched markets are. We test whether the means of the explanatory variables differ between the treated and the control markets. The null hypothesis is equality in the observed characteristics between the treated markets and the control markets. We find that none of the characteristics differ significantly in the models of store-level assortment size (Panel I) and market-level product count (Panel III). In the case of assortment similarity, we need to match store pairs to each other. PSM does not perform well: The propensity score, population density, the share of Hispanics and land prices differ significantly between merger and comparison markets (Panel II). Therefore we do not use PSM for the estimation of the similarity effect but use the broad comparison group instead.

Table A.12: Description of Control Variables

Control Variable	Description	Frequency	Source
Population Density	Number of inhabitants per square mile	yearly	IRI Data
Hispanics	Share of Hispanics (in %)	yearly	IRI
HHI	Hirschman-Herfindahl Index	yearly	own computations (using IRI data)
Income	Median Household Income	yearly	IRI
House Prices	Average value of owner-occupied houses	quarterly	Lincoln Institute of Land Policy
Land Prices	Home value less the replacement cost	quarterly	Lincoln Institute of Land Policy

Table A.13: Propensity Score Matching Quality

	Mean		t-test	
Variable	Treated	Control	t	P>|t|
I. Store-Level Assortment				
Propensity Score	.05612	.05653	-0.07	0.945
Income	53377	56332	-0.89	0.376
Population Density	1486.4	1366	0.58	0.562
Hispanics	3.0588	2.7636	1.10	0.276
House Prices	1.6964	1.6909	0.28	0.784
Land Prices	1.8991	1.9774	-0.71	0.477
HHI	.28451	.31513	-1.51	0.134
Number of Firms	6.8125	5.9167	1.36	0.176
II. Store-Pair Similarity				
Propensity Score	.54814	.25517	4.20	0.000
Income	53254	54929	-0.54	0.593
Population Density	1673.1	2183.5	-2.17	0.033
Hispanics	4.9746	6.8475	-1.69	0.094
House Prices	1.7379	1.6884	1.55	0.126
Land Prices	2.4956	2.0464	2.93	0.004
HHI	.23474	.23154	0.14	0.889
Number of Firms	17.125	17.35	-0.10	0.923
III. Market-Level Product Count				
Propensity Score	.07379	.07344	0.03	0.976
Income	52117	48079	2 1.31	0.193
Population Density	1488.3	1494.9	-0.03	0.977
Hispanics	3.3103	3.7947	-0.80	0.424
House Prices	1.6984	1.7153	-0.69	0.491
Land Prices	1.8961	2.1146	-2.11	0.038
HHI	.39568	.40235	-0.20	0.842
Number of Firms	8.9231	10.231	-1.07	0.288

Comparison of mean characteristics between treated stores and control stores.

Chapter 3

Consumer Stockpiling and Sales Promotions

3.1 Introduction

The design of promotions is a key concern for both marketing practitioners and marketing research. Since the 1970s, price promotions have become the main marketing instrument in many industries (Currim and Schneider 1991; Blattberg, Briesch and Fox 1995). Promotions are particularly important in grocery retailing. In 2016, they accounted for almost 66% of the marketing budget of consumer packaged goods (Bhardwaj et al. 2016), by far exceeding advertising expenditures.

It is well known that price promotions in storable-goods markets induce consumer stockpiling: When retailers offer promotions, consumers strategically buy large amounts at low prices and store them for future consumption. Previous research developed rich models to describe and estimate such strategic consumer behavior (e.g. Erdem, Imai and Keane 2003; Hendel and Nevo 2006*a*, 2013; Su 2010). What is much less well understood is how firms *should* design promotions in response to consumer stockpiling.

In this paper, we are the first to study the effect of promotion length vs. promotion depth on firms' long-term revenue in a structural framework. To do so, we use rich panel data from the U.S. market for laundry detergent. Detergent can be stored long before and after its first use, it comes in large packs associated with substantial storage cost and, thus, is a typical product in which to study the dynamics of stockpiling (Bell, Iyer and Padmanabhan 2002; Hendel and Nevo 2006*b*; Seiler 2013; Pires 2016).[1] We use a dynamic discrete-choice model of strategic consumer stockpiling. In our model, consumers are forward-looking with rational expectations over future prices. They can choose to stockpile but incur storage costs. We estimate consumer preferences, price sensitivities and storage costs. Using these estimates, we can simulate how changes in promotion depth and length affect consumer purchase decisions and seller revenues.

Our results suggest that in the detergent market, shorter but deeper promotions are preferred over longer, shallower promotions. We find that the revenue elasticity with respect to promotion depth is, *ceteris paribus*, about four times higher than with respect to promotion length. Our findings provide general insights into promo-

[1] There are many other product categories that also have these characteristics, e.g. soda (Hendel and Nevo 2013), ketchup (Pesendorfer 2002; Erdem, Imai and Keane 2003; Sun, Neslin and Srinivasan 2003), coffee (Neslin, Henderson and Quelch 1985), as well as razors and razor blades (Hartmann and Nair 2010).

tion design that can prove helpful for industry practitioners: Compared to longer promotions, deeper promotions tend to perform better in markets with substantial heterogeneity in storage costs, large heterogeneity in price sensitivity, and steady consumption rates.

The rest of the paper is organized as follows: In Section 3.2, we give a brief overview of the literature. In Section 3.3, we describe the data. We lay out the model in Section 3.4 and describe our identification and estimation strategy in Section 3.5. We discuss the estimation results in Section 3.6 and the counterfactual simulations in Section 3.7. Finally, we conclude in Section 3.8.

3.2 Literature

The core contribution of this paper is to quantify and compare the revenue effects of promotion length and promotion depth. In doing so, we add to a large literature on the impact of promotional price cuts. Since the mid-1990s, the economics and marketing literature has used structural models to investigate how sellers should set their prices (Kadiyali 1996; Besanko, Gupta and Jain 1998; Sudhir 2001; Chintagunta 2002; Verboven 2002; Besanko, Dubé and Gupta 2003; Draganska and Jain 2006; Pancras and Sudhir 2007; Richards 2007). The majority of these pricing models are based on static demand models in which consumers are not forward-looking and remain unaware of the fact that their present-day decisions will affect their future payoffs.

However, it is well-known that consumers *do* behave in a forward-looking fashion in a vast array of situations and markets.[2] In particular, researchers have found strong evidence that consumers stockpile in storable-goods markets when they face temporary price cuts, anticipating that prices will increase shortly after (Blattberg, Eppen and Lieberman 1981; Neslin, Henderson and Quelch 1985; Mela, Jedidi and Bowman 1998; Pesendorfer 2002). One well-documented example is the market for laundry detergent (Bell, Iyer and Padmanabhan 2002; Hendel and Nevo 2006*a*,*b*; Seiler 2013; Pires 2016).

Previous work in economics and marketing research has developed sophisticated structural models to estimate consumer stockpiling (Erdem, Imai and Keane 2003;

[2] For example, consumers display forward-looking behavior when booking flight tickets (Li, Granados and Netessine 2014), buying college textbooks (Ching and Osborne 2015), and making career decisions (Keane and Wolpin 1997).

Sun, Neslin and Srinivasan 2003; Hendel and Nevo 2006*a*; Osborne 2010; Hartmann and Nair 2010; Seiler 2013; Pires 2016). These studies model forward-looking strategic stockpiling by incorporating a storage cost parameter and consumer price expectations. The main strength of these structural models is that they allow us to perform counterfactual simulations. Consequently, the literature looks at various demand-side counterfactuals, for example how household purchases change when storage costs are lowered.

There is limited empirical work on pricing for markets with dynamic consumer demand. One reason for this is that, even in toy models, it is computationally challenging to solve for optimal prices in a framework with both a dynamic demand side and a dynamic supply side. We are aware of only one paper doing this: Hendel and Nevo (2013) develop a simple dynamic demand model in order to empirically quantify the impact of intertemporal price discrimination on profits in the soda market. They find that sales capture 25-30% of the gap between non-discriminatory profits and third-degree price discrimination profits. Their model relies on simplifying assumptions, e.g. storage cost is assumed to be zero and products are assumed to be perishable. These assumptions allow them to derive an optimal pricing strategy, but make the model unsuitable for many markets.

In related work, Nair (2007) solves for optimal prices in a durable goods market with forward-looking consumers. More specifically he investigates different pricing policies in the market for video games: After the introduction of a new game, consumers may choose to wait for the price of the game to drop. Nair solves for optimal prices in a market with one monopolist and two types of consumers who differ in their product valuation. Such an approach could theoretically be taken to storable goods data. However, product storability comes with additional computational challenges: It requires keeping track of households' inventories because, unlike in markets of durable goods, consumers do not drop out of the market after making a purchase. In practice, it is computationally challenging if not infeasible to incorporate this in the estimation.

Instead of solving for optimal prices, Osborne (2010) simulates different pricing regimes. This is the paper closest to our work. Osborne studies how changes in frequency and depth of promotions affect revenues in the canned tuna category. He finds that increasing promotion depth significantly increases sold quantity whereas an increase in promotion frequency has a much smaller effect. Our work differs from his in that we investigate promotion length instead of promotion frequency. This difference is important because, unlike a change in frequency, a change in promotion length will not affect the fixed costs of running promotions.

Our work is also related to a large literature on sales. This literature proposes a multitude of rationales for sales. One important explanation is that sales take advantage of some dimension of consumer heterogeneity. This may be heterogeneity in information about prices (Varian 1980), preferences and brand loyalty (Sobel 1984; Narasimhan 1988; Raju, Srinivasan and Lal 1990; Hendel and Nevo 2013), storage levels (Hong, McAfee and Nayyar 2002; Pesendorfer 2002), storage costs (Blattberg, Eppen and Lieberman 1981; Jeuland and Narasimhan 1985), or heterogeneity in the stores that shoppers visit (Salop and Stiglitz 1982).[3] In our paper, we investigate heterogeneity in storage costs and storage levels as the main rationale for sales.

3.3 Data

3.3.1 IRI Panels

The market research company Information Resources Inc. provided us with U.S. retail data on households' laundry detergent purchases in the years 2001 to 2004. Laundry detergent comes in two main forms: liquid and powder. In order to be able to compare pack sizes, we have to choose one form of detergent. In the following, we study liquid laundry detergent because it has a market share of 94.9% in our sample period. The household panel contains 6,000 to 10,000 households (the number of panelists varies by year) in Eau Claire, Wisconsin, and Pittsfield, Massachusetts. These households use handheld scanners to scan their purchases after every shopping trip. We observe all detergent purchases of the panelists, and for each purchase, we observe the paid price, the date of the purchase, the store identifier, the chain identifier, and the characteristics of the household. The average household purchases detergent every six weeks, buys only one pack per trip, and switches between two different brands.

[3]There is also a vast business and operations research literature investigating seller-side rationales for sales, e.g. inventory management (e.g. Whitin 1955; Petruzzi and Dada 1999) or loss-leading (e.g. Mason and Mayer 1984; Lal and Matutes 1994). In this paper, we abstract from such motives for sales.

Table 3.1 displays brand summary statistics. The market is relatively concentrated, with the largest three firms capturing almost half of the market. We focus our analysis on the 12 best-selling brands, which have an accumulated market share of 82.5%. All residual brands are collected into one composite brand called "OTHER". In the supermarket retailing context, detergent is a relatively expensive product. More than half of the brands in our sample have at least one variety that retails at more than 10 USD.

Table 3.1: Summary Statistics: Brands

	Price				
Brand	Mean	Std.Dev.	Min	Max	Market Share
ALL[1]	4.974	1.534	.65	15.4	9.794
CHEER	6.115	1.558	3.5	27.96	1.713
DYNAMO	4.119	1.909	1.49	9.99	7.529
ERA	5.002	1.922	1.48	15.13	8.481
FAB	4.891	1.344	1.98	6.79	1.067
GAIN	5.776	1.245	2.44	10.59	.86
OTHER	3.844	1.654	.34	14.89	17.499
PUREX	4.096	1.549	1.8	11.69	12.39
SURF	5.656	.891	3.09	9.98	.466
TIDE	7.774	3.19	1.28	24.59	19.468
TREND	2.797	.156	2.69	3.75	.933
WISK	6.087	1.946	1.5	18.51	6.504
XTRA	2.842	.74	.5	5.99	12.323
YES	4.334	1.693	1.99	6.99	.972

[1] ALL is the name of a brand.

We also use an auxiliary store panel to supplement our estimation with information on prices that we do not observe in the household panel (see Section 3.5.1). The store panel contains check-out scanner data from 1,588 U.S. supermarkets. We observe the weekly sales quantity for every laundry detergent brand and size that was sold at least once, together with prices, promotions, and information on the stores' location.[4]

[4]For more details on the IRI data set see Bronnenberg, Kruger and Mela (2008).

3.3.2 Preliminary Analysis of Stockpiling

We observe a distinctive price pattern in the store level data. Prices are not kept constant, but vary across time. Figure B.1 and B.2 show how the average prices for 100-ounce and 200-ounce packs change over time across all stores. It shows that promotions do not occur in predictable intervals and that consumers consequently face uncertainty with respect to future prices and promotions.

In the following, we provide preliminary evidence that consumers stockpile strategically, i.e. they purchase for *future* consumption. In contrast, myopic consumers make purchases only with current consumption in mind. The household panel shows that during promotions, more ounces of detergent are sold. Imagine that this sales spike is the result of consumers only caring about the present. Then present-day purchase decisions should be independent of previous or following purchases. Interestingly, we find that a) the duration since the previous purchase is shorter during a promotion; and b) the duration before the next purchase is longer during a promotion. Both effects are significant at the 10% level (see Table B.1). This finding is inconsistent with the hypothesis that consumers do not have strategic stockpiling motives. Instead, it suggests that households build stock during promotions by purchase acceleration, i.e. by making earlier purchases than initially planned.[5]

Furthermore, we find that households rarely buy multiple packs per shopping trip; instead, the vast majority of households (84.7%) buys only one pack of detergent per trip. We also find that consumers stockpile by buying *larger* packs: The share of large packs increases considerably during promotional periods (see Table B.2). Consumers switch frequently between pack sizes; 54.0% of all households buy more than one pack size, and on 12.6% of all shopping trips, the household chooses a different pack size than the pack size it purchased last time. All in all, our preliminary findings suggest that consumers stockpile primarily through purchase acceleration and by buying larger pack sizes.

[5] For a more detailed investigation into the evidence of stockpiling in the market for laundry detergent see Hendel and Nevo (2006*b*) and Pires (2016).

3.4 Model

In each period t, household $i = 1, ..., I$ can buy one pack of laundry detergent.[6] We define a product j as a unique combination of brand $b = 1, ..., B$ and size $z = 1, ..., Z$, for example "Tide 50-ounce pack" or "Xtra 100-ounce pack".[7] The household can also choose to buy no detergent ($j = 0$). In the following, we denote a household's decision to buy product j in period t as $d_{it} = j$. The flow utility of $d_{it} = j$ at the time of the purchase decision is:

$$U_{ijt} = \begin{cases} \underbrace{v(c_i) - C(k_t) + \alpha p_{jt} + x_{jt}\beta}_{=u_{ijt}} + \varepsilon_{ijt} & \text{if } j = 1, ..., J \\ v(c_i) - C(k_t) + \varepsilon_{i0t} & \text{if } j = 0, \end{cases} \tag{3.1}$$

where p_{jt} is the retail price, x_{jt} is a vector of the observable characteristics, and ε_{ijt} is an individual-product-time-specific demand shock. The parameter α is the disutility from price, β is the taste parameter. The term u_{ijt} denotes the flow utility minus the idiosyncratic error. Households consume the product at a household-specific constant rate c_i and receive utility $v(c_i)$ from consumption. Any units that are not consumed enter the household's inventory. An inventory of k_t units creates storage costs $C(k_t)$. Both the consumption rate and the storage costs do not depend on the composition of brands in stock, i.e. product differentiation occurs at the moment of purchase, not at the moment of consumption.[8]

In the following, we denote by s_{it} a vector that captures the current level of inventory and the current prices of all brand-size combinations at time t. The vector $\varepsilon_{it} = (\varepsilon_{ijt})_{j=1,...,J}$ stacks the household-time-specific shocks for all brands and pack sizes. Together, s_{it} and ε_{it} describe the so-called *state*, i.e. all the information that is relevant for a household's decision. In each period, the household fully observes its state and forms expectations over future states.

[6] We make this assumption since the vast majority of households in our sample never buys more than one pack per trip. For a further discussion see Section 3.5.2.

[7] In this paper, we do not model retailer choice because it adds significantly more complexity to the consumer decision (Farley 1968; Fotheringham 1988; Rhee and Bell 2002; Smith 2004; Lu 2016). We assume that during any given shopping trip, a household's choice set includes only the products at the visited store.

[8] We borrow this model specification from Hendel and Nevo (2006*a*). It simplifies our estimation because it implies that not the brand composition of storage but only the total quantity in stock matters for consumption.

We assume that households form rational expectations, i.e. their expectations are correct in equilibrium. This is a standard assumption in the literature (for a survey of the dynamic discrete-choice literature, see Aguirregabiria and Mira (2010)). In general, the rational-expectations assumption is made for identification purposes because observed choices may be explained by multiple specifications of expectations and beliefs. This assumption can be relaxed if the researcher has data on elicited beliefs (e.g Van der Klaauw and Wolpin 2008; Arcidiacono, Hotz and Kang 2012).

Households understand that future inventory depends on how much detergent they buy, how much they consume, and whether they run out of stock in the meantime. Consequently, inventory evolves according to

$$k_t = \max(k_{t-1} - c_i + q_{it}, 0), \tag{3.2}$$

where q_{it} is the amount of new detergent that enters storage if household i buys a pack of detergent in period t. While the evolution of inventory is deterministic, consumers face uncertainty with regards to future prices and future utility shocks. This implies that, in order to compute expected future values, we have to integrate over prices and utility shocks. Since this is computationally burdensome, we make two assumptions to reduce the dimensionality of the problem.

Firstly, we make assumptions on the error structure. Without any assumptions, we have to numerically integrate over the errors for every future value term. To avoid this, we make a simplifying assumption that was initially proposed by Rust (1987) and since then has become a standard in the literature:

Assumption 1 *The demand shocks ε_{ijt} are independently and identically extreme-value I distributed.*

This assumption is popular in the literature on dynamic discrete-choice models (see Aguirregabiria and Mira (2010)) because it delivers a closed-form solution of expectations of future utility conditional on the states and, thus, significantly reduces the computational burden.

Secondly, we aim to simplify consumer expectations of future prices. A model in which price expectations depend on the (infinitely long) history of previous prices is neither realistic nor tractable. Instead, the literature makes simplifying assumptions of various degrees, for example assuming that price expectations are conditional on a small number of price lags (Hendel and Nevo 2006*a*) or the identity of the store (Seiler 2013). In particular, we follow Pires (2016) and make the following assumption:

Assumption 2 *Prices are identically and independently distributed.*

This assumption implies that consumers do not condition their price expectations on past prices. Figures B.1 and B.2 show that prices indeed do not display clear promotional patterns. Therefore, it seems difficult for consumers to form sophisticated price expectations based on lagged prices. This is supported by the fact that, in general, consumers have been found to have poor knowledge of the prices that they pay; even seconds after selecting a product, only about 50% of shoppers are able to correctly recall its price (Dickson and Sawyer 1990; Wakefield and Inman 1993). Furthermore, a number of important theoretical studies suggests that prices can be intertemporally independent since the optimal decision to conduct a sale may involve randomization (e.g. Varian 1980; Salop and Stiglitz 1982; Narasimhan 1988; Raju, Srinivasan and Lal 1990; Pesendorfer 2002; Su 2010).

In each period, the household makes a decision according to a decision rule r_{it} which assigns a decision – whether to purchase anything and, if so, which product – to each possible state $(s_{it}, \varepsilon_{it}) \in \mathcal{S}$. Let a policy be a sequence of decision rules for each point in time. For a given policy $\pi_i = (r_{i1}, r_{i2}, ...)$, the discounted value of utilities is

$$V(\pi_i, s_{i0}, \varepsilon_{i0}) = E\left[\sum_{t=1}^{\infty} \tau^{t-1} U_{ijt}(s_{it}, \varepsilon_{it}, r_{it}(\pi_i, s_{it}, \varepsilon_{it}))|s_{i0}, \varepsilon_{i0})\right], \tag{3.3}$$

where τ is the discount factor. Since the time horizon remains infinite in every period and the transition probabilities are stationary, we do not have to find the optimal policy but only the optimal decision rule. For the same reason, we can drop the time subscript in the following. For notational simplicity we also drop the household subscript, i.e. the following solution of the dynamic discrete choice problem relates to a specific household i. Given that the set of decision rules is finite, there is an optimal decision rule $r^*(s, \varepsilon)$. The associated value function satisfies

$$\begin{aligned} V(s, \varepsilon) &= \max_{m \in \mathcal{R}} \{u(s, m) + \varepsilon(m) + \tau E(V(s', \varepsilon')|s, m)\} \\ &= \max_{m \in \mathcal{R}} \{u(s, m) + \varepsilon(m) + \tau \int V(s', \varepsilon') f(s'|m) g(\varepsilon') ds' d\varepsilon'\}, \end{aligned} \tag{3.4}$$

where $\mathcal{R}$ denotes the set of all possible decision rules. s' and ε' denote s and ε in the next period, respectively. $f(\cdot)$ and $g(\cdot)$ are probability distribution functions. We reformulate the problem in order to characterize this optimality condition as a function of s only. Before the demand shock ε is realized, the consumer expects

$$W(s) = \int V(s, \varepsilon) g(\varepsilon) d\varepsilon. \tag{3.5}$$

If we plug this expression into Equation (3.4), we obtain

$$V(s, \varepsilon) = \max_{m \in \mathcal{R}} \left\{ u(m, s) + \varepsilon(m) + \tau \int W(s') f(s'|s, m) ds' \right\}. \tag{3.6}$$

Finally, if we take the expectation on both sides of Equation 3.6, we obtain the integrated Bellman equation which depends only on s:

$$W(s) = \int \max_{m \in \mathcal{R}} \left\{ u(m, s) + \varepsilon(m) + \tau \int W(s') f(s'|s, m) ds' \right\} g(\varepsilon) d\varepsilon. \tag{3.7}$$

The right-hand side of this equation defines a contraction mapping $\Gamma : \mathcal{B} \to \mathcal{B}$. According to the Banach fixed-point theorem, the equation has a unique solution that must equal the expected value function $W = \Gamma(W)$. Once we have solved for $W(s)$, we can define the choice-specific value function

$$V_k(s) = u_k(s) + \tau \int W(s') f(s'|s, k) ds' + \varepsilon_k, \tag{3.8}$$

where the subscript k denotes the choice of k in the present period. Due to Assumption 1, the choice probabilities have the simple logit form. Choosing option k has the probability

$$\Pr(k|s) = \frac{\exp(V_k(s))}{\sum_{j=0}^{J} \exp(V_j(s))}. \tag{3.9}$$

In general, dynamic discrete-choice models with infinite-time horizons do not have closed-form solutions but must be solved numerically. We use the nested fixed-point algorithm proposed and made popular by Rust (1987). This is an iterative gradient search method to obtain the maximum-likelihood estimates of the structural parameters. It nests two loops: The inner loop solves the dynamic optimization problem for a given set of structural parameters, and the outer loop solves for the parameters that maximize the likelihood.

3.5 Estimation

3.5.1 Estimating Non-Observed Data

Prices In order to obtain purchase probabilities, we need to know what price a household *would have* paid for any other alternative at the store it was shopping at. To fill in these prices, we use the IRI store-level panel data. It records weekly prices for all products that were sold at least once during each week. However, not all prices can be matched like this because not every product is sold at every supermarket at least once a week. For 16,905 out of 35,132 brand-pack-size-store-week combinations we are able to retrieve the corresponding store prices. For the remaining cases – mostly smaller brands and exotic varieties – we replace the missing price with the corresponding weekly or monthly average price across all stores of the chain. We are able to fill in all missing prices this way.

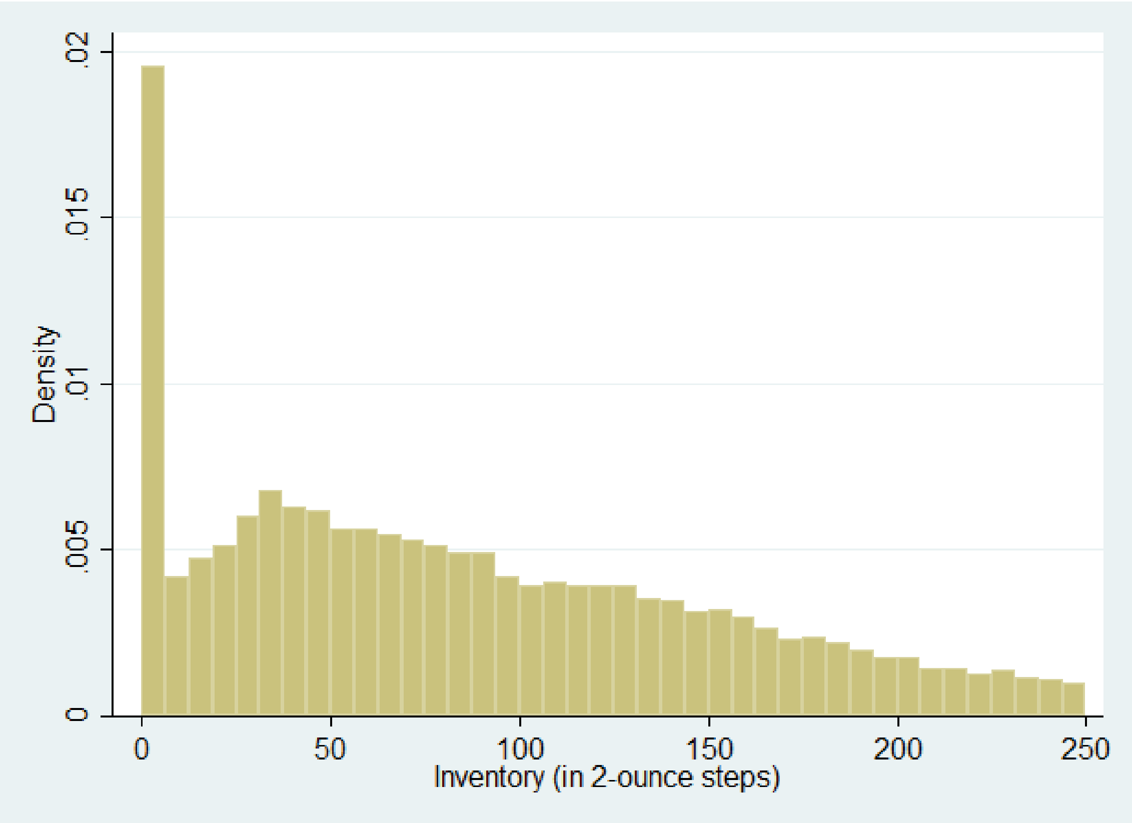
Inventory Consumer inventory and consumption are unobserved in our data. However, if we knew both the consumption rate and the initial inventory of a household, we could use the observed purchases to construct the series of inventories. Because the consumption of laundry detergent is of a relatively stable nature, we can compute the weekly consumption rate of a household as the total sum of purchases during the sample period divided by the total number of weeks (e.g. Erdem, Imai and Keane 2003).

Then, we follow Seiler (2013) and assume that households start with zero inventory before the first observed purchase in our sample.[9] The impact of the initial inventory will fade over time because consumption is not constant; instead, it drops to zero when stocks are depleted. In the estimation, we follow Hendel and Nevo (2006*a*) and Pires (2016) and drop the first ten observations of each household in order to mitigate the effect of the initial inventory. Finally, in order to reduce the state space in the inventory level dimension, we keep only those observations with inventories that are less or equal to 500 liquid ounces, i.e. about 14.785 liters. This

[9] We run robustness checks of this and use an initial inventory equal to three, five, or ten times the consumption rate. Our results are robust to these changes.

assumption is not too restrictive, as we lose only 0.67% of observations.[10] Figure 3.1 shows the final distribution of estimated inventories. There is a spike at zero inventory - 1.9% of the households face stock-outs. Higher levels of inventory are increasingly less likely to occur. In the estimation, we discretize inventory in units of 2 liquid ounces. We do so in order to make the problem tractable; it does not imply that serving size equals 2 ounces.

Figure 3.1: Estimated Inventory Distribution 2001-2004



3.5.2 Reducing Dimensionality

Applications to real-world data often have the problem of dimensionality of the state space. Consumers (theoretically) face infinitely many possible inventory levels[11] and

[10] If a household at one point in time has an inventory above 500 ounces, we drop only the observations during that time.

[11] We only have to track the total amount of inventory and not the brand-size composition of it. This is because in our model product differentiation takes place at the time of purchase and not at the time of consumption (e.g Hendel and Nevo 2006*a*).

many possible brand-size combinations that are all offered at a variety of prices. This extends the state space so much that the estimation may become computationally infeasible. To reduce the state space, we adopt an approach proposed by Hendel and Nevo (2006*a*) that simplifies the estimation by decomposing the problem into a static discrete-choice part and a dynamic discrete-choice part. We further reduce the state space by dropping a selection of households that show outlier behavior.

Decomposition Approach In the following, we detail the steps of the decomposition approach of Hendel and Nevo (2006*a*). We break the problem down into 1) optimal brand choice and 2) optimal size choice. This means that the probability of choosing brand b and size z can be written as the probability of choosing a pack size times the probability of choosing a brand given this pack size, i.e.

$$\Pr(d_t = (b, z)|p_t, k_t) = \Pr(d_t^{brand} = b|p_t, z_t, k_t) \cdot \Pr(d_t^{size} = z|p_t, k_t), \tag{3.10}$$

where k_t is the inventory carried in period t and p_t is the price, z_t denotes the pack size chosen in period t, and d_t^{brand} and d_t^{size} denote the choice of brand and size, respectively.

Firstly, we estimate a static discrete-choice model in which we restrict the choice of options to products of the same size as the one that was actually purchased. This estimation yields the static demand parameters, namely the price sensitivity and taste preferences. We then use the estimates from this first step to compute for each pack size the inclusive value, i.e. the expected utility from choosing that pack size:[12]

$$\omega_{zt} = \log\left(\sum_{b=1}^{B} \exp(\alpha p_{zbt} + x_{zbt}\beta)\right). \tag{3.11}$$

Since we collapse brand valuation into the inclusive value, the inclusive value serves as a size-specific adjusted price-index. In the following, we only need to track one inclusive value per pack size instead of tracking one price for each brand-size combination. This considerably reduces the state space. We use the empirical distribution of inclusive values as the probability distribution with which households face a certain inclusive value. This implies that consumer expectations over future inclusive

[12] The inclusive value originates from the nested logit (e.g. McFadden 1980) but has gained popularity in many applications with a nested structure.

values are time-invariant; this assumption is analogous to Assumption 2 of the original problem (see Section 3.4). The utility function of the simplified dynamic problem is

$$\tilde{U}_{izt} = \begin{cases} v(c_i) - C(k_{it}) + \omega_{zt} + \varepsilon_{izt} & \text{if } z = 1, ..., Z \ , \\ v(c_i) - C(k_{it}) + \varepsilon_{i0t} & \text{if } z = 0. \end{cases} \tag{3.12}$$

where $z = 0$ corresponds to the no-purchase option. This modified utility function corresponds to the utility function in Equation 3.1 and can be solved analogously.

Household Selection We drop households that make excessively many or few purchases because our model may not be able to properly describe their shopping behavior.[13] More specifically, we drop households that, on average, make less than one purchase every six months or more than one purchase every two weeks. We conduct robustness checks with different thresholds and find that our results do not change substantially.

In order to reduce the state space and, thus, the computational burden, we limit households' choice sets. We assume that a household's choice set includes only the pack sizes that it buys over the total sample period, i.e. from 2001 to 2004. We further limit our analysis to households that consider only the two dominant pack sizes: 100-ounce packs ($\approx$ 2.957 liters) and 200-ounce packs ($\approx$ 5.915 liters) which have a combined market share of 68.95%. In other words, we drop a household from our analysis if it purchases another pack size at least once during our sample period.

The large majority of households (84.68%) buys only one pack on each trip, conditional on buying detergent (see Table B.3). This implies that, in the detergent market, consumers do not stockpile by increasing the number of purchased packs per trip. Instead, they keep buying only one pack per trip but increase their purchase frequency. In the following, we restrict our analysis to households that buy at most one pack per shopping trip. We drop all households that at some point in time buy more than one pack.

Our final sample contains 243 households over a span of 208 weeks. This is a very standard sample size in the literature (see for example Hendel and Nevo (2006*a*)

[13]For example, households with extremely few detergent purchases may regularly visit a laundromat. Households with extremely frequent purchases may be buying not only for private consumption but also for resale or on behalf of someone else, such as a relative, friend, or organization.

with 218 households or Osborne (2010) with 299 households). Table B.4 details how much each selection step reduces the sample size. Table B.5 shows the summary statistics of the households in the final sample.

3.5.3 Identification

We informally discuss the empirical identification of both static and dynamic parameters.[14] The parameters to be estimated are the price coefficient α, the taste parameter β, the consumption rate c_i, and the parametrized function $C(k)$. The utility of consumption $v(c_i)$ is not well-identified since households in our model always consume a constant amount of detergent, unless they face a stock-out. Therefore, we follow Seiler (2013) and define $v(c_i) = c_i$. Previous research establishes that the discount factor is difficult to identify (e.g. Rust 1994; Magnac and Thesmar 2002). Therefore, we follow the literature and set τ equal to 0.975, which corresponds to an annual interest rate of about 2.56%.

The identification of the static parameters is standard. Price sensitivity is identified by variation in prices. Brand and size preferences are identified by variation in shares of products. Heterogeneity in the price sensitivity is driven by variation in relevant household characteristics (here: family size) and heterogeneity in consumer response to promotions.

Identification of the dynamic parameters comes from variation in interpurchase duration and the extent to which consumers exploit price cuts: Higher storage costs decrease the consumer's ability to benefit from sales and decreases interpurchase duration. Imagine two households that face the same prices, have identical consumption rates, and buy the same total amount of detergent over a given period of time. One household buys only small packs, the other household buys only large packs. The household that buys small packs is then characterized by higher storage cost. For the storage costs, we assume the functional form $C(k_t) = \theta_1 k_t + \theta_2 k_t^2$. Doing so, we follow the literature that typically assumes a linear or quadratic cost function, for example Erdem, Imai and Keane (2003) in the market for ketchup, Osborne (2010) for canned tuna, as well as Hendel and Nevo (2006*a*), Seiler (2013), and Pires (2016) for liquid detergent.[15]

[14]For a formal discussion of identification in dynamic discrete-choice models, see Rust (1996) and Magnac and Thesmar (2002).

[15]We are aware that storage costs may depend not only on the total amount of liquid but also on the total number of detergent bottles. However, modelling this would require either further

Jointly, the two sets of estimates from stages one and two determine consumer response. Consequently, both sets of estimates are crucial to the simulation of pricing counterfactuals. The static parameters from the first stage determine the short-term effects of how consumers substitute between brands. The storage cost parameters from the second stage determine long-term effects of how consumers substitute in quantities.

3.6 Results

Table 3.2 shows a selection of parameters from the static estimation of brand choice conditional on chosen pack size. In all four specifications, price coefficients are negative and significant at the 1% level. We find that price sensitivity is heterogeneous; in particular, it tends to be larger for families with a larger income. We include brand fixed-effects and brand-size fixed-effects and find that they are almost all significant (see Table B.6 for the complete table of results). In the following, we continue with the estimates from Model 4, the richest specification.

Storage cost can be interpreted as the opportunity cost of the storage space. This cost is likely to decrease as total available storage space increases. Housing size should therefore be an important determinant of households' storage costs. However, we do not observe direct measures of housing size in our data. Instead, we use the existence of children as a proxy because families with children tend to live in larger homes. In our sample, 34.16% of the households have at least one child in their home (see Table B.5). We sort households into two types, those without children (type 1) and those with children (type 2).

data or strong assumptions on how households consume their stock, i.e. whether they consume bottle by bottle or spread consumption evenly over the bottles in stock.

Table 3.2: Estimation Results: Static Parameters

Variables	Model 1	Model 2	Model 3	Model 4
Price	-0.0560***	-0.482***	-0.525***	-0.436***
× Household Income				yes
Brand Dummies		yes		
Brand-Size Dummies			yes	yes
Observations	104,730	104,730	104,730	104,730

The symbols *, ** and *** denote significance at the 1%, 5%, and 10% level, respectively.

Table 3.3: Estimation Results: Dynamic Parameters

Variable	Mean	Standard Error	P-Value
Type 1: No Children			
Storage Cost (linear)	-0.0182	0.0091	0.03421
Storage Cost (quadratic)	-1.9128e-05	0.0018	0.4958
Type 2: Children			
Storage Cost (linear)	-0.0179	0.0077	0.0119
Storage Cost (quadratic)	-1.9114e-05	0.0018	0.4958

Table 3.3 shows the results from the dynamic choice problem, split by household type. We find that storage costs increase in a linear way because the quadratic cost term is not significantly different from zero. Furthermore, households with children, i.e. households that are more likely to have larger homes, incur a lower storage cost. To provide an idea of the economic relevance of these estimates, consider the storage cost for a 100-ounce pack at zero inventory. It is given by 50 (units of 2 ounces) $\times 0.017925 = 0.895$ USD for a household with children and $50 \times 0.018208 = 0.91$ USD for a household without children. Buying a 100-ounce pack doubles this storage cost.

3.7 Counterfactual Promotion Policies

In this section, we simulate how counterfactual promotional pricing would affect consumer decisions and seller revenue.[16] We look at two types of changes in pricing: Firstly, we study an increase in promotion length, i.e. existing promotions are extended in time. Secondly, we study an increase in promotion depth, i.e. promotional prices are further lowered. In both counterfactual simulations, consumers adjust their price expectations and face new dynamic programming problems. We compute counterfactual seller revenues as follows:

1. Change prices from observed p to counterfactual $\tilde{p}$.
2. Compute new inclusive values using the new prices and the previously estimated preference parameters $\hat{\alpha}$ and $\hat{\beta}$:

$$\tilde{\omega}_{zt} = \log\left(\sum_{b=1}^{B} \exp(\hat{\alpha}\tilde{p}_{zbt} + x_{zbt}\hat{\beta})\right). \tag{3.13}$$

3. Compute empirical probability of each inclusive value.
4. Compute value function $\tilde{W}(s)$ given new inclusive values $\tilde{\omega}$ and new empirical probability of inclusive values.
5. Use $\tilde{W}(s)$ to simulate household choices of quantity. Note that for every household and in each period this choice is affected by a random shock. Therefore, we simulate each household decision 1000 times, each time drawing a random shock from an extreme-value I distribution of shocks.
6. Simulate household brand choice conditional on quantity choice. Again, we use $n = 1000$ draws of the extreme-value I distributed error term. For each draw n, household i, and chosen pack size z, we compute :

$$\tilde{U}_{ibtn} = c_i - (\hat{\theta}_1 k_t + \hat{\theta}_2 k_t^2) + \tilde{\omega}_{zt} + \varepsilon_{ibtn} \quad \forall b \in B^z, \tag{3.14}$$

[16] We do not study profits because we do not observe marginal costs. Theoretically, we could back out marginal costs from a structural model – in this case, a model with both a dynamic demand and a dynamic supply side. However, this is so technically challenging that we are not aware of any paper in the literature that does this. One alternative is to make assumptions on marginal costs. For example, Nair (2007) assumes a constant marginal cost of 12 USD per video game. But since the packaging and production of detergent are less standardized than those of CD-ROM disks, it is much more difficult to make assumptions on marginal costs of detergent. Therefore we decide to focus, like Osborne (2010), on revenues instead of profits.

where B^z are all brands that are available in pack size z. A household i chooses brand b if $\tilde{U}_{ibtn} > \tilde{U}_{iltn}\ \forall\ l \in B^z$. We average household choices over $n = 1000$ draws.

7. Compute retailer revenue across products and time.

3.7.1 Promotion Length

We first simulate how a change in promotion length affects quantities and revenues. Since we observe prices and purchases on a weekly level, we can only vary promotion length in steps of one week. In the following, we simulate an extension of all promotions in our sample period (2001-2004) by one week. Table 3.4 displays how this affects purchase probabilities. Note that our counterfactuals are simulated for a time span of four years, i.e. they show long-term effects. We see that with longer promotions fewer households choose not to buy anything. This is because an extension of promotion length affects price expectations. When consumers expect lower product prices, the outside option of not buying anything becomes relatively less attractive. We find that, with longer promotions, consumers buy more packs of both sizes. In the baseline as well as in the extended-promotions scenario, type 2 households with children buy more of both pack sizes than childless type 1 households. This is because type 2 households tend to be larger and therefore consume more detergent.

3.7.2 Promotion Depth

We now simulate a change in promotion depth. In order to ensure that we can compare counterfactuals in promotion length and depth, we compare counterfactuals of equal promotion value. The promotion value is defined as the sum of the price discounts across products. We find that the promotion value of prolonging all sales by one week corresponds to the promotion value of an additional 2.307% price cut on all sales prices. Table 3.4 displays the results for this increase in promotion depth. Similar to the previous counterfactual – and following the same logic – we find that increased promotion depth leads to fewer households choosing the outside option. Again, type 2 households generally buy more detergent than type 1 households.

Table 3.4: Promotion Depth vs. Length: Quantities

Model	Small Pack (%)	Large Pack (%)	No Pack (%)
	Type 1: No Children		
Baseline	0.1276	0.017702	0.8546
1 week longer	0.1281	0.017703	0.8541
2.307% off	0.1308	0.017209	0.8519
	Type 2: Children		
Baseline	0.1344	0.019367	0.8461
1 week longer	0.1354	0.019733	0.8448
2.307% off	0.1392	0.018987	0.8419

3.7.3 Comparison: Length Vs. Depth

When we compare the two counterfactuals, we see that consumers react more strongly to an increase in promotion depth. We compute the revenues (see Table 3.5) and find that the elasticity of revenue[17] is 0.0314 for a change in promotion depth and 0.0082 for a change in promotion length, i.e. making a promotion deeper such that the promotion value increases by 1% will lead to a 3.14% increase in revenue. This effect is about four times larger than for an increase in promotion length. Our estimated revenue elasticities are of the same magnitude as those by Osborne (2010).

In the following, we discuss why a change in promotion depth is more effective than a change in promotion length. We have different dimensions of heterogeneity among households. Firstly, households carry different levels of inventory. Secondly, households differ in price sensitivity. Thirdly, we have two types of households with inherently different storage costs per unit. Lastly, households experience an idiosyncratic demand shock ε_{ijt} such that, even under identical conditions, a household may make different decisions on two days. These four dimensions of heterogeneity affect consumer behavior differently in the two pricing counterfactuals.

When promotions are made deeper by 2.307%, price $p_{\text{promo},t}$ is replaced by $\tilde{p}_{\text{promo},t} = p_{\text{promo},t} \cdot (1 - 0.02307)$. Now, there are households that would not buy at price $p_{\text{promo},t}$ but would buy at price $\tilde{p}_{\text{promo},t}$. Note that both the level of household inventory and the idiosyncratic demand shock ε_{ijbt} are unaffected by the price change. Instead, the influx of new buyers is driven by households who were previ-

[17] The elasticity of revenue with respect to sales value is computed by $\frac{\triangle \text{revenue}}{\triangle \text{sales value}} \cdot \frac{\text{sales value}}{\text{revenue}}$.

Table 3.5: Promotion Depth vs. Length: Revenues

Model	Revenue Type 1	Revenue Type 2	Total Revenue	Elasticity (of Revenue)
Baseline	20535.26	11305.69	31840.95	-
1 week longer	20687.36	11415.45	32102.81	0.0082
2.307% discount	21144.61	11729.06	32873.67	0.0314

ously too price-sensitive and/or had too high storage costs to make a purchase but can afford the purchase after the additional cut to sale prices.

Now look at the case in which promotions are extended by one week. Consider households that do not buy at price $p_{\text{promo},t}$ but do buy in the following week at price $p_{t+1} = p_{\text{promo},t}$. This influx of new buyers can be explained by two dimensions of heterogeneity and their interplay: A purchase may suddenly become attractive in period $t+1$ if the demand shock for a purchase is sufficiently large compared to the previous period and/or if household inventory i_{t+1} has sufficiently dropped due to mean-time consumption and, thus, lowered total storage costs.

Importantly, an increase in promotion depth is not generally more effective than an increase in promotion length: The relative effectiveness of the two promotion policies varies from market to market. In general, a change in promotion depth will be effective in a market in which storage costs and price sensitivities are relatively heterogeneous. A change in promotion length will be more effective in markets in which idiosyncratic shocks vary heavily and in which inventory can drastically change from one period to the next. The latter typically applies to markets that are strongly affected by demand shocks. Examples include the market for ice-cream (with weather-specific demand spikes), baking powder (rarely consumed on a daily basis, instead used irregularly for baking), and champagne (with demand spiking due to festive events).

3.8 Conclusion

In this paper, we target a core question of both marketing researchers and industry practitioners: How should one design promotions? In particular, we are the first to investigate how the length of a promotion affects its effectiveness and how this compares to changes in promotion depth. We study this in the context of a storable-goods market that is characterized by forward-looking consumers who strategically stockpile. We develop a dynamic, structural model of consumer stockpiling and apply it to the U.S. market for laundry detergent.

We find that in this market, shorter but deeper promotions generate more revenue than longer, shallower promotions. However, this is not a general result; instead, marketers need to tailor promotion policies to product markets. Our results suggest that shorter, deeper promotions are generally preferable in markets with relatively heterogeneous storage costs and price sensitivities. Longer, shallower promotions are better suited for markets with strong demand shocks and unsteady consumption rates.

3.9 Acknowledgement

I am grateful for insightful advice from Jaap Abbring, Ulrich Doraszelski, Jean-Pierre Dubé, Anthony Dukes, Tomaso Duso, Jana Friedrichsen, Germain Gaudin, Fedor Iskhakov, Mia Lu, Matthew Osborne, Stephan Seiler, and Hannes Ullrich. I also thank seminar participants at DIW Berlin and DICE for their helpful comments and Mike Kruger from Information Resources Inc. for his help with understanding the data.

Bibliography

Aguirregabiria, Victor, and Pedro Mira. 2010. "Dynamic Discrete Choice Structural Models: A Survey." *Journal of Econometrics*, 156(1): 38–67.

Arcidiacono, Peter, V. Joseph Hotz, and Songman Kang. 2012. "Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals." *Journal of Econometrics*, 166(1): 3–16.

Bell, David R., Ganesh Iyer, and V. Padmanabhan. 2002. "Price Competition Under Stockpiling and Flexible Consumption." *Journal of Marketing Research*, 39(3): 292–303.

Besanko, David, Jean-Pierre Dubé, and Sachin Gupta. 2003. "Competitive Price Discrimination Strategies in a Vertical Channel Using Aggregate Retail Data." *Management Science*, 49(9): 1121–1138.

Besanko, David, Sachin Gupta, and Dipak Jain. 1998. "Logit Demand Estimation under Competitive Pricing Behavior: An Equilibrium Framework." *Management Science*, 44(11): 1533–1547.

Bhardwaj, Bhanu, Olga Casabona, Joy Joseph, and Howard Shimmel. 2016. "Drive Growth with Media Parity." IRI Technical Paper.

Blattberg, Robert C., Gary D. Eppen, and Joshua Lieberman. 1981. "A Theoretical and Empirical Evaluation of Price Deals for Consumer Nondurables." *Journal of Marketing*, 45(1): 116–129.

Blattberg, Robert C., Richard Briesch, and Edward J. Fox. 1995. "How Promotions Work." *Marketing Science*, 14(3): 122–132.

Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela. 2008. "Database Paper-The IRI Marketing Data Set." *Marketing Science*, 27(4): 745–748.

Ching, Andrew, and Matthew Osborne. 2015. "Identification and Estimation of Forward-Looking Behavior: The Case of Consumer Stockpiling." Working Paper.

Chintagunta, Pradeep K. 2002. "Investigating Category Pricing Behavior at a Retail Chain." *Journal of Marketing Research*, 39(2): 141–154.

Currim, Imran S., and Linda G. Schneider. 1991. "A Taxonomy of Consumer Purchase Strategies in a Promotion Intensive Environment." *Marketing Science*, 10(2): 91–110.

Dickson, Peter R., and Alan G. Sawyer. 1990. "The Price Knowledge and Search of Supermarket Shoppers." *The Journal of Marketing*, 54(3): 42–53.

Draganska, Michaela, and Dipak C. Jain. 2006. "Consumer Preferences and Product-Line Pricing Strategies: An Empirical Analysis." *Marketing Science*, 25(2): 164–174.

Erdem, Tülin, Susumu Imai, and Michael P. Keane. 2003. "Brand and Quantity Choice Dynamics Under Price Uncertainty." *Quantitative Marketing and Economics*, 1(1): 5–64.

Farley, John U. 1968. "Dimensions of Supermarket Choice Patterns." *Journal of Marketing Research*, 5(2): 206–208.

Fotheringham, A. Stewart. 1988. "Consumer Store Choice and Choice Set Definition." *Marketing Science*, 7(3): 299–310.

Hartmann, Wesley R., and Harikesh S. Nair. 2010. "Retail Competition and the Dynamics of Demand for Tied Goods." *Marketing Science*, 29(2): 366–386.

Hendel, Igal, and Aviv Nevo. 2006*a*. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, 74(6): 1637–1673.

Hendel, Igal, and Aviv Nevo. 2006*b*. "Sales and Consumer Inventory." *RAND Journal of Economics*, 37(3): 543–562.

Hendel, Igal, and Aviv Nevo. 2013. "Intertemporal Price Discrimination in Storable Goods Markets." *American Economic Review*, 103(7): 2722–2751.

Hong, Pilky, R. Preston McAfee, and Ashish Nayyar. 2002. "Equilibrium Price Dispersion with Consumer Inventories." *Journal of Economic Theory*, 105(2): 503–517.

Jeuland, Abel P., and Chakravarthi Narasimhan. 1985. "Dealing-Temporary Price Cuts by Seller as a Buyer Discrimination Mechanism." *Journal of Business*, 58(3): 295–308.

Kadiyali, Vrinda. 1996. "Entry, its Deterrence, and its Accommodation: A Study of the US Photographic Film Industry." *RAND Journal of Economics*, 27(3): 452–478.

Keane, Michael P., and Kenneth I. Wolpin. 1997. "The Career Decisions of Young Men." *Journal of Political Economy*, 105(3): 473–522.

Lal, Rajiv, and Carmen Matutes. 1994. "Retail Pricing and Advertising Strategies." *Journal of Business*, 67(3): 345–370.

Li, Jun, Nelson Granados, and Serguei Netessine. 2014. "Are Consumers Strategic? Structural Estimation from the Air-Travel Industry." *Management Science*, 60(9): 2114–2137.

Lu, Anna W. 2016. "Inference of Consumer Consideration Sets." Working Paper.

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

Mason, J. Barry, and Moris L. Mayer. 1984. *Modern Retailing : Theory and Practice.* Business Publications, Plano, TX.

McFadden, Daniel. 1980. "Econometric Models for Probabilistic Choice Among Products." *Journal of Business*, 53(3): 13–29.

Mela, Carl F., Kamel Jedidi, and Douglas Bowman. 1998. "The Long-Term Impact of Promotions on Consumer Stockpiling Behavior." *Journal of Marketing Research*, 35(2): 250–262.

Nair, Harikesh. 2007. "Intertemporal Price Discrimination with Forward-Looking Consumers: Application to the US Market for Console Video-Games." *Quantitative Marketing and Economics*, 5(3): 239–292.

Narasimhan, Chakravarthi. 1988. "Competitive Promotional Strategies." *Journal of Business*, 61(4): 427–449.

Neslin, Scott A., Caroline Henderson, and John Quelch. 1985. "Consumer Promotions and the Acceleration of Product Purchases." *Marketing Science*, 4(2): 147–165.

Osborne, Matthew. 2010. "Frequency Versus Depth: How Changing the Temporal Process of Promotions Impacts Demand for a Storable Good." Working Paper.

Pancras, Joseph, and K. Sudhir. 2007. "Optimal Marketing Strategies for a Customer Data Intermediary." *Journal of Marketing Research*, 44(4): 560–578.

Pesendorfer, Martin. 2002. "Retail Sales: A Study of Pricing Behavior in Supermarkets." *Journal of Business*, 75(1): 33–66.

Petruzzi, Nicholas C., and Maqbool Dada. 1999. "Pricing and the Newsvendor Problem: A Review with Extensions." *Operations Research*, 47(2): 183–194.

Pires, Tiago. 2016. "Costly Search and Consideration Sets in Storable Goods Markets." *Quantitative Marketing and Economics*, 14(3): 157–193.

Raju, Jagmohan S., Venkatesh Srinivasan, and Rajiv Lal. 1990. "The Effects of Brand Loyalty on Competitive Price Promotional Strategies." *Management Science*, 36(3): 276–304.

Rhee, Hongjai, and David R. Bell. 2002. "The Inter-Store Mobility of Supermarket Shoppers." *Journal of Retailing*, 78(4): 225–237.

Richards, Timothy J. 2007. "A Nested Logit Model of Strategic Promotion." *Quantitative Marketing and Economics*, 5(1): 63–91.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, 55(5): 999–1033.

Rust, John. 1994. "Structural Estimation of Markov Decision Processes." In *Handbook of Econometrics.* , ed. Robert F. Engle and Daniel L. McFadden, 3081–3143. North-Holland, Amsterdam.

Rust, John. 1996. "Numerical Dynamic Programming in Economics." In *Handbook of Computational Economics.* , ed. H. M. Amman, D. A. Kendrick and J. Rust, 619–729. North-Holland, Amsterdam.

Salop, Steven, and Joseph E. Stiglitz. 1982. "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents." *American Economic Review*, 72(5): 1121–1130.

Seiler, Stephan. 2013. "The Impact of Search Costs on Consumer Behavior: A Dynamic Approach." *Quantitative Marketing and Economics*, 11(2): 155–203.

Smith, Howard. 2004. "Supermarket Choice and Supermarket Competition in Market Equilibrium." *Review of Economic Studies*, 71(1): 235–263.

Sobel, Joel. 1984. "The Timing of Sales." *Review of Economic Studies*, 51(3): 353–368.

Sudhir, K. 2001. "Competitive Pricing Behavior in the Auto Market: A Structural Analysis." *Marketing Science*, 20(1): 42–60.

Sun, Baohong, Scott A. Neslin, and Kannan Srinivasan. 2003. "Measuring the Impact of Promotions on Brand Switching When Consumers are Forward-Looking." *Journal of Marketing Research*, 40(4): 389–405.

Su, Xuanming. 2010. "Intertemporal Pricing and Consumer Stockpiling." *Operations Research*, 58(4): 1133–1147.

Van der Klaauw, Wilbert, and Kenneth I. Wolpin. 2008. "Social Security and the Retirement and Savings Behavior of Low-Income Households." *Journal of Econometrics*, 145(1): 21–42.

Varian, Hal R. 1980. "A Model of Sales." *American Economic Review*, 70(4): 651–659.

Verboven, Frank. 2002. "Quality-Based Price Discrimination and Tax Incidence: Evidence from Gasoline and Diesel Cars." *RAND Journal of Economics*, 33(2): 275–297.

Wakefield, Kirk L., and J. Jeffrey Inman. 1993. "Who Are the Price Vigilantes? an Investigation of Differentiating Characteristics Influencing Price Information Processing." *Journal of Retailing*, 69(2): 216–233.

Whitin, Thomson M. 1955. "Inventory Control and Price Theory." *Management Science*, 2(1): 61–68.

Table B.1: Duration Since Last Purchase and Till Next Purchase

Promotion	Obs	Mean	Std. Err.	Std. Dev.	95% Conf. Interval	
DURATION TILL NEXT PURCHASE						
0	51,564	7.6009	.0455	10.3280	7.5118	7.6900
1	39,032	7.0751	.0516	10.1905	6.9740	7.1762
Diff=mean(0)-mean(1)		.5258	.0689		.3908	.6608
DURATION SINCE LAST PURCHASE						
0	51,328	7.2769	.04469	10.1257	7.1892	7.3645
1	39,268	7.5019	.0528	10.4595	7.3984	7.6053
Diff=mean(0)-mean(1)		-.2250	.0689		-.3560	-.0900

3.11 Appendix

Table B.2: Shares of Pack Sizes

Pack Size (ounces)	Regular Price	Promotional Price
50	20.13%	2.09 %
100	71.30 %	86.06%
200	8.57 %	11.85%
Total	100.00 %	100.00%

This table shows the distribution of different pack sizes during regular-price periods (column 2) and promotional periods (column 3). The entries in the second and third column are column percentages.

Table B.3: Number of Purchased Packs Per Shopping Trip

Number of Packs per Trip	Number of Trips	%
1	83204	84.678
2	11209	11.408
3	3,010	3.063
4	392	0.399
5	113	0.115
6	221	0.225
7	7	0.007
8	18	0.018
9	34	0.035
10	14	0.014
12	11	0.011
13	1	0.001
14	1	0.001
15	21	0.021
16	1	0.001
21	1	0.001
30	1	0.001
Total	98259	100

Table B.4: Household Selection and Sample Size

Treatment: Keep households that...	Number Households
	8289
... make >7 but <105 purchases in 2001-2004	4397
... buy only 100- and 200-ounce packs	534
... never buy more than one pack	243

Table B.5: Summary Statistics: Household Characteristics

Group	Number	%
HOUSEHOLD SIZE		
1 person	50	20.58
2 people	110	45.27
3 people	31	12.76
4 people	34	13.99
5 people	14	5.76
6 people	4	1.65
CHILDREN		
No children	160	65.84
At least one child	83	34.16
ANNUAL HOUSEHOLD INCOME (IN USD)		
$<$ 9,999	6	2.47
10,000 to 11,999	9	3.70
12,000 to 14,999	10	4.12
15,000 to 19,999	13	5.35
20,000 to 24,999	22	9.05
25,000 to 34,999	20	8.23
35,000 to 44,999	35	14.40
45,000 to 54,999	27	11.11
55,000 to 64,999	19	7.82
65,000 to 74,999	27	11.11
75,000 to 99,999	33	13.58
$\geq$100,000	22	9.05
Total	243	100

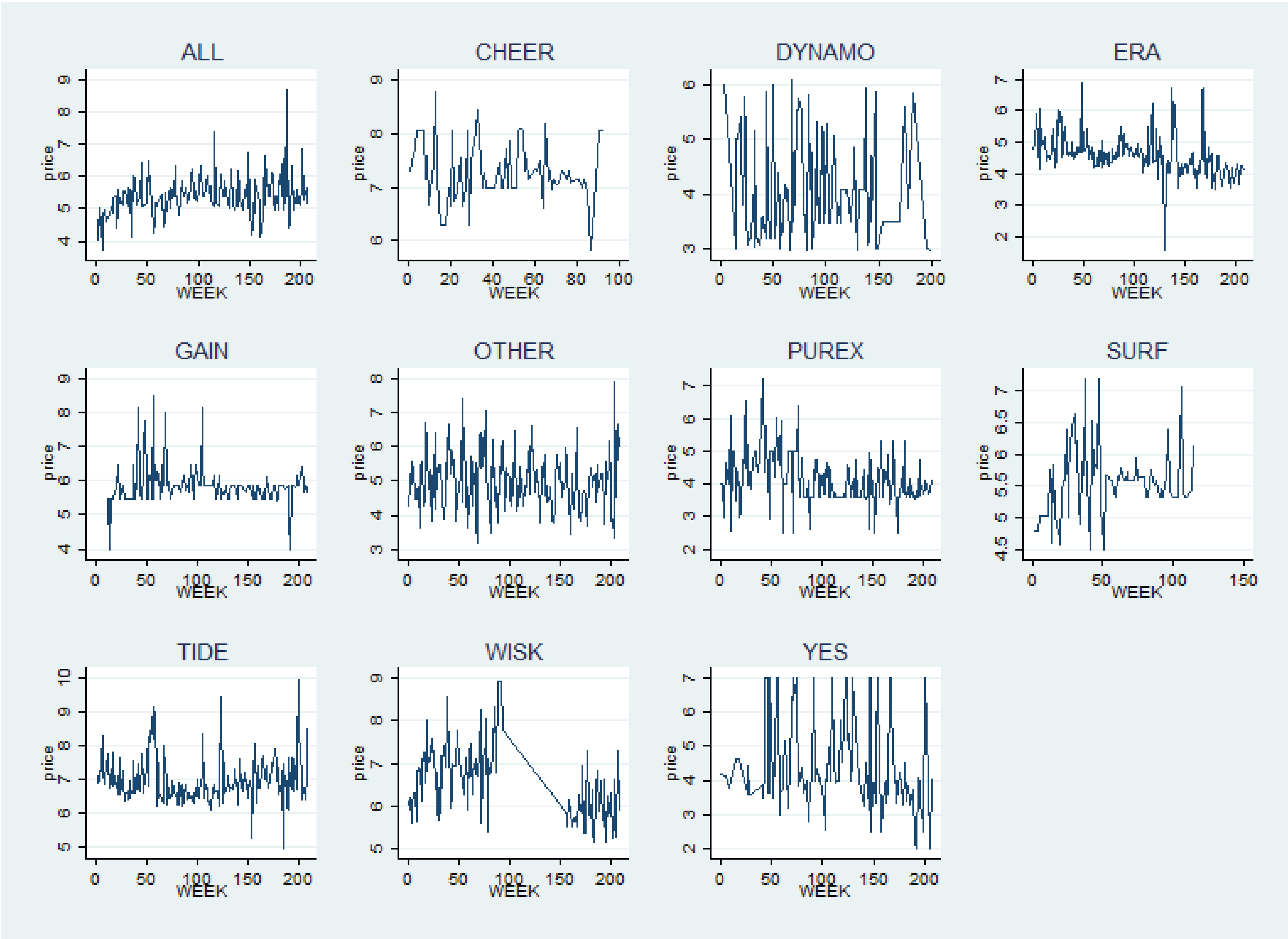


Figure B.1: Price Series 100-Ounce Packs

Price development over sample period across stores.

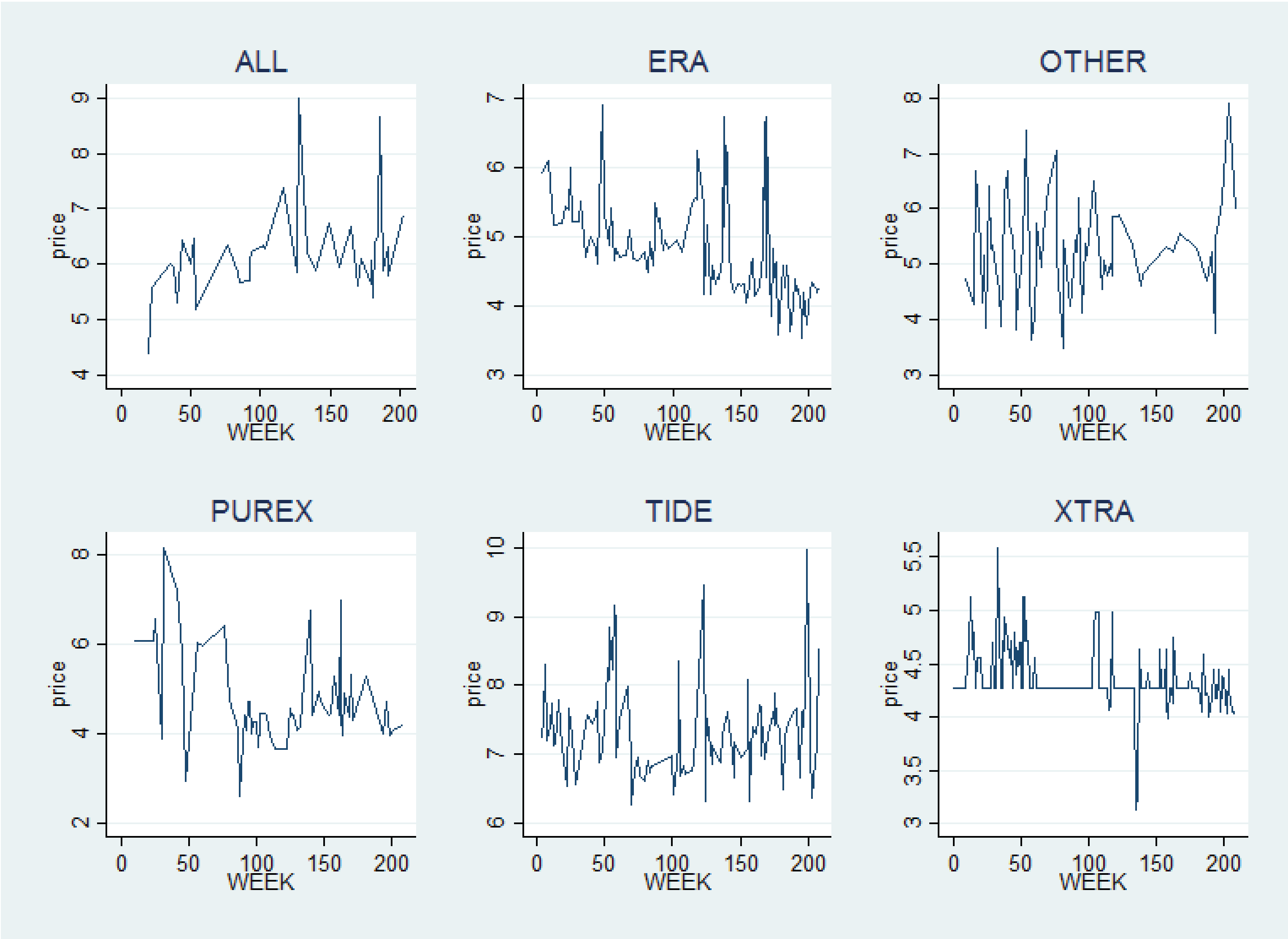


Figure B.2: Price Series 200-Ounce Packs

Price development over sample period across stores.

Table B.6: Estimation Results: Static Parameters (Full Table)

Variables	(1)	(2)	(3)	(4)
Price	-0.0560***	-0.482***	-0.525***	-0.436***
	(0.00806)	(0.0139)	(0.0284)	(0.0547)
Brand 2		-2.125***		
		(0.283)		
Brand 3		-4.063***		
		(0.321)		
Brand 4		-0.207***		
		(0.0741)		
Brand 5		-3.836***		
		(0.381)		
Brand 6		-1.601***		
		(0.163)		
Brand 7		-0.971***		
		(0.0881)		
Brand 8		-2.703***		
		(0.158)		
Brand 9		-1.878***		
		(0.161)		
Brand 10		2.435***		
		(0.0626)		
Brand 11		-0.965***		
		(0.124)		
Brand 12		-5.576***		
		(0.709)		
Brand 13		-4.205***		
		(0.412)		
Brand-Size 2			-1.048***	-1.069***
			(0.195)	(0.196)
Brand-Size 3			-2.183***	-2.184***
			(0.289)	(0.289)
Brand-Size 4			-4.270***	-4.271***
			(0.323)	(0.323)
Brand-Size 5			-0.243***	-0.240***
			(0.0772)	(0.0775)
Brand-Size 6			-1.805***	-1.824***
			(0.318)	(0.318)
Brand-Size 7			-3.990***	-3.990***
			(0.382)	(0.382)
Brand-Size 8			-1.716***	-1.709***
			(0.165)	(0.165)
Brand-Size 9			-0.989***	-1.010***
			(0.0952)	(0.0957)
Brand-Size 10			-1.919***	-1.930***
			(0.228)	(0.228)
Brand-Size 11			-2.767***	-2.792***
			(0.166)	(0.166)
Brand-Size 12			-4.099***	-4.102***
			(0.712)	(0.712)
Brand-Size 13			-1.923***	-1.921***
			(0.166)	(0.166)
Brand-Size 14			-2.815***	-2.866***
			(0.721)	(0.721)
Brand-Size 15			2.308***	2.310***
			(0.0790)	(0.0792)
Brand-Size 16			3.137***	2.978***
			(0.219)	(0.221)
Brand-Size 17			-1.054***	-1.045***
			(0.130)	(0.130)

continued on next page

continued from previous page

Brand-Size 18			-1.187**	-1.295**
			(0.533)	(0.534)
Brand-Size 19			-5.773***	-5.786***
			(0.710)	(0.710)
Brand-Size 20			-4.378***	-4.377***
			(0.412)	(0.412)
Income Group 2 × Price				-0.272***
				(0.0832)
Income Group 3 × Price				-0.181**
				(0.0823)
Income Group 4 × Price				-0.246***
				(0.0693)
Income Group 5 × Price				0.0293
				(0.0582)
Income Group 6 × Price				-0.203***
				(0.0632)
Income Group 7 × Price				-0.162***
				(0.0557)
Income Group 8 × Price				-0.0114
				(0.0557)
Income Group 9 × Price				-0.0132
				(0.0562)
Income Group 10 × Price				-0.0892
				(0.0561)
Income Group 11 × Price				-0.00125
				(0.0536)
Income Group 12 × Price				-0.0312
				(0.0584)
Observations	104,730	104,730	104,730	104,730

Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5% and 10% level, respectively.

Chapter 4

Inference of Consumer Consideration Sets

4.1 Introduction

In the modern market place, consumers face a large variety of products. While people generally value variety, the proliferation of alternatives may pose a complicated decision problem: Consumers need to engage in costly search in order to evaluate and compare alternatives all the while being constrained by cognitive limitations. To simplify the decision problem, consumers have been found to reduce the global set of objectively available alternatives to a subset of "relevant" alternatives. In the marketing and psychology literature, this concept is well established, and the smaller subset of relevant alternatives is typically referred to as the "consideration set" (Howard and Sheth 1969; Bettman 1979; Hauser and Wernerfelt 1990; Roberts and Lattin 1997; Malhotra, Peterson and Kleiser 1999).

Due to their cognitive nature, consideration sets are typically unobserved. Consequently, most studies in the demand estimation literature have to assume a model of consideration. For example, the economics literature usually assumes that consumers consider the global set of alternatives (Berry 1994; Berry, Levinsohn and Pakes 1995; Nevo 2000) whereas the marketing literature often uses a two-staged consideration set approach.[1] Importantly, both literatures generally motivate their choice of the consideration model with intuition but rarely support it with statistical evidence.

It is important to choose a consideration model that closely matches actual consumer behavior. This is because misspecified models of consideration can lead to biases in the demand estimates (shown for example by Bronnenberg and Vanhonacker 1996; Sovinsky 2008; Draganska and Klapper 2011; Conlon and Mortimer 2013), and this bias will carry over to supply side estimates and policy evaluations because they require demand estimates as an input.

In this paper, we propose a framework which is able to formally test competing models of consideration against one another. Our test follows the intuition of the so-called "menu approach" which is used to infer unobserved firm conduct and compares the equilibrium outcome in an industry to theoretical predictions of a finite set of alternative models . Our test has relatively modest data requirements. In addition to sales data, it requires only data on marginal cost-shifters. At least on an aggregate level, such data is widely available for many industries.

[1]See for example Allenby and Ginter (1995), Bronnenberg and Vanhonacker (1996), Chiang, Chib and Narasimhan (1998), Draganska and Klapper (2011), and Barroso and Llobet (2012).

We illustrate our approach in an application to the grocery retailing industry. Specifically, we test the model of global consideration sets against a two-stage model of consideration. We apply our test to the categories of milk and coffee, both of which have been extensively studied in the literature.[2] Our results show that the consideration process fundamentally differs across product categories: While the assumption of global consideration sets performs well in the market for coffee, it performs poorly in the market for milk. Instead, buyers of milk seem to consider milk only at the store at which they are currently shopping. We explain this discrepancy between the two markets with differences in demand and supply conditions, for example in terms of consumer perception of the product category, the level of product differentiation, retailer pricing, and advertising. Our results suggest that the assumption of global consideration sets is better suited for hedonic goods like coffee or wine, i.e. goods that provide emotional responses like excitement or pleasure. In contrast, the two-stage model is better suited for utilitarian goods, i.e. primarily functional goods like milk, sugar, or flour.[3]

The remainder of this paper is organized as follows: First, we give a brief overview of the related literature in Section 4.2. We develop our model in Section 4.3 and describe the data and patterns of consumer behavior in Section 4.4. In Section 4.5, we describe the identification strategy, our estimation procedure, and how we allow for household heterogeneity in our estimation. We present and discuss the estimation results in Section 4.6. Finally, we conclude in Section 4.7.

4.2 Related Literature

We contribute to a large literature in economics and marketing that aims to infer consumer preferences from revealed choices. In this literature, discrete-choice methods have gained wide-spread use (for a review see Train (2009)). The central premise of discrete-choice models is that consumers are utility-maximizers, i.e. when faced with a finite number of alternatives, they choose the alternative that gives them the largest utility. A discrete-choice model has to specify two things: Firstly, it needs to

[2]See for example Guadagni and Little (1983), Krishnamurthi and Raj (1988), Draganska, Klapper and Villas-Boas (2010), Draganska and Klapper (2011), and Bonnet and Bouamra-Mechemache (2015).

[3]The difference between hedonic and utilitarian goods is well-established in the marketing literature. For a discussion see Holbrook and Hirschman (1982) or Dhar and Wertenbroch (2000).

specify the utility function in the form of parametric and distributional assumptions. Secondly, it needs to specify the set of alternatives from which the consumer makes her choice.

Economics and marketing have traditionally made different assumptions on the set of products that a consumer considers. In economics, it is typically assumed that consumers consider the global set of products (e.g. Berry 1994; Berry, Levinsohn and Pakes 1995; Nevo 2001). This modeling assumption is, to a large part, driven by the limited availability of individual-level data in many economics applications. In contrast, a large share of the marketing literature studies consumer packaged goods for which detailed individual-level is often available; marketing researchers thus tend to be able to use richer models of consumer choice. A dominant belief in marketing is that consumers seek to simplify their decision problem by reducing the set of objectively available options to a subset of "relevant" options. The actual choice is then made only from this subset, i.e. the so-called "consideration set".[4] Marketers use two-staged models of consideration to study the determinants of consideration sets, such as advertising (Allenby and Ginter 1995; Mitra 1995; Sovinsky 2008; Draganska and Klapper 2011; Honka, Hortaçsu and Vitorino 2017), promotions (Siddarth, Bucklin and Morrison 1995), or search costs (Mehta, Rajiv and Srinivasan 2003; De los Santos, Hortaçsu and Wildenbeest 2012; Seiler 2013).

What complicates demand estimation is the fact that consideration sets are rarely observed and therefore prone to misspecification. This in turn can bias demand estimates (shown by, for example, Bronnenberg and Vanhonacker 1996; Sovinsky 2008; Draganska and Klapper 2011; Conlon and Mortimer 2013). A small literature circumvents this problem by collecting direct information on consideration sets. This is typically done via questionnaires in which participants state which products they considered, for example for hypothetical purchases in a virtual supermarket (Van Nierop et al. 2010), or coupled with actual purchase decisions, e.g. in the German coffee market (Draganska and Klapper 2011) or in the U.S. auto insurance industry (Honka 2014). In related work, Bruno and Vilcassim (2008) and Conlon and Mortimer (2013) combine sales data with information on the physical availability of products in supermarkets and vending machines, respectively.

[4]For a review of the literature on consideration sets see Roberts and Lattin (1997) or Malhotra, Peterson and Kleiser (1999).

A recent strand of the literature proposes methods for demand estimation under unobserved consideration sets that does not require the specification of the consideration process. Lu (2016) develops a moment inequalities approach which returns bounds on preference estimates. The basic idea of his approach is that if a consumer chooses product j, the true consideration set must be bounded by the largest and the smallest possible consideration set that contains j. Crawford, Griffith and Iaria (2017) develop an estimation approach for panel data which returns point estimates. It requires that past choices carry a sufficient amount of information about present-day consideration sets. The authors propose multiple scenarios of sufficient intertemporal correlation of consideration sets, e.g. once a consumer chooses a product, it remains in her consideration set for all subsequent periods.

We develop a novel approach that complements the existing methods. For given choice data, our method is able to pick the best-fitting consideration model from a finite set of competing consideration models. Our method follows the intuition of a so-called menu approach. This approach was developed to test for unobserved competitive conduct in an industry. Its basic idea is that it compares the equilibrium outcome in an industry to the theoretical predictions of a finite set (a "menu") of alternative models of competition, and uses a model selection test to identify the model which matches the data best. Our approach is most closely related to Villas-Boas (2007) who tests for different models of vertical relationships. Other notable examples include Bresnahan (1987), Gasmi, Laffont and Vuong (1992), Kadiyali, Vilcassim and Chintagunta (1996), and Nevo (2001).

Compared to the approach of Crawford, Griffith and Iaria (2017), our approach has less restrictive assumptions on the intertemporal correlation between consideration sets. Since it is able to test any two consideration models against one another, it is able to accommodate models with any intertemporal structure. Compared to the bounds approach of Lu (2016), our approach allows for point identification of demand parameters. However, the performance of our method critically relies on the quality of the menu of consideration models. This is because we cannot identify the true model but only the *best model within the menu*. Generally, our method is best for markets in which we can make reasonably good guesses of consideration models. When there is evidence on how consideration sets are linked across time, the approach of Crawford, Griffith and Iaria (2017) performs better. When a market is generally not well understood, the method of Lu (2016) is preferable because it makes the weakest assumptions.

4.3 Model

In this paper, we test two prominent demand models against each other: a single-stage mixed logit and a more flexible two-stage mixed logit. The former is the standard model in empirical industrial organization, the latter is popular in marketing research. In the following, we develop both models and derive the corresponding choice probabilities, and then describe how we test two consideration set models against one another.

4.3.1 Single-Stage Decision Process

We study a market in which each of R competing firms sells at least one of J substitute products. In every period t, consumer i chooses one product j from J partially differentiated competing products with respective prices $p_{1t}, ..., p_{Jt}$. The consumer obtains a utility equal to

$$U_{ijt} = \alpha_i p_{jt} + x_{jt}\beta + \varepsilon_{ijt}, \quad i = 0, \dots, I, \quad t = 1, \dots, T, \tag{4.1}$$

where x_{jt} is a K-dimensional vector of observed product characteristics and ε_{ijt} is a zero-mean, i.i.d. extreme-value I distributed individual-specific random shock.[5] The coefficient α_i is consumer i's marginal disutility of price and β is a K-dimensional vector of marginal utilities with respect to the K observed product characteristics. Consumers can choose not to buy any of the J products. Since the mean utility from the outside good is not identified, we normalize it to zero. The utility from this outside option is then

$$U_{i0t} = \alpha_i p_{0t} + x_{0t}\beta + \varepsilon_{i0t}. \tag{4.2}$$

In our estimation, we take household heterogeneity into account in multiple ways. Firstly, we let store choice depend on household travel distance. More specifically,

[5] This is a distributional assumption that since McFadden (1978) has become extremely popular in demand estimation because it provides closed-form solutions of the probabilities.

travel distance enters the product characteristics x_{jt} because it varies across choice options, depending on which of the R sellers offers it. There is rich evidence from the marketing literature that the probability of choosing a store is inversely related to distance from the consumer's home. In fact, travel distance has been found to be a major driver of store choice (e.g. Arnold, Oum and Tigert 1983; Smith 2004; Briesch, Chintagunta and Fox 2009). In contrast, the literature finds that other household characteristics like household income or household size do not significantly affect store choice (Leszczyc, Sinha and Timmermans 2000; Cleeren et al. 2010).

Secondly, we allow for heterogeneity in price sensitivity. Price sensitivity is modeled to contain a mean coefficient and a varying component which depends on observed household characteristics. The individual price coefficient α_i is distributed with density $f(\alpha|\phi)$, where ϕ collectively refers to the parameters of this distribution. We assume that

$$\alpha_i = \alpha_0 + d_i\xi + \sigma^\alpha \nu, \quad \nu \sim N(0,1), \tag{4.3}$$

where α_0 denotes the mean price response across all consumers, σ^α is the parameter of the random consumer-specific taste variation ν, and d_i is a vector of household characteristics. In particular, we allow d_i to include household income because both economic theory and the empirical literature suggest that it is a major – if not the most important – determinant of price sensitivity (e.g. Berry, Levinsohn and Pakes 1995; Hoch et al. 1995; Nevo 2001; Wakefield and Inman 2003). ξ captures how strongly observed household characteristics enter price sensitivity.

From the logit structure it follows that the probability L_{ijt} of consumer i choosing product j at time t conditional on the consumer-specific taste variation ν is

$$L_{ijt}(\nu) = \frac{\exp(V_{ijt}(\nu))}{1 + \sum_{k=1}^{J} \exp(V_{ikt}(\nu))}, \tag{4.4}$$

where $V_{ijt} = \alpha_i p_{jt} + x_{jt}\beta$.

When we have panel data, we observe a sequence of household decisions. The probability of a consumer making this sequence of decisions is the product of the probabilities across the T periods

$$L_i(\nu) = \prod_{t=1}^{T} \frac{\exp(V_{ijt}^{chosen}(\nu))}{1 + \sum_{k=1}^{J} \exp(V_{ikt}(\nu))}, \tag{4.5}$$

where V_{ijt}^{chosen} denotes the indirect utility from the alternative that was chosen by individual i in period t. The unconditional probability of observing the sequence of T choices corresponds to the integral over all possible values of ν:

$$P_i = \int \left(\prod_{t=1}^{T} \frac{\exp(V_{ijt}^{chosen}(\nu))}{1 + \sum_{k=1}^{J} \exp(V_{ikt}(\nu))} \right) \phi(\nu) d\nu. \tag{4.6}$$

Lastly, we maximize the log-likelihood $\sum_{i=1}^{N} \ln(P_i)$ with respect to the coefficients α_0, β, and σ^α.

4.3.2 Two-Stage Decision Process

The two-stage model is very similar to the single-stage model. However, each consumer i makes two consecutive choices per period. In the first stage, she chooses a time- and individual-specific consideration set θ from Θ possible consideration sets. Each consideration set θ contains a different subset of the J products in the market. Not all combinations of products have to be available. The utility from choosing consideration set θ is given by

$$U_{i\theta t} = X_{\theta t}\gamma + \eta_{i\theta t}, \quad i = 1, \ldots, I, \quad t = 1, \ldots, T, \quad \theta = 1, \ldots, \Theta, \tag{4.7}$$

where $X_{\theta t}$ is a vector of observed consideration set characteristics and $\eta_{i\theta t}$ is a vector of i.i.d. extreme-value I distributed shocks. We include travel distance in $X_{\theta t}$ because it varies across stores.

In the second stage, the consumer chooses a product j from her consideration set θ. The corresponding utility is

$$U_{ijt} = \alpha_i p_{jt} + x_{jt}\beta + \varepsilon_{ijt}, \quad j \in \theta, \quad i = 0, \ldots, I, \quad t = 1, \ldots, T, \tag{4.8}$$

where x_{jt} is a vector of K observed product characteristics and p_{jt} denotes the price of product j at time t. ε_{ijt} is a zero-mean, i.i.d. extreme-value I distributed individual-specific random shock. α_i is again distributed as specified in Equation 4.3.

Let L_{ijt} be the probability of consumer i choosing product j conditional on the random consumer-specific taste variation ν. Using Bayes' rule, L_{ijt} can be computed as $\sum_\theta L_{ijt|\theta} L_{i\theta t}$, where $L_{ijt|\theta}$ denotes the probability of choosing product j

conditional on having consideration set θ and $L_{i\theta t}$ denotes the probability of choosing consideration set θ. The two probabilities are given by

$$L_{i\theta t} = \frac{\exp(V_{i\theta t})}{1 + \sum_{l=1}^{\Theta} \exp(V_{ilt})} \tag{4.9}$$

and

$$L_{ijt|\theta}(\nu) = \frac{\exp(V_{ijt}(\nu))}{1 + \sum_{k=1}^{J_\theta} \exp(V_{ikt}(\nu))}, \tag{4.10}$$

where J_θ is the set of products included in θ. $L_{i\theta t}$ is equal to one under the standard assumption of consumers choosing from all products in the market. $L_{ijt|\theta}$ is zero if product j is not included in consideration set θ.

The unconditional probability of a consumer making the sequence of observed choices of considerations sets and products is then

$$P_i = \int \left(\prod_{t=1}^{T} \frac{\exp(V_{i\theta t})}{1 + \sum_{l=1}^{\Theta} \exp(V_{ilt})} \cdot \frac{\exp(V_{ijt}(\nu))}{1 + \sum_{k=1}^{J_\theta} \exp(V_{ikt}(\nu))} \right) \phi(\nu) d\nu. \tag{4.11}$$

Again, we maximize the log-likelihood $\sum_{i=1}^{N} \ln(P_i)$ with respect to the coefficients $\alpha_0, \sigma^\alpha, \beta$ and γ.

4.3.3 Testing

In this section, we describe how we test the two consideration set models against each other. Our test follows the idea of a so-called menu approach which is used to estimate the typically unobserved competitive conduct in an industry. The idea is to compare the equilibrium outcome in an industry to the theoretical predictions of a finite set (a "menu") of different models of competition, and then use a model selection test to identify the model which provides the best match with the observed market outcomes. This approach has been used for example to test for collusive vs. competitive behavior (Bresnahan 1987; Gasmi, Laffont and Vuong 1992; Nevo 2001), or for Stackelberg vs. Cournot competition (Kadiyali, Vilcassim and Chintagunta 1996).

In particular, our testing approach is closely related to Villas-Boas (2007) who tests for different models of vertical relationships between grocery retailers and yogurt manufacturers. For each vertical model, she uses consumer demand estimates to retrieve the corresponding set of marginal costs. She then regresses each set of implied marginal costs on input prices collected from supplemental data, and uses a non-nested selection test to identify the model with the best fit.

For any two competing consideration models, we estimate consumer demand. We then assume a model of seller price-setting behavior and recover a set of marginal costs implied by each demand model's estimates. Finally, we regress each set of marginal costs on marginal cost-shifters and use a non-nested model selection test à la Vuong (1989) to test the null hypothesis that two models perform equally well. The identification of the best model comes from the overidentifying restriction that the marginal costs have to be well-explained by supplemental data on input prices.

The main caveat of our testing procedure is that we have to make assumptions on the model of competition. This is not a problem in markets in which competitive conduct is well known, for example from previous research, reports by competition authorities, etc. However, when we know little about how firms compete in a market, we can only jointly identify the model of competition and consideration: If we have a set of candidate models of competition $A = \{1, 2, 3, ..., n_A\}$ and a set of candidate models of consideration $B = \{1, 2, 3, ..., n_B\}$, we have to test $n_A \cdot n_B$ model combinations to identify the best-fitting pair (a^{best}, b^{best}), where $a^{best} \in A$ and $b^{best} \in B$.

We now formally describe the testing procedure and detail how we recover marginal cost estimates. Each consideration model $z \in \{1, ..., Z\}$ returns a different $J \times 1$ vector of marginal cost estimates c^z. To compute these marginal costs, we need to assume a model of seller competition. In our application, we assume Bertrand-Nash competition (for a discussion of this assumption see Section 4.5.1). In the following, we set up and solve the maximization problem of the seller. We omit time subscripts t and model subscripts z because the problem is invariant across time and model.

Each seller r sets prices for all products in her assortment S_r which is a non-empty set of products. The seller obtains profits

$$\Pi_r(p) = \sum_{j \in S_r} (p_j - c_j) s_j(p), \tag{4.12}$$

where c_j is the marginal cost of selling product j, p is a vector of prices $(p_1, p_2, ..., p_J)$, and s_j is the market share of product j. The seller sets her prices such that she maximizes profit Π_r. The corresponding first-order condition is

$$s_j + \sum_{m \in S_r} (p_m - c_m) \frac{\partial s_m}{\partial p_j} = 0. \tag{4.13}$$

For notational simplicity, we switch to matrix notation in the following. Let T denote the $J \times J$ seller ownership matrix where element $T(j,k)$ is equal to 1 if products j and k are sold by the same firm and 0 otherwise. Let Δ be a $J \times J$-matrix of first derivatives of all market shares with respect to all prices, i.e. element $\Delta(j,k)$ is defined as $\partial s_k / \partial p_j$. Stacking up the first-order conditions for all products and rearranging terms, we obtain the $J \times 1$-vector of marginal costs

$$c = p + (T * \Delta) s(p), \tag{4.14}$$

where c is a $j \times 1$-vector of marginal costs, p is a $J \times 1$-vector of prices, $s(p)$ is a $J \times 1$-vector of market shares, and $*$ denotes element-wise matrix multiplication. The marginal cost is identified from the market shares $s(p)$ and the ownership matrix T which we observe in the data, and by the matrix Δ which we obtain from our demand estimates $(\alpha, \sigma^{\alpha}, \beta, \gamma)$. We repeat this procedure to recover c^z for each model $z = 1, ..., Z$.

The most accurate consideration set model z^{best} will return the most accurate set of marginal cost estimates c^{best}. To evaluate the "goodness" of a marginal cost vector and to identify the best model, we use external data on cost-shifters. Specifically, we regress each marginal cost vector c^z on a set of marginal cost-shifters

$$c^z = \xi \delta + \mu, \tag{4.15}$$

where ξ is a $J \times L$-matrix of cost-shifters, L is the number of different cost-shifters, δ is a $L \times 1$-vector of cost-shifter weights, and μ is a $J \times 1$-vector of mean-zero i.i.d. errors. The regression returns the $L \times 1$-vector of estimated parameters $\hat{\delta}$.

Lastly, we use a model selection test to identify the model with the best fit for the estimation of Equation (4.15). Specifically, we use the closeness test proposed by Vuong (1989). This test does not require any of the competing models to be correctly specified. Instead, it indicates which model is closest to the true data generation process. The Vuong test states that under the null hypothesis that two non-nested models 1 and 2 fit the true data generation process equally well, the log-likelihood ratio statistic LR asymptotically follows a normal distribution. The Vuong closeness test statistic for two competing models 1 and 2 is computed as

$$V(1,2) = \frac{LR_N(\hat{\delta}_1, \hat{\delta}_2)}{\sqrt{N}\omega_N} \longrightarrow N(0,1), \tag{4.16}$$

where

$$LR = L1_N(\hat{\delta}_1) - L2_N(\hat{\delta}_2) - \frac{K1 - K2}{2} \cdot log(N). \tag{4.17}$$

In equation (4.16), ω_N denotes the variance of LR and N denotes the sample size. $L1_N$ and $L2_N$ denote the likelihoods of the two models, and $K1$ and $K2$ are the numbers of estimated coefficients in model 1 and 2, respectively. In the final step, we compare the sample value of $V(1,2)$ with critical values of the standard normal distribution.

4.4 Data

We use German household scanner panel data provided by the market research company GfK. Our data cover all milk and coffee purchases of 1,251 German households in 2010. All households in our sample live in North Rhine-Westphalia, the most populous state of Germany. Each observation in our sample corresponds to one purchase of milk or coffee. We observe the date of the purchase, the retail chain, the paid price, the brand, the characteristics of the product and the sociodemographic characteristics of the household. In total, we observe 31,387 milk purchases and 4,240 coffee purchases.

The German supermarket landscape is characterized by a highly concentrated market structure. In the following, we focus on the seven largest chains which together capture almost 90% of the market (Bundeskartellamt 2013). Four of the seven biggest chains are full-line retailers, the rest are discounters. Discounters are popular, with a smaller store size and a narrower assortment, typically with bare-bones store designs and a large share of private labels. Table 4.1 shows that the discounters in our sample have systematically smaller category assortments: On average, a discounter (full-line retailer) carries 11 (50) milk and 22 (55) coffee varieties (column 4). In general, discounters carry more private labels than national brands.

Table 4.1: Summary Statistics: Retail Chains

Retailer	Format	Market share	#Products	$\frac{\#\text{National Brands}}{\#\text{Private Label}}$
		MILK		
Retail Chain 1	Discount	32.1%	6	0
Retail Chain 2	Full-Line	9.0%	59	4.36
Retail Chain 3	Full-Line	6.5%	38	3.75
Retail Chain 4	Discount	17.4%	8	0.6
Retail Chain 5	Discount	6,1%	19	0.27
Retail Chain 6	Full-Line	9.4%	50	1.94
Retail Chain 7	Full-Line	19.5%	51	2.4
		COFFEE		
Retail Chain 1	Discount	39.0%	16	0
Retail Chain 2	Full-Line	4.9%	57	not defined
Retail Chain 3	Full-Line	3.6%	40	not defined
Retail Chain 4	Discount	25.6%	27	1.75
Retail Chain 5	Discount	3.0%	24	2
Retail Chain 6	Full-Line	7.3%	54	not defined
Retail Chain 7	Full-Line	17.0%	67	13

Source: GfK

The five columns show retail chains, their formats, market shares, assortment size, and brand penetration, i.e. the number of national brands over the number of private labels. For four retail chains, brand penetration is not defined because the store does not carry private labels. For confidentiality reasons we cannot disclose the identity of the chains.

We define a product as a unique combination of characteristics. In the milk market, we define a product as a combination of retail chain, brand, a private label dummy, fat content, a UHT dummy, and an organic dummy. In the coffee market, we define it as a combination of retail chain, brand, a private label dummy, an organic dummy, a fair trade dummy, and a dummy for decaffeinated coffee. Table 4.2 shows descriptive statistics for the 50 best-selling products in the milk category and the 30 best-selling products in the coffee category.

Milk (see Panel I in Table 4.2) is typically sold in cardboard cartons of 1 liter and is almost always pasteurized, i.e. subjected to heating for a short time in order to increase its shelf-life. Different pasteurization procedures yield either fresh milk (with a market share of 64.8%) or ultra-high temperature (UHT) processed milk, the two of which differ in shelf life and taste.[6] Milk usually comes in two different fat levels: 1.5% (semi-skimmed) and 3.5% (full-fat), with roughly equal market shares. Organic milk is a niche market and makes up less than 3% of the total sales. The milk market is largely dominated by private label products: 95% of all milk is sold under a private label, national brands capture only a small share of the market. In particular, discounters sell none or very few national brands (see column 1 in Table 4.1). Promotions are rarely offered for milk: Only about 1% of all milk sales have promotional prices. In general, milk is a relatively cheap product with an average price of 53.6 euro cents/liter.

Panel II of Table 4.2 displays the summary statistics for the coffee market. Ground coffee is typically sold in vacuum-sealed packs of 500 grams. It is a storable[7] good with a relatively strong presence of national brands: About 45% of all sold products are branded, and three out of four full-line retailers do not carry any private label coffee (see column 5 in Table 4.1). Promotions are frequent and popular in the coffee category. In more than 30% of all cases, consumers purchased coffee that was on sale. Mild coffee varieties, i.e. varieties with a lighter roast, have a market share of 30.8%. Decaffeinated (5.7%), organic (1.9%), and fair trade (1.5%) varieties have small market shares.

[6]Heating milk for about 15 seconds up to 75 °C produces what is termed regular fresh milk. Heating milk for 1-4 seconds up to 135-150 °C yields so-called UHT milk.

[7]We are aware that consumers may stock products and that there could be an upward bias in the price coefficient. A dynamic stockpiling model (see for example Erdem, Imai and Keane 2003; Hendel and Nevo 2006; Lu 2017) is currently beyond the scope of this paper.

Table 4.2: Summary Statistics: Product Characteristics and Cost-Shifters

Variable	Mean	Std. Dev.	Min.	Max.
I. Milk				
Price (euro cents)	53.6	8.590	25	109
Private Label (0=brand, 1=private label)	0.950	0.217	0	1
Organic (0=conventional, 1=organic)	0.028	0.165	0	1
Fresh (0=UHT, 1=fresh)	0.648	0.478	0	1
Fat (%)	2.315	1.021	0.1	3.8
Promotion (1= yes, 0=no)	.010	.098	0	1
Retailer 1 (Discounter)	.349	.477	0	1
Retailer 2 (Full-Line)	.076	.265	0	1
Retailer 3 (Full-Line)	.061	.241	0	1
Retailer 4 (Discounter)	.189	.392	0	1
Retailer 5 (Discounter)	.061	.240	0	1
Retailer 6 (Discounter)	.080	.272	0	1
Retailer 7 (Full-Line)	.182	.386	0	1
Local Market Share	.160	.084	0	.40625
Number of Observations	31,387			
II. Coffee				
Price (euro cents)	346.92	102.312	119	999
Private Label (1=PL, 0=NB)	.552	.497	0	1
Fair (1=fair, 0=conventional)	0.015	0.123	0	1
Organic (1=organic, 0=conventional)	0.019	0.135	0	1
Decaf (1=decaffeinated, 0=caffeinated)	.057	.232	0	1
Mild (1=mild, 0=not mild)	.308	.462	0	1
Promotion (1= yes, 0=no)	.309	.462	0	1
Retailer 1 (Discounter)	.369	.482	0	1
Retailer 2 (Full-Line)	.079	.271	0	1
Retailer 3 (Full-Line)	.048	.214	0	1
Retailer 4 (Discounter)	.220	.414	0	1
Retailer 5 (Discounter)	.050	.218	0	1
Retailer 6 (Discounter)	.070	.256	0	1
Retailer 7 (Full-Line)	.164	.370	0	1
Local Market Share	.385	.092	.15	.65
Number of Observations	4240			
III. Marginal Cost-Shifters				
German Raw Milk Price (euro cents/liter)	31.004	2.568	27.95	34.65
Arabica Coffee (USD/kg, world market price)	4.320	.687	3.480	5.471
Robusta Coffee (USD/kg, world market price)	1.736	.207	1.483	2.074
Paper (index)	102.655	3.201	98.2	107
Diesel (index)	100.043	3.307	92.7	106.6
Electricity (index)	99.998	0.689	98.7	100.8
Labor Costs (index)	102.173	5.01	94.085	112.493

For the households in the panel we observe two key characteristics. Firstly, we observe the ZIP code of the household's home. Since Germany is divided into 28,683 post code areas, five-digit ZIP codes are a relatively precise measure of location. Secondly, we observe net monthly household income in brackets. We divide households into groups of low income (less than 1,751 euros per month), medium income (more than 1,750 euros but less than 2,751 euros per month), and high income (more than 2,750 euros per month) such that the groups are of roughly the same size (see Table C.1 in Appendix).

We construct a variable to capture how accessible a retail chain is to a household. To do so, we collect all supermarket locations from the German Yellow Pages (2010 edition). We then compute chain r's accessibility to household i as the number of r's outlets divided by the total number of retail outlets in a 10 km radius around household i's home.[8] Values of accessibility vary across households from 0, i.e. a chain not being in a household's shopping radius at all, to 0.406. No retailer is a local monopolist by being the only one to have outlets in the shopping radius of some households.

Finally, we add industry-wide data on marginal cost-shifters. They are provided by the German Federal Statistical Office. We use monthly price indices for the inputs raw milk, coffee beans of the two most popular species (*Coffea arabica* and *Coffea robusta*), paper, diesel, electricity, and labor (see Panel III of Table 4.2). Marginal cost data of a higher (lower) collection frequency result in a more (less) powerful model selection test.

[8]This radius is an approximation of how far consumers are willing to travel to do their shopping. We are aware that it neglects cases in which households do their shopping far from home, for example during travel or next to their work place. However, those cases are difficult to observe because linked store-consumer data is often not available. Consequently, the radius assumption has become widely popular and is used both in research (e.g. Villas-Boas 2007) and by antitrust authorities like the FTC (Ellickson, Grieco and Khvastunov 2016) and the German Cartel Office, e.g. `www.bundeskartellamt.de/SharedDocs/Entscheidung/DE/Entscheidungen/Fusionskontrolle/2010/B2-52-10.pdf` or `www.bundeskartellamt.de/SharedDocs/Entscheidung/DE/Entscheidungen/Fusionskontrolle/2015/B2-96-14.pdf` (both last accessed on 21 March 2017).

4.5 Estimation

4.5.1 Identification

In the following, we informally discuss identification. The main contribution of this paper is to identify the best-fitting model of consideration among a set of competing models. Each consideration model comes with a different set of estimated marginal costs. Identification of the best model comes from the fact that the marginal costs have to be well-explained by externally collected marginal cost-shifters, i.e. we use the cost-shifter data to construct an overidentifying restriction.

We have to make assumptions on the supply side in order to identify marginal costs. Our first assumption is that retailers compete in Bertrand-Nash fashion.[9] Indeed, German retail chains compete fiercely in prices; the press regularly refers to retail competition as a "price war".[10] Farmers frequently protest against downward pressure on wholesale prices[11] and brand manufacturers express concerns that low retail prices may harm brand reputation.[12] Prices are set simultaneously, documented by the fact that price changes typically occur on Mondays. The rare exception are special promotions, e.g. promotions valid only on the weekend.

Our second assumption is that retail chains know true consumer consideration sets. This is supported by large retailer investments in understanding consumer behavior. In Germany for example, more than 2.5 billion euros were spent in 2015 on market research alone.[13] In applications to markets in which retailers are less invested in market research, our assumption can be relaxed. For example, one could model retailers to observe true consideration sets with a measurement error.

[9] We do not model vertical relationships between retailers and suppliers. This is not a limitation to our estimation because wholesale prices – the result of vertical relationships – will be included in the marginal costs that we back out.

[10] 2010. "Preiskampf der Discounter geht weiter." *Frankfurter Allgemeine Zeitung*, 14 January. `www.faz.net/aktuell/wirtschaft/unternehmen/lebensmittel-einzelhandel-preiskampf-der-discounter-geht-weiter-1596161.html`. Last accessed on 16 March 2017.

[11] 2016. "Preiskampf zwischen Aldi und Lidl bedroht Bauern." *Focus*, 5 May. `www.focus.de/finanzen/news/milchpreis-im-freien-fall-billige-milchprodukte-gefaehrden-existenz-von-bauern_id_5503602.html`. Last accessed on 16 March 2017.

[12] 2015. "Unilever kritisiert Aldi, Lidl und Co." *Handelsblatt*, 20 July. `www.handelsblatt.com/unternehmen/handel-konsumgueter/preiskampf-im-einzelhandel-unilever-kritisiert-aldi-lidl-und-co-/12079182.html`. Last accessed on 16 March 2017.

[13] Statista: `de.statista.com/statistik/daten/studie/161551/umfrage/umsatz-der-marktforschungsinstitute-in-deutschland`. Last accessed on 16 March 2017.

Taste parameters are identified by variation in product characteristics (see Table 4.2). Store-fixed effects explain why consumers may choose a store which offers products at worse conditions than its competitor. The error term is individual-, time- and alternative-specific. It rationalizes why, on two different shopping trips, a consumer may choose differently even when all conditions remain exactly the same. The error term captures, among others, the momentary mood of the consumer, advertising exposure, and end-of-aisle displays.[14]

4.5.2 Estimation Technique

We estimate demand using a simulated maximum likelihood estimator (see appendix 4.9.4). We draw the price coefficient from a lognormal distribution. We do not specify an outside option; instead, demand is estimated conditional on purchase. We do so because milk and coffee are both important staple goods, and their consumption remains remarkably stable despite price variations (see Figures C.1 and C.2).

For both product markets, we test two models (A) and (B) against one another. Model (A) corresponds to the single-stage model of global consideration sets which is described in Section 4.3.1. Model (B) corresponds to a two-stage approach in which consumers first choose a supermarket and afterwards select a product from the chosen supermarket; it is described in Section 4.3.2.

In both models it is implicit that, if consumers consider a supermarket, they are aware of all products sold at that supermarket. This is not a necessary assumption. The model can easily be extended by an additional stage in which consumers choose a within-store consideration set. These consideration sets could be modeled as a function of marketing instruments (Sovinsky 2008; Draganska and Klapper 2011) or search costs (Seiler 2013). Importantly, our assumption does not affect the mechanism of our test; instead, it is able to test any two models against each other, regardless of their level of complexity.[15]

[14]We are aware that the error term may be correlated with the price. For example, marketing instruments such as advertising can increase both the price and the demand of a product. We run robustness checks in which we tackle endogeneity using the control function approach proposed by Petrin and Train (2010) (see Appendix 4.9.2) and find that model selection is not affected by it.

[15]In particular, the two models do not have to be nested in each other.

4.6 Results

4.6.1 Results in the Milk Category

Tables 4.3(a) and 4.3(b) present the estimation results in the milk category for the single-stage and the two-stage model. As expected, we find a negative price coefficient in both models. However, it is in absolute terms smaller in the single-stage model. The reason lies in the inflexibility of the single-stage model: When consideration sets are assumed to be global and consumer i does not react to a price change in product j, this is rationalized by consumer i having a low price sensitivity. Once we allow for a two-stage consideration process, a non-reaction can also be attributed to consumer i not including product j in her consideration set.

Our estimates show a substantial level of household heterogeneity: The standard deviation of the price coefficient is significantly different from zero, thus indicating that price sensitivity varies significantly across households. We find that, in both models, households with low or medium incomes are more price-sensitive than households with a high income. Also, distance plays a crucial role: Households are more likely to select a chain if they live close to its outlets.

Plugging the demand estimates into Equation (4.14), we recover marginal costs and retailer margins for both models (see Panel I of Table 4.4). The estimated median margin is 5.3 euro cents in the single-stage model and 4.2 euro cents in the two-stage model. The single-stage model yields a higher margin because it estimates a lower price sensitivity than the two-stage model, which in turn implies that retailers are more able to raise prices. Our estimates are of the same order of magnitude as those from industry reports.[16]

[16] `www.ife-ev.de/index.php/ife-publikationen`. Last accessed on 9 March 2017.

Table 4.3: Estimation Results: Milk

(a) Single-Stage Model

Mean		
Retailer 1	0.9224***	(0.0207)
Retailer 2	-0.0066***	(0.0345)
Retailer 3	-0.2666***	(0.0373)
Retailer 4	1.0070***	(0.0240)
Retailer 5	-1.0372***	(0.0321)
Retailer 6	0.2396	(0.0382)
Local Market Share	3.8725***	(0.1385)
Freshness	0.4802***	(0.0124)
Fat	2.4263***	(0.0799)
Private Label	-3.0270***	(0.1786)
Organic	4.7643***	(0.2944)
Price	-1.0303***	(0.0435)
× Income (Low)	-0.0056	(0.0044)
× Income (Medium)	-0.0170***	(0.0040)
Standard Deviation		
Price	3.4283	(2.3954)
No. of Households	1251	
No. of Choice Occasions	31387	

(b) Two-Stage Model

Mean		
Retailer 1	0.4255***	(0.0180)
Retailer 2	-0.7432***	(0.0250)
Retailer 3	-0.4873***	(0.0344)
Retailer 4	0.2214***	(0.0200)
Retailer 5	-0.7201***	(0.0299)
Retailer 6	-0.2537***	(0.0319)
Local Market Share	3.7837***	(0.1355)
Freshness	0.4306***	(0.0129)
Fat	1.9468***	(0.2300)
Private Label	-0.5588	(0.5263)
Organic	-0.4028	(0.8670)
Price	-1.7200***	(0.2531)
× Income (Low)	-0.1376***	(0.0037)
× Income (Medium)	-0.0865***	(0.0038)
Standard Deviation		
Price	3.5415***	(0.0595)
No. of Households	1251	
No. of Choice Occasions	31387	

Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5%, and 10% level, respectively.

Next, we regress estimated marginal costs on observed cost-shifters, i.e. prices of input factors (raw milk, diesel, electricity, labor, and paper), product characteristics (fat content and dummies for private label, organic, and fresh milk), and retailer- and product-dummies. Table C.3 shows the results. While no cost-shifter has a significant coefficient in the single-stage model, several cost-shifters are significant in the two-stage model, e.g. the retailer dummies and the price indices for paper, diesel, and electricity.

Table 4.4: Estimation Results: Marginal Costs of Milk and Coffee

Model	Marginal Cost in euro cents	Retailer Margin in euro cents	Price in euro cents
I. Milk			
Single-Stage Model	47.476	5.282	55.968
Two-Stage Model	47.861	4.217	55.968
II. Coffee			
Single-Stage Model	256.850	45.170	299
Two-Stage Model	267.920	36.670	299

We display the median values instead of the means because of outliers in the marginal cost estimates.

We perform a Vuong model selection test and find that the best-performing model is the model in which consumers consider only the products of the retailer they currently shop at: The two-stage model outperforms the single-stage model at a 1%-significance level (see Appendix 4.9.6). This is economically relevant: If we fail to allow for consideration sets, we obtain margins that are overestimated by 26.2%.

4.6.2 Results in the Coffee Category

Tables 4.5(a) and 4.5(b) show the demand estimates in the coffee category for the single-stage model and the two-stage model. Like in the results for the milk category – and following the same intuition – the price coefficient in the coffee category is in absolute terms larger for the two-stage model. Again, we find that households with low and medium incomes tend to be more price-sensitive. In both models, the standard deviation of the price is significantly different from zero, indicating substantial household heterogeneity in price sensitivity.

The average consumer prefers, *ceteris paribus*, regular coffee over decaffeinated coffee. The coefficient for mild roasts is insignificant in the single-stage model but significantly positive in the two-stage model. The coefficient for private labels is significantly negative in the single-stage model but insignificant in the two-stage model. Interestingly, unlike in the case of milk, the local availability of a retailer does not seem to affect coffee choice in either of the two models.

Panel II in Table 4.4 shows the estimated marginal costs and retailer margins. To offer coffee on its shelves, the average supermarket incurs a total cost of around 2.6 euros per pack. The median retailer margin is 45.2 euro cents in the single-stage model and 36.7 euro cents in the two-stage model. These estimated retailer margins are close to those from industry reports.[17]

We regress both sets of marginal cost estimates on the input prices (for Arabica and Robusta beans, diesel, electricity, labor, and paper), product characteristics (dummies for mild, decaffeinated, and private label coffee), and retailer- and product-dummies (see Table C.4 in Appendix). The two models have very similar R-squared statistics and yield similar coefficients, which already suggests that none of the models significantly outperforms the other. Finally, we compute the Vuong test statistic and find that, indeed, the single-stage model and the two-stage model do not perform significantly differently from each other in the market for ground coffee (see Appendix 4.9.6).

[17] 2013. "Brennpunkt Kaffee." *Brand Eins*, 27 May. `www.brandeins.de/fileadmin/redaktion/wissen/presse/2013_05_27_focus.pdf`. Last accessed on 9 March 2017.

Table 4.5: Estimation Results: Coffee

(a) Single-Stage Model

Mean		
Retailer 1	1.3245***	(0.0865)
Retailer 2	0.0842	(0.0759)
Retailer 3	-0.2004**	(0.1167)
Retailer 4	0.2703***	(0.0676)
Retailer 5	-0.3676***	(0.1215)
Retailer 6	-0.3479***	(0.0902)
Local Market Share	0.3666	(0.3866)
Decaffeinated	-0.5451***	(0.0689)
Mild Roast	0.0107	(0.0374)
Private Label	-0.1248**	(0.0675)
Price	-2.1214***	(0.1139)
× Low Income	-0.0078***	(0.0009)
× Medium Income	-0.0001	(0.0009)
Standard Deviation		
Price	2.2507***	0.1202
No. of Households	318	
No. of Choice Occasions	4240	

(b) Two-Stage Model

Mean		
Retailer 1	0.9536***	(0.0524)
Retailer 2	-0.9905***	(0.0744)
Retailer 3	-1.7386***	(0.1140)
Retailer 4	0.3574***	(0.0514)
Retailer 5	-1.9184***	(0.1129)
Retailer 6	-0.8502***	(0.0876)
Local Market Share	0.2229	(0.3666)
Decaffeinated	-0.3585***	(0.0728)
Mild Roast	0.0885**	(0.0391)
Private Label	0.0696	(0.0685)
Price	-3.1748***	(0.1528)
× Low Income	-0.0064	(0.0018)
× Medium Income	-0.0137	(0.0022)
Standard Deviation		
Price	2.2701***	(0.1167)
No. of Households	318	
No. of Choice Occasions	4240	

Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5%, and 10% level, respectively.

4.6.3 Discussion

In this section we discuss the discrepancy between the findings for coffee and the findings for milk: While the two-stage model outperforms the single-stage model in the milk category, it does not perform significantly better in the coffee category. This is because the two-stage model imposes a timing structure: Consumers first choose a store and then a product. This structure implicitly assumes that store choice is not affected by product choice. In the following we explain why this assumption is likely to be violated in the coffee market but appropriate in the milk market.

Firstly, coffee is subject to frequent price promotions which tend to be heavily advertised. At the same time, it is a relatively expensive grocery item, i.e. taking advantage of sales promotions can yield large absolute savings. As a result, consumers have an incentive to collect information about which supermarkets offer coffee promotions, and they may want to select supermarkets depending on their promotional coffee prices. Press reports support the notion that German consumers have a strong preference for bargain-hunting.[18] Secondly, coffee is a product category that is heavily differentiated, both vertically and horizontally. Previous literature has found that coffee is linked to strong brand loyalty on the part of consumers (Krishnamurthi and Raj 1988). This suggests that consumers may choose stores based on whether they carry the preferred coffee brands and varieties.

Unlike coffee, milk is a relatively cheap product with barely any price promotions; in our sample, only 1% of all milk is sold under a promotion. Advertising is similarly rare, mostly because of the dominance of private labels. National brands have a weak position in the milk market, as only 5% of all sales are branded products. In fact, the taste of milk sold by different manufacturers is virtually indistinguishable.[19] All of this suggests that the decision to buy milk is unlikely to affect supermarket choice, i.e. consumer behavior in this market is consistent with the two-stage model.

In general, consideration set formation is driven by both demand and supply conditions. On the demand side, consumer tastes determine whether they care enough about a product to make supermarket choice conditional on it. Consumers care about some product categories more than about others. This fact is closely

[18] Heidtmann, Jan. 2016. "Den Deutschen können Lebensmittel nicht billig genug sein." *Süddeutsche Zeitung*, 30 May. `www.sueddeutsche.de/wirtschaft/ernaehrung-den-deutschen-koennen-lebensmittel-nicht-billig-genug-sein-1.3012509`. Last accessed on 20 March 2017.

[19] For example, Joubert and Poalses (2012) find that perceived taste differences between milk brands can be explained by brand reputation and disappear in blind tests.

related to the marketing concept of hedonic and utilitarian products: While hedonic products provide emotional responses like excitement and pleasure, utilitarian goods are primarily functional (Dhar and Wertenbroch 2000). In the food context, coffee, wine or cheese tend to be hedonic goods whereas milk, flour, and salt fall into the utilitarian category. On the supply side, the price level as well as the frequency and advertising of promotions determine whether price-sensitive consumers select stores in order to take advantage of a promotion in the relevant category.[20]

4.7 Conclusion

Understanding consideration sets is important. For policy-makers, it is valuable in many applications; for example, antitrust authorities may reach very different conclusions about welfare implications depending on their assumptions about consideration sets (Sovinsky 2008; Conlon and Mortimer 2013). Also, a wide range of policies directly targets consideration sets and consequently affects consumer welfare, e.g. advertising bans (Honka, Hortaçsu and Vitorino 2017), regulation of choice in public services like health (Gaynor, Propper and Seiler 2016), or improved education of doctors and its effects on which treatment options they discuss with patients (Fiebig et al. 2015).

In order to infer typically unobservable consideration sets, we construct a test that can compare any two models of consideration and identify which model fits the data better. We use external data on marginal cost-shifters to construct overidentifying restrictions. Our approach has limited data requirements: Next to increasingly accessible household-level purchase data we require only widely available cost-shifter data. We illustrate our approach with an application to supermarket shopping and test two models against each other: a single-stage model with global consideration sets, and a two-stage model in which consumers first choose a store and then a product.

Our results suggest that the single-stage model performs better in hedonic product categories with strong product differentiation, high levels of brand loyalty, and frequent and well advertised promotions. On the other hand, the two-stage model

[20]To be precise, the supply conditions are themselves a result of the demand conditions: Firms optimize prices, promotions and advertising conditional on primitives of consumer demand, such as category-specific sensitivity to advertising and promotions.

tends to perform better in functional product categories with little product differentiation and few promotions. Importantly, our findings show that there is no "one-size-fits-all" model of consideration. Instead, researchers need to carefully tailor their demand models to the product markets they study.

4.8 Acknowledgement

I would like to thank Céline Bonnet, Federico Ciliberto, Anthony Dukes, Tomaso Duso, Jana Friedrichsen, Germain Gaudin, Justus Haucap, Florian Heiss, Ali Hortaçsu, Tobias Klein, Mia Lu, Roland Rathelot, Vincent Réquillart, André Romahn, Philipp Schmidt-Dengler, Stephan Seiler, Matthew Shi, Hannes Ullrich, Lars Zeigermann, and seminar participants at the Toulouse School of Economics, at the EARIE, EEA, and VfS Conferences, the University of Virginia, DICE, and DIW for helpful comments and suggestions. Financial support from the German Academic Exchange Service is gratefully acknowledged.

Bibliography

Allenby, Greg M., and James L. Ginter. 1995. "The Effects of In-Store Displays and Feature Advertising on Consideration Sets." *International Journal of Research in Marketing*, 12(1): 67–80.

Arnold, Stephen J., Tae H. Oum, and Douglas J. Tigert. 1983. "Determinant Attributes in Retail Patronage: Seasonal, Temporal, Regional, and International Comparisons." *Journal of Marketing Research*, 20(2): 149–157.

Barroso, Alicia, and Gerard Llobet. 2012. "Advertising and Consumer Awareness of New, Differentiated Products." *Journal of Marketing Research*, 49(6): 773–792.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, 25(2): 242–262.

Bettman, James R. 1979. *Information Processing Theory of Consumer Choice.* Reading, MA: Addison-Wesley Educational Publishers Inc.

Bonnet, Céline, and Zohra Bouamra-Mechemache. 2015. "Organic Label, Bargaining Power, and Profit-sharing in the French Fluid Milk Market." *American Journal of Agricultural Economics*, 98(1): 113–133.

Bresnahan, Timothy F. 1987. "Competition and Collusion in the American Automobile Industry: the 1955 Price War." *Journal of Industrial Economics*, 35(4): 457–482.

Briesch, Richard A., Pradeep K. Chintagunta, and Edward J. Fox. 2009. "How Does Assortment Affect Grocery Store Choice?" *Journal of Marketing Research*, 46(2): 176–189.

Bronnenberg, Bart J., and Wilfried R. Vanhonacker. 1996. "Limited Choice Sets, Local Price Response and Implied Measures of Price Competition." *Journal of Marketing Research*, 33(2): 163–173.

Bruno, Hernán A., and Naufel J. Vilcassim. 2008. "Research Note-Structural Demand Estimation with Varying Product Availability." *Marketing Science*, 27(6): 1126–1131.

Bundeskartellamt. 2013. "Sector Inquiry Into the Food Retail Sector."

Chiang, Jeongwen, Siddhartha Chib, and Chakravarthi Narasimhan. 1998. "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity." *Journal of Econometrics*, 89(1): 223–248.

Cleeren, Kathleen, Frank Verboven, Marnik G. Dekimpe, and Katrijn Gielens. 2010. "Intra- and Interformat Competition Among Discounters and Supermarkets." *Marketing Science*, 29(3): 456–473.

Conlon, Christopher T., and Julie Holland Mortimer. 2013. "Demand Estimation under Incomplete Product Availability." *American Economic Journal: Microeconomics*, 5(4): 1–30.

Crawford, Gregory S., Rachel Griffith, and Alessandro Iaria. 2017. "Estimating Demand Parameters with Choice Set Misspecification." *Mimeo.*

De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest. 2012. "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior." *American Economic Review*, 102(6): 2955–2980.

Dhar, Ravi, and Klaus Wertenbroch. 2000. "Consumer Choice Between Hedonic and Utilitarian Goods." *Journal of Marketing Research*, 37(1): 60–71.

Draganska, Michaela, and Daniel Klapper. 2011. "Choice Set Heterogeneity and the Role of Advertising: an Analysis with Micro and Macro Data." *Journal of Marketing Research*, 48(4): 653–669.

Draganska, Michaela, Daniel Klapper, and Sofia B. Villas-Boas. 2010. "A Larger Slice or a Larger Pie? an Empirical Investigation of Bargaining Power in the Distribution Channel." *Marketing Science*, 29(1): 57–74.

Ellickson, Paul B., Paul L.E. Grieco, and Oleksii Khvastunov. 2016. "Measuring Competition in Spatial Retail." Working Paper.

Erdem, Tülin, Susumu Imai, and Michael P. Keane. 2003. "Brand and Quantity Choice Dynamics Under Price Uncertainty." *Quantitative Marketing and Economics*, 1(1): 5–64.

Fiebig, Denzil G., Rosalie Viney, Stephanie Knox, Marion Haas, Deborah J. Street, Arne R. Hole, Edith Weisberg, and Deborah Bateson. 2015. "Consideration Sets and Their Role in Modelling Doctor Recommendations About Contraceptives." *Health Economics*, 26: 54–73.

Gasmi, Farid, Jean Jacques Laffont, and Quang Vuong. 1992. "Econometric Analysis of Collusive Behavior in a Soft-Drink Market." *Journal of Economics & Management Strategy*, 1(2): 277–311.

Gaynor, Martin, Carol Propper, and Stephan Seiler. 2016. "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service." *American Economic Review*, 106(11): 3521–3557.

Guadagni, Peter M., and John D. C. Little. 1983. "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science*, 2(3): 203–238.

Hauser, John R., and Birger Wernerfelt. 1990. "An Evaluation Cost Model of Consideration Sets." *Journal of Consumer Research*, 393–408.

Hendel, Igal, and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, 74(6): 1637–1673.

Hoch, Stephen J., Byung-Do Kim, Alan L. Montgomery, and Peter E. Rossi. 1995. "Determinants of Store-Level Price Elasticity." *Journal of Marketing Research*, 32(1): 17–29.

Holbrook, Morris B., and Elizabeth C. Hirschman. 1982. "The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun." *Journal of Consumer Research*, 9(2): 132–140.

Honka, Elisabeth. 2014. "Quantifying Search and Switching Costs in the US Auto Insurance Industry." *RAND Journal of Economics*, 45(4): 847–884.

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino. 2017. "Advertising, Consumer Awareness, and Choice: Evidence from the US Banking Industry." *RAND Journal of Economics*, forthcoming.

Howard, John A., and Jagdish N. Sheth. 1969. *The Theory of Buyer Behavior.* New York: Wiley.

Joubert, Johan P.R., and Jacolize Poalses. 2012. "What's in a Name? The Effect of a Brand Name on Consumers' Evaluation of Fresh Milk." *International Journal of Consumer Studies*, 36(4): 425–431.

Kadiyali, Vrinda, Nauel J. Vilcassim, and Pradeep K. Chintagunta. 1996. "Empirical Analysis of Competitive Product Line Pricing Decisions: Lead, Follow, or Move Together?" *Journal of Business*, 69(4): 459–487.

Krishnamurthi, Lakshman, and S.P. Raj. 1988. "A Model of Brand Choice and Purchase Quantity Price Sensitivities." *Marketing Science*, 7(1): 1–20.

Leszczyc, Peter T.L., Ashish Sinha, and Harry J.P. Timmermans. 2000. "Consumer Store Choice Dynamics: an Analysis of the Competitive Market Structure for Grocery Stores." *Journal of Retailing*, 76(3): 323–345.

Lu, Anna. 2017. "Consumer Stockpiling and Sales Promotions." Working Paper.

Lu, Zhentong. 2016. "A Moment Inequality Approach to Estimating Multinomial Choice Models with Unobserved Consideration Sets." Working Paper.

Malhotra, Naresh K., Mark Peterson, and Susan B. Kleiser. 1999. "Marketing Research: a State-Of-The-Art Review and Directions for the Twenty-First Century." *Journal of the Academy of Marketing Science*, 27(2): 160–183.

McFadden, Daniel. 1978. *Modelling the Choice of Residential Location.* Institute of Transportation Studies, University of California.

Mehta, Nitin, Surendra Rajiv, and Kannan Srinivasan. 2003. "Price Uncertainty and Consumer Search: a Structural Model of Consideration Set Formation." *Marketing Science*, 22(1): 58–84.

Mitra, Anusree. 1995. "Advertising and the Stability of Consideration Sets over Multiple Purchase Occasions." *International Journal of Research in Marketing*, 12(1): 81–94.

Nevo, Aviv. 2000. "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Economics & Management Strategy*, 9(4): 513–548.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-To-Eat Cereal Industry." *Econometrica*, 69(2): 307–342.

Petrin, Amil, and Kenneth Train. 2010. "A Control Function Approach to Endogeneity in Consumer Choice Models." *Journal of Marketing Research*, 47(1): 3–13.

Roberts, John H., and James M. Lattin. 1997. "Consideration: Review of Research and Prospects for Future Insights." *Journal of Marketing Research*, 34(3): 406–410.

Seiler, Stephan. 2013. "The Impact of Search Costs on Consumer Behavior: a Dynamic Approach." *Quantitative Marketing and Economics*, 11(2): 155–203.

Siddarth, S., Randolph E. Bucklin, and Donald G. Morrison. 1995. "Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data." *Journal of Marketing Research*, 32(3): 255–266.

Smith, Howard. 2004. "Supermarket Choice and Supermarket Competition in Market Equilibrium." *Review of Economic Studies*, 71(1): 235–263.

Sovinsky, Michelle. 2008. "Limited Information and Advertising in the US Personal Computer Industry." *Econometrica*, 76(5): 1017–1074.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation.* Cambridge University Press.

Van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel, and Philip H. Franses. 2010. "Retrieving Unobserved Consideration Sets from Household Panel Data." *Journal of Marketing Research*, 47(1): 63–74.

Villas-Boas, Sofia B. 2007. "Vertical Relationships Between Manufacturers and Retailers: Inference with Limited Data." *Review of Economic Studies*, 74(2): 625–652.

Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*, 57(2): 307–333.

Wakefield, Kirk L., and J. Jeffrey Inman. 2003. "Situational Price Sensitivity: the Role of Consumption Occasion, Social Context and Income." *Journal of Retailing*, 79(4): 199–212.

4.9 Appendix

4.9.1 Household Characteristics

Table C.1: Summary Statistics: Net Monthly Household Income

Income Group	No. Households	%
COFFEE		
1 (income <1750 euros)	1,781	36.08
2 (1750 euros ≤ income ≤ 2750 euros)	1,871	37.91
3 (2750 euros < income)	1,284	26.01
MILK		
1 (income <1750 euros)	9,186	29.27
2 (1750 euros ≤ income ≤ 2750 euros)	11,703	37.29
3 (2750 euros < income)	10,498	33.45

4.9.2 Control Function Approach

In this section, we describe our application of the control function approach proposed by Petrin and Train (2010). The key idea is that if we can derive a proxy variable that captures the part of the price that depends on the error term, then the remaining variation in the price will be independent of the error and thus allow standard estimation. In the first step, we use an ordinary least squares estimator to regress the potentially endogenous price on a number of instruments and exogenous variables:

$$p_{jt} = \delta J_{jt} + \gamma W_{jt} + \eta_{jt}. \tag{4.18}$$

J_{jt} and W_{jt} are vectors of product characteristics and cost-shifters, respectively. J_{jt} includes the fat content, a private label dummy, a fresh milk dummy, an organic dummy and retailer dummies. W_{jt} includes the price indices for raw milk, diesel, and electricity. η_{jt} is an i.i.d. mean-zero error term. Table C.2 displays the regression results which are all consistent with economic intuition.

In the second step, we obtain the residual from (4.18) and plug it into the utility function:

$$U_{ijt} = \alpha_i p_{jt} + x_{jt}\beta_i + \tau\hat{\eta}_{jt} + \bar{\varepsilon}_{ijt}, \tag{4.19}$$

Table C.2: Estimation Results: Control Function

Variable	Mean	Standard Error
German Raw Milk Price Index	0.217***	(0.0101)
Diesel Price Index	0.0177*	(0.00856)
Electricity Price Index	0.528***	(0.00762)
Private Label	-16.16***	(0.110)
Fresh Milk	-0.348***	(0.0452)
Fat Content (in %)	2.933***	(0.0211)
Organic	34.11***	(0.129)
Retailer 2	0.298***	(0.0893)
Retailer 3	0.0934	(0.0963)
Retailer 4	0.536***	(0.0629)
Retailer 5	-0.232**	(0.0773)
Retailer 6	0.00378	(0.0927)
Retailer 7	1.385***	(0.0895)
N	37799	
adj. R^2	0.995	

Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5%, and 10% significance level, respectively.

where ε_{ijt} equals $\bar{\varepsilon}_{ijt} + \tau\hat{\eta}_{jt}$ and is extreme-value I distributed. Equation 4.19 can now be estimated with standard methods. In our robustness checks we find that the coefficient of the control variable is statistically significant but economically irrelevant and does not affect our model selection results.

4.9.3 Grid-Search: Filling in Prices

In order to fill in price p_{jtig} of good j on day t paid by household i living in postcode g, we search for households in the same postcode that purchased the product on the same day. If we find such households, we replace p_{jtig} with the average price paid by households in the postcode area. If we do not find any such households, we increase the searched time period to a week and repeat the procedure. We gradually and alternatingly increase both the time period and the geographical area until we find a matching household that purchased the same product.

4.9.4 Simulated Maximum Likelihood

One complication of the mixed logit model is that there is no analytic solution to the integral in Equations 4.6 and 4.11. We approximate both equations via simulation. The simulated probability is:

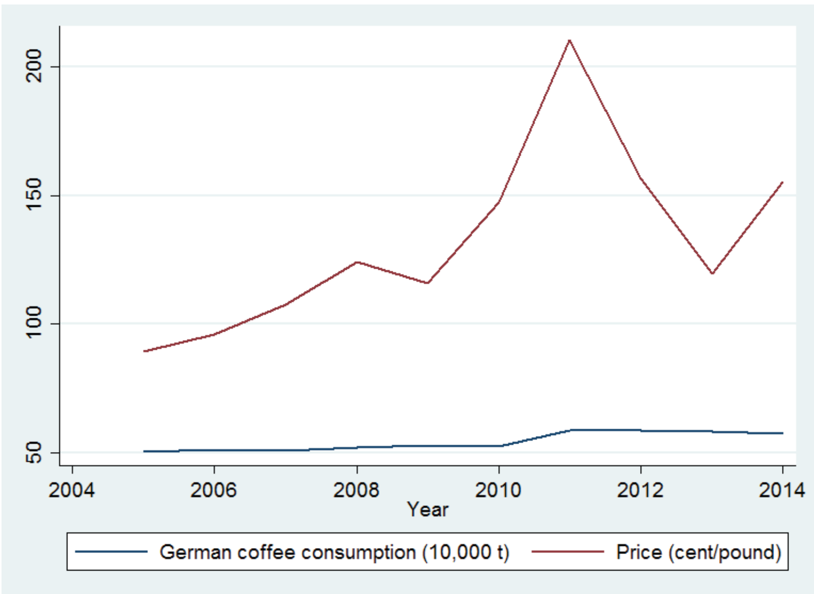
$$SP_i = \sum_{r=1}^{R} L_i(\nu^r), \tag{4.20}$$

where R is the number of simulations and ν^r is the r^{th} draw from the standard-normal distribution. We use Halton draws for faster convergence. The simulated log-likelihood function is

$$SLL = \sum_{i=1}^{N} \ln(SP_i). \tag{4.21}$$

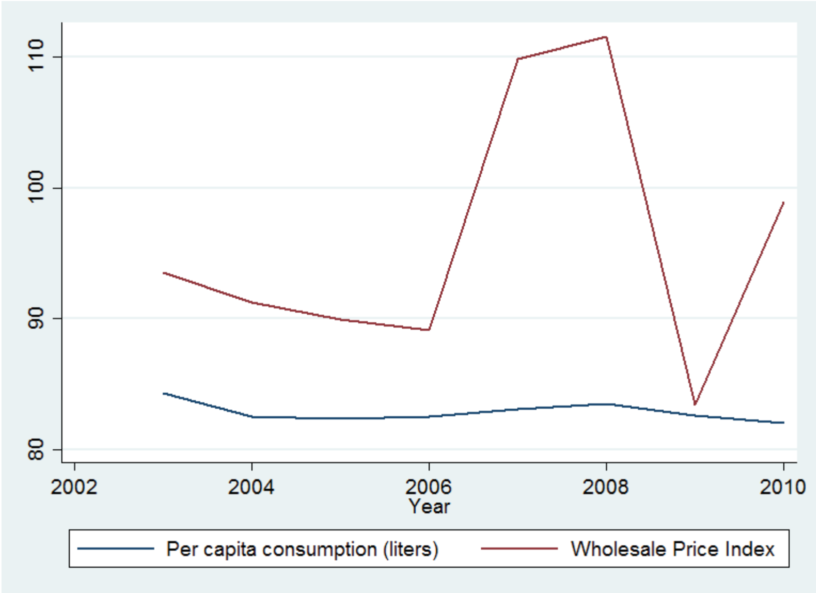
4.9.5 Coffee and Milk Consumption

Figure C.1: Coffee Prices and Coffee Consumption



The figure shows variation in the coffee world market price and in German coffee consumption. Data Source: International Coffee Organization, German Coffee Association.

Figure C.2: Milk Prices and Milk Consumption



The figure shows variation in the German wholesale price index for milk (base year = 2010) and in the annual German per capita milk consumption. Data Source: German Federal Statistical Office.

4.9.6 Model Selection Test

Table C.3: Regression Results of Marginal Costs on Cost-Shifters for Milk

Variables	Single-Stage Model	Two-Stage Model
German Raw Milk Price Index	-0.897	-575.8
	(10.12)	(2,175)
Diesel Price Index	-2.716	2,898***
	(3.153)	(677.9)
Electricity Price Index	4.052	-5,691***
	(8.446)	(1,816)
Labor Cost Index	0.937	-775.3
	(2.477)	(532.6)
Paper Price Index	2.146	-8,626***
	(14.78)	(3,177)
Fresh	3.651	6,026
	(31.19)	(6,707)
Fat Content	-1.375	1,251
	(10.20)	(2,193)
Private Label	-7.037	-1,369
	(32.95)	(7,085)
Organic	12.68	-8,802
	(53.43)	(11,489)
Retailer 2	-20.32	46,135***
	(50.54)	(10,868)
Retailer 3	-11.98	46,670***
	(46.47)	(9,992)
Retailer 4	-10.60	42,404***
	(47.48)	(10,208)
Retailer 5	-3.569	49,770***
	(49.29)	(10,598)
Retailer 6	-8.975	51,460***
	(65.54)	(14,092)
Retailer 7	-15.23	44,046***
	(56.98)	(12,252)
Constant	377.5	-69,068
	(996.7)	(214,307)
Observations	600	600
R-squared	0.153	0.167

Product dummies are not displayed due to their large number. Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5%, and 10% level, respectively.

Table C.4: Regression Results of Marginal Costs on Cost-Shifters for Coffee

Variables	Single-Stage Model	Two-Stage Model
Arabica Beans	-13.67	-15.01
	(16.43)	(16.40)
Robusta Beans	57.77	62.44*
	(36.36)	(36.30)
Diesel Price Index	-0.405	-0.239
	(0.957)	(0.956)
Electricity Price Index	-20.19*	-21.40**
	(10.50)	(10.48)
Labor Cost Index	0.474	0.593
	(0.541)	(0.540)
Paper Price Index	5.632	5.887
	(3.718)	(3.711)
Private Label	-54.66***	-57.51***
	(8.990)	(8.974)
Mild	100.8***	104.0***
	(8.990)	(8.974)
Decaffeinated	86.90***	88.06***
	(12.71)	(12.69)
Retailer 2	15.79	13.05
	(12.71)	(12.69)
Retailer 3	-50.18***	-51.23***
	(12.71)	(12.69)
Retailer 4	46.64***	52.43***
	(8.990)	(8.974)
Retailer 5 (omitted)	-	-
	-	-
Retailer 6	-17.60	-22.64*
	(12.71)	(12.69)
Retailer 7	-43.79***	-48.03***
	(12.71)	(12.69)
Constant	1,674**	1,740**
	(807.8)	(806.3)
Observations	360	360
R-squared	0.833	0.836

Product dummies are not displayed due to their large number. Standard errors are in parentheses. The symbols *, ** and *** denote significance at the 1%, 5%, and 10% level, respectively.

Table C.5: Vuong Test Statistic

Model	Vuong Test Statistic	Result
MILK		
$V(B, A)$	2.807	B $\succ$ A
COFFEE		
$V(B, A)$	0.011	B $\approx$A

(A) Homogeneous consideration sets. (B) Heterogeneous consideration sets. The test is carried out at a 1% significance level, with the corresponding χ^2-distributed comparison value being 2.326.

Chapter 5

Conclusion

This thesis consists of three essays in empirical industrial organization that study how strategic behavior of firms and consumers shapes market outcomes in the modern grocery retailing landscape. Methodologically, I blend a wide variety of techniques; some are at the forefront of empirical research, e.g. dynamic discrete-choice models of stockpiling, and some are widely established, e.g. reduced-form estimation. I also borrow methods from other fields, such as clustering algorithms from machine learning and cosine similarity measures from marketing.

The first essay studies variety competition between supermarkets and investigates how supermarkets reposition their assortments after a change in market concentration. I find evidence that supermarkets adjust not only assortment size but also assortment composition. In particular, I find that a series of local U.S. mergers in 2010 increased assortment overlap between supermarkets by 3%. This result suggests that merger control should consider not only mergers effects on assortment size but also on substitutability between stores.

The second essay studies strategic consumer stockpiling in storable goods markets and its implications for retailer pricing. In the U.S. market for laundry detergent, I find evidence of consumer stockpiling. This has important implications for retailers: They have to take forward-looking behavior into account when they design price promotions. Specifically, I study how promotion length and promotion depth affect revenues, and find that, in the detergent market, an increase in promotion depth is four times more effective in raising revenue than an increase in promotion length. Whether practitioners should increase promotion length or depth depends crucially on the demand conditions of the product market.

Unlike the first two essays, the third essay focuses more on the limitations to strategic decision-making. In particular, I study how consumers struggle to make choices from excessively large assortments and how they simplify the decision problem by limiting their attention to subsets of alternatives, i.e. consideration sets. I develop an approach to test for consumer consideration processes. I illustrate my test with an application to the German markets for milk and coffee, and find that consumer consideration differs fundamentally across product markets. This suggests that researchers need to carefully tailor their demand models to the markets they study.

In conclusion, this thesis provides an in-depth analysis of the grocery retailing industry. It delivers insights into the strategic decisions of firms and consumers, and their implications for policy-makers and industry practitioners. My work connects to a large body of literature that studies an industry that is fascinating in many ways – due to, among others, its immediate importance to consumers, a wide range of important policy questions, the availability of high-quality data sets and powerful estimation tools, and, last but not least, valuable research synergies with many other fields, such as marketing, operations research, and psychology.

Eidesstattliche Versicherung

Ich, Anna W. Lu, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig, und ohne unzulässige fremde Hilfe, unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist.

Berlin, 22. Mai 2017

Unterschrift